Studies in Theoretical Psycholinguistics 46

Lyn Frazier
Edward Gibson *Editors*

Explicit and Implicit Prosody in Sentence Processing

Studies in Honor of Janet Dean Fodor

Springer

Studies in Theoretical Psycholinguistics

Volume 46

The goal of this series is to bring evidence from many psychological domains to the classic questions of linguistic theory. The fundamental question from which the others flow is: What is the mental representation of grammar? Evidence from all aspects of language are relevant. How is the grammar acquired? How is language produced and comprehended? How is the grammer instantiated in the brain and how does language breakdown occur in cases of brain damage? How does second language acquisition and processing differ from first language acquisition and processing? A satisfactory theory of language calls for articulated connections or interfaces between grammar and other psychological domains. The series presents volumes that both develop theoretical proposals in each of these areas and present the empirical evidence needed to evaluate them.

More information about this series at http://www.springer.com/series/6555

Lyn Frazier • Edward Gibson
Editors

Explicit and Implicit Prosody in Sentence Processing

Studies in Honor of Janet Dean Fodor

Springer

Editors
Lyn Frazier
Department of Linguistics
University of Massachusetts
Amherst
MA
USA

Edward Gibson
Department of Psychology
Massachusetts Institute of Technology
Cambridge
MA
USA

ISSN 1873-0043
Studies in Theoretical Psycholinguistics
ISBN 978-3-319-12960-0 ISBN 978-3-319-12961-7 (eBook)
DOI 10.1007/978-3-319-12961-7

Library of Congress Control Number: 2015936311

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To Janet with love

Contents

Part II Implicit Prosody

Contributors

Markus Bader Goethe-University, Frankfurt a. M., Germany

Claire Beyssade Institut Jean Nicod, CNRS-ENS-EHESS, Paris, France

Jason Bishop Linguistics Program, City University of New York (College of Staten Island and The Graduate Center), New York, NY, USA

Mara Breen Department of Psychology and Education, Mount Holyoke College, South Hadley, MA, USA

Meredith Brown Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

Andrés Buxó-Lugo Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA

Katy Carlson Department of English, Morehead State University, Morehead, KY, USA

Charles Clifton Department of Psychology, University of Massachusetts, Amherst, MA, USA

Eva M. Fernández Linguistics and Communication Disorders, Queens College, City University of New York, Flushing, NY, USA

Graduate Center, City University of New York, New York, NY, USA

Fernanda Ferreira Department of Psychology, Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA

Caroline Féry Institut of Linguistics, Goethe University Frankfurt, Frankfurt am Main, Germany

Anouschka Foltz School of Linguistics and English, Bangor University Language, Bangor, Gwynedd, UK

Lyn Frazier Department of Linguistics, University of Massachusetts, Amherst, MA, USA

Edward Gibson Department of Brain and Cognitive Sciences, MIT, Cambridge, MA, USA

Barbara Hemforth Laboratoire de Linguistique Formelle (LLF), CNRS-Université Paris Diderot, Paris, France

Sun-Ah Jun Department of Linguistics, University of California Los Angeles, Los Angeles, CA, USA

Tom Juzek Faculty of Linguistics, Philology, and Phonetics, University of Oxford, Oxford, UK

Hossein Karimi Department of Psychology, Institute for Mind and Brain, University of South Carolina, Columbia, SC, USA

Chigusa Kurumada Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

Roger Levy Department of Linguistics, University of California, San Diego, CA, USA

Jean-Marie Marandin Laboratoire de Linguistique Formelle (LLF), CNRS-Université Paris Diderot, Paris, France

Robin Melnick Department of Linguistics, Stanford University, Stanford, CA, USA

Cristel Portes Laboratoire Parole et Langage (LPL), CNRS-Aix-Marseille Université, Aix-en-Provence, France

Christoph Scheepers Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK

Irina A. Sekerina Graduate Center, City University of New York, New York, NY, USA

College of Staten Island, City University of New York, Staten Island, NY, USA

Dominique C Simmons Department of Psychology, University of California Riverside, Riverside, CA, USA

Shari R. Speer Department of Linguistics, The Ohio State University, Columbus, OH, USA

Michael K. Tanenhaus Department of Brain and Cognitive Sciences, University of Rochester, Rochester, NY, USA

Thomas Wasow Department of Linguistics, Stanford University, Stanford, CA, USA

Duane G Watson Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA

Bo Yao School of Psychological Sciences, University of Manchester, Manchester, UK

Hanzhi Zhu Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA, USA

Introduction

Lyn Frazier and Edward Gibson

A group of prominent psycholinguists of various persuasions eagerly responded to a call for papers to honor Janet Dean Fodor, our beloved mentor, colleague, and friend. Those papers appear here. Janet has made numerous important contributions to the field of psycholinguistics. In the area of adult language processing, she is perhaps best known for her work on implicit prosody, but she has also contributed to our understanding of gap-filling, syntactic reanalysis, and cross-language processing (see Ferreira this volume for a fuller description). In the area of language acquisition, Janet has worked on parameter setting models, and the role of parsing in acquisition among other more specific topics. She was also the visionary behind the CUNY Conference on Human Sentence Processing, which meets annually and not necessarily in New York. It is difficult to exaggerate the role that this conference has had on organizing the field of psycholinguistics as a community involving linguists, psychologists, and computer scientists. It would not be an overstatement to claim that this community exists largely because of Janet, her vision, and her sustained efforts to make that vision a reality.

Most of the papers in this volume were first presented at a workshop held in Janet's honor in Amherst in May 2013. Taken jointly the papers present a glimpse of the state of the art concerning prosody, both explicit and implicit, in sentence processing. Only a few decades ago, it was unclear whether there were rules governing prosody. Lehiste's (1973) pioneering empirical work simply tried to ascertain whether various types of syntactic ambiguities could be prosodically disambiguated. By the time of Nespor and Vogel's (1986) seminal study of prosodic disambiguation, the notion of a hierarchy of prosodic phrases (the "prosodic hierarchy") was becoming established, in large part due to them, and in their book an explicit

L. Frazier (✉)
Department of Linguistics, University of Massachusetts, 01003 Amherst, MA, USA
e-mail: lyn@linguist.umass.edu

E. Gibson
Department of Brain and Cognitive Sciences, MIT, 43 Vassar Street, rm 3035, 02139 Cambridge, MA, USA
e-mail: egibson@mit.edu

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_1

phonological theory of prosodic phrasing was offered which actually predicted in advance which syntactic structures could be disambiguated prosodically. They tested the theory in Italian and the results strongly supported the theory. Today it is generally assumed that there are rules governing prosody though they are often stated in terms of violable phonological constraints, essentially with a default that the prosody reflects the syntax (Match constraints) when no higher ranked constraint countermands the syntactic constituency (see Féry this volume, Selkirk 2011).

With respect to the role of prosody in silent reading ("implicit prosody"), Janet proposed that prosodic preferences for equal-sized prosodic units, or for the default prosody of the language, might influence the syntactic analysis readers assign. Focusing on relative clause attachment ambiguities, the idea was that conditions favoring a prosodic boundary before the relative clause would favor high attachment, since the prosodic boundary serves to close off the current—lowest—phrase, thereby leaving higher attachment as the only option. Bader (1998, this volume) also argued for the existence of implicit prosody. He showed that readers assume function words are unstressed and this in turn can determine the preferred syntactic analysis of an ambiguous sentence. Before these proposals, there were various investigations of the role of phonology in silent reading (see Clifton, this volume), but they concerned the role of phonology in lexical access or in maintaining a sentence representation in memory long enough to permit further processing of it, such as drawing nonlinguistic inferences from it. What is different about the implicit prosody hypothesis is that the prosodic representation at the phrase and sentence level is hypothesized to play a causal role in syntactic processing.

The current volume examines grammatical issues (Féry, Bessayde, et al., Ferreira and Karimi), processing issues (most chapters), and brain representation (Yao and Scheepers) concerning explicit and implicit prosody. Féry (this volume) investigates extraposition in German and argues that it is constrained by prosody. Avoidance of an ungrammatical prosody, "prosodic monsters," forces extraposition to take place under predictable circumstances. Beyssade, Hemforth, Marandin and Portes (this volume) investigate the prosody of answers to broad focus and narrow focus ("partial") questions in French. They show that two prosodic properties, Nuclear Pitch Accent and an Initial Rise on the resolving/answering constituent are in play: both properties may be present in the answer to a narrow focus question, or only one of the properties may be present. They suggest that the Initial Rise marks a constituent with any of a number of discourse roles, including answering a narrow focus question, whereas the Nuclear Pitch Accent marks a discourse update, and as a consequence the two prosodic properties may co-occur. These studies illustrate just how closely connected prosody and intonation are to other systems of language, be it phonology and syntax, as in Féry's study, or pragmatics, as in Beyssade et al's chapter.

Carlson (this volume) examines focus in the semantic sense, i.e., where a focused constituent introduces semantic alternatives (Rooth 1992). She investigates whether focus conveyed by clefts and by other means such as a pitch accent behave alike in processing. In self-paced reading and auditory questionnaire studies of English, she shows that empirically they do behave similarly with respect to their effects on interpretation of ambiguous sentences. What is striking is that the particular

means used to convey the contrast does not seem to matter, e.g., in Carlson's study. This is reminiscent of Cutler and Fodor (1979), where comparable effects are found for prosodically conveyed and for semantically conveyed focus (though see also, Drenhaus et al. 2010 for evidence that the focus introduced by clefts, pitch accent, and *only* differ semantically.)

Ferreira and Karimi (this volume) highlight the important issue of how one analyzes empirical data concerning duration and pauses. Since the durational properties of an utterance reflect both the prosodic phrasing assigned to the utterance and any planning/performance pauses, it is essential to separate the two if we are to properly evaluate linguistic and psycholinguistic accounts of prosody. Ferreira and Karimi's own approach is to investigate individual differences in working memory, inhibitory control, and lexical difficulty. They show that individuals with less capacity are more likely to produce sentence internal breaks—at positions unexpected according to prosodic theory. (See Jun and Bishop, this volume, for additional discussion of individual differences.).

Watson et al. (this volume) are also concerned with understanding word durations in natural production. They evaluate whether the lengthening of a discourse-focused word is due to difficulties in phonological encoding, by comparing results of a task where participants produced pairs of words against the predictions of a simple recurrent network applied to the same task. They found that both the network and the experimental participants experienced the most errors for word pairs at word onset in cases where initial fragments overlapped in two words and at points of word nonoverlap. They therefore propose that word lengthening may be partly a result of the phonological encoding system needing processing time. They discuss these effects in part in terms of uniform information density (Aylett and Turk 2004; Levy and Jaeger 2006), whereby speakers lengthen and shorten words to facilitate robust communication with listeners.

Wasow et al. (this volume) report a corpus investigation of a previously understudied phenomenon in English that they call the "do-be construction" (DBC). In line with earlier work on optional "that" that provided support for uniform information density (Jaeger 2010; Wasow et al. 2011), they found that factors that contribute to the processing difficulty of a DBC sentence increased the probability of the use of optional "to." In addition, they found that "to," which is almost always unstressed, sometimes serves to prevent two stressed syllables from appearing adjacent to one another ("stress clash"; Liberman and Prince 1977). An important theoretical consequence of this work is that the prosodic effects on lexical selection favor the interactivist view over a serial, modularist view of the lexical-selection and phonological-encoding stages of language production. These results provide support for a view of moment-by-moment language production as being crucially guided by considerations of communicative optimality (Levy and Jaeger 2006; Jaeger 2010).

Tanenhaus et al. (this volume) are concerned with mapping prosody onto intentions. The relevant intentions vary with the context of an utterance (e.g., the speaker's goals) and the realization of prosodic contours varies across speakers, accents, and speech conditions. They propose that listeners map acoustic information

onto prosodic representations using (rational) probabilistic inference, in the form of generative models, which are updated on the fly based on the match between predictions and the input. They review some ongoing work, motivated by this framework, focusing on the "It looks like an X" construction, which, depending on the pitch contour and context, can be interpreted as "It looks like an X and it is" or "It looks like an X and it isn't." Using this construction, they show that pragmatic processing exhibits the pattern of adaptation effects that are expected if the mapping of speech onto intentions involves rational inference.

Turning to implicit prosody, Clifton (this volume) provides an elegant review of what's known about the role of phonology in silent reading. Knowledge of the mapping of orthography onto phonology appears to be important in skilled reading, and this knowledge is applied very early in the process of recognizing words in isolation. The same is true when one is reading sentences and texts, and the creation of a phonological representation of a text is a critical determinant of eye movement patterns during reading. Phonological representations beyond the level of the individual word, including prosodic representations, also seem to play an important role in guiding parsing and in integrating discourse-level information.

Breen (this volume) follows with a history of Fodor's implicit prosody hypothesis (Fodor 2002) and discusses a variety of studies which have demonstrated that implicit phrasing, accentuation, and rhythm play a role in syntactic parsing. Breen suggests that the field needs to explore more subtle aspects of implicit prosody, including its relationship to overt prosody, its interaction with other information sources, and how an implicit prosodic representation serves to assist a reader in understanding written language.

Bader (this volume) reports new studies showing that default assumptions about stress/accent play a role in implicit prosody. One involves a manipulation so that the reader places stress on the more distant potential head for an extraposed relative clausein German. These studies add to his earlier work showing that manipulating stress in the implicit prosody has effects on the syntactic analysis assigned. The work highlights some of the questions about how readers assign prosody during reading. Do they use a strategy of assigning whatever prosody/intonation they would assign to the sentence in the spoken language, or are there circumstances where readers assign some minimal prosody, e.g., postulating a prosodic phrase or pitch accent only where necessary, with the consequence that the absence of a boundary (triggered by the minimality assumption) might dictate a particular syntactic analysis?

Jun and Bishop (this volume) provide an overview of the work on implicit prosody and relative clause attachment. Janet Fodor's work stimulated a large range of work on the role of prosody in processing relative clauses with more than one potential attachment site (*The daughter of the colonel who was on the balcony…*, Cuetos and Mitchell 1988). Jun and Bishop note that much of the research targeting the explicit prosody assigned to this structure uses the method of having participants read sentences out loud in the laboratory. The results of this method may not be representative of natural speech. Instead Jun and Bishop introduce an implicit priming method where ambiguous target sentences are preceded by three sentences with a prosodic boundary in a particular location. With silent primes, using the

length of constituents to manipulate hypothesized prosodic boundaries, no effect was found on the interpretation of the ambiguous target, though participants with a low working memory span did show more high attachment than other participants (as in Swets et al. 2007). With overt primes, an effect of prosodic boundary location was observed. Interestingly, the effect was the opposite of that predicted: In the configuration NP1 NP2 RC, a prosodic boundary after NP2 resulted in MORE NP2 attachments, not fewer (as in the original implicit prosody hypotheses). The reason is that in the experimental materials, which had equivalent accents on the two NPs, the prosodic boundary after NP2 leads to the accent on NP2 being interpreted as the Nuclear Accent since it is the final accent in its prosodic phrase. Consequently, the accent is perceived as being stronger than the accent on NP2 and thus it attracts the relative clause, similar to what has been found in Focus attraction studies (Schafer et al. 1996).

Fernández and Sekerina (this volume) also address issues concerning how to verify the assumptions about explicit prosody that are invoked in studies of implicit prosody. They are primarily concerned with developing a new methodology that can be used with different populations. In their study, ambiguous questions of the form *What color is the < part > of the < shape > that has a < image > in the middle?* were answered verbally with respect to a visually present context, and eye movements were recorded. What varied was whether the image appeared in the part of the shape (biasing for high attachment of the relative clause) or in the shape (biasing for low attachment) or in both (ambiguous). The question itself could have a prosodic break after NP1 or after NP2, biasing toward high or low attachment, respectively. In the preliminary results, there was an advantage for a break after NP1, which was more strongly biasing than the break after NP2. The authors argue that what is special about their technique is that the questions involve a very shallow semantics (lack of the usual real world biases), and the task can be accomplished by any population of participants.

Speer and Foltz (this volume) investigate corrective contrast in overt and implicit prosody using a priming technique where a contrastive focus assumed for the implicit prosody of a mini-discourse involving a correction does or does not match an end of discourse probe spoken with or without an accent appropriate for a corrective contrast. When analyzed as a group, no matching effect was observed. But an analysis based on each participant's own pronunciation when reading the discourse aloud revealed that there was indeed a match between the participants own pronunciation and the accent on the probe. This is taken to reveal properties of the auditory image created during silent reading.

The final chapter reports a fascinating new line of investigation of reported speech.

It has been known for a long time that quoted speech has different linguistic properties than indirectly reported speech: *John said "I want you to come here" vs John said he wanted me to go there*. Yao and Scheepers (this volume) review evidence suggesting that voice-activated regions of the brain are active when silently reading quotations but not when reading indirectly reported speech. The evidence suggests that silent readers are supplying covert prosody for quotations,

likely involving auditory imagery. The authors then take up the difficult issue of how to think about the relation of this covert prosody and the implicit prosody familiar from studies of relative clause attachment, such as those discussed in many of the chapters of the present volume. Understanding the representational basis for prosody in quotation and its similarities to the prosodic representation in implicit prosody is likely to advance our understanding of both areas of investigation and it is sure to sharpen the questions that are asked.

Taken jointly, the chapters reflect the state of the art with respect to the effects of explicit and implicit prosody in sentence processing. There now exist highly developed analyses of the interaction of prosodic structure and other aspects of language, including syntax, semantics, discourse structure, and pragmatics. And, as the chapters demonstrate, current research on prosody now goes well beyond simply demonstrating that prosody plays some role in speech and reading comprehension. There are now detailed proposals, raising new questions in domains with rich empirical data. This research has also spurred the development of new methodologies and novel arguments, as illustrated throughout this volume. The existence of these new proposals together with current analysis techniques allows a range of languages to be investigated in the context of cross-language differences. Though the state of the art is currently one without clear answers to many central questions, interesting explicit hypotheses are now being formulated and tested using a wide array of complementary types of experimental techniques, arguments, and evidence.

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*(1), 31–56.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Fereirra (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht: Kluwer.

Cuetos, F., & Mitchell, D. (1988). "Cross linguistic differences in parsing" restrictions on the use of the late closure strategy in Spanish. *Cognition, 3,* 73–105.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7,* 49–59.

Drenhaus, H., Zimmermann, M., & Vasishth, S. (2010). Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics, 24*(3), 320–337.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East Linguistic Society 32.* GSLA, University of Massachusetts, Amherst.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23–62.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa, 7,* 102–122.

Levy, R. P., & Jaeger, T. F. (2006). Speakers optimize information density through syntactic reduction. *Proceedings of the 20th conference on neural information processing systems (NIPS)* (pp. 849–856).

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8,* 249–336.

Nespor, M., & Vogel, I. (1986). Prosodic structure above the word. In A. Cutler & R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 123–140). Berlin: Springer-Verlag.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1,* 75–116.

Schafer, A., Carter, J., Clifton, C., Jr., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes, 11,* 135–163.

Selkirk, E. O. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory* (2nd ed.). Oxford: Blackwell.

Swets, B., Demset, T., Hambrick, D., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136*(1), 64–81.

Wasow, T., Jaeger, T. F., & Orr, D. (2011). Lexical variation in relativizer frequency. In H. Simon & H. Wiese (Eds.), *Proceedings of the 2005 DGfS workshop "Expecting the unexpected: Exceptions in Grammar"* (pp. 175–195). Berlin: De Gruyter Mouton.

Part I
Explicit Prosody

Extraposition and Prosodic Monsters in German

Caroline Féry

Abstract In this chapter, the implications of extraposition for syntax–prosody interface are examined in a recursive theory of prosodic structure. It is shown that extraposition in German often improves the prosodic structure of a sentence. The prosodic grammar has its own rules and constraints, which can have an impact on syntax in the following way: If two syntactic structures are in competition for expressing the same content, and at the same time one of them is clearly preferred in terms of prosodic structure, the latter one is chosen. Only a theory allowing recursivity on a regular basis can reveal the formal influence of prosody on syntax. If entire syntactic constituents are parsed in entire prosodic constituents, a clause located in the middle field violates *Layeredness* and *Equal Sisters*. Such a constellation is called a "prosodic monster." In the case of prepositional phrases (PP) extraposition, recursion of prosodic domains is avoided, but no prosodic monster is at play. Extraposition is not always available: it is blocked by an accented constituent intervening between the antecedent or reconstructed position and the extraposed constituent. In the last part of the chapter, an optimality-theoretic approach is proposed that accounts for extraposition as a prosody-driven operation.

Keywords Syntax–prosody interface · Extraposition · Optimality theory · Recursivity in prosodic structure

1 Introduction

This chapter explores Fodor's insight that prosody plays a crucial role in language processing. It is assumed here that the role of prosody in processing reflects its role in grammar. It focuses on extraposition in German, which presents a clear application of this insight. In a version of grammar inherited from the T-model of grammar (see, for instance, Chomsky 1981), phonology cannot influence syntax. According to this model, the interpretation of a sentence should derive from the lexicon and the syntactic structure, and prosody should only redundantly interpret the established

C. Féry (✉)
Institute of Linguistics, Goethe University Frankfurt, Frankfurt am Main, Germany
e-mail: caroline.fery@gmail.com

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_2

meaning. Chomsky and Halle (1968) initiated a line of research in which morphosyntactic constituents are mapped into prosodic domains in a cyclic way. Lexical phonology (Kiparsky 1982) distinguishes between lexical and postlexical phonology, and the creation of phonological domains proceeds from small to large. Within each cycle, morphosyntax affects phonology but not vice versa. Nearly all models investigating the syntax–prosody interface have done so by choosing a specific syntactic structure and showing how prosody is mapped to it, avoiding in this way the question of how prosody can shape syntax. Early syntax–prosody-mapping models like Nespor and Vogel's (1986) relation-based account also allow only one direction of mapping, and the so-called readjustment rules or rhythmic rules are strongly limited by syntax. Similarly, edge-based models (Selkirk 1986) do not allow a symmetric interaction between syntax and prosody. Prosodic domains are created from syntactic inputs. In accounts using syntax–phonology-mapping constraints like *Align* and *Wrap* (Truckenbrodt 1995a), readjustment and variation are not easy to handle; as a consequence of the evaluation, there is one optimal candidate that cannot be changed. Information structure is shown to play a role, but mostly in respecting the prosodic domains created by syntax: Focus and givenness can only delete existing prosodic domains or create additional ones. As a result, the unique function of prosody is to represent and interpret sentence structure. If this view is correct, it is unexpected that prosody may influence one or the other reading of an ambiguous sentence or that it could influence syntax at all.

In important studies on syntactic parsing in reading, Fodor (1998, 2002a, b) refutes the view that prosody is limited to interpretation of the syntax, even in silent reading. She discusses concrete examples showing how "implicit" prosody affects syntactic decisions. In the implicit prosody hypothesis (IPH), a reader projects a prosodic structure onto what is read silently. This hypothesis claims that the projected prosodic structure may affect the interpretation of a sentence. Fodor (2002b) gives the following example: "A reader may create a boundary for one reason (e.g., optimal phrase length), but the boundary may be understood as present for another reason (e.g., alignment with syntax). Under the latter construal, the prosodic break can be relevant to syntactic structure assignment: it can bias the resolution of a syntactic ambiguity just as a prosodic break in a spoken sentence does." An area of application of this hypothesis concerns ambiguous attachment of relative clauses, as for example in the sentence *Mary met the friend of the actress who was drinking tea,* where the relative clause can be attached to *friend* (high attachment) or to *actress* (low attachment). A prosodic break between *actress* and the relative clause increases the probability of high attachment. If a specific language assigns a left boundary at the beginning of a relative clause for reasons other than for disambiguation, then the preference will be for high attachment in general. This is because in many languages, the presence of a prosodic break in this position correlates with a high attachment preference.

Fodor and Nickels (2011) examine cases of "heavily nested" syntactic structure in two center-embedded relative clauses, like *The elegant woman that the man I love met lives in Barcelona*. They propose that such sentences can be adjusted to create a flat structure for prosody. Where phrase length cooperates with syntactic

alignment, no mismatch takes place, and comprehension is facilitated. This is what they call "productive interaction between syntax and prosody online." Problems appear when phrase lengths induce a prosodic structure that mismatches the syntactic structure. Fodor (2002b) suggests that the *AlignR XP* constraint in English (Selkirk 2000) is an instance of a more general right-alignment phenomenon sensitive to the number of right-edge syntactic brackets between adjacent words. She interprets this constraint as a graded constraint that reflects the configurational relations in the syntactic tree: "the pressure to insert a prosodic break (and perhaps the intensity of the acoustic realization of the break) is greater where the structural discontinuity in the tree is greater (i.e., more right brackets together)."

Some aspects of Fodor's IPH are straightforwardly adopted in the remainder of the chapter. Additionally, it will be shown that syntax can be modeled by prosody, in the same way as prosody is modeled by syntax. There is thus a shift in perspective between Fodor's main interest and the point of view of the major part of the literature concerning the role of prosody in grammar on the one hand, and the role of prosody in the elaboration of syntactic structures as it is examined in the present chapter on the other hand. It will be shown with extraposition in German that prosody not only has an effect in the processing of sentences, ambiguous or not, but that it also influences syntax in production. If a constituent may be optionally extraposed, prosody is often the motor behind the decision to extrapose. Not extraposed ("in-situ" or "intraposed") embedded clauses create prosodic structures that mismatch the syntactic structure, such as those described by Fodor. Extraposition is applied to avoid such mismatches, in which case prosody acts as a facilitating factor for a syntactic operation. This can be compared on the one hand to the example discussed by Fodor in which the presence of a prosodic boundary before a relative clause facilitates high attachment (a syntactic structure), and on the other hand with Fodor and Nickels' center-embedded examples which may cause disruption between syntax and prosody. In both cases, the interface between syntax and the formation of prosodic domains has a role to play and this interface between syntax and prosody may be facilitated or disrupted.

Despite extensive evidence to the contrary (see, for instance, Ladd 1990; Ishihara 2003; Féry 2011), non-recursivity has been a guiding theme in mainstream prosody research. Due to the fact that prosody is realized in real time and that the speech stream cannot easily represent hierarchical structure, it has been assumed that prosodic structures cannot be recursive. This assumption is a consequence of the fact that most of the data considered for the creation of syntax–prosody interaction are structurally very simple. Once it is recognized that recursion is a feature of prosody, the similarity between recursion in syntax and in prosody becomes obvious and possible interactions between the two can no longer be denied; see, for instance, Kentner (2012) and Kentner and Féry (2013) for subtle interactions between syntax and prosody.

The present chapter is dedicated to the role of prosody in extraposition. Studies investigating the choice between extraposition and in-situ position in German have been heavily influenced by the work of Hawkins (1994), who shows that in English the distance between the head of a relative clause and the relative clause

itself plays a more important role than the length of the relative clause.[1] This result was reproduced for German by Uszkoreit et al. (1998), who verify Hawkins' locality-based prediction by analyzing relative clauses in two written corpora. They demonstrate that the probability that the relative clause is in-situ increases when the distance between the head and the relative clause increases. Uszkoreit et al. (1998) and Konieczny (2000) show that speakers nevertheless prefer in-situ relative clauses, even when extraposition only crosses one word (a participle). This result may be due to the fact that perception of the sentences investigated in the form of spoken speech was not involved. Speakers had to judge written sentences, and normative factors may have played a role. Once spoken data are involved, extraposed relative clauses are often judged better than non-extraposed ones (see Poschmann and Wagner 2014).

The prosodic theory developed in this chapter locates itself in approaches seeking to replace performance accounts based on length by more detailed models that allow different grammatical factors to figure into the preference for extraposition over in-situ position. Further factors, not investigated in detail here, are information structure, the difference between restrictive and nonrestrictive relative clauses, more precisely the question of how the relative clause is related to the at-issueness of the main clause, and the syntactic relation between the main verb and the head of the relative clause.

In Section 2, it is shown that clause extraposition may be (partly) interpreted as a prosody-driven syntactic effect repairing a less than perfect syntax–prosody interface. In the version of the syntax–prosody interface used in the present chapter, that is, recursively embedded prosodic domains corresponding one-to-one to syntactic constituents, an in-situ clause triggers a prosodic structure in which an intonation phrase (ι-phrase) is embedded into a lower prosodic constituent, a prosodic phrase (Φ-phrase). The result is an ill-formed prosodic structure called a "prosodic monster." One way of resolving the problematic structure explored in this chapter is to extrapose the embedded clause. However, if the prosodic structure of a sentence with an in-situ clause does not contain a prosodic monster, there is no pressure to extrapose the clause, or the pressure decreases. This happens when the final portion of the main clause, located after the embedded clause, is heavy enough to form a Φ-phrase all by itself. A further factor acting on the decision whether to extrapose or not is the need to keep an embedded clause adjacent to its antecedent. This applies to relative clauses, or to complement clauses with a nominal antecedent, but not to complement clauses, which can be located before or after the verb: they are adjacent to the verb in both cases.

Section 3 examines extraposition of prepositional phrases (PP), an optional operation. When the PP is a possessive attributive or an argument, *Non-Recursivity,* another well-formedness constraint on the prosodic structure, is violated in the case

[1] I do not dwell on proposals for English, since extraposition in German is truly different from extraposition in English, due to the verb-final properties of German. Uszkoreit et al. (1998) observe that most German extraposed relative clauses are separated from their antecedent by the verb only. In English, extraposition usually crosses an adverb, like *yesterday*. The kinds of constituents that can be extraposed also differ in the two languages.

of in-situ location of the PP, causing a mild pressure to extrapose. The pressure to extrapose is even milder when the PP is an adjunct.

If extraposition delivers better prosodic patterns than in-situ position, this option should be allowed on a principled basis. But this is not what is observed. In many cases, extraposition produces a structure that is less acceptable than the in-situ one. In Section 4, the limit of extraposition is addressed. It is shown that an accented noun intervening between a relative clause and its antecedent or between a PP and its reconstructed position heavily degrades the structure. The data discussed in Section 4 demonstrate that prosody can also have a blocking influence on a syntactic operation. If extraposition renders parsing more difficult than non-extraposition, or if its application degrades the prosodic structure, extraposition does not apply.

Section 5 returns to the original question, namely whether prosody merely serves an interpretative function or whether it can generate structure independently. In a first step, it is shown that purely syntactic accounts, which assume either movement from a preverbal underlying position, or base generation in the postverbal position in all cases, are largely inconclusive. A plausible alternative approach allows relative clauses and complement clauses to be generated in different positions in the sentence, in which case several options as to the linearization of constituents may be considered as equivalent from the point of view of syntax. In a second step, an optimality-theoretic (OT) account is proposed: *Match* constraints regulate the syntax–prosody interface, and a number of well-formedness constraints further act on the prosodic structure. In short, prosody plays an important role in grammar and is integrated as an active component of grammar.

The language investigated is German, because of word order issues that render extraposition particularly productive and interesting in this language. Some of the generalizations are relevant for English grammar, too, but others do not hold in English, see footnote 1.

2 Extraposition of Clauses and Prosody

Before showing the prosodic role of extraposition, it is important to make a strict distinction between three kinds of postfield positions in German, because extraposition is only one of them. Altmann (1981); Averintseva-Klisch (2006) and Ott and de Vries (to appear) distinguish between extraposition, right dislocation, and afterthought, in German and in other Germanic languages, and show how they differ in their syntactic and prosodic properties. Of the three constructions, only extraposition is described by these authors as being a true constituent of the main clause. It is intonationally integrated into its host sentence, i.e., it continues the tone movement of the host sentence. Neither right dislocation nor afterthought is part of the intonation contour of the main clause. Both of them build a separate prosodic unit (optionally separated from the clause by a pause). A right-dislocated constituent may have a clause-like accent of its own, or not, and it often triggers clitic-doubling. An afterthought always has an accent of its own. A further important difference

between extraposition and the other two constructions is that in the latter cases an adverb like *nämlich* "namely," *also* "well," or *ich meine* "I mean" can be inserted after the main clause, whereas this insertion is not possible in extraposition. In the following, we are only concerned with extraposition. The constituents that can be extraposed are extremely limited: Clausal complementizer phrases (CP) and PP can easily be extraposed, but nominal, adjectival, and verbal phrases (DPs, APs, and VPs) cannot, or only exceptionally.[2]

The examples in (1) show that a sentence containing a *dass*-complement clause is much more acceptable when the complement clause is postverbal, as in (1a), than when it is in-situ, as in (1b). An in-situ clausal complement is often heavily degraded as compared to its extraposed version (but see Sternefeld 2008 and Section 5 for examples of in-situ complement sentences that are acceptable).

(1) a. Sie hat niemandem erzählt, dass sie an dem Tag spät nach Hause kam.
she has nobody told that she on that day late to home came
"She didn't tell anybody that she came home late on that day."
b. *?Sie hat niemandem, dass sie spät nach Hause kam, erzählt.

In (3), the same sentences as in (1) are provided with prosodic structure, assuming that syntactic and prosodic constituents are subject to a strict one-to-one mapping, as proposed by Féry (2011) for German. In the following, the *Match* constraints proposed by Selkirk (2011) are used for demonstrating the prosodic properties of clause extraposition. These constraints are used because of their simplicity and straightforwardness. The *Match* constraints are formulated in (2). They assume that a grammatical word, a syntactic phrase and a clause roughly correspond to the three higher prosodic constituents, prosodic word (ω), prosodic phrase (Φ), and intonation-phrase (ι), respectively.

(2) Match Constraints (Selkirk 2011, p. 439)
a. *Match Clause*
A clause in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ι, in phonological representation.
b. *Match Phrase*
A phrase in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it Φ, in phonological representation.
c. *Match Word*
A word in syntactic constituent structure must be matched by a corresponding prosodic constituent, call it ω, in phonological representation.

Match Phrase requires a constituent formed by a predicate and its arguments (the VP) to be phrased in a prosodic phrase (Φ-phrase). However, in (3b), this Φ-phrase partly consists of the complement clause, itself an intonation phrase (ι-phrase) by

[2] An example of an exceptional DP extraposition appears in (12) below.

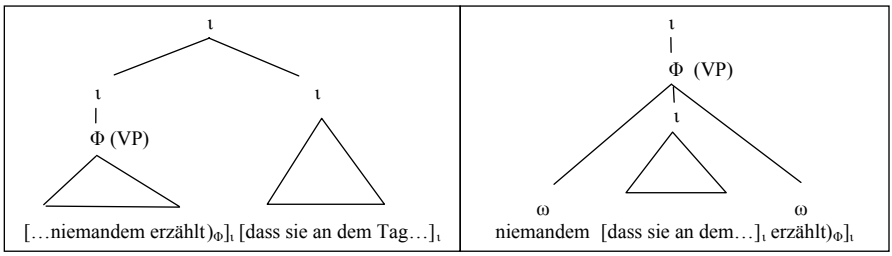


Fig. 1 Extraposition as avoidance of a prosodic monster (*dass*-complement)

virtue of *Match Clause*.[3] As a result, a hierarchically higher prosodic constituent is embedded in and dominated by a lower level constituent. It should be noticed that the function words *sie* and *hat* are too light to form their own Φ-phrase and are included in the adjacent Φ-phrase.

(3) a. [(Sie hat niemandem t_i erzählt)$_\Phi$]$_\iota$ [dass sie an dem Tag spät nach Hause kam]$_\iota$.
she has nobody told that she on that day late to home came
"She didn't tell anybody that she came home late on that day."
b. *?[((Sie hat niemandem)$_\Phi$, [dass sie spät nach Hause kam]$_\iota$ erzählt$_\omega$)$_\Phi$]$_\iota$.

Figure 1 illustrates how the prosodic structure favors extraposition of complement clauses: When the *dass*-clause is extraposed, as in (3a) and Fig. 1 left panel, the main clause and the embedded clause each form their own ι-phrase. They project ι-phrases at the higher level of the hierarchy by virtue of being clauses.[4] The sequence of two ι-phrases itself forms a larger recursive ι-phrase. However, when the complement clause is in-situ as in (3b) and the right panel of Fig. 1, the Φ-phrase formed on the VP *niemandem erzählt* "told nobody" is interrupted by the ι-phrase formed by the complement clause. The verb does not form a Φ-phrase by itself; it is only a ω-word. In this case, besides the Φ-phrase on the object, the Φ-phrase mapped to the VP dominates a ω-word and an ι-phrase. In her paper on extraposition in German, Hartmann (2013) shows with numerous naturally occurring examples that sentences like (3b) are avoided in German. She assumes that a final single ω-word cannot be parsed into the preceding prosodic constituent, and that it does not form a Φ-phrase all by itself.[5] These assumptions are taken for granted here. In the present proposal, the verb is parsed into a larger Φ-phrase. The ungrammaticality is a result of the prosodic imbalance between the prosodic constituents and the way they are layered. In particular, a constraint called LAYEREDNESS (from Selkirk

[3] Selkirk (2011, p. 453) makes a distinction between *Match* (illocutionary clause, ι) and the more general *Match* (clause, ι). In the following, the distinction between the two doesn't play any role and is ignored in the remainder of the chapter.

[4] This differs from many accounts in the literature in which the apprehension of prosodic constituents is guided by the phonetic cues associated with them (see Schubö 2010; Elfner 2012; Myrberg 2013, etc.).

[5] However, it is not clear in her approach why the same structure does not lead to ungrammaticality in the case of relative clauses.

1996), prohibiting a category of a certain level to dominate a higher category, is violated. Moreover, an additional constraint called EqualSisters (from Myrberg 2013) is also violated in this configuration. EqualSisters requires that the prosodic constituents dominated by a higher constituent are at the same level; see Section 5 for formal definitions and further illustrations.

The examples in (4) show that a relative clause can also appear in-situ, i.e., right after its antecedent, or be extraposed, in which case it is postverbal. Both the extraposed and the in-situ locations are felicitous in German, even though the in-situ version (4b) is degraded as compared to the extraposed variant (4a). Due to *Match Phrase*, the object of the main verb and the relative clause form an additional Φ-phrase by virtue of being a DP, albeit a complex one.

(4) a. $[[(\text{Sie hat ihre Mutter getroffen})_{\Phi}]_{\iota}\ [\text{die an dem Tag mit Freunden unterwegs war}]_{\iota}]_{\iota}$
she has her mother met who on that day with friends out was
"She met her mother who was out with friends on that day."

b. $^{?}[[(\text{Sie hat }((\text{ihre Mutter})^{\Phi}\ [\text{die an dem Tag mit Freunden unterwegs war}]_{\iota})^{\Phi}$
$\text{getroffen})_{\Phi}]_{\iota}]_{\iota}$

Figure 2 illustrates the difference in prosodic structure between the two versions of (4). In the left panel, the relative clause is extraposed, and the Φ-phrase formed by the object and the transitive verb is not interrupted. As before, both the main clause and the embedded clause project ι-phrases at the higher level of the hierarchy. But when the relative clause is in-situ, as in the right panel of Fig. 2, Layeredness is violated in the DP. The object of *getroffen,* thus *ihre Mutter,* forms a Φ-phrase because of *Match Phrase,* and the relative clause forms an ι-phrase because of *Match Clause*. Additionally, the DP plus the relative clause also form a Φ-phrase.[6] The verb by contrast does not form a Φ-phrase by itself. *Equal Sisters* is violated twice, in the Φ-phrase formed by the DP and in the Φ-phrase formed by the VP.

Comparing extraposition of a sentential complement with extraposition of a relative clause, it is striking that extraposition improves the acceptability of sentences with a sentential complement much more than in the case of a relative clause. There is a difference in acceptability between the two versions of (1), which is absent in (4). Extraposition of a relative clause is never obligatory: A preverbal relative clause may be degraded but is always acceptable. Besides the difference in prosodic structure, to which we return in Section 5, it must also be noticed that the relative clause has an antecedent, as opposed to a complement clause, which has none. The presence of an antecedent provides a strong syntactic motivation for a

[6] In the example, the relative clause is nonrestrictive because of the antecedent *Mutter* "mother," denoting a unique person, and I assume that, in this case, *ihre Mutter* "her mother" is a Φ-phrase mapped to the DP to which the relative clause is adjoined. A restrictive relative clause would be attached to the N *Mutter,* forming a prosodic monster one level down the hierarchy. It is sometimes assumed that a restrictive relative clause extraposes more easily than a non-restrictive relative clause. This may be due to the difference in the level at which the prosodic monster is formed (see also Section 4 for some comments on the influence of accent structure, definiteness and restrictivity on extraposition).

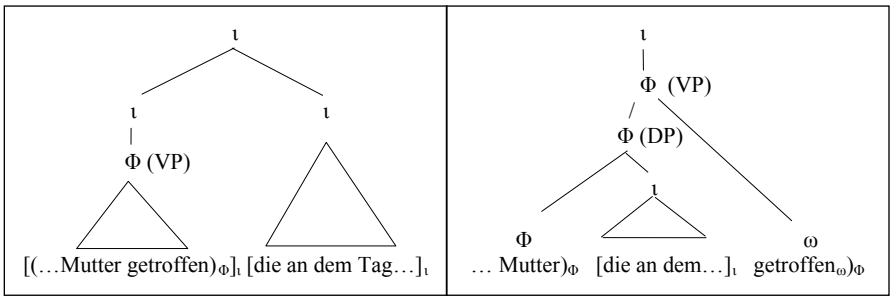


Fig. 2 Extraposition as avoidance of a prosodic monster (relative clause)

relative clause to be adjacent to its antecedent. The antecedent anchors the entire object with its relative clause in the preverbal position, as they form a syntactic and a prosodic constituent together, as shown in Fig. 2. This constituent is lacking in the case of a complement clause. In the OT account in Section 5, the preference for a relative clause and its antecedent to be adjacent is captured by a constraint called ADJACENCY, formulated in (27).

That this analysis is on the right track is further confirmed by the following observation. A *dass*-complement can follow a noun, a demonstrative or a quantifier, as shown in (5) with a noun. In this case, the complement clause behaves like a relative clause and can remain head-adjacent, even if the part of the main clause following the complement clause is very short and consists of only one ω-word. The embedded clause is thus quite acceptable in the preverbal position. The extraposed version is of course even better; see (5b).

(5) a. ?[Anna hat (die Behauptung, [dass sie in der Nacht ihre Mutter im
Anna has the claim that she in the night her mother on.the
Treppenhaus gesehen hat]$_{\iota}$, bestritten)$_{\Phi}$]$_{\iota}$.
staircase seen has denied
"Anna denied the claim that she met her mother on the staircase that night."

b. [Anna hat (die Behauptung bestritten)$_{\Phi}$]$_{\iota}$, [dass sie in der Nacht ihre Mutter im Treppenhaus gesehen hat]$_{\iota}$.

In further cases, the in-situ version of sentences with embedded clauses sounds at least as good or even better than the extraposed version. Consider (6), in which the final part of the main clause consists of two words, *nicht erzählt* "not told," instead of just one. Augmenting the verb with an adverb improves the in-situ variant of this sentence. This is because now the adverb plus the verb form a Φ-phrase. In (6) and Fig. 3, the relative clause is inserted *between* two Φ-phrases. The adverb carries the nuclear stress of the main sentence, which is then adjacent to the verb. The top Φ dominates two lower Φ-phrases and an ι, and a prosodic monster is avoided. Additionally MINIMALBINARITY (MINBIN) is fulfilled, a constraint to the effect that a Φ-phrase needs at least two ω-words to be well formed (Ghini 1993; Selkirk 2000). It is fulfilled in Fig. 3 by *nicht erzählt*. See Section 5 for a formal demonstration.

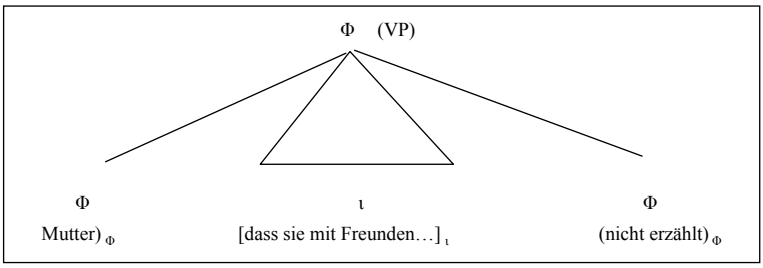


Fig. 3 No prosodic monster in a *dass* complement with nominal antecedent: *MinBin* is fulfilled

(6) [(Sie hat ihrer Mutter)$_{\Phi}$ ([dass sie mit Freunden unterwegs war]$_{\iota}$ (nicht erzählt)$_{\Phi}$)$_{\Phi}$]$_{\iota}$
she has her mother that she with friends out was not told
"She did not tell her mother that she was out with friends."

It has been shown in this section that a prosodic account of extraposition can explain the difference between nearly obligatory extraposition of complement clauses and optional extraposition of relative clauses. Extraposition is nearly obligatory when an embedded structure creates a prosodic monster. In the case of a relative clause, extraposition is optional because extraposition destroys the preferred adjacency between the antecedent and the relative clause. In this case, the need for continuous constituents conflicts with the need to avoid prosodic monsters.

3 The Prosodic Structure of PP Extraposition

As in the case of clauses, extraposition of prepositional phrases improves the prosodic structure of the sentence as a whole. However, it is rarely obligatory, and only rarely preferred. The prosodic structure of the in-situ versions of PPs involves a Φ-phrase mapped to the PP and often embedded into a larger Φ-phrase, depending on the syntactic role of the PP. Recursion of Φ-phrases is often found in German and does not lead to ungrammaticality by itself. Nevertheless, a PP readily extraposes, creating in this way a prosodically balanced structure, as shown below.

In illustrating PP extraposition, a syntactic distinction will be adopted from Frey (2012), who distinguishes between attributive, argumental, and adverbial PPs. Both syntactic and prosodic structures differ between these three kinds of PP. We start with attributive PPs, as in *von ihrer Mutter* "of her mother" in (7). The attributive PP is part of the DP whose head it characterizes, and it is embedded into the larger DP when it is in-situ. Such a PP can be extraposed, as in (7a), or in-situ, as in (7b); there is not much difference in acceptability.

(7) a. [(Maria)$_{\Phi}$(wollte (das Kleid)$_{\Phi}$ tragen)$_{\Phi}$ (von ihrer Mutter)$_{\Phi}$]$_{\iota}$
Maria wanted the dress wear of her mother
"Maria wanted to wear her mother's dress."
b. [(Maria)$_{\Phi}$(wollte (das Kleid (von ihrer Mutter)$_{\Phi}$)$_{\Phi}$ tragen)$_{\Phi}$]$_{\iota}$

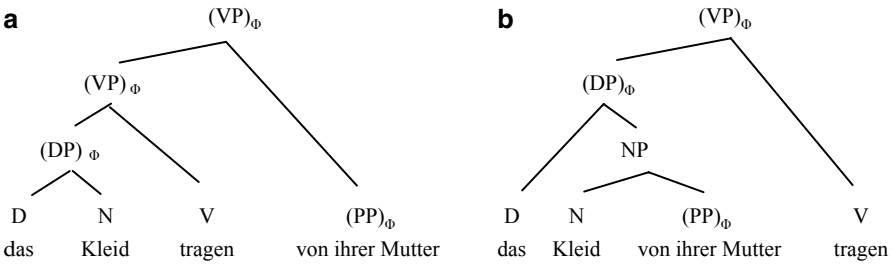


Fig. 4 Extraposition of an attributive PP

Compare Fig. 4 illustrating the prosodic structure of the two versions with the syntactic structure added. In the case of PP extraposition (left panel and (7a)), there is a lower segment of the VP consisting of the head of the object and the verb, thus *das Kleid tragen,* allowing them to build a Φ-phrase to the exclusion of the attributive PP. Even if *Kleid* "dress," the head of the argument noun phrase, does not carry the nuclear accent in this case, it has a special role in being preverbal: it is the head of the argument-predicate complex. The higher VP segment includes both the lower VP and the attributive PP. In the in-situ case (right panel and (7b)), the entire argument is immediately preverbal. The VP is complete with the PP intervening between the noun *Kleid* "dress" and the verb. As a result, the head noun *Kleid* is separated from the verb by the possessive attributive, which carries the default nuclear accent. In this case, the PP *von ihrer Mutter* "of her mother" is a Φ-phrase, embedded in the Φ-phrase of the entire object, which is itself embedded into the Φ-phrase of the VP. In both cases, recursion of the Φ-phrase applies, although in different ways. Both before and after extraposition, the PP is a subpart of the prosodic constituent from which it originates, i.e., the Φ-phrase matching the higher VP segment. However, it is recursively embedded in the case of preverbal location and juxtaposed in the case of extraposition.

The first version contains two more or less equally balanced Φ-phrases (the VP and the PP), but the PP is separated from the noun it modifies; see Section 5 for a more formal analysis. The second version contains one long recursive Φ-phrase (the VP). The two versions elicit subtle differences in meaning. When the attributive is discourse-given, (7a) is much better; (7b) is avoided when both the attributive and the final verb are unaccented. If the attributive PP is new, both versions are fine.[7]

If the possessive attributive constituent is a genitive DP as in (8), extraposition is ungrammatical.

[7] In the example (i) from Haider (2010, quoted from Max Frisch), the extraposed PP is unaccented, and thus potentially right dislocated.

(i) (Sie will (nichts mehr)$_{\Phi}$ wissen)$_{\Phi}$(davon)$_{\Phi}$
she wants nothing more know it-of
"She does not want to know anymore about this."

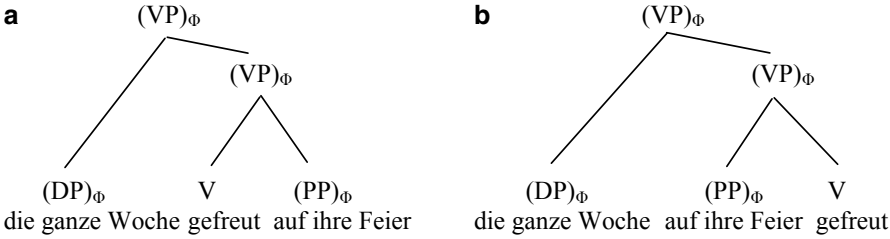


Fig. 5 Extraposition of an argumental PP

(8) a. [(Maria)$_{\Phi}$ (wollte (das Kleid (ihrer Mutter)$_{\Phi}$)$_{\Phi}$ tragen)$_{\Phi}$]$_{\iota}$
Maria wanted the dress her.GEN mother wear
"Maria wanted to wear her mother's dress."
b. *[(Maria)$_{\Phi}$(wollte ((das Kleid)$_{\Phi}$ tragen)$_{\Phi}$)$_{\Phi}$(ihrer Mutter)$_{\Phi}$]$_{\iota}$

I assume that the explanation for the ungrammaticality of (8b) is located in syntax, and not in prosodic structure, since the DP in (8) has the same prosodic form as the PP in (7): the reason for the ungrammaticality of (8b) is that a genitive complement has to be adjacent to its head, and thus it cannot be extraposed on independent grounds. Notice also that the genitive DP in (8) gets its case from the noun and not from the verb, so that an explanation in terms of case assignment should be general enough to account for such a restriction. And a DP is introduced by a functional element, the article, in the same way as PPs and CPs are also introduced by functional elements, so that this cannot explain the difference between the extraposability of the constituents, at least not without additional stipulations.

The next example, in (9), involves an argumental PP. (9a) shows an extraposed PP *auf ihre Feier* "to her party," and (9b) an in-situ one.

(9) a. [(Anna)$_{\Phi}$ (hatte sich (die ganze Woche)$_{\Phi}$(gefreut (auf ihre Feier)$_{\Phi}$)$_{\Phi}$)$_{\Phi}$]$_{\iota}$
Anna had REFL the whole week rejoiced on her party
"Anna had been looking forward to her party the whole week."
b. [(Anna)$_{\Phi}$ (hatte sich (die ganze Woche)$_{\Phi}$ ((auf ihre Feier)$_{\Phi}$ gefreut)$_{\Phi}$)$_{\Phi}$]$_{\iota}$

Except for the relative position of the verb and its argument and thus the position of the metrical head, there is no difference in phrasing between (9a) and (9b), see Fig. 5. In both cases, *auf ihre Feier* forms a Φ-phrase together with the verb, in addition to forming its own Φ-phrase. In other words, the argument and the verb are prosodically integrated in a joint Φ-phrase. There is thus one level of embedding less than in the case of an attributive PP. Notice that the adverbial *die ganze Woche* "the whole week" and the participle do not form a single Φ-phrase together to the exclusion of the argument (see Gussenhoven 1992; Truckenbrodt 2006 and Féry 2011 for the difference in phrasing between arguments and adjuncts).

The extraposed version in (9a) is again preferred when the argument is given in the context, or at least when it has a different information structural role from the preceding constituent. When the argument is immediately preverbal, as in (9b), the

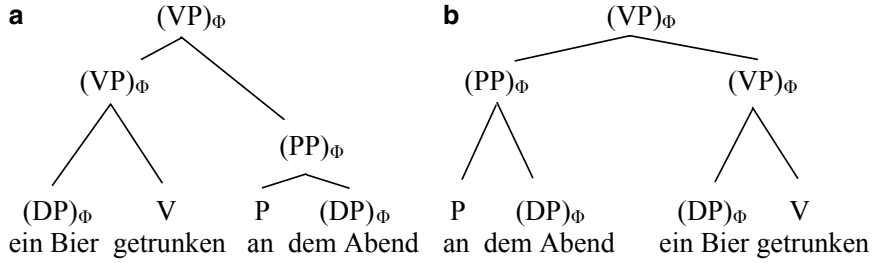


Fig. 6 Extraposition of an adverbial PP

nuclear accent of the argument–predicate complex is located on this preverbal constituent. This version is preferred in a context where the sentence is all-new, and it is slightly awkward when the argument is given and the verb needs the nuclear accent.

Finally, when the PP is an adverbial adjunct, as in (10), no integration between verb and adjunct is expected. In other words, the PP and the verb are in different Φ-phrases from the start, regardless of word order. In (10a), the temporal adverbial *an dem Abend* "in the evening" is extraposed, and in (10b), it is in-situ. The argument *ein Bier* "a beer" is preverbal in both cases, and this argument forms a Φ-phrase with the verb. The adverbial PP is located before the object in the in-situ word order, as shown in (10b). The prosodic versions are shown in Fig. 6. It is unsurprising that both orders, the non-extraposed and the extraposed one, are more or less equivalent in their acceptability. Argument and verbal head are adjacent in both cases. And as before, the extraposed version is the best one if the adjunct is given. In the b version it may also be a (contrastive) topic; see Frey (2004).

(10) a. $[(\text{Anna})_{\Phi}\ ((\text{hatte}\ (\text{ein Bier})_{\Phi}\ \text{getrunken})_{\Phi}\ (\text{an dem Abend})_{\Phi})_{\Phi}]_{\iota}$
Anna had a beer drunk at the evening
"That evening, Anna had a beer."
b. $[(\text{Anna})_{\Phi}\ (\text{hatte}\ (\text{an demAbend})_{\Phi}\ ((\text{einBier})_{\Phi}\ \text{getrunken})_{\Phi})_{\Phi}]_{\iota}$

A directional or locational adverb is usually preverbal, i.e., located after the argument of the verb and it thus intervenes between the argument and the predicate. It is often unaccented, and does not block the integration between object and verb; see Féry (2011) for an OT analysis. Haider (2010) cites a sentence with an extraposed locational adjunct PP, (11a), from Thomas Mann. The locational PP can also appear between the preverbal object and the verb, as in (11b), forming a recursive prosodic structure. However, and differently from the attributive PP in (7), the adjunct can be unaccented even if it is not part of the background.

(11) a. $(\text{Morgen})_{\Phi}\ (\text{soll}\ \text{ich}\ (\text{den Dienst})_{\Phi}\ \text{antreten})_{\Phi}\ (\text{in diesem Haus})_{\Phi}$
Tomorrow shall I the service begin in this house
"Tomorrow I shall begin my service in this house."
b. $(\text{Morgen})_{\Phi}\ (\text{soll ich}\ (\text{den Dienst})_{\Phi}\ (\text{in diesem Haus})_{\Phi}\ \text{antreten})_{\Phi}$

In the case of an adjunct PP, the ability to extrapose is probably due to the adverbial status rather than to the PP status. In (12), an adverbial DP is extraposed. Such adverbial DPs bear intrinsic (nonstructural) case, as opposed to the structural case of arguments as discussed for (8).

(12) [Weil ((Maria)$_{\Phi}$ (geschlafen hat)$_{\Phi}$)$_{\Phi}$ (den ganzen Vormittag)$_{\Phi}$]$_{\iota}$
Because Maria slept has the whole morning
"Because Maria slept the whole morning."

In summary, the answer provided in this chapter for extraposition is based on the prosodic needs of a sentence, which may conflict with the syntactic preferences. On the syntactic side, there is a strong preference for constituents to be continuous, and for arguments to be on the left side of the verb in order to be properly governed, at least in the embedded word order. There is also a preference for the verb to be sentence final. On the prosodic side, extraposition results in fulfillment of LAYEREDNESS and *Non-Recursivity* and an overall more balanced prosodic structure than in the case of in-situ; see Section 5. The choice between extraposition and non-extraposition of PP can be the result of a trade-off between these conflicting tendencies.

4 Prosodic Limits of Extraposition

So far, it has been shown that extraposition may improve the prosodic structure of an entire sentence. In this section, we turn to examples that show that extraposition of a relative clause or of a PP may lead to less acceptable results than an in-situ version of the same sentence. This happens when a potential intervener is located between an extraposed constituent and its antecedent or its reconstructed position. A potential intervener is an accented full maximal projection (XP), usually a DP. A constraint called NOINTERVENER is formulated in (13) for ease of reference. This constraint forbids the presence of an accented intervener—the accented XP in (13)—between the antecedent of an extraposed constituent or its reconstructed position (… t_i…) and its actual position (…YP_i…). It is to be interpreted as a violable OT constraint, thus as expressing a preferred option, rather than a strict prohibition.[8]

(13) NOINTERVENER: No intervener between antecedent or reconstructed position and extraposed relative clause
×
*… t_i…(XP)$_{\Phi}$(….YP_i…)$_{\Phi}$

The main idea of NOINTERVENER is to account for the fact that the distance between an extraposed constituent relative to its reconstructed position is not as relevant as the presence of an intervening potential antecedent.

[8] The absence of prosodic boundaries around t_i leaves it open whether there are additional boundaries. Moreover, ι-phrase boundaries separate XP and YP in the case of clause extraposition.

Consider first an example involving PP extraposition, adapted from Truckenbrodt (1995b, p. 510). NoIntervener accounts for the difference in felicity between (14a) and b in the following way: Sentence (14a) satisfies NoIntervener because there is no intervener between the reconstructed position *t* and the extraposed constituent. Sentence (14b) does not satisfy NoIntervener because the DP *Buch* "book" intervenes between the reconstructed position *t* and the extraposed constituent. In such a constellation, the participle is not accented. Accented words are indicated with small caps. Note that the status of the extraposed constituent as accented or not is immaterial.

(14) a. (Anna)$_{\Phi}$ (hat einem KOLLEGEN)$_{\Phi}$ (ein BUCH *t* gekauft)$_{\Phi}$ (von Chomsky)$_{\Phi}$.
Anna has a.DAT colleague a book bought by Chomsky
"Anna has bought a book by Chomsky for a colleague."
b. ??/*(Anna)$_{\Phi}$ (hat einem KOLLEGEN *t*)$_{\Phi}$ (ein BUCH gekauft)$_{\Phi}$ (aus Italien)$_{\Phi}$.
Anna has a.DAT colleague a book bought from Italy
"Anna has bought a book for a colleague from Italy."

The following examples illustrate relative clause extraposition. In (15a), there is no intervener between the relative clause and its antecedent, whereas there is one in (15b), and it is this difference that accounts for the ill-formedness of (15b).

(15) a. [(Linda)$_{\Phi}$ (hat dem KIND)$_{\Phi}$(das KLEID *t* geschenkt)$_{\Phi}$]$_{\iota}$ [(das sie selbst
Linda has the.DAT child the dress given that she herself
ausgesucht hatte)$_{\Phi}$]$_{\iota}$.
chosen had
"Linda gave the child the dress that she had chosen herself."
b. ??/*[(Linda)$_{\Phi}$ (hat dem KIND *t*)$_{\Phi}$ (das KLEID geschenkt)$_{\Phi}$]$_{\iota}$ [(das gestern
Linda has the.DAT child the dress given who yesterday
geweint hat)$_{\Phi}$]$_{\iota}$.
cried has
"Linda gave the dress to the child who cried yesterday."

Numerous similar cases of ill-formed extraposition of a relative clause are well known from the literature, some of which are reproduced here. In all examples, the source of the infelicity is the intervener separating the relative clause from its antecedent; see also Bader, this volume, and Poschmann and Wagner (2014) for experimental confirmation of this observation for German. (16a–b) are from Haider (1994). (16c) is from Lenerz (1977, p. 34); see also Altmann (1981, p. 176).[9]

[9] The account presented here contrasts with the formula (i) proposed by Truckenbrodt (1995b:503), which claims that only the distance in terms of prosodic constituents counts for extraposition.
(i) [$_{\pi}$… XP…] → [$_{\pi}$… t$_{i}$…] [$_{\pi}$ XP$_{i}$].
Extraposed constituents are separated from their base position by exactly one phonological constituent of the same size as themselves. When the movement is too short or too long, extraposition is no longer allowed. XP is a syntactic category that is mapped into the prosodic category π. π is either a Φ-phrase or an ι-phrase: An extraposed PP is a Φ-phrase and an extraposed clause is an ι-phrase. However, Frey (2009) shows that (i) both overgenerates and undergenerates. For

(16) a. ??/*[(Maria)$_{\Phi}$ (hat dem KOLLEGEN *t*)$_{\Phi}$(ihre FREUNDIN vorgestellt)$_{\Phi}$]$_{\iota}$
Mary has the.DAT colleague her.ACC friend introduced
[(der im LOTTO gewonnen hat)$_{\Phi}$]$_{\iota}$.
who in-the lottery won has
"Mary introduced her friend to the colleague who won in the lottery."

b. [Maria hat ihre FREUNDIN dem KOLLEGEN *t* vorgestellt]$_{\iota}$ [der im Lotto gewonnen hat]$_{\iota}$.

c. ??/*[(Peter hatte der FRAU *t*)$_{\Phi}$ (eine ROSE geschenkt)$_{\Phi}$]$_{\iota}$ [die schwanger war]$_{\iota}$.
Peter had the woman a rose given who pregnant was
"Peter gave a rose to the woman who was pregnant."

Altmann (1981) and Inaba (2007) claim that every intervening DP can in principle block extraposition; see also Bolinger (1992) for similar remarks for English. However, Kathol and Pollard (1995) cite the following exception to the general blocking by any intervening DP: directional or locational adverbs can be unaccented, even when they are new in the context (see Féry 2011 for a prosodic analysis of such adverbials); compare (17) and also (18) from Truckenbrodt (1995b). In such cases, NOINTERVENER is fulfilled since there is no accented potential antecedent intervening between the extraposed constituent and the antecedent.

(17) Wir haben das BUCH *t* ins Regal gestellt [das ich gestern gekauft habe]$_{\iota}$
we have the book on.the shelves put that I yesterday bought have
"We put the book that I bought yesterday on the shelves."

(18) [Anna hat zwei BÜCHER *t* auf einen Tisch gelegt]$_{\iota}$ [die sie am Dienstag
Anna has two books on the table put which she on Tuesday
aus Italien mitgebracht hat]$_{\iota}$.
from Italy brought has
"Anna put two books that she brought from Italy on Tuesday on the table."

Lenerz (1977, p. 35) observes that the relative clause can be extraposed across a full DP when the determiner is accented, as in (19a).[10] Wiltschko (1997, p. 387) makes the same claim and cites the pair in (19 b–c). She attributes the grammaticality of (19b) to the restrictiveness of the relative clause. It is true that an accent on the determiner strongly correlates with a restrictive reading. However, the reason for the improvement of these sentences relative to those in (16) is the absence of an accented DP between the antecedent and the relative clause, as the reader with knowledge of German can verify. The nuclear status of the accent on the determiner

instance, in both cases in (15), the extraposed relative clause is adjacent to the ι-phrase containing the antecedent and thus to the ι-phrase from which the relative clause originates, in agreement with (i), which thus predicts that both versions of (15) should be equally acceptable. The same comment holds for all sentences in (16).

[10] Bader (this volume) finds that accent on the determiner improves the acceptability of a sentence with extraposition, as compared to accent on the noun.

correlates with the absence of postnuclear accents after the pitch accent. If there is no accented intervener, NoIntervener is satisfied.

(19) a. Peter [hatte DER Frau eine Rose/sie geschenkt] [die schwanger war]$_t$.
b. [Den Mann *t* gesehen] hat Peter gestern auf der Party [der Bier trinkt]$_t$.
the man seen has Peter yesterday on the party who beer drinks
"Peter saw the man who drinks beer at the party yesterday."
c. ??/*[Den Mann *t* gesehen] hat Peter gestern auf der Party [der Bier trinkt]$_t$.

NoIntervener often accounts for the sequencing of embedded clauses. The sentences in (20), adapted from Wiltschko (1997, p. 381), contain two extrapositions, a relative clause and a *dass*-complement. NoIntervener accounts for the preference for (20a) over (20b) if it is assumed that embedded clauses contain at least one accented word.

(20) a. weil Anna einer Frau t_1 t_2 gesagt hat [die sie KANNTE$_1$]$_t$ [dass
because Anna a.DAT woman said has who.FEM she knew that
sie JEMANDEN getroffen hat$_2$]$_t$.
she someone met has
"Because Anna told a woman she knew that she met someone."
b. * weil Anna einer Frau *t* gesagt hat [dass sie JEMANDEN getroffen hat]$_t$
because Anna a.DAT woman said has that she someone met has
[die sie KANNTE]$_t$.
who.FEM she knew

If the reconstructed position is the same, two extraposed clauses can come in both orders. The following examples are again from Wiltschko (1997, p. 381). t_1 t_2 may come in both orders. Notice that the in-situ version, with both embedded sentences in the preverbal position, is barely acceptable if at all. One embedded clause is a *dass*-complement and the other one is a comparative clause.

(21) a. Peter hat schneller t_1 t_2 gesagt, [$_{ARG}$ dass er sich langweilt]$_t$ [$_{COMPAR}$ als ich
Peter has more.quickly said that he REFL bored.is]$_t$ than I
erwartet hatte.
expected had
"Peter said more quickly than I had expected that he was bored."
b. Peter hat schneller gesagt, [$_{COMPAR}$als ich erwartet hatte]$_t$ [$_{ARG}$dass er sich langweilt]$_t$.

To sum up, this section has been concerned with accented elements intervening between an extraposed element and its antecedent (in the case of a relative clause) or its reconstructed position (in the case of a complement clause or a PP). It has been shown that such an intervener always drastically reduces the grammaticality of extraposition, and that in the case of a relative clause an accented DP is particularly problematic. Some further prosodic effects also play a role, such as accents on other constituents. Additional syntactic and semantic principles influencing the order of two extraposed clauses, like binding and information structure, cannot be discussed in this chapter.

5 An OT Analysis of Extraposition

5.1 The Role of Syntax

If it is assumed that the canonical licensing direction for verbs in German is to the left (see for instance Frey 2012; Haider 2010; Hartmann 2013 and Sternefeld 2008 for this claim), extraposition of sentential complements is bound to be a syntactically marked construction as compared to in-situ location of complements. As a result, the near obligatoriness of extraposition in (1) is difficult to explain in a purely syntactic model. To make this point even clearer, compare the preverbal position of a nominal argument with the ungrammatical extraposition of this argument in (22). As shown in (22c), topicalization is not a problem for argumental DPs in German.

(22) a. Anna hat ihrer Mutter die Geschichte erzählt.
Anna has her mother the story told
"Anna told her mother the story."
b. *Anna hat ihrer Mutter erzählt die Geschichte.
c. Die Geschichte hat Anna ihrer Mutter erzählt.

All approaches assuming that extraposition is movement to the postfield must assume at the same time that it is a less marked syntactic option for a relative clause to be adjacent to its head than to be extraposed. It can safely be claimed that no movement approach has ever considered an extraposed constituent as syntactically better than its in-situ counterpart. As a result, it has sometimes been claimed that extraposition is a postsyntactic phenomenon; see Chomsky (1986, p. 40) for the view that "extraposition is indeed a phonetic form (PF) rule."[11]

The difficulty of finding a straightforward explanation for extraposition in purely syntactic terms is even broader. It is fair to say that although the body of literature on the subject is huge, it is largely inconclusive: Neither A-movement nor Ā-movement nor base-generation delivers satisfactory explanations. One reason for this relates to the diversity and the complexity of the individual factors bearing on extraposition in syntax, as shown by several authors (see, for instance, Büring and Hartmann 1997 and Culicover and Rochemont 1990).

According to Haider (2010, p. 205) and Frey (2012), even though the canonical direction of licensing by verbs is to the left, extraposed PPs may be base-generated postverbally as locally dependent elements, with an obligatory antecedent relation. This may also hold for relative, complement, comparative, and resultative clauses. An argument for the view that argument clauses (CP complements) are generated after the verb is illustrated with an example from Müller (1998, p. 166) that shows that a preverbal argument clause can be ungrammatical rather than merely infelicitous.

[11] Chomsky lists three reasons for this judgment. First, the usual constraints on movement operations are not operational for extraposition. Second, the movement has no configurational-structural effect. Third, extraposition is in principle not obligatory, but optional.

(23) a. (Ich weiß nicht) wen_i er gesagt hat [$_{CP}$ dass Claudia t_i geküsst hat].
I know not whom he said has that Claudia kissed has
"I don't know who he said that Claudia has kissed."
b. *(Ich weiß nicht) wen_i er [*CP* dass Claudia t_i geküsst hat] gesagt hat.

To assume only one underlying position for each constituent may be misleading. An alternative solution is to assume that the position of a CP argument is intrinsically optional. In this case, the position of a dependent clause is not regulated once and for all in syntax; rather in many cases, both pre- and postverbal locations are possible options. In this case, the decision as to which surface position a clause occupies in a specific case may be driven by prosody (or by semantic, or information structural factors) in an OT fashion. Such an approach is in line with the prosodic approach developed in this chapter, which claims that the prosodic factors entering into the decision to extrapose a phrase or not may be decisive.

5.2 The Role of Prosody

This section proposes an OT approach to PP and clause extraposition to account for the prosodic component of the operation; see Prince and Smolensky (1993/2004) and McCarthy and Prince (1993a, b) for OT. Several OT constraints have been introduced above. It is now time to show how they interact formally in grammar and how they affect the data on extraposition. It is assumed that syntax delivers alternative linearizations of the constituents under examination, providing in this way the candidates to be evaluated. In other words, the candidates shown in the tableaux below are the result of the possible linearizations according to the syntactic constraints on linearization. The fact that the syntactic constraints are not shown in the tableaux does not imply that they are lower or higher ranking than the prosodic constraints. On the contrary, it is assumed here that syntax and prosody are working hand in hand and simultaneously. We concentrate in this chapter on the prosodic constraints, and ignore the syntactic constraints. In all the examples considered below, syntax provides two linearizations of embedded clauses and PPs: in-situ and extraposed. Note that there may be further relevant candidates delivered by the syntax, but they are of no concern here.

The candidates are assigned a prosodic structure through the effect of the *Match* constraints. The *Match* constraints from Selkirk (2011) were formulated in (2). They straightforwardly assume that a grammatical word, a syntactic phrase, and a clause roughly correspond to the three higher prosodic constituents, ω-word, Φ-phrase, and ι-phrase, respectively. The effects of the *Match* constraints are counterbalanced by well-formedness constraints imposing restrictions on the form of the relevant prosodic domains, as well as on the relations between them. These constraints evaluate the resulting prosodic domains, and choose among several candidates those that fulfill the well-formedness constraints best. Some of the well-formedness constraints were introduced and illustrated above; they are formulated

in (24); see Ghini (1993); Nespor and Vogel (1986), and Selkirk (1996, 2000) for the original formulations.

(24) a. Non-Recursivity: A prosodic constituent C_n does not dominate another constituent of the same level C_n.

b. Layeredness: A prosodic constituent C_m does not dominate a constituent of a higher level C_n, n>m.

c. Headedness: A constituent C_n dominates a constituent of the immediately lower level C_{n-1}. (A prosodic constituent has a head on the immediately higher level.)

d. Exhaustivity: No C_n immediately dominates C_{n-2}. (No prosodic constituent is skipped.)

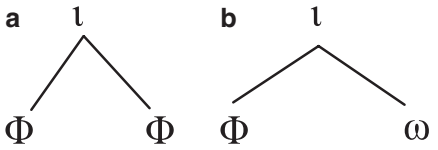
e. MinimalBinarity: A prosodic constituent C_n dominates at least two C_s. (A prosodically binary constituent is better balanced than a simple one.)

Myrberg (2013) proposes the constraint EqualSisters, which posits that the sister constituents of a dominating prosodic constituent are at the same level of the prosodic hierarchy. Headedness and Exhaustivity in (24c) and d independently account for the fact that the two sisters are preferably of the immediately lower category.

(25) EqualSisters (Myrberg 2013, p. 75)
Sister nodes in prosodic structure are instantiations of the same prosodic category.

(26)a fulfills EqualSisters, and (26)b violates this constraint.

(26)



As was shown in Section 2, a prosodic monster violates both Layeredness and EqualSisters.

An additional constraint in (27), called Adjacency, requires adjacency between a relative clause or an attributive PP and its nominal head (the antecedent). The fact that the relative clause or the possessive attributive is to the right of its head is regulated by independent (syntactic) principles that are of no concern here.

(27) Adjacency: A relative clause or a possessive attributive is adjacent to its antecedent.

In the following, it is shown how in-situ complement clauses and relative clauses violate the relevant well-formedness constraints, and how they are thus suboptimal as compared to the corresponding extraposed versions. In Tableau 1 (T1) the input

T1 DP+V+*dass*-Compl(1)		Layered	EqSis	Adj	MinBin
a. ☞	Ex (1a): [(…DP V)$_{\Phi}$]$_{\iota}$ [*dass*-Compl]$_{\iota}$				
b.	In (1b): [(…DP [*dass*-Compl]$_{\iota}$ V$_{\omega}$)$_{\Phi}$]$_{\iota}$	*!	**		
c.	In (1b): [(…DP [*dass*-Compl]$_{\iota}$(V)$_{\Phi}$)$_{\Phi}$]$_{\iota}$	*!	**		*

In-situ version: [(Sie hat niemandem, [dass sie spät nach Hause kam]$_{\iota}$ erzählt$_{\omega}$)$_{\Phi}$]$_{\iota}$

consists of a verb, an argument of the verb, and a complement clause of the verb, not linearized relative to each other. The extraposed and the in-situ versions of sentence (1) are the candidates to be evaluated for their prosodic well-formedness. The *Match* constraints are high ranking (see below for some additional remarks to this effect), and they are not violated in the tableaux of this section. For reasons of space, they are not shown. However, see below for elements of a solution to the problem of prosodic monsters implying violation of *Match*.

Candidate a. is the extraposed version, and it does not violate any of the well-formedness constraints in the tableau. It does violate Non-Recursivity and Headedness but these constraints are relatively low ranking in German.

Candidates b. and c., the suboptimal in-situ versions, violate Layeredness and EqualSisters. The latter constraint is violated twice in each candidate, because the top Φ-phrase dominates a Φ-phrase, an ι-phraseand a ω-word; see Fig. 1. Each adjacent pair of constituents constitutes a violation of EqualSisters. Candidate c shows that the main verb in the in-situ version is too light to form a Φ-phrase all by itself: it violates MinimalBinarity. It is important to realize that the well-formedness constraints are violable. It is proposed here that Layeredness is higher ranked than EqualSisters, Adjacency, and MinimalBinarity, though the exact ranking maybe subject to revision when more structures are considered.[12]

In Tableau 2 (T2), the input consists of a DP, a relative clause, and a verb, not linearized. Again, the two linearizations shown in candidates a. and b. are the results of the syntactic constraints. The extraposed version a. violates Adjacency, and the in-situ version b. violates Layeredness once and EqualSisters twice. This time, one of the violations of EqualSisters is caused by the Φ-phrase of the VP, which dominates a Φ-phrase and a ω-word, while the second violation comes from the Φ-phrase of the DP, which dominates a Φ-phrase and an ι-phrase; see Fig. 2. To account for

T2 [DP+*RelCl*+V]$_{VP}$(4)		Layered	EqSis	Adj	MinBin
a. ☞	Ex (4a): […(DP V)$_{\Phi}$]$_{\iota}$ [*RelCl*]$_{\iota}$			*	
b.	In (4b): […(DP [*RelCl*]$_{\iota}$ V$_{\omega}$)$_{\Phi}$]$_{\iota}$	*!	**		

In-situ version: [Sie hat ((ihre Mutter [die an dem Tag mit Freunden unterwegs war]$_{\iota}$)$_{\Phi}$ getroffen)$_{\Phi}$]$_{\iota}$

[12] Below, Non-Recursivity is added for PP extraposition. For now, this constraint is ignored: it is violated a number of times in all candidates. However, it is relatively low ranking and it never decides between the candidates in Tableaux 1, 2, and 3.

T3 *dass-Compl*+Φ-phr. (6)		Layered	EqSis	Adj	MinBin
a. ☞	Ex (6a): …$(\Phi)_{\Phi}]_{\iota}$ [*dass*-Compl]$_{\iota}$				
b.	In (6b): …[*dass*-Compl]$_{\iota}(\Phi)_{\Phi}]_{\iota}$		*!*		

In-situ version: [(Sie hat ihrer Mutter)$_{\Phi}$([dass sie mit Freunden unterwegs war]$_{\iota}$ (nicht erzählt)$_{\Phi}$)$_{\Phi}$]$_{\iota}$

the fact that the in-situ version may sometimes be preferred, it is assumed that other constraints may play a role; see below for the role of NoIntervener.

The in-situ versions considered in Tableaux 1 and 2 display unbalanced prosodic structures. It was shown with sentence (6) that as soon as the second part of the main clause is a Φ-phrase, there is no prosodic monster anymore; see Fig. 3 illustrating this. In the in-situ version, the adverb *nicht* separates the verb from its complement. The extraposed version is optimal in T3 since the in-situ variant violates EqualSisters. As in the case of T1, other constraints not considered here may force the in-situ version to be chosen in some circumstances. In-situ version b. does not violate Layeredness since the highest ι-phrase dominates one ι-phrase and two Φ-phrases (Table 3).

Since violation of the well-formedness constraints resulting in a prosodic monster is dependent on the result of *Match,* and *Match* requires that a clause is always mapped by an ι-phrase, and that a VP or a DP is always mapped by a Φ-phrase, a different kind of solution is conceivable, namely one that changes the prosodic constituency of the syntactic elements, in violation of *Match*. More generally, the question here is whether a clause could be downgraded to a Φ-phrase, or an XP could be upgraded to an ι-phrase, so that no prosodic monster arises in those configurations. In such a case, there would be no violation of Layeredness and no pressure of the prosody on syntax. Prosodic downgrading is expressed in (28).

(28) Prosodic Downgrading: $(\ldots\ [\ldots]_{\iota}\ \ldots)_{\Phi} \rightarrow (\ldots\ (\ldots)_{\Phi}\ \ldots)_{\Phi}$

Infinitive-CPs may escape the need to form ι-phrases, in which case they are indeed downgraded, as illustrated in (29) (from Sternefeld 2008, p. 410). As a result, they do not obligatorily extrapose. Infinitives form verbal complexes with the finite verbs of the main clause, especially modals. This also happens in syntax and in semantics.[13]

(29) [(Weil er (es zu vernichten)$_{\Phi}$ anordnete)$_{\Phi}$]$_{\iota}$
because he it to destroy ordered
"Because he ordered it to be destroyed."

Another conceivable option to escape violation of Layeredness consists in downgrading the Φ-phrase formed on the syntactic phrase comprising the embedded clause to an ι-phrase, as shown in (30). In a conceivable but different implementation of the *Match* constraints, as soon as a Φ-phrase contains an ι-phrase, it would

[13] Generally it can be said that the less embedded sentences participate in the at-issueness of the main clause, the more likely they are to be separate ι-phrases (see Potts 2005 and Selkirk 2011), and vice versa. However, since at-issueness is not prosodic, we do not try to address it in detail here.

become an ι-phrase itself, and thus respect LAYEREDNESS. However, except for the sake of avoiding a violation of LAYEREDNESS, there is no reason for such a step, in particular no intonational one. This solution strikes me as ad-hoc.

(30) Prosodic upgrading: $(\ldots\ [\ldots]_{\iota}\ \ldots)_{\Phi} \rightarrow [\ldots\ [\ldots]_{\iota}\ \ldots]_{\iota}$

Various models of syntax–prosody disallow embedding of prosodic constituents in each other, and eliminate this possibility from the start. A strict application of ALIGNMENT and NON-RECURSIVITY is illustrated in (31) (see Selkirk 2000 and Truckenbrodt 2006 among others). The result is a sequence of prosodic constituents of the same size, but no embedding of constituents into each other. Such an approach denies that the syntactic structure is reflected in the prosodic structure and favors a flat and non-isomorphic model of prosodic structure. Selkirk (2000) and Truckenbrodt (2006) propose that prosodic constituency may be deleted in postfocal and postnuclear material as a result of the absence of metrical prominence in this part of a sentence.

(31) Result of ALIGNMENT + NON-RECURSIVITY: $(\ldots\ [\ldots]_{\iota}\ \ldots)_{\Phi} \rightarrow (\ldots)_{\Phi}\ [\ldots]_{\iota}(\ldots)_{\Phi}$

Turning now to extraposition of PPs, no prosodic monster is at play here. It was shown in Section 4 that LAYEREDNESS is not violated by in-situ PPs, and that the pressure to extrapose a PP is much less than in the case of clauses. As a result, extraposition of PPs is always optional. To account for PP extraposition, NON-RECURSIVITY and ADJACENCY are equally ranked. MINIMALBINARITY is not shown anymore since it is irrelevant for the following cases. Tableaux 4, 5, 6 illustrate the three types of PPs that were discussed in Section 4. T4 shows a possessive attributive PP, T5 an argument PP, and T6 an adjunct PP. In T4, one candidate violates NON-RECURSIVITY and the other ADJACENCY. In T5, both candidates violate NON-RECURSIVITY, and in T6, no constraint is violated. As a result, in each case, both candidates are optimal.

T4 DP of PP+V (7) Attrib. PP	LAYERED	EQSIS	ADJ	NORECURS
a. ☞ Ex (7a): $(\text{DP V})_{\Phi}(PP)_{\Phi}]_{\iota}$			*	
b. ☞ In (7b): $(\text{DP }(PP)_{\Phi}\ \text{V})_{\Phi}]_{\iota}$				*

Ex version: $[(\text{Maria})_{\Phi}(\text{wollte }((\text{das Kleid})_{\Phi}\ \text{tragen})_{\Phi})_{\Phi}(\text{von ihrer Mutter})_{\Phi}]_{\iota}$

T5 *PP*+V (9) Argument PP	LAYERED	EQSIS	ADJ	NORECURS
a. ☞ Ex (9a): $(\text{V }(PP)_{\Phi})_{\Phi}]_{\iota}$				*
b. ☞ In (9b): $((PP)_{\Phi}\ \text{V})_{\Phi}]_{\iota}$				*

Ex version: $[(\text{Anna})_{\Phi}(\text{hatte sich (die ganze Woche})_{\Phi}(\text{gefreut})_{\Phi}(\text{auf ihre Feier})_{\Phi})_{\Phi}]_{\iota}$

T6 *PP*+VP (10) Adjunct PP	LAYERED	EQSIS	ADJ	NORECURS
a. ☞ Ex (10a): $(\text{VP})_{\Phi}(PP)_{\Phi}]_{\iota}$				
b. ☞ In (10b): $(PP)_{\Phi}(\text{VP})_{\Phi}]_{\iota}$				

Ex version: $[(\text{Anna})_{\Phi}((\text{hatte (ein Bier})_{\Phi}\ \text{getrunken})_{\Phi}(\text{an dem Abend})_{\Phi})_{\Phi}]_{\iota}$

T7 DP [DP *RelCl*]$_{DP}$ V (15a)	NoInterv	Layer	EqSis	Adj
×× a. ☞Ex (15a): (DP)$_{\Phi}$(DP *t* V)]$_{\iota}$ [(*Rel Cl*)]$_{\iota}$				*
×× b. In (15a): (DP)$_{\Phi}$(DP [(*R Cl*)]$_{\iota}$V)$_{\Phi}$]$_{\iota}$		*!	**	

Without interv.: [(Linda)$_{\Phi}$(hat dem Kind *t*)$_{\Phi}$(das Kleid*t* geschenkt)$_{\Phi}$]$_{\iota}$ [(das sie selbst ausgesucht hatte)$_{\Phi}$]$_{\iota}$

T8 [DP *RelCl*]$_{DP}$ DP V (15b)	NoInterv	Layer	EqSis	Adj
××	*!			*
a. Ex (15b): (DP *t*)$_{\Phi}$(DP V)]$_{\iota}$ [(*R Cl*)]$_{\iota}$				
××		*	**	
b. ☞In (15b): (DP)$_{\Phi}$(DP *t* V)$_{\Phi}$]$_{\iota}$ [(*R Cl*)]$_{\iota}$				*

With intervener: [(Linda)$_{\Phi}$(hat dem Kind *t*)$_{\Phi}$(das Kleid*t* geschenkt)$_{\Phi}$]$_{\iota}$ [(das gestern geweint hat)$_{\Phi}$]$_{\iota}$

The selection of one candidate over the other must be made by other constraints (regulating the information structure for instance), not shown here.

Finally, let us turn briefly to the effect of NoIntervener, as formulated in (13). This constraint is higher ranking than the other prosodic well-formedness constraints. Tableau 7 illustrates first a sentence without an intervener. It is assumed that each DP is accented, as shown in the candidates. The relative clause is not separated from its antecedent by an accented DP, whether the relative clause is in-situ or extraposed. As a result, the evaluation takes place as in T2 and the extraposed candidate is optimal. Tableau 8 shows a similar sentence, but this time with an intervener: The relative clause is separated from its antecedent by an accented DP. NoIntervener eliminates the candidate with an extraposed relative clause, even though it violates Layeredness and EqualSisters.

This short overview of an OT approach to the prosodic facts considered here ends the technical part of this chapter. The last section contains a short conclusion.

6 Conclusion

This chapter has investigated the prosodic aspects of extraposition of clauses and PPs, and their influence on syntax. First, a facilitation factor has been identified. It has been shown that extraposition takes place when the prosodic structure of the entire sentence improves, in the sense that the in-situ version violates some well-formedness constraints on the prosodic structure that the extraposed version does not. Avoidance of a prosodic monster, defined as a constellation violating Layeredness and EqualSisters, is achieved by extraposition. A prosodic monster arises

in a sentence containing a relative clause or an argument complement, where a Φ-phrase dominates the ι-phrase mapped on the embedded clause. Moreover, it was also shown that extraposition of a PP usually improves the prosodic structure of the sentence, but that extraposition of a PP is nonetheless always optional. When a PP is in-situ, no well-formedness constraint is fatally violated.

The second factor is a blocking one. Prosodic constraints can limit or block extraposition. First, syntax can block extraposition, as was shown with the ungrammaticality of extraposing an argumental DP in (22b). Second, extraposition is not available when an accented XP intervenes between an antecedent and relative clause, i.e., between an extraposed constituent and its reconstructed position; see (15) and the other examples of Section 4.

There is a long tradition in syntax of explaining extraposition from a purely syntactic perspective. However, syntactic approaches have to choose between movement and base-generated theory, and it has been amply demonstrated in the literature that neither approach can account for all cases of extraposition. As already claimed by Fodor (1998, 2002a, b), a view of prosody that limits its role to the interpretation of syntax is not satisfactory, because the effects demonstrated in this chapter are not due exclusively to syntax; see also Frazier et al. (2006) for the role of prosody in general. What we need is a theory of the syntax–prosody interface that allows a true interaction between the two.

An OT model has been proposed in the chapter that achieves this aim. When speakers elaborate a syntactic structure, they need to plan the corresponding prosodic structure at the same time. A theory like *Match* sketched above requires the syntactic structure to be mapped to abstract prosodic structures, which are layered, headed and recursive (see Féry 2011 for an analysis along these lines for German). Prosody has fewer constituents than syntax, although the constituents are organized in a stricter way than those of syntax. The prosodic structure is regulated by well-formedness constraints. In planning a sentence, a speaker tries to fulfill NoIntervener, Layeredness, Headedness, EqualSisters, MinimalBinarity, and Non-Recursivity, as well as other constraints regulating well-formedness of prosodic constituency. If these principles would be violated too badly in a concrete case, the speaker produces an alternative.

In this chapter, it has been amply demonstrated that prosody plays a role in choosing between competing syntactic structures. Fodor's work has opened a new avenue of research in this direction and this chapter has proposed a new application.

Acknowledgments The ideas developed in this chapter were presented for the first time at the workshop for Janet Fodor organized by Lyn Frazier and Ted Gibson in May 2013. I am very grateful to the organizers and to the audience for the vivid discussion. I also thank Werner Frey, Sara Myrberg, Lisa Selkirk, Fabian Schubö, and especially Shin Ishihara and Gerrit Kentner for comments and discussion. Many thanks to Markus Bader and an anonymous reviewer, as well as to Lyn Frazier, for detailed and helpful comments on a first version of this chapter. Last but not the least, many thanks to Kirsten Brock for correcting my English and other things.

References

Altmann, H. (1981). *Formen der 'Herausstellung' im Deutschen: Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen*. Tübingen: Niemeyer.

Averintseva-Klisch, M. (2006). The 'separate performative' account of the German right dislocation. In C. Ebert & C. Endriss (Eds.), *Proceedings of the Sinn und Bedeutung 10. ZAS Papers in Linguistics 44*. Berlin: ZAS.

Bolinger, D. (1992). The role of accent in extraposition and focus. *Language, 16,* 265–324.

Büring, D., & Hartmann, K. (1997). Doing the right thing. *The Linguistic Review, 14,* 1–42.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Barriers*. Cambridge: MIT.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper and Row.

Culicover, P. W., & Rochemont, M. S. (1990). Extraposition and the complement principle. *Linguistic Inquiry, 21*(1), 23–47.

Elfner, E. (2012). *Syntax-prosody interactions in Irish*. (Doctoral dissertation, University of Massachusetts, Amherst).

Féry, C. (2011). German sentence accents and embedded prosodic phrases. *Lingua, 121,* 1906–1922.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D. (2002a). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the thirty-second annual meeting of the North-Eastern linguistic society* (pp. 113–137). Amherst: University of Massachusetts.

Fodor, J. D. (2002b). *Psycholinguistics cannot escape prosody*. Proceedings of the speech prosody 2002 conference, Aix-en-Provence, pp. 83–88.

Fodor, J. D., & Nickels, S. (2011). *Prosodic phrasing as a source of center-embedding difficulty*. Poster presented at the 2nd experimental and theoretical approaches to prosody conference. McGill University, Canada. Montreal.

Frazier, L., Carlson, K., & Clifton, C. (2006). Prosodic phrasing is central to language comprehension. *TRENDS in Cognitive Sciences, 10*(6), 244–249.

Frey, W. (2004). A medial topic position for German. *Linguistische Berichte, 198,* 153–190.

Frey, W. (2009). *Extrapositionals PF movement: Further arguments in its favour*. Berlin: Handout ZAS.

Frey, W. (2012). *Über die strukturelle Position von extraponierten Konstituenten. Handout Nichtsententiale Konstituenten im deutschen Nachfeld*. Workshop. ZAS Berlin.

Ghini, M. (1993). Ø-formation in Italian: A new proposal. In C. Dyck (Ed.), *Toronto working papers in linguistics 12.2* (pp. 41–78). Toronto: Department of Linguistics, University of Toronto.

Gussenhoven, C. (1992). Sentence accents and argument structure. In I. Roca (Ed.), *Thematic structure. Its role in grammar* (pp. 79–106). Berlin: Foris.

Haider, H. (1994). *Detached clauses—The later the deeper*. Technical report 41, SFB 340. University of Stuttgart.

Haider, H. (2010). *The syntax of German*. Cambridge: Cambridge University Press.

Hartmann, K. (2013). Prosodic constraints on extraposition in German. In G. Webelhuth, M. Sailer, & H. Walker (Eds.), *Rightward movement in a comparative perspective*. Amsterdam: John Benjamins.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Inaba, J. (2007). *Die Syntax der Satzkomplementierung: Zur Struktur des Nachfeldes im Deutschen* [*Studia Grammatica 66*]. Berlin: Akademie Verlag.

Ishihara, S. (2003). *Intonation and interface conditions*. (Doctoral Dissertation, Massachusetts Institute of Technology).

Ishihara, S. (2014). *On match constraints*. Frankfurt a. M.: Ms. University of Frankfurt.

Kathol, A., & Pollard, C. (1995). *On the left periphery of German subordinate clauses*. In West coast conference on formal linguistics14. Stanford University. CSLI Publications/SLA. .

Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition, 123,* 1–20.
Kentner, G., & Féry, C. (2013). A new approach to prosodic grouping. *The Linguistic Review 2013, 30*(2), 277–311. doi:10.1515/tlr-2013-0009
Kiparsky, P. (1982). From cyclic phonology to lexical phonology. In H. van der Hulst & N. Smith (Eds.), *The structure of phonological representations* (pp. 131–175). Dordrecht: Foris.
Konieczny, L. (2000). Locality and parsing complexity. *Psycholinguistic Research,* 29*,* 627–645.
Ladd, D. R. (1990). Metrical representation of pitch register. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and the physics of speech* (pp. 35–57). Cambridge: Cambridge University Press.
Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Narr.
McCarthy, J. J., & Prince, A. S. (1993a). *Prosodic morphology I: Constraint interaction and satisfaction*. Amherst: Ms. University of Massachusetts.
McCarthy, J. J., & Prince, A. S. (1993b). Generalized alignment. In G. Booij & J. van Marle (Eds.), *Yearbook of morphology 1993* (pp. 79–153). Dordrecht: Kluwer.
Müller, G. (1998). *Incomplete category fronting*. Dordrecht: Kluwer Academic Publishers.
Myrberg, S. (2013). Sisterhood in prosodic branching. *Phonology, 30,* 73–124.
Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.
Ott, D., & de Vries, M. (to appear). Right-dislocation as deletion. Natural language and linguistic theory.
Poschmann, C., & Wagner, M. (2014). Relative clause extraposition and prosody. To appear in natural language and linguistic Theory.
Potts, C. (2005). *The logic of conventional implicatures*. Oxford: Oxford University Press.
Prince, A., & Smolensky, P. (1993/2004). *Optimality theory: Constraint interaction in generative grammar*. Malden: Blackwell.
Schubö, F. (2010). *Recursion and prosodic structure in German complex sentences*. Master thesis. Potsdam, 2010.
Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook, 3,* 371–405.
Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 187–214). Mahwah: Erlbaum.
Selkirk, E. (2000). The Interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 231–261). Dordrecht: Kluwer.
Selkirk, E. (2011). The syntax-phonology interface. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The handbook of phonological theory*, (2nd ed.). West Sussex: Wiley Blackwell.
Sternefeld, W. (2008). *Syntax. Eine morphologisch motivierte generative Beschreibung des Deutschen*. Tübingen: Stauffenburg.
Truckenbrodt, H. (1995a). *Phonological phrases: Their relation to syntax, focus and Prominence*. (Unpublished doctoral dissertation. MIT. Cambridge).
Truckenbrodt, H. (1995b). *Extraposition from NP and prosodic structure*. Proceedings of NELS 25, pp. 503–517.
Truckenbrodt, H. (2006). Phrasal stress. In K. Brown (Ed.), *Theencyclopedia of languages and linguistics* (2nd ed., Vol. 9, pp. 572–579). Oxford: Elsevier.
Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft, 7,* 129–133.
Wiltschko, M. (1997). Extraposition, identification and precedence. In D. Beerman, D. LeBlanc, & H. van Riemsdijk (Eds.), *Rightward movement, Vol 17 of linguistik aktuell* (pp. 357–395). Amsterdam: John Benjamins Publishing.

Prosodic Realizations of Information Focus in French

Claire Beyssade, Barbara Hemforth, Jean-Marie Marandin and Cristel Portes

Abstract In this chapter, we provide empirical evidence on the prosodic marking of information focus (IF) in French. We report results from an elicitation experiment and two perception experiments. Based on these experiments, we propose that phrases that resolve a question are set off by two types of intonational markers in French: they host the nuclear pitch accent (NPA) on their right edge and/or they are intonationally highlighted by an initial rise (IR). These intonational markers are very often realized conjointly but can also be applied separately thus leading to considerable variation in our elicitation data. We will propose that some of the variation can be explained by differences in the function of NPA and IR: NPA placement is sensitive to the informational/illocutionary partitioning of the content of utterances, while IRs are sensitive to different types of semantic or pragmatic salience. We also suggest that "question/answer" pairs provide a criterion to identify the IF only if the answer is congruent. Answers may, however, contribute to implicit questions resulting in different prosodic realizations.

Keywords Information focus · French · Initial rise

This chapter is a largely extended version of a conference proceedings paper by Beyssade et al. 2009. Authors appear in alphabetical order.

C. Beyssade (✉)
Institut Jean Nicod, CNRS-ENS-EHESS, 29, rue d'Ulm, 75005 Paris, France
e-mail: claire.beyssade@ens.fr

B. Hemforth · J.-M. Marandin
Laboratoire de Linguistique Formelle (LLF), CNRS-Université Paris Diderot, Place Paul Ricoeur, 75013 Paris, France
e-mail: barbara.hemforth@linguist.univ-paris-diderot.fr

J.-M. Marandin
e-mail: marandin@linguist.univ-paris-diderot.fr

C. Portes
Laboratoire Parole et Langage (LPL), CNRS-Aix-Marseille Université, 5 avenue Pasteur, 13100 Aix-en-Provence, Francer
e-mail: cristel.portes@lpl-aix.fr

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_3

1 Introduction

The organization of information in a sentence is a central issue in the sentence processing literature as much as in theoretical linguistics. Focusing an element of the current utterance by syntactic or prosodic means contributes to what is perceived as the implicit or explicit question under discussion. The focused element has a high chance of being picked up as the topic of the discourse unit that follows (Dahan et al. 2002). The role of information structure has been studied extensively for phenomena such as clefting, left- or right-dislocation and other syntactic constructions (Colonna et al. 2012; 2014; Drenhaus et al. 2011; de la Fuente and Hemforth 2013). More recently, syntactic realizations of topic and focus are more and more studied in interaction with prosodic realizations using systematic empirical methods (e.g., Repp and Drenhaus in press; Carlson, this volume). A common outcome of these studies is a surprising variability in the way that intonational features such as nuclear pitch accent (NPA) are used across apparently parallel linguistic contexts, but also a variability of choices within individual speakers, and most clearly across languages (Zimmermann and Onea 2011). Only a few languages, however, have been studied in enough detail so far. For most languages intonational means for focus marking are largely understudied from an empirical perspective. Crosslinguistic evidence is, however, indispensible if we want to know which aspects of focus marking are generalizable across languages and which are language specific.

This chapter takes up this issue by investigating the distribution of two intonational markers in French, initial rises (IRs) on the left of focused XPs and NPAs on their right edge. IRs usually take the shape of an H tone on the left but not necessarily on the first syllable of the XP, while NPAs can take a variety of shapes (we will suggest H*, L*, H+L*, and H*+L as possible NPAs). In a question–answer pair like (1), a typical prosodic realization in French includes a pitch rise on the first syllable of Bernadette (the IR, marked by small capitals here) as well as a variant of an NPA on the last syllable (marked by capitals). As we will spell out in more detail in Sect. 2.2, these two accent types are rather different and we will suggest in Sect. 5 that they serve rather different purposes.

(1) A: Qui est-ce que tu as rencontré hier soir?
Who did you meet last night?
B: J'ai rencontré BERnarDETTE hier soir.
I met BernarDETTE last night.

The use of these markers will be investigated in a staged reading aloud experiment. Our results will show that, beyond considerable variation, IRs and NPAs have a tendency to occur jointly in marking the content resolving a wh-question such as *Who wrote the famous paper on implicit prosody in 2002?* (as opposed to the broad question *What happened in psycholinguistics in 2002?*). In two perception studies, we will, however, show that they contribute separately to the marking of an XP as the information focus (IF) of the sentence, i.e., as the answer to a wh-question.

The paper proceeds as follows. We briefly establish our terminology in Sect. 2. In Sect. 3, we describe the corpus obtained via a production experiment and present an analysis assuming the working hypothesis that resolving XPs are information foci. In Sect. 4, we report the results of two perceptual experiments designed to find out whether speakers recognize the two distinct marking strategies observed in the production corpus and relate them to the resolution of questions. In Sect. 5, we present a more comprehensive analysis, which accounts for both intonational marking strategies.

2 Descriptive Framework

2.1 Information Focus

There is general agreement that, beyond all sorts of more detailed variations, IF in the answer to a wh-question in English can be realized roughly as exemplified in the short question-answer pairs in (2) and (3). For these examples as well as for the rest of the chapter, we take IF to be the XP that resolves a question. In question–answer pairs with speakers A and B as in (2) and (3), the XP resolving the question (Bill in 2, Sue in 3) are generally marked by an NPA. Placing the NPA on an XP different from the one resolving the question strongly reduces the felicity of the answer.

(2) A: Who did Paul introduce to Sue?
B: a. Paul introduced BILL to Sue
b. # Paul introduced Bill to SUE

(3) A: Who did Paul introduce Bill to?
B: a. Paul introduced Bill to SUE
b. # Paul introduced BILL to Sue

There is, however, much less consensus about the phonology of IF in French. Assuming that the full sentence is the answer to the broad question in (4) and that *Marie* is the answer to the wh-question in (5), there is broad consensus that these two answers are prosodically different, but much less so with respect to the nature of this difference.

Broad question:
(4) A: Qu'est-ce qui s'est passé?
What happened?
B: [Marie est venue]$_F$
Marie came

Wh-question:
(5) A: Qui est venu?
Who came?
B: [Marie]$_F$ est venue.

At first glance, the phenomenology of the prosodic/intonational realization of resolving XPs in answers is actually varied. As stated by Jun and Fougeron (2002) as well as Fonagy et al. (1979), some aspects of prosodic marking are optional and seem more probabilistic in nature. This variation can be partly, but not fully explained by structural parameters. The segmental structure and the length of the accentual (or phonological) phrase yield different surface realizations of an underlying prosodic structure as well as rhythmic constraints such as constraints on stress clashes. IRs can, for example, be found on the second instead of the first syllable when the first one is a function word (*le* MAUvais *garÇON ment à sa mère*, the bad boy lies to his mother, Jun and Fougeron 2000) or not at all when the prosodic unit (the accentual phrase (AP)) is only one or two syllables long (*Non, MaRIE est arrivée.* No, Mary has arrived.) (see Sect. 2.2 for details on the segmental structure assumed for French). In this chapter, we will discuss the question of how far beyond these structural parameters, parts of the systematicity in the prosodic variation is related to IF in French. Jun and Fougeron (2002) admit that part of the variation may be due to meaning (e.g., semantic importance) and information structure (e.g., contrastive focus). We will claim that at least some of the diversity can be explained by the interplay of the two distinct marking strategies introduced before: the placement of the NPA in the utterance and the intonational highlighting (IR) on the left of phrases. We will, moreover, suggest that these two strategies may be variants related to the semantic/pragmatic status of IF, differentiating the status of being specifically asserted and that of being salient in the content conveyed in the assertion.

In the rest of this chapter, we will report the results of three experiments that contribute evidence relevant to the choice between the competing descriptive or analytical claims currently debated. We assume that the question/answer pair yields a criterion to identify the IF in utterances: the IF is the part of the content of answers that resolves the question. We put this definition to use in the design of several experiments whose results are presented here.

2.2 Terminology for the Question/Answer Pair

Let's consider the two dialogues (6a) and (6b), involving discourse participants A and B. As before, we call the question in (6a) a broad question and that in (6b) a wh-question.

(6) a. A: What happened? B: [Jean invited Marie to the party last night.]$_F$
b. A: Who did Jean invite? B: Jean invited [Marie]$_F$ to the party last night.

In (6a), the resolving XP (R-XPs for short henceforth) is the whole sentence; in (6b), it is the Object NP. Under the assumption that IF is the part of content that resolves the question, the IF is contributed by the whole sentence in (6a) and by the Object NP in (6b). Answer (6a) is an answer to a broad question and (6b) an answer to a wh-question (a. o. Lambrecht 1994; Vallduví and Engdahl 1996). It must be kept in mind that the equation "R-XP = IF" is only valid in congruent answers; congruent answers are answers that strictly convey a value for the parameter introduced

in the question (Krifka 2001; Kadmon 2001); i.e., they give a precise answer to the question explicitly asked. This excludes over- or underinformative answers of any type. This limitation will turn out to be important for the comprehensive analysis of the data we will present in Sect. 5.

It is usually assumed in the literature that resolving the question is an appropriate criterion for IF because it is a criterion for the newness of the content it contributes. The notion of new (vs. old) is, however, notoriously vague. Here, we take it that "new" means the content the speaker proposes for updating the part of the Common Ground under discussion. Accordingly, new is closely linked to the working of the assertion in declaratives: what is new is the part of the content that is specifically asserted by the Speaker (Jacobs 1984). We strictly restrict ourselves to the question/answer pair here. We do not consider corrections or denials which bring about contrastive content (as, e.g., in Jun and Fougeron 2000, or Dohen and Loevenbruck 2004). We assume that the intonational correlates of contrast are different from those of IF (Beyssade et al. 2004; Selkirk 2009).

2.3 Prosodic Framework

Our analysis borrows some basic ideas from the Aix-en-Provence school (Di Cristo 1999; Rossi 1999) but is couched in the autosegmental-metrical framework (AMT, Post 2000; Jun and Fougeron 2000, 2002). There is consensus that French intonation has at least two levels of phrasing: the AP, also called phonological phrase, and the intonational phrase (IP) (see Example 7 taken from Jun and Fougeron 2000, p. 215, for the partitioning of a sentence into APs). Moreover, several frameworks assume a third level of intermediate phrase (ip); its relevance has been argued by Michelas and D'Imperio (2012) for reasons independent of focus marking.

(7) Le désagréable garçon ment à sa mère. 'The unpleasant boy lies to his mother'
{L Hi L H*}{ }

The AP is structured by two tonal events: an IR LHi and a final rise LH*. The surface realization of these tones, however, varies depending on different factors such as the number of syllables of the AP and the speech rate, giving rise to the following surface patterns: LHiLH*, LH*, LLH*, LHiH*, and LHiL* (Jun and Fougeron 2002). IPs are marked by boundary tones that may be low L% or high H%.

In Jun and Fougeron (2002), the LHiL* pattern surfaces when the H* cannot be realized due to undershoot. The inventory of possible pitch accents is therefore H* or L* in their proposal. In order to account for the tonal phenomenology at the right edge of the AP, we assume two more pitch accents borrowed from Ladd (2008, p. 122).[1] These are bitonal pitch accents: H+L* accounts for patterns where the pitch peak occurs on the penultimate preaccentual syllable, H*+L codes for a rising–falling movement on the last accented syllable. Therefore, our inventory of

[1] For further arguments in favor of this coding, see Portes and Beyssade (to appear).

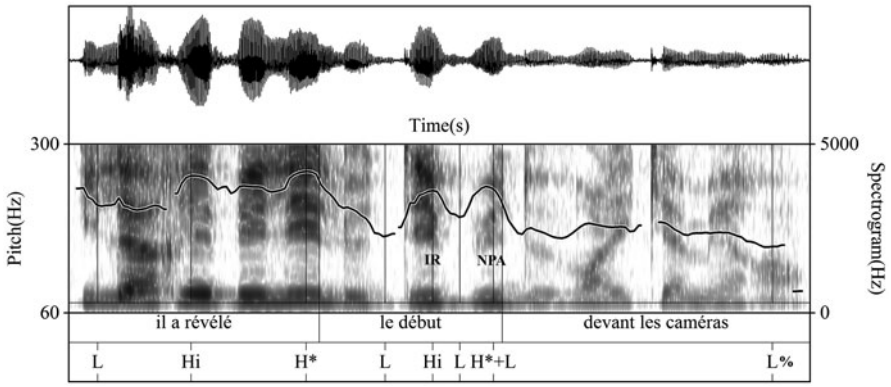


Fig. 1 Answer to a wh-question with the direct object "le début" (the beginning) as the resolving XP in the sentence "il a révélé le début devant les caméras" (he revealed the beginning in front of the cameras)

pitch accent occurring in nuclear position (i.e., as the last pitch accent in the IP) is the following: H*, L*, H+L*, and H*+L.

Several authors have observed that narrow information foci or contrastive foci are followed by "deaccented" phrases up to the end of the utterance: they are intonationally realized with high or low constant pitch depending on the height of the previous tonal target.[2] Here, we model deaccenting through the spreading of the boundary tone, which copies the final tone of the NPA: it is an H% when the NPA ends with an H tone (H*) and an L% when the NPA ends with an L tone (L*, H+L*, and H*+L).

Therefore, in case of deaccenting, the NPA (H*+L in Fig. 1) is moved back to the right edge of the focused phrase while the boundary tone spreads through the deaccented following phrases. The global contour is, thus, preserved as well as its dialogical meaning (Beyssade and Marandin 2007; Portes and Beyssade to appear).

Moreover, several authors have proposed that the optional IR LHi should play a role in the marking of IF. Di Cristo (1999) claims that the IR is more often realized at the left edge of the focused constituent. In this case it may surface with a wider pitch range, giving rise to a specific "accent emphatique" ou "accent de contraste" (emphatic accent or contrastive accent).[3] German and D'Imperio (2010) also found that LHi is more likely to occur at the left edge of a contrastive focus domain, without mentioning any scaling differences. In this study, we assume the tonal marking of the left edge of the focused constituent under the descriptive label "IR." It has been observed that it may form an "accentual arch" with the following rising accent LH*(or LH+L*, or LH*+L), or trigger a high plateau up to the following accent

[2] See, however, Féry (2014) for evidence that deaccenting or compression may be restricted to adjuncts. (see Di Cristo & Jankowski 1999, for an analysis in favor of compression). Since post-focal elements in our study are never arguments, this distinction does not apply here.

[3] Concerning the phonetic realization of the initial rise, we refer the reader to Astésano (2001), which is the most comprehensive and detailed approach to our knowledge.

when the intermediate L tone is not realized. The IR or the high plateau may be implemented quite high in the pitch range.

2.4 Focus Marking

Following our brief description of the prosodic framework, we assume Beyssade et al.'s (2004) analysis of IF marking: the phrase contributing the IF hosts the NPA on its right edge. This assumption follows Di Cristo's proposal that the right edge of XPs contributing the IF provides the site for anchoring the nuclear accent. According to Di Cristo, the nuclear accent is a low tone in declarative sentences. Beyssade et al. take up Di Cristo's claim and generalize it: on the basis of corpus observations, they claim that the right edge of focal XPs may anchor the whole repertory of NPAs in French (H*, L*, H+L*, and H*+L), as the choice of the NPA is independently determined by dialogical parameters. Such a claim directly explains some of the variability of the occurrence of the nuclear contour: as it occurs at the right edge of the phrase contributing the IF, we expect it within the utterance when IF is narrow and the focused XP is non-final and at the end of the utterance when IF is broad. In both approaches, IF marking is identical for IF in wh- and broad questions: in the former case, IF is contributed by a phrase while it is contributed by the whole sentence in the latter. Moreover, Di Cristo and Beyssade et al. also observe that an IR may occur on the left syllable(s) of the phrase conveying narrow IF. Di Cristo proposes that IR marks the left edge of the narrow focal XP: he speaks of bilateral marking of Focus. As for Beyssade et al., they speculate that IR can be related to contrastive focus (following Rossi 1999; see Experiment III for a more detailed discussion of this proposal).

The goal of this chapter is to test the following claim

- IF marking in French resorts to two means: placement of the NPA and IR.

Furthermore, we want to investigate the interplay of these two types of marking. Finally, we address the question of whether this double strategy is functionally equivalent or associated with distinct roles for the discourse.

3 Production and Comprehension of Prosodically Marked XPs

3.1 Experiment I: Elicitation of IF in a Staged Reading Aloud Task

Our first experiment used staged reading aloud as the experimental paradigm. We thus created a corpus of answers to broad and wh-questions.

Methods

Participants Fourteen participants from the University of Paris Descartes volunteered to take part in this study:[4] ten of them were psychology students who received course credits for participation and four were psychological staff. None of the participants had any training in linguistics. Participants were naïve with respect to our research question.

Materials The corpus of answers we analyze here has been elicited via a production experiment. In the full corpus, we varied not only broad and wh-questions, but also prosodic/intonational realizations of the associate of the adverb *seulement (only)* (Beyssade et al. 2008). All in all, we created 32 sets of items such as (9) within short contexts such as (8) with eight variants each. The eight variants were distributed across eight lists following a Latin square design. Only two of the conditions directly relating to the question of IF with four items per condition per participant will be discussed in this chapter. In these two conditions, we varied the question type: wh-questions (bearing on the direct Object) (9a) and broad questions (bearing on the whole sentence) (9b). Importantly, all entities in the answer were introduced in the context, including the direct Object, which was introduced as part of a set of alternatives. For wh-questions, the XP following the Object was always introduced in the question thus making it impossible to consider it as part of IF or as additional information related to IF (in the sense of a *side* structure, Klein and Stutterheim 2002).

Procedure The short texts, involving a description of the context such as (8) were presented to the subjects visually as well as auditorily with one of the two types of questions: a wh-question (bearing on the Object) (9a) or a broad question (bearing on the whole sentence) (9b). The subjects' task was to read aloud answers as if they were actually participating in a dialogue. The participants' answers were recorded in a sound-attenuated room. The 32 target items were pseudorandomly interspersed with 32 fillers, partly from a different experiment on the prosody of coordinations. Four practice items were presented to familiarize participants with the experimental set-up. Out of the 112 (8 * 14) answers we recorded, we only analyzed 107: Five answers were not taken into account in our quantitative analyses because of disfluencies or production errors.

(8) Context [translated]: Richard is a policeman. He has to treat various documents (videos, leaflets, K7s) seized in a terrorist cache.

(9) a. Le responsable: Qu'as-tu visionné la nuit dernière?
What did you screen last night?
Richard: J'ai visionné les vidéos la nuit dernière.
I screened the videos last night
b. Le responsable: Où en es-tu dans ton enquête?
What's up with your investigation?

Richard: J'ai visionné les vidéos la nuit dernière.
I screened the videos last night

[4] We actually recorded four more participants whose data we could not include because they turned out to be bilingual or very disfluent readers.

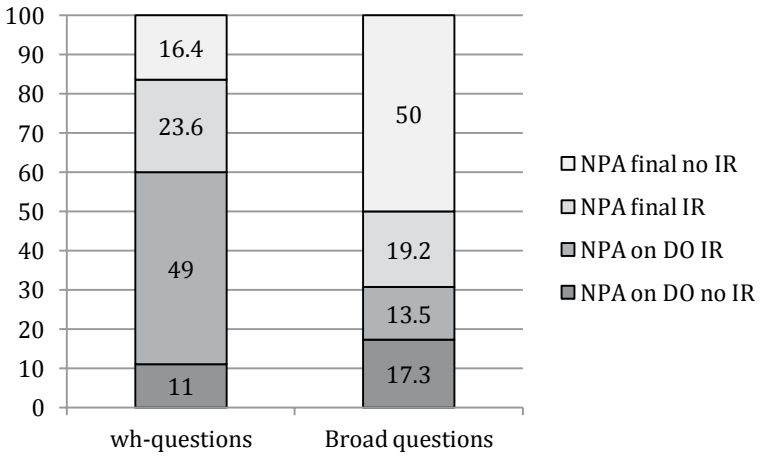


Fig. 2 Prosodic realizations for wh- and broad questions in percent

3.2 Results

All answers were coded blindly by the three native French authors without access to the experimental conditions in which they had been produced. In cases of disagreement, the respective answers were discussed in joint meetings until agreement was reached. All answers were coded for IRs on the Object, NPAs on the right edge of the Object, and sentence final NPA. Four different contours were identified: sentence final NPA with IR on the Object, NPA on the right edge of the Object with IR on the left of the Object (not necessarily on the left edge), NPA on the right edge of the Object without IR, and sentence final NPA without IR. We will first look at the general distribution of the different contours for the two question types before presenting NPA and IR distributions separately in more detail. Figure 2 shows the distribution of the different contours for wh- and broad questions. The different contours are obviously not equally distributed (Chisquare(3)=38,83, $p<0.001$). While the production of final NPA plus IR on the Object was not significantly different across conditions ($p>0.40$), NPA on the Object plus IR was clearly much more dominant for wh-questions (Chisquare(1)=19.44, $p<0.001$). No significant difference was established for NPA on the Object without IR ($p>0.20$). Final NPA without IR was, however, significantly more frequent for broad questions (Chisquare(1)=17.515, $p<0.001$).

Looking separately at the occurrence of NPA and IR, we can see that NPA was produced significantly more often sentence finally for broad questions than for wh-questions (Chisquare(1)=7.72, $p<0.01$), while the inverse was true for NPA on the right edge of the direct object (Chisquare(1)=9.24, $p<0.01$) (Fig. 3).

IR on the left of the direct object was realized significantly more often for wh-questions than for broad questions (Chisquare(1)=15.01, $p<0.001$).

Summing up, we can say that the Object noun phrases in answers to wh-questions are distinguished in three different ways:

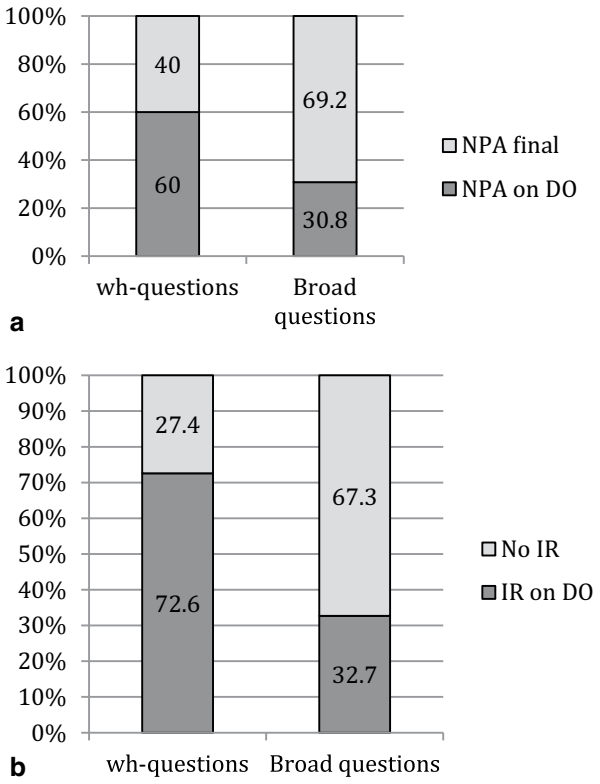


Fig. 3 **a** Realization of NP for wh- and broad questions. **b** Realization of initial rise (IR) for wh- and broad questions

(10) a. The direct Object hosts the NPA on its right edge with an IR (initial rise) on the left (Fig. 4);
b. The direct Object hosts the NPA on its right edge without IR (Fig. 5);
c. The direct Object shows an IR, while the NPA occurs at the end of the utterance (Fig. 6).

Pattern (10a) conjoins the placement of NPA and IR. It is the most frequent pattern with 49% of all answers. NPA placement and IR appear separately in the two other patterns (10b and 10c). Pattern (10b) features the placement of the NPA on the Object with the corresponding deaccenting of the PP to the right.[5] It is the least attested pattern (11% of the all answers). Pattern (10c) highlights the Object, while the NPA occurs at the end of the utterance. Crucially, the PP to the right of the Object is not deaccented. This pattern is well represented in the corpus: 23.6% of all answers.

[5] See, however, Féry (2014) who shows that deaccentuation does not necessarily occur in postfocal regions. It does at least not seem to be obligatory for verbal arguments. Since postfocal elements in our study are never arguments, this distinction does not apply here.

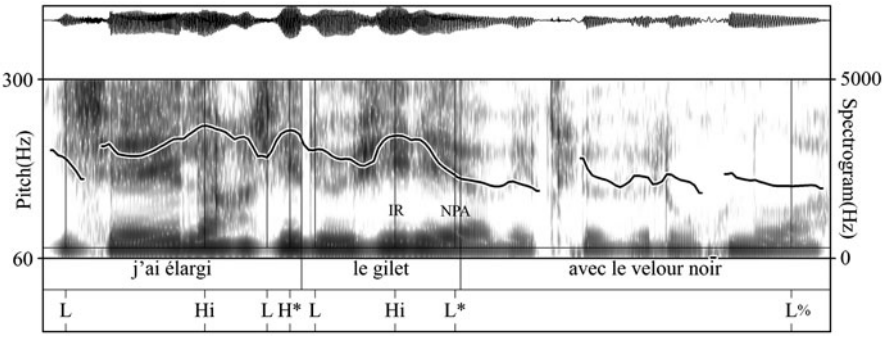


Fig. 4 Answers with pattern 10a: initial rise (IR) on direct Object (with a high implemented initial accent) and Object-final nuclear pitch accent (NPA)

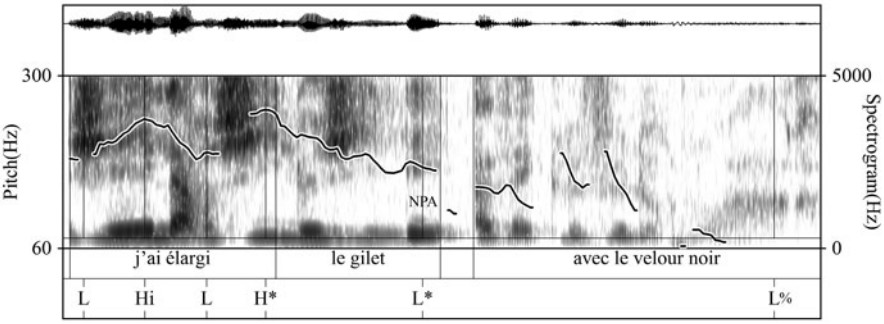


Fig. 5 Answers with pattern 10b: Object-final nuclear pitch accent (NPA) and no initial rise (IR)

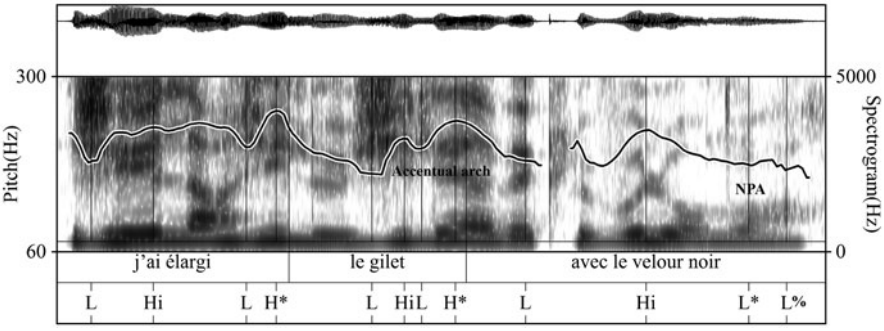


Fig. 6 Answers with initial rise (IR) on the direct object (realization of an accentual arch Hi-H*) and utterance-final nuclear pitch accent (NPA) (10c)

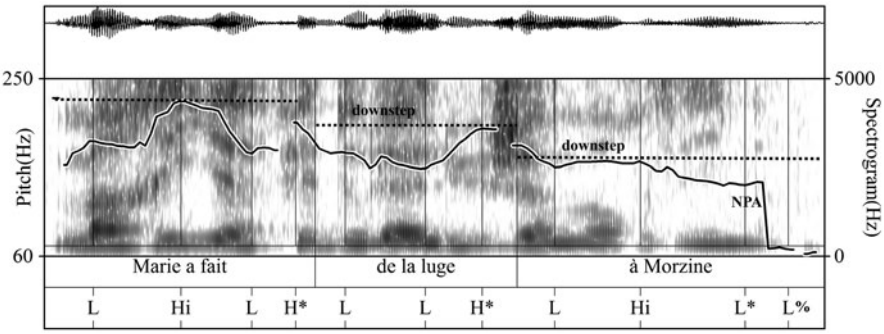


Fig. 7 Answers with utterance-final nuclear pitch accent (NPA) and downstep of the second and third accentual phrases (Aps). Downstep is modeled as a reference base line defined by the H targets (*dashed line* in bold) as proposed by van den Berg et al. (1992)

Finally, there are 16.4 % of the answers in which the Object is not set off by any means: we come back to them in Sect. 5 below.

Answers to Broad Questions

Let's now look at answers to broad questions: 69.2 % of the answers to broad questions show pattern (11):

(11) NPA occurs at the right edge of the utterance (NPA is utterance final).

Pattern (11) mostly gives rise to a regular downstep of the APs following the initial AP (Fig. 7). No constituent is highlighted: no high implemented initial accent (IR) can be seen. This pattern corresponds to 50 % of all answers to broad questions. The remaining answers feature one of the patterns described in (10) for answers to a wh-question. 30.8 % of all answers show the NPA on the right edge of the Object, which corresponds to patterns (10a = with IR: 13.5 %) and (10b = without IR: 17.3 %). Moreover, 19.2 % of the answers with the NPA on the right edge of the utterance feature a highlighted Object, which corresponds to pattern (10c). We come back to those two cases in Sect. 5 below.

NPA Contours

In our corpus, we found several types of NPA contours at the right edge of IF, which corroborates Beyssade et al.'s (2004) generalization. Three types of nuclear pitch movement are attested in the corpus:

1. Falls (corresponding to Di Cristo's B or Beyssade et al.'s L*) (Figs. 4, 5 and 6) above.
2. Falls from the penultimate, which corresponds to Ladd's (2008) H + L*: the pitch peak occurs on the penultimate syllable and the following valley on the last syllable. This is illustrated for IF in an answer to a wh-question in Fig. 8.

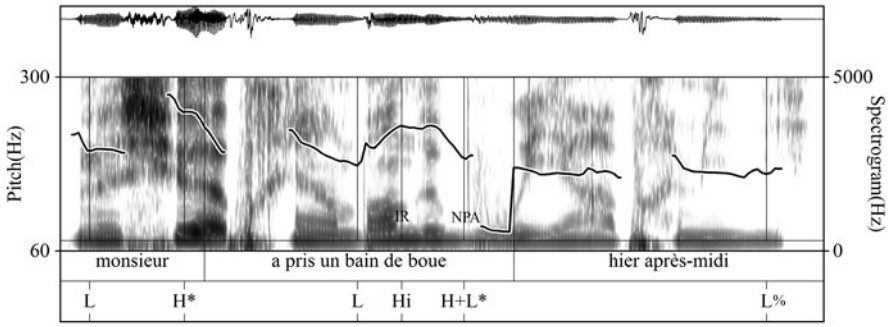


Fig. 8 Answer to a wh-question with a fall from the penultimate (H+L*) nuclear pitch accent (NPA) occurring at the right edge of the focused Object "bain de boue." Note that an initial accent occurs on "bain" immediately followed by the leading tone H+of the H+L* pitch accent on the penultimate syllable "de" which contains a schwa.

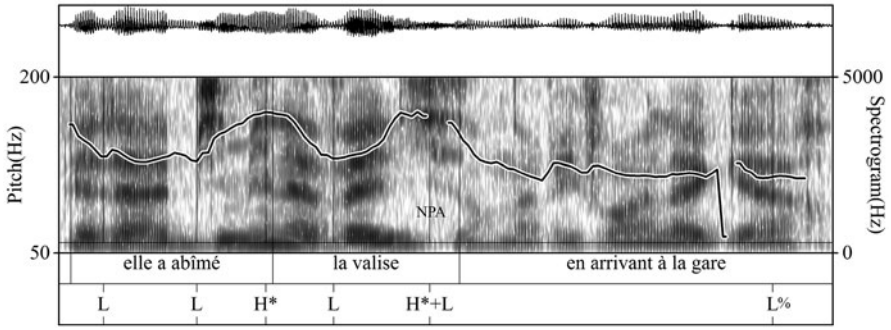


Fig. 9 Answer to a wh-question with a rise–fall (H*+L) nuclear pitch accent (NPA) occurring at the right edge of the focused Object NP "la valise".

3. Rise–falls (H*+L in Ladd 2008 and Portes and Beyssade to appear) for which the pitch peak and the following valley occur on the last syllable as illustrated for IF in answers to wh-questions in Fig. 9.

3.3 Discussion

We first analyze the patterns we observed in the data assuming the working hypothesis that the resolving XPs (R-XPs) are IFs and the intonational approach to IF marking as defined in (11) as proposed by Di Cristo (1999) and Beyssade et al. (2004).

(12) XPs contributing the Information Focus host the Nuclear Pitch Accent on their right edge.

Claim (12) is corroborated in the majority of the cases in our production experiment: 60 % of the answers to a wh-question show the NPA at the right edge of the Object whereas 69.2 % of the answers to a broad question show the NPA on the right edge of the utterance. Three different NPA contours were moreover found in our corpus (falls, falls from the penultimate, and rise–falls) as predicted by Beyssade et al. (2004).

Nevertheless, there are facts that do not fit the picture predicted by (12) and call for another analysis: 72.6 % of the answers to a wh-question show an IR on the object, which is compatible with, but not predicted by (12). Among them, 23.6 % show only the IR on the Object, while the NPA is docked at the right edge of the sentence (corresponding to pattern 10c) for the remaining 49 % the NPA is on the right edge of the direct object (pattern 10a).

We propose the hypothesis in (13) to account for the use of IR in answers:

(13) The XP resolving a narrow question may be marked by NPA placement or by IR.

We devote the next section to the corroboration of (13).

4 The Role of the IR

We ran two perception experiments in order to test hypothesis (13). In Experiment II, we are testing whether IRs alone can be recognized as a way of marking the XP resolving a question. In Experiment III, we asked whether IR is linked to the expression of Contrast (as suggested by Rossi 1999 and taken up by Beyssade et al. 2004).

4.1 Experiment II

Methods

Participants The experiment involved 24 participants, native speakers of French, first-year undergrad students in Humanities at U. Paris Diderot. All participants were naïve as to the research questions of our experiments.

Materials and Procedure We selected 20 answers from the preceding corpus, all of them particularly clear examples of the different intonational contours we wanted to examine: ten realizations with NPA at the end of the sentence and no IR that are supposed to be identified as answers to broad questions, ten with marking of the Object (five with NPA and five with IR only) which, conversely, are predicted to be identified as answers to wh-questions. The sentences were presented in two blocks. The first block is composed of five answers with final NPA (hypothesized answers to broad questions) and five sentences with final NPA and IR on the object NP (hypothesized answers to wh-questions). The second block is composed of five answers with final NPA (expected answers to broad questions) and five answers with NPA on the object NP (expected answers to wh-questions). The ten sentences

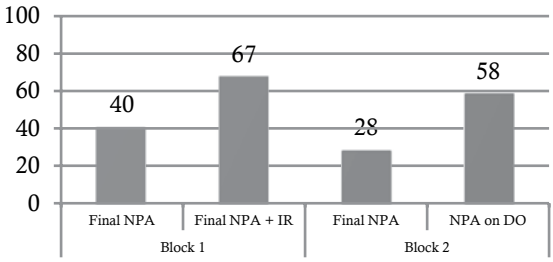


Fig. 10 Results of Experiment II. Percentages of wh-question choices associated with each prosodic pattern

composing each block were presented in random order. The subjects had to listen to the selected items and to judge to which of two visually presented questions the current sentence had been produced as an answer (14). Each session was run in a quiet room within the Paris Diderot library, where we recruited our participants. The sessions lasted at most 15 min, which made recruiting voluntary participants relatively easy. We also kept the experiment short to avoid habituation effects.

(14) Questions: 1. Pour finir, qu'est-ce que tu as élargi avec du velours noir?
Finally, what have you let out with the black velvet?
2. Pour finir, tu t'en es sorti comment?
Finally, how did you get by?

Answer: J'ai élargi le gilet avec du velours noir.
I let out the vest with black velvet

Results

Figure 10 shows how often participants chose wh-questions as relevant for the heard answer. Participants clearly distinguished answers with Final NPA and answers with highlighted Objects (IR on direct Object (DO)) in block 1, as well as between with NPA at the end (Final NPA) and answers with NPA at the right edge of DO (NPA on DO) in block 2. They chose the wh-questions reliably more often for answers with IR on DO than for answers with final NPA (67 vs. 40 %; $F(1,24) = 19.54$; $p < 0.001$). They also chose the wh-question reliably more often for answers with NPA on DO (58 %) than for answers with final NPA (28 %, $F1,24 = 23.93$; $p < 0.001$). No reliable difference between answers with IR on DO and those with NPA on DO could be established.

Conclusion

The results of Experiment II corroborate our Hypothesis (13): utterances with NPA on the direct Object or with IR on the direct Object are similarly recognized as answers to wh-questions bearing on the direct Object. However, the data also show

that the intonational marking does not lead to unambiguous interpretations. Sentence final NPA is still considered as compatible with a wh-question in 34 % of the cases on average (40 % in Block 1 and 28 % in Block 2), and IR or NPA on the direct Object is considered as compatible with a broad question in 37.5 % of the cases on average (33 % in Block 1 and 42 % in Block 2). In the following section, we will present data on the role of sets of alternatives for IF in the preceding context.

4.2 Experiment III

The presence of IR in our production data concerns 72.6 % of all responses to wh-questions. Looking for an explanation, we linked this massive occurrence to the systematic presence of a set of alternatives in all of the eliciting contexts (see, for example, "films, leaflets, K7" in (8) above). We, thus, designed a second perception experiment in order to test the hypothesis that IR is related to the expression of Contrast as formulated by Rossi (1999, see also Beyssade et al. 2004). We define the notion of contrast as a membership relation in a set of alternatives activated in the immediate context (Chafe 1974).

Methods

The only difference between Experiments II and III is that we added a sentence presenting a set of alternatives in the description of the context before the presentation of the question. Otherwise, the procedure was identical. For example, context (15) in which the phrase "le gilet et la veste" corresponds to a set of two possible choices has been added to (14). If the presence of a set of alternatives plays a central role for the presence of IR, we expect the choice of wh-questions to increase for sentence with IR on the direct Object compared to Experiment II. The experiment involved 17 participants, native speakers of French, first-year undergraduate students in Humanities at U. Paris Diderot, who had not participated in Experiment II. The experiment was run using the same procedure and under the same circumstances as Experiment II.

(15) Pierre ne rentre plus dans son costume: le gilet et la veste sont trop serrés. Comme il est tailleur, il va faire les retouches.
His suit does not fit Pierre any longer: the vest and the jacket are too tight. As he is a tailor, he will alter them.

Results

Figure 11 shows the percentage of wh-questions chosen by participants to be consistent with the heard answer. The pattern is nearly identical to that of Experiment II. The 17 subjects chose the wh-question reliably more often for answers with

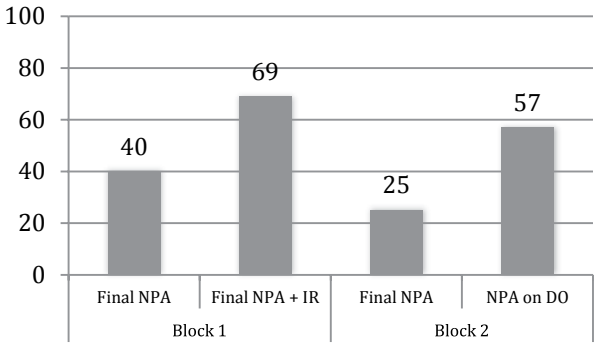


Fig. 11 Results of Experiment III: percentages of wh-question choices associated with each prosodic pattern

IR on NP (67%) than for answers with final NPA (69 vs. 40%; $F(1, 17) = 8.86$, $p < 0.01$). They also chose the wh-question reliably more often for answers with NPA on NP than for answers with final NPA (57 vs. 25%, $F(1,17) = 5.12$, $p < 0.04$). No reliable difference between answers with IR on NP and those with NPA on NP could be established.

Discussion

The presence of alternatives in the immediate context does not influence the choice of the question types corresponding to different intonational contours of the R-XPs. The results were actually nearly identical to those of Experiment II.

4.3 Conclusion of Perception Experiments

Both experiments show that speakers recognize the highlighting of the Object as a cue to its saliency in the answer as it resolves the question. Accordingly, we conclude that hypothesis (13) is corroborated. Participants did, however, interpret the intonational marking fairly directly and independent of the sets of alternatives provided in the context.

5 General Discussion

In our elicitation and perception experiments, we were able to show the variable marking of IF in French. Direct objects serving as answers to wh-questions can be marked by IRs on their left, by NPAs on their right edge or both. IR and NPA on the direct object are independently perceived as cues for IF although they most often occur conjointly. IR can even occur on the direct object in answers to broad questions. This observed variation in the data can of course be just evidence for variation

with not much more to add. We would, however, like to propose that some of the variation may be explained by the assumption that NPA placement and IR do not necessarily cue the same phenomenon. We will propose that NPA placement is sensitive to the illocutionary import of the content of the utterance, while IR is considered a polyvalent means to give intonational prominence to the content of a phrase.

5.1 Background: "Congruent" vs. "Noncongruent" Answers

In Sect. 2.1, we took up the accepted distinction between congruent vs. noncongruent answers. The equation between IF and resolving XP (R-XP) holds only in congruent answers. But, we know that in naturally occurring contexts, dialogue participants quite often answer in a noncongruent way: they contribute underinformative or over-informative answers (Krifka 2001). This is easily explained by reasons of cooperation or default of cooperation. Speakers infer the current question under discussion, which may be the explicit question but can as easily be an implicit underlying question (Ginzburg 1995a, b; Roberts 1996; see also Clifton and Frazier 2012 for processing evidence). For example, it is very common that speakers offer apparently over-informative answers anticipating underlying reason for the question on the part of the questioner. This is the case with over-informative answers in (16) and (17) below: in (16), the speaker does not produce a direct answer to the polar question "Est-ce que quelqu'un t'a contacté?", but she produces an answer to the wh-question "Qui t'a contacté?", and this answer implies that the answer to the polar question is positive. In (17), the answer resolves the question and contributes more precise information about the issue raised by the question.

(16) A: Est-ce que quelqu'un t'a contacté? B: Bernadette.
Did someone contact you? *Bernadette did.*

(17) A: Qui t'a contacté? B: Bernadette m'a envoyé un mail.
Who contacted you? *Bernadette sent me an email.*

A case of underinformative response is given in (18): the answer does not resolve the question, while it contributes relevant information about the question.

(18) A: Qui t'a contacté?
Who contacted you?
B: II n'y a pas eu d'appel.
There was no call.

To recapitulate, discourse participants—when they answer—do not simply resolve the explicit question of the interlocutor; they have their own agenda and the answers they offer are a trade-off between what is required by the interlocutor's question, what they think is required and which information they are able/willing to give. In experiments, in the lab, one does not control that aspect of the answers all that well, nor do we necessarily do so in natural dialogues. Accordingly, we do expect that not all answers we have elicited are answers to the explicitly asked questions.

5.2 Proposal

Phrases that resolve a question (be they a constituent in a clause or the entire sentence) have a double status: a semantic status in that they resolve the question, but also a pragmatic status in that they contribute the new content, viz., that part of content that makes up the update brought forth by the assertion.

It is currently assumed that those two statuses are interdependent and coincide. They certainly do in congruent answers. Now, part of the working of noncongruent answers can be explained by the fact that they can be dissociated. For example, in (17), *Bernadette* resolves the question while the whole answer contributes the update brought over by the answer. If the statuses can be teased apart, their cueing possibly can be too. Hence, we propose that:

(19) NPA placement cues the part of the content that contributes to the update brought by the answer.

(20) [Provisory] IR cues the constituent that resolves the question.

We will use the label "pragmatic marking" for (19) and "semantic marking" for (20). The proposal in (19) is just a reformulation in dialogical terms of Jacobs' 1984 definition of free focus (see also Beyssade et al. 2004). In terms of the contrast "new vs. old" relativized to the working of the assertion, only the NPA placement is sensitive to the newness of the content.

We are now in a position to account for the distribution of the patterns we observe in the corpus including the answers that at first blush do not abide by our hypothese (13: The XP resolving a narrow question may be marked by NPA placement or by IR.).

5.3 Analysis of Answers to a Wh-Question

Assuming (19) and (20), the analysis of patterns (10) can be made explicit for answers to a wh-question:

- Pattern (10a: The Object hosts the NPA on its right edge and an IR on its left) conjoins both the semantic and pragmatic markings.
- Pattern (10b: The Object hosts the NPA on its right edge without an IR) only marks the pragmatic update.

Accordingly, the intonation of answers in pattern (10a) and (10b) fits the working of the question–answer pair: they are intonationally congruent.

- Pattern (10c: The Object shows an IR, while the NPA occurs at the end of the utterance) disjoins the statuses: the semantic relation is marked while the whole content is presented as making up the update of the answer.

Accordingly, the intonation of answers in pattern (10c) is partly noncongruent.

Finally, 16.9% of the answers that we left aside in Sect. 3 feature pattern (11): No IR and the NPA occurs at the end of the sentence. As such, the intonation does not cue the semantic relation holding with the question and they sound like All Focus answers. They make up a clear case of intonational noncongruence. This is probably why there are so few of these patterns in the corpus.

5.4 Analysis of Answers to a Broad Question

At first blush, the analysis of answers to a broad question should be simpler, since only the placement of NPA is relevant: we expect NPA at the end of the sentence, which corresponds to pattern (11: NPA occurs at the right edge of the utterance). And indeed, 69.2% of the answers in the corpus show pattern (11).

We left aside 30.8% of the answers in Sect. 3. They show NPA at the right edge of the Object, which indeed corresponds to patterns (10a) or (10b), which we observed for answers to a wh-question. In other words, those answers are intonationally realized as answers to a wh-question. As such, they make up a case of intonational noncongruence. Their number in the corpus is relatively high. We may speculate that is in line with a tendency observed in naturally occurring contexts: speakers tend to offer answers which are more informative than those that are required by polar or broad questions. Such a speculation will have to be consolidated by experimental evidence.

5.5 Reanalysis of IR

Now, we observe that 19.2% of the answers to a broad question show a highlighted Object while the NPA is at the right edge of the sentence, which corresponds to pattern (10c). According to (20), we should analyze them as resolving a question. Assuming a hierarchical model of dialogue *à la* Büring (2003) or Roberts (1996), we could posit a covert intermediary question as we did in the informal analysis of (16). But, this is not necessarily the intuition triggered by those answers. One of the more received views on initial rise is that of a marker of empathy as it can be found in exclamations or emphatic expressions more generally ("C'est MERveilleux!" This is wonderful! "Je le DEteste!" I hate him!) (Féry 2001; Grammont 1933). Correspondingly, the intuition is that IR in those answers may have an expressive flavor: a marker of empathy with an element of the content (21a) (Kuno 2004). It may also be used as a centering marker for the discourse topic to come (19c).

(21) Qu'est-ce qu'il s'est passé?
a. Martine a abimé la valise à la GARE.
a. Martine has damaged the suitcase at the train station.
b. Je lui avais recommandé de prendre un sac à dos.
b. I had told her to take a backpack.
c. C'était ma valise préférée.
c. It was my favorite suitcase.

If this intuition is correct, we would have to generalize (20) into (22):

(22) IR sets off a constituent that is salient at the semantic or pragmatic level.

Claim (22) means that IR is functionally underspecified. Its specific function, probably some sort of semantic or pragmatic distinction needs to be specified by the context. Resolving a question would be one among other prominent statuses of phrases. These hypotheses are for the moment primarily based on our intuitions and clearly need to be further tested by corpus studies and experiments. However, Beyssade et al. 2008 observed that IR is also used to cue the associate of the restrictive adverb *seulement* ("only"). However, the results of Experiment III prevent an analysis of IR as a marker of Contrast (i.e., membership in an activated set of alternatives): IR is most probably compatible with Contrast, but not a Contrast marker. According to (22), its use with associative adverbs would precisely be to set off the phrase that plays the role of associate.

6 Conclusion

We have identified three sources of variations in the marking of Informational Focus in French. Firstly, there are two strategies to mark the IF of an utterance: initial rises (for narrow IF) and NPA placement (for both narrow and broad IF). Secondly, each strategy has its own phonotactic and pragmatic constraints (that are independent from Focus marking): they account for most of the surface variations. Thirdly, the partition of utterance content into Ground and Focus is not deterministically fixed by the context: it crucially depends on the choice of the Speaker. This is particular true when discourse participants answer questions. In the last part of the paper, we have proposed that IRs and NPA placement are not specialized for the marking of IF. In short, they are not focus markers.

Placement of NPA in the utterance (most often correlated with deaccentuation of XPs to the right) and IRs are two ways of setting off a phrase in French. Both are used in answers, but with different roles. NPA placement marks the part of content that is specifically asserted, which counts for the new content with respect to the working of the assertion. In that respect, placement of NPA is the primary way of marking what is new in answers, and more generally in assertions. On the other hand, IR sets off a phrase for a variety of semantic or pragmatic reasons. It may be used to mark a phrase that resolves the question, thus cueing the semantic relation between questions and answers, but also a phrase endowed with other discourse roles, in particular with respect to the generation of the discourse topic.

References

Astésano, C. (2001). *Rythme et Accentuation en Français: Invariance et Variabilité Stylistique*. Collection Langue et Parole, Recherches en Sciences du Langage, dirigée par Henry Boyer, Editions L'Harmattan, Paris, 337 p.

Berg, R., Gussenhoven, van den, C., & Rietveld, A. (1992). Downstep in Dutch: Implications for a model. In G. J. Docherty & D. R. Ladd (eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 335–359). Cambridge: Cambridge University Press.

Beyssade, C., & Marandin, J.-M. (2007). French intonation and attitude attribution. In P. Denis, E. McCready, A. Palmer, & B. Reese (eds.), *Proceedings of the 2004 Texas Linguistics Society Conference: Issues at the Semantics-Pragmatics Interface*.

Beyssade, C., Delais-Roussarie, E., Doetjes, J., Marandin, J.-M., & Rialland, A. (2004). Prosody and Information in French. In F. Corblin & H. de Swart (eds.), *Handbook of French semantics* (pp. 477–499). Stanford: CSLI.

Beyssade, C., Hemforth, B., Marandin, J.-M., & Portes, C. (2008). The prosody of restrictive seulement in French. *Third TIE Conference on Tone and Intonation*. Barcelone (pp. 15–17), September 2008.

Beyssade, C., Hemforth, B., Marandin, J.-M., & Portes, C. (2009). Prosodic Markings of Information Focus in French. Interface Discours & Prosodie Paris, 9–11 Septembre 2009.

Büring, D. (2003). On D-trees, beans, and B-accents. *Linguistics & Philosophy, 26*(5), 511–545.

Chafe, W. (1974). Language and consciousness. *Language, 50*(1), 111–133.

Clifton, C., & Frazier, L. (2012). Discourse integration guided by the 'question under discussion'. *Cognitive Psychology, 65*(2), 352–379.

Colonna, S., Schimke, S., & Hemforth, B. (2012). Information structure effects on anaphora resolution in German and French: A cross-linguistic study of pronoun resolution. *Linguistics, 50,* 991–1013.

Colonna, S., Schimke, S., & Hemforth, B. (2014). Information structure and pronoun resolution in German and French: Evidence from the visual-world paradigm. In B. Hemforth, B. Schmiedtova, & C. Fabricius-Hansen (eds.), *Psycholinguistic approaches to meaning and understanding across languages* (pp. 175–195). Series in Theoretical Psycholinguistics. Heidelberg: Springer.

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47,* 292–314.

De la Fuente, I., & Hemforth, B. (2013). Effects of clefting and left dislocation on subject and object pronoun resolution in Spanish. In J. Cabrelli Amaro, G. Lord, A. de Prada Pérez, & J. Elana Aaron (eds.), *Selected Proceedings of the 16th Hispanic Linguistics Symposium*. Cascadilla Proceedings Project, Somerville, MA, USA, pp. 27–45.

Di Cristo, A. (1999). Le cadre accentuel du français contemporain. *Langues, 3*(2), 184–205; *Langues, 4*(2), 258–267.

Di Cristo, A., & Jankowski, L. (1999). Prosodic organisation and phrasing after focus in French. *Proceedings of XIVth ICPhS*. San Francisco: USA, pp. 1565–1568.

Dohen, M., & Loevenbruck, H. (2004). Pre-focal Rephrasing, Focal Enhancement and Post- focal Deaccentuation in French. *Proceedings of the 8th International Conference on Spoken Language Processing* (ICSLP), pp. 1313–1316, http://www.isca-speech.org/archive/interspeech_24.

Drenhaus, H., Zimmermann, M., & Vasishth, S. (2011). Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics, 1/2,* 11.

Féry, C. (2001). Focus and phrasing in French. In C. Féry & W. Sternefeld (eds.), *Audiatur Vox Sapientiae. A Festschrift for Arnim von Stechow* (pp. 153–181). Berlin: Akademie-Verlag.

Féry, C. (2014). Final compression in French as a phrasal phenomenon. In S. Katz Bourns & L. L. Myer (eds.), *Perspectives on linguistic structure and context* (pp. 133–156). Amsterdam: Benjamins.

Fonagy, I., Fonagy, Y., & Sap, J. (1979). A la recherche de traits pertinents prosodiques du français parisien. *Phonetica, 36*, 1–20.

German, J., & D'Imperio, M. (2010). Focus, phrase length, and the distribution of phrase-initial rises in French. Proceedings of 5th International Conference on Speech Prosody (2010; May 11–14: Chicago, USA), 1–4.

Ginzburg, J. (1995a). Resolving questions, I. *Linguistics and Philosophy, 18*(5), 459–527.

Ginzburg, J. (1995b). Resolving questions, II. *Linguistics and Philosophy, 18*(6), 567–609.

Grammont, M. (1933). *Traité de phonétique*. Paris: Librairie Delagrave.

Jacobs, J. (1984). Funktionale Satzperspektive und Illokutionsemantik. *Linguistische Berichte, 91,* 25–58.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation. In A. Botinis (ed.), *Intonation: Analysis, modeling and technology* (pp. 209–242). Dordrecht: Kluwer.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus, 14,* 147–172.

Kadmon, N. (2001). *Formal pragmatics: Semantics, pragmatics, presupposition, and focus*. Malden/Oxford: Blackwell.

Klein, W., & Stutterheim, C. von. (2002). Quaestio and L-perspectivation. In C. F. Graumann & W. Kallmeyer (eds.), *Perspective and perspectivation in discourse* (pp. 59–88). Amsterdam: John Benjamins.

Krifka, M. (2001). For a structured meaning account of questions and answers. In C. Fery & W. Sternefeld (eds.), *Audiatur Vox Sapientia. A Festschrift for Arnim von Stechow* (pp. 287–319). Berlin: Akademie Verlag.

Kuno, S. (2004). Empathy and direct discourse perspective. In L. Horn & G. Ward (eds.), *The handbook of pragmatics* (pp. 315–343). Blackwell.

Ladd, R. D. (2008). Intonational Phonology, 2nd edition. Cambridge: Cambridge University Press.

Lambrecht, K. (1994). *Information structure and sentence form: Topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.

Michelas, A., & D'Imperio, M. (2012). When syntax meets prosody: Tonal and duration variability in French accentual phrases. *Journal of Phonetics, 4*(6), 816–829.

Portes, C., & Beyssade, C. (to appear). *Is intonational meaning compositional?* Verbum.

Post, B. (2000). *Tonal and phrasal structures in French intonation*. Holland Academic Graphics.

Repp, S., & Drenhaus, H. (2015). Intonation influences processing and recall of sentences by indicating information-structural status of referents as topic vs. focus. *Language, cognition and neuroscience.* formerly: *Language and cognitive processes 30*(3), 324–346.

Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. In J. H. Yoon & A. Kathol (eds.), OSU Working Papers in Linguistics, Volume 49. The Ohio State University Department of Linguistics. Reprinted in the 1998 version in Semantics and Pragmatics, Volume 5, 2012, http://dx.doi.org/10.3765/sp.5.6.

Rossi, M. (1999). *L'intonation: le système du français*. Paris: Ophrys.

Selkirk, L. (2009). A New Paradigm for Studying the Prosodic Distinction between Contrastive Focus and Discourse-New. Presentation at IDP 9. this volume.

Vallduví, E., & Engdahl, E. (1996). The linguistic realization of information packaging. *Linguistics, 34*(3), 459–519.

Zimmermann, M., & Onea, E. (2011). Focus marking and focus interpretation. *Lingua, 1*/2011; *121*(11), 1651–1670.

Clefting, Parallelism, and Focus in Ellipsis Sentences

Katy Carlson

Abstract There are multiple ways to overtly indicate the position of contrastive focus in English, from pitch accents in prosody to clefting in syntax. But how comparable are these distinct focus indicators in their effects during processing? Ambiguous ellipsis sentences whose resolution is sensitive to focus provide a testing ground for this question, showing where perceivers choose to locate a contrast when given single or multiple focus indicators. In a self-paced reading experiment and an auditory questionnaire, syntactic and prosodic focus markers both influenced ellipsis interpretation. However, no single focus indicator fully disambiguated the sentences, illustrating the optionality of using focus-marked elements to resolve ellipsis structure. Further, the studies show a need for a detailed prosodic and semantic representation of noun phrase (NP) features to be held across clause boundaries.

Keywords Focus · Accents · Clefts · Ellipsis

To understand the processing of ellipsis sentences, in which missing material is filled in from surrounding complete clauses, we need to understand what types of information are retained across clause boundaries. In this project, different focus indicators as well as parallelism are shown to bias the resolution of ambiguous ellipsis sentences, such as *It was Shirley who counseled Naomi during the flight, not Donna*. They contribute to a calculation of similarity between noun phrases (NPs) and thus influence their likelihood of contrasting with each other.

1 Background on Focus

Focus is a way to formalize the fact that not everything in a sentence is equally important. Most of the time, some things in a sentence are already known or have already been discussed (this information is said to be "given"). Other parts of a

K. Carlson (✉)
Department of English, Morehead State University, 419 Combs Building,
Morehead, KY 40351, USA

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_4

entence are new and therefore add something to a discourse or conversation. Generally, the new information in a sentence should have what is known as informational focus or just focus. Every sentence has a focused element, and a sentence can have more than one. A simple test for the position of focus is to consider what wh-question a sentence would naturally answer. If *John went home* seems like a good answer to the question *Who went home?*, then *John* is focused; if it seems like a good answer to *Where did John go?*, then *home* is focused. Rooth (1992) presented a widely accepted proposal for what focus does in semantic interpretation, suggesting that focused items trigger consideration of possible alternatives.

Although English does not have a syntactic position in which focused elements always appear (unlike languages like Hungarian; Kiss 1998), there are places where it is more natural to have informational focus in a sentence. This is where focus will tend to be assumed in silent reading or in the absence of contextual information to the contrary, sometimes called default focus (Bader 1998; Cinque 1991; Gussenhoven 1994; Selkirk 1984; Stolterfoht et al. 2007). In particular, it is more common for informational focus to appear on the object of a sentence, or the last argument within the verb phrase (VP), than on the subject. Subjects are more likely to be topical and given, while the VP is where new information is often presented (Rooth 1992; Schwarzschild 1999).

The third category of information (besides given and new) is contrastive: Contrastive information may be new or given, but it counters or presents an alternative to something that has been stated or implied. Contrastive focus is optional, so not every sentence has a contrastive focus. Some theories of focus consider informational and contrastive focus to be the same, with contrastive focus just being focus in a noticeably contrastive context; others consider contrastive focus to be a distinct notion (see discussions in Kadmon 2001; Kiss 1998; Kratzer 2004; Rooth 1992; Schwarzschild 1999).

A sentence needs at least one accent and at least one focus, so much of the time focused items are accented prosodically in speech. But several other focus markers exist in English, including clefting and focus particles, as shown in (1).

(1) I heard that Bill died.
a. No, it was John that died, not Bill. (clefting)
b. No, only John died, not Bill. (focus particle)
c. No, JOHN died, not Bill. (pitch accent)

The clefting structure in (1a) places the focused element (*John*) in a specific syntactic position and backgrounds the rest of the sentence. The clefted element, *John,* is asserted to exhaust the set of items having the expressed property (exhaustivity: Kiss 1998; Rooth 1992, 1996). A focus particle like *only* highlights the constituent it precedes and similarly asserts that this phrase is the single contextually relevant possessor of the property expressed by the rest of the sentence. The placement of a pitch accent as indicated by capital letters in (1c) can indicate focus position on the accented word or phrase (Pierrehumbert and Hirschberg 1990; Schwarzschild 1999).

Like the controversy over whether contrastive focus is the same as informational focus, there is similar disagreement over whether there is a special accent type just for indicating contrastive focus or not. Usually, focused items in English receive an H* accent, which has a high target F0 (fundamental frequency) on or near the stressed syllable of an accented word (Beckman and Elam 1997; Pierrehumbert 1980). A L+H* accent is similar but preceded by a low F0 target, leading to a steep rise and fall for this accent type, and it is often higher than a standard H*. Some prosody researchers believe that the H* and L+H* accents are not distinct categories, but are part of a general continuum of accents which are more or less steep and more or less high (Bartels and Kingston 1996; Krahmer and Swerts 2001; Ladd 1996, 2008; Ladd and Schepman 2003). Whether the L+H* is a category or one end of a continuum, particularly high and steep accents are said to indicate contrastive focus (Pierrehumbert and Hirschberg 1990), and Ito and Speer (2008) have shown that L+H*s produce different behaviors than H* accents when instructing people to move items in contrastive and non-contrastive contexts.

All of the focus markers have additional functions and properties besides indicating the position of focus within a sentence. Pitch accents, for example, are influenced by the phonological structure of the utterance they appear in: A longer word or phrase will often have more accents than a shorter one, as well as more prosodic phrases. Both *only* and clefting are said to indicate exhaustivity (Kadmon 2001; Kiss 1998; Rooth 1992), as noted above. Interestingly, Drenhaus et al. (2011) present processing evidence suggesting that exhaustivity has a different status for *only* than in clefts. Specifically, they show that violating the exhaustivity requirement in German sentences with *only* resulted in a different event-related potential (ERP) signal than violating exhaustivity in sentences with clefts. They suggest that exhaustivity is not part of the truth-functional meaning of clefts, but is for *only*. Finally, clefting clearly affects the syntactic structure of an utterance more than either focus particles or pitch accents.

2 Focus in Processing

Turning to sentence processing, one line of focus research has found that focused elements in unambiguous sentences or sentence pairs receive greater attention, are remembered better, and are processed faster than unfocused elements (Birch and Garnsey 1995; Birch and Rayner 1997; Cutler 1976; Cutler and Fodor 1979; Gernsbacher and Jescheniak 1995). Focus has been indicated by various means in these studies, including accents, prior questions, and clefting. Cutler (1976) even showed that a constituent which should be accented, but was replaced with an unaccented rendition, received the focus benefit. Another line of research has found that processing is facilitated when a focus indicator (*only* or accent) correctly indicates an upcoming, unambiguous contrast (Bock and Mazzella 1983; Carlson 2013; Paterson et al. 2007; Stolterfoht et al. 2007). Similarly, processing is facilitated when accents appear on new rather than given elements (Birch and Clifton 1995, 2002; Nooteboom and Kruyt 1987).

In recent research, visual world eye-tracking studies have explored the processing of accents on line. Sedivy et al. (1995) found that contrastive accent on an adjective (e.g., *LARGE red square*) produced eye movements showing that perceivers expected a corresponding contrast between large and small items of the same type in the display. Similar results were found in German for color adjectives by Weber et al. (2006). Dahan et al. (2002) studied eye movements while perceivers heard accented or deaccented nouns with the same initial syllable (e.g., *candy* and *candle*) which were either already given or new in the context. They found looks to the new item occurring as soon as or even before an accented version of *CAN-* was heard, and similar looks to the given item in the deaccented condition. Studies of this sort show that perceivers can integrate information from pitch accents with the given/new/contrastive status of objects in a visual display or a discourse, and that they do so quickly.

Focus or accents can also affect the interpretation of ambiguous sentences. Several researchers have looked at whether *only* can help favor the reduced relative interpretation of the main clause/reduced relative ambiguity (Ni et al. 1996; Paterson et al. 1999; Liversedge et al. 2002; Sedivy 2002; Filik et al. 2005; but see also Clifton et al. 2000). Most researchers found that the presence of the focus particle eased the processing of the reduced relative interpretation of material following a head noun (e.g., *only businessmen loaned money*…; Sedivy 2002). Accents have been shown to influence relative clause attachment (Schafer et al. 1996; Lee and Watson 2011) and sentences with an indirect question/relative clause ambiguity (Schafer et al. 2000). Focused elements seem to be favored as antecedents in pronoun resolution across sentences (Almor 1999; Foraker and McElree 2007; see also Grosz et al. 1995; Cowles and Garnham 2005). Within sentences, though, Colonna et al. (2012) found that focused, clefted antecedents for subject pronouns in a subordinate clause were dispreferred in French and German, because they would require a shift in the topic of the sentence; topicalized antecedents, however, were preferred.

Ellipsis sentences, because they are focus-sensitive and often ambiguous, seem to be ideal for probing the interpretive effects of accents and other focus indicators. These sentences share the property of having an incomplete constituent elided under identity with parts of a nearby (usually preceding) clause. Theories of ellipsis resolution critically use the position of focus within the unelided antecedent clause to generate the meaning of the elided material (Merchant 2001; Rooth 1992; Sag 1980). For example, in order to interpret a replacive ellipsis (or bare argument ellipsis) sentence like (2), a focused argument in the first clause is abstracted over. This creates an open proposition, as in (a) or (b), which can then be copied around the remnant (*the senator*) and provide it with context.

(2) The judge joined the diplomat for coffee, not the senator.
a. x joined the diplomat for coffee (makes *the senator* a subject)
b. the judge joined x for coffee (makes *the senator* an object)

Syntactically, there is believed to be a full clause around the remnant that is just not pronounced (though, see Reinhart 1991), but in processing, one needs to copy or at least reactivate overt structure from the first clause in order to have something to

interpret. Syntactic and prosodic discussions of ellipsis state that certain focus and/or accent patterns are necessary to these sentences (e.g., Fery and Hartmann 2005; Kehler 2001; Kuno 1976; Rooth 1992). When there are multiple possible ways to resolve the ellipsis, as in (2), the perceived meaning of the sentences shows in part where perceivers have placed focus.

Processing research on ellipsis sentences so far has found that overt focus indicators are important in their resolution, though they do not appear to fully determine their interpretation (Carlson 2001, 2002; Carlson et al. 2005, 2009; Frazier and Clifton 1998; Paterson et al. 2007; Stolterfoht et al. 2007). The processing of sluicing, comparative ellipsis, gapping, replacives, stripping, and VP ellipsis suggests that the placement of pitch accents and focus indicators like *only* interacts with a bias toward focus within the VP. In particular, there is a persistent bias to resolve ellipsis towards object interpretations (e.g., with *the senator* in (2) as an object of *join*, contrasting with *the diplomat*) in all of these ellipsis types, which Frazier and Clifton (1998) and Carlson et al. (2009) suggested is due to the effects of expectations about focus position.

Since speakers have various ways to indicate focus, and there are positions where focus usually appears, how do people reading or listening to a sentence decide where focus actually is? One hypothesis is that the position of focus is fully determined for a perceiver by any overt focus indicators in the sentence, such as pitch accents and focus particles. Alternatively, expectations about the usual position of focus might still be active even in the presence of overt focus markers, as suggested by Frazier and Clifton (1998) and Carlson et al. (2009). In either case, it could also be that some focus indicators are more effective than others and indicate the position of focus more strongly. In fact, Kiss (1998) proposed that pitch accents do not convey the same exhaustive, contrastive focus that *only* or clefting do, which suggests that they might be processed differently. Pierrehumbert and Hirschberg (1990) claimed instead that one particular pitch accent, L+H*, conveys contrastive focus while other accents do not. In this research, we explore whether different focus markers (clefting and pitch accents) are more or less effective in influencing ellipsis sentence interpretation. Clarifying whether focus indicators have similar effects, as well as how expectations about focus positions interact with overt indicators in processing, is essential to developing a full theory of focus perception.

3 Experiment 1: Clefts in Self-Paced Reading

Experiment 1 studied the on line and off line interpretation of replacive ellipsis sentences, using self-paced reading followed by end of sentence interpretation questions. The sentences varied in featural parallelism between NPs and which argument was clefted. If overt focus completely determines the interpretation of these sentences, then the clefted argument should always be chosen to abstract over (as the correlate, or contrasting argument, for the ellipsis remnant). If there is still room for other possibilities, then other factors might also affect interpretation, such as the usual position of focus or similarities between the NPs to be contrasted.

3.1 Method

Participants The participants were 32 native English-speaking undergraduate students at Northwestern University enrolled in lower-division linguistics classes. They received course credit for their participation.

Materials The replacive sentences used in this project all have a first clause containing a subject, verb, object, and prepositional phrase (PP), followed by a negative remnant phrase (as in (3)).The negative *not* introduces a remnant NP (*Donna*) which contrasts with some NP in the first clause. The subject interpretation refers to the sentence meaning where the remnant contrasts with the first-clause subject, while the object interpretation is the meaning where the remnant contrasts with the object.

The self-paced reading study contained the four conditions in (3).

(3) a. It was *Shirley*/who counseled/a client/during the flight,/not *Donna*,/amazingly.
(Subject Cleft, Subject Parallel)
b. It was a client/who counseled/*Shirley*/…,/…,/…(Subject Cleft, Object Parallel)
c. It was a client/who *Shirley*/counseled/…,/…,/…(Object Cleft, Subject Parallel)
d. It was *Shirley*/who a client/counseled/…,/…,/…(Object Cleft, Object Parallel)

These sentences varied in which first-clause argument, the subject or object, was moved to the cleft position. They also included an additional factor, parallelism, or similarity in lexical form between NPs, which has been shown to affect ellipsis processing (Carlson 2001, 2002). Specifically, processors prefer to assign parallel syntactic positions to a remnant (e.g., *Donna* in (3)) and a first-clause argument (*Shirley* or *a client*) if they are more like each other in syntactic and semantic features than like other NPs in the sentence. In these sentences, the remnant was always a proper name of the same gender as one first-clause argument (the subject or object: shown italicized in (3)), while the nonparallel argument was an indefinite or definite description. The two factors, clefting and parallelism, were crossed so that conditions (a) and (d) had both factors biased in the same direction, and conditions (b–c) had conflicting biases. There were 16 items (see list in Appendix A).

Procedure and Equipment This experiment was carried out on a personal computer (PC) running PCEXPT software created by Charles Clifton, Jr. Each trial began when a participant pressed the space bar. This brought a preview of the sentence on the screen, with underscores replacing each letter but spaces and punctuation intact. The sentences were broken up into seven phrasal segments, as shown by the slashes in (3). Pressing the space bar again caused the letters of the first segment to appear on screen, and subsequent presses brought each following segment up. Although the segmentation used in presentation is shown in (3), the second and third presentation regions were collapsed for analysis purposes. The reaction times thus analyzed the whole region from *who* up to the start of the prepositional phrase (*during*), regardless of whether the object or subject was missing from it.

Participants were instructed to read at a comfortable pace that allowed them to comprehend the sentences, and a short practice session familiarized them with the procedure. The experimental and filler sentences appeared in one of four pseudo-randomized orders such that no two consecutive items were of the same type, and all items had similar segmentations. Each participant saw an equal number of items in each condition over the experiment. The experiment contained a total of 136 sentences, including various types of fillers.

Each trial was followed by a visually presented end-of-sentence question and two possible answers, as shown in (4) for the item in (3). The answers varied to match the roles of the arguments in the experimental conditions: Conditions (3a/c) had *Shirley* in the subject role, conditions (3b/d) had *a client* as subject.

(4) What did you find out about Donna?
a/c. Donna didn't counsel a client. (subject) vs. Donna wasn't counseled by Shirley. (object)
b/d. Donna didn't counsel Shirley. (subject) vs. Donna wasn't counseled by a client. (object)

Participants had been instructed to press the button on the right of the keyboard (marked with a red sticker) for the answer on the left side of the screen, or a button on the right (marked with a blue sticker) for the answer on the right. The position of answers was counterbalanced between items so that answers with the remnant contrasting with the subject (e.g., *Donna didn't counsel Shirley*) appeared on the right and left sides equally often over the experiment.

The answers differed in structure as shown here, with the object interpretation presented in the passive. The passive structure is more difficult than the simple active structure shown for the subject interpretation answer. Since sentences of this type usually have an object bias in interpretation (Carlson 2002), though, the difference in answer structures would work against this bias. If participants tended to choose the answer with the simpler structure, they would provide more subject answers. There was no reason to think this tendency would interact with the experimental manipulations.

Since the experimental sentences were ambiguous, no feedback was given to participants regarding their answers. The experiment lasted between 30 and 60 min, with short breaks for stretching provided.

3.2 Results and Discussion

The results included reading time measures from the on line interpretation of the sentences as well as the final interpretation percentages. We will begin with the reading times. In analysis, any reading time under 200 ms or over 4000 ms were dropped. To compensate for any length differences in segments, the raw times were subjected to a linear regression for each subject predicting reading times as a function of the characters in each segment. The residual reading times shown in Fig. 1 are the observed deviations from the predicted time for each segment. (Segment 2 in

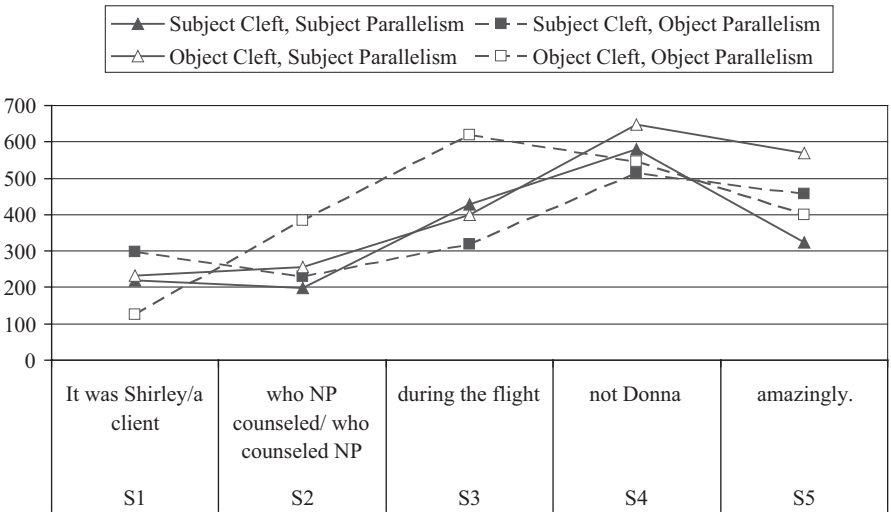


Fig. 1 Self-paced reading times in regressed ms, experiment 1

the figure and analysis refers to the combined segment from the start of the relative clause to the PP.)

At segments 2 and 3, the object cleft conditions were read slower than the subject clefts, as shown in significant main effects of cleft position in each segment (seg. 2: $F1(1,31)=8.5$, $p<0.01$, $F2(1,15)=6.3$, $p<0.05$; seg. 3: $F1(1,31)=5.3$, $p<0.05$, $F2(1,15)=4.8$, $p<0.05$) with no significant interaction in segment 2. By segment 3, there was also a significant interaction of parallelism and clefting, since condition (d), the object cleft with object parallelism, was slower than the rest ($F1(1,31)=11.8$, $p<0.01$; $F2(1,15)=7.1$, $p<0.05$). No significant differences were found in segment 4, but in segment 5, the two conditions with conflicting parallelism and cleft positions were slower than the other two, shown in a significant interaction between clefting and parallelism (interaction, $F1(1,31)=5.9$, $p<0.05$; $F2(1,15)=6.1$, $p<0.05$), while no main effects were significant.

The differences early in the sentences, with object clefts read slower than subject clefts, are most likely explained by the fact that object clefts are more difficult structures to process than subject clefts (Gordon et al. 2001). Further, the slowed processing for the object cleft with object parallelism condition probably reflects the combination of the more difficult object cleft structure with a dispreferred pattern of givenness (on the theory of a givenness hierarchy for NPs; Ariel 1990; Warren and Gibson 2002): This condition had an indefinite or definite subject NP but a proper name as the clefted object. Since there is a general preference for subjects to be higher on the givenness scale than objects, the interpretation of the object cleft would be even harder than usual. Thus, the effect in Segment 3 does not relate to parallelism, as such, since the remnant had not yet been encountered. Finally, the

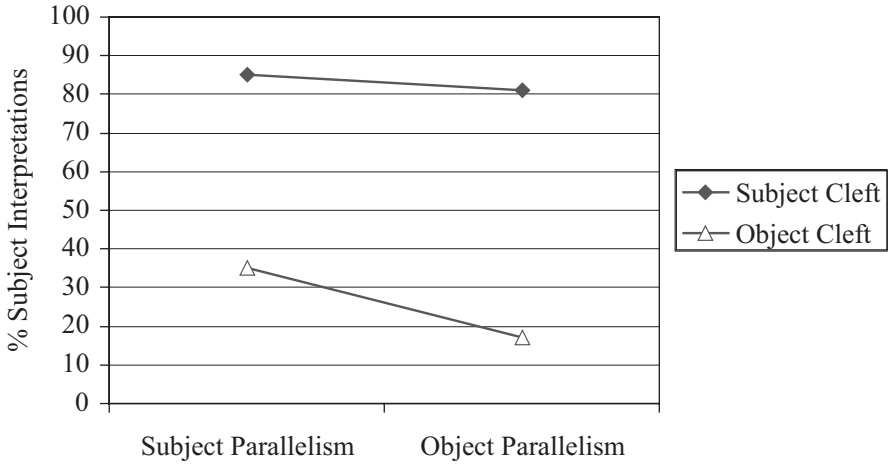


Fig. 2 Interpretation percentages, experiment 1

results later in the sentences find participants taking slightly longer when the parallelism and clefting cues do not point to the same interpretation, reasonably enough. A more parallel pair of NPs in a contrastive relationship makes for a more natural comparison. Clefting an NP marks it as contrastively focused, making it a natural choice for the contrast with the ellipsis remnant.

The end-of-sentence interpretation question results are consistent with these online measures, with parallelism having a significant but smaller effect than the position of clefting (see Fig. 2).

The subject-clefted conditions received primarily subject interpretations, over 80 % regardless of parallelism, while the object-clefted conditions received under 40 % subject responses. There was a significant main effect of which argument was clefted ($F1(1,31) = 161$, $p < 0.001$; $F2(1,15) = 149$, $p < 0.001$) and a significant main effect of parallelism ($F1(1,31) = 10$, $p < 0.005$; $F2(1,15) = 9$, $p < 0.01$), but a marginal to nonsignificant interaction between the factors (p's = 0.06–0.12). Numerically, the parallelism manipulation appeared to be more effective in the object clefts.

Overall, clefting the subject was a strong indication that the subject NP should contrast with the remnant, even overriding the general object bias usually seen for ambiguous ellipsis sentences. Clefting the object tended to produce more object interpretations, especially when object parallelism also made that analysis attractive. The clefted structure did not absolutely determine the interpretation, however, leaving room for the extra-grammatical factor of parallelism to also affect interpretations. The timing measures showed that syntactic difficulty and givenness mattered early in processing, before the contrastive remnant could be compared to the earlier NPs. At the remnant and following final segment, though, the parallelism between the remnant and the clefted NP could be computed, and increased

parallelism with the clefted argument was favored. It is interesting to note that clefting an NP places it earlier in the clause than any other syntactic position, increasing its distance from the remnant. Therefore, the fact that clefting increases interpretations in which the clefted argument contrasts with the remnant, rather than decreases them, means there is no strong recency preference: The processor does not favor locality of contrast.

4 Experiment 2: Auditory Clefts

In the auditory study, pitch accents and clefting were both applied to replacive sentences. This allowed for comparison between two different methods of indicating overt focus within the same sentences. Although it might seem that the clefted element is the only natural place to put a pitch accent, Gundel and Fretheim (2003) note that accents can occur within the post-cleft clause as well. Indeed, any sentence might have multiple foci, indicated with multiple accents. The question here is which focus is taken to indicate the contrast with the replacive remnant, and thus what interpretation is chosen.

4.1 Method

Participants The participants were 28 native English-speaking Northwestern University undergraduates in introductory linguistics classes. They received course credit for their participation.

Materials For this experiment, the parallelism manipulation was removed so that only clefting and pitch accents were varied. Thus, the sentences from experiment 1 were amended to make all three arguments proper names or all three definite descriptions, as in (5). This eliminated any increased syntactic or semantic similarity between the remnant and the subject or object. The final adverbial segment, which had functioned as a wrap-up segment for self-paced reading, was also omitted.

(5) a. It was SHIRLEY who counseled Naomi during the flight, not Donna.
b. It was Shirley who counseled NAOMI during the flight, not Donna.
c. It was Shirley who NAOMI counseled during the flight, not Donna.
d. It was SHIRLEY who Naomi counseled during the flight, not Donna.

Conditions (a–b) were subject clefts and conditions (c–d) were object clefts. The position of accent was crossed with clefting so that conditions (a) and (d) had accents marking the clefted argument, while conditions (b–c) accented the non-clefted argument. It was expected that the responses for conditions (b–c) would be most interesting, revealing whether the syntactic manipulation of clefting outweighed

Table 1 Average F0 measurements (and standard deviations, SDs) for peak heights and boundary tones, in Hz

	Cleft peak	Subject or object peak	Boundary tones (L−H%)	Remnant peak
S cleft, S accent	352 (20)	174 (8)	150 (7), 221 (21)	288 (31)
S cleft, O accent	247 (16)	328 (16)	149 (5), 231 (18)	277 (23)
O cleft, S accent	236 (13)	331 (16)	147 (6), 228 (25)	263 (27)
O cleft, O accent	347 (20)	176 (9)	149 (7), 226 (20)	278 (25)

Table 2 Average duration measurements (and SDs) for primary NPs, in ms

	Cleft NP	Subject or object NP	Remnant NP
S cleft, S accent	591 (146)	404 (129)	642 (116)
S cleft, O accent	483 (142)	517 (134)	629 (123)
O cleft, S accent	477 (152)	503 (155)	621 (122)
O cleft, O accent	570 (143)	411 (138)	629 (116)

the pitch accents in signaling the position of the relevant contrastive focus. The remnants in all sentences were also accented (e.g., *Donna*).

There were 16 cleft sentences in this auditory experiment (shown in Appendix B), and 120 sentences in total. The experimental sentences were produced in four prosodic conditions as in (5), with the position of contrastive accents varied, and analyzed for prosodic consistency. Average acoustic measurements for the conditions are shown in Tables 1 and 2, and a sample set of pitch tracks with tones and break indices (ToBI) analysis is shown in Fig. 3. These reveal that the clefted elements were accented in all conditions, for naturalness, but with small, non-prominent H* accents in conditions (b–c). The conditions where the clefted element had the only prominent accent in its clause had L+H* accents. Thus only the type of accent (or the extremeness of the accent) varied.

Procedure and Equipment This experiment was carried out on a PC running Superlab. Participants were seated at a desk in a soundproof booth in front of a computer, with a pen and an answer sheet. They wore headphones and pressed the space bar on the keyboard to hear each experimental item. After hearing each sentence, they looked down to a printed answer sheet and circled the answer which best fit their understanding of the sentence. The questions and answers corresponded to those used in the self-paced reading experiment, except that all critical NPs within any one sentence were proper names or definite descriptions. An instruction sheet and a short practice round familiarized participants with the procedure. The items appeared in one of four pseudo-randomized orders such that no two consecutive items were of the same type. The different answer choices also appeared equally often in first and second position. The entire experiment lasted approximately half an hour.

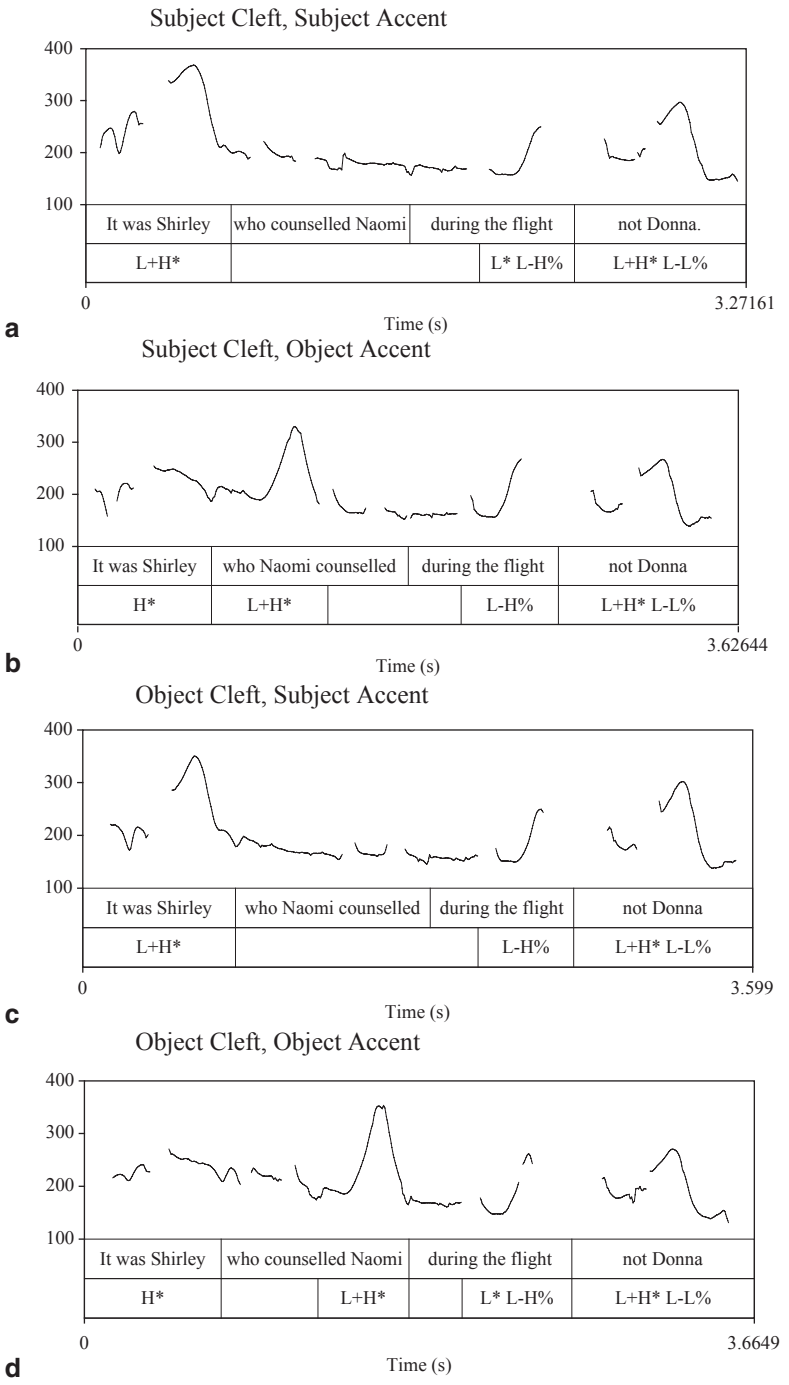


Fig. 3 Pitch tracks for an item in experiment 2. **a** Subject cleft, subject accent. **b** Subject cleft, object accent. **c** Object cleft, subject accent. **d** Object cleft, object accent

4.2 Results

The response percentages from the interpretation questions following each sentence are shown in Fig. 4.

With the subject clefted and also prominently accented, subject responses reached almost 90%, while a major accent on the non-clefted object instead lowered the percentage to under 60%. Object clefts with object accent received under 20% subject responses, which rose to almost 40% with the subject accented. There was a significant main effect of which argument was clefted ($F1(1,27)=68$, $p<0.001$; $F2(1,15)=146$, $p<0.001$) and a significant main effect of accent position ($F1(1,27)=19$, $p<0.001$; $F2(1,15)=27$, $p<0.001$), with a marginal to non-significant interaction (p's$=0.051-0.14$). The position of accent was important to the resolution of these sentences, though which argument was clefted was even more decisive.

There is a possibility that the conditions with non-matching clefting and accent position were less acceptable than the others, leading to confusion among listeners. This could then manifest in interpretation percentages closer to 50%. It is true that conditions with one NP clefted and another accented were more complex than the matching conditions. However, the theoretical literature on focus does allow for accents after a clefted element (Gundel and Fretheim 2003). Further, the conditions with non-matching clefts and accents did not have the same interpretation percentages, but showed a greater effect of clefting and a smaller effect of accent.

4.3 Discussion for Experiments 1–2

These two experiments examined the use of clefting and pitch accents in focusing an argument to contrast with an ellipsis remnant. In the self-paced reading study, there were slower reading times for the object-clefted conditions (c–d) than the

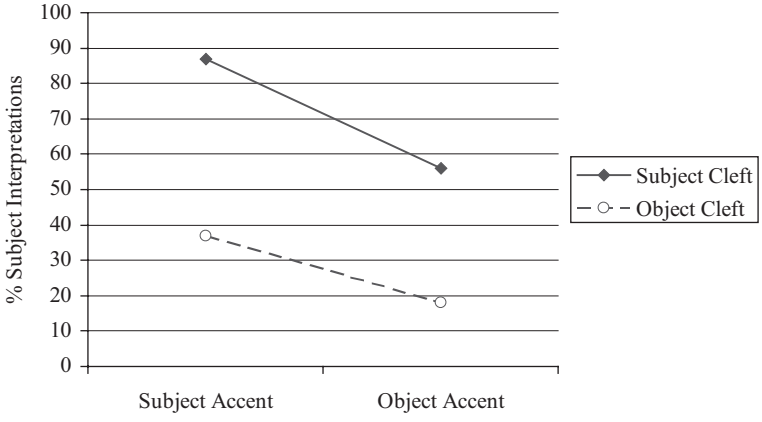


Fig. 4 Interpretation percentages for experiment 2

subject clefts early on in the sentences, consistent with prior work on cleft processing showing object clefts to be difficult (e.g., Gordon et al. 2001). The object cleft penalty was also heightened initially in the condition with a clefted object higher on the givenness hierarchy than the actual sentence subject. By the final segments, the slowest conditions were those where parallelism and clefting conflicted (b–c), suggesting that both factors were active at that point. End-of-sentence interpretation questions showed that parallelism ultimately had a fairly weak effect, with both subject-clefted conditions (a–b) producing over 80% subject responses, though the object clefts varied from 35% subject responses (c) to 18% subject responses (d). It makes sense that parallelism would be less influential than clefting, since focus indicators affect the semantic interpretation of a sentence directly while similarity between contrasting NPs is an extra-grammatical and optional factor.

Experiment 2, the auditory study, contained similar sentences without parallelism manipulations (i.e., all arguments were proper names). Instead, this experiment crossed accent position and cleft position. The subject-clefted conditions produced a majority of subject responses even when the accent placement did not support that analysis. The position of accent was still a significant determinant of an interpretation, but the syntactic cleft structure was more definitive in leading to a particular ellipsis resolution for perceivers.

5 Conclusions

These experiments show that clefting and pitch accents both affect the perceived focus structure of sentences, but that they do not determine the intended sentence meaning even in contrastive ellipsis sentences. Instead, each affected the interpretation of the sentences to a certain extent. When two different focus indicators marked the same argument, then the intended position of contrastive focus was fairly clear to the perceiver and tended to be used in ellipsis resolution. When different focus indicators were placed in different positions, results were intermediate. It is likely that both arguments were interpreted as focused in such cases, since multiple foci are allowed in any sentence. Gernsbacher and Jescheniak (1995), for example, found that probe recognition in sentences with two accented arguments was facilitated for both arguments to the same extent as a single stressed argument. Here, perceivers then had a choice as to which focus position would be taken as relevant to the ellipsis resolution.

The overall pattern of results is consistent with the various focus indicators being roughly equally effective and functioning additively. While this is not necessarily surprising, syntactic focus indicators like clefting might have been taken as clearer signals of focus position than accents. This would have harmonized with Kiss (1998)'s view of their semantics, and the fact that prosodic accents are much more common than cleft structures and used for a variety of purposes. Certainly, the different focus indicators do have different semantic properties, and other effects on sentences besides focus. Kiss (1998) may be right that clefting and *only* convey exhaustivity while even the potentially contrastive L+H* pitch accents do not; Drenhaus et al. (2011) may be right in suggesting that exhaustivity has a different

relationship to the meaning of clefts versus *only*. But for the purposes of establishing a contrastively focused element in a sentence which will contrast with an ellipsis remnant, all of the focus indicators behave quite similarly to each other.

In related research, Carlson (2013) has shown that ambiguous replacive sentences (e.g., *The curator embarrassed the gallery owner in public, not the artist*) are also affected by the presence and position of the focus particle *only*. The ellipsis remnant (*not the artist*) was read faster in sentences with *only* on the first-clause subject or object than in sentences without the overt focus marker, and the position of *only* influenced the choice of interpretations. Further, NPs in the target sentence focused by a question in a preceding context were also read faster and chosen as the contrast with the remnant more often. The overall level of subject interpretations in this *only* research was lower (with a maximum around 67% with subject biases) than in the clefting research; this raises the possibility that clefting has a stronger effect on ellipsis resolution than the particle *only* or pitch accents.

It is worth noting that the ambiguous regions of the ellipsis sentences in the current experiments were identical across conditions (all remnants were accented in the auditory study, and none were clefted). All differences in interpretation were therefore produced by manipulations within the unambiguous first clauses, rather than in the ambiguous remnants. Thus, the results of the experiments demonstrate the need for a global sentence representation which retains detailed prosodic and semantic information over clause boundaries (and in experiment 2, over intonational phrase boundaries). Features of the first-clause NPs, from prosodic accenting to similarities in NP form, aided in determining what would be the most natural contrast with the remnant. Just comparing the remnant to the nearest NP, the object, would not be enough to produce the parallelism effects in experiment 1 and the accent effects in experiment 2. Those required that the remnant be compared to both first-clause NPs on a range of features, and the more features that were similar, the more likely the remnant would be placed in a parallel syntactic position to that NP.

The high rate of subject interpretations with subject clefts, especially with subject parallelism or subject accenting, shows that these factors are able to overcome default focus expectations. And there is no evidence of a strong recency bias in these sentences. If there were, it would have worked against the cleft manipulation, which focuses an argument by placing it earlier in the sentence that it otherwise would be. The overall pattern of results shows that perceivers do notice multiple focus indicators in a sentence, and each affects the choice of a position to abstract over during ellipsis resolution. Focused elements are favored as contrasts regardless of the source of the focus marking, but no overt focus marker completely disambiguated these ellipsis sentences.

Acknowledgments I would like to thank Catherine Anderson, Michael Walsh Dickey, and Ann Bradlow for assistance in the setup and running of the experiments. I also am indebted to Lyn Frazier, Chuck Clifton, Mike Dickey, Joseph Tyler, and three anonymous reviewers for commenting on drafts of this work. This research was partially supported by an institutional development award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health under grant number 5P20GM103436-13 and by the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under Award Number R15HD072713. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

Appendix A

Materials for experiment 1, visual clefts. Vertical slashes indicate the presentation regions for self-paced reading.

1a. It was Dr. Waters|who saved|a lifeguard|from drowning,|not Dr. Green,|interestingly.
1b. It was a lifeguard|who saved|Dr. Waters|from drowning,|not Dr. Green,|interestingly.
1c. It was a lifeguard|who Dr. Waters|saved|from drowning,|not Dr. Green,|interestingly.
1d. It was Dr. Waters|who a lifeguard|saved|from drowning,|not Dr. Green,|interestingly.
2a. It was George|who described|a neighbor|for the newspapers,|not Charles,|reportedly.
2b. It was a neighbor|who described|George|for the newspapers,|not Charles,|reportedly.
2c. It was a neighbor|who George|described|for the newspapers,|not Charles,|reportedly.
2d. It was George|who a neighbor|described|for the newspapers,|not Charles,|reportedly.
3a. It was Shirley|who counseled|a client|during the flight,|not Donna,|amazingly.
3b. It was a client|who counseled|Shirley|during the flight,|not Donna,|amazingly.
3c. It was a client|who Shirley|counseled|during the flight,|not Donna,|amazingly.
3d. It was Shirley|who a client|counseled|during the flight,|not Donna,|amazingly.
4a. It was William|who bored|a technician|with his life story,|not Anthony,|thankfully.
4b. It was a technician|who bored|William|with his life story,|not Anthony,|thankfully.
4c. It was a technician|who William|bored|with his life story,|not Anthony,|thankfully.
4d. It was William|who a technician|bored|with his life story,|not Anthony,|thankfully.
5a. It was Mandy|who presented|the teacher|with an award,|not Julie,|naturally.
5b. It was the teacher|who presented|Mandy|with an award,|not Julie,|naturally.
5c. It was the teacher|who Mandy|presented|with an award,|not Julie,|naturally.
5d. It was Mandy|who the teacher|presented|with an award,|not Julie,|naturally.
6a. It was Abbie|who amazed|the soloist|after the concert,|not Gloria,|obviously.
6b. It was the soloist|who amazed Abbie|after the concert,|not Gloria,|obviously.
6c. It was the soloist|who Abbie amazed|after the concert,|not Gloria,|obviously.
6d. It was Abbie|who the soloist amazed|after the concert,|not Gloria,|obviously.
7a. It was Karl|who coached|the prodigy|before the test,|not Andrew,|strangely.
7b. It was the prodigy|who coached|Karl|before the test,|not Andrew,|strangely.
7c. It was the prodigy|who Karl|coached|before the test,|not Andrew,|strangely.
7d. It was Karl|who the prodigy|coached|before the test,|not Andrew,|strangely.
8a. It was Reggie|who reported|the cheater|to the authorities,|not Douglas,|apparently.
8b. It was the cheater|who reported|Reggie|to the authorities,|not Douglas,|apparently.
8c. It was the cheater|who Reggie|reported|to the authorities,|not Douglas,|apparently.
8d. It was Reggie|who the cheater|reported|to the authorities,|not Douglas,|apparently.
9a. It was Maude|who informed|the fireman|about the damage,|not Felicia,|actually.
9b. It was the fireman|who informed|Maude|about the damage,|not Felicia,|actually.
9c. It was the fireman|who Maude|informed|about the damage,|not Felicia,|actually.
9d. It was Maude|who the fireman|informed|about the damage,|not Felicia,|actually.
10a. It was Lindsay|who called|the dentist|after the party,|not Dolores,|fortunately.
10b. It was the dentist|who called|Lindsay|after the party,|not Dolores,|fortunately.
10c. It was the dentist|who Lindsay|called|after the party,|not Dolores,|fortunately.
10d. It was Lindsay|who the dentist|called|after the party,|not Dolores,|fortunately.
11a. It was Jeremy|who hired|a manager|for the resort,|not Peter,|surprisingly.
11b. It was a manager|who hired|Jeremy|for the resort,|not Peter,|surprisingly.
11c. It was a manager|who Jeremy|hired|for the resort,|not Peter,|surprisingly.
11d. It was Jeremy|who a manager|hired|for the resort,|not Peter,|surprisingly.
12a. It was Lyle|who defended|a prosecutor|in court,|not Jack,|interestingly.
12b. It was a prosecutor|who defended|Lyle|in court,|not Jack,|interestingly.
12c. It was a prosecutor|who Lyle|defended|in court,|not Jack,|interestingly.
12d. It was Lyle|who a prosecutor|defended|in court,|not Jack,|interestingly.
13a. It was Patricia|who interviewed|an athlete|last December,|not Caroline,|apparently.

13b. It was an athlete|who interviewed|Patricia|last December,|not Caroline,|apparently.
13c. It was an athlete|who Patricia|interviewed|last December,|not Caroline,|apparently.
13d. It was Patricia|who an athlete|interviewed|last December,|not Caroline,|apparently.
14a. It was Judith|who insulted|the waiter|during dinner,|not Ashley,|reportedly.
14b. It was the waiter|who insulted|Judith|during dinner,|not Ashley,|reportedly.
14c. It was the waiter|who Judith|insulted|during dinner,|not Ashley,|reportedly.
14d. It was Judith|who the waiter|insulted|during dinner,|not Ashley,|reportedly.
15a. It was Megan|who advised|an actress|about makeup,|not Denise,|thankfully.
15b. It was an actress|who advised|Megan|about makeup,|not Denise,|thankfully.
15c. It was an actress|who Megan|advised|about makeup,|not Denise,|thankfully.
15d. It was Megan|who an actress|advised|about makeup,|not Denise,|thankfully.
16a. It was Marcus|who invited|the jerk|to the party,|not Joshua,|naturally.
16b. It was the jerk|who invited|Marcus|to the party,|not Joshua,|naturally.
16c. It was the jerk|who Marcus|invited|to the party,|not Joshua,|naturally.
16d. It was Marcus|who the jerk|invited|to the party,|not Joshua,|naturally.

Appendix B

Materials for experiment 2, auditory clefts. Capital letters indicate position of contrastive accent within the first clause.

1a. It was Dr. WATERS who saved Dr. Miller from drowning, not Dr. Green.
1b. It was Dr. Waters who saved Dr. MILLER from drowning, not Dr. Green.
1c. It was Dr. WATERS who Dr. Miller saved from drowning, not Dr. Green.
1d. It was Dr. Waters who Dr. MILLER saved from drowning, not Dr. Green.
2a. It was GEORGE who described Travis for the newspapers, not Charles.
2b. It was George who described TRAVIS for the newspapers, not Charles.
2c. It was GEORGE who Travis described for the newspapers, not Charles.
2d. It was George who TRAVIS described for the newspapers, not Charles.
3a. It was SHIRLEY who counseled Naomi during the flight, not Donna.
3b. It was Shirley who counseled NAOMI during the flight, not Donna.
3c. It was SHIRLEY who Naomi counseled during the flight, not Donna.
3d. It was Shirley who NAOMI counseled during the flight, not Donna.
4a. It was WILLIAM who bored Daniel with his life story, not Anthony.
4b. It was William who bored DANIEL with his life story, not Anthony.
4c. It was WILLIAM who Daniel bored with his life story, not Anthony.
4d. It was William who DANIEL bored with his life story, not Anthony.
5a. It was the PRINCIPAL who presented the superintendent with an award, not the teacher.
5a. It was the principal who presented the SUPERINTENDENT with an award, not the teacher.
5d. It was the PRINCIPAL who the superintendent presented with an award, not the teacher.
5d. It was the principal who the SUPERINTENDENT presented with an award, not the teacher.
6a. It was the VIOLINIST who amazed the soloist after the concert, not the conductor.
6b. It was the violinist who amazed the SOLOIST after the concert, not the conductor.
6c. It was the VIOLINIST who the soloist amazed after the concert, not the conductor.
6d. It was the violinist who the SOLOIST amazed after the concert, not the conductor.
7a. It was the teacher's PET who coached the prodigy before the test, not the bookworm.
7b. It was the teacher's pet who coached the PRODIGY before the test, not the bookworm.
7c. It was the teacher's PET who the prodigy coached before the test, not the bookworm.
7d. It was the teacher's pet who the PRODIGY coached before the test, not the bookworm.
8a. It was the DETECTIVE who informed the fireman about the damage, not the reporter.
8b. It was the detective who informed the FIREMAN about the damage, not the reporter.
8c. It was the DETECTIVE who the fireman informed about the damage, not the reporter.

8d. It was the detective who the FIREMAN informed about the damage, not the reporter.
9a. It was REGGIE who reported Scott to the authorities, not Douglas.
9b. It was Reggie who reported SCOTT to the authorities, not Douglas.
9c. It was REGGIE who Scott reported to the authorities, not Douglas.
9d. It was Reggie who SCOTT reported to the authorities, not Douglas.
10a. It was LINDSAY who met Cheryl after the party, not Dolores.
10b. It was Lindsay who met CHERYL after the party, not Dolores.
10c. It was LINDSAY who Cheryl met after the party, not Dolores.
10d. It was Lindsay who CHERYL met after the party, not Dolores.
11a. It was JEREMY who hired Matthew for the resort, not Peter.
11b. It was Jeremy who hired MATTHEW for the resort, not Peter.
11c. It was JEREMY who Matthew hired for the resort, not Peter.
11d. It was Jeremy who MATTHEW hired for the resort, not Peter.
12a. It was LYLE who defended Chris in court, not Jack.
12b. It was Lyle who defended CHRIS in court, not Jack.
12c. It was LYLE who Chris defended in court, not Jack.
12d. It was Lyle who CHRIS defended in court, not Jack.
13a. It was a CELEBRITY who interviewed a writer last December, not a journalist.
13b. It was a celebrity who interviewed a WRITER last December, not a journalist.
13c. It was a CELEBRITY who a writer interviewed last December, not a journalist.
13d. It was a celebrity who a WRITER interviewed last December, not a journalist.
14a. It was the DISHwasher who insulted the waiter during dinner, not the cook.
14b. It was the dishwasher who insulted the WAITER during dinner, not the cook.
14c. It was the DISHwasher who the waiter insulted during dinner, not the cook.
14d. It was the dishwasher who the WAITER insulted during dinner, not the cook.
15a. It was a STAGEHAND who advised an actress about makeup, not an understudy.
15b. It was a stagehand who advised an ACTRESS about makeup, not an understudy.
15c. It was a STAGEHAND who an actress advised about makeup, not an understudy.
15d. It was a stagehand who an ACTRESS advised about makeup, not an understudy.
16a. It was the STUDENTS who invited the professor to the party, not the TAs.
16b. It was the students who invited the PROFESSOR to the party, not the TAs.
16c. It was the STUDENTS who the professor invited to the party, not the TAs.
16d. It was the students who the PROFESSOR invited to the party, not the TAs.

References

Almor, A. (1999). Noun–phrase anaphora and focus: The informational load hypothesis. *Psychological Review, 106,* 748–765.

Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 3–48). Dordrecht: Kluwer.

Bartels, C., & Kingston, J. (1996). Salient pitch cues in the perception of contrastive focus. In M. W. Dickey & S. Tunstall (Eds.), *UMOP 19: Linguistics in the laboratory* (pp. 1–26). Amherst: GLSA.

Beckman, M., & Elam, G. A. (1997). *Guidelines for ToBI transcription, version 3*. Columbus: Ohio State University. http://www.ling.ohio-state.edu/~tobi/ame_tobi/. Accessed 9 April 2009.

Birch, S., & Clifton, C., Jr. (1995). Focus, accent, and argument structure: Effects on language comprehension. *Language and Speech, 38,* 365–391.

Birch, S., & Clifton, C., Jr. (2002). Effects of varying focus and accenting of adjuncts on the comprehension of utterances. *Journal of Memory and Language, 47,* 571–588.

Birch, S. L., & Garnsey, S. (1995). The effect of focus on memory for words in sentences. *Journal of Memory and Language, 34*, 232–267.
Birch, S., & Rayner, K. (1997). Linguistic focus affects eye movements during reading. *Memory and Cognition, 25,* 653–660.
Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory and Cognition, 11,* 64–76.
Carlson, K. (2001). The effects of parallelism and prosody on the processing of gapping structures. *Language and Speech, 44,* 1–26.
Carlson, K. (2002). *Parallelism and prosody in the processing of ellipsis sentences* (*Outstanding dissertations in linguistics series*). New York: Routledge.
Carlson, K. (2013). The role of *only* in contrasts in and out of context. *Discourse Processes, 50,* 249–275.
Carlson, K., Dickey, M. W., & Kennedy, C. (2005). Structural economy in the processing and representation of gapping sentences. *Syntax, 8,* 208–228.
Carlson, K., Dickey, M. W., Frazier, L., & Clifton, C., Jr. (2009). Information structure expectations in sentence comprehension. *Quarterly Journal of Experimental Psychology, 62,* 114–139.
Cinque, G. (1991). A null theory of phrase and compound stress. *Linguistic Inquiry, 24,* 239–297.
Clifton, C., Bock, J., & Rado, J. (2000). Effects of the focus particle only and intrinsic contrast on comprehension of reduced relative clauses. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 591–619). Amsterdam: Elsevier.
Colonna, S., Schimke, S., & Hemforth, B. (2012). Information structure effects on anaphora resolution in German and French: A crosslinguistic study of pronoun resolution. *Linguistics, 50,* 991–1013.
Cowles, H. W., & Garnham, A. (2005). Antecedent focus and conceptual distance effects in category noun-phrase anaphora. *Language and Cognitive Processes, 20,* 725–750.
Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20,* 55–60.
Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7,* 49–59.
Dahan, D., Tanenhaus, M., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language, 47,* 292–314.
Drenhaus, H., Zimmermann, M., & Vasishth, S. (2011). Exhaustiveness effects in clefts are not truth-functional. *Journal of Neurolinguistics, 24,* 320–337.
Fery, C., & Hartmann, K. (2005). The focus and prosodic structure of German right node raising and gapping. *The Linguistic Review, 22,* 69–116.
Filik, R., Paterson, K. B., & Liversedge, S. P. (2005). Parsing with focus particles in context: Eye movements during the processing of relative clause ambiguities. *Journal of Memory and Language, 53,* 473–495.
Foraker, S., & McElree, B. (2007). The role of prominence in pronoun resolution: Active versus passive representation. *Journal of Memory and Language, 56,* 357–383.
Frazier, L., & Clifton, C., Jr. (1998). Comprehension of sluiced constituents. *Language and Cognitive Processes, 13,* 499–520.
Gernsbacher, M. A., & Jescheniak, J. D. (1995). Cataphoric devices in spoken discourse. *Cognitive Psychology, 29,* 24–58.
Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 1411–1423.
Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics, 21,* 203–226.
Gundel, J. K., & Fretheim, T. (2003). Topic and focus. In G. Ward & L. Horn (Eds.), *Handbook of pragmatic theory* (pp. 175–196). Oxford: Blackwell.
Gussenhoven, C. (1994). Focus and sentence accents in English. In P. Bosch & R. van der Sandt (Eds.), *Focus and natural language processing* (pp. 83–92). Heidelberg: IBM Deutschland.
Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58,* 541–573.

Kadmon, N. (2001). *Formal pragmatics: Semantics, pragmatics, presupposition, and focus*. Oxford: Blackwell.
Kehler, A. (2001). Coherence and the resolution of ellipsis. *Linguistics and Philosophy, 23,* 533–575.
Kiss, K. E. (1998). Identificational focus vs. information focus. *Language, 74,* 245–273.
Krahmer, E., & Swerts, M. (2001). On the alleged existence of contrastive accents. *Speech Communication, 34,* 391–405.
Kratzer, A. (2004). Interpreting focus: Presupposed or expressive meanings? *Theoretical Linguistics, 30,* 123–136.
Kuno, S. (1976). Gapping: A functional analysis. *Linguistic Inquiry, 7,* 300–318.
Ladd, D. R. (1996). *Intonational phonology*. Cambridge: Cambridge University Press.
Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.
Ladd, R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: Experimental evidence. *Journal of Phonetics, 31,* 81–112.
Lee, E.-K., & Watson, D. G. (2011). Effects of pitch accents in attachment ambiguity resolution. *Language and Cognitive Processes, 26,* 262–297.
Liversedge, S. P., Paterson, K. B., & Clayes, E. L. (2002). The influence of only on syntactic processing of "long" relative clause sentences. *Quarterly Journal of Experimental Psychology, 55A,* 225–240.
Merchant, J. (2001). *The syntax of silence: Sluicing, islands, and identity in ellipsis*. Oxford: Oxford University Press.
Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes, 11,* 283–334.
Nooteboom, S. G., & Kruyt, J. G.(1987). Accents, focus distribution, and the perceived distribution of given and new information: An experiment. *Journal of the Acoustical Society of America, 82,* 1512–1524.
Paterson, K. B., Liversedge, S. P., & Underwood, G. (1999). The influence of focus operators on syntactic processing of "short" reduced relative clause sentences. *Quarterly Journal of Experimental Psychology, 52A,* 717–737.
Paterson, K. B., Liversedge, S. P., Filik, R., Juhasz, B. J., White, S. J., & Rayner, K. (2007). Focus identification during sentence comprehension: Evidence from eye movements. *Quarterly Journal of Experimental Psychology, 60,* 1423–1445.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.
Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of intonation in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge: MIT.
Reinhart, T. (1991). Elliptic conjunctions—Non-quantificational LF. In A. Kasher (Ed.), *The Chomskyan turn* (pp. 360–384). Cambridge: Blackwell.
Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1,* 75–116.
Rooth, M. (1996). On the interface principles for intonational focus. In T. Galloway & J. Spence (Eds.), *Salt VI* (pp. 202–226). Ithaca: Cornell University.
Sag, I. (1980). *Deletion and logical form*. New York: Garland Publishing.
Schafer, A. J., Carter, J., Clifton, C., Jr., & Frazier, L. (1996). Focus in relative clause construal. *Language & Cognitive Processes, 11,* 135–163. doi:10.1080/016909696387240.
Schafer, A., Carlson, K., Clifton, C., Jr., & Frazier, L. (2000). Focus and the interpretation of pitch accents: Disambiguating embedded questions. *Language and Speech, 43,* 75–106.
Schwarzschild, R. (1999). Givenness, avoid F and other constraints on the placement of accent. *Natural Language Semantics, 7,* 141–177.
Sedivy, J. C. (2002). Invoking discourse-based contrast sets and resolving syntactic ambiguities. *Journal of Memory and Language, 46,* 341–370.

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1995). Using intonationally-marked presuppositional information in on-line language processing: Evidence from eye movements to a visual model. In J. D. Moore & J. F. Lehman (Eds.), *Proceedings of the 17th annual conference of the cognitive science society* (pp. 375–380). Hillsdale: Erlbaum.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT.

Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition, 104,* 565–590.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85,* 79–112.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49,* 367–392.

The Effect of Phonological Encoding on Word Duration: Selection Takes Time

Duane G Watson, Andrés Buxó-Lugo and Dominique C Simmons

Abstract In this chapter, we investigate whether the process of phonological encoding plays a role in determining the duration of a word. We explore whether points of complexity in word production as predicted by a simple recurrent network also predict points within a word at which speakers slow down. Simple recurrent networks were trained to produce two different words under two conditions: In the first condition, the two words in the sequence overlapped in their initial morphemes (e.g., layover layout) and in the second condition, the words overlapped in their final morpheme (e.g., overlay outlay). The network experienced the most error for words that overlapped initially and at points of word non-overlap. Participants who produced these same sequences in a repetition task exhibited lengthening at points of complexity predicted by the network. We propose that lengthening may be partly a result of the phonological encoding system needing processing time.

Keywords Prosody · Production · Phonological encoding · Simple recurrent network · Duration · Modeling · Phonology

It is a well-known phenomenon that speakers lengthen words that are new, informative, or not predictable in a conversation and shorten words that are given, predictable, or non-informative (e.g., Aylett and Turk 2004; Bell et al. 2009; Fowler and Housum 1987; Jurafsky et al. 2001; Lam and Watson 2010; Pluymaekers et al. 2005 and many others). A puzzle for linguists, psychologists, and computer scientists who are interested in prosody is understanding why.

There are two varieties of explanations. One is that speakers lengthen and shorten words to facilitate robust communication with listeners. This idea has been described

D. G. Watson (✉) · A. Buxó-Lugo
Department of Psychology, University of Illinois Urbana-Champaign, 603 E. Daniel St., 61820 Champaign, IL, USA
e-mail: dgwatson@illinois.edu

A. Buxó-Lugo
e-mail: buxo2@illinois.edu

D. C. Simmons
Department of Psychology, University of California Riverside, Riverside, CA, USA
e-mail: dsimm002@ucr.edu

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_5

within formal frameworks like the uniform information density hypothesis (e.g., Jaeger 2010) and the smooth signal hypothesis (Aylett and Turk 2004): Speakers lengthen linguistic information with high information content and shorten words with low information content to create a uniform information density across utterances. The other explanation is that the duration of words partly reflects the complexity of underlying production processes. Speakers produce words that are new or informative with longer duration because those words are actually more difficult to say. The extra time provided by lengthening the segments facilitates the production process.

It is important to note that these two explanations are not incompatible. It is possible that duration choices facilitate the workings of mechanisms that are engaged in production while at the same time optimizing word length for robust communication. However, a challenge for both of these approaches is mapping out the underlying mechanisms.

The current chapter explores the algorithms that underlie production-centered theories of reduction and lengthening. There have been some proposals for how reduction and lengthening might facilitate production (e.g., see Bell et al. 2009; Kahn and Arnold n.d., 2012), though typically the mechanism is framed in terms of activation of the routines associated with production. If a word has been recently produced or is highly predictable, its resting activation will be higher, and consequently, it will be easier to produce, and the word will be shortened. Similarly, because new words will have lower activation, the increased effort required for articulation results in longer production times. A drawback of this type of explanation is that it remains unspecified as to why lengthening a difficult word (or reducing a highly activated word) would facilitate language production. If lengthening is linked to planning difficulty, why does it not occur before the critical target word? Once one begins to utter a word, presumably its meaning and lemma have already been accessed. What benefit could a speaker derive from lengthening a new word once articulation has already begun?

The answer may lie in theories of phonological encoding. In some models of word production, phonological selection is a serial process (Sevald and Dell 1994; although see O'Seaghdha and Marin 2000). Once a word is accessed, phonemes are accessed in an order that corresponds with the order in which they appear in a word, starting with phonemes at the beginning of the lexical item. There is empirical support for this type of architecture. Sevald and Dell (1994) found that rapid repetition of two alternating words was faster when those words shared their rhymes ("TICK" vs. "PICK") than when they shared onsets ("PICK" vs. "PIN") (see Jaeger et al. n.d., for similar effects of phonological overlap on lexical selection). These results can be accounted for within an interactive activation model like the Dell (1986) model in which low-level phonemic representations send feedback to higher-level lexical representations. In such a model, the shared onset activates both words, which increases competition between the lexical items and inhibits the correct selection of the target. In contrast, words that share rhymes are not burdened by inhibition early in the selection process, which facilitates production.

If phonological encoding is a serial process, or at least a process that is not entirely completed at the point of articulation, this may explain why words that are new

are lengthened. Lengthening could provide more time for phonological selection to take place at the point of articulation. Similarly, word reduction could be the result of faster phonological selection. There are empirical data that suggest these effects are in fact driven by phonological processes. Kahn and Arnold (2012) found that both non-mentioned, conceptually given words and mentioned words are reduced, but mentioned words exhibit greater reduction. Similarly, Lam and Watson (n.d.) found that repeated words, but not repeated referents, lead to reduction. Although the results in Kahn and Arnold (2012) and Lam and Watson (n.d.) do not by themselves suggest that a serial production process underlies these effects, they do suggest that these effects originate at the phonological or articulatory level.

In this chapter, we explore whether the dynamics of a phonological production system that serially encodes linguistic information can explain changes in duration in word production. The strategy we use is to first understand whether a serial selection model predicts complexity at varying points within a word and across words. Then we test to see whether English speakers' durational choices match predicted points of complexity by the model. If predicted points of complexity and lengthening overlap, it will suggest that duration effects might be linked to phonological encoding processes.

We use a model inspired by Dell et al. (1993). It is a simple recurrent network (SRN), originally designed to model speech errors. This model was used for two reasons. The first is that it allowed us to easily encode phonological selection as a serial process that occurs across time. As in all SRNs, Dell et al.'s model has a set of context units that encodes activation of hidden nodes on previous time steps. The second motivation for using this model was that it allowed us to test whether representational similarity across words impacts difficulty of production while making minimal assumptions about the architecture of the model.

As in Sevald and Dell (1994), the model was trained to produce words in which the output overlaps in its initial part (e.g., "layover layout") or overlaps in its final part ("overlay outlay"). We used words with morphological overlap instead of overlap in subsyllabic components like the rime and onset (as in Sevald and Dell 1994). This was done to increase the amount of overlap across words in order to amplify the size of any potential effect that this might have on word production in our human production data. Manipulating morphological overlap also has the added advantage of simplifying the learning goals of the model: Rather than learning mappings between a lemma and a string of phonological features or phonemes, the model learns a simple mapping between a lemma and two parts of a word, allowing us to focus our question on how linguistic overlap generally impacts production. Finally, by making the unit of overlap a morpheme, we can more easily measure duration differences across words in human productions.

The prediction from Sevald and Dell's (1994) results is that words that overlap initially should be more difficult to produce than those that overlap finally. Critically, we will see where within the words the model predicts the greatest point of complexity, and determine whether these predictions correspond with speaker duration preferences.

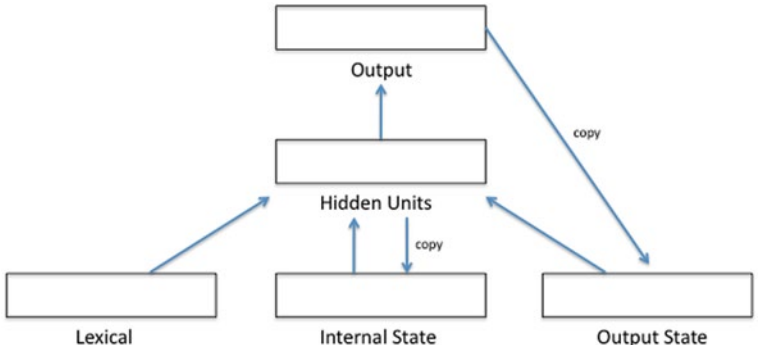


Fig. 1 The architecture of the simple recurrent network

1 Models

The architecture of the model was similar to that used by Dell et al. (1993). The primary components are illustrated in Fig. 1.

The network consists of an input layer that represents the lexical context, a hidden layer, an output layer that generates a morpheme based on the lexical input it is receiving, and two context layers that represent the state of the hidden layer on the previous time step and the state of the output layer on the previous time step. These layers function as a memory for the model. Consequently, its context layers allow it to produce sequential structures.

On each cycle of the model, activation propagates from the lexical layer to the hidden layer and from the hidden layer to the output layer. Activation to each node was computed using the logistic activation function. After each cycle, the activations of nodes in the hidden layer are copied to the internal state layer. The activations of nodes in the output layer are copied to the output state layer. On the next cycle, activation from both the internal state layer and the output state layer propagate their activation along with that of the lexical layer to the hidden units. This allows for the state of the units in the hidden and output layers on previous cycles to influence processing of hidden units on the current cycle, giving the model a memory (see Elman 1990; Jordan 1986). Dell et al. (1993) tested versions of the model in Fig. 1 with both internal state and output state layers and with just an internal state layer. They found that errors produced by the former more closely matched the errors produced by speakers, so we use the same architecture here.

The model was trained using the back-propagation learning algorithm (Rumelhart et al. 1986). The input layer consisted of two nodes, one for each lexical item to be produced (e.g., layout vs. layover). The hidden layer consisted of seven nodes, as did the internal state layer. The output layer consisted of four nodes, as did the output state layer. The four nodes of the output layer corresponded to each of the morphemes in the two-word vocabulary (e.g., lay, over, out) as well as a node that corresponded with a word boundary.

The training vocabulary consisted of two words. The model was trained to produce the two words in alternation (layover–layout–layover–layout–etc.). Models were trained on two-word vocabularies in order to determine, in general, how final and initial overlap impact production difficulty. Although we would expect to find similar effects in models with larger vocabularies, by examining models with two words, we can focus specifically on effects of overlapping representations on production rather than effects of other factors such as interactions across lexical items or model memory constraints. In addition, a two-word vocabulary allowed the learning phase of the model to more closely match the task performed by participants, which we discuss below, a production task in which two lexical items are produced in sequence.

On each cycle, the input node corresponding to the target word was activated. This activation occurred across three time steps to produce the two morphemes of the word and the word boundary (e.g., lay, over, word boundary, in that order). Training ended after 200 epochs, which included two productions of the two compound words each (i.e., "layout layover layout layover" 200 times).

Two types of models were trained. One group of models was trained to produce two words that overlapped in their initial morphemes (e.g., layout and layover). Another group of models was trained to produce two words that overlapped in their final morphemes (e.g., outlay and overlay). At test, the models were given a target two-word sequence to produce. We used the mean summed squared error of the output nodes as an indicator of overall model difficulty in producing each of the morphemes. Figure 2 displays the average summed squared error for ten models trained on words that overlap initially and for ten models trained on words that overlap finally.

The overall pattern replicates what one would expect from Sevald and Dell's (1994) results: Words that overlap in their initial segments are more difficult to produce than those that overlap in their final segments, and summing the squared error over the three regions yields in error of 1.1879 for the initial overlap condition and 1.0061 for the final overlap condition. The second thing to note is that both models predict more difficulty at points at which the words do not overlap than at points at which they do, predicting that the most distinctive part of the word should be the one that is the most difficult to generate for speakers.

As in the model proposed by Sevald and Dell (1994), the serial nature of phonological encoding readily explains this pattern. Representational similarity creates more difficulty when it occurs earlier in the word. In the SRN, retrieved material on the previous cycle serves as a partial cue for retrieving material at the present cycle. Thus, the input to the hidden layer for "layover" and "layout" is similar at the points of the second morpheme, and this representational similarity leads to more difficulty in producing it. In contrast, for words that overlap finally, this representational similarity does not occur until the word is near completion, and thus, does not create interference. Thus, the difference in difficulty in producing words that overlap initially and finally is the result of competing representations between words as suggested by Sevald and Dell (1994) in their interactive model.

Thus, a production model based on a simple recurrent network architecture mirrors the performance of human speakers (see Sevald and Dell 1994). However, the

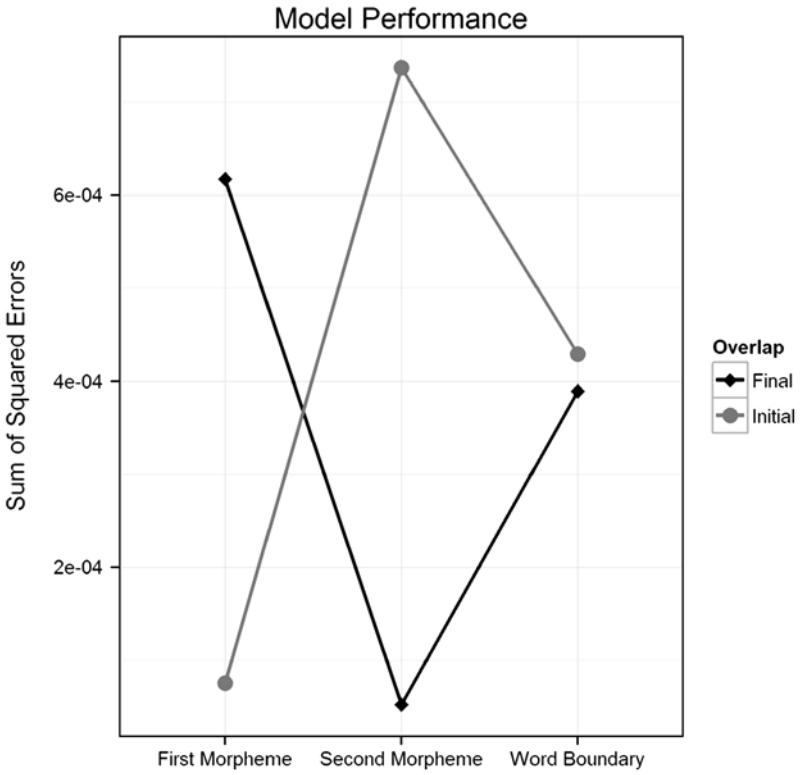


Fig. 2 The average summed squared error across output nodes for ten models trained on words that overlap in their initial morpheme (layout–layover) and ten models trained on words that overlap in their final morpheme (outlay-overlay)

present question is whether the relative duration of points within a word reflects this speaker difficulty. If it is the case that the sequential nature of phonological encoding influences the timing of word production, we would expect points at which the model has difficulty producing a word like "layout" to predict the relative duration of points within the word. For both sets of models, producing the nonoverlapping part of the word resulted in the most error. This is likely, in part, a frequency effect: The overlapping morpheme was far more frequent in the input than the nonoverlapping morpheme. In addition, nonoverlapping morphemes were the most distinctive parts of the words, so the contexts that preceded them (and their inputs to the hidden layer) tended to be similar resulting in increased difficulty, another interference effect.

Thus, two effects are predicted: (1) Speakers should produce words that overlap initially with longer duration than words that overlap finally and (2) we expect greater slow downs on parts of the word that do not overlap in the two conditions. In the experiment described below, speakers produced alternating word sequences that overlapped either initially or finally such as "layout layover layout layover…" or "outlay overlay outlay overlay…." We investigated how this overlap affected word duration across these two conditions.

2 Method

2.1 Participants

Fifteen undergraduates from the University of Illinois Urbana–Champaign participated in the study for class credit.

2.2 Materials

Because we were interested in the effects of word overlap on duration, it was critical to control for other factors known to affect word duration such as lexical stress, metrical stress, and phonological context. To control for the phonological context of the target words, we created word sets in which reversing the order of the morphemes in the compound produced English words that overlapped either initially or finally such as in (1) below, so that the strings produced across conditions were matched and varied only in their order:

1a) layover layout
1b) overlay outlay

There were a total of six items within each condition. The words differed across conditions, but as in (1), these words were matched with respect to the morphological components that were used.

In addition, unlike in the modeling data, we could not simply compare performance across the two conditions. Word duration is affected by metrical stress, syllable position, and other potential confounding factors. Thus, target words, such as "layout" and "layover," were compared to baseline conditions that included one of the critical words. This word was paired with a compound with which it did not overlap morphologically. Thus, for the string "layout layover," the duration of "layout" in critical trials was compared to "layout" in a baseline condition "layout handover." Similarly, a baseline was constructed for the other member of the target set: The baseline for "layover" was "layover handout." Baseline conditions were also constructed for the final overlap conditions. Thus, the data presented below represent the difference in duration between the target words in initial and final overlap conditions and their respective baseline conditions. All the conditions and their baselines are listed in (2) in an example item (all six critical items and their baseline conditions are listed in the Appendix):

2a) Initial overlap: *lay*over *lay*out
2b) Final overlap: over*lay* out*lay*
2c) Initial baseline 1: *layover* handout
2d) Initial baseline 2: *layout* handover
2e) Final baseline 1: *overlay* handout
2f) Final baseline 2: *outlay* handover

A within-participant design was used. In addition, all participants produced both conditions for each item as well as their associated baseline conditions. This was done to reduce potential inter-speaker differences in pronunciation and speech rate. These stimuli were presented to subjects with 50 distractor items that consisted of unrelated compounds (e.g., horseshoe nightshade, hindsight staircase).

Two factors were counterbalanced across participants. One was the order in which the critical items were presented to participants (layover layout vs. layout layover), which yielded two lists. Critical and baseline items were randomized within these two lists. In order to counterbalance the order of presentation, two more lists were constructed with the items presented in reverse order, yielding a total of four lists.

2.3 Procedure

Each trial consisted of a single word pair that participants were told to say aloud as quickly as possible and as many times as possible without errors. Participants were given 8 s to speak. Participants completed several practice trials before proceeding to the main body of the experiment, which consisted of a total of 86 trials.

To analyze durations, each morpheme of each compound was labeled in Praat (Boersma and Weenink 2012), a speech-analysis platform. Morpheme duration was automatically extracted using a script.

3 Results

Out of 15,316 morphemes, a total of 323 word errors were made. These errors were excluded from the analysis, leaving 97.89% of the data in the analysis. Errors included disfluencies within a word, coughs during production, producing an incorrect word, producing only part of the target compound, and producing a correct word but with neighboring incorrect words. The remaining data were analyzed using multilevel linear mixed effects models with fixed effects of trial type (target vs. baseline), location of the morpheme in the word (first vs. second), and where in the word the morphological overlap occurred (final vs. initial morpheme). All three factors were centered. Reported p values were obtained by assuming that, given the number of observations, the t-distribution approximated a z-distribution. Following the recommendations of Barr et al. (2013), the maximal random effects structure was used.

The fixed effects are presented in Table 1. Overall, there was a bias towards producing the second morpheme with longer duration than the first. This was true in all conditions, and probably reflects a metrical structure imposed on the words by the participants given the repetitive nature of the task.

Critically, there was a reliable three-way interaction between trial type, morpheme position, and overlap ($t = 1.97$, $p < 0.05$). The morpheme durations are displayed in Fig. 3.

Table 1 Fixed effect estimates for multi-level model of participant durations

	Estimate	Standard error	*t* value
Intercept	0.2596	0.0079	32.80
Overlap location	0.0022	0.0096	0.23
Target (vs. baseline)	0.0021	0.0047	0.46
Morpheme location	0.0472	0.0107	4.42
Overlap * target	0.0095	0.0099	0.96
Overlap * morpheme location	0.0155	0.0194	0.80
Target * morpheme location	0.0004	0.0118	0.03
Overlap * target * morpheme location	0.0471	0.0239	1.97

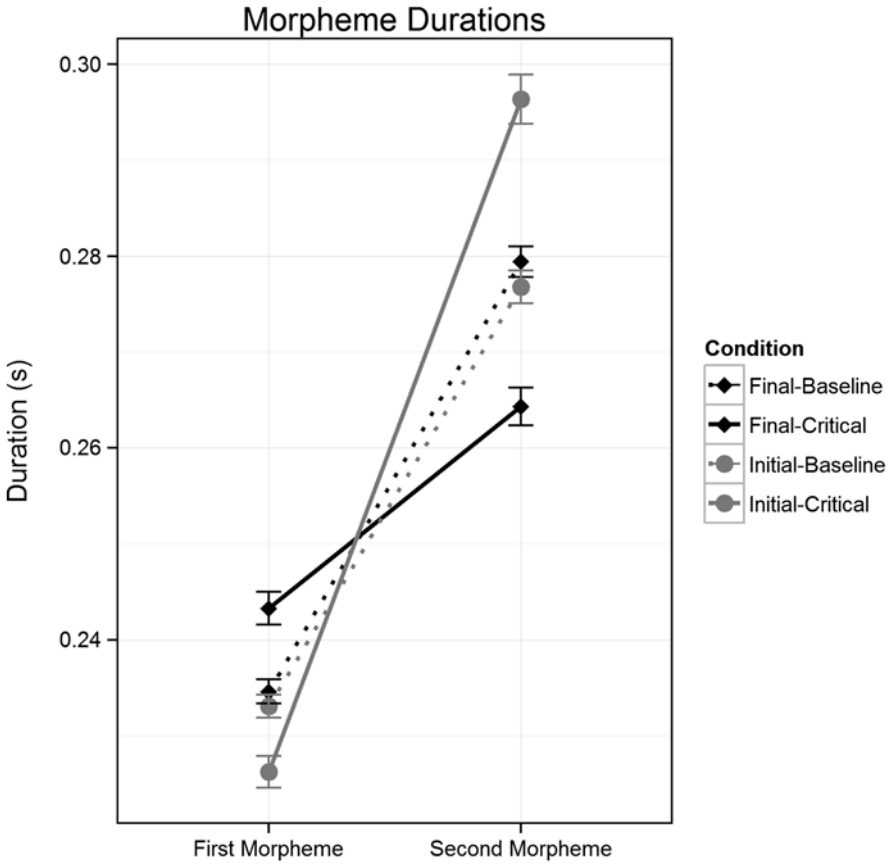


Fig. 3 The mean duration in seconds of morphemes across trial. Error bars show standard errors

In the condition in which morphemes overlap initially, target durations were longer than the corresponding baseline condition at the second morpheme. In contrast, in the condition in which morphemes overlap finally, target durations were shorter

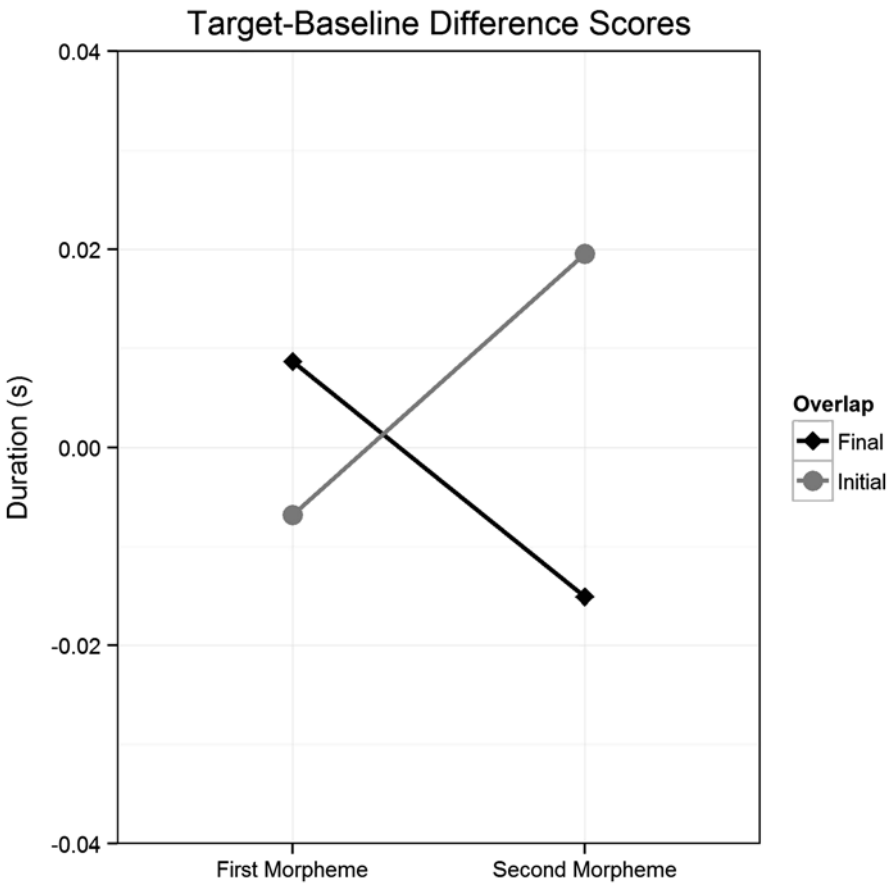


Fig. 4 A difference score between the durations of the morphemes in the initial and final overlap conditions and their corresponding baseline conditions

than the corresponding baseline condition at the second morpheme. The opposite pattern is true at the first morpheme: The initial overlap condition is shorter than its corresponding baseline control while the final overlap condition is longer than its corresponding baseline control. To highlight these differences, Fig. 4 displays a graph of the durations of initial and final overlap conditions with the duration of their baselines subtracted.

Note that the relative durations of the first and second morpheme in both conditions closely match the error predictions of the model in Fig. 2: The model experiences the most difficulty on the portion of the word that does not overlap, and we find that this matches the human duration data.

Finally, contrary to predictions, there was no overall difficulty effect of overlap. It was not the case that the overall difference between the initial overlap condition and its baseline ($M = 0.0036$s) was significantly different than the final overlap con-

dition and its baseline ($M = -0.0051$s) ($t = 0.96$, $p > 0.05$). The effect was numerically in the right direction though it was not reliable. Previous results have found that initial overlap leads to shorter durations than final overlap (e.g., Sevald and Dell 1994; O'Seaghdha and Marin 2000). It is possible that this effect did not reach significance in the current study because of power: Only six items could be used in each condition because there are very few compound pairs in English that overlap initially and, when reversed, overlap finally and still yield real English words. These design constraints may have limited our ability to detect an effect if it was present.

4 Discussion

This chapter began with a puzzle: If variability in the duration of a word is linked to production difficulty, and lengthening difficult words facilitates production, why does this lengthening occur during word production rather than before it?

The goal of this chapter was to demonstrate that production processes involved in phonological selection can provide a partial explanation for duration differences both within and across words. If phonological selection is a serial process that inevitably varies in complexity at different time points, lengthening at points of uncertainty might facilitate production by giving the system more time to converge on selecting the correct phonemes. The simple recurrent network presented above predicted that words that overlap initially should be more difficult to produce than words that overlap finally. It also predicted that production should be more difficult in the regions of the word that do not overlap. We found that participants that produced word pairs in contexts similar to the model slowed down in exactly the points at which the model predicted production difficulty. The fact that points of complexity correspond with lengthening suggests that some durational choices by speakers may be attributable to the process of phonemic encoding.

As discussed above, we did not replicate the effect of overlap found in previous studies (e.g., Sevald and Dell 1994; O'Seaghdha and Marin 2000), though this was predicted by the model. This may have been due to insufficient power. However, it is encouraging that the SRN correctly accounts for the findings of previous work: Initial overlap leads to more difficulty than final overlap. Furthermore, the model correctly predicts that the nonoverlapping morphemes of the compounds should be produced with longer durations than overlapping morphemes.

Note that we are not arguing that production constraints are the only factors that affect word duration. Factors like word or lemma frequency, speech rate, and communicative factors such as those outlined in Aylett and Turk's (2004) smooth signal hypothesis almost certainly contribute to the duration of a word. Nevertheless, complexity in the production system could help explain why at least some of these factors contribute to the changes in word durations.

Another question is understanding the level of production at which these duration effects arise. Above, we attribute the duration effects to mechanisms linked to serially ordering phonological information; however, these data are also consistent

with complexity in the ordering of any sub-lexical linguistic production process (e.g., phonological, articulatory, morphological, or syllabic representations). Furthermore, these data are also compatible with certain types of non-serial production processes. That is to say, these data are compatible with sub-lexical linguistic production routines not reaching completion or occurring at different stages. The process of selection might only be partially serial, and still yield the types of differential lengthening we see across words. For example, the phonemes that are most highly activated might be selected first while phonemes that are less activated are only selected at a later stage. Both processes might occur during articulation, but the system itself is not entirely serial. Such an architecture is consistent with the larger point being made here: That duration choices allow time for linguistic selection, but they do not necessarily assume a fully serial architecture. We leave the question of what level of representation and the degree of seriality in the production system open to future investigation.

Finally, it is important to note that the SRN presented above is not meant to be a model of the representations that are engaged in language production. Although the model demonstrates that the difficulty of ordering linguistic information is sensitive to overlap between compounds and that these map onto speakers' durational choices, it does not necessarily model the actual mechanisms that are engaged in language production. One approach for developing a process model would be to adapt Dell's (1986) model so that it captures some of the effects in this chapter, as well as the effects reported in Sevald and Dell (1994) and O'Seaghdha and Marin (2000). This might serve as a useful next step in understanding how production algorithms lead to differences in duration.

Overall, the SRN and the behavioral data point towards a link between production processes and duration. Although there are claims in the literature that such a link exists, up until now, there has been relatively little work specifying exactly how lengthening and reduction are linked to the process of speaking. The work here represents a first step in spelling out the mechanisms that underlie this link.

Acknowledgments We would like to thank John Hummel and Gary Dell for their comments and advice on the modeling component of the chapter. This work was supported by NIH grant R01 DC008774 and a grant from the James S. McDonnell foundation.

Appendix

Items

Initial overlap: *lay*over *lay*out
Final overlap: over*lay* out*lay*

(1) Initial baseline: layover handout
(1) Final baseline: overlay handout

(2) Initial baseline: layout handover
(2) Final baseline: outlay handover

Initial overlap: *turn*down *turn*over
Final overlap: down*turn* over*turn*

(1) Initial baseline: turndown takeover
(1) Final baseline: downturn takeover
(2) Initial baseline: turnover takedown
(2) Final baseline: overturn takedown

Initial overlap: *set*off *set*up
Final overlap: off*set* up*set*

(1) Initial baseline: setoff holdup
(1) Final baseline: offset holdup
(2) Initial baseline: setup handoff
(2) Final baseline: upset handoff

Initial overlap: *over*cross *over*hang
Final overlap: hang*over* cross*over*

(1) Initial baseline: overcross crisscross
(1) Final baseline: hangover crisscross
(2) Initial baseline: overhang uphang
(2) Final baseline: crossover uphang

Initial overlap: *out*stand *out*break
Final overlap: stand*out* break*out*

(1) Initial baseline: outstand daybreak
(1) Final baseline: standout daybreak
(2) Initial baseline: outbreak kickstand
(2) Final baseline: breakout kickstand

Initial overlap: *out*look *out*sell
Final overlap: look*out* sell*out*

(1) Initial baseline: outlook undersell
(1) Final baseline: lookout undersell
(2) Initial baseline: outsell overlook
(2) Final baseline: sellout overlook

References

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47*, 31–56.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure in mixed-effects models: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language, 60,* 92–111.

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer [Computer program]. Version 5.3.32. http://www.praat.org/. Accessed 17 Oct 2012.

Dell, G. S. (1986). A spreading activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S., Juliano, C., & Govindjee, A. (1993). Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science, 17,* 149–195.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211.

Fowler, C. A., & Housum, J. (1987). Talkers' signaling of "new" and "old" words in speech and listener's perception and use of the distinction. *Journal of Memory and Language, 26,* 489–504.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61,* 23–62.

Jaeger, T. F., Furth, K., & Hillard, C. (n.d.). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* (in press).

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach. Tech. Rep. No. 8604*. San Diego: University of California, Institute for Cognitive Science.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee & P. Hopper (Eds.). Frequency and the emergence of linguistic structure (pp. 229–254). Amsterdam: John Benjamins.

Kahn, J., & Arnold, J. E. (n.d.). Speaker-internal processes drive durational reduction. *Language and Cognitive Processes* (in press).

Kahn, J., & Arnold, J. E. (2012). A processing-centered look at the contribution of givenness to durational reduction. *Journal of Memory and Language, 67,* 311–325.

Lam, T. Q., & Watson, D. G. (n.d.). Repetition reduction? *Journal of Experimental Psychology: Learning, Memory, & Cognition* (in press).

Lam, T. Q., & Watson, D. G. (2010). Repetition is easy: Why repeated referents have reduced prominence. *Memory & Cognition, 38,* 1137–1146.

O'Seaghdha, P. G., & Marin, J. W. (2000). Phonological competition and cooperation in form-related priming: Sequential and nonsequential processes in word production. *Journal of Experimental Psychology: Human, Perception, and Performance, 26,* 57–73.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America, 118,* 2561–2569.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propogation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge: MIT.

Sevald, C. A., & Dell, G. S. (1994). The sequential cuing effect in speech production. *Cognition, 53,* 91–127.

Prosody and Intention Recognition

Michael K. Tanenhaus, Chigusa Kurumada and Meredith Brown

Abstract Listeners face multiple challenges in mapping prosody onto intentions: The relevant intentions vary with the general context of an utterance (e.g., the speaker's goals) and how prosodic contours are realized varies across speakers, accents, and speech conditions. We propose that listeners map acoustic information onto prosodic representations using (rational) probabilistic inference, in the form of generative models, which are updated on the fly based on the match between predictions and the input. We review some ongoing work, motivated by this framework, focusing on the "It looks like an X" construction, which, depending on the pitch contour and context, can be interpreted as "It looks like an X and it is" or "It looks like an X and it isn't." We use this construction to investigate the hypothesis that pragmatic processing shows the pattern of adaptation effects that is expected if the mapping of speech onto intentions involves rational inference.

Keywords Pragmatics · Contrastive focus · Adaptation · Probabilistic inference

In a note to the speakers before the workshop, Lyn Frazier encouraged us to flag particular aspects of our proposals that we thought were novel, promising, or suspicious. Lyn also encouraged us to flag unidentified problematic assumptions in the field. In response to Lyn's suggestions, we begin with an example to illustrate some of the challenges involved in understanding the mapping of prosody onto intentions.

We thank the members of MTan Lab, Anne Pier Salverda, Delphine Dahan, and T. Florian Jaeger for the valuable discussion, and Chelsea Marsh and Olga Nikolayeva for the support with participant testing. This research was supported by NICHD grants HD27206 and HD073890 (MKT), a JSPS postdoctoral fellowship (CK), and an NSF graduate research fellowship (MB).

M. K. Tanenhaus (✉) · C. Kurumada · M. Brown
Department of Brain and Cognitive Sciences, University of Rochester, Meliora Hall, Rochester, NY 14627, USA
e-mail: mtan@bcs.rochester.edu
e-mail: mktanenhaus@gmail.com

C. Kurumada
e-mail: ckurumada@bcs.rochester.edu

M. Brown
e-mail: mbrown@bcs.rochester.edu

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_6

In the introduction to his book, *Arenas of Language Use,* Herb Clark (1992) gives a lovely example that illustrates the richness and subtlety of the context-specific-based inferences that are required for a listener to map an utterance onto the speaker's intended meaning. Clark describes a situation in which he addressed the utterance, "I'm hot," to his (then) school-age son, Damon. Clark notes that none of the plausible pre-compiled interpretations of "I'm hot" (e.g., I'm lucky; I'm on a roll, I'm uncomfortably warm; I'm saying the one thing that no child wants to hear a parent say, etc.) captures the intended (and immediately understood) meaning of his utterance. Herb and Damon were playing poker and Damon was about to make a large bet. Herb, who had uncharacteristically been winning most of the hands, was warning Damon that he should think twice about making that bet.

As this anecdote illustrates, context plays a central role in determining speaker meaning. Therefore, understanding pragmatic inference requires us to tackle some of the most difficult questions in real-time language processing. For example, how do listeners (and speakers) determine what aspects of a context are relevant? And, since pragmatic inference involves considerations of not only what the speaker said but also what she chose not to say instead, how do listeners determine those likely alternatives (Grice 1975)?

Examining how listeners map prosody onto likely speaker intentions adds an additional set of challenges. Herb is unlikely to have uttered "I'm hot" with the "canonical" prosodic contour that he would use in an utterance that was intended to be a simple assertion, as in Fig. 1a. Instead, some aspects of the prosodic contour, including choice of pitch accents and boundary tones, probably signaled that his utterance should be interpreted in a noncanonical way (e.g., Fig. 1b or c). There is no way to know exactly *how* Herb said what he said, but it would likely interact with other factors, for example whether he raised his eyebrows, how he might have gestured, and various Herb-centric aspects of his speech. It seems unlikely, for example, that Janet (Fodor) would use exactly the same prosodic contour, let alone the same linguistic expression, were she in a similar situation and intending to convey the same information.

A fundamental difficulty in studying prosody-based intention recognition is that there are few, or perhaps no, discrete units in the prosodic signal itself. The generally accepted phonological categories for prosodic analysis, such as pitch accents and boundary tones, consist of bundles of features (e.g., pitch, duration, intensity), all of which shift in a gradient manner. Consequently, many of the representations exhibit overlapping acoustic features and/or overlapping interpretations (e.g., Watson et al. 2008). Furthermore, actual realization of prosody varies along multiple dimensions such as the gender, age, and social background of speakers. The same speaker also shifts prosodic uses according to contexts and speech conditions (e.g., child-directed vs. adult-directed speech). To our knowledge, how listeners can extract subtle but pragmatically meaningful acoustic variations in the presence of this substantial variability in the acoustic realization of speech remains an unresolved question.

In an ongoing line of research, we have been addressing a set of related questions on how listeners use prosodic information as they process an unfolding

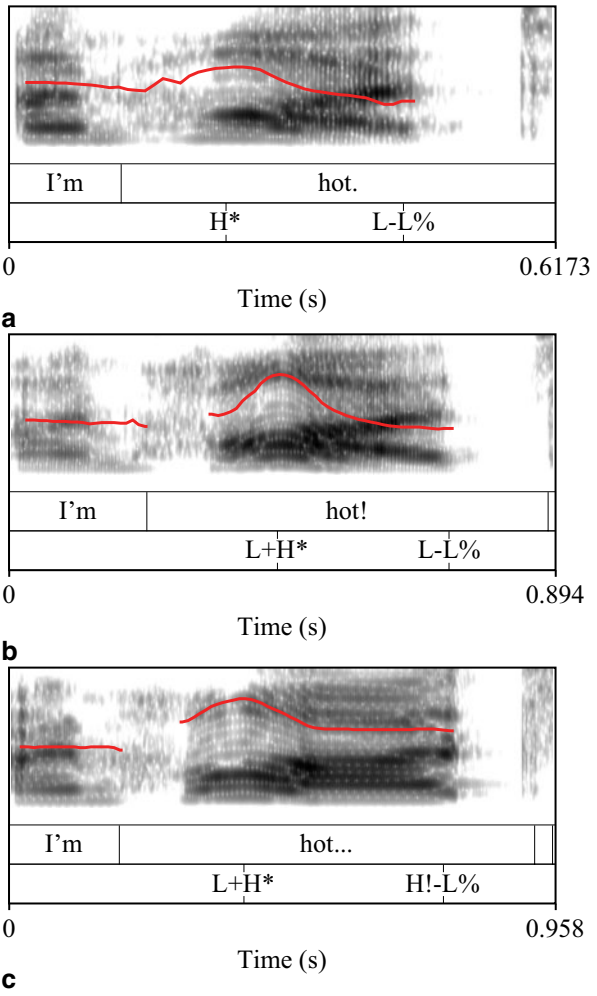


Fig. 1 Spectrograms, pitch contours, and ToBI labels for the phrase "I'm hot" uttered with three different prosodic contours. **a** Depicts "canonical" statement prosody, whereas the less canonical prosodic contours in **b**–**c** may be more likely to trigger pragmatic inferences (e.g., that the speaker is surprised or is advising caution). *ToBI* tones and break indices

utterance which might trigger a pragmatic inference. What is the nature of prosodic representations that support pragmatic inferences? How does the prosody (in relation to the lexical content of the sentence) serve as a cue to unstated speaker meaning? And, how do we arrive at a particular pragmatic interpretation when we hear the particular combination of prosodic features? Underlying these questions is the core inquiry about the mechanism of a language comprehension system: How can we achieve a robust mapping between the realization of speech sounds and phonetic or phonological representations as well as the mapping between these representations and possible interpretation given the constraints provided by the relevant context? We begin by sketching out some of the assumptions that have been guiding our work.

1 Our Framework

As we mentioned above, one of the biggest challenges to any classification model of prosodic categories is the ambiguity arising from the continuous and variable nature of the information. In developing our approach to prosody, we have built upon recent work on phonetic categorization (e.g., Clayards et al. 2008; McMurray and Jongman 2011; and especially Kleinschmidt and Jaeger 2011, 2012; 2015). We draw three analogies between prosodic interpretations and speech perception. First, prosodic representations, such as different pitch accent types, boundary tones, and contours, are best characterized as distributions of relevant acoustic cue values. Therefore, just as distributions of acoustic cue values for phoneme representations (e.g., voice onset time for /p/ and/b/) show some overlap, and vary with surrounding acoustic cues, e.g., the duration of the preceding and following vowel, prosodic categories form overlapping distributions that can vary with the phonetic context (Liberman and Pierrehumbert 1984). The distributional hypothesis explains how acoustically highly variable, and sometimes ambiguous, input can be grouped into two or more functionally contrasted abstract representations.

Second, categorical perception of prosodic representations is an outcome of inferences rather than a property of the acoustic signal itself. It is widely established that perception and recognition of phonemes are dependent not only on the acoustic information but also on a wide range of contextually derived expectations. For instance, listeners integrate information such as lexical status of the carrier word (Ganong 1990; Connine and Clifton 1987; Miller et al. 1984), lexical effects on compensation for coarticulation (Elman and McClelland 1988), gender of the speaker (Kraljic and Samuel 2007; Strand and Johnson 1996), and information structure of the sentence (Brown et al. 2015). Most generally, even the most robust cue-integration mechanisms cannot account for how a contrast such as voicing is perceived without taking into account expectations (McMurray and Jongman 2011). We hypothesize that, in prosodic processing too, listeners integrate the bottom-up prosodic cues and top-down contextual expectations to inferentially arrive at a particular representation. More specifically, contextual information is expected to serve two important roles. First, it enables the listener to predict what interpretations are likely to be conveyed and how they will be encoded via prosody. For instance, an utterance following a question (e.g., What about beans? Who ate them? (Jackendoff 1972)) is likely to contain a pitch movement signaling an informational focus (e.g., JOHN ate the beans). Second, contextual information is used to resolve perceptual ambiguity resulting from prosodic variability and noise. Even when prosodic information is ambiguous or partially lost in noise, listeners can recover an intended interpretation relying on their contextual knowledge.

Third, prosodic processing is highly plastic: Listeners flexibly adapt their prosodic expectations according to recent experiences. As stated above, we hypothesize that listeners predict upcoming prosodic input based on contextual information. Upon receiving the input, listeners match it up with their prediction and compute how much their expectations deviated from the input. The amount of deviation is

then used to update expectations for future input. Indeed, it has been demonstrated that listeners adapt their percepts of phoneme categories through rapid perceptual learning (e.g., Norris et al. 2003; Kraljic and Samuel 2006; Vroomen et al. 2007; Clayards et al. 2008). We hypothesize that a similar mechanism in prosodic processing allows the comprehension system to maintain mappings between the variable perceptual input and more abstract prosodic categories that are more or less constant across speakers and contexts.

Based on these three assumptions, we propose that the pragmatic interpretation of prosodic contours can best be understood as rational inference over noisy input. Most generally, we are assuming a "data explanation framework" in which perceptual systems seek to provide an explanation for sensory data using "generative models" (Kleinschmidt and Jaeger 2015; Fine et al. 2013; Fine and Jaeger 2013; Farmer et al. 2013; Brown et al. 2015; Brown, Dilley & Tanenhaus 2015). These models evaluate hypotheses about the state of the world according to how well they could have given rise to ("generated") the observed perceptual properties. In the case of prosodic processing, a model integrates acoustic cues such as pitch, duration, and intensity to infer which pragmatic interpretation has generated the data at hand.

To better predict future input, hypotheses are continually adjusted based on the observed differences between the predicted state of the world and the observed state of the world, with the goal of minimizing prediction error. The prediction error provides a signal that is used during learning to continuously update the generative model (for similar proposals within a connectionist framework, see Chang et al. 2006; Dell and Chang 2013). This approach has been successful in explaining how listeners converge on coherent percepts in phoneme identification and speech perception—a domain in which a lack of categorical mappings between acoustic signals and linguistic categories has been investigated in great depth. It has also been applied to work in syntactic processing (Fine et al. 2013; Fine and Jaeger 2013) and the role of expectations in segmentation (e.g., Brown et al. 2012). In experiments presented below, we ask whether listeners (1) integrate a multitude of prosodic and contextual cues to constrain their inferences and (2) adjust weights of these cues according to recent exposure.

2 Does it *Look Like* Speech Adaptation?

We now present an overview of four experiments from an ongoing project that uses the construction "It looks like an X" as a case study to test and refine our hypotheses about how listeners map acoustic cues onto prosodic categories. This construction has a number of desirable properties. First and foremost, it can evoke different pragmatic meanings depending on its prosodic realization (Kurumada 2013). A canonical accent placement (as illustrated in Fig. 1, left panel, henceforth noun-focus prosody) typically elicits an affirmative interpretation (e.g., It looks like a zebra and I think it is one), hereafter the "It is" interpretation. In the context that we investigate, when the verb "looks" is lengthened and emphasized with a contrastive accent

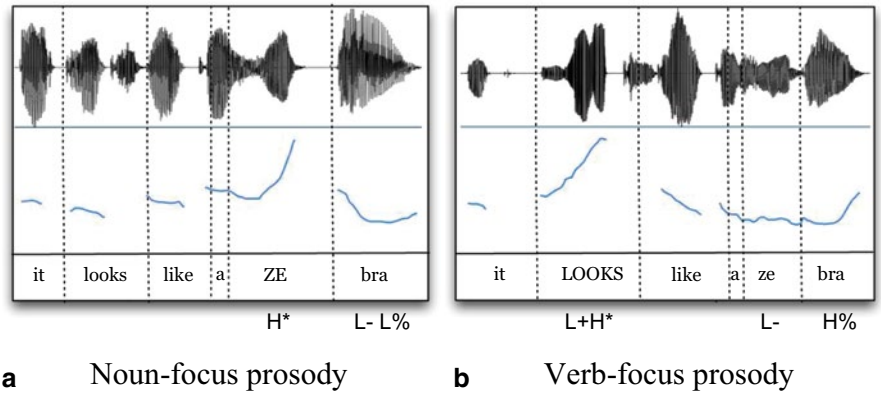


Fig. 2 Examples of waveforms (*top*) and pitch contours (*bottom*) for the utterance *It looks like a zebra*. In the context that we investigate, the affirmative interpretation "It is a zebra" is typically conveyed by the pattern on the left (**a**), while the negative interpretation "It is not a zebra" is conveyed by the pattern on the right (**b**)

(L+H*) and the utterance ends with an L–H% boundary tone (Fig. 2, right, hereafter verb-focus prosody), it can trigger a negative interpretation (e.g., It LOOKS like a zebra but it's actually not; see also Dennison and Schafer 2010). The fact that these interpretations map onto different referents, e.g., a zebra or something that only resembles a zebra, makes it possible to determine which interpretation a participant has chosen.

Second, with verb-focus prosody, we are investigating a contour that is known to evoke a contrastive interpretation: the contrastive pitch accent (fall–rise: often annotated as L+H* in the tones and break indices (ToBI) convention (e.g., Silverman et al. 1992)) followed by a rising boundary tone (L–H%). This contour can signal a contrast between referents (e.g., we have pie L+H* L"H% (but no cake); Ward and Hirschberg 1985) or predicates (e.g., Lisa HAD L+H* the bell $_{\text{L–H\%}}$ (but she no longer has one); Dennison and Schafer 2010).

The fact that both the pitch accent and the boundary tone contribute to the contrastive meaning means that we can vary the reliability of two asynchronous cues. Moreover, online comprehension of the L+H* accent has been studied extensively and it has been shown to trigger immediate eye movements to visually represented contrast items (e.g., Ito and Speer 2008; Watson et al. 2008; Weber et al. 2006). For example, as soon as listeners encounter the L+H* on a color adjective (e.g., "Pick up a blue ball. Now, pick up a YELLOW L+H*…"), they begin to fixate color-contrasted items that belong to the same object category as the previous referent.

In all of the work we will be describing we presented participants with variations on a scenario in which the "It looks like an X" utterance occurs in a context in which an adult (e.g., a teacher or a parent) is looking at a picture book with a young child. We created pairs of pictures in which one picture had a common name (e.g., a picture of zebra) and the other picture was similar-looking but did not have a common name (e.g., a picture of an okapi). The participants were instructed that the adult is

referring to a picture in the book as he or she addresses an utterance to the child. The participant then made a response to indicate which picture is being referred to.

This cover story was crucial in constraining the range of interpretations that listeners can draw. Without contextual constraints, the "It looks like an X" construction allows the speaker to express a range of nuanced meanings such as uncertainty, or a contrast between visual similarity and similarities in other domains (e.g., If it looks like a duck, walks like a duck, and quacks like a duck…). However, in this current scenario, we are implicitly imposing the assumption that the speaker (the mother) knows what the identity of a referent is, and is giving a hint to the child so that he can make an appropriate inference, i.e., it is an X or it is not an X. It is an important independent question how interlocutors negotiate underlying assumptions about speaker knowledge and domains (and subdomains) that determine the saliency and likelihoods of possible interpretations within an actual conversational context.

Another topic of interest, which we do not address here, is the extent to which the contrastive inference "It looks like an X but it is not" is conventionalized. The construction "It looks like an X" occurs more frequently than expressions like "smells like" or "sounds like," and it often conveys the fact that an appearance of an object conflicts with its identity. However, a corpus analysis of child-directed speech by Hansen and Markman (2005) found that adult speakers use "looks like" to talk about both appearance and identity, and the interpretation is most often largely context dependent. For instance, if a child were to say "It's a zebra," and an adult were to answer "It LOOKS like a zebra," then the preferred interpretation is "but it isn't a zebra." However, if the child instead had said, "It's not a zebra," then the preferred interpretation of "It LOOKS like a zebra" would be "It is a zebra." This observation has been confirmed experimentally (Bibyk et al. (in preparation)). This suggests that the interpretation of the "looks like" construction is not completely conventionalized and hence the uncertainty in the interpretation needs to be resolved through contextual inference.

We first summarize the manipulations and results of a series of off line studies which used an online crowd-sourcing platform (Amazon's Mechanical Turk) to test: (a) our assumption that verb-focus and noun-focus prosody with "It looks like an X" probabilistically maps onto different interpretations and (b) the hypothesis that listener adapt their mapping of the these contours onto interpretations based on the statistics of the input (study 1). We first established that listeners preferentially map noun- and verb-focus prosody onto the interpretations "it is" and "but it's not," respectively. We then asked whether listeners would adapt by (a) shifting the strength of their preferences when they were presented with evidence that a speaker often used a stronger alternative to signal the "it is" interpretation (study 2) and (b) down-weighting prosodic cues when exposed to a speaker who used prosody unreliably (study 3).[1]

[1] A preliminary report of these three studies, including the methodological details and results, is presented in Kurumada et al. (2012). A longer manuscript reporting these results is under review (Kurumada et al. n.d.).

3 Study 1: Prosodic Representations as Distributions of Acoustic Cues

Our most basic claim is that prosodic representations involve distributions of relevant acoustic cues such as pitch, duration, and intensity. Acoustic cues should therefore map probabilistically onto different interpretations and listeners should be sensitive to properties of the distribution. Crucially, listeners should therefore adapt the mapping of contours onto interpretations according to the distribution of tokens in the input.

In order to test these hypotheses, we selected a clear exemplar of a noun-focus utterance and a clear exemplar of a verb-focus utterance for each item (e.g., zebra) and resynthesized a 12-step continuum of prosodic contours. The stimuli were divided into six regions corresponding to each of the four initial words (i.e., it | looks | like | a) and the portions of the final word associated with each of the two tonal targets (i.e., the H* and L–L% in the noun-focus contour and the L and H% in the verb-focus contour). The turning point in the f0 contour within the final word was used to delineate the final two regions. The f0 of each region was sampled at 20 equally spaced time points, and measures from each time point were aggregated across items to derive mean f0 contours for noun-focus and verb-focus utterances (following Isaacs and Watson 2010). Likewise, the durations of each region were averaged across items by contour type. Twelve-step continua for each item were derived from these mean f0 contours and durations by interpolating between values within each region and then manipulating the f0 and duration of each recording to match the interpolated values using the pitch-synchronous overlap-and-add algorithm implemented in Praat (Moulines and Charpentier 1990; Boersma and Weenink 2008). A schematic of the f0 contours for a sample item are presented in Fig. 3.

We first established a categorization function for the continua illustrated in Fig. 3. We used these results to postulate distributions for how the phonetic contours map onto the "it is" and "but it's not" interpretations. These are illustrated in Fig. 4, panel (a). We then selected two distributions of tokens for presentation to separate groups of participants. The shaded areas in panels (a) and (b) of Fig. 4 correspond to the values along the *X*-axis used in exposing a new set of participants to either the distribution of contours presented in Fig. 4c (the *affirmative-bias* condition) or Fig. 4d (the *negative-bias* condition). During exposure, participants heard an "It looks like an X" utterance and chose which picture they believed the teacher was intending to refer to. After making a picture selection, participants heard a second clause that disambiguated the intended referent (e.g., "because it has black and white stripes" or "but it isn't because it only has stripes on its legs"). Participants were then presented with 12 new tokens and made picture selections without feedback. The distribution of exposure tokens in the affirmative-bias condition (Fig. 4c) was chosen to mirror the distribution that we postulated based on the norming results. In the negative-biased condition, illustrated in Fig. 4d, we presented ambiguous tokens from steps 7, 8, and 9 with feedback indicating that the speaker intended

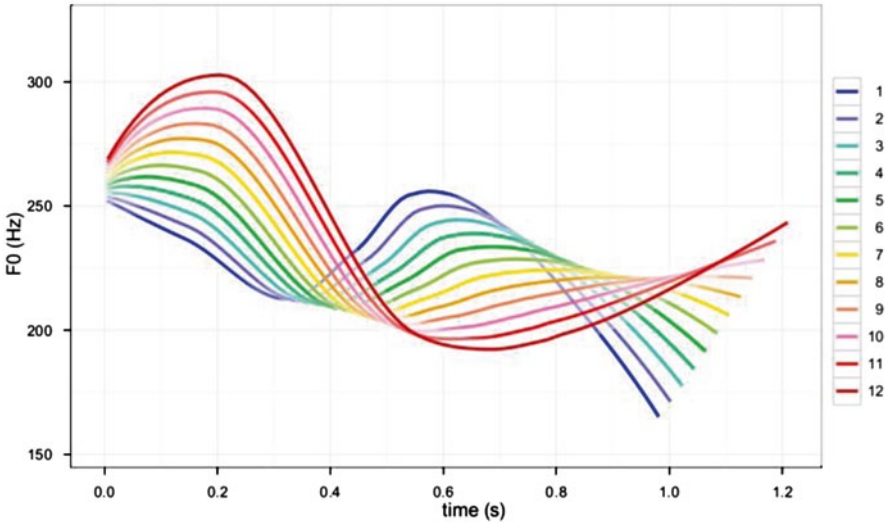


Fig. 3 Schematic illustration of the manipulation of f0 and duration in resynthesizing a 12-step continuum of prosodic contours. The numbers represent the 12 steps. Step 1 and step 12 represent mean prosodic cue values associated with typical noun-focus and verb-focus prosodic contours, respectively

to refer to the atypical referent (e.g., the okapi). The predicted effect of exposure on the distributions is illustrated in Fig. 5. Figure 5a shows the categorization function from the norming study and the predicted shift in the categorization function after exposure to the negative-biased tokens.

Figure 6 presents the categorization functions for the norming study, the affirmative-bias exposure condition, and the negative-bias exposure condition. The affirmative-bias exposure condition was chosen to mirror the assumed preexposure distribution, whereas the negative-bias exposure condition was predicted to shift participants' categorization functions. The results closely mirror our predictions.

In sum, the results of this study establish that listener's mapping of prosodic tokens onto interpretations is probabilistic and malleable according to recent exposure. Most crucially, listeners are sensitive to the distribution of new tokens, showing the predicted adaptation effects.

4 Effect of Alternatives

Another source of variability is the fact that the pragmatic interpretation of speaker meanings relies on the listener's estimates of what kind of lexical, syntactic, and prosodic elements the speaker could have produced. For example, the contrastive interpretation of "it looks like an X" depends upon an implied contrast between

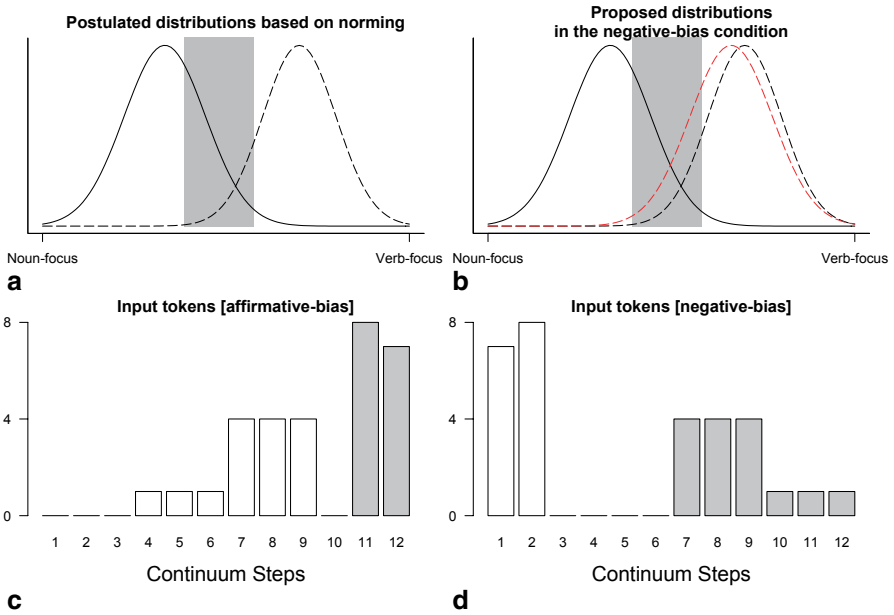


Fig. 4 **a** A schematic representation of distributions of prosodic cue values postulated based on the results of native speakers' judgments. The *solid line* and the *dashed line* represent the distributions of prosodic cue values for the "it is" and "but it's not" interpretations, respectively. **b** Proposed experimental manipulation of contour distributions. **c** and **d** Input frequencies of tokens sampled from each step of the continuum in the training phase of the affirmative-bias and negative-bias conditions. *X*-axis: continuum steps. *Y*-axis: Token frequencies of input utterances. Tokens indicated as *white bars* were disambiguated as affirmative interpretation and those indicated as *shaded bars* were disambiguated as negative interpretation

potential alternative predicates. Specifically, the contrastive accent on "LOOKS" signals a contrast between "(it) looks like (an X)" and its semantically stronger alternative (e.g., "it is (an X)"). This contrast supports the reasoning that the speaker could have said, "it is a zebra" but did not, which implicates that the speaker meant it was not a zebra. The availability of the contrastive interpretation of "it looks like a zebra" thus hinges on the listener's belief about how likely the speaker would say "it is an X" if that is what she meant. If it is likely, the form "it looks like an X" is more strongly associated with the contrastive interpretation. On the other hand, if the listener does not believe that the speaker would use "it is an X," the formal contrast does not support the contrastive inference.

To test this hypothesis, in study 2, we directly tested the effect of semantic alternatives. In this experiment, we used only prototypical instances of the noun-focus and the verb-focus prosody. Participants were presented with a cover story in which a male teacher described animals and objects in an encyclopedia with pictures that were not directly accessible to his students. In response to a question from a child about what he saw on the page, the teacher said, "It looks like an X" (e.g., It looks like a zebra). The participants' task was to judge whether the teacher was referring

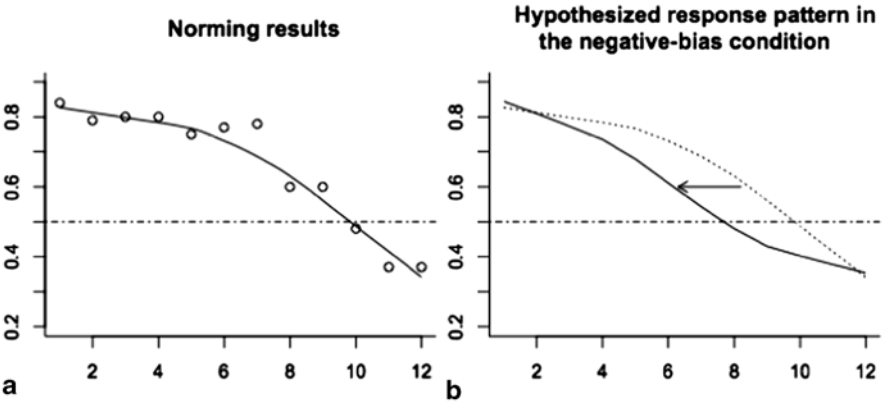


Fig. 5 **a** Proportion of a target picture chosen (affirmative interpretation) in the norming study. *X*-axis: Continuum steps (1 = prototypical noun-focus prosody, 12 = prototypical verb-focus prosody). *Solid line* represents LOWESS smoothing and *dashed line* indicates where the stimuli elicit most ambiguous responses (50 % chance of a target picture chosen); **b** a hypothesized pattern of category recalibration in the negative-bias condition

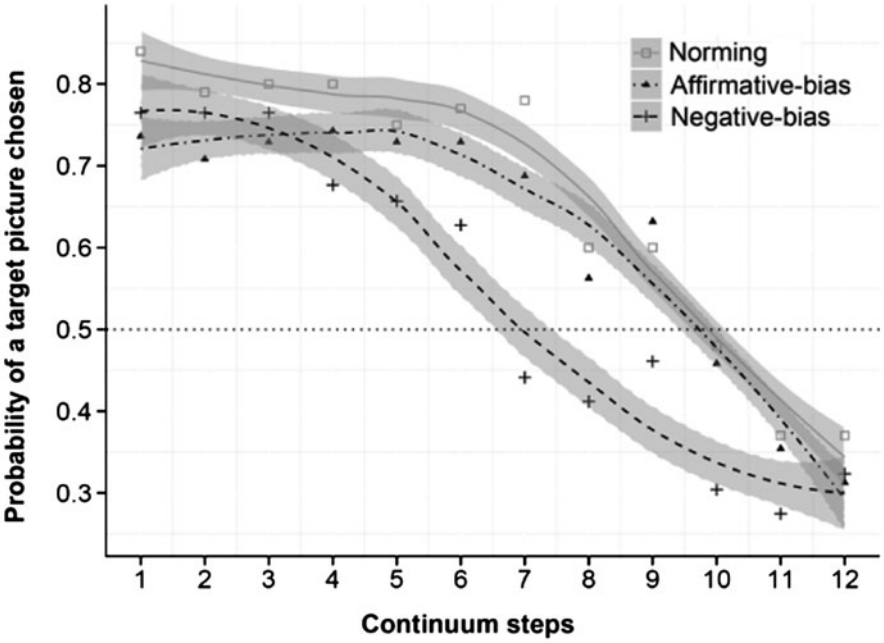


Fig. 6 Proportions of target pictures (e.g., zebra) chosen in the test phase. *Dotted, solid,* and *dashed lines* represent responses from the norming, the affirmative-bias and the negative-bias conditions, respectively. Continuum steps are plotted on the *X*-axis (step 1: prototypical noun-focus prosody; step 12: prototypical verb-focus prosody)

to the typical or the atypical referent. Each participant heard 24 utterances, 12 each with noun-focus and verb-focus prosody. The items were rotated through presentations lists so that different participants heard the version of the utterance with noun- and verb-focus prosody, respectively.

For utterances with noun-focus prosody, participants preferred the typical referent, choosing it on about 70 % of the trials, whereas with verb-focus prosody participants preferred the atypical referent, choosing the typical referent on only 40 % of the trials. Thus, the mapping of the contours onto interpretations was again probabilistic, with noun-focus prosody preferentially mapping onto the "it is" interpretation and verb-focus prosody preferentially mapping onto the "but it's not" interpretation. These results set the stage for testing our prediction that listeners would adapt by shifting their representations if they were presented with evidence that the speaker will use a less ambiguous, i.e., stronger alternative when he is expressing the "It is" interpretation.

We tested this hypothesis by replacing 8 of the 12 noun-focus utterances with the stronger statement, "It is an X" (e.g., "It's a zebra!"). We then compared the judgments for the subset of items that could be directly compared to the prior study, by excluding the results from the "It is an X" trials. For details of the design, see Kurumada et al. (2012; n.d.). As predicted, the stronger alternative shifted the preferred referent of "It looks like an X" towards the atypical referent for both noun-focus and verb-focus prosody. The typical referent was now chosen on only 40 % of the trials for noun focus prosody and 20 % of the trials for verb focus prosody. These results are consistent with our proposals that direct evidence for use of "it is an X" increases the likelihood that listeners will derive the contrastive inference based on the construction "it looks like an X." In sum, the same prosodic contour produced in the same context can be interpreted differently depending on the listener's expectation about what kind of lexical, syntactic, and prosodic means the speaker could use to express a particular intention. These results are not predicted by approaches that derive an intonational interpretation based solely on the mappings between a prosodic contour, and/or combination of pitch accent and boundary tone (e.g., L+H* L−H %), and a pragmatic meaning or category, e.g., contrast.

5 Study 3: Speaker Reliability

As we discussed in the introduction, acoustic realizations of prosodic information varies across speakers and contexts. When talking to a young baby, adult speakers tend to use a wider pitch range with more peaks and troughs in a pitch contour (Fernald and Kuhl 1987). In such a context, a small excursion of pitch may not be pragmatically meaningful and, hence the listener needs to suspend their pragmatic interpretations, which would be warranted in adult–adult conversation. Thus, the pragmatic interpretation of prosody requires an effective "down-weighting" of prosodic information that is not a reliable indication of the speaker's pragmatic intentions.

In order to evaluate the effect of speaker reliability, we used a design in which an exposure phrase (16 items) was followed by a test phase (10 items). Participants were randomly assigned to the reliable-speaker or the unreliable-speaker condition. In either condition, during the 16-utterance exposure phase, participants made a judgment based on a "It looks/LOOKS like an X" utterance, and then heard a disambiguation continuation that either indicated the "it is" interpretation or the "but it isn't" interpretation as in study 1. In the *reliable-speaker* condition, the continuation disambiguated all eight noun-focus utterances in favor of the "it is" interpretation and all eight verb-focus utterances in favor of the "but it isn't," interpretation. In the *unreliable-speaker* condition, half of the noun-focus and half of the verb-focus utterances were followed by phrases that disambiguated the utterance in favor of each interpretation. Thus, the participants were receiving feedback that the speaker's use of prosodic patterns did not provide reliable information about her intended referent.

The exposure phrase was followed by a ten-utterance test phase in which no feedback was provided after the participant's judgment. We predicted that participants in the unreliable condition would place less weight on the prosodic information in their judgments. This prediction was confirmed. With the reliable speaker, the typical referent was chosen on 82% of the utterances with noun-focus prosody compared to only 18% of the utterances with verb-focus prosody. In contrast, with the unreliable speaker, the typical referent was chosen for 78% of the utterances with noun-focus prosody and 55% of the utterances with verb-focus prosody.

Taken together, the results of the three studies provide strong support for the adaptive nature of the mechanism employed for intonation interpretation. We have shown that listeners are sensitive to the distribution of tokens, modulating their mapping of prosodic contours onto categories just as listeners adapt phonetic categories based on new distributions. Moreover, studies 2 and 3 suggest that listeners seem to be constantly adjusting their interpretations based on their estimates of how likely it is for a particular linguistic signal, defined with lexical and prosodic information, to convey either of the possible speaker meanings (i.e., it is an X vs. it is not an X).

The offline judgment paradigms demonstrate that listeners adapt over the course of multiple utterances. However, these studies cannot tell us how the reliability information accumulated over time affects real-time online language comprehension. We hypothesize that online language comprehension actively employs a generative model available each point of time, and any discrepancy between a predicted pattern and an actual input signal would generate an error signal. It is this error signal that allows the listener to adapt to the statistics of the input. Thus, we predict that listeners will (a) generate expectations based on cues to prosodic contours and (b) adapt these expectations based on the statistics of the input. We now briefly describe some ongoing work that examines the hypothesis listeners generate expectations based on prosodic information as an utterance unfolds, and modulate their expectations based on the reliability of the speaker.

6 Study 4: Expectations in Real-Time Processing

Our studies take advantage of the fact that the intonation contour that most naturally maps onto the contrastive "but it isn't" interpretation consists of both an L+H* pitch accent on the verb "looks" and an L−H% boundary tone. We first established that listeners process an information contour predictively when the L+H* is reliably paired with the L–H% boundary tone. In other words, when there is a unique contrast pair in a visual scene, listeners would launch an eye movements to a less nameable referent as soon as they heard "(it) *LOOKS*…," indicating that they had generated a prediction about the likely referent. In order to do so, we designed a visual world study (Cooper 1974; Tanenhaus et al. 1995) using displays that contained either one or two contrast sets, as illustrated in Fig. 7.

The logic of the study was based on previous work using contrast and contrastive prosody. A line of visual world studies initiated by Sedivy and colleagues (e.g., Sedivy et al. 1999) established that when listeners hear a prenominal scalar adjective, e.g., "Pick up the tall glass," they immediately look at the taller member of a contrast set (e.g., the taller of two glasses) upon hearing "tall" when a display contains only a single contrast pair. However, if there are multiple contrast sets (two glasses, and two boxes), listeners do not begin to look at potential referents (e.g., the taller of the objects in the contrast sets) until they hear the noun (Hanna et al. 2003; Heller et al. 2008). As we discussed above, several visual world studies have established that listeners are sensitive to contrastive prosody (Ito and Speer 2007; Watson et al. 2008; Weber et al. 2006).

We reasoned that if upon hearing the fall–rise contour (L+H*) on "LOOKS," listeners incrementally develop a contrastive interpretation, then with a single contrast set, participants should make anticipatory eye movements to the less nameable referent (e.g., the okapi) about 200 ms or so after the onset of the contrastive pitch accent. We confirmed this prediction in a study conducted in collaboration with Sarah Bibyk and Daniel Pontillo (Kurumada et al. 2014a). As can be seen in Fig. 8, fixations to the nonprototypical target (e.g., an okapi) based on the verb-focus prosody increased even before the segmental information of the final noun became fully available. This result demonstrates that listeners processed the acoustic cues in the contour incrementally, generating predictions before they encountered either the beginning of the noun (e.g., zebra) or the boundary tone.

We then tested the hypotheses that listeners would down-weight the information provided by the fall–rise contour, if prior exposure established that a speaker used contrastive focus unreliably (Kurumada et al. 2014b). The experiment consisted of an exposure and a test phase. The test phase was identical to the experiment described above. In the exposure phase, the same speaker used contrastive prosody with prenominal adjectives either reliably or unreliably. We used prenominal adjectives because this allowed us to expose listeners to information about whether or not the speaker reliably used the L+H* accent to signal contrast, without giving them experience with the "It looks like an X" construction. In the prosody-reliable condition, the speaker provided instructions such as "Click on the blue circle. Now, click

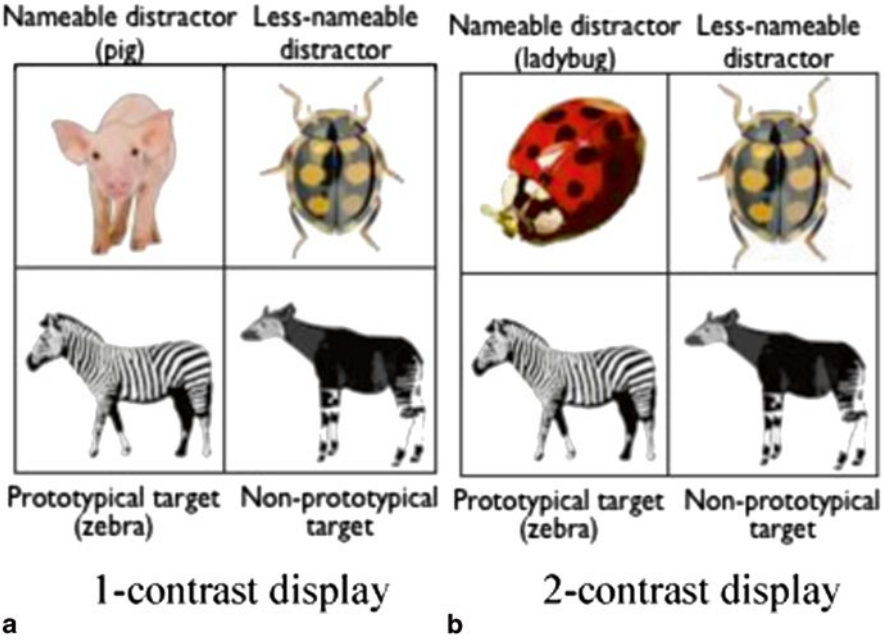


Fig. 7 Sample visual displays for the one-contrast trials (**a**) and the two-contrast trials (**b**)

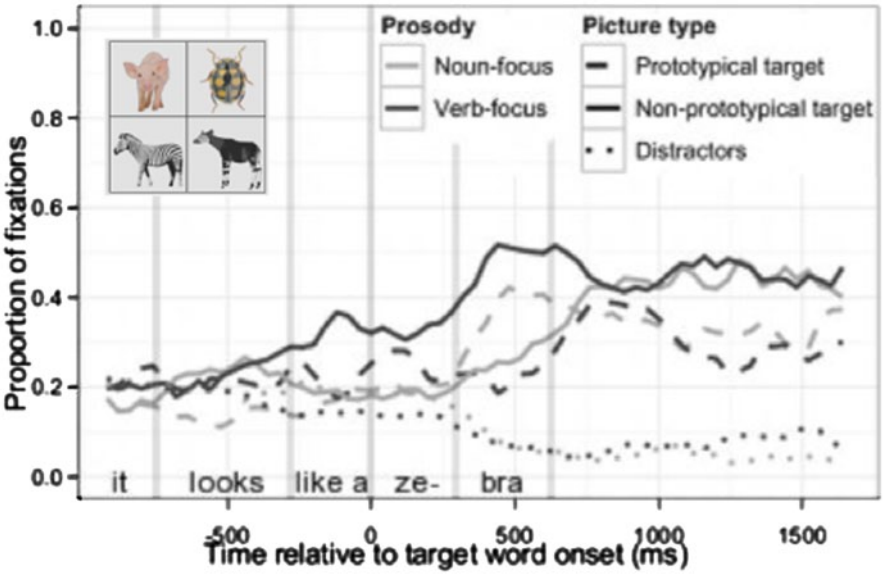


Fig. 8 Proportions of fixation to pictures in response to noun-focus (*gray lines*) and verb-focus prosody (*black lines*) in one-contrast displays. The *X*-axis indicates time with respect to the onset of the final noun

on the YELLOW L+H* circle," in which the L+H* accent highlighted a contextual contrast between two objects (Ito and Speer 2008). In the prosody-unreliable condition, the speaker used an L+H* accent on a wrong constituent, or did not use one when it would have been informative. After being exposed to 12 of such exposure items, participants responded to the same items from the original eye-tracking experiment. In the reliable condition, participants again made anticipatory eye movements when they heard LOOKS L+H*. However, in the unreliable condition, listeners did not make anticipatory eye movements as they heard LOOKS L+H* but rather waited until the disambiguating noun with the boundary tone. These results demonstrate that the adaptation effect present in offline judgments does indeed affect the time course of prosodic interpretation as an utterance unfolds. Listeners generate expectations incrementally based on reliable cues about a prosodic contour, but modify their expectations when a cue becomes unreliable. It is important to note that unlike the earlier offline experiments, the effects of reliability transferred from one construction in which contrast was signaled by an L+H* on an adjective to a construction in which the interpretation is conveyed by a prosodic contour with both contrastive focus on a verb and a subsequent boundary tone.

7 Summary, Conclusions, and Implications

Let us return to our original question. Hearing Herb utter "I'm hot" with a particular prosodic contour, how would Damon have arrived at the particular interpretation Herb intended to convey? In the four studies described in this chapter, we argued that interpretation of intonation makes use of various sources of information about the speaker, context, and the prosodic features experienced in past exposure. We suggested that the discourse context biases the listener's expectations for particular speaker meanings. In playing poker, one would expect the speaker to say something "relevant" to the situation at hand. With this expectation, the comprehension system compares possible speaker meanings to identify which speaker meaning would be most likely to generate the given utterance. This is done through comparison between alternative hypotheses (speaker meanings) as well as alternative linguistic elements, including lexical, syntactic, and prosodic information, which the speaker could apply to convey those possible speaker meanings. While sentence prosody generally plays an important role in this process, it is effectively discounted when recent experience indicates that it does not reliably predict the speaker's pragmatic intentions.

This approach is distinguished from the general view that one can posit a one-to-one mapping between given prosodic features of speech (e.g., pitch movement) and a particular pragmatic interpretation or function. Such an approach would not predict that listeners would process and interpret the same acoustic input differently depending on their expectations and recent experiences. We acknowledge that it was a deliberate decision for past researchers to begin their phonological analyses assuming categories abstracted away from their phonetic specifications. Many

researchers have, in fact, noted that each phonological category contains substantial phonetic variation (e.g., Ladd 2008). Our proposal, however, goes beyond claiming that listeners are normalizing phonetic variability across different instances of speech. Rather, we are arguing that prosodic interpretation is part of a principled inference process to arrive at speaker meanings based on noisy and variable data. In this process, each cue—contextual or linguistic—is weighted according to its reliability, which probabilistically shifts as an outcome of the inference. The prosodic features of speech alone, therefore, cannot reliably predict pragmatic interpretations of utterances, as has been assumed in previous research. Prosodic information can effectively support inferences only when it is interpreted against appropriate models with highly structured knowledge about discourse contexts and linguistic expressions.

As we mentioned earlier, variability is ubiquitous in speech, and the human language comprehension system needs to deal with it at all levels of linguistic representations. Recent studies have begun to address how listeners integrate recent experiences to adjust their inferences about intended messages. These attempts include investigations of speech perception (e.g., Norris et al. 2003), syntactic parsing (e.g., Fine and Jaeger 2013; Jaeger and Snider 2013), and semantic comprehension of quantifiers (e.g., Degen 2013; Yildirim et al. 2013). The evidence of prosodic adaptation outlined in this chapter demonstrates that listeners' sensitivity to the characteristics of the input extends to the mapping between prosodic profiles of speech and abstract pragmatic information. This points to the exciting possibility of a unified model for language comprehension, encompassing low-level speech perception through high-level intention recognition. At all levels, listeners infer an underlying representation that generated the observed input. Along with other recent studies in the same spirit, we argue that this inferential association between the signal and the representation is the key to robust language comprehension in the face of substantial variability in the input.

We acknowledge that the work we have presented is only a first step towards demonstrating the promise of this framework. As with any line of research examining adaptation, the first step is to determine whether adaptation does indeed occur. It then becomes important to determine the scope and generality of the effects, including when and how listeners generalize the learned knowledge across constructions, prosodic contours, and contexts. For example, when listeners observe that a speaker uses an L+H* accent unreliably to signal contrast, they can generalize the information in more than one way. The particular speaker might be unreliable with respect to all prosodic uses, or only to L+H* (e.g., a proficient nonnative adult speaker might fail in marking contrast in prosody, yet can be otherwise fully competent). The speaker could also be incapable of reliably recognizing a contextual contrast, but otherwise capable of producing expected prosodic patterns. Therefore, the rate and outcome of adaptation will likely be modulated by the listener's beliefs about the language, speaker, and context. Ultimately, it will be important to provide a principled account for how listeners generalize from their experience.

We also will need to provide more specific, testable, quantitative models of how listeners combine different types of cues, before concluding that they can be

accounted for naturally within a rational inference framework. Perhaps the most difficult challenge is to integrate research on how listeners combine acoustic and phonetic cues, which can be observed and measured, with models of how interlocutors generate expectations for what types of intended meanings are relevant to a particular context or class of contexts, and what utterance types are likely to convey those intentions. That said, we are willing to make a substantial bet that (a) this line of investigation is likely to prove promising and that (b) adaptation will play a crucial role in how we solve the class of variability and cue-integration problems that arise. We would even make a small wager that this approach might prove useful in understanding how readers generate and use implicit prosody.

References

Beckman, M. E., & Ayers, G. M. (1994). Guidelines for ToBI labeling. http://www.ling.ohio-state.edu/research/phonetics/E ToBI. Accessed 23 Sept 2008.

Beckman, M. E., & Hirschberg, J. (1994). The ToBI annotiation conventions. http://www.ling.ohio-state.edu/~tobi/ame tobi/annotation conventions.html. Accessed 23 Sept 2008.

Bibyk, S., Kurumada, C., & Tanenhaus, M. K. (n.d.). Context constraints on intonation interpretation (in preparation).

Boersma, P., & Weenink, D. (2008). Praat: Doing phonetics by computer (version 5.0.26) [computer program]. http://www.praat.org/. Accessed 16 June 2008.

Brown, M., Dilley, L. C., & Tanenhaus, M. K. (2015). Real-time expectations based on context speech rate can cause words to appear or disappear. In Proceedings of the 34th Annual Conference of the Cognitive Science Society (pp. 1374–1379).

Brown, M., Salverda, A. P., Gunlogson, C., & Tanenhaus, M. K. (2015). Interpreting prosodic cues in discourse context. *Language, Cognition and Neuroscience* , *30,* 149–166.

Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272.

Clark, H. H. (1992). *Arenas of language use*. Chicago: University of Chicago Press.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition, 108*(3), 804–809.

Connine, C. M., & Clifton, C., Jr. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human perception and performance, 13*(2), 291.

Cooper, R. M. (1974). Control of eye fixation by meaning of spoken language: New methodology for real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6,* 84–107.

Degen, J. (2013). Alternatives in pragmatic reasoning. PhD dissertation, University of Rochester.

Dell, G., & Chang, F. (2013). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B, 369*.

Dennison, H. Y., & Schafer, A. (2010). Online construction of implicature through contrastive prosody. Proceedings of 5th Speech Prosody Conference.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27,* 143–165.

Farmer, T., Brown, M., & Tanenhaus, M. (2013). Prediction, explanation, and the role of generative models in language processing. *Behavioral and Brain Sciences, 36,* 211–212.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development, 10*(3), 279–293.

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science, 37*(3), 578–591.

Fine, A. B., Jaeger, T. F., Farmer, T., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE, 8*(10), e77661. doi:10.1371/journal.pone.0077661.

Ganong, W. F. (1990). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics, 3,* 41–58.

Hanna, J., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43–61.

Hansen, M. B., & Markman, E. M. (2005). Appearance questions can be misleading: Adiscourse-based account of the appearance-reality problem. *Cognitive Psychology, 50*(3), 233–263.

Heller, D., Grodner, D., & Tanenhaus, M. K. (2008). The role of perspective in identifying domains of reference. *Cognition, 108*(3), 831–836.

Isaacs, A., & Watson, D. (2010). Accent detection is a slippery slope: Direction and rate of f0 change drives listeners comprehension. *Language Cognitive Processes, 25*(7), 1178–1200.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58,* 541–573.

Jackendoff, R. (1972). *Semantics in generative grammar*. Cambridge: MIT Press.

Jaeger, T. F., & Snider, N. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*(1), 57–83.

Kleinschmidt, D., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In ACL workshop on cognitive modeling and computational linguistics. Portland.

Kleinschmidt, D., & Jaeger, T. F. (2015). Robust speech perception: Recognizing the familiar, generalizing to the similar, and adapting to the novel. *Psychological Review, 122*(2), 148–203.

Kleinschmidt, D., Fine, A., & Jaeger, T. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. Proceedings of the 34th annual conference of the cognitive science society.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin and Review, 13*(2), 262–268.

Kurumada, C. (2013). Navigating variability in the linguistic signal: Learning to interpret contrastive prosody. PhD dissertation, Stanford University.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Prosody and pragmatic inference: It looks like speech adaptation. Proceedings of the 34th Annual Conference of the Cognitive Science Society.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2013). Incremental processing in the pragmatic interpretation of contrastive prosody. Proceedings of the 35th Annual Meeting of the Cognitive Science Society.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014a). Is it or isn't it: Listeners make rapid use of prosody to infer speaker meanings. *Cognition, 133,* 335–342.

Kurumada, C., Brown, M., Bibyk, S., Pontillo, D., & Tanenhaus, M. K. (2014b). Rapid adaptation in online pragmatic interpretation of contrastive prosody. Proceedings of the 36th Annual Meeting of the Cognitive Science Society.

Kurumada, C., Brown, M., & Tanenhaus, M. K. (n.d.). Probabilistic inferences in pragmatic interpretation of English contrastive prosody (submitted).

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.

McClelland, J. L., & Elman, J. L. (1986). The trace model of speech perception. *Cognitive Psychology, 18*(1), 1–86.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118*(2), 218–246.

Miller, J. L., Green, K., & Schermer, T. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics, 36,* 329–337.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9*(5–6), 453–467.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238.

Pierrehumbert, J. & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions and plans in communication and discourse* (pp. 271–311). Cambridge: MIT Press.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition,* 71*,* 109–147.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., et al. (1992). ToBI: A standard for labeling English prosody. In International conference on spoken language processing (Vol. 2, pp. 867–870). Banff.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268,* 1632–1634.

Ward, G., & Hirschberg, J. (1985). Implicating uncertainty: The pragmatics of fall-rise intonation. *Language, 61,* 747–776.

Watson, D., Gunlogson, C., & Tanenhaus, M. (2008). Interpreting pitch accents in on-line comprehension: H* vs L+H*. *Cognitive Science, 32,* 1232–1244.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech, 49,* 367–392.

Yildirim, I., Degen, J., Tanenhaus, M. K. & Jaeger, T. F. (2013). Linguistic variability and adaptation in quantifier meanings. In Proceedings of the Thirty-Fifth Annual Conference of the Cognitive Science Society.

Prosody, Performance, and Cognitive Skill: Evidence from Individual Differences

Fernanda Ferreira and Hossein Karimi

Abstract If a pause occurs in the middle of a sentence, is it attributable to prosodic structure, planning problems, or both? And if both prosodic representation and performance constraints conspire to cause a speaker to divide a sentence into two units, can the durational effects that result be parsed into those two different sources? In this chapter, we argue that prosody and performance are theoretically and empirically distinct, and that durational effects may arise from two distinct sources: from the implementation of a grammatical representation, and from performance limitations. A range of empirical evidence is presented to support this distinction. Studies investigating the effects of working memory, inhibitory control, and lexical difficulty indicate that individuals with less cognitive capacity are more likely to produce sentence-internal breaks, and these are not conditioned by characteristics of a prosodic representation. This finding suggests that performance units are not necessarily prosodic units, and that an adequate theory of sentence production must incorporate mechanisms for implementing prosodic structure as well as strategies for managing processing load during speech.

Keywords Language production · Prosody · Timing · Working memory · Inhibition

1 Introduction

When speakers pause in the middle of a sentence, is the pause attributable to the speaker's implementation of a prosodic representation, or do speakers pause for some performance reason—for example, to buy more time to plan the upcoming stretch of speech? Or, is the correct answer "both"? That is, speakers sometimes not only need time to plan or in some way manage their cognitive resources but they

F. Ferreira (✉) · H. Karimi
Department of Psychology, Institute for Mind and Brain,
University of South Carolina, Columbia, SC 29201, USA
e-mail: Fernanda@sc.edu

H. Karimi
e-mail: karimi@email.sc.edu

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_7

also use prosodic information to achieve their performance goals in a linguistically principled way. These are the questions we address in this chapter. Before we begin, however, we would like to highlight the extraordinary influence that Janet Fodor has had on the field of psycholinguistics, not just due to her work on prosody, the focus of the present volume, but also through her contributions to numerous other debates as well. Whether the subject is the online processing of phrase structure (Frazier and Fodor 1978; Fodor and Frazier 1980), the establishment of filler–gap relations (Fodor 1978), the reanalysis of garden-path sentences (Fodor and Inoue 1994, 1998, 2000; Fodor and Ferreira 1998), or the critical role of prosody in written and spoken language (Fodor 2002a, b), Janet Fodor's arguments have sharpened the issues and allowed psycholinguists from a variety of perspectives and theoretical orientations to design coherent and theoretically focused experiments and to draw conclusions that genuinely move the field forward. This is very clearly true when it comes to the role of prosody in language processing. Let us now turn to this topic.

It is probably fair to say that psycholinguists tend to be biased towards what we might term "naturalized prosody"—that is, they are predisposed to believe that prosodic effects arise, at least in part, due to factors related to performance in sentence planning. But in our own work (Ferreira 1991, 1993, 2007), we have adopted the strong position that prosody and performance effects must be distinguished in any psycholinguistic model. In that early work (Ferreira 1991, 1993), we found empirical evidence suggesting that the left and right contexts surrounding a word have markedly different effects on word and pause durations: The complexity of upcoming material influenced the likelihood of a pause but did not lead to word lengthening. In addition, pause durations patterned with sentence initiation times. In contrast, the prosodic complexity of the context to the left of a word affected that word's duration and what we characterized as grammar-based pauses: Pauses of a relatively short duration that tend to co-occur with phrase-final lengthening. We also argued that these pauses arise in part because a syllable reaches the limits of its "stretchability," and as a result, the speaker is unable to maintain a timing pattern with lengthening alone (Ferreira 1993; Selkirk 1984). We therefore concluded that acoustic effects associated with material to the left of a potential prosodic boundary are related to implementation of a metrical representation, whereas those associated with material to the right are attributable to planning and performance factors. This model which assumes prosodic effects from left context and planning effects from right context has been challenged based on new processing models that offer more sophisticated accounts of how performance constraints might lead to prosodic effects (Watson and Gibson 2004). Nonetheless, we have maintained that these newer algorithms and findings are not entirely persuasive, in part because the success of any algorithm depends critically on the choice of sentences used to evaluate it (Ferreira 2007), and most studies do not employ a design in which left and right contexts are systematically and orthogonally manipulated.

In contrast to performance, we view prosody as a linguistic system with its own grammar. The grammar has a metrical component, which causes an utterance to have a distinct and grammatical rhythm, and an intonational component, which is meant to capture changes in pitch across an utterance. Both are a function of

prosodic constituency, which is derived from rules that define the syntax–phonology interface. These prosodic domains in turn determine the application of rules linked to phrase-final lengthening and pausing, as well as the placement of pitch accents associated with different communicative intentions. Thus, a set of grammatical constraints defines prosodic structure and the rules that apply to those structures. One interesting aspect of prosody is that the application of rules is often graded, with optimality theory approaches (Prince and Smolensky 1993) well suited to capturing the idea that rules do not apply in an all-or-none manner, but instead apply with a certain probability depending on the precise balance of conflicting constraints. In addition, prosodic constituents are created from a syntactic structure, but the two forms of representation are not isomorphic. Both these points will become relevant when at the end of the chapter we consider the viability of a hybrid approach relating prosody and performance.

On the other hand, performance effects are often poorly behaved with respect to any semantic or syntactic constraints that might govern the production of a sentence. To take a clear example, if a person pauses after a sentence initial *the,* which is fairly common in spontaneous speech (Boomer 1965; Maclay and Osgood 1959), that pause has no obvious grammatical motivation. Indeed, the pause would usually be treated as a disfluency, and a speaker aiming to speak fluently would avoid it. This is not to say that the disfluency is random and conveys no information to the listener. A fair bit of research has shown that listeners in fact can use disfluencies as information concerning what might be coming up next, and these predictions are based on listeners' knowledge of typical co-occurrences between, for example, difficult concepts and the need to pause to allow time for lexical retrieval (Arnold et al. 2007). Nonetheless, few would think of such a pause as prosodically conditioned.

This case seems clear-cut, but the picture gets a little more complicated when we consider pauses in other sentence locations, especially near the middle of an utterance. Consider this example: *Mary ordered salad because < pause > she is trying to eat more healthily*. The pause after *because* is not in the syntactically most prominent location; *because* is part of the second clause, and therefore the pause after it separates *because* from the syntactic constituent of which it is a part. Based on this criterion, it might be tempting to view a pause in this location as also non-prosodic, like the one after a sentence-initial *the*. On the other hand, if we make reference to constraints on prosodic rather than syntactic constituency, then that pause location is perfectly fine. As was argued decades ago for cases such as *This is the cat that chased the rat that swallowed the cheese…* (Chomsky and Halle 1968), the rules of phonology seem to have the effect of simplifying and flattening a syntactic structure. And as argued more recently by Selkirk (1984), a particular intonational phrasing is acceptable as long as the resulting phrases obey the sense unit condition, which states that the constituents inside an intonational phrase must be in a head–modifier or head–argument relation. In addition, the phrasing that groups *because* with the first clause has the additional virtue of dividing the sentence into two parts of roughly equal size—two balanced sisters, to use Fodor's terminology (Fodor 1998, 2002a, b). Thus, in this sort of example, it is more difficult to tell whether the pause is prosodic, performance based, or both. Careful experiments are required

to distinguish the two potential sources of lengthening and pausing—the grammar versus disfluency.

In this chapter, we approach the distinction between prosody and performance in a somewhat novel way: We will motivate the distinction by reviewing evidence from new research that links performance effects to cognitive skill. These studies use an individual differences approach to assess whether people with lower working memory (WM) capacity, weaker inhibitory control, or lower intelligence quotients (IQs) are also more likely to need a break point within a sentence compared to those with more robust cognitive systems. The logic of the approach is to assume that there is no principled reason to expect that prosodic effects will be influenced by cognitive skill; the grammar is the grammar whether a person has high or low WM capacity. Of course, in cases in which the grammar presents the language system with a choice between more than one linguistic structure, cognitive factors will play an important role in making the linguistic choice. The grammar presents options, and the cognitive system selects from among them on the basis of a range of factors, including performance constraints. In contrast, performance is clearly affected by cognitive skill. For example, a person who has a shorter WM span would seem to be more likely to break up a sentence into smaller performance units than would someone with a longer span. We turn to these studies next.

2 Working Memory and Implicit Prosody

One of the most influential and important ideas to emerge from psycholinguistics in the past decade or so is the notion that prosody is not confined to spoken language: Readers also generate a prosodic representation for written sentences. This proposal is compatible with decades of research in cognitive psychology showing that, fundamentally, reading is the translation of visual symbols into a phonological code (Berent and Perfetti 1995). Visually presented words activate their phonological forms, as demonstrated by phenomena such as tongue-twister effects in reading (McCutchen and Perfetti 1982) as well as interference from homophones (Van Orden 1987). For instance, using a semantic categorization task, Van Orden demonstrated that homophones associated with a target significantly increased false positive categorization rates. He observed that the word *rows* was sometimes miscategorized as a kind of flower, indicating not only that the phonological representation of words are activated during reading but also that this representation might mediate access to the word's semantic representation. Reading also seems to involve an "inner voice" that generates an ongoing phonological representation of text, and which even has speech characteristics such as gender (Quinn et al. 2000; Slowiaczek and Clifton 1980; Stolterfoht et al. 2007). What Fodor added to our theoretical understanding of phonological processing during reading is critical for psycholinguistics: She explicitly argued that the sounds we hear as we read include prosodic information, and are governed by a principled representation of prosodic structure. Many of these arguments were based on studies of garden-path

reanalysis, most of which were conducted with visual materials, and many of which suggested that revision of an incorrect syntactic structure is more difficult when the new analysis requires the generation of a different prosodic form for the sentence. In one seminal study using self-paced reading, Bader (1998) used focus operators to manipulate the prosodic structure of local garden-path ambiguities and showed that prosodic structure can influence recovery from a misanalysis independent of syntactic structure, suggesting that reanalysis is prosodically constrained, and more importantly for our purposes, providing evidence for implicit prosody in reading.

Armed with this theoretical innovation, we conducted a large-scale individual differences study designed to investigate the relationship between WM capacity and attachment decisions (Swets et al. 2007). Our research strategy was to identify a long sentence type that would likely need to be spoken as more than one production unit. For this purpose we chose the relative clause attachment structure illustrated in *The maid of the princess who scratched herself in public was terribly embarrassed*. This sentence seems to allow for two possible break points: one after *public,* at the subject–verb phrase boundary, and the other after *princess,* before the relative clause. These options are rank ordered, of course: The location between the subject and verb phrase is the one that is structurally most preferred, and the location before the relative clause might also be exploited if an individual has such limited processing capacity that he/she must divide the sentence into more than two performance units. In addition, as has been widely discussed, the sentence is globally ambiguous because the relative clause can attach either high, to the first noun (N1, *maid*), or low, to the second and more recent noun (N2, *princess*). The preference for N1 or N2 attachment seems to vary crosslinguistically: Dutch has about a 60–40 % bias for N1 attachments (Desmet et al. 2002), whereas English has a 40–60 % bias for N2 attachments (Cuetos and Mitchell 1988). These crosslinguistic differences have been explained by appealing to implicit prosody: Whereas speakers of Dutch tend to put a prosodic break between the complex noun phrase (NP) and the relative clause in sentences like these, English speakers tend to leave out this break and prefer to place a break after the relative clause instead of before it. If a speaker does insert a prosodic break before the relative clause, as Dutch speakers tend to do, the result is a bias towards higher attachment decisions for spoken sentences (Carlson et al. 2001). The prosodic break is assumed to induce N1 preferences because it can be interpreted as a "structural discontinuity in the syntactic tree" (Fodor 2002a, p. 4). This interpretation supports the formation of a tree in which the entire NP is modified by the relative clause rather than just N2, resulting in a high-attachment (N1) preference.

Speculation has also centered around whether the preference for N1 versus N2 attachment might be related to WM capacity. The intuitive idea is that recency favors N2 attachment, and those with smaller working memories might be more biased to use a recency strategy to make attachment decisions, as reported by Felser et al. (2003) for 6–7-year-olds. This possibility has even occasionally been invoked to explain the crosslinguistic differences in attachment preference mentioned above, with the argument going something like this: There is a tendency for the N1 preference to be found in experiments which include participants attending European

universities, and for the N2 preference to emerge in experiments with students from American universities, especially large public institutions. If we assume that selectivity is correlated with WM capacity (and there is evidence that WM and IQ are positively correlated), then perhaps what appears to be a crosslinguistic difference is actually a confound caused by testing participant groups with different individual difference characteristics. This would be an unfortunate interpretation, but fortunately, the results of Swets et al. (2007) allow us to rule it out, as we will see shortly.

The study was unusual (perhaps unique at the time) for adopting a psychometric approach to these psycholinguistic questions concerning attachment preference and implicit prosody. A psychometric approach attempts to establish relationships among variables that occur naturally and that naturally vary (i.e., WM capacity), in contrast with variables that can be experimentally manipulated. The statistical method is then to test for correlations using sophisticated quantitative techniques such as structural equation modeling. An important requirement of such work is that sample sizes be adequate to ensure there is sufficient power to conduct continuous analyses because continuous analyses allow researchers to evaluate the relationship between two variables across the full range of scores and allow them to avoid the problems inherent in the use of so-called extreme-groups designs (i.e., the testing of only the subjects with the highest and the lowest WM scores, so that WM is treated as a categorical variable in statistical analyses). To that end, 150 Michigan State University undergraduates, all native speakers of English, were tested along with 96 undergraduates from Ghent University, all of whom were native speakers of Dutch. Each person's WM capacity was assessed using a reading span task and a separate spatial span task. Then participants were shown sentences individually, and after each sentence, the participants answered a question such as *Who scratched herself in public,* with the options represented by N1 and N2 attachments shown one above the other.

The critical manipulation in this study was conducted between experiments as well as between subjects. In Experiment 1, each sentence was presented on a single line, so that nothing about the visual presentation encouraged the inclusion of a prosodic break within the sentence. With this setup, Swets et al. (2007) replicated previous work showing that Dutch participants prefer to attach to N1 and English participants to N2; however, although the effect was statistically significant, it was quite a bit smaller than in previous studies, amounting to no more than a 3–4% difference in attachment decisions. Much more surprising was the effect of WM: Contrary to the recency principle, we observed that the smaller a participant's WM capacity, the more likely he or she was to prefer N1 attachments. Moreover, this effect of WM was statistically identical for Dutch and English participants, suggesting that it was entirely independent of any crosslinguistic factors. Moreover, if the participants are divided into two equal *n* groups based on their WM capacities, the pattern that emerges is that the participants with the lowest spans preferred N1 attachments whether they were English or Dutch, and the participants with the highest spans preferred N2 attachments, again regardless of what language they spoke.

How do we explain this counterintuitive result? Our account made a critical appeal to the notion of prosodic chunking in silent reading. Imagine that high-span

subjects can "chunk" more information together while reading. These higher-span individuals are able to treat the entire subject of the sentence as a single "processing unit". Low-span readers, in contrast, may have to break up the subject because of its length. A likely boundary for such a break point is right before the relative clause. And these breaks in turn will encourage N1 attachments, for the reasons described earlier. This hypothesis that chunking strategies underlie the individual differences observed in our first experiment was tested in the second experiment. This time, the sentences were presented in three successive displays. The first included the words before the relative clause (*The maid of the princess*), the second consisted of the entire relative clause (*who scratched herself in public*), and the third consisted of the entire verb phrase (*was terribly embarrassed*). Our prediction was that this segmented presentation method would induce readers to prosodically chunk the sentences into three units, including one that separated both potential attachment sites from the relative clause. As a result, all participants would be turned into low-capacity readers; based on the presentation format, all participants would generate an implicit prosodic phrasing that isolated the relative clause, and based on the principles mentioned earlier, this would lead to an overall preference for N1 attachments.

These predictions were clearly confirmed. Although we once again replicated the slight preference for N1 attachments in Dutch and for N2 attachments in English, we no longer observed a significant effect of WM capacity in either group. Not only did everyone regardless of WM capacity prefer N1 attachments but also the overall N1 preference was much larger than has been reported in any previous work: 71 % for English speakers and 75 % for Dutch speakers (the two groups did not differ significantly). Thus, if we manipulate presentation format so that all participants are induced to read the way we hypothesize low-capacity readers do, we dramatically enhance the N1 attachment preference.

Two important conclusions can be drawn from this study. First, we have discovered some of the strongest evidence we know of for the reality of implicit prosody in reading. Moreover, in pilot work we are currently conducting in our laboratory, we are measuring WM capacity once again, but this time asking participants to say the sentences out loud. Participants are being asked to read and learn the sentences, and then to repeat them from memory upon receipt of a cue. We additionally varied whether the verb in the relative clause was high or low frequency (e.g., *glorified* vs. *idolized*), because we predicted that greater lexical difficulty would increase the chances of a performance break, particularly before that relative clause. Our preliminary data suggest that people with lower spans are more likely to require two break points within these same sentences, and that they are also more affected by the frequency manipulation. It thus appears that our findings concerning implicit and explicit prosody dovetail nicely: Regardless of whether people speak out loud or read silently, it seems that those with smaller WM spans are more likely to divide a sentence up into multiple performance units. This is our first important conclusion. Second, given this relationship between WM capacity, which is a cognitive ability factor, and the tendency to break up a sentence, we think it makes a great deal of sense to think of these units not as prosodic constituents but as performance

units. We base this argument on the idea that prosodic constituency has no obvious connection to cognitive capacity; there is no theoretical reason for believing that WM span is in any way related to the way the grammar of prosody is applied or implemented. In contrast, there are very compelling theoretical reasons for linking WM and performance; indeed, in multiple domains it has been observed that those with larger spans chunk information more effectively and are able to pack more information into a single chunk (Ottem et al. 2007).

In short, the chunks formed during silent reading are affected by WM capacity, as would be expected if performance units reflect cognitive skill. This in turn motivates a separation between prosodic and performance-based effects in language processing.

3 Inhibitory Control and Planning in Production

Next, we turn to research we have conducted investigating the relationship between the integrity of inhibitory systems and speakers' fluency. Broadly speaking, inhibition as a cognitive skill can be defined as the suppression of inappropriate responses or intervening memories when the context changes (Aron et al. 2004). In other words, cognitive inhibition is a mechanism whereby prepotent behavioral responses are constrained when the expression of such responses is inappropriate or incorrect (Burle et al. 2004). A powerful method for investigating inhibition is again to use an individual differences approach—in this case, to compare performance in individuals suffering from attention deficit hyperactivity disorder (ADHD) to the performance of demographically matched controls (people of approximately the same age and social/educational background). A large number of studies suggest that people with ADHD have impaired inhibitory systems, leading to problems in tasks such as the anti-saccade and Stroop task, both of which require participants to squelch a prepotent response. For example, in the anti-saccade task, participants are instructed to look away from a visual stimulus such as a cross or dot as soon as it appears on the screen (Hallet 1978; Nieuwenhuis et al. 2001). Because such an abrupt onset of visual stimulus is known to automatically capture attention and eye movements (Theeuwes et al. 1998), efficient anti-saccade performance requires inhibition of the reflexive eye movement towards the abrupt-onset stimulus (Nieuwenhuis et al. 2001). Our work with this population has also shown that ADHD is characterized by more focused inhibitory deficits related specifically to language planning. In one study (Engelhardt et al. 2010), we asked individuals with ADHD as well as matched control subjects to generate a sentence from two objects (one animate, one inanimate), together with a printed verb. The verb either was ambiguous between simple past and past participle (*moved*) or was unambiguously the past participle (*ridden*) form, and presentation of the animate object (e.g., the girl) either preceded or followed the presentation of the inanimate object (e.g., the bicycle). Thus, in the past participle condition where the animate entity was presented first, the participants could start uttering the sentence "the girl…" and then realize at this

point that the sentence needs to be in passive form ("The bicycle was ridden by the girl."). As predicted, we observed that both groups of subjects were less fluent producing sentences with the participle verb, particularly when an animate object was shown before the inanimate object. This is because the participle nearly forces the generation of a passive form (the past perfect is a legitimate alternative, but our participants seemed to be unaware of this), which in turn forces the inanimate entity to serve as the sentential subject. In addition, as would be expected, given that people with ADHD tend to have problems with inhibitory control, this effect was larger for those with ADHD. The effect was particularly pronounced for self-repairs, suggesting that problems with inhibition lead individuals with ADHD to begin speaking before they have planned out the entire utterance and know it will be grammatical and semantically appropriate. Post hoc analyses also revealed that lower IQ scores were associated with more disfluencies overall, perhaps because one component of the IQ is vocabulary knowledge, which presumably relates to the ease of retrieving information from the lexicon.

In follow-up research, Engelhardt et al. (2011) asked healthy subjects and matched individuals with ADHD to describe networks of colored circles so that another person could draw the networks based on those descriptions. The resulting utterances had this character: *First there is an orange dot, and above it is a red dot. To the left of the red dot is a green dot and a blue dot,* etc. Successful description of these networks required some planning because the networks contained branches and choice points, and therefore speakers had to decide which branch of the network to describe next, and they had to make sure they remembered the choice circle so they could return to it to describe its other branches. In contrast to the previous study, this one taps into sentence planning at a level higher than grammatical encoding. Based on our other work, however, we expected to find that people with ADHD generated less fluent descriptions, and this is what we reported in the study: People with ADHD paused more often and generated more self-repairs than did normal controls. These differences were observed even though the two groups were matched on age, IQ, years of education, and even reading ability.

Thus, it appears that weaker inhibitory systems are associated with more errors and pauses in language production. We will make the same argument concerning inhibition that we made earlier with respect to WM: There is no theory of prosody from which predictions concerning effects of inhibition fall out naturally. Again, prosody is part of the grammar, and the grammar does not appeal to factors relating to cognitive skill. In contrast, there are compelling reasons for thinking that cognitive skill—in this case, inhibitory control—would be associated with performance and the need to pause or break during language production. This leads us to conclude that prosody and performance are distinct phenomena: Prosody is about the grammar, whereas performance is influenced by individual difference characteristics relating to cognitive skill.

Finally, in a recent study of individual differences among 106 normal participants, we used structural equation modeling to assess the relationships between various cognitive skills and the tendency to be disfluent during production (Engelhardt et al. 2013). This study included a range of measures of both intelligence (e.g.,

processing speed, vocabulary) and executive control (e.g., a stop-signal reaction time task and a Stroop task). We observed no significant effects related to IQ once correlated relationships with executive control were statistically removed, but we found a moderate effect of executive functioning, suggesting that those with poorer executive control and, in particular, those with weaker inhibitory systems tended to be less fluent. Thus, it is not only in clinical populations that we find a relationship between cognitive skill and fluency but we also see that within a large group of normal speakers, those with less intact cognitive systems are more likely to have performance problems during production.

In summary, then, factors that are not naturally thought of as part of the grammar seem to have a strong effect on language performance. We have seen that both smaller WM capacity and weaker inhibitory control systems cause speakers to produce more pauses and breaks when they speak. From these data, we argue that prosody and performance are distinct phenomena, and therefore no adequate theory of language production or of prosody in psycholinguistics can conflate them—to do so would be to blur important distinctions among representational types and processing mechanisms.

4 Bringing Prosody and Performance Together

Having laid out our arguments for distinguishing prosody and performance, we now want to consider how we can think about the interactions between the two, and the way both affect the auditory characteristics of a sentence. As we argued previously (Ferreira 1993, 2007), if we measure a variable such as pause duration, any effects are likely to be a mixture of both planning and rhythm—some of the pause time is attributable to the need to plan upcoming material, and some of it is attributable to the implementation of a prosodic representation and the need to insert pauses in order to maintain a specified rhythm.

Planning-based pauses are typically longer than prosodic pauses, and also tend to correlate with other planning-based variables such as sentence initiation time. In addition, these pauses will tend to get shorter and eventually disappear as a speaker becomes more practiced and fluent with a particular utterance. In contrast, rhythmic pauses are shorter, correlated with other prosodic effects such as phrase-final lengthening, and, by hypothesis, cannot be deleted without harming the prosodic well-formedness of the utterance. One way to think about the distinction is with an analogy to music: When a musician plays a piece of music, she will insert pauses at particular locations as she attempts to implement the musical score, and of course rests in specific places and of specific durations are as integral to any musical piece as the notes are. But if she struggles a bit with a certain sequence of notes and pauses before trying to execute them in order to plan the movements, that pause is a performance-based pause and ultimately needs to be smoothed out if the musician wants to give a performance that will be viewed as competent and aesthetically pleasing. Rests, then, remain in the performance, but silences due to cognitive

challenges are essentially errors and need to be eliminated if the musician wants to be viewed as a skilled musician. The same idea applies to language production: Rhythmic pauses must be maintained in an utterance, but planning-based pauses are performance effects and should not be included in the most fluent renditions of the utterance.

At the same time, spontaneous speech will almost invariably be a mixture of both planning and rhythmic effects. Where we want to end this chapter is with a theoretical speculation: Speakers do indeed sometimes need time to plan or in some way manage their cognitive resources when they produce spontaneous utterances, but they also have the ability to use a prosodic representation to achieve their performance goals in a linguistically principled way. As mentioned previously in our description of the prosodic system, the grammar defines a set of prosodic constituents and rules that apply to the resulting representation. One of the aspects of prosody that makes it attractive to psycholinguists is that both constituent structure and rule application tend to be graded, as are almost all phenomena related to human cognition. For example, the division of a sentence into intonational phrases involves both obligatory and optional constraints. The border between a subordinate and a main clause in a sentence must be marked by an intonational phrase boundary, but when it comes to the division between subject and verb phrase, the speaker has the option to place an intonational phrase boundary there or not. Most often we think of the decision to place the boundary as pragmatically conditioned; speakers use intonational phrase boundaries to convey their communicative intentions, including features such as focus, backgrounding, and mood. But the decision may also sometimes be driven by performance considerations: If a sentence is long and the speaker needs to divide it up to say it easily, then he/she might exercise the option of placing an intonational phrase break at the subject–verb phrase boundary. This break would enable him to recover from executing the subject and would also provide time for planning of the rest of the sentence, while at the same time perhaps conveying information to the listener related to the difficulty of the utterance. Thus, the grammar would be available to define an ideal break point from the perspective of prosodic constituency, and performance factors would help to determine whether the option was actually taken.

In addition, because of the nature of the interface between syntax and prosody, the two representational forms are not necessarily isomorphic. One important difference is that prosody (specifically, the sense unit condition) may allow the subject and verb to occur as part of one prosodic constituent and the postverbal constituents to make up another. An example might be a sentence such as *The noisy students left/after we ran out of beer,* which could naturally be spoken in such a way that the prepositional phrase constitutes its own prosodic phrase. This freedom to deviate from syntactic constituency means that the prosodic system presents the speaker with another tool for managing cognitive load during production: If the subject is relatively short and postverbal material is long, the speaker can create an utterance with two balanced sisters by exercising the option to break after the verb instead of before it. This would result in a more prosodically appealing rendition of a sentence, because sisters that are mismatched in length sound a bit odd, and it would also

permit a more even distribution of information over a sentence, an idea consistent with the so-called uniform information density (UID) hypothesis (Jaeger 2010), which assumes that speakers try to avoid major peaks and troughs of information in their utterances, and instead attempt to distribute information more evenly.

Yet another situation that may arise and that is more complex than the others is one that highlights the potential dependencies among different break locations. Let us consider again the relative clause attachment sentences that we focused on earlier; e.g., *The maid of the princess ^ who scratched herself in public ^ was terribly embarrassed,* with the ^ symbol indicating the two potential sites for a prosodic boundary. As we saw previously, a speaker with more limited WM resources might not be able to handle the entire subject as one prosodic unit, and might therefore place a break before the relative clause. But notice that if a speaker chooses this particular site for a prosodic boundary, he has also committed himself to placing a boundary at the subject–verb phrase location as well. This is because (*The maid of the princess, who scratched herself in public was embarrassed*) is not a well-formed prosodic phrasing; it violates rules of prosodic constituency, perhaps creating a "prosodic monster" (Féry, this volume). Thus, the speaker might choose the earlier boundary for reasons related to constraints on cognitive processing, but he/she might then choose the later boundary to maintain prosodic integrity. The first break site would thus be planning based, and the other would be prosodically motivated and even forced. Perhaps these two sources for the two breaks would cause the boundaries to have different properties relating to pitch and other prosodic features, although any differences would be hard to distinguish from those associated with the break locations within the prosodic constituency. Thus, we can argue that the role of the grammar is to create a prosodic representation that gives the cognitive system options when it needs to select.

References

Arnold, J. E., Hudson Kam, C. L., & Tanenhaus, M. K. (2007). If you say thee uh—you're describing something hard: The on-line attribution of disfluency during reference comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 914–930.

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8,* 170–177.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor, & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 1–46). Dordrecht: Kluwer Academic.

Berent, I., & Perfetti, C. A. (1995). A rose is a reez: The two cycles model of phonology assembly in reading English. *Psychological Review, 102,* 146–184.

Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech, 8,* 148–158.

Burle, B., Vidal, F., Tandonnet, C., & Hasbroucq, T. (2004). Physiological evidence for response inhibition in choice reaction time tasks. *Brain and Cognition, 56,* 153–164.

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language, 45,* 58–81.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Cuetos, F., & Mitchell, D. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30,* 73–105.

Desmet, T., Brysbaert, M., & Da Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology: Section A, 55*(3), 879–896.

Engelhardt, P. E., Corley, M., Nigg, J. T., & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory & Cognition, 38*(5), 617–628.

Engelhardt, P. E., Ferreira, F., Nigg, J. T. (2011). Language production strategies and disfluencies in multi-clause network descriptions: A study of adult Attention/hyperactivity disorder. *Neuropsychology, 25*(4), 442–453.

Engelhardt, P. E., Nigg, J. T., & Ferreira, F. (2013). Is the disfluency of language outputs related to individual differences in intelligence and executive function? *Acta Psychologica, 144,* 424–432.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11,* 127–163.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language, 30,* 210–233.

Ferreira, F. (1993). The creation of prosody during sentence processing. *Psychological Review, 100,* 233–253.

Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes, 22*(8), 1151–1177.

Fodor, J. D. (1978). Parsing strategies and constraints on transformations. *Linguistic Inquiry, 9,* 427–473.

Fodor, J. D. (1998). Learning to parse?. *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D. (2002a). Prosodic disambiguation in silent reading. In: M. Hirotani (Ed.), *NELS 32* (pp. 113–132). Amherst: GLSA Publications.

Fodor, J. D. (2002b). Psycholinguistics cannot escape prosody. *In Proceedings of the SPEECH PROSODY 2002 Conference*, Aix-en-Provence, France, April 2002.

Fodor, J. D., & Ferreira, F. (Eds.). (1998). *Reanalysis in sentence processing*. Dordrecht: Kluwer Academic Publishers.

Fodor, J. D., & Frazier, L. (1980). Is the human sentence processing mechanism an ATN?. *Cognition, 8,* 417–459.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23*(4)*,* 405–432.

Fodor, J. D., & Inoue, A. (1998). Attach Anyway. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing*. Dordrecht: Kluwer.

Fodor, J. D., & Inoue, A. (2000). Garden path re-analysis: Attach (anyway) and revision as last resort? In M. D. Vincenzi & V. Lombardo (Eds.), *Cross-linguistic perspective on language processing*. Dordrecht: Kluwer.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

Hallet, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18,* 1279–1296.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61,* 23–62.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15,* 19–44.

McCutchen, D., & Perfetti, C. A. (1982). Coherence and connectedness in the development of discourse production*, Text, 2,* 113–119.

Nieuwenhuis, S., Ridderinkhof, R. K., Blom, J., Band, G. P. H., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38,* 752–760.

Ottem, E. J., Lian, A., & Karlsen, P. J. (2007). Reasons for the growth of traditional memory span across age. *European Journal of Cognitive Psychology, 19,* 233–270.

Prince, A., & Smolensky, P. (1993). *Optimality Theory: Constraint Interaction in Generative Grammar*, ms., Rutgers University, New Brunswick, and University of Colorado, Boulder.

Quinn, D., Abdelghany, H., & Fodor, J. D. (2000). *More evidence of implicit prosody in reading: French and Arabic relative clauses*. Poster presented at the 13th Annual CUNY Conference on Human Sentence Processing, La Jolla, March 30–April 1.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.

Slowiaczek M. L., & Clifton C. (1980). Subvocalization and reading for meaning. *Journal of verbal learning and verbal behavior, 19*(5), 573–582.

Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during silent reading: Differential ERP effects. *Cognition, 104*(3), 565–590.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136,* 64–81.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science, 9,* 379–385.

Van Orden GC. (1987). A rows is a rose: Spelling, sound, and reading. *Memory & Cognition, 15,* 181–198.

Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes, 19,* 713–755.

Processing, Prosody, and Optional *to*

Thomas Wasow, Roger Levy, Robin Melnick, Hanzhi Zhu and Tom Juzek

Abstract The infinitival marker *to* is optional in many instances of the *do-be* construction, exemplified by sentences like *All I want to do is (to) go to work* However, it has not previously been investigated what factors govern speakers' choices in *to* use and omission. Here, we analyze nearly 10,000 such examples from the Corpus of Contemporary American English (COCA), using mixed-effects logistic regression to determine the respective contributions of a range of factors including phrasal complexity, wordform frequency and predictability, and prosody in predicting *to* use. We found that *to* use rate increases as phrasal complexity increases and as wordform frequency and predictability decrease, consistent with established psycholinguistic theory and data on the use of other optional function words. We also find the first quantitative corpus-based evidence for a role of prosody in governing optional function-word use: *to* is used more frequently when both the immediately preceding and the immediately following syllables carry some stress. This suggests that speakers use the intervening unstressed *to* to prevent stress clash. This result holds in writing as well as in speech, lending support to Janet Fodor's proposal that implicit prosody plays a role in sentence processing.

T. Wasow (✉) · R. Melnick
Department of Linguistics, Stanford University, Stanford, CA, USA
e-mail: wasow@stanford.edu

R. Melnick
e-mail: rmelnick@stanford.edu

R. Levy
Department of Linguistics, University of California, San Diego, CA, USA
e-mail: rlevy@ucsd.edu

H. Zhu
Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA, USA
e-mail: hanzhi@mit.edu

T. Juzek
Faculty of Linguistics, Philology, and Phonetics, University of Oxford, Oxford, UK
e-mail: tom.juzek@googlemain.com

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_8

Keywords Do-be construction · Optional function words · Prosody · Frequency · Predictability · Phrasal complexity · Uniform information density · Corpus analysis · Mixed-effects models · Infinitive · Implicit prosody · Information theory · Length · Stress clash

1 Introduction

Flickinger and Wasow (2013) discuss a previously understudied phenomenon in English that they call the "do-be construction" (DBC). This widely used construction is characterized by a remarkably rich and interconnected set of constraints. Before enumerating them, we present a few examples:[1]

(1) a. the thing that I tried to do was to keep the score close
b. the least we should do is make it as much fun as possible
c. what the CBO does is takes Congress's promises at face value
d. what we have done is taken military action in Bosnia through NATO
e. all he's been doing is going over legal papers

Flickinger and Wasow identify the following as the characteristic properties of DBC:

(2) a. The top verb in the construction is a specificational copula—that is, a form of *be* stipulating identity between the denotations of its subject and its complement.
b. The subject contains a relative clause headed by one of the following seven words: *what, thing, all, best, worst, most, or least.*
c. A form of the word *do* occurs within the relative clause.
d. The complement of the copula is a verb phrase (VP).
e. The understood subject of the post-copula VP (PCVP) is the same as the understood subject of *do*.
f. The form of the post-copula verb (PCV) depends on the form of *do* in the subject.

There are many questions one might ask about this construction, including how to analyze it within a particular theory of grammar (Flickinger and Wasow do this for Head-driven Phrase Structure Grammar), what its discourse function is, how it relates to other constructions (e.g., pseudoclefts), how it differs across dialects and registers of English, and what its history is. We will not address any of these here; rather, we are concerned with what conditions the presence or absence of the infinitival *to* at the beginning of the post-copula verb phrase (PCVP).

As noted in (2f), the form of the post-copula verb (PCV) is constrained. Specifically, there are three possible inflectional forms for the PCV: the same form as *do*

[1] Except where otherwise noted, examples in this chapter are from the Corpus of Contemporary American English (http://corpus.byu.edu/coca/), or COCA for short. We have truncated many of the examples, keeping only what is needed to make our point. Hence, most of our examples are presented without initial capitalization or sentence-final punctuation. Invented examples begin with capital letters and end with periods.

in the subject, base (that is, uninflected), or infinitival (that is, *to* followed by an uninflected verb).[2]

This is illustrated in the contrasts between (1) and (3):

(3) a. The thing that I tried to do was keep/*keeps/*kept/*keeping the score close.
b. The least we should do is to make/*makes/*made/*making it as much fun as possible.
c. What the CBO does is take/to take/*taken/*taking Congress's promises at face value
d. What we have done is take/to take/*took/*taking military action in Bosnia through NATO.
e. All he's been doing is?go/?to go/*went/*gone over legal papers.

Thus, whenever the PCV can be in base form (without *to*), it could just as well be in infinitival form (with *to*), and vice versa. To see the apparent interchangeability of these forms, consider the examples in (4), all of which were taken from Corpus of Contemporary American English (COCA), but only half of which had *to* in the original:

(4) a. what we're here on earth to do is (to) celebrate humanity
b. what I would do is (to) call upon the press to police yourselves
c. the other thing that it'll do is (to) facilitate getting Chinese troops into Tibet as well
d. the most important thing that Bretton Woods did was (to) create two institutions for international cooperation on monetary international problems
e. all they can do is (to) circumvent themselves
f. all I want to do is (to) go to work

Audiences to whom we have presented these examples do not have clear intuitions about which examples had *to* in the original.[3]

This raises the question of what factors lead people to use *to* in the DBC when they do. The bulk of this chapter describes a study aimed at answering this question and discussing why the answer is of theoretical interest. Of particular note in the context of this volume is the fact that one important factor influencing *to* use in the DBC is prosodic, and that the influence of prosody is evident in writing as well as in speech.

2 Data Extraction and Annotation

We conducted a corpus study using COCA, a 450-million word web-based collection, roughly equally divided among speech (radio and television interviews), newspapers, magazines, fiction, and academic writing, dating from 1990 to 2012.[4]

[2] Flickinger and Wasow claim that if the form of *do* is a present participle (that is, *doing*), then the PCV also has to be a present participle, citing invented examples like the following, which they judge unacceptable:

(i) The thing I'm doing is (to) try to learn from my mistakes.

But the corpus studies we report here turned up enough real examples similar to (i) to convince us that Flickinger and Wasow were mistaken.

[3] Examples b, d, and f had *to* in the original.

[4] The data in our statistical model were collected in the summer of 2012, when the corpus was somewhat smaller (425 million words) and did not yet have data from 2012.

COCA is tagged for part of speech, but not syntactically parsed. It has a user-friendly web interface, which extracts examples based on patterns that may include parts of speech, particular words, disjunction, and wildcards. A small window of context around the matching text can also be extracted.

An earlier pilot study (Wasow et al. 2012) of optional *to* in the DBC had involved hand-coding 1000 randomly selected examples from the spoken portion of COCA for a variety of factors that we thought might correlate with *to* use. For this chapter, our dataset was much larger, including written as well as spoken examples. By using computational tools for extraction, culling, and annotation, we were able not only to obtain considerably more data but also to consider more factors than in the pilot. These factors are described in Sect. 3; in the remainder of this section, we describe the extraction, culling, and annotation process.

We initially extracted all examples that included some form of the verb *do,* followed by some form of the verb *be,* optionally followed by *to,* and (obligatorily) followed by any verb in base form.[5] The extraction pattern allowed up to two[6] words to intervene between any two of these words—that is, it could be abbreviated as:

DO (W)(W) BE (W)(W) (*to* (W)(W)) V[base]

where "DO" means any form of *do,* "W" means one word, and "BE" means any form of *be*. The resulting sample was then parsed with the Stanford parser (Klein and Manning 2003). Through trial and error, we developed a tgrep2 (Rohde 2005) pattern to help us cull out examples that were not in fact instances of DBC.

Annotation was done with Perl scripts, some of which made use of the parses. The most obvious need for the parses was in measuring the lengths of constituents, since that required assigning constituent structure. But the parses were also used in identifying such things as the occurrences of *do* and *be* whose forms we thought might influence *to* use. The annotations provided by the scripts were subsequently used to automatically code the data for the factors we considered for use in modeling. In some cases, the annotations could simply be used as codings (for example, the form of *do* in the subject and whether the example was written or spoken), but in others some additional computation was required—e.g., the measure of subject length was computed by subtracting the position of the subject's head noun in the sentence (that is, its distance from the start of the sentence) from the position of *do* in the sentence. These computations were carried out by an R script, which also renamed some of the annotations and removed unused fields.

Some codings (e.g., ones that did not give one of our seven nouns as the head noun of the subject or that gave the number of words between *do* and *be* as more

[5] COCA has two distinct tags verb.BASE and verb.INF for uninflected nonfinite verbs. We have not been able to discern a consistent basis for this distinction, although verb.INF seems to appear after *to* at a considerably higher rate than verb.BASE. In all of our searches, we used the disjunction of these two tags to search for what we call base forms of verbs. For the purposes of this chapter, we treated the two COCA tags as interchangeable. That is, when we say a verb's form is base, we mean it is uninflected and not preceded by *to*; and when we say a verb is infinitival, we mean it is preceded by *to.*

[6] The limitation of at most two intervening words was required for computational reasons.

than two) triggered hand checks of particular examples, and additional random hand checks were performed. Altogether, we hand-checked hundreds of examples and discarded examples that were not the type of DBC sentences we were investigating. Our final dataset contained 10,116 examples, but 143 of them had uncoded values for some variable used in our analysis. Furthermore, only one example involved the *were* form of the copula. We dropped these 144 examples before statistical analysis so that our analyses involved 9972 examples. In a random check of over 100 examples, all were examples of the DBC with base or infinitival PCV.

We used the same pipeline to extract and annotate DBC examples from the Fisher corpus (Cieri et al. 2004). Fisher consists of telephone conversations on designated topics; it is far smaller (about 22 million words) than COCA. Analysis of the Fisher dataset was qualitatively consistent with the COCA model we report below, but the number of examples extracted (861) was too small to show reliable effects for many of the significant factors in our COCA data. Consequently, we provide a detailed accounting only of the COCA study.

3 Factors in Our Analysis

Based on earlier work on optional *that* in both relative clauses and complement clauses (see Jaeger 2010; Wasow et al. 2011, inter alia), we expected that similar factors might influence the presence or absence of *to*. In particular, we expected factors that contribute to the processing difficulty of a DBC sentence would increase the probability of *to* use. These factors include long and/or syntactically complex phrases within the sentence. They also include the use of relatively infrequent words or word forms.

Why should processing difficulty encourage *to* use? The obvious answer is that the extra little word takes time, giving the speaker an extra fraction of a second for planning the remainder of the utterance and lexical retrieval. The extra time is also useful for the listener, providing more time for parsing and lexical retrieval. The work on *that* suggests that these effects show up in writing as well as in speech, even though our hypotheses about why they occur are based on the temporal pressures on speakers and listeners. This could be due either to habits of speech being preserved in writing, or to similar temporal pressures on readers. We will not attempt to resolve this question here.

We coded measures of phrasal complexity and word frequency based on the parts of the utterance most closely connected with the site of optional *to* and thus most likely a priori to influence speaker's choice, where by "connected" we mean parts of the utterance that are components of the DBC (see (2) above) and/or are close to optional *to* in terms of linear ordering. For phrasal complexity, this led us to code the amount of material in (i) the subject NP between the head noun and *do,* (ii) between *do* and *be,* (iii) between *be* and the PCV, and (iv) in the PCVP. We expected that in all cases, more material would lead to greater utterance complexity and thus greater preference for *to*. Both length and complexity can, of course, be measured

in multiple ways. For length, we used number of words, though number of syllables might have been as good or better, as might duration (for speech) or number of characters (for writing). There is a substantial body of literature (see, e.g., Hawkins 1994; Wasow 2002) that has found number of words to be a good proxy for more sophisticated measures of complexity. Complexity measures tend to depend on the parse assigned, and the ones we had were not very reliable. Moreover, since complexity is highly correlated with length, the only complexity measure we looked at was number of verbs in a phrase. This turned out to be highly collinear with length and a less reliable predictor, so we ended up relying only on number of words for our length/complexity measures.

We also examined the effects of wordform frequencies[7] for critical components of the DBC: the head of the subject NP, the form of *do,* the form of the specificational copula *be,* and the PCV. For the first three of these, only a small number of wordforms are possible, so in our analysis we directly modeled the *to* use preference associated with each wordform and performed exploratory visualizations of the relationship between preference and in-construction frequency of the word form (Sect. 4.3).

In contrast, there are many different PCVs; furthermore, the PCV is distinctive among DBC components in that there is strong reason from a mathematically precise theory for predicting that its frequency will affect *to* use preference. Namely, the theory of uniform information density (UID; Levy and Jaeger 2007; Jaeger 2010) posits that communicative efficiency is optimized if information is transmitted at a uniform rate, and that speakers take advantage of the grammatical opportunities afforded to them to smooth this information rate out. The notion of "information" here is based on information theory (Shannon 1948), and is measured as log of inverse probability (equivalently, negative log-probability) or *surprisal*. It follows that optional function words like *that* and *to* are more likely to be inserted in environments where, without them, there might be an information peak.[8] To understand how this applies to the DBC, we can use reasoning directly analogous to that developed by Levy and Jaeger (2007) for *that*-use in relative clauses; the key is that the PCV is often the first point in the utterance where it becomes clear that the utterance must involve a DBC. Consider the variant of Example (3a) without *to:*

(5) what we're here on earth to do is celebrate humanity

Before the PCV *celebrate,* there are alternative ways that the utterance could continue that do not involve the DBC:

(6) a. what we're here on earth to do is a complete mystery
b. what we're here on earth to do is unique
c. what we're here on earth to do is not what you think we're here to do

[7] We used frequencies of these forms in our sample, rather than in the whole of COCA.

[8] To test whether people employ this UID strategy in actual usage using corpus studies has required computing information at critical points in utterances on the basis of very local information, usually immediately preceding *n*-grams for some very small *n*.

Therefore, in the *to*-free variant, the PCV conveys two distinct pieces of information about the structure and content of the utterance: (i) the fact of the DBC, and (ii) the identity of the PCV of the DBC. These can be measured information-theoretically as follows:

$$\log \frac{1}{P(\text{PCV}, \text{DBC} \mid \text{Context})} = \log \overbrace{\frac{1}{P(\text{DBC} \mid \text{Context})}}^{\text{(i)}} + \log \overbrace{\frac{1}{P(\text{PCV} \mid \text{DBC}, \text{Context})}}^{\text{(ii)}}$$

The use of *to* separates out these two pieces of information: *to* conveys (i), whereas after *to* the PCV conveys only (ii). Therefore, the optimal distribution of optional *to* from an information-density perspective would be to use it when (i), (ii), or both are large. With respect to PCV information content, this line of reasoning predicts that *to* use in the DBC will be higher when the PCV is less predictable, and (ii) is thus large. In principle, (ii) should be measured with respect to the complete context; but a reasonable and convenient first simplification is to assume that $P(\text{PCV}|\text{DBC}, \text{Context}) \approx P(\text{PCV}|\text{DBC})$—namely, that in-construction PCV frequency allows us to approximate the information content of (ii).

We also expected that priming could increase the probability of *to,* so we expected that, when *do* in the subject was in infinitival form (that is, preceded by *to*) the rate of *to* before the PCV would be increased.

An expectation that was not derivative from the work on optional *that* was that some phonological factors might influence the use of *to*. This idea was suggested to us by Arto Anttila, who has shown the influence of prosody on other syntactic alternations in English (e.g., Anttila et al. 2010). He also brought to our attention a book published over a century ago entitled *Rhythm in English Prose* (van Draat 1910) with a chapter entitled "The infinitive *with* and *without* preceding *to,*"[9] which argued that *to* could have a prosodic function. Based on Anttila's suggestion, we considered whether *to,* which is virtually always unstressed, might sometimes serve to prevent two stressed syllables from appearing adjacent to one another, a situation known to be disfavored and referred to as "stress clash" (see Liberman and Prince 1977). To understand how this might influence speaker choice regarding *to* production, consider the following two examples from the spoken section of COCA, with the presumably[10] stressed syllables in bold (see the next paragraph regarding stress status of the copula *is*):

(7) And **one** of the **best ways** to **do** it **is** (to) **break bread** with **them**.
(8) **All** I can **do is** (to) con**ti**nue to be**have** in a **way** that **earns** your **trust**.

In both cases, the inclusion or omission of *to* has no bearing on the grammaticality of the sentence. However, speakers' *to* use decisions could affect the prosodic optimality of the utterances. In (7), omitting *to* would cause a stress *clash*—a sequence

[9] Interestingly, all of the cases discussed in van Draat's chapter, except the complement of *help* now strike us as categorically either requiring or prohibiting *to*.

[10] No sound files are available for this corpus, so our assignments of stress in these examples are based on our own intuitions.

of two consecutive stressed syllables, *is* and *read*—that would be avoided by the use of *to*. In (8), on the other hand, including *to* would cause a stress *lapse*—a sequence of two consecutive unstressed syllables, *to* and *con-*, which would be avoided by the omission of *to*. If speakers are sensitive to this potential prosodic function of *to,* they should tend to include *to* when its omission would cause stress clash, and omit *to* when its inclusion would cause stress lapse. (In fact, *to* was used in (7) and omitted in (8) in COCA.) Note that in cases where nothing (other than *to*) intervenes between the copula and the PCV, the predictions of clash avoidance and lapse avoidance are identical: A PCV with initial stress should favor *to* use more than a PCV with noninitial stress. Since this covers a large majority of our examples, we conflated clash avoidance and lapse avoidance into one factor, which we refer to as clash avoidance, or simply (potential) stress clash.

We determined stress clash by annotating (i) the PCV for whether it had initial stress and (ii) the word immediately preceding the PCV (or the word immediately preceding *to,* if *to* is used) for whether it had final stress. Since the word immediately preceding the PCV is normally a copula—usually *is*—our ability to investigate potential effects of stress clash hinges on whether the copula is stressed. Is it? If so, it is not clearly audible. On the other hand, the fact that *is* in the DBC is never contracted (and sounds quite unacceptable when contracted, e.g., **All you need to do's pay attention*) suggests that it does carry some stress. We thus considered the copula as stressed in our dataset.

In cases where something (other than *to*) intervenes between the copula and the PCV, these arguments based on prosody depend on the stress pattern of the intervening material. The situation is somewhat complicated by the fact that, when *to* appears, intervening material can appear before and/or after *to*. Hence, when there is intervening material but no *to,* there may be multiple locations where insertion of *to* would be grammatical, and the effect on prosody would be different in each one. To avoid this complication, we made the simplifying assumption that, in cases without *to,* the alternative we were comparing the actual sentence to was one with *to* immediately preceding the PCV. Because the vast majority of examples do not have material intervening between the copula and the PCV, this simplification is unlikely to have materially affected our results.

In addition to prosody, we thought segmental phonology might, conceivably, influence the use of *to*. Our reasoning was that, if the initial segment of the PCV is too similar phonologically to the final segment of the preceding word, the word boundary might be obscured. We conjectured that such a situation might favor *to* use. Consequently, we coded for the initial segment of the PCV and for the final segment of the preceding word.

The final factor we thought might affect *to* use is whether the sentence in question is spoken or written. Our sample contained roughly equal numbers of examples from speech (4865) and writing (5251), and we included this factor as one we considered. We did not actually know what to expect in terms of this factor's effects. On the one hand, the DBC occurs at a much higher rate in speech than in writing (recall that COCA is 80% written), and written language tends to employ more complex structures and longer sentences than speech. These factors would lead to

the expectation of a higher rate of *to* use in writing than in speech. On the other hand, if one of the reasons for using *to* is to buy time in production, then it should appear more frequently in speech. It turns out that speech favors *to* use: *to* occurs in 38 % of our spoken examples, compared to 29 % of written examples.

The following is a list of the factors we used in our analyses:

- Head noun of the subject. We had four values for this: ALL, THING, WHAT, and SUPER (for "superlative"), where the last category includes the relatively rare head nouns *best, worst, most,* and *least,* plus the handful of examples with something else heading the subject.
- Subject length. This was measured in words, from the head noun to *do*.
- Form of *do*. We considered seven forms: base, infinitive (with *to*), present tense nonthird person *do, does, did, done,* and *doing*.
- Form of the copula. The vast majority of the examples have *is,* but *was* also occurs with some frequency, and there are some examples with *are*.
- Number of intervening words between *do* and *be*. In the COCA data, this could be zero, one, or two, with most cases being zero.
- Number of intervening words between *be* and the PCV. Again, in the COCA data, this could be zero, one, or two, with most cases being zero.
- PCVP length. This was measured in words, including the PCV (but excluding *to* when present). It relied on the parse tree to find the end of the PCVP.
- Frequency of the PCV in our sample. As is standard in corpus studies, we used the log of the frequency.
- Stress clash. This would occur (without *to*) if the PCV had initial stress and the preceding word had final stress. We treated the copula has having final stress.
- Segmental phonology. We classified the initial segment of the PCV and the final segment of the preceding word (not counting *to,* when present) into one of four categories: vowels, sibilants, sonorants, and other. We then coded each example for whether the two segments in question were of the same or different categories.
- Speech versus writing.

Figure 1 shows univariate statistics for four of these factors: number of *do-be* and *be*-PCB interveners, segmental phonology, and stress clash. Univariate statistics for other factors can be found in Sects. 4.3 and 4.4.

4 Model of the Data

4.1 Mixed Logit Models

To analyze the effects of these various factors in our data, we use *mixed-effects* (sometimes also called *hierarchical* or *multilevel*) *logistic regression* analysis (or *mixed logit* analysis for short; Pinheiro and Bates 2000; Bresnan et al. 2007; Baayen et al. 2008; Jaeger 2008). Mixed logit analysis uses data to infer the dependence of

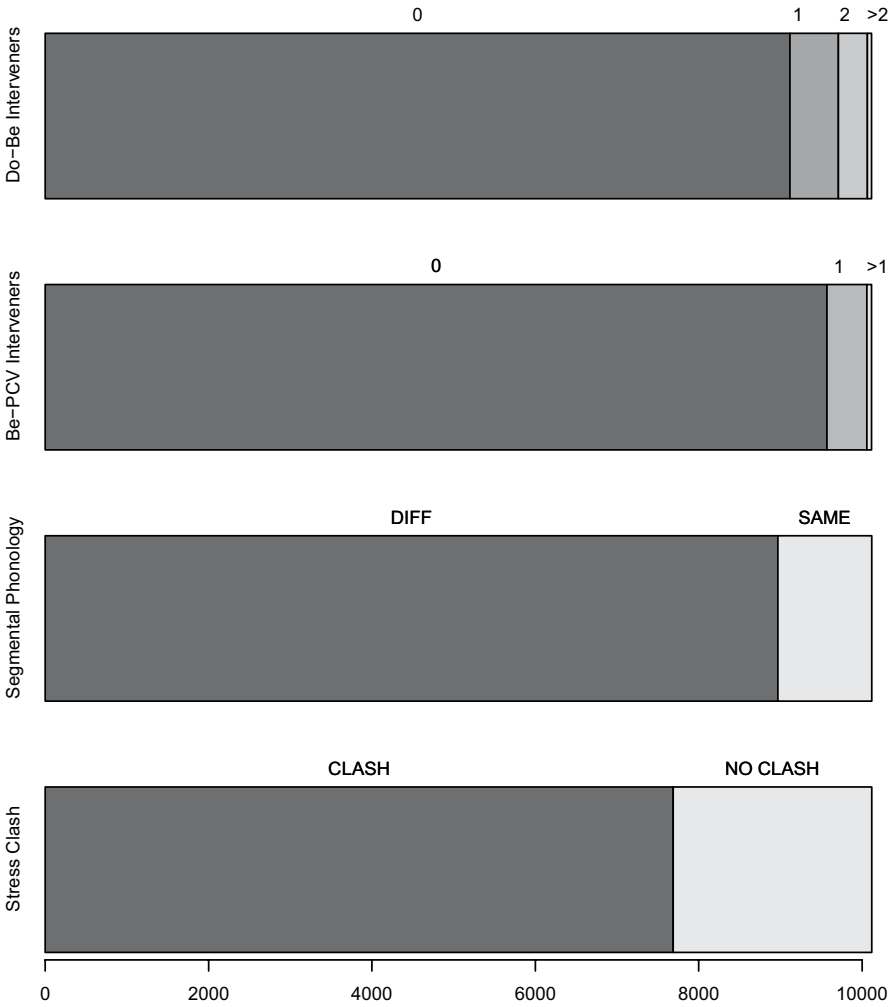


Fig. 1 Univariate statistics for *do-be* interveners, *be-PCV* interveners, segmental phonology of initial PCV segment and final segment of preceding word, and potential stress clash. *PCV* post-copula verb

a single, dichotomous *response variable*—in our case, whether the optional word *to* is used in a given utterance—on one or more *predictors,* allowing for the possibility that different factors may have overlapping and even interacting influences on the response variable. In particular, mixed logit analysis follows the assumption of basic logistic regression (Cedergren and Sankoff 1974; Agresti 2002) that all effects of and interactions among predictors can be expressed in terms of additive effects on the *log odds* of the outcome of the dependent variable; these effects are the *regression coefficients* and inferred from data. For example, consider two hypothetical utterances differing only in the head noun of the subject:

(9) a. All she did was (to) stare and smile
b. What he did was (to) stare and smile

If the difference in the regression coefficients associated with *what* and *all* were, for example, log(4) = 1.39, then the difference in the log odds of *to* use between the examples would also be 1.39. Additive effects on log odds can equivalently be expressed as multiplicative effects on odds, so the ratio of the odds of *to* use in the two examples would be $e^{\log(4)} = 4$: if the odds of *to* use were 1:2 for (i) (33 % chance of using *to*), then the odds for (ii) would be 2:1 (67 % chance); if the odds were 1:1 for (i), then the odds for (ii) would be 4:1 (80 % chance), and so forth. We code our dependent variable and predictors such that positive regression coefficients indicate favoring *to* use, whereas negative coefficients indicate favoring *to*-omission.

Mixed logit analysis extends this picture by adding to the "fixed effects" of ordinary logistic regression a set of "random effects": idiosyncratic departures from the "overall" population norm in baseline behavior and sensitivity to predictor variables that vary across meaningfully clustered subsets. In our case, it is the PCV that makes mixed logit analysis essential. Since the presence or absence of *to* is a feature of an utterance highly local (in both linear and structural terms) to the PCV, it is quite plausible that different PCVs might possess idiosyncratic preferences regarding baseline level of *to* use due to historical accident and/or systematic pressures that we have not measured and included in our model. Furthermore, PCVs have a nearly Zipfian distribution (Zipf 1936) in our dataset (Fig. 2) so that some PCVs are attested in dozens or even hundreds of utterances. If PCVs do in fact have idiosyncratic *to* use preferences—for example, if *get,* which occurs in over 900 observations, idiosyncratically prefers *to* more strongly than *make,* which occurs in nearly 500—not including such preferences in our model will interfere with the inferences we draw regarding the effects of other predictors. Finally, note that several theoretically critical predictors in our model—including in-construction PCV frequency, potential stress clash, and segmental phonology—are nearly completely determined by which PCV occurs in the construction.[11] By including a by-PCV "random intercept" in our model, we avoid the "language as fixed effect fallacy" (Clark 1973; Barr et al. 2013) and ensure that, if we conclude that these predictors reliably affect *to* use, it is above and beyond any apparent patterns that might emerge due to idiosyncratic PCV-specific preferences alone. For the same reasons, we include a by-PCV "random slope" for the effect of corpus type in our model, since there could be PCV-specific differences between speech and writing in *to* use preferences. The complete formal specification of our mixed logit analysis is as follows: the probability of *to* use in a given utterance with fixed-effects predictors denoted by $x_1, \ldots, x_n$ is

$$P(to) = \frac{e^{\eta}}{1 + e^{\eta}}$$

[11] We say *nearly* because in the infrequent cases when material such as adverbs intervene between the copula and the PCV, stress-clash and segmental phonology predictors are determined by that material, not by the PCV.

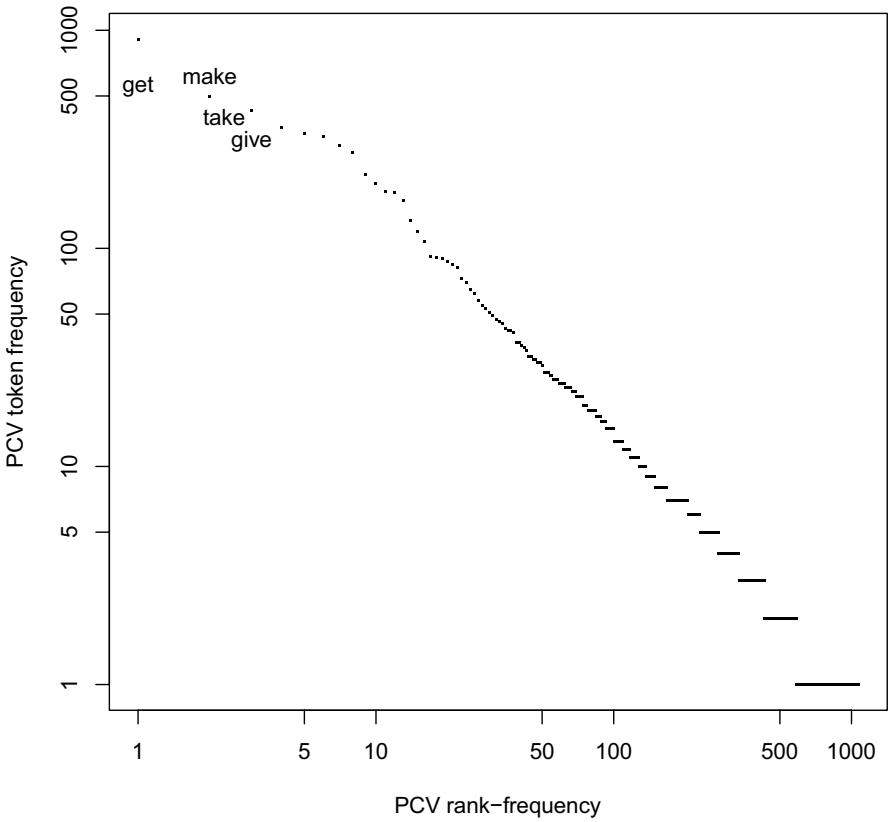


Fig. 2 The near-Zipfian distribution of post-copular verbs in the do-be construction (DBC)

where the *linear predictor* η is

$$\eta = \alpha + \beta_1 x_1 + \cdots + \beta_n x_n + a_{PCV} + b_{PCV}\, CorpusType$$

and a_{PCV} and b_{PCV} are jointly normally distributed PCV-specific regression coefficients. We fit our model using version 0.999999-2 of the lme4 package of R (Bates et al. 2013), which estimates mixed logit models by maximizing Laplace-approximated data likelihood.

4.2 Base Model Results

Our fitted regression model is given in Table 1. Several of the factors in our model are nonnumeric, specifically: head noun of the subject, form of *do,* form of the copula, stress clash, and speech versus writing. For Table 1, we employed what is known as "treatment coding" (Chambers and Hastie 1991) for these factors: one

Table 1 Overall model fit

```
Random effects:
 Groups Name                Variance Std.Dev. Corr
 PCV    (Intercept)         0.990401 0.99519
        corpus.typeWRITTEN  0.028768 0.16961  0.565
Number of obs: 9972, groups: PCV, 1060

Fixed effects:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)          -2.287661   0.166618 -13.730  < 2e-16 ***
Subj.NP.headSUPER     2.579917   0.108405  23.799  < 2e-16 ***
Subj.NP.headTHING     1.871227   0.086696  21.584  < 2e-16 ***
Subj.NP.headWHAT      1.421681   0.082513  17.230  < 2e-16 ***
Subj.length.posthead  0.081163   0.020336   3.991 6.58e-05 ***
do.be.interveners     0.430031   0.048972   8.781  < 2e-16 ***
be.PCV.interveners   -0.021152   0.095580  -0.221    0.825
PCVP.length           0.034689   0.004176   8.307  < 2e-16 ***
do.typedid            0.547730   0.126449   4.332 1.48e-05 ***
do.typedoes           0.219956   0.176829   1.244    0.214
do.typedoing          3.158128   0.595429   5.304 1.13e-07 ***
do.typedone           1.661955   0.131475  12.641  < 2e-16 ***
do.typefinite do      0.172856   0.135887   1.272    0.203
do.typeinf do         0.939312   0.081400  11.539  < 2e-16 ***
be.formare            0.303157   0.547914   0.553    0.580
be.formwas            0.836636   0.075825  11.034  < 2e-16 ***
corpus.typeWRITTEN   -0.275712   0.060109  -4.587 4.50e-06 ***
seg.phonSAME         -0.038220   0.122963  -0.311    0.756
stressNO CLASH       -1.012925   0.106449  -9.516  < 2e-16 ***
log(PCV.freq.in.DBC) -0.235447   0.038228  -6.159 7.32e-10 ***
```

value of the factor is arbitrarily selected as the baseline, and each of the other possible values appears as a separate predictor in the regression, with the coefficient representing the difference in effect on *to*-preference between the value in question and the baseline value for the factor. These baseline values are ALL for subject NP head, base *do* for *do*-type, *is* for *be*-form, SPOKEN for corpus type, and CLASH for stress. For the continuous predictors, in our initial model fit, we assume simple linear effects on the linear predictor, but explore possible nonlinear effects on *to* use later in this section. Positive values in the first column of Table 1 indicate that the predictor correlates positively with *to* use, all other predictors being held constant; negative values indicate a negative correlation with *to* use. The absolute value of the regression parameter estimate in the first column indicates the strength of the predictor's effect, and the final column gives a measure of statistical significance based on the Wald *z* statistic.[12]

[12] Each major predictor statistically significant in Table 1 is also significant by a likelihood-ratio test in which the null hypothesis includes a random by-PCV slope for the predictor (results not shown).

We first summarize those results that can readily be understood from Table 1, and later proceed to explain results that require further visualization. To begin with, the random effects part of our fitted model assigns considerable idiosyncratic variability across PCVs in *to*-preference not captured by other predictors in our model: the random by-PCV intercept has standard deviation 1.00 (units on the logit scale).[13] However, the idiosyncratic difference in *to*-preference of any given PCV in written versus spoken usage is very small (standard deviation 0.17): PCV-specific *to* use preferences are consistent across genre.

Moving on to fixed effects, we see several critical pieces of evidence supporting our general predictions. Our general prediction from the perspective of utterance planning was that factors increasing memory load and planning difficulty should also increase the rate of *to* use. In Table 1 we see this prediction confirmed in the positive parameter estimates for the effects of the post-head length of the subject NP, the number of words intervening between *do* and *be,* and the length of the PCVP. All of these estimates differ significantly from zero at $p<0.005$ or more highly significant. We also see confirmation of the more specific prediction of UID: the higher the conditional log-probability of the PCV given the preceding context (here, crudely approximated by conditioning on the fact of being in the DBC), the less likely *to* is to be used. For the continuous predictors and for speech versus writing, we can see from the table that the correlations are all in the direction predicted. Thus, UID and the more general hypothesis that difficulty in utterance planning favors *to* use receives broad empirical support. The exception is that the presence and number of interveners between *be* and the PCV has no effect on *to* use preference in this model—but see Sect. 4.5 for further discussion of this predictor.

We also explored two predictions regarding the effects of phonological predictors on *to* use. The predicted effect of segmental phonology—namely, that the first segment of the PCV and the final segment of the immediately preceding word being of the same type would promote *to* use—was not borne out. However, the predicted effect of prosody—that when the first syllable of the PCV and the final syllable of the immediately preceding word are both stressed, *to* use would be favored to eliminate stress clash—was strongly confirmed. This can be seen in Table 1 from the fact that the parameter estimate associated with NO CLASH is negative (with respect to the baseline level of CLASH).

4.3 Categorical Predictors in Greater Detail

We now examine in greater detail the effects of categorical predictors with more than two levels: subject NP head, *do* type, and *be* form. Although Table 1 contains all the information necessary to reconstruct the effect of each of these predictors, it is not the easiest format in which to visualize these effects, in particular because

[13] To perhaps give a better sense of effect sizes seen in our regression model, a difference of one unit on the logit scale is equivalent to the difference between *to* use probabilities of, for example, 0.02 and 0.05, between 0.05 and 0.12, between 0.12 and 0.27, or between 0.27 and 0.5.

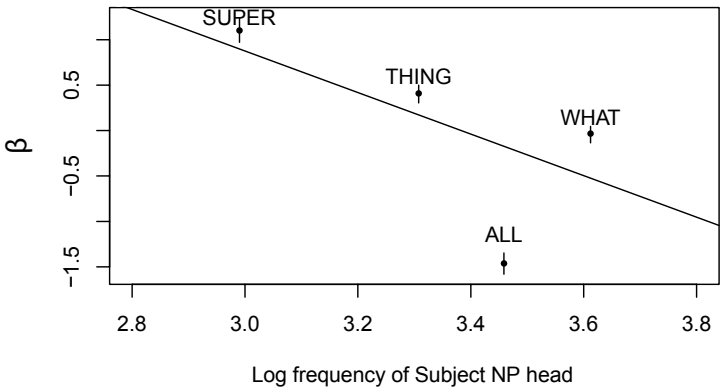


Fig. 3 The effects of different subject NP heads on *to* use preference. Error bars are standard errors of the regression parameter estimate

the degree of confidence in the size and direction effect of the "baseline" level of each predictor is not visible. For this reason, the next series of figures provide visualizations of these effects based on "sum" or "deviation" coding of predictor levels, where the effect of each predictor level is estimated subject to the constraint that the sum is zero. This representation also allows us to explore our general hypothesis that low-frequency material favors *to* use due to difficulty in utterance planning and production. Because the subject NP head, *do,* and the copula are all critical components of the *DBC,* it is likely that they would have similar influences as the PCV: The more frequently the particular variant of each component occurs in the construction, the more it should favor *to*-omission. We visualize the extent to which each predictor's effects conform to this hypothesis by passing weighted best-fit lines through the estimated effects in each plot (Fig. 3 through Fig. 5).[14] Since each of these three components has only a small number of variants, our results regarding relationship of variant frequency against *to* use preference are necessarily exploratory but, as will be seen momentarily, are provocatively consistent with our general theoretical predictions.

Figure 3 shows the effects of different subject NP heads on *to* use preference. As predicted by our general hypothesis, we see a general trend for more frequent subject NP heads to dispreferr *to* use more strongly. However, the dispreference of the head *all* for *to* is far stronger than would otherwise be expected from this tendency. We have no explanation at present for this exception.

The form of *do* is a particularly interesting factor, as shown in Fig. 4. Our prediction that more frequent forms[15] of *do* would have lower rates of *to* use holds up well,

[14] The weights for the best-fit line are the inverses of the squared standard errors of each parameter estimate.

[15] Frequency is measured as the number of occurrences of the form in question as the obligatory *do* of the DBC in our dataset.

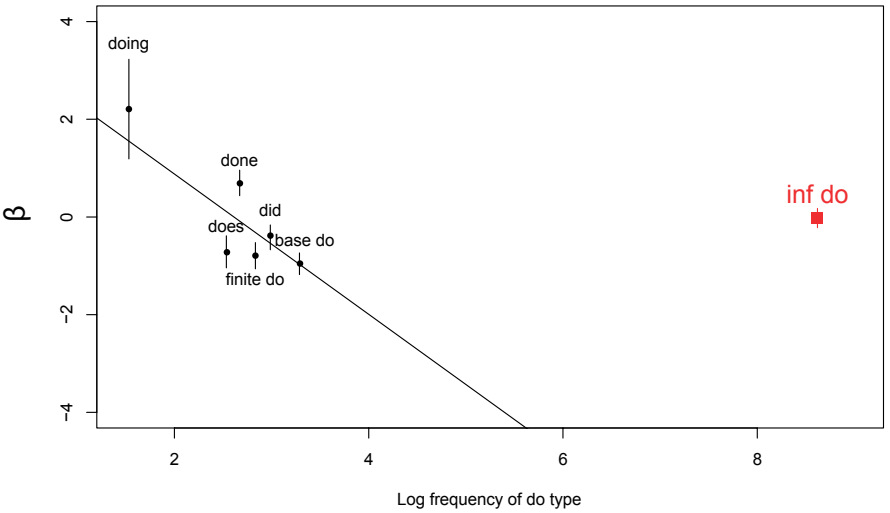


Fig. 4 The effects of different *do* types on *to* use preference. Error bars are standard errors of the regression parameter estimate

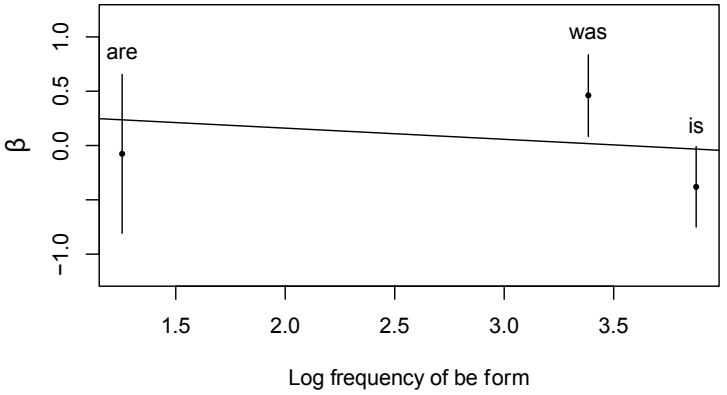


Fig. 5 The effect of *be* form on *to* use preference. Error bars are standard errors of the regression parameter estimate

except for one severe exception: When *do* is infinitival (that is, *to do*), the use of *to* with the PCV is much higher than would be predicted on grounds of frequency. This exception, however, is consistent with another psycholinguistically motivated prediction we made: that the *to* in the infinitival *do* primes the later use of *to.* Thus, Fig. 4 nicely matches our predictions.

Figure 5 shows the effects of different forms of the copula on *to* use preferences. Although there are only three distinct forms in our dataset,[16] and one of them, *are,* is so infrequent that our model has little confidence in its precise effect, the general trend, driven by relative preferences for *is* and *was*, is for more frequent copula forms to be associated with less *to* use, once again consistent with our general hypothesis.

In sum, in all four "critical" components of the DBC—the subject NP head, the form of subject NP-internal *do,* the form of the main clause copula, and the choice of PCV—we find the same pattern emerging: The lower the in-construction frequency of the variant of a component, the more strongly the variant favors *to* use. One clear exception to this generalization, that infinitival *do* does not disfavor *to* despite its being far and away the most common *do* form, has an independent explanation, namely that it induces repetition priming of *to* use in the main clause. Hence, we see broad support from construction component frequencies for our hypothesis that utterance complexity favors *to* use.

4.4 Continuous Predictors in Greater Detail

We now move on to a more detailed investigation of the effects of the continuous predictors for which we found significant effects on *to* use preference in the base model of Table 1. Our depth of understanding of these effects is limited by the assumption built into this base model that the effects of these predictors are linear in log-odds space. For each of these predictors, we explored their effects on *to* use in more depth by relaxing this assumption: We removed the predictor from the basic model of Table 1 and put in its place a richer version of the predictor using restricted cubic splines (Green and Silverman 1994), which allow the model to learn nonlinear effects of the predictor on the log odds of *to* use. Figure 6 through Fig. 8 depict these effects, together with 95 % confidence intervals, controlling the effects of other predictors; as with Table 1, more positive values indicate stronger preference for *to* use. At the bottom of each figure is a summary of the data distribution for the predictor in question: for discrete predictors (subject and PCVP length), a histogram of counts among the 9972 total in the model, and for the continuous predictor of in-construction PCV frequency, a kernel density estimate.

Figure 6 shows the results for the post-head length of the subject NP. This figure reveals a regularity invisible in the base model of Table 1: that the general pattern of longer subject NPs favoring *to* use is reversed for very short subject NPs with four or fewer post-head words. The reason for this reversal is currently unclear to us. One speculative suggestion is as follows: In many utterances with three or four post-head subject NP words, the only material beyond the minimum (which is two words: a single-word subject of the relative clause, and *do*) is auxiliary and/or

[16] Note that we discarded the one instance of a *were* copula since one instance is insufficient data to estimate that form's effect.

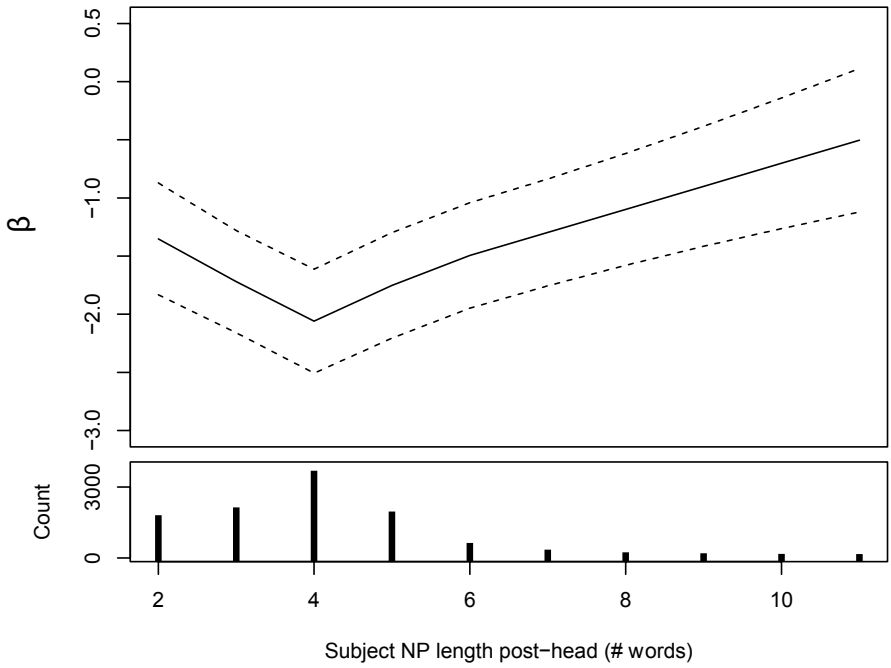


Fig. 6 Effect of the number of post-head words in the subject NP on *to* use preference

modal verbs preceding *do,* which may not add appreciably to utterance complexity (Warren and Gibson 2002). This speculation would require further research to investigate seriously, however.

The effect of the length of the PCVP is illustrated in Fig. 7. We see a near-linear effect of PCVP length on *to* use preference throughout its range; the linearity assumption of the base model in Table 1 was in fact reasonable.

Figure 8 shows the effect of in-construction PCV log-frequency. As with PCVP length, we see a near-linear effect throughout the range of PCV frequency (though model confidence in effect shape drops off for the sparse, highest-frequency PCV range), validating the linearity assumption of the base model in Table 1, which ultimately derived from the theory of UID.

In sum, more in-depth spline-based analyses of our continuous predictors largely validate the linearity assumption implicit in the base model of Table 1. The one exception is that there is a reversal of the subject NP post-head length effect for the shortest subject NPs, a pattern whose source we have speculated on but would require further research to understand more fully.

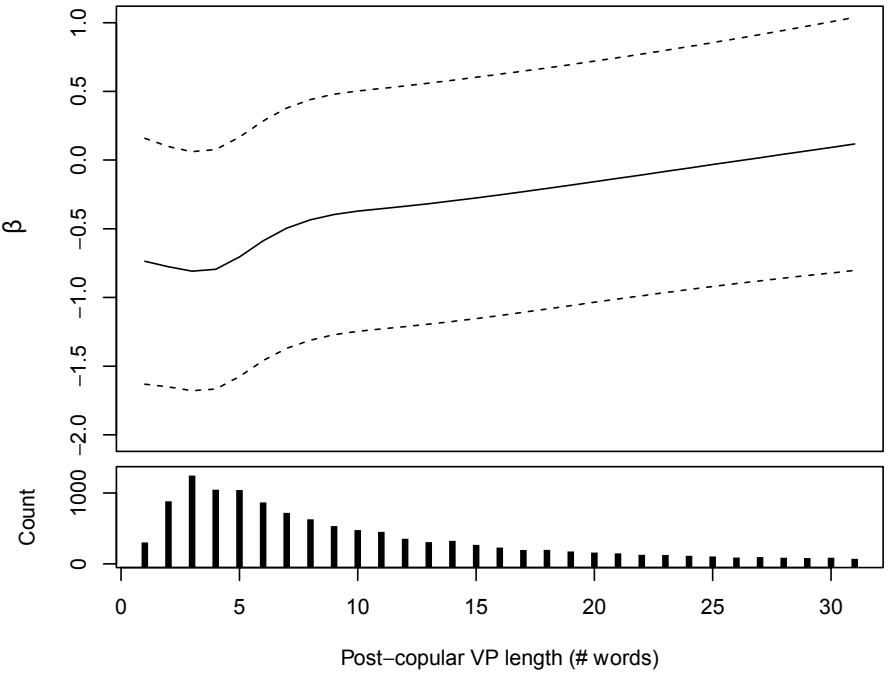


Fig. 7 Effect of PCVP length on *to* use preference. PCVP post-copula verb phrase

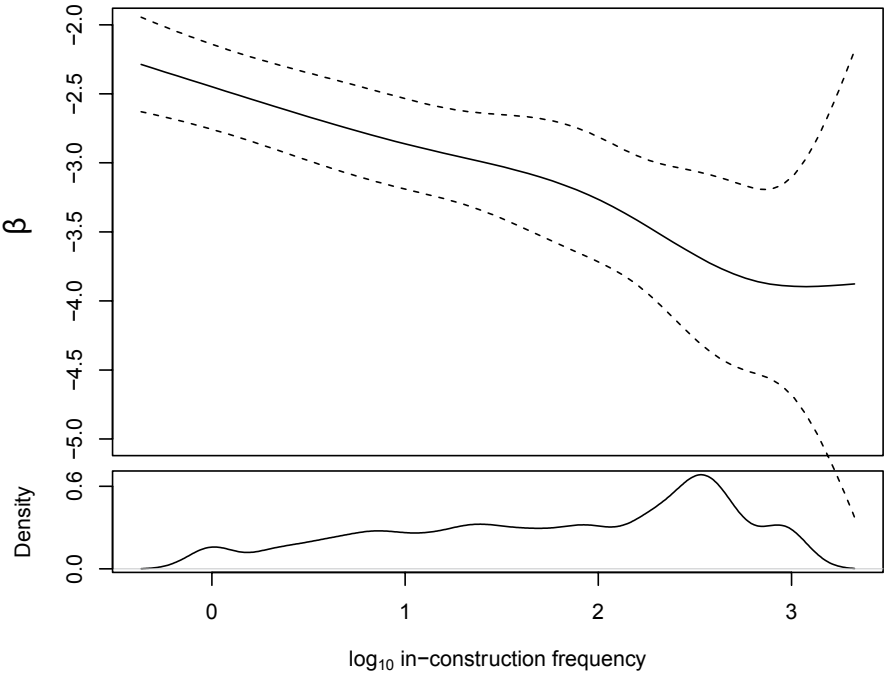


Fig. 8 The effect of in-construction PCV frequency on *to* use preference. *PCV* post-copula verb

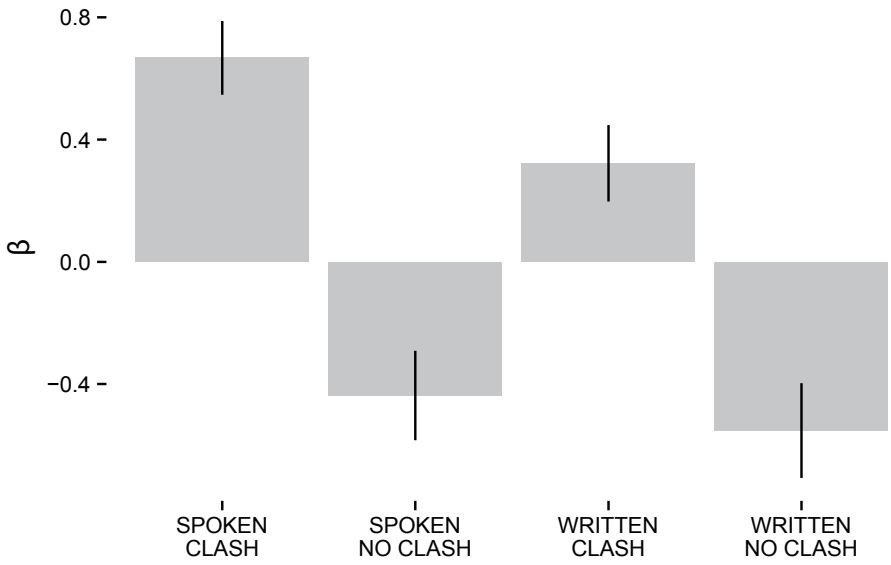


Fig. 9 Effect of potential stress clash on *to* use preference in speech and writing. Error bars show standard errors of the regression parameter estimates

4.5 Interactions with Corpus Type

Written language shows at least one difference from speech in its lower overall *to* use rates; moreover, it is not subject to the same real-time production pressures as speech and does not normally indicate what should be stressed (other than occasional marking of contrastive stress through devices like boldfacing). Thus, it is natural to ask whether the effects of complexity and stress clash avoidance in our model may differ between speech and writing. To answer this question, we tested for significant interactions between corpus type and each of the other predictors in our model, using likelihood-ratio test model comparison in each case between the base model of Table 1 and a minimally enriched model in which only an interaction between corpus type and the predictor in question was added. Our predictors fell into three categories. In the first category are predictors that did not interact significantly with corpus type: subject NP head, post-head subject NP length, form of the copula, and stress clash. Of the predictors in this first category, stress clash deserves more lengthy discussion, because there is a numerical interaction between corpus type and stress clash that is marginal in statistical significance ($p = 0.06$), with stress clash mattering less in writing than in speech,. More importantly, however, the effect of stress clash is highly significant in each corpus type individually ($p << 0.001$), with the same qualitative effect: potential stress clash favors *to* use. Figure 9 illustrates this effect, with effect estimates and standard errors derived from a sum-coding representation of the interaction. The existence of the effect

in the written data provides support for the part of Janet Fodor's (1998) "Implicit Prosody Hypothesis" that Fodor (2002) formulates as follows: "In silent reading, a default prosodic contour is projected onto the stimulus…." Moreover, it suggests that writers are influenced in their wording choices by this implicit prosody.

In the second category are predictors that interact significantly with corpus type, but not in ways that lead to a qualitative change in our overall picture: *do* type, number of interveners between *do* and *be,* and PCVP length (all $p < 0.01$). Number of *do*-*be* interveners and PCVP length have the same effect in speech and writing (with more of each favoring *to* use), but in each case the effect is stronger in writing than in speech. In the case of *do* type, the interaction involved the forms *did* and *done* favoring *to* use more strongly in speech than in writing. These past-tense forms are less common in speech than in writing, so this result is also consistent with our theoretical picture of less frequent components of the construction favoring *to* use.

The sole predictor in the third category was the number of interveners between *be* and the PCV, which interacted significantly ($p << 0.001$) with corpus type in a theoretically important way. Recall that in the base model of Table 1, *be*-PCV interveners had no effect on *to* use preference. However, adding an interaction with corpus type resulted in a far better fitting model (likelihood-ratio test $p << 0.001$). To understand this interaction, we nested *be*-PCV interveners inside corpus type, and added random by-PCV slopes of *be*-PCV interveners and its interaction with corpus type; in this model, we found that more *be*-PCV interveners favored *to* use in written English ($\beta = 0.66$, $p < 0.001$) but marginally disfavored *to* use in spoken English ($\beta = -0.21$, $p = 0.09$). A likelihood ratio test confirmed that the interaction between *be*-PCV interveners and corpus type is highly significant ($p < 0.001$) in this model with maximal random effects structure with respect to this critical interaction (see Barr et al. 2013).

Why would the effect of *be*-PCV interveners, unlike all our other measures of utterance complexity, differ qualitatively between speech and writing? Consider this: additional material inserted between *be* and the PCV, unlike additional material in the NP subject or the postverbal part of the PCVP, is in the same position as optional *to*. While some types of *be*-PCV interveners may be semantically "full"—obligatory in order for the utterance to convey the speaker's intended meaning—others may be semantically "empty," and the speaker's use of them may be driven by the same considerations—utterance planning and prosodic optimization—that drive *to* use. The following pair (both from the spoken part of COCA, italics indicate the intervener) illustrates this potential contrast, the first semantically "full" and the second "empty":

(10) a. all we have to do is not continue the $ 100-billion-a-year increase that Obama and the Democrats put into domestic discretionary spending
b. all it has to do is just jump down that hill right there

On this view, semantically "empty" material may sometimes be used instead of *to* and thus disfavor it. If this view is correct, and such semantically "empty" material disfavoring *to* is more common in speech than in writing, it could explain the

Table 2 Rate of *to* use in speech versus writing for cases with a one-word *be*-PCV intervener

	Spoken COCA	Written COCA
Singleton intervener is NOT *just*	34.6%	44.6%
Singleton intervener is *just*	14.7%	43.5%

discrepancy seen in the effect of *be*-PCV across the corpus types: a true underlying effect of semantically "full" material that favors *to* could be obscured by a higher incidence of semantically "empty" material in speech. We explored this hypothesis by focusing on the single-most common *be*-PCV intervener, the word *just*. Although it is difficult to judge when and to what extent *just* is semantically "full" versus "empty," there are few if any interveners that are likely to be "empty" more often than *just*. As it turns out, the behavior of *just* is highly revealing. Table 2 shows the rate of *to* use in speech and writing among utterances with single-word *be*-PCV interveners.[17] In writing, the rate of *to* use is approximately the same for *just* and other single-word interveners. In speech, however, *just* disfavors *to* use far more strongly than other single-word interveners.

This speech-specific dispreference of *just* for *to* use provides initial confirmation of our hypothesis. We tested the hypothesis more rigorously by fitting a model with both the intervener by corpus type interaction, a main effect of single-word *just*, and an interaction between *just* and corpus type (with a maximal random effects structure with respect to these parameters). In this model, *just* significantly disfavored *to* use in speech ($\beta = -0.68$, $p << 0.001$) but had no effect in writing ($\beta = -0.01$, $p = 0.96$); more *be*-PCV interveners still favored *to* use in writing ($\beta = 0.64$, $p < 0.005$) but now had no effect in speech ($\beta = -0.03$, $p = 0.78$). That is, simply by accounting for the possible effect of *just* as behaving differently from other *be*-PCV interveners, the reverse effect of interveners in speech disappeared altogether. We speculate that the underlying effect of semantically "full" interveners may be to favor *to* in speech as in writing, but remains obscured by a longer tail of other semantically "empty" interveners individually less frequent that *just*. We leave assessment of this speculation as an open question for future research.

5 Conclusions and Directions for Future Work

Our study of optional *to* in the DBC suggests that processing factors familiar from the study of optional *that* play a major role in determining where *to* is used. These factors include measures of structural complexity and in-construction word frequency, including the specific prediction from the theory of UID that in-construction frequency of the post-copular verb will be negatively associated with *to* use. These findings support the idea that these factors apply quite generally to language production and are likely to influence the use of other optional function

[17] In speech, 40% of these one-word interveners are *just*; in writing, the figure is 31%.

words in similar ways. We also found that prosody, a factor not included in models of optional *that,* seems to play an important role in determining whether speakers use *to* before the PCV in DBC sentences. The fact that the same factor affects the use of *to* in writing provides support for Fodor's notion of implicit prosody.

One broad theoretical consequence of our results is that they constitute evidence against a serial, modularist view of the lexical-selection and phonological-encoding stages of language production. Production is commonly seen as a cascaded process in which lexical selection precedes phonological encoding (Levelt 1993). On a serial, modularist version of this view, preferences stated in terms of representations from the later stage of phonological encoding cannot affect decisions in the earlier stage; on an interactivist view, such effects are possible through self-monitoring and feedback (see Goldrick 2006; Jaeger et al. 2012, for discussion). Our evidence for prosodic effects on lexical selection favors the interactivist view.

A second broad theoretical consequence regards the nature of these interactivist effects. Our key empirical findings all involve the speaker making *to* production decisions that optimize the communicative properties of the utterance. These properties include the time available to prepare or recover from syntactically complex parts of the utterance, the information-density profile of the utterance, and the prosodic contour of the utterance. Our results thus support a view of moment-by-moment language production as being crucially guided by considerations of communicative optimality (Levy and Jaeger 2007; Jaeger 2010). Our results do not, however, speak directly to the familiar question of audience design (Clark and Murphy 1982): Do the effects we see on *to* production reflect speaker-centric production pressures, or effort on the part of the speaker to optimize the utterance for the addressee? This question is beyond the scope of the present chapter.

As another test of the generality of the influence of prosody on optional *to* use, we did a very preliminary check of *to* use in another construction where it is optional, namely, after the verb *help*. As the examples in (11) show, *help* can take VP complements that are either base or infinitival, irrespective of whether an object NP intervenes:

(11) a. a lot of people helped to find you
b. she has helped find dozens of people
c. it did help Austin to find her voice
d. he could help Luke find the gateway

We searched COCA for uses of the verb *help* followed by a verb, with or without an intervening personal pronoun. We did this separately for the spoken and written portions of the corpus.

We made the following working assumptions: *help* is normally stressed; *to* and personal pronouns in this position are typically unstressed; and a large majority of the verb tokens in our searches probably have initial stress.[18] Given these

[18] These assumptions need verification, and are deliberately stated with hedges. Obviously, many verbs are not stress-initial. But more frequent words tend to be shorter, so a high percentage of the verb tokens will be monosyllabic and hence stress initial; and many polysyllabic verbs are also stress initial. The reasoning leading to our predictions does not go through when the pronoun gets

Table 3 *to* Use After *help* in COCA

	Spoken hits	Written hits
HELP V	6989 (78%)	38,000 (77%)
HELP *to* V	1957 (22%)	11,225 (23%)
HELP PPRO V	5637 (88%)	22,012 (90%)
HELP PPRO *to* V	746 (12%)	2578 (10%)

assumptions and the fact that both stress clash and stress lapse are disfavored, we expected to see a far higher rate of *to* use when no pronoun intervenes between *help* and the following verb. Including *to* after a pronoun puts two unstressed syllables next to each other, resulting in stress lapse. On the other hand, including *to* when no pronoun is present often prevents stress clash. Table 3 gives the results of these searches.

In both speech and writing, the rate of *to* use after a personal pronoun is about half of what it is when no pronoun is present. This is what we predicted. Of course, the role of prosody in *to* use after *help* needs to be studied much more carefully, minimally by including further factors (like verb frequency), checking the actual stress patterns of the verbs, and by distinguishing between *helping* and the other (monosyllabic) inflections of *help*. But the pattern in Table 2 strongly suggests that prosody plays a role in the use of optional *to* after *help,* just as it does in the DBC. Moreover, the effect appears to hold in both speech and writing, providing additional support for Fodor's Implicit Prosody Hypothesis.

Returning to the DBC, while our study has made progress towards explaining why *to* is used where it is, a great deal of the variability remains unaccounted for. Our model indicates that individual PCVs have different likelihoods of being preceded by *to,* over and above what can be explained by their in-construction frequencies. Assuming that these differences are not arbitrary lexical idiosyncrasies, we would like to discover what properties of verbs are associated with being preceded by *to* at higher rates.

We conjecture that verb semantics may be relevant, and we have begun investigating one semantic property, namely stativity. This was based in part on a claim of Lakoff (1966), who used the DBC (with *what* as the subject head) as a diagnostic for nonstativity; that is, he claimed that stative verbs could not appear as PCVs in the DBC, giving examples like (12), which he prefixed with asterisks:

(12) a. What I did was hear the music.
b. What Harry did was know the answer.

Our dataset includes many counterexamples to Lakoff's categorical claim, for example (13):

(13) a. what we want you to do is hear some stories of the real-life people
b. one thing you need to do before you go in is know your rights

contrastive stress, or when the form of *help* used is *helping*. But we are confident that our assumptions hold for enough of the data to make this a meaningful preliminary test.

But Lakoff's claim was not entirely off base. He listed 28 stative and 28 nonstative verbs at the end of his paper, and a check of our dataset shows that the nonstative ones occur in our collection at about six times the rate of the stative ones: 1378 for the nonstatives (out of about five million total occurrences of these verbs in COCA) versus 163 for the statives (out of about four million total occurrences of these verbs).

This suggests that there is a semantic incongruence between the DBC and stative predicates, which might make the combination harder to produce and comprehend. If so, this could lead to higher rates of *to* use in DBC examples with stative PCVs. Testing this requires some independent means of assessing the stativity of verbs. And since stative verbs have low frequency in the DBC, we will have to determine whether any effect of stativity on *to* use is already covered in our model by in-construction frequency. We are beginning to investigate these issues, but do not yet have results to report.

Much remains to be done before we know all the factors that influence the use of *to* in the DBC. And a true understanding of the phenomenon will require explanations of why these factors influence *to* use as they do.

Acknowledgment We are grateful to two anonymous reviewers for thoughtful comments on earlier versions of this chapter.

References

Agresti, A. (2002). *Categorical data analysis* (2nd ed.). New Jersey: Wiley.

Anttila, A., Adams, M., & Speriosu, M. (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes, 25*(7/8/9), 946–981.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D. M., Maechler, M., & Bolker, B. lme4: Linear mixed-effects models using S4 classes. R Package version 0.999999-2. 2013. ht tp. *cran. r-project. org/web/packages/lme4.*

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, H. (2007). Predicting the dative alternation. In G. Boume, I. Kraemer, & J. Zwarts (Eds.), *Cognitive foundations of interpretation* (pp. 69–95). Amsterdam: Royal Netherlands Academy of Science.

Cedergren, H. J., & Sankoff, S. (1974). Variable rules: Performance as a statistical reflection of competence. *Language, 50*(2), 333–355.

Chambers, J. M., & Hastie, T. J. (1991). Statistical models. In J. M. Chambers & T. J. Hastie (Eds.), *Statistical models in S* (Chap. 2, pp. 13–44). London: Chapman and Hall.

Cieri, C., Miller, D., & Walker, K. (2004). The fisher corpus: A resource for the next generations of speech-to-text. *LREC, 4,* 69–71.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335–359.

Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. L. Ney & W. Kintsch (Eds.), *Language and comprehension* (Vol. 9, pp. 287–297). Amsterdam: North Holland Publishing.

Flickinger, D., & Wasow, T. (2013). A corpus-driven analysis of the do-be construction. In P. Hofmeister & E. Norcliffe (Eds.), *The core and the periphery: Data-driven perspectives on syntax inspired by Ivan A. Sag* (pp. 35–63). Stanford: CSLI Publications.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D., (2002). Prosodic disambiguation in silent reading. *Proceedings of NELS 32*, M. Hirotani (Ed.). Amherst: GLSA, University of Massachusetts.

Goldrick, M. (2006). Limited interaction in speech production: Chronometric, speech error, and neuropsychological evidence. *Language & Cognitive Processes, 21*(7–8), 817–855.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models: A roughness penalty approach*. London: Chapman & Hall.

Hawkins, J. A. (1994). *A peformance theory of order and constituency*. Cambridge: Cambridge University Press.

Jaeger, T. F. (2008) "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models." *Journal of memory and language* 59(4), 434-446.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology, 61*(1), 23–62.

Jaeger, T. F., Furth, K., & Hilliard, C. (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 38*(5), 1439–1449.

Klein, D., & Manning, C. D. (2003). Accurate unlexicalized parsing. *Proceedings of the 41st Meeting of the Association for Computational Linguistics*, pp. 423–430.

Lakoff, G. (1966). Stative adjectives and verbs in English. In A. G. Oettinger (Ed.), *Mathematical linguistics and automatic translation*. Cambridge: Harvard University. Report NSF 19, computation laboratory.

Levelt, W. J. M. (1993). *Speaking: From intention to articulation*. Cambridge: MIT Press.

Levy, R. P., & Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. In J. Platt & T. Hoffman (Eds.), *Advances in neural information processing systems* (pp. 849–856) Cambridge: MIT Press.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8,* 249–336.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-PLUS*. Berlin: Springer.

Rohde, D. L. (2005). Tgrep2 user manual. http://tedlab.mit.edu/~dr/Tgrep2/tgrep2.pdf.

Shannon, C. E. (1948). A mathematical theory of communications. *Bell Systems Technical Journal, 27,* 623–656.

van Draat, P. F. (1910). *Rhythm in English prose*. Heidelberg: Carl Winter's Universitätsbuchhandlung.

Warren, T., & Gibson, E. (2002). The influence of referential processing on sentence complexity. *Cognition, 85*(1), 79–112.

Wasow, T. (2002). *Postverbal behavior*. Stanford: CSLI Publications.

Wasow, T., Jaeger, T. F., & Orr, D. (2011). Lexical variation in relativizer frequency. In H. Simon & H. Wiese (Eds.), *Expecting the unexpected: Exceptions in grammar* (pp. 175–195). Berlin: De Gruyter.

Wasow, T., Greene, R., & Levy, R. (2012). Optional *to* and Prosody. Poster at the 25th annual CUNY Conference on Human Sentence Processing. New York, March 2012.

Zipf, G. (1936). *The Psychobiology of Language*. London: Routledge.

Part II
Implicit Prosody

The Roles of Phonology in Silent Reading: A Selective Review

Charles Clifton, Jr.

Abstract This chapter presents a selective review of evidence about how phonological representations are involved in silent reading. Knowledge of the mapping of orthography onto phonology appears to be important in skilled reading, and this knowledge is applied very early in the process of recognizing words in isolation. The same is true when one is reading sentences and texts, and the creation of a phonological representation of a text appears to play a critical role in guiding the movement of the eyes during reading. Phonological representations beyond the level of the individual word, including prosodic representations, also seem to play an important role in guiding parsing and in integrating discourse.

Keywords Implicit prosody · Prosody · Sentence processing · Reading · Syntactic ambiguity

1 The Roles of Phonology in Reading: A Selective Review

It is tempting to say that written language is speech made visible. This formulation is too simple. Written language loses important aspects of speech, but makes possible types of symbol manipulation that are not available in speech (see Elbow 2012, Chap. 1, for useful discussion). But the formulation has a large grain of truth. Beginning readers learn to map visual symbols onto sounds, gaining access to the form of language that they already know (see Rayner et al. 2012, Chaps. 10–11 for a review). Skilled readers rely on this learned mapping, perhaps obligatorily, perhaps as a fallback procedure when visual shortcuts fail (Frost 1998). Readers commonly experience something like heard speech when they read—the "inner voice" (Huey 1908/1968). This inner speech, or at least the mental representation that supports it, guides and aids comprehension of a written text (Slowiaczek and Clifton 1980; Fodor 2002a). Skilled writers know that good writing is writing that sounds good when spoken aloud (Chafe 1988; Elbow 2012). In this chapter, I review a selection of the evidence supporting some of these claims, emphasizing work done in the University

C. Clifton, Jr. (✉)
Department of Psychology, University of Massachusetts, 01003 Amherst, MA, USA

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_9

of Massachusetts laboratories, and work stimulated by Janet Fodor's analyses of the effects of implicit prosody. The review will largely be limited to work done on reading of English, although some of the basic results have been shown to hold true for languages whose mapping of orthography onto phonology is less transparent than for English, or more transparent (see Rayner et al. 2012, especially Chaps. 2, 3, and 5).

2 Phonology and the Recognition of Written Words

Many readers report hearing an "inner voice" even while reading silently. This convinced Huey (1908/1968)—and many others—that readers convert print into subjective speech as an essential component of the reading process. Others have been more cautious, claiming only that the conversion of print into a phonological representation (which could, but need not, be realized subjectively) may sometimes play a role. A great deal of experimental work has been directed at questions like, is such phonological recoding a necessary part of reading, and, does such recoding precede or follow the identification of the word being read?

Many clever demonstrations show that phonological recoding is ubiquitous. For instance, readers tend to identify a homophonic word like *meet* as referring to a member of the category of edible things (Van Orden 1987). Meyer et al. (1974) found that reading a word like *couch* slowed the decision that a following *touch* (with a different vowel pronunciation) is a real word. While both these demonstrations show that phonology can get in the way of identifying a word, they do not show that phonological recoding is a necessary part of identifying a word or even that phonological recoding precedes word recognition. A demonstration by Van Orden et al. (1988) does rule out the possibility that phonological recoding is purely post-lexical: *sute* is misclassified as an article of clothing, even though there is no lexical entry for *sute*. Still, there is no knockdown evidence that print is always converted to sound on the way to recognizing a word (but see Halderman et al. 2012, for a recent defense of the importance of phonology in identifying words).

This state of affairs has led many theorists to accept a position that says that there are two routes to identifying a word: a print-to-sound route and a direct visual route. The former is governed by a reader's knowledge of the relation between spelling and sound; the latter is based on familiarity with the written form of a particular word (see Frost 1998, for a useful discussion). Some versions of this position claim that a reader's spelling–sound knowledge takes the form of rules. Other versions claim that such apparent knowledge results from a generalization (perhaps analogical) of particular instances of how words sound and are spelled (Coltheart et al. 2001; Glushko 1979; Harm and Seidenberg 2004). Some theorists claim that word recognition involves a race between a print-to-sound route and a direct visual route, with the visual route being strong enough to typically win the race for familiar words. Others claim that the two routes interact, supporting each other (Carr and Pollatsek 1985). Still others claim that the phonological route is "initial and primary" (Lukatela and Turvey 1994). The resolution of these competing claims has

important theoretical and practical applications. For instance, should children be taught to "sound out" words (including novel words) or quickly recognize familiar words on the basis of their form (see Rayner et al. 2012, Chaps. 10 and 11, for extensive discussion).

While all these claims still have their defenders, some recent research, largely from the University of Massachusetts Amherst laboratories, highlights the importance of the phonological route. One line of research uses a technique of "display change" while measuring eye movements (Rayner 1975). A subject in an experiment reads a sentence in which one word is initially replaced by another "preview" word or a string of random letters until the eye moves into the area of the word, at which time the correct word replaces the preview. The preview word appears outside the eye's fovea, in the "parafovea," where words are generally not identified (Rayner 1998). The change takes place during the saccade from one word to the next, when the eye is functionally nearly blind. While readers do occasionally report seeing something happen when the display changes, they almost never report seeing the preview word. The basic finding is that if the actual word in a sentence is given as a parafoveal preview, it is identified faster once the eye lands on it than if a different word, or random letters, had been used as the preview.

This display change technique has been used to show an early involvement of phonological coding in reading. Pollatsek et al. (1992) showed that previewing a homophone of the actual word in the sentence resulted in faster reading of the target word than previewing a visually similar but non-homophonic word. For example, readers fixate a shorter period of time on the target word *beach* if the parafoveal preview had been *beech* than if it had been *bench*. If one accepts that most words are not fully identified before they are fixated (Rayner 1998), this shows that the sound of a word is (at least sometimes) identified before the word is accessed. Rayner (1998) has argued that parafoveal words are not fully identified. If a parafoveal word changes to a semantically related word when it is fixated, reading is not speeded up (e.g., previewing *pies* does not speed reading of *cake*). This is true even though one would prime the other if each were fixated. However, Schotter (2013) has provided evidence that challenges this claim: Parafoveal preview of one word does speed reading of a synonym. Thus, it is possible that the phonological representation that allows parafoveal preview of one word to speed reading a homophone was created as a result, not a precursor, of identifying the parafoveal word. Nonetheless, one can conclude that the preview of a phonologically similar word did speed identification of the actual word in the sentence, and thus that phonological information can be involved in rapid word identification.

A related technique, "fast priming" (Sereno and Rayner 1992) provides similar, and possibly more constraining, evidence that phonology is involved in word recognition. In this technique, preview of a target word in a sentence is blocked by replacing it parafoveally by a nonword, generally random letters or letter-like forms. When the eye moves into this nonword, it is replaced by either the target word or a related "prime" word for a very brief period of time (e.g., 36 ms), after which the target word is presented. The presentation time is too brief for readers to identify the word, and they generally do not report having seen it. Nonetheless, if the prime

word is a homophone of the target word, the target word is identified more quickly (fixated for a shorter period of time) than if the prime word had been a visually matched nonhomophonic control (Rayner et al. 1995). This result is very similar to the effect of a homophonic parafoveal preview, but makes it even less likely that the phonological representation that facilitated identification of the target word followed on, rather than preceding, identification of the prime.

The parafoveal preview technique has been used to go beyond demonstrating the importance of phonology and to explore the nature of the phonological representation created by the parafoveally presented prime word. This raises a question of interest in itself—how rich and complete is the representation? How far is it abstracted from the identity of the prime word?—but the results also turn out to provide evidence that the phonological representation is built up directly from the orthography of the visual prime rather than being a result of accessing the lexical representation of the prime. Ashby and Rayner (2004) asked whether the phonological representation contains not only information about the sequence of individual segments but also information about syllable structure. They presented nonword parafoveal primes that matched or mismatched the syllable structure of the target word, and measured the time to read the target word in a sentence. For instance, the target word *device* (whose syllable structure is de + vice) was preceded by either *de_ πxw* or *dev_πx*. The former, but not the latter, matched the syllable structure of the target word, and resulted in faster reading. Comparable results were obtained when the target word had a three-letter initial syllable (e.g., *balcony*): A three-letter parafoveal prime resulted in faster reading than a two-letter prime.[1] This result shows that the phonological representation resulting from the parafoveal prime—and by extension, required by the target word—contains suprasegmental information. It is not a minimal (Frost 1998) representation of the phonological segments that constitute the word. Other research using the parafoveal preview technique provides evidence that the representation contains additional information about the relations between phonological segments. Ashby et al. (2006) used the fact that, in English, the pronunciation of a vowel is affected by the following consonant. For instance, the vowel *a* is typically pronounced differently in the nonwords *rall* and *raff*. If a word like *rack* is parafoveally primed by a nonword sharing the pronunciation of its vowel (*raff*), it is read faster than if the prime had had a different vowel pronunciation (*rall*).

Similar results can be obtained using a different technique, measuring scalp electrical responses (event related potentials, ERPs) to single words presented after a brief masked prime. Ashby (2010) found that the N1 (a negative-going electrical potential, taking place 100–160 ms after the onset of a word) was reduced when a two-syllable word (e.g., *balcony*) was preceded by a very brief (44 ms) masked partial-word prime that matched in syllable structure (*bal*) compared to when the prime mismatched (*ba*; see Ashby and Martin 2008, for similar results, using both ERP and lexical decision reaction time; RT). Ashby et al. (2009) used ERP to provide

[1] Ashby (2006) found that this facilitation was limited to low-frequency words, suggesting that phonological processing may play a smaller role in recognizing very familiar words.

complementary evidence that the representation contains not only suprasegmental information but also information about the phonological features that constitute a segment. They showed a reduction in an early (80 ms) negative-going response if a target word was preceded by a brief nonword masked prime whose last letter shared the voicing feature of the target's last letter. For instance, if the target word *fad* was preceded by *faz,* there was a smaller early negativity than if it had been preceded by *fap*, with the opposite result if the target word had a voiceless final letter, e.g., *fat*.

All these results show that reading results in the creation of a phonological representation, and that the phonological representation is a rich one. It contains information about subsegmental phonological features, about the constraints that exist between adjacent segments, and about suprasegmental (syllabic) structure, in addition to information about the segmental makeup of a word. Further, the results make it very unlikely that the phonological representation is post-lexical. A rich phonological representation is constructed even when a nonword is presented, and it is created when the presentation of a word or a nonword is so brief that it is probably not identified, certainly not consciously. Finally, the fact that activating an appropriate phonological representation speeds the reading of a word, and reduces a scalp electrical event that signals the occurrence of a new word, leads one to believe that creating a phonological representation helps one identify the word.

But does the phonological representation play any role in processing language beyond identifying words? The remainder of this chapter reviews a selection of evidence that provides an affirmative answer and gives some suggestions about the roles that it plays (see also Breen, this volume, for an extensive analysis of the role of implicit prosody in sentence processing).

3 Identifying Words in Text

The emphasis of the research reviewed to this point is on how phonology affected the recognition of individual words. What follows shifts the emphasis to how the phonology of a word that is read in a text affects the reading of the text, especially how the eyes move through the text. In analyzing the relation between eye movements and text comprehension, I adopt the framework spelled out in the E–Z Reader model (Reichle et al. 1998, 2003, 2009). In this framework, the movement of the eyes quite directly reflects successful recognition of a word (and in Reichle et al. 2009), integration of the word into text. Thus, shorter fixations on a word, and fewer regressive eye movements from the word, provide evidence for faster identification and integration of the word into the representation of the text.

There are interesting old demonstrations that the sound of a spoken text has global effects on how the text is silently read. The following sentence looks like nonsense, but try listening to it when you "say it in your head," and it makes perfect sense: *The bouy and the none tolled hymn that they had scene and herd a pear of bear feat* (LaBerge 1972, acknowledging Jay Samuels). Experimental demonstrations involving "tongue twisters" make much the same point. Readers are slowed

when they silently read something like *Barbara burned the brown bread badly* (Haber and Haber 1982; McCutchen and Perfetti 1982; Warren and Morris 2009). The phonological similarity among the words seems to interfere with some aspect of reading them. These old demonstrations may rely on conscious subvocal articulation, but, even so, they do show that the phonological properties of words in text can influence silent reading.

Recent experimental research has shed some light on how one aspect of phonology, syllabic structure, influences silent reading, even in the likely absence of conscious subvocalization. One finding comes from Fitzsimmons and Drieghe's (2012) research on when words are skipped during silent reading. Previous results of experiments measuring eye movements during reading (e.g., Drieghe et al. 2005) have shown that short, high-frequency, predictable words are skipped fairly often. They are presumably identified parafoveally, eliminating the need to fixate them. Fitzsimmons and Drieghe examined the effect of number of syllables in a word on skipping frequency. Five-letter words were skipped less often when they contained two syllables (e.g., *cargo*) than when they contained a single syllable (e.g., *grain*), even though the words were carefully matched on word frequency and predictability within the sentence. Only information gathered from the parafovea (together with preceding context) can affect skipping, and since syllabic structure is a phonological property of the word that is skipped, it apparently is picked up parafoveally. Thus, this finding is consistent with previously reviewed evidence that readers extract phonological information parafoveally, and goes beyond it to show that this information affects where the eyes move as well as how long it takes to identify a word.

Interestingly, in the Fitzsimmons and Drieghe (2012) data, the number of syllables in a word did not affect how long it was fixated, just how often it was skipped. The fixation time result is consistent with an ancillary result reported by Ashby and Clifton (2005): While two-syllable words took longer to pronounce than monosyllables, they were not fixated any longer during silent reading. The primary result from the Ashby and Clifton research was different: Four-syllable words with two stressed syllables (e.g., *RA-di-A-tion*) were read more slowly than four-syllable words with a single stressed syllable (e.g., *ge-O-me-try*). The result appeared in gaze duration during silent reading (the sum of the duration of all fixations in a word from first fixating on it until first leaving it); it did not appear in the duration of the initial fixation on the word. Basically, two-syllable words often required a second fixation, increasing gaze duration.

While this pattern of results does implicate the importance of phonology in controlling eye movements during silent reading, it presents an explanatory challenge. Ashby and Clifton suggested that the effect of number of syllables on reading time might be related to a phenomenon observed by Sternberg et al. (1978). These researchers had people memorize lists of words that varied in number of syllables and number of stressed syllables. The latency to begin reciting the list in response to a "start" signal increased as a function of the number of stressed syllables, not the total number of syllables (although that latter increases the total time to recite the words). They suggested that the latency reflected the time to prepare production units, and

proposed that a phonological foot (a group of syllables headed by a stressed syllable) was the unit in terms of which production units are prepared. It is possible that during normal reading, the eye moves out of a word once such a production unit is prepared, rather than waiting for the unit to actually be produced. In fact, such production may not normally occur. The experience of inner speech (see the following section) may be triggered just by the intention to produce, and subvocal production may not be required. Why, then, are words with a single syllable skipped more often than words with two syllables (one stressed) in the Fitzsimmons and Drieghe (2012) data? All one can say with any certainty is that, while phonological factors do affect both skipping and reading time, they apparently affect them differently. It would be very interesting to know whether the number of stressed syllables in a parafoveal word affects skipping rate, over and above the sheer number of syllables. If not, one might be tempted to conclude that the decision to skip a word is based on phonological processing that is less complete than the processing required to make the decision to move on after fixating it.

While the results just reviewed were presented as evidence that the syllabic structure of a word affects how it is recognized, it is possible that some aspect of a word's phonology or spelling that is confounded with syllabic structure is actually responsible for the observed effects. For instance, transitional probability between segments might differ within and across syllable boundaries, or across the boundaries between stressed and unstressed syllables, and these differences in transitional probability could affect reading. However, these particular confounds do not apply to some other lines of research showing the importance of syllabic structure in reading text. One such line of research builds on demonstrations that listeners can expect a particular pattern of stressed syllables in spoken language, and that their expectations influence what they perceive (Dilley and McAuley 2008). Expected patterns of stressed syllables also affect how text is read. Breen and Clifton (2011) demonstrated such an effect by measuring eye movements, while people read limericks, judging whether they were "dirty" or not. A limerick, of course, has a very rigid rhythmic pattern. Some of the limericks Breen and Clifton had people read, honored this rhythmic pattern; in others, the last word of the second line had the wrong stress pattern. An example of the beginning of an acceptable limerick is:

> There once was a clever young gent
> Who had a nice talk to present….

Present needs, and has, stress on its second syllable. But consider the following unacceptable beginning:

> There once was a penniless peasant
> Who went to his master to present….

Here, the rhythmic context requires the last word of the second line to have stress on its first syllable. However, it has to be a verb, and the verb form of *present* has second syllable stress. Silent reading was disrupted on *present* in limericks like the second example above. This suggests that readers have expectations for the metrical structure of text, that a word is initially represented as fitting the metrical structure,

and that a critical part of reading the word is revising an inappropriate stress pattern so that it is correct for the word's actual lexical entry. The Breen and Clifton results can be taken as providing further evidence for Ashby and Clifton's (2005) speculation that the eyes move on in text when an appropriate production unit is prepared, and that production units are defined in terms of groups headed by a stressed syllable. More generally, they suggest an extension to the E–Z Reader model (Reichle et al. 1998) in which the eyes move on only when the reader has prepared a lexical representation that is phonologically veridical in the sense of being consistent with the required word's representation in the mental lexicon.

Evidence for a related effect, in which expectations for lexical stress patterns are based on syntactic and lexical rather than rhythmic information, comes from a second experiment that Breen and Clifton (2011) presented (see also, Breen and Clifton 2014). This experiment relied on the fact that some noun–verb pairs are identical (e.g., *report*), whereas some pairs are pronounced with different stress patterns (e.g., *ABstract* vs. *abSTRACT).* Breen and Clifton had people silently read words like these in sentence contexts that syntactically biased readers to take the critical word to be a noun, and then followed the critical word with text that allowed it to remain a noun versus that forced it to be taken to be a verb. Consider the following examples:

The brilliant report/abstract was accepted at the prestigious conference.
The brilliant report/abstract the best ideas from the things they read.

In both examples, it is critical that the word *brilliant* is overwhelmingly taken at first to be an adjective, not a noun. If it is taken to be an adjective, the critical word—*report* or *abstract*—must be a noun. This is consistent with the continuation of the first example, …*was accepted*…. However, in the second example, the material that follows the critical word, …*the best ideas*…, requires a preceding verb. The word *brilliant* must be taken to be a noun, resulting in a revision of the syntactic structure of the sentence, and the critical word, *report* or *abstract,* must be taken to be a verb. Breen and Clifton found that this revision did disrupt reading, as expected. However, crucially to the present discussion, disruption was greater when the stress pattern of the critical word had to be changed (from *ABstract* to *abSTRACT)* than when no such change was needed (*rePORT*). Interestingly, this added disruption showed up on the critical word itself, even though only the following word determined that it had to be a verb. However, the disruption was limited to when the following word (almost always a short-function word) was skipped, indicating that the disambiguating word had been identified, while the critical word was fixated.[2]

Once again, this result indicated that the phonology of a word—here, its stress pattern, unconfounded with any other lexical factors—is included in a reader's representation of a word. This supports the claim that a veridical representation of a word is required before the eyes move on in a text. Together, the results reviewed in this section support several conclusions. Phonology is deeply involved in silent

[2] Breen and Clifton 2014 in press, reported that when parafoveal preview of the disambiguating word was prevented, the disruption appeared later, after the disambiguating word was actually fixated.

reading. While it is not likely to require explicit subvocalization, the phonological representation is not impoverished. It includes not only information about syllabification of individual words but also information about their metrical structure. Further, phonology is involved in controlling the movement of the eyes, perhaps by specifying how a word is to be articulated even if the articulation does not take place. Phonology may have to be viewed as a critical part of the veridical representation of a word that a reader requires before moving ahead in a text.

4 The Roles of the "Inner Voice" in Reading

The previous section introduced one aspect of prosody, namely metrical structure, as part of the phonological representation that is created in silent reading. This final section turns to other aspects of prosody, including its rhythm, segmentation, melody, and rate. The section's title refers to the "inner voice," generally viewed as being the conscious experience of something like a heard voice, but inside one's head. The data to be reviewed do not actually require a commitment to such a subjective experience. As was the case in recognizing words, some of the effects may reflect processes that precede, or even occur independently of, the relevant experience. But because many of the effects do have counterparts in subjective experience, and because the subjective experience is so richly prosodic, I will continue to speak of the inner voice when convenient.

Some clever demonstrations buttress the intuition that we hear an inner voice while reading. Kosslyn and Matt (1977; see also Alexander and Nygaard 2008) had subjects listen to the supposed author of a written text, speaking aloud, before they read the text. If the author was a fast talker, subjects read the text more rapidly than if the author was a slow talker. Readers' inner voice apparently mimics the supposed author of what is being read. An auditory image of a particular speaker is not required, however. Yao and Scheepers (2011; see also Scheepers, this volume) had people silently read a passage that implied that a protagonist was speaking rapidly (e.g., was anxious or upset) or slowly (e.g., was lethargic or ill), and ended with a direct quotation of what the protagonist said. This direct quotation was read more quickly in the rapid-speaking than the slow-speaking context, suggesting again that the inner voice mirrored the presumed actual voice of the source of the written material.

These demonstrations, while very interesting, do not tell us much about the possible function of the inner voice. That very interesting question was brought to psycholinguists' attention by a series of papers by Janet Fodor (especially Fodor 1998, 2002a, 2002b; see Breen, this volume, for more extensive discussion). As part of a broader analysis of how prosody affects language comprehension, Fodor advanced the implicit prosody hypothesis (Fodor 2002b):

> Implicit Prosody Hypothesis (IPH): In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour of the construction.

Earlier work did suggest that implicit prosody might play an important role in language comprehension. For instance, Slowiaczek and Clifton (1980) had people silently read passages, while engaging in activity designed to block subvocalization and presumably, inner speech (rapidly counting 1–10 or saying *cola* repeatedly). Blocking subvocalization did not impair comprehension of the propositions contained in the passages, as indexed by accuracy of recognizing clauses from the passage with content words replaced by synonyms. However, compared to conditions without the subvocalization block (or compared to conditions in which subjects heard rather than read the passages), comprehension that required inferences or integration across multiple clauses was impaired. These data do suggest that the inability to "hear the inner voice" impaired the processes involved in creating a high-level representation of the passage. Slowiaczek and Clifton speculated that these processes crucially involve the passage's prosody, which could enable readers to maintain a representation of relations such as subordination and relative emphasis among the passage's parts.

In contrast to Slowiaczek and Clifton, Fodor (1998, 2002a) presented data that rather directly implicate the role of prosody in silent reading. Much of her evidence involved the placement of prosodic, and by inference syntactic, boundaries. In 1998, she advanced the "same-size-sister" hypothesis. According to this hypothesis, a sentence is ideally divided into prosodic phrases of equal size, and this division can affect the syntactic analysis of a sentence. One structure Fodor analyzed was the much-studied relative clause attachment ambiguity (Cuetos and Mitchell 1988), in which the relative clause *who was on the porch* in *Everybody liked the daughter of the colonel who was on the porch* can be taken to modify either *daughter* or *colonel*. Fodor provided evidence from several languages that the relative length of the phrases (and in fact, the language's preferred patterns of prosodic phrasing) affects how this ambiguity is resolved. A short relative clause tends to be taken to modify a short head phrase, and a long one to modify a long phrase. Similar-length phrases are taken to be syntactic sisters of each other. There is likely a prosodic basis for this. A long relative clause is generally pronounced as a separate prosodic phrase, and, if so, it would ideally be conjoined to a prosodic phrase of similar length.[3]

One example of this kind of evidence comes from unpublished research conducted by B. Hemforth and colleagues (for a preliminary report, see Walter et al. 1999). They showed (studying English, German, French, and Spanish) that increasing the length of a relative clause in a sentence like *The son of the colonel who died wrote five books on tropical disease* from *who died* to *who tragically died of a stroke* increased the number of times it was reported as modifying the longer phrase *the son of the colonel* as compared to modifying the short phrase *the colonel*. *The colonel* is similar in length to *who died*; *The son of the colonel* is similar in length to *who tragically died of a stroke*.[4]

[3] See Jun (2010) for evidence that overt prosody may not have the properties assumed by Fodor's implicit prosody hypothesis (IPH). In particular, English readers tended to place a prosodic boundary immediately before a relative clause (RC) in an NP–NP–RC configuration. However, as Jun notes, implicit prosody may well be different. See also Jun (this volume).

[4] To be sure, factors other than length affect the resolution of the relative clause ambiguity, most saliently, whether the modified noun phrase (NP) is subject or object of a sentence. In the Hem-

Luo et al. (2012) manipulated prosodic phrasing in a very different way, again affecting the resolution of a syntactic ambiguity. They measured eye movements, while people silently read Chinese sentences that contained an ambiguity between a conjunction of two noun phrases (NPs) and the introduction of a prepositional phrase. This ambiguity appears to cause difficulty in comprehension, but can be eliminated in spoken language by introducing a prosodic break at different points. In the experiment, the visual presentation of the sentence was preceded by an unintelligible, low-pass filtered, spoken version of the sentence that did or did not contain a disambiguating phonological phrase boundary. Luo et al. found that the visual sentence was read more rapidly following a speech melody that contained a disambiguating boundary than one that did not. The authors interpreted their finding as indicating that readers imposed the prosody of the speech melody onto the sentence as it was silently read, assuming that a syntactically ambiguous sentence is read more slowly than one that is disambiguated (by the prosodic phrasing of the inner voice).[5]

Aspects of implicit prosody besides prosodic phrasing have been shown to affect silent reading. An important paper by Bader (1998) showed that whether a word did or did not receive a pitch accent affected the syntactic analysis of a sentence (see Bader, this volume, for descriptions of new experiments that show online effects of preferred stress and accent patterns). Breen (this volume) provides a more detailed description of Bader's (1998) experiment, but briefly, German readers of a sentence including *...dass man (sogar) ihr Geld...* were induced (by the occurrence of the focusing word *sogar)* to place an implicit accent on the head of the phrase *ihr GELD* (meaning, by default, *her gold*, with *ihr* playing the role of possessive pronoun). If the following material forced a reanalysis in which *ihr* was taken to be a referring pronoun, the dative object of a higher verb, it would have to be prosodically reanalysed to receive the accent. This prosodic reanalysis slowed reading time compared to a condition without *sogar*. The result is in some ways parallel to Breen and Clifton's (2011) prosodic reanalysis finding, discussed earlier. The difference is that while Breen and Clifton's research involved the reassignment of lexical stress, the Bader results involves the reassignment of a focus-expressing pitch accent. Similarly, Kentner (2012; described in Breen, this volume) has provided evidence indicating that readers avoided having two implicitly stressed syllables adjacent to each other by changing whether or not a function word was implicitly accented. The presumed accent on the function word, in turn, affected how it was interpreted.

I will conclude with a brief description of some recent findings that may (or may not) reflect the operation of implicit prosody. Benatar and Clifton (2014) asked

forth et al. research, the often-discussed differences among languages were rather minor, and could be attributed to factors such as how the different languages encode information status.

[5] This latter assumption is apparently inconsistent with Clifton and Staub's (2008) review of the syntactic processing literature that concluded that syntactic ambiguity has no effect on reading time, or even speeds it. It is possible that prosodic disambiguation has a different effect than contextual or morpho-syntactic disambiguation, or that the demand characteristics of the present experiment induced readers to attempt to use the prosody of the preceding speech melody to resolve the ambiguity of the target sentence.

whether the information status of a word would affect how rapidly it was silently read. They measured eye movements, while subjects read short dialogs containing words that were given versus new in the discourse context (using Schwarzschild's 1999, analysis of givenness). For example, consider sentence (a):

a. Kyle cares about Natalie but he doesn't show it.

The word *Natalie* is given in (a) when that sentence follows (b), but not when it follows (c):

b. I'm confused, does Kyle care about Natalie?
c. Natalie is confused, does Kyle care about someone?

The target word *Natalie* in (a) was read more slowly when it was non-given (following c) than when it was given (following b). This effect was even larger when the non-given term had "corrective focus," e.g., the word *John* in *Did you tell Mary to go home early? I told John but I don't know if it was a good idea.*

Benatar and Clifton interpreted these results as indicating that when new information must be added to a mental representation of a discourse, or especially when old information has to be corrected, reading is slowed. But an alternate, or additional, possible interpretation emphasizes prosody. A non-given word in spoken English must receive a pitch accent. Perhaps words that receive a pitch accent in implicit prosody receive additional attention, which in addition to speeding decisions about the form of the word (Cutler 1976; Cutler and Fodor 1979), may increase the time during which it is attended. This suggestion is superficially at odds with the common assumption (e.g., Reichle et al. 2009) that the eyes move on past a word as soon as its identification is imminent. But a variety of data show that eye movements in reading reflect more than just the identification of words (e.g., Clifton et al. 2007; see Reichle et al. 2009, for an extension of the E–Z Reader model designed to deal with this observation), and the possibility that implicit prosody directly impacts eye movements is well worth exploring.

5 Conclusions and Prospects

It is abundantly clear that the phonology of words plays a role in skilled reading. It is also abundantly clear that we are not able to say that each and every instance of identifying a written word requires constructing or accessing its phonology: It is impossible to demonstrate that word identification always requires phonological processing. What it is possible to do is to sort out just what roles phonology plays in recognizing words. A great deal of progress has been made along these lines. We now know that not only phonological segments but also phonological features, larger phonological units such as rhymes and syllables, and metrical structure of words contribute to word recognition. We also know that at least some of these aspects of phonology are identified in the parafoveal, before a reader fixates on a word, and that they are identified very quickly. What we do not know enough about

is the conditions affecting how major a role these various aspects of phonology play in identifying words. For instance, although we know that phonology plays some role not only in languages with transparent (Spanish) or messy (English) or mappings of orthography onto phonology but also in languages widely viewed as logographic (Chinese; Pollatsek et al. 2000), we know little about whether and how these languages differ in their use of phonology in reading. We do not know enough about how visual and phonological constraints are coordinated in the process of recognizing a word. We do not know enough about how the brain does whatever it does (but see Price and Devlin 2011, for some interesting speculations about the interaction of phonological expectations and visual processing).

Beyond the word, we know even less. Again, we do not, and really cannot, know whether inner speech or some other extended phonological representation accompanies and guides each and every act of silent reading. We cannot confidently say that the subjective experience of hearing an inner voice is universal, but that is only because an individual's report of a subjective experience (or its absence) is privileged—nobody else has access to such an experience. But there are shreds of intriguing evidence that a phonological representation that extends beyond the individual word does play a role in guiding and facilitating language comprehension. Excessive phonological similarity confuses a reader, interference with subvocalization disrupts global comprehension of a text, the rhythm of a sentence affects how words are identified, the preferred division of a sentence into prosodic units and the preferred placement of stress and pitch accents affect its interpretation, etc. One expects that much of interest remains to be learned. Do languages with different metrical properties exhibit different effects of implicit prosody? Are differences in writing genre, or differences in reading goals, associated with different inner speech phenomena? Can differences in implicit prosody be distinguished from differences in how a text updates a reader's understanding of what the text conveys? How is inner speech reflected in brain activity, and can measures of brain activity help pin down the time course and even logical sequence of its effects? In the acquisition of reading skill, how is oral reading fluency related to silent reading ability?

References

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 446–459. doi:10.1037/0096-1523.34.2.446.

Ashby, J. (2006). Prosody in skilled silent reading: Evidence from eye movements. *Journal of Research in Reading, 29* (3), 318–333.

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychonomic Bulletin & Review, 17,* 95–100.

Ashby, J., & Clifton, C., Jr. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96*(3), B89–100. doi:10.1016/j.cognition.2004.12.006.

Ashby, J., & Martin, A. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance, 34,* 224–236.

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes, 19,* 391–426.

Ashby, J., Treiman, R., Kessler, B. T., & Rayner, K. (2006). Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory and Cognition, 32,* 416–424.

Ashby, J., Sanders, L. D., & Kingston, J. (2009). Skilled readers begin processing phonological features by 80 ms: Evidence from ERPs. *Biological Psychology, 80,* 84–94.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht: Kluwer.

Benatar, A., & Clifton, C., Jr. (2014). Newness, givenness and discourse updating: Evidence from eye movements. *Journal of Memory and Language, 71*(1), 1–16. doi:10.1016/j.jml.2013.10.003.

Breen, M., & Clifton, C., Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language, 64*(2), 153–170. doi:10.1016/j.jml.2010.11.001.

Breen, M., & Clifton, C., Jr. (2014). Stress matters revisited: A display change experiment. *Quarterly Journal of Experimental Psychology*. doi:10.1080/17470218.2013.766899.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In T. G. Waller, D. Besner, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (Vol. 5, pp 2–82). New York: Academic.

Chafe, W. (1988). Punctuation and the prosody of written language. *Written Communication, 5,* 396–426.

Clifton, C., Jr., & Staub, A. (2008). Parallelism and competition in syntactic ambiguity resolution. *Language and Linguistics Compass, 2,* 234–250.

Clifton, C., Jr., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. V. Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movement research: Insights into mind and brain* (pp. 341–371). New York: Elsevier.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204–256.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30,* 73–105.

Cutler, A. (1976). Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20,* 55–60.

Cutler, A., & Fodor, J. A. (1979). Semantic focus and sentence comprehension. *Cognition, 7,* 49–59.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59,* 294–311.

Drieghe, D., Rayner, K., & Pollatsek, A. (2005). Eye movements and word skipping during reading revisited. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 954–969.

Elbow, P. (2012). *Vernacular eloquence*. Oxford: Oxford University Press.

Fitzsimmons, G., & Drieghe, D. (2012). How fast can predictability influence word skipping during reading? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. doi:10.1037/a0030909.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D. (2002a). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East Linguistics Society* (Vol. 32, pp. 112–132). Amherst: GSLA.

Fodor, J. D. (2002b). *Psycholinguistics cannot escape prosody*. Paper presented at Speech Prosody 2002, Aix-en-Provence, France.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: Rue issues and false trails. *Psychological Bulletin, 123*(1), 71–99.

Glushko, R. J. (1979). The organization and activation of orthographic information in reading. *Journal of experimental Psychology: Human Perception and Performance, 5,* 674–691.

Haber, R. N., & Haber, L. R. (1982). Does silent reading involve articulation? Evidence from tongue-twisters. *American Journal of Psychology, 95,* 409–419.

Halderman, L. K., Ashby, J., & Perfetti, C. A. (2012). Phonology: An early and integral role in identifying words. In J. S. Adelman (Ed.), *Visual word recognition: Vol. 1. Models and methods, orthography, and phonology* (Vol. 1, pp. 207–228). London: Psychology Press.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720.

Huey, E. B. (1908/1968). *The psychology and pedagogy of reading*. Cambridge: MIT Press.

Jun, S.-A. (2010). The implicit prosody hypothesis and overt prosody in English. *Language and Cognitive Processes, 25*(7–9), 1201–1233.

Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition, 123,* 1–20.

Kosslyn, S. M., & Matt, A. M. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society, 9*(4), 250–252.

LaBerge, D. (1972). Beyond auditory coding. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and by eye* (pp. 241–248). Cambridge: MIT Press.

Luo, Y., Yan, M., & Zhou, X. (2012). Prosodic boundaries delay the processing of upcoming lexical information during silent sentence reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 915–930. doi:10.1037/a0029182.

Lukatela, G., & Turvey, M. T. (1994). Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 107–127.

McCutchen, D., & Perfetti, C. (1982). The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior, 21,* 672–687.

Meyer, D. E., Schvaneveldt, R. W., & Ruddy, M. G. (1974). Functions of graphemic and phonemic codes in visual word-recognition. *Memory & Cognition, 2*(2), 309–321. doi:10.3758/BF03209002.

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 148–162.

Pollatsek, A., Tan, L. H., & Rayner, K. (2000). The role of phonological codes in integrating information across saccadic eye movements in Chinese character identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 607–633. doi:10.1037/0096-1523.26.2.607.

Price, C. J., & Devlin, J. T. (2011). The interactive account of ventral occipitotemporal contributions to reading. *Trends in Cognitive Sciences, 15*(6), 246–253. doi:10.1016/j.tics.2011.04.001.

Rayner, K. (1975). Parafoveal identification during a fixation in reading. *Acta Psychologica, 39,* 172–282.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422.

Rayner, K., Sereno, S. C., Lesch, M. F., & Pollatsek, A. (1995). Phonological codes are automatically activated during reading: Evidence from an eye movement paradigm. *Psychological Science, 6,* 26–32.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C., Jr. (2012). *Psychology of reading* (2nd ed.). New York: Psychology Press.

Reichle, E. D., Pollatsek, A., Fisher, D. F., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–156.

Reichle, E., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26,* 445–526.

Reichle, E., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher-level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16,* 1–21.

Sereno, S. C., & Rayner, K. (1992). Fast priming during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 173–184.

Schotter, E. R. (2013). *Constraints on semantic preview benefit*. Poster presented at the Psychonomic Society, Toronto, Canada.

Schwarzschild, R. (1999). Givenness, avoid and other constraints on the placement of accent. *Natural Language Semantics, 7,* 141–177.

Slowiaczek, M. L., & Clifton, C., Jr. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning & Verbal Behavior, 19*(5), 573–582.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparison of speech and typerwriting. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). San Diego: Academic.

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition, 15,* 181–198.

Van Orden, G. C., Johnston, J. C., & Hale, B. L. (1988). Word identification in reading proceeds from spelling to sound to meaning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 371–386.

Walter, M., Clifton, C., Jr., Frazier, L., Hemforth, B., Konieczny, L., & Seelig, H. (1999). *Prosodic and syntactic effects on relative clause attachment in German and English*. Poster presented at AMLaP 1999, Edinburgh.

Warren, S., & Morris, R. K. (2009). *Phonological similarity effects in reading*. Paper presented at the European Conference on Eye Movements, Southampton, England.

Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition, 121*(3), 447–453. doi:10.1016/j.cognition.2011.08.007.

Empirical Investigations of Implicit Prosody

Mara Breen

Abstract Fodor's introduction of the implicit prosody hypothesis (IPH; 2002) inspired a series of studies exploring how readers' "inner voice" influences sentence comprehension. In this chapter, I review the history of the IPH and a variety of studies which have demonstrated that implicit phrasing, accentuation, and rhythm appear to play a role in syntactic parsing. I explore how work moving forward might address the question of the psychological reality of the "inner voice," and how we can investigate the relative contribution of implicit prosody to sentence processing in consideration of other known information sources.

Keywords Implicit prosody · Prosody · Sentence processing · Reading · Syntactic ambiguity

When sentence processing research was in its infancy, researchers focused primarily on reading. The factors they considered to be important to parsing were those language features which were discernable on the written page. These factors initially included lexical and syntactic information (Frazier 1979; Frazier and Rayner 1989), though researchers soon began to consider the role of semantic information, like discourse context (Tanenhaus et al. 1995) and argument structure (Trueswell et al. 1994) in online sentence processing.

As the field progressed, researchers examined the specific contribution that characteristics of spoken language make to ambiguity resolution, focusing primarily on the role of prosody in lexical and syntactic ambiguity resolution (Cutler et al. 1997; Wagner and Watson 2010). Most recently, researchers have begun to explore the overlap between these two information sources: written text and spoken language. That is, the field has begun to investigate to what extent sound representations are activated during silent reading. Janet Fodor, one of the pioneers in the field of sentence processing, termed this phenomenon *implicit prosody.*

Fodor's idea is certainly not a new one. Scholars have been curious about the role of the "inner voice" in silent reading for more than 100 years (Huey 1908/1968;

M. Breen (✉)
Department of Psychology and Education, Mount Holyoke College, 01075 South Hadley, MA, USA
e-mail: mbreen@mtholyoke.edu

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_10

Chafe 1988). Indeed, one of the earliest experimental explorations of the inner voice was that of Slowiaczek and Clifton (1980) who had participants repeat "colacola-cola" while reading or listening to short discourses. Readers' high-level interpretation of discourses was impaired more than listeners', which Slowiaczek and Clifton interpreted as indicating that the suppression of subvocalization interfered with the reader's generation of an inner voice (see Clifton, this volume). However, Fodor (1998) was one of the first to bring the idea of the inner voice to the forefront of sentence processing research.

In her 1998 paper, Fodor recounts her interest in the role of implicit prosody in sentence processing as arising from a challenge to late closure (Frazier 1979). Late closure is a sentence processing heuristic which maintains that when readers encounter a new constituent to be added into the current syntactic parse, they will first pursue the parse in which the new constituent attaches to already built structure. Evidence for late closure comes from the finding that readers encounter difficulty at *fell* in (1) (from Frazier and Rayner 1982). When readers encounter *the sock*, they are more likely to interpret it as the object of the verb in the current verb phrase (*mending the sock*) than as the subject of the main clause. The former interpretation does not require the reader to build additional structure while the latter does, which leads to difficulty for the reader, who has to reanalyze and build the missing structure.

(1) While Mary was mending the sock fell off her lap.

Late closure was assumed to be a universal parsing constraint until the discovery that Spanish readers (among others) favor high attachment (Cuetos and Mitchell 1988). This discovery left researchers with a puzzle: How to reconcile this apparent case of language-specific parsing with the desire to identify a set of universal parsing constraints, to put it on par with other aspects of universal grammar.

Cuetos and Mitchell's (1988) solution was the *tuning hypothesis,* which holds that language users are sensitive to the frequency of syntactic constructions in their language, and, given that sensitivity, will make parsing decisions that are in line with the majority of those in their language. That is, Spanish readers, encountering more ambiguous attachments that are resolved in favor of high attachment than low attachment, use that information to make on line parsing decisions.

Subsequent research has demonstrated that Spanish readers are not the only group who prefer high attachment. Other languages which appear to have a high attachment preference, for at least some types of syntactic constructions, include French (Frenck-Mestre and Pynte 2000), Dutch (Brysbaert and Mitchell 1996), German (Hemforth et al. 1994), and Japanese (Kamide and Mitchell 1997). Some of these researchers adopted Cuetos and Mitchell's (1998) tuning hypothesis to account for these crosslinguistic differences, while others proposed alternate mechanisms which rely on two separate parsing principles, depending on the type of constituent in question. For example, Gibson et al. (1996) proposed that parsing preferences are dictated by *both* recency (akin to late closure) and predicate proximity, which maintains that the parser prefers to attach incoming constituents close to the verb. Hemforth et al. (2000) argued that, in addition to something like late closure, parsing is also subject to a constraint whereby anaphors prefer to attach to their antecedents,

which accounts for certain cases of high attachment. Finally, Frazier and Clifton (1996) argued that attachment decisions can, under some circumstances, be influenced by non-syntactic factors, which can promote high attachment.

In contrast to the proposals above, Fodor suggested that prosodic packaging may offer a more parsimonious way to account for the general low attachment preference across languages that could also account for those languages which demonstrate a high attachment preference. In this way, principles governing the packaging of phrases could have implications for syntactic phrasing. She proposed that, before syntactic parsing takes place, a prosodic packager shuttles through the sentence, dividing it up into phrases which will then be fed to the syntactic parser. In her words:

> I assume that in silent reading a prosodic contour is imposed on the input string, and that the syntactic parser is sensitive to the prosodic phrase boundaries even though they were fabricated by the perceptual system itself (p. 303).

Under Fodor's view, this prosodic packager operates with the goal of dividing the sentence into roughly equal parts, a feature that she terms the *same-size sister constraint*. This packaging can lead to different attachment preferences for similar ambiguous constituents, which vary minimally in length. For example, in (2) the reader must determine who is divorced: The bishop or the daughter.

(2) a. The divorced bishop's daughter
b. The recently divorced bishop's daughter
c. The recently divorced bishop's daughter-in-law

Fodor argues that readers should have no strong intuition in (2a) because there is no need for the constituent to be divided into multiple phrases. In (2b), in contrast, Fodor argues that readers will divide the phrase into the same-sized phrases *recently divorced* and *bishop's daughter*. In this way, they would produce an implicit phrase boundary between *divorced* and *bishop,* blocking an attachment between *divorced* and *bishop*, thereby leading to an interpretation that it is the daughter who is divorced. Finally, in (2c) the interpretation should shift with the reader dividing the constituent into *the recently divorced bishop's* and *daughter-in-law.* Now, Fodor argues, readers can attach *divorced* and *bishop* within the phrase and assume that the bishop is divorced, not the daughter-in-law.

Fodor's same-size sister constraint is based, in part, on proposals about how readers make phrasing decisions in overt production. For example, Gee et al. (1979) argued that the probability of a break at a location was determined by two factors: pressure to group syntactic dependents and pressure to balance the length of phrases. Gee and Grosjean (1983) added an additional constraint to the former model, that boundaries could not be produced within phonological phrases. However, in contrast to these models, Fodor (1998) argued that, rather than being the main determiner of phrasing, balancing operates in cases where the syntax leaves the option of phrasing open. Fodor's same-size sister constraint has received some empirical support, both in German (Augurzky 2008) and Japanese (Hirose 2003).

Fodor (2002) formalized the framework of which the same-size sister constraint was only a part when she proposed the *implicit prosody hypothesis:*

> The Implicit Prosody Hypothesis (IPH): In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction.

Moreover, she laid out a procedure that sentence processing researchers could follow to investigate implicit prosody:

1. Find a factor F which can be manipulated in an experiment, and which measurably affects the OVERT prosody of a sentence.
2. Show that the overt prosodic difference caused by F measurably influences an ambiguity resolution preference in parsing.
3. Show (or claim?) that F does not affect parsing DIRECTLY.
4. Include F in a silent reading task. Is ambiguity resolution affected by F as it is the listening task?

In the following sections, I will briefly review recent empirical work[1] which, inspired by Fodor (1998) and Bader (1998; see below), has demonstrated effects of implicit prosody on comprehension across four types of prosodic phenomena: phrasing, stress and accent, rhythm, and intonation.

1 Implicit Phrasing

Fodor's exhortation to sentence processing researchers in 1998 was limited to discovering how implicit prosodic factors might serve to rescue universal parsing principles. In this way, she focused on the aspects of implicit prosody that would likely have implications for syntactic parsing, namely prosodic phrasing. Indeed, there is now considerable evidence that implicit phrasing can influence parsing decisions, from a wide variety of languages, including Japanese (Kitagawa and Fodor 2006), English (Quinn et al. 2000; Swets et al. 2007), German (Augurzky 2006), Croatian (Lovric 2003), Hindi (Vasishth et al. 2005), Dutch (Wijnen 2004), French (Pynte and Colonna 2000), and Korean (Hwang and Schafer 2009).

Another set of studies inspired by Fodor (1998) have explored to what extent prosodic length influences attachment decisions. For example, Hirose (2003) and Hwang and Schafer (2009) have demonstrated, in Japanese and Korean, respectively, that readers' interpretations of ambiguous phrases are affected by length manipulations. Specifically, readers were less likely to pursue local attachments to longer noun phrases than shorter ones, presumably due to their need to place an implicit phrase boundary after the long noun phrase.

Others, however, have tried, and failed, to find the effects of length on implicit phrasing decisions. For example, Foltz et al. (2011) conducted two experiments to investigate whether individual speakers' overt phrasing of globally ambiguous sentences patterns with their interpretation of those same sentences. Consistent

[1] see also Breen (2014) for a more detailed review.

with prior work, the researchers found length effects on overt productions such that speakers produced larger prosodic boundaries after a noun which preceded a long relative clause (RC *bridegroom* in 3b) than a noun which preceded a short RC (*bridegroom* in 3a). However, these length effects did not appear to influence these same speakers' interpretations of the ambiguity. That is, the production data would predict that participants should be more likely to attach the relative clause *who swims like a fish* to *brother* in (3b) than (3a) because, in (3b), an implicit phrase boundary between *bridegroom* and *who* should block the attachment. However, Foltz et al. observed no significant effect of length on participants' attachment preferences.

(3) a. The brother of the bridegroom who swims was last seen on Friday night.
b. The brother of the bridegroom who swims like a fish was last seen on Friday night.

Despite the lack of an effect in their first experiment, these researchers observed effects of implicit phrasing on attachment decisions when they required speakers to indicate their interpretation of the attachment of the relative clause *prior* to producing the sentence aloud. Foltz et al. (2011) argue that this result indicates different phrasing strategies for familiar and unfamiliar sentences.

The inconsistent results of off line investigations of implicit phrasing have been partially addressed by explorations of on line effects of length on implicit phrasing using event-related potentials (ERP). These studies have generally utilized the closure positive shift (CPS), an ERP component observed in response to both overt and implicit phrase intonational phrase boundaries (Steinhauer 2003). Steinhauer and Friederici (2001) had participants read locally ambiguous sentences which could be disambiguated by the presence of a prosodic boundary. Before reading, participants listened to a filtered version of an overt production of the target sentence, which maintained the prosodic contour but not the lexical material, and were directed to apply the prosodic contour to the subsequent read sentence. The results revealed a CPS in the locations where the prosodic contour would have induced readers to postulate an implicit prosodic boundary during silent reading, suggesting that the CPS was reflecting implicit prosody.

Most recently, Hwang and Steinhauer (2011) had participants read locally ambiguous Korean sentences which are effectively disambiguated by the presence of an overt phrase boundary after an initial noun phrase (NP). The authors varied the length of the initial NP, and observed a CPS only after the long sentence-initial NP. Furthermore, subsequent explicit disambiguation to the less preferred interpretation of the sentence elicited a smaller ERP marker of syntactic difficulty (the P600) when the NP was long. The authors argue that, when the initial NP was long, readers imposed an implicit prosodic boundary, which served to ameliorate subsequent garden path effects (see Liu et al. 2010 for similar results from Chinese).

2 Implicit Stress and Accent

Concurrent with crosslinguistic work on attachment, researchers have also explored what aspects of prosody *apart* from phrasing might be driving sentence comprehension. For example, at the same time that Fodor was arguing for implicit phrasing, Bader (1998; this volume) argued that implicit prosodic focus influences sentence processing. He proposed the *prosodic constraint on reanalysis* (*PCR*), which holds that any required syntactic reanalysis will be more difficult if it requires a concurrent prosodic reanalysis. Bader's initial evidence for the PCR came from an eye-tracking study, in which German participants read sentences like those in (4):

(4) *Zu mir hat Maria gesagt,*
to me Maria has said
"Maria said to me"
a. *…daß man (sogar) ihr Geld anvertraut hat.*
…that one (even) her money entrusted has
"…that someone entrusted money (even) to her."
b. *…daß man (sogar) ihr Geld beschlagnahmt hat.*
…that one (even) her money confiscated has
"…that someone confiscated (even) her money."

In (4a), *ihr* functions as an object pronoun, as the indirect object of *entrusted*. In (4b), *ihr* functions as a possessive pronoun, in that it specifies whom the money belongs to. Bader argues that this latter interpretation of *ihr* is the preferred one and that the default phrasing of both interpretations, without the inclusion of *sogar*, is the same with stress on *man* but not on *ihr*. Critically, the phrasing of the two sentences changes with the addition of the focus particle *sogar*. Bader argues that the readers first interpret *ihr* in (4a) as a possessive pronoun. As function words are usually not accented, the reader would leave *ihr* unaccented and place the nuclear accent on *Geld*. However, upon encountering *anvertraut,* the reader would have to reanalyze both syntactically and prosodically, first reinterpretating *ihr* as an object pronoun, and second, shifting the accent from *Geld* to *ihr*. That is, *sogar* serves to direct semantic focus to *ihr*, and as such it (*ihr*) requires a focal accent. Indeed, Bader observed that reanalysis in (4a) was more difficult with the presence of the focus particle than without as evidenced by longer reading times on the disambiguating region comprised of the final two words, which he argued was due to the additional cost of updating the implicit prosodic representation (cf. Bader, this volume).

Two more recent studies provide further support for Bader's claim that readers represent prosodic focus during silent reading. Stolterfoht et al. (2007) performed an ERP study in which participants read sentences which required either a focus structural revision (from wide as in (5a) to narrow as in (5b)) or both a focus structural and prosodic revision (as in 5c). The required revisions are evident by a comparison of (5b) and (5c) to (5a), which exemplifies the default focus structure assignment. Upon encountering *den Lehrer* in (5b), the reader realizes that she must revise her focus representation, as the object replacive structure indicates a narrow object focus interpretation rather than a wide focus interpretation. In this case, there is no need to shift the implicit nuclear accent, because it would already be assigned to *den Schuler,* where it could project focus to the entire sentence. Conversely, upon

encountering *der Lehrer* in (5c), the reader must not only revise her focus structure (from wide to narrow subject focus) but also revise the placement of an implicit pitch accent, moving it from *den Schuler* to *der Direktor.*

(5) a. [Am Dienstag hat der Direktor den SCHÜler getadelt]$_F$
On Tuesday has the principal$_{nom}$ the pupil$_{acc}$ criticized
"On Tuesday, the principal criticized the pupil."

b. Focus structural revision only
Am Dienstag hat der Direktor [den SCHÜler]$_F$ getadelt, und nicht [den LEHrer]$_F$
On Tuesday has the principal$_{nom}$ the pupil$_{acc}$ criticized, and not the teacher$_{acc}$
"On Tuesday, the principal criticized the pupil, and the principal did not criticize the teacher."

c. Focus structural + prosodic revision
Am Dienstag hat [der DiREKtor]$_F$ den Schüler getadelt, und nicht [der LEHrer]$_F$
On Tuesday has the principal$_{nom}$ the pupil$_{acc}$ criticized, and not the teacher$_{nom}$
"On Tuesday, the principal criticized the pupil, and the teacher did not criticize the pupil."

Indeed, compared to sentences that required no focus structural revision, where the focus particle *nur* (*only*) preceded *den Schuler*, the final noun in sentences like (5c) (*den Lehrer*) elicited a late positivity. However, compared to sentences that required no focus structural revision, the final noun in sentences like (5c) (*der Lehrer*) elicited an earlier negativity, which the authors interpreted as evidence of implicit prosodic reanalysis (cf. Stolterfoht and Bader 2004). Crucially, in addition, Kitagawa et al. (2013) demonstrated higher acceptability judgments for Japanese sentences where an explicit pitch accent disambiguated focus location than for sentences without the pitch accent where, presumably, readers were silently assigning a default (implicit) nuclear pitch accent to a non-focused element.

These results suggest that readers are generating implicit nuclear accents during silent reading. A related set of studies has explored to what extent lexical stress is also a feature of silent reading. Clifton (this volume) details the specifics of these studies, which include Ashby and Clifton's (2005) demonstration of longer reading times for four-syllable words with two stressed syllables than for four-syllable words with only one stressed syllable. In addition, Breen and Clifton (2011, 2013) demonstrated that syntactic reanalysis of a stress-shifting noun–verb homograph like *permit* (*PERmit* as a noun; *perMIT* as a verb) is more costly than revising the wrong interpretation of non-shifting noun-verb homograph like *report*.

A further area of exploration in this vein is that of a correspondence between the auditory emphasis provided by accents and written emphasis provided by font devices like *italics,* underlining, or CAPITALIZATION. A recent paper by Fraundorf et al. (2013) offers some circumstantial evidence that words presented with font emphasis provided by italics or capitalization activate representations similar to those activated by overtly accented words. Specifically, readers had better memory for, and were better at rejecting alternatives for, a target word presented in CAPS. These results parallel what Fraundorf et al. (2010) observed in an earlier study in which target words were presented auditorily with L+H* accents. One explanation for these similar results is that font emphasis and acoustic emphasis tap into the same discourse processes and do not involve the inner voice; however, another

viable explanation, as Fraundorf et al. (2013) suggest, is that font emphasis leads to the generation of implicit accents.

3 Implicit Rhythm

While Clifton and colleagues (Ashby and Clifton 2005; Breen and Clifton 2011, 2013) have primarily focused their investigation on the role of lexical stress patterns in single words, others have recently begun to explore whether readers' interpretations are affected by a proclivity to place implicit stresses at regular, isochronous, intervals. Although there is debate about how frequently spontaneous speech is isochronous (see Arvaniti 2012, for a recent review), there is increasing evidence that globally rhythmic patterns can influence auditory language comprehension. For example, Niebuhr (2009) was able to influence German listeners' interpretation of global sentence rhythm by manipulating the local *F*0 peak of a target to align with the other *F*0 peaks in the sentence to create one or another global rhythmic pattern. In related work in English, Dilley and her colleagues (Dilley and McAuley 2008; Dilley et al. 2010; Brown et al. 2011, 2012) have repeatedly demonstrated that global *F*0 and duration patterns can influence the interpretation of local word segmentation. For example, Dilley et al. (2010) presented listeners with strings like *banker helpful tie mer der bee* in which the final four syllables could be perceived as *timer derby* or *tie murder bee*. Listeners' report of the final word they heard in these sequences (*derby* or *bee*) was influenced by the global rhythm determined by an alternating high–low *F*0 pattern across the first five syllables of the sentence. Brown et al. (2012) demonstrated that a similar global rhythmic pattern can influence listeners' expectations about upcoming stress patterns; they used global sentence rhythm to induce listeners to interpret the syllable *ʤʊə* as either the first unstressed syllable of *giraffe,* or as the first stressed syllable of *jury*.

Inspired by Dilley's work, Gumkowski and Breen (2013) demonstrated that global rhythm influences listeners' interpretations of syntactically ambiguous lexical material. They embedded stress-shifting noun-verb homographs in ambiguous sentence fragments like (6) in which the target could be interpreted as either a noun (*PROduce*) or a verb (*proDUCE*). Participants listened to acoustically manipulated auditory productions of the fragments, and then provided a written continuation of the sentence. Example (6a) exemplifies the prosodic context designed to encourage listeners to interpret *produce* as a noun (i.e., noun-primed prosody), while (6b) exemplifies the pattern designed to encourage interpretation of *produce* as a verb (i.e., verb-primed prosody). Indeed, listeners were more likely to provide a noun-consistent sentence continuation when the fragment was presented with noun-primed prosody than when it was presented with verb-primed prosody.

(6) Mothers know the good produce…

(a)	H-L	H	L	H-L	H-L
	Mothers	know	the	good	produce…
(b)	L-H	L	H	L	H-L
	Mothers	know	the	good	produce…

There is also evidence that local rhythmic characteristics can influence production. Speakers of English and German prefer to produce speech with alternating strong and weak syllables (Hayes 1995; Selkirk 1984). As such, they tend to avoid stress clashes, in which two adjacent syllables are stressed (Kelly and Bock 1988; Kelly 1988; Anttila et al. 2010; Wasow, this volume; Lee and Gibbons 2007). Inspired by these demonstrations that both global and local rhythmic context guide segmentation and lexical access, researchers have begun to explore whether implicit rhythm has similar effects.

Kentner (2012) manipulated the context surrounding the ambiguous word *mehr*, which can be interpreted as either part of the temporal adverbial *nicht mehr* (7a) or a comparative quantifier (7b).

(7) Der Polizist sagte, dass man.
The policeman said that one.
a…. nicht mehr NACHweisen/erMITteln kann, wer der Täter war.
... couldn't prove/determine anymore who the culprit was.
b…. nicht MEHR NACHweisen/erMITteln kann, als die Tatzeit.
... couldn't prove/determine more than the date of the crime.

In an unprepared reading task, where participants began reading aloud without first silently reading the sentence, readers were more likely to accent *mehr* when it preceded *ermitteln* than when it preceded *nachweisen,* a result that Kentner attributes to readers' avoidance of a stress clash between *mehr* and the first syllable of *nachweisen*. This effect was also evident in a silent reading task, such that reading times were longer on the disambiguating final phrase of the sentence in (7b) for *ermitteln* than *nachweisen*. Kentner argues that the implicit rhythmic structure of *mehr nachweisen* would lead to an initial interpretation of *mehr* as an (unstressed) temporal adverbial, an interpretation which would have to be reanalyzed upon encountering the final phrase, which signals the need for reanalysis of *mehr* as a quantifier.

Other recent studies have corroborated Kentner's finding. For example, Breen and Kenter (2014) demonstrated the effects of rhythm on syntactic disambiguation in the types of sentences used by Gumkowski and Breen (2013). Recall that listeners' interpretation of the ambiguous phrase *good produce* was affected by the overt global rhythmic context in which it occurred. In Breen and Kentner's follow-up study, readers provided written continuations of sentence fragments which contained versions of these ambiguous phrases in which the number of syllables in the ambiguous adjective/noun *good* was manipulated (e.g., *good produce*, *damaged produce*). Readers were more likely to provide a continuation indicating a noun interpretation of *produce* when it was preceded by a trochaic word (*damaged*) than when it was preceded by a monosyllabic word (*good*), and also more likely to provide a continuation consistent with a verb interpretation of *produce* when it was preceded by a monosyllabic word (*good*) than when it was preceded by a trochaic word (*damaged*). This result provides more evidence that readers are sensitive to rhythmic context during silent reading.

As of yet, studies demonstrating a role for implicit rhythm in ambiguity resolution have provided little answer to the question of how specified these rhythmic representations might be. For example, are implicit rhythmic effects the result of the

implicit representation of isochronous (whenever possible) lexical stresses which act as a prime for words with stresses in the predicted locations? If so, how do these lexical stresses interact with implicit nuclear accents, supported by Bader (1998) and Stolterfoht et al. (2007), among others? How, too, are they affected by implicit phrasing? There is certainly more work required to understand how these implicit features interact, if at all.

4 Implicit Intonation

There are, to date, no published studies reporting online effects of implicit intonation. However, two studies have explored whether implicit representations of intonation can affect off line processing. For example, Abramson (2007) demonstrated facilitated lexical decision when the overt intonational pattern of an auditorily presented target matched the implicit intonational contour of a visually presented prime. Participants silently read sentences which were declarative or interrogative and spoken by a man or a woman, as in (8):

(8) a. He/She said: "Do you want to open the package?"
b. He/she said: "I want to open the package."

In a subsequent auditory lexical decision task, participants were faster to recognize targets (e.g., *package*) that were presented in a tonally consistent way (i.e., with rising or falling intonation). This result suggests that readers represent specific intonational contours during silent reading.

Speer and Foltz (this volume) generalized this effect to accents with specific tonal contours using a method similar to that of Abramson (2007). In their study, readers read prime sentences which included corrective contrast, as would be produced on Belinda in (9):

(9) Jacquelyn didn't pass the test. Belinda passed the test.

Subsequently, participants who, in an off line production task, produced L+H* accents on corrective contrast, were faster to identify auditorily presented names which had appeared in the prime sentence if they matched the readers' hypothesized tonal pattern. That is, participants were faster to recognize *Belinda* produced with a L+H* accent than without an accent.

5 Future Directions in Implicit Prosody Research

Having argued here, and elsewhere (Breen 2014) that implicit prosody plays a functional role in sentence comprehension, in that it can influence syntactic ambiguity resolution and reanalysis, I believe we can now consider some more specific questions about the phenomenon. As I see it, research moving forward should be occupied with considering two big questions about implicit prosody: First, how simi-

lar is the inner voice to the overt voice? Second, how do implicit prosodic factors interact with other sources of information known to influence sentence processing?

The first question has often been discussed within the context of the larger question of embodied cognition or perceptual simulation, which is concerned with understanding to what extent people perceptually simulate physical actions described in read text. Some of the most influential work on this topic has demonstrated that brain areas responsible for specific actions are activated even when a participant is only reading about an action. For example, Hauk et al. (2004) demonstrated greater activation in areas of motor cortex specific to verbs that would activate that area of cortex (e.g., reading "kick" activated the area of motor cortex devoted to the feet and legs; cf. Martin and Chao 2001).

Early behavioral work supports the claim that readers also engage in some form of acoustic perceptual simulation. For example, Kosslyn and Matt (1977) demonstrated that readers read text faster when they thought it had been written by a person with a fast speaking rate than when they thought it had been written by someone with a slower speaking rate (cf. Alexander and Nygaard 2008). In fact, Kurby et al. (2009) argue that these effects mean that these perceptual representations of textual speech can serve to influence the comprehension. A similar finding comes from Stites et al. (2013), who explored how readers read direct quotes. They found that readers spent less time reading quotes from speakers described as speaking quickly than quotes from speakers described as speaking slowly (cf. Yao and Scheepers 2011).

Researchers have begun to investigate the nature of the inner voice by following the lead of embodied cognition researchers by investigating to what extent auditory cortex is activated during silent reading. For example, Yao et al. (2011) demonstrated greater activation of voice-selective areas of auditory cortex when participants read direct speech (*Mary said: "Gosh! The movie was terrible!"*) than when they read indirect speech (*Mary said that the movie was terrible;*) (but see Jancke and Shah 2004, for evidence that auditory imagery is dependent in part on training). Moreover, Perrone-Bertolotti et al. (2012) used evidence from intra-cranial recordings of epileptic patients to explore the time course of speech area activation in silent reading. They observed activation of auditory cortex within 500 ms of activation of visual cortex.

These studies offer great insight into the nature of perceptual simulation during reading, but there are surprisingly few having to do specifically with prosody. That is, there is very little evidence about the extent to which prosodic features of accents, phrasing, and rhythm are realized implicitly during silent reading. The challenge here, I believe, is in deciding what kind of evidence is required to show that implicit accents, for example, are the same as, or at least qualitatively similar to, overt accents. And, relatedly, deciding how similar to overt prosodic features we think that implicit prosodic features are. For instance, the fact that silent reading is faster than reading aloud (Ashby et al. 2012) would immediately rule out the possibility that implicit prosody is identical to overt prosody unless we want to allow for the possibility that implicit prosody is highly similar but operates on a faster time course.

Studies of individual differences may inform our understanding of implicit prosodic representations. Certainly, we know that speakers differ in their speech rates, and, to a certain extent, their attachment preferences, so it follows that one readers' implicit prosodic representations should pattern with his/her overt behavior. Jun (2003), for example, has demonstrated that individual speakers' relative clause attachment preferences are consistent with their overt prosodic phrasing. Further, results from Swets et al. (2007) suggest that these individual differences in attachment decisions may be due to the differences in working memory capacity. Speer and Foltz (this volume) found that only those individuals who produced L+H* accents on contrastive constituents demonstrated memory facilitation for constituents produced with implicit L+H* accents. Finally, there is preliminary evidence from work on reading development of a strong connection between children's prosodic production and reading comprehension in that children who produce fluent prosody (characterized by few disfluencies and the presence of well-formed intonational contours) and are also better comprehenders (Schwanenflugel et al. 2004).

Finally, I believe that the next steps in this research should include exploration of how implicit prosody works in conjunction with other information sources during sentence comprehension. Just as researchers continue to explore interactions between lexical, syntactic, and discourse context during on line processing, we are now in a position to begin to explore how implicit prosody contributes.

One example of the type of work I envision moving forward is that of McCurdy et al. (2013), who pitted implicit rhythm effects against discourse context. To do this, they modified the stimuli from Kentner (2012), who manipulated readers' interpretation of the ambiguous *nicht mehr* ("not more") which, as discussed above, is interpreted as a temporal modifier (not anymore) when *nicht* is stressed and *mehr* is unstressed, but as a comparative clause ("not more…than") when *nicht* is unstressed and *mehr* is stressed. Kentner demonstrated that the rhythmic context of the sentence influences readers' interpretation of *nicht mehr* such that readers pursued the interpretation which resulted in an isochronous metrical structure. McCurdy, et al. manipulated the discourse context in which Kentner's (2012) sentences appeared. Specifically, they preceded target sentences, which included the ambiguous *nicht mehr* region, with context sentences which served to prime either the comparative or temporal meaning. They found that both the discourse manipulation and the implicit rhythmic manipulation influenced readers' behavior in an off line rating task. Critically, they found evidence of implicit stress clash avoidance in an eye-tracking study even when discourse context should have ruled out the possibility of stress clash.

Useful starting points for more studies along these lines will be those that have compared overt prosodic features to other sources of information, For example, McDonald et al. (1993) demonstrated that rhythmic well-formedness influenced recall when two members of a conjunct were both inanimate, such that subjects recalled *surprise and sin,* which is rhythmically well-formed in that it contains consistently alternating strong and weak syllables, more than they recalled *sin and surprise,* which is ill-formed due to the presence of a medial stress lapse. However, rhythmic well-formedness did not apply when one member of the conjunct was animate, as

evidenced by the fact that subjects were just as likely to recall the ill-formed *children and room* (as they were to recall the well-formed *horse and tower*). Indeed, finding similar interactions between implicit prosodic features and, e.g., semantic ones, as have been observed for overt prosodic features, would also inform our understanding of the reality of the inner voice.

The past 15 years have seen an increasing interest in the role implicit prosody plays in normal sentence processing. The studies explored here demonstrate tentative evidence for the influence of all prosodic features (phrasing, accent, rhythm, and intonation) on comprehension. Moving forward, I believe the field needs to continue to explore more subtle aspects of implicit prosody, including its relationship to overt prosody, and its interaction with other information sources. Moreover, we can explore how an implicit prosodic representation serves to assist a reader in most effectively understanding written language.

References

Abramson, M. (2007). The written voice: Implicit memory effects of voice characteristics following silent reading and auditory presentation. *Perceptual and Motor Skills, 105,* 1171–1186.

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance, 34*(2), 446–459. doi:10.1037/0096–1523.34.2.446.

Anttila, A., Adams, M., & Speriosu, M. (2010). The role of prosody in the English dative alternation. *Language and Cognitive Processes, 25,* 946–981.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics, 40,* 351–373.

Ashby, J., & Clifton, C., Jr. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96*(3), B89–100. doi:10.1016/j.cognition.2004.12.006.

Ashby, J., Yang, J., Evans, K., & Rayner, K. (2012). Eye movements and the perceptual span in silent and oral reading. *Attention, Perception and Psychophysics, 74*(4), 634–640.

Augurzky, P. (2006). Attaching relative clauses in German: The role of implicit and explicit prosody in sentence processing. MPI Series in Human Cognitive and Brain Sciences 77. Leipzig, Germany.

Augurzky, P. (2008). Prosodic balance constrains argument structure interpretation in German. Poster presented at the 14th conference on architectures and mechanisms for language processing.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht: Kluwer.

Breen, M. (2014). Empirical investigations of the role of implicit prosody in sentence processing. *Language and Linguistics Compass, 8*(2), 37–50.

Breen, M., & Clifton, C., Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language, 64*(2), 153–170. doi:10.1016/j.jml.2010.11.001.

Breen, M., & Clifton, C. Jr. (2013). Stress matters revisited: A display change experiment. *Quarterly Journal of Experimental Psychology, 66*(10), 1896–1909.

Breen, M., & Kentner, G. (2014). Sentence completion to the beat: Effects of implicit prosodic rhythm in English and German. Posted presented at the *27th Annual CUNY Conference on Human Sentence Processing*, Columbus, Ohio. March, 2014.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review, 18*(6), 1189–1196.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2012). Preceding prosody influences metrical expectations during online sentence processing. Poster presented at the 25th annual CUNY conference on human sentence processing.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *The Quarterly Journal of Experimental Psychology: Sect. A, 49*(3), 664–695.

Chafe, W. (1988). Punctuation and the prosody of written language. *Written Communication, 5,* 396–426.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition, 30,* 73–105.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40,* 141–201.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59,* 294–311.

Dilley, L., Mattys, S. L., & Vinke, L. (2010). *Journal of Memory and Language, 63,* 274–294.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27,* 285–319.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. In M. Hirotani (Ed.), *Proceedings of the North East linguistics society* (Vol. 32, pp. 112–132). Amherst: GSLA.

Foltz, A., Maday, K., & Ito, K. (2011). Order effects in production and comprehension of prosodic boundaries. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 39–68). Dordrecht: Springer.

Fraundorf, S. H., Watson, D. G., & Benjamin, A. S. (2010). Recognition memory reveals just how CONTRASTIVE contrastive accenting really is. *Journal of Memory and Language, 63,* 367–386.

Fraundorf, S. H., Benjamin, A. S., & Watson, D. G. (2013). What happened (and what did not): Discourse constraints on encoding of plausible alternatives. *Journal of Memory and Language, 69,* 196–227.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington: Indiana University Linguistics Club.

Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Frazier, L., & Rayner, K. (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning. Memory, and Cognition, 15,* 779–790.

Frenck-Mestre, C., & Pynte, J. (2000). Resolving syntactic ambiguities: Cross-linguistic differences?. *Cross-linguistic perspectives on language processing* (pp. 119–148). Netherlands: Springer.

Gee, J., & Grosjean, F. (1983). Performance structures: A psycholinguistic appraisal. *Cognitive Psychology, 15,* 411–458. doi:10.1016/0010–0285(83)90014–2.

Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59*(1), 23–59.

Grosjean, F., Grosjean, L., & Lane, H. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology, 11,* 58–81.

Gumkowski, N., & Breen, M. (2013). Effects of distal prosody on perceived word stress and syntactic ambiguity resolution. Poster presented at the 26th annual CUNY conference on human sentence processing.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron, 41,* 301–307.

Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. Chicago: University of Chicago Press.

Hemforth, B., Konieczny, L., & Scheepers, L. (1994). Probabilistic or universal approaches to sentence processing: How universal is the human language processor? In H. Trost (Ed.), *KONVENS-94* (pp. 161–170). Berlin: Springer.

Hemforth, B., Konieczny, L., & Scheepers, C. (2000). Syntactic attachment and anaphor resolution: Two sides of relative clause attachment. In M. Crocker, M. Pickering, & C. Clifton, Jr. (Eds.), *Architectures and mechanisms for language processing*. Cambridge: Cambridge University Press.

Hirose, Y. (2003). Recycling prosodic boundaries. *Journal of Psycholinguistic Research, 32,* 167–195.

Huey, E. B. (1908/1968). *The psychology and pedagogy of reading*. Cambridge: MIT Press.

Hwang, H., & Schafer, A. J. (2009). Constituent length affects prosody and processing for a dative NP ambiguity in Korean. *Journal of Psycholinguistic Research, 38,* 151–175.

Hwang, H., & Steinhauer, K. (2011). Phrase length matters: The interplay between implicit prosody and syntax in Korean garden path sentences. *Journal of Cognitive Neuroscience, 23*(11), 3555–3575.

Jancke, L., & Shah, N. J. (2004). Hearing syllables by seeing visual stimuli. *European Journal of Neuroscience, 19,* 2603–2608.

Jun, S. (2003). The effect of phrase length and speech rate on prosodic phrasing. In M. J. Solé, D. Recansens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences*. Barcelona, Spain.

Kamide, Y., & Mitchell, D. C. (1997). Relative clause attachment: Non-determinism in Japanese parsing. *Journal of Psycholinguistic Research, 26,* 247–254.

Kelly, M. (1988). Rhythmic alternation and lexical stress differences in English. *Cognition, 30,* 107–137.

Kelly, M., & Bock, J. (1988). Stress in time. *Journal of Experimental Psychology: Human Perception and Performance, 14*(3), 389–403.

Kentner, G. (2012). Linguistic rhythm guides parsing decisions in written sentence comprehension. *Cognition, 123,* 1–20.

Kitagawa, Y., & Fodor, J. (2006). Prosodic influence on syntactic judgments. In G. Fanselow, C. Fery, R. Vogel, & M. Schlesewsky (Eds.), *Gradience in grammar: Generative perspectives*. Oxford: Oxford University Press.

Kitagawa, Y., Tamaoka, K., & Tomioka, S. (2013). Prosodic matters in intervention effects in Japanese: An experimental study. *Lingua, 124,* 41–63.

Kosslyn, S. M., & Matt, A. M. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society, 9*(4), 250–252.

Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. Invited review of the literature. *Reading Research Quarterly, 45,* 232–253.

Kurby, C., Magliano, J., & Rapp, D. (2009). Those voices in your head: Activation of auditory images during reading. *Cognition, 112,* 457–461.

Lee, M. W., & Gibbons, J. (2007). Rhythmic alternation and the optional complementiser in English: New evidence of phonological influence on grammatical encoding. *Cognition, 105,* 446–456.

Liu, B., Wang, Z., & Zhixing, J. (2010). The effects of punctuations in Chinese sentence comprehension: An ERP study. *Journal of Neurolinguistics, 23,* 66–80.

Lovric, N. (2003). Implicit prosody in silent reading: Relative clause attachment in Croatian. Doctoral dissertation, CUNY Graduate Center.

Martin, A., & Chao, L. L. (2001). Semantic memory and the brainstructure and processes. *Current Opinion in Neurobiology, 11,* 194–201.

McCurdy, K., Kentner, G., & Vasishth, S. (2013). Implicit prosody and contextual bias in silent reading. *Journal of Eye Movement Research, 6,* 1–17.

McDonald, J. L., Bock, K., & Kelly, M. (1993). Word and world order: Semantic, phonological, and metrical determinants of serial position. *Cognitive Psychology, 25,* 188–230.

Niebuhr, O. (2009). F0-based rhythm effects on the perception of local syllable prominence. *Phonetica, 66,* 95–112.

Perrone-Bertolotti, M., Kujala, J., Vidal, J., Hamame, C., Ossandon, T., et al. (2012). How silent is silent reading? Intracerebral evidence for top-down activation of temporal voice areas during reading. *The Journal of Neuroscience, 32,* 17554–17562.

Pynte, J., & Colonna, S. (2000). Decoupling syntactic parsing from visual inspection: The case of relative clause attachment in French. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a Perceptual Process*. Oxford: Elsevier.

Quinn, D., Abdelghany, H., & Fodor, J. D. (2000). More evidence of implicit prosody in silent reading: French, English and Arabic relative clauses. Poster presented at 13th annual CUNY conferences on human sentence processing.

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology, 96*(1), 119–129.

Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.

Slowiaczek, M. L., & Clifton, C., Jr. (1980). Subvocalization and reading for meaning. *Journal of Verbal Learning & Verbal Behavior, 19*(5), 573–582.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 86,* 142–164.

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: The closure positive shift in ERPs as a universal marker of prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30,* 267–295.

Stites, M., Luke, S., & Christianson, K. (2013). The psychologist said quickly, "Dialogue descriptions modulate reading speed!". *Memory & Cognition, 41,* 137–151.

Stolterfoht, B., & Bader, M. (2004). Focus structure and the processing of word order variations in German. In A. Steube (Ed.), *Information structure: Theoretical and empirical aspects* (pp. 259–276). Berlin: De Gruyter.

Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition, 104*(3), 565–590.

Swets, B., Desmet, T., Hambrick, D., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136,* 64–81.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information during spoken language comprehension. *Science, 268,* 1632–1634.

Trueswell, J., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285–318.

Vasishth, S., Agnihotri, R. K., Fernandez, E. M., & Bhatt, R. (2005). Noun modification preferences in Hindi. In Proceedings of Construction of Knowledge conference. Vidya Bhawan Society, Udaipur.

Wagner, M., & Watson, D. (2010). Experimental and theoretical advances in prosody: A review. *Language and Cognitive Processes, 25,* 905–945.

Wijnen, F. (2004). The implicit prosody of jabberwocky and the relative clause attachment riddle. In H. Quene & V. van Heuven (Eds.), *On speech and language. Studies for Sieb G. Nooteboom* (pp. 169–178). Utrecht: Landelijke Onderzoeksschool Taalwetenschap. (LOT occasional Series 2).

Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition, 121*(3), 447–453.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience, 23,* 3146–3152.

How Prosody Constrains First-Pass Parsing During Reading

Markus Bader

Abstract This chapter reviews several experiments (originally presented at CUNY Conferences on Sentence Processing) that have investigated the role of implicit prosody for syntactic ambiguity resolution. The subject of these experiments is the syntactic ambiguity that arises when a matrix clause in German contains two potential antecedent NPs for an extraposed relative clause. In such cases, the second NP is the preferred construal site for the relative clause, probably for reasons of recency. Experimental evidence shows that the first NP becomes more easily accessible as a construal site if readers are given a reason to stress the first NP. The results of these experiments indicate that stress assignment during reading can affect the syntactic structure chosen during first-pass parsing. The chapter explores this finding in the wider context of effects of implicit prosody on first and second pass parsing.

Keywords Implicit prosody · Parsing · Garden-path recovery · German · Extraposition

1 Introduction

Recently I got stuck when reading a newspaper article about the Arabic version of American Idol. The relevant part of this article is given in (1).

(1) Für Samar steht fest: Sie will zur Begrüßung von Assaf fahren, an die Grenze
for S. stands firm she wants to-the welcoming of A. drive to the border

zu Ägypten – auch wenn die Feier in Rafah ein offizieller Termin der Hamas ist.
to Egypt even if the party in Rafah an official date of-the Hamas is

Doch schließlich habe sie für Assaf gestimmt und nicht die Regierung,
but ultimately has she for A. voted and not the government

sagt die junge Oppositionelle.
says the young member-of-the-opposition

M. Bader (✉)
Institut of Linguistics, Goethe University Frankfurt, Frankfurt am Main, Germany
e-mail: bader@em.uni-frankfurt.de

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing*,
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_11

> "Samar has made up her mind: She wants to go to the welcoming of Assaf, to the border of Egypt—even if the party in Rafah is an official event of the Hamas. But ultimately it was her who voted for Assaf and not the government, the young member of the opposition says."

What happened was that I read the last sentence with the intonation shown in (2a), which is the default intonation for such a sentence in German. With this intonation, the contrast phrase *die Regierung* ("the government") did not fit into the already built syntactic structure because then this NP should have been preceded by the preposition *für*. The correct intonation for this sentence is the one shown in (2b)—an unusual intonation that contains a stressed pronoun.

(2) a. Doch schließlich habe sie für ASSAF gestimmt und nicht [*für*] die REGIERUNG.
but ultimately has she for A. voted and not for the government
"But ultimately she voted for ASSAF and not for the GOVERNMENT."

b. Doch schließlich habe SIE für Assaf gestimmt und nicht die REGIERUNG.
but ultimately has she for A. voted and not the government
"But ultimately SHE voted for Assaf and not the GOVERNMENT."

Experimental evidence for the type of processing difficulty associated with an example as in (1) has been provided by Stolterfoht et al. (2007). In an ERP study, Stolterfoht et al. had participants read sentences as in (3) and (4).

(3) Am Dienstag hat der Direktor [den SCHÜler]$_F$ getadelt, und nicht [den LEHrer]$_F$
On Tuesday has the$_{NOM}$ principal the$_{ACC}$ pupil criticized, and not the$_{ACC}$ teacher
"On Tuesday, the principal criticized the pupil, but he did not criticize the teacher."

(4) Am Dienstag hat [der DIREKTOR]$_F$ den Schüler getadelt, und nicht [der LEHrer]$_F$
On Tuesday has the$_{NOM}$ principal the$_{ACC}$ pupil criticized, and not the$_{NOM}$ teacher
"On Tuesday, the principal criticized the pupil, but the teacher did not criticize the pupil."

By default, sentence accent in German falls on the constituent directly preceding the clause-final verb, which is the object in subject–object sentences. The accent pattern shown in (3) is thus the default pattern and the sentence-final contrast phrase must bear accusative case in order to match the focused object. If the final phrase bears nominative case and thus contrasts with the subject, as in (4), the sentence must receive a non-default accent pattern in which the subject bears main sentence accent. Stolterfoht et al. (2007) found evidence that readers initiated processes of reanalysis when encountering the sentence final phrase in (4). Because the syntactic structure of these examples was in no way ambiguous, this provides strong evidence that readers read the sentences with default stress on the object. Since the implicit prosodic structure assigned during first-pass parsing was contradicted by the sentence-final contrast phrase, a prosodic revision became necessary.

The examples discussed so far were examples of pure prosodic ambiguities. In other examples, a prosodic ambiguity goes hand in hand with a syntactic ambiguity. Due to the lack of a one-to-one mapping between syntactic and prosodic structure, two types of syntactic ambiguities can be distinguished. For some syntactic ambiguities, the alternative syntactic structures are associated with different prosodic structures. In this situation, revising the syntactic structure calls for a revision of the associated prosodic structure. For other syntactic ambiguities, the alternative syntactic structures share one and the same prosodic structure. In such a situation,

syntactic revisions can be accomplished without prosodic revisions. Based on these observations, Bader (1994, 1998) proposed the Prosodic Constraint on Reanalysis (PCR) given in (5).

(5) *Prosodic Constraint on Reanalysis (PCR)*
Revising a syntactic structure is difficult if it necessitates a concomitant reanalysis of the associated prosodic structure.

In the examples considered so far, the prosodic ambiguity always concerned the location of the main sentence accent. The PCR is not confined to ambiguities of this kind, however. It also applies to the other major domain of prosody, namely the structuring of sentences by means of prosodic boundaries. To see how the PCR differentiates between weak and strong garden-path effects, consider the two sentences in (6) and (7).

(6) In order to help *the little boy* put down the package he was carrying. strong GP

(7) Peter knew *the answer* would be false. weak GP

Whereas (6) gives rise to a strong garden-path effect which has been claimed to be consciously perceivable, the garden-path effect caused by (7) is a weak one which normally goes unnoticed (e.g., Gorrell 1995; Pritchett 1992).

Consider first the processing of the difficult garden-path sentence in (6). In (6), the locally ambiguous phrase *the little boy* starts the main clause. This contradicts its preferred attachment as the object of *to help* within the embedded clause. For both the preferred and the unpreferred attachment, the relationship between attachment site and intonational phrasing is shown in (8). Here, round brackets indicate the prosodic structure and "IP" stands for intonation phrase.

(8) a. ($_{IP}$ In order to help *the little boy* $_{IP}$) ($_{IP}$ Jill put down the package she was carrying. $_{IP}$)
b. ($_{IP}$ In order to help $_{IP}$) ($_{IP}$ *the little boy* put down the package he was carrying. $_{IP}$)

In (8), the fronted embedded clause and the following main clause are separated by an intonation break because topicalized clauses usually form an intonation phrase of their own. As pointed out by Wagner and Watson (2010) in a recent overview of prosody and sentence processing, experimental investigations of auditory sentence comprehension have repeatedly shown that prosody reliably helps hearers in avoiding the garden-path that occurs when reading such sentences (e.g., Speer et al. 1996; Warren et al. 1995).

The consequences of the intonational phrasing in (8) for the processing of the garden-path sentence in (6) is shown in (9)

(9) a. ($_{IP}$ [$_{CP}$ In order to help *the little boy* $_{CP}$] $_{IP}$) …? ⇒ Integrate next word: *put*

b. ($_{IP}$ [$_{CP}$ In order to help *the little boy* $_{CP}$] $_{IP}$) put …? ⇒ IProsodic Revision

c. ($_{IP}$ [$_{CP}$ In order to help $_{CP}$] $_{IP}$) *the little boy* put …

During first-pass parsing of (6), the NP *the little boy* is attached as object of the preceding verb *to help* and a syntactic as well as a prosodic boundary is inserted after this NP. When the parser then tries to integrate the next word, the main clause verb *put*, it will not find an appropriate integration site. Such an integration site will only become available after a prosodic reanalysis has removed the NP *the little boy* from

the sentence-initial intonation phrase. It is this prosodic reanalysis which leads to the impression of a strong garden-path effect according to the PCR.

Consider next the processing of sentence (7), which causes only a weak garden-path effect. (10) shows the prosodic structure assigned to this sentence and to an alternative sentence in which the locally ambiguous NP *the answer* is attached in the preferred way, namely as the direct object of the preceding verb.

(10) a. ($_{IP}$ Peter knew *the answer* immediately $_{IP}$)

b. ($_{IP}$ Peter knew *the answer* would be false $_{IP}$)

Both sentences in (10) constitute a single intonation phrase. This implies that neither the preferred nor the unpreferred structure contains a major prosodic break between the verb *knew* and the following NP *the answer*. Although the sentences differ at the lower level of the prosodic phrase, experimental evidence shows that hearers cannot reliably differentiate between the two structures (e.g., Watt and Murray 1996). As discussed by Wagner and Watson (2010), in examples of this kind phrasing is also affected by the subcategorization preferences of the main clause verb (e.g., Tily et al. 2009), but this does not affect the main argument made here.

If we assume that the prosodic level most relevant for the PCR is the level of the intonation phrase, arriving at the correct structure for sentence (7) is less costly than it was in the case of sentence (6). This is shown in (11).

(11) a. ($_{IP}$ [$_{CP}$ *Peter knew the answer* …⇒ Integrate next word: *would*

b. ($_{IP}$ [$_{CP}$ *Peter knew* [$_{CP}$ *the answer* would …

No prosodic reanalysis at the level of the intonation phrase becomes necessary on encountering the disambiguating auxiliary *would* in sentence (7). This sentence is accordingly easy to process despite the need to revise the initial syntactic structure.

The PCR does in no way imply that ease of garden-path recovery is only a matter of prosody. While the revision of the syntactic structure built on first-pass parsing was not associated with noticeable costs in the weak garden-path example discussed above, this is not always so. For example, the two sentences in (12) exhibit a local syntactic ambiguity concerning the case of the clause-initial object—dative case in (12a) and accusative case in (12b).

(12) a. ¿Menschen$_{DAT}$, die in Not sind, sollte man helfen.
people who in distress are should one help
'One should help people who are in distress.'

b. Menschen$_{ACC}$, die in Not sind, sollte man unterstützen.
people who in distress are should one support
'One should help people who are in distress.'

This syntactic ambiguity has no correspondence in the prosodic domain. Experimental results provided by Hopf et al. (1998) nevertheless show that sentence (12a) is a garden-path sentence. That is, revising the initial assignment of accusative case on encountering a dative-assigning verb in clause-final position is not cost-free. As far as the PCR is correct, one of the challenges for research into syntactic ambiguity resolution is to tease apart the different sources that jointly determine ease of garden-path recovery.

The PCR follows as a special case from the Implicit Prosody Hypothesis (IPH) proposed by Janet Fodor (Fodor 2002).

(13) *The Implicit Prosody Hypothesis (IPH)*
In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for the construction.

The IPH is broad enough to subsume effects of implicit prosody both on first- and second-pass parsing. As discussed in more detail by Breen (Empirical Investigations of Implicit Prosody), the work of Janet Fodor and her colleagues on implicit prosody has provided important insights into how prosodic phrasing affects syntactic ambiguity resolution during reading. The aim of the current chapter is to show that stress assignment does not only constrain garden-path recovery, as claimed by the PCR, but that the effects of stress assignment are more general, as envisioned by the IPH, affecting both first- and second-pass parsing. To this end, this chapter presents a case study on the parsing of sentences containing extraposed relative clauses.

2 Extraposed Relative Clauses

A relative clause in German can occur either adjacent to its head noun, as in (14a), or extraposed behind the clause-final verb, as in (14b).

(14) a. Ich glaube, dass der Lehrer *ein Buch*, *das langweilig ist*, empfohlen hat.
I believe that the teacher a book that boring is recommended has
"I believe that the teacher recommended a book that is boring."

b. Ich glaube, dass der Lehrer *ein Buch* empfohlen hat, *das langweilig ist*.
I believe that the teacher a book recommended has that boring is
"I believe that the teacher recommended a book that is boring."

The conditions governing relative clause extraposition have been studied both by means of corpus analysis (Hawkins 1994; Shannon 1992; Uszkoreit et al. 1998) and by experimental means (Konieczny 2000). The syntactic position of extraposed relative clauses is controversial, but since this issue is not crucial for the upcoming discussion it will not be discussed further here.[1]

The empirical research on relative clause extraposition in German has revealed two major generalizations. First, when the NP containing the relative clause and thus the relative clause itself occur directly in front of the clause-final verb(s), extraposition is strongly preferred. Second, the probability of extraposition decreases when additional material separates the relative clause from the clause-final verb(s).

For simple sentences with a subject and an object occurring in that order, this implies that a relative clause modifying the object is preferentially extraposed,

[1] A detailed discussion of extraposition in German can be found in Haider (2010). A review of the various syntactic approaches to relative clause extraposition, mainly based on data from English, is provided by Webelhuth et al. (2013).

whereas a relative clause modifying the subject preferentially stays in situ and thus adjacent to its head noun. To see why, consider (15) and (16).

(15) Preferred position of a relative clause modifying the object
C° $NP_{Subject}$ [NP_{Object}] V° RC
Ich glaube, dass der Lehrer *ein Buch* empfohlen hat, *das langweilig ist.*
I believe that the teacher a book recommended has that boring is
"I believe that the teacher recommended a book that is boring."

(16) Preferred position of a relative clause modifying the subject
C° [$NP_{Subject}$ RC] NP_{Object} V°
Ich glaube, dass *der Lehrer, der langweilig ist,* ein Buch empfohlen hat.
I believe that the teacher that boring is a book recommended has
"I believe that the teacher that is boring recommended a book."

Since an object occurs directly in front of the clause-final verb, extraposition as in (15) is the preferred option for relative clauses modifying an object. The subject, in contrast, is separated from the clause-final verb by the intervening object. A relative clause modifying the subject therefore preferentially stays in situ, as in (16).

Several accounts have been proposed for the pattern of relative clause extraposition described above. According to Hawkins (1994), the principle of *Early Immediate Constituents* prefers word orders that allow for the most rapid online construction of a phrase-structure representation. In Hawkins' more recent version of his theory (Hawkins 2004), the principle of Early Immediate Constituents still plays an important role, but various types of syntactic and semantic dependencies are now taken into account, too. Inspired by the Dependency Locality Theory of Gibson (2000), Temperley (2007) and Gildea and Temperley (2010) have developed a performance account stated directly in terms of dependency length. According to this account, word orders are preferred that minimize the length of the various syntactic dependencies within a clause.

There are several grammar-based alternatives to these performance-based theories. An information-structural account was proposed by Shannon (1992). According to Shannon, relative clauses that modify a topic stay adjacent to their head noun whereas relative clauses that modify a focus are extraposed. The focus constituent in German tends to occur late in the clause, typically directly in front of the clause-final verb, whereas the topic typically occupies an early position in the clause. This explains the basic finding that the probability of extraposition is high when the antecedent is adjacent to the clause-final verb, but declines when additional material intervenes between antecedent NP and verb. A prosodic account of extraposition is proposed by Féry (Extraposition and Prosodic Monsters in German). According to this account, extraposition is a means to avoid a prosodic structure in which an intonation phrase (corresponding to the relative clause) is embedded within a prosodic phrase (corresponding to the VP of the matrix clause). Such a structure violates the Strict Layer Hypothesis and is therefore disprefered for prosodic reasons. Extraposition of a constituent is blocked, however, when an accented noun intervenes between the constituent and the clause-final verb.

Performance-based and grammar-based explanations of extraposition do not exclude each other. In fact, work on extraposition phenomena in English has revealed cases where the decision to extrapose is influenced both by weight and by grammar-internal properties. For example, Arnold et al. (2000) have shown that the rate of

heavy-NP shift in English is higher both for longer NPs and for NPs that are new in the discourse.

To my knowledge, research on relative clause extraposition in German has not yet addressed the question of whether different factors—in particular weight, information structure, and prosody—are involved, and if so, what the contribution of each factor is. There are good reasons, however, for assuming that an adequate theory of relative clause extraposition will subsume elements of both performance accounts and information-structural accounts, as for related structures in English. On the one hand, a theory like that of Shannon (1992) does not capture the finding that the probability of extraposition declines in a gradient fashion with increasing length of the intervening material. Shannon's theory also gives no account of the finding that the probability of extraposition increases with increasing length of the relative clause. On the other hand, performance-based theories are incomplete insofar as they leave open why under identical weight conditions speakers sometimes extrapose and sometimes donot. It is at this point where information structure might come to help.

For example, Shannon's hypothesis that relative clauses are extraposed when they modify a focused constituent could account for those rare cases where extraposition occurs across an intervening object. This should be possible if a constituent in front of the object is focused. As shown in (17), when a subject is focused by means of a focus particle, extraposing a relative clause modifying the subject becomes quite natural, but only when the article gets main stress (17a), and not when the noun is stressed (17b).

(17) Extraposition from a focused subject

a. Ich glaube, dass nur DER Lehrer ein Buch empfohlen hat, der langweilig ist.
I believe that only the teacher a book recommended has that boring is
"I believe that only the teacher that is boring recommended a book."

b.?Ich glaube, dass nur der LEHRER ein Buch empfohlen hat, der langweilig ist.
I believe that only the teacher a book recommended has that boring is
"I believe that the teacher that is boring recommended a book."

The reason why main stress on the determiner is necessary for extraposition follows from the different interpretative effects brought about by stressing the determiner or the noun. (17a) can be used in a situation in which it is already known that at least one teacher from a group of teachers has recommended a book, but it is unknown which teacher(s). Here, the noun is deaccented because the set of teachers is a given information, and a relative clause is necessary in order to identify the particular teacher who has recommended a book. (17b), in contrast, is appropriate in situations in which a group of people is under discussion, with only a single member of this group being a teacher. Here, the noun gets main stress in order to identify the teacher as the one who recommended a book. A relative clause is superfluous in this case because there is only a single teacher.

The judgments shown in (17) reflect my own intuitions. As the discussion in Féry (Extraposition and Prosodic Monsters in German) makes clear, these judgments are not shared by everyone. Since empirical evidence that could settle this issue is lacking, I will stick to the assumption that stress on the determiner is needed in order to extrapose from the subject across an intervening object.

With the necessary grammatical background at hand, we can now turn to the question of how readers process sentences containing an extraposed relative clause that is ambiguous with respect to its host NP. For purposes of illustration, consider the example in (18).

(18) Ich glaube, dass der Lehrer einen Film empfohlen hat, der langweilig ist.
I believe that the teacher a film recommended has that boring is
"I believe that the teacher that is boring recommended the film."

In accordance with the IPH of Fodor (2002), assume that readers project a default prosodic contour onto this sentence. In this default prosodic contour, main stress falls onto the noun of the object NP, because the object is located directly in front of the clause-final verb, and the position in front of the clause-final verb is the default position for main sentence accent in German. With this intonation pattern, the parser must construe the extraposed relative clause as modifying the object. Associating the relative clause with the subject is excluded for prosodic reasons. For a sentence as in (18), a clear preference for the object construal of the relative clause is thus predicted.

Consider next sentence (19) in which the focus particle *gerade* precedes the subject.

(19) Es scheint, dass gerade [$_F$ der Lehrer] den Film empfohlen hat, der langweilig ist.
It seems that just the teacher the film recommended has that boring is
"I seems that just the teacher that is boring recommended the film."

Due to the presence of the focus particle, the subject must be a focus. As explained above, the subject NP *der Lehrer* ("the teacher") can be made a focus either by accenting the determiner or by accenting the noun. Only an accent on the determiner licenses an extraposed relative clause. The reason is that an accent on the determiner presupposes that a group of teachers is under discussion and the relative clause serves to pick out a particular teacher. If the accent is put on the noun, an extraposed relative clause is not licensed.

By default, an NP gets main stress at its right edge, which is the noun in simple NPs like *der Lehrer*. If readers assign default stress, as stated by the IPH, they should accordingly stress the noun and not the determiner. Because this does still not license a relative clause in extraposed position, readers must be given a trigger in the input in order to deviate from default stress. The focus particle itself is already such a trigger, which directs the main sentence stress away from its default position (the object) and onto the subject. While readers indeed seem to assign default stress also NP internally, they sometimes deviate from default stress for rhythmic reasons. In particular, reading time data presented in Bader (1998) shows that the chance of stressing the word directly following a focus particle increases when the focus particle ends in one or more unstressed syllables (e.g., *gerade* "just"; *ausschließlich* "exclusively"). The most probable reason for this is that readers thereby avoid a lapse—a long stretch of unstressed syllables, which is a marked configuration for rhythmic reasons (Nespor and Vogel 1989). In the following experiments, focus particles of this type will be used in the experimental material.

In sum, an extraposed relative clause can either modify the subject or the object. If the extraposed relative clause modifies the subject, the subject's determiner has to be stressed. This is not necessary if the relative clause modifies the object. If readers project a default prosodic structure, as claimed by the IPH, determiners will not be stressed, and a strong preference for associating the relative clause with the object should result. However, if the subject is preceded by an appropriate focus particle, readers have a reason to put main stress on the subject's determiner, and associating the relative clause with the subject should become possible.

3 Experiment 1

Experiment 1 tests the two predictions that were derived above concerning the processing of extraposed relative clauses. First, extraposed relative clauses are preferentially construed as modifying the most recent NP, which is the object in simple subject–object sentences. Second, this preference diminishes when readers are given a reason to stress *die* from the beginning. Such a reason can be provided by putting a focus particle in front of the subject, with the consequence that the subject becomes a focus. However, a focus particle before a definite NP only means that either the article or the noun must be accented. Only the former case will lead to the expectation that the NP is modified by a relative clause. Therefore, a preceding focus particle will change the preferred construal of an ambiguous extraposed relative clause only in a probabilistic way.

4 Method

Participants Forty-four students of the University of Jena participated for course credits or payment. In this and all other experiments, participants were always native speakers of German and naive with respect to the purpose of the experiment.

Materials Twenty-four sentences were constructed for Experiment 1. Each sentence appeared in four versions according to the two factors association site (object vs. subject) and focus particle (without vs. with). A complete stimulus set is shown in Table 1.

Procedure Experiment 1 used a word-by-word noncumulative self-paced reading procedure. Participants read sentences on a computer screen using a moving window display in which all nonspace characters of the sentence were initially replaced by underlines (Just et al. 1982). Participants pressed a key on the keyboard to see each new word of the sentence. On each key press, a new word was uncovered and the previous word was again replaced by underlines. The time between successive

Table 1 A complete stimulus set of Experiment 1

Without focus particle	
Object RC	Der Direktor wunderte sich darüber, dass die Professorin The director wondered himself about that the professor einige Studenten besucht hat, die letzte Woche krank waren some students visited has who last week sick were "The director was surprised that the professor visited some students who were sick last week"
Subject RC	Der Direktor wunderte sich darüber, dass die Professorin The director wondered himself about that the professor einige Studenten besucht hat, die letzte Woche krank war some students visited has who last week sick was "The director was surprised that the professor who was sick last week visited some students"
With focus particle	
Object RC	Der Direktor wunderte sich darüber, dass gerade die Professorin The director wondered himself about that just the professor einige Studenten besucht hat, die letzte Woche krank waren some students visited has who last week sick were "The director was surprised that just the professor visited some students who were sick last week"
Subject RC	Der Direktor wunderte sich darüber, dass gerade die Professorin The director wondered himself about that just the professor einige Studenten besucht hat, die letzte Woche krank war some students visited has who last week sick was "The director was surprised that just the professor who was sick last week visited some students"

key presses was recorded automatically. The key press terminating the last word of the sentence either revealed the next sentence or a yes–no-question which had to be answered by pushing the "j"-key for "Ja" ("yes") or the "n"-key for "Nein" ("no"). Participants received no feedback for their answers. To become acquainted with the procedure, participants read four training sentences before the experiment started.

5 Results

The reading times on the clause-final disambiguating verb of the relative clause are shown in Fig. 1. Two-way ANOVAs revealed that the factor association site was significant ($F1(1,43)=18{,}31$, $p<0.001$; $F1(1, 23)=27{,}75$, $p<0.001$) whereas the factor focus particle was not ($F1(1, 43)=0.87$, $p>0.1$; $F2(1, 23)=,64$, $p>0.1$). The interaction of focus particle and association site was also significant ($F1(1,43)=4{,}82$, $p<0.05$; $F2(1,23)=9.65$, $p<0.01$).

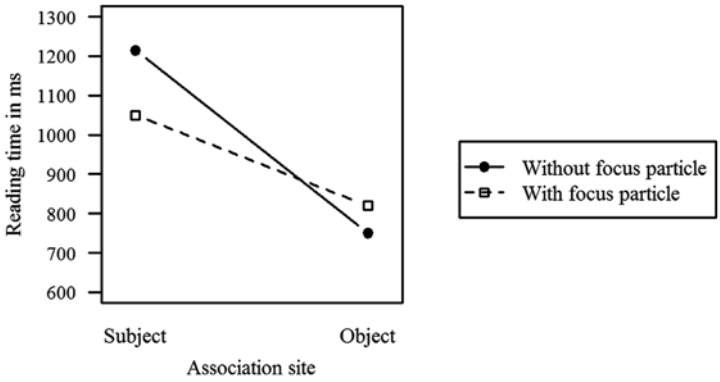


Fig. 1 Reading times on the clause-final verb of the relative clause in Experiment 1

6 Discussion

The results of Experiment 1 confirm the predictions made at the outset. First, sentences in which the extraposed relative clause modified the object were generally much easier to process than sentences in which the extraposed relative clause modified the subject. Second, introducing a focus particle in front of the subject made sentences with subject-modifying relative clauses easier and sentences with object-modifying relative clauses more difficult to process.

Because Experiment 2 replicates Experiment 1 with a different procedure, a more thorough discussion of the results of Experiment 1 is postponed to the end of Experiment 2.

7 Experiment 2

Experiment *2* investigates the same material as Experiment 1, but uses the procedure of end-of-sentence speeded grammaticality judgments instead of a self-paced reading procedure. This procedure has been used before in research on syntactic ambiguity resolution (e.g., Ferreira and Henderson 1991; Warner and Glass 1987). In the current context, this method is of special value because it yields straightforward information concerning garden-path strength.

8 Method

Participants and Materials Twenty-four students of the University of Jena participated for course credits or payment. The stimulus material consisted of a subset of the 20 sentence quartets from Experiment 1.

Procedure Sentences were presented visually using the DMaster software developed by K. Forster and J. Forster at Monash University and the University of Arizona. Participants sat in front of a computer monitor. Their task was to read sentences on the computer screen and judge the grammaticality of each sentence as quickly and accurately as possible. The concept of grammaticality was explained with the help of examples. Participants initiated each trial by pressing the spacebar which triggered three fixation points to appear in the center of the screen for 1050 ms. Thereafter, the sentence appeared on the screen word-by-word, with each word appearing at the same position (mid-screen). Each word was presented for 225 ms plus additional 25 ms for each character to compensate for length effects. There was no interval between words. Immediately after the last word of a sentence, three red question marks appeared on the screen, signaling to participants that they now had to make their judgment. Participants indicated their judgment by pressing either the left or the right shift key on a computer keyboard. If participants did not respond within 2000 ms, a red warning line "zu langsam" ("too slow") appeared on the screen and the trial was aborted.

9 Results

The results for Experiment 2 are shown in Fig. 2.[2] Two-way ANOVAs revealed a significant main effect of association site with subject association judged poorer than object association ($F1(1,23)=78.76$, $p<0.01$; $F2(1,19)=104.24$, $p<0.01$). The factor focus particle was not significant (both F-values <1). The interaction of association site and focus particle was significant ($F1(1,23)=38.07$, $p<0.01$;

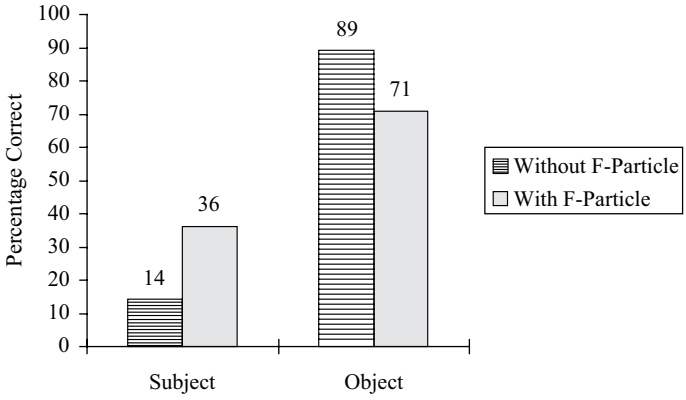


Fig. 2 Percentages of correct responses in Experiment 2

[2] Reaction times were also measured. The reaction time data are not shown here because they do not provide additional information for the experiments presented in this chapter.

F2(1,19) = 11.97, $p < 0.01$). For sentences with subject-association, sentences containing a focus particle were judged as grammatical more often than sentences without a focus particle (F1(1,23) = 17.67, $p < 0.01$; F2(1,19) = 17.55, $p < 0.01$). For sentences with object-association, the reversed pattern was found (F1(1,23) = 12.65, $p < 0.01$; F2(1,19) = 12.56, $p < 0.01$).

10 Discussion

Experiment 2 replicates the result pattern observed in Experiment 1. Sentences with subject-modifying extraposed relative clauses are more difficult to process than sentences with object-modifying extraposed relative clauses, and introducing a focus particle in front of the subject enhances the probability of associating an extraposed relative clause with the subject, thereby making subject-modifying relative clauses less difficult and object-modifying relative clauses more difficult.

Taken together, the first two experiments show that it is quite difficult for readers to construe an extraposed relative clause with the subject across an intervening object. This is in particular true for sentences in which the subject was not preceded by a focus particle. As shown by Experiment 2, sentences of this kind were rejected as ungrammatical in about 90 % of the time. Thus, reanalysis seems to be particularly hard in these cases, as predicted by the PCR. When a focus particle preceded the subject, performance was substantially better, but the absolute level of acceptance was still at only 30 %. A probable reason for this is that a focus particle is only suggestive of putting an accent on the adjacent determiner, but does not force readers to do so. Thus, even with a focus particle in front of the subject, participants will often misconstrue the relative clause on first-pass parsing, which results in a strong garden-path effect due to the difficult reanalysis.

Since the object follows the subject, the preference for associating the extraposed relative clause with the object is a locality or recency preference of the sort that has often been found in research on syntactic ambiguity resolution (see Frazier and Fodor 1978 and much following work). The preference for object modification would thus also follow from one of the principles proposed to account for recency effects in human syntactic parsing, like the Recency Preference Principle proposed by Gibson et al. (1996).

(20) *Recency Preference*
Preferentially attach structures for incoming lexical items to structures built more recently.

While it is surely no accident that the most recent NP is the preferred association site for the extraposed relative clause, recency alone is not sufficient to account for the present findings. In particular, recency gives no account of the strength of the observed preference. Consider for comparison the two sentences in (21) and (22). These two sentences illustrate the well-known attachment ambiguity that arises when a relative clause follows a complex NP.

(21) a. Ich kenne die Töchter *der Gräfin*, die in London wohnt.
I know the dauthers of-the countess who in London lives
"I know the daughters of the countess who lives in London."
b. Ich kenne *die Töchter* der Gräfin, die in London wohnen.
I know the dauthers of-the countess who in London live
"I know the daughters of the countess who live in London."

For German, as for many others, but not for all languages, a preference for high attachment has been found (see Pickering et al. 2006 for an overview). Since this is just the opposite of a recency preference, the Recency Preference Principle cannot play a general role in German. In particular, the Recency Preference Principle cannot be invoked to explain the strong preference for object-association observed for extraposed relative clauses. In fact, this is not necessary because the IPH provides a viable alternative.

The results of Experiments 1 and 2 are also in accordance with the *Focus Attraction Hypothesis* proposed by Schafer et al. (1996).

(22) *Focus Attraction Hypothesis*
It is more likely that a phrase that is neither a complement nor syntactically obligatory will be taken to modify a phrase P if P is focused than if it is not, grammatical and pragmatic constraints permitting.

The evidence for the Focus Attraction Hypothesis provided by Schafer et al. (1996) comes from auditory language processing, whereas Experiments 1 and 2 were reading experiments. Under the assumption that it is the prosodic manifestation that is relevant for the Focus Attraction Hypothesis, and not just focus as a formal feature, the finding of focus related effects both in listening and reading provides further evidence for the hypothesis that prosody can have an influence on syntactic processing during reading.

A question left open by the results discussed so far is whether the difficulty seen with extraposed relative clauses when they modify a subject across an object is just a problem of reanalysis, or whether integrating the relative clause with the subject is a difficult operation per se, even when reanalysis is not at stake. This question is addressed in the next two experiments.

11 Experiment 3

Experiment 3 takes advantage of the fact that the German determiner *diejenige* is strongly biased toward occurring together with a relative clause. Sentences (23) is an example sentence containing this determiner.

(23) Ich glaube, dass **diejenige** Schwimmerin gewinnen wird, die am meisten trainiert.
I believe that that swimmer win will who at most exercises
"I believe that that swimmer will win that exercises most."

Without the relative clause, sentence (23) sounds incomplete, although it would probably not be considered as completely ungrammatical. Under specific contextual conditions, it seems to be possible to use a sentence like (23) without the relative clause, in particular by using the *diejenige* NP in a deictic way. Without such a specific context, *diejenige* creates a strong expectation of an upcoming relative clause.

Because of this expectation, any processing difficulties that are nevertheless observed for sentences with the determiner *diejenige* can be attributed to difficulties of integrating a relative clause across intervening material. Experiment 3 compares sentences in which the subject is headed by the determiner *diejenige* to sentences in which the subject is a definite NP preceded by a focus particle, as investigated in the two prior experiments. This comparison makes it possible to see how effective the focus particle is in narrowly focusing the adjacent article and thereby creating an expectation for a relative clause.

12 Method

Participants and Procedure Ninety-nine students of the University of Jena participated for course credits or payment. The same speeded grammaticality judgment procedure was used as in Experiment 2.

Materials The sentences investigated in Experiment 3 were identical to the sentences of Experiment 2 with one exception. In the condition "without focus particle", the sentences of Experiment 3 contained the determiner *diejenige* instead of the definite article *die*. Table 2 shows a complete stimulus set for Experiment 3.

Table 2 A complete stimulus set of Experiment 3

Diejenige	
Object RC	Der Direktor wunderte sich darüber, dass diejenige Professorin The director wondered himself about that professor einige Studenten besucht hat, die letzte Woche krank waren some students visited has who last week sick were "The director was surprised that that professor visited some students who were sick last week"
Subject RC	Der Direktor wundertesich darüber, dass diejenige Professorin The director wondered himself about that professor einige Studenten besucht hat, die letzte Woche krank war some students visited has who last week sick was "The director was surprised that that professor who was sick last week visited some students"
Die + focus particle	
Object RC	Der Direktor wunderte sich darüber, dass ger*ade* die Professorin The director wondered himself about that just the professor einige Studenten besucht hat, die letzte Woche krank waren some students visited has who last week sick were "The director was surprised that just that professor visited some students who were sick last week"
Subject RC	Der Direktor wunderte sich darüber, dass ger*ade* die Professorin The director wondered himself about that just the professor einige Studenten besucht hat, die letzte Woche krank war some students visited has who last week sick was "The director was surprised that just that professor who was sick last week visited some students"

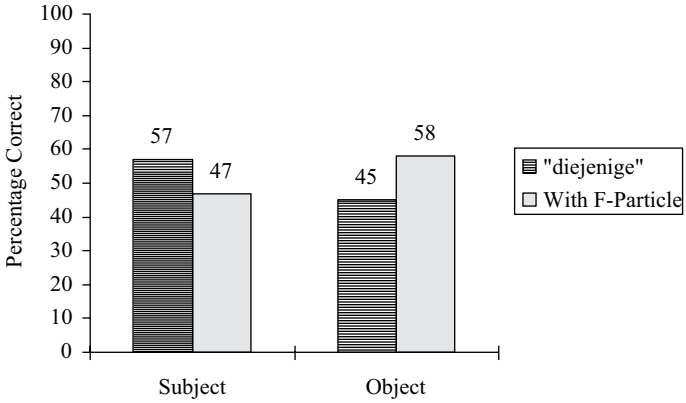


Fig. 3 Percentages of correct responses in Experiment 3

13 Results

The results of Experiment 3 are shown in Fig. 3. The two main factors were not significant, but the interaction between association site and type of the first NP was ($F1(1,98) = 26.88$, $p < 0.01$; $F2(1,19) = 11.80$, $p < 0.01$). For sentences with *diejenige*, sentences with subject-association were judged as grammatical more often than sentences with object-association ($F1(1,98) = 19.39$, $p < 0.01$; $F2(1,19) = 9.37$, $p < 0.01$). For sentences with focus particle, sentences with object-association were judged as grammatical more often than sentences with subject-association ($F1(1,98) = 11.56$, $p < 0.01$; $F2(1,19) = 5.93$, $p < 0.05$).

14 Discussion

When the determiner of the subject was *diejenige*, sentences with a subject-modifying relative clause were judged better than sentences with an object-modifying relative clause. This is the opposite to what was found in the preceding experiment and, therefore, shows that *diejenige* had the expected effect—creating an expectation for an upcoming relative clause and thereby easing the association of the extraposed relative clause with the subject. In absolute terms, however, performance was poor even for sentences with *diejenige*.

For sentences in which the subject was preceded by a focus particle, the results of Experiment 3 replicate the pattern found in Experiment 2. Sentences with an object-modifying relative clause were judged better than sentences with a subject-modifying relative clause. In absolute terms, sentences with subject-modifying relative clauses received higher and sentences with object-modifying relative clause lower judgments than in Experiment 2. This was probably a side effect of the presence of sentences with *diejenige*, which provided participants with a model for associating an extraposed relative clause with a distant subject NP.

A comparison between sentences with *diejenige* and sentences with a focus particle reveals only small differences. The judgments for sentences with subject-modifying relative clauses were about 10 % higher with the determiner *diejenige* than with a focus particle. For sentences with object-modifying relative clauses, a reverse difference of about the same size was found. This shows that the focus particles used in the present materials were almost as effective in creating an expectation for a relative clause as the determiner *diejenige,* which creates this expectation as an inherent lexical property.

15 Experiment 4

In all prior experiments, the extraposed relative clause was locally ambiguous insofar as it could be associated with either the subject or the object until the clause-final auxiliary disambiguated the sentence by means of subject-verb agreement. As shown in (24), the local ambiguity of the relative clause was caused by the morphological ambiguity of the relative pronoun *die* which is ambiguous between the feature specification [feminine, singular] and the feature specification [plural].

(24) dass die Professorin einige Studenten besucht hat, die … war/waren that the professor some students visited has who was/were *fem.sing plural fem.sing/plural sing/plural*

Experiment 4 will contrast sentences with feminine *diejenige* with sentences containing masculine *derjenige* instead of *diejenige* as determiner of the subject and *der* instead of *die* as relative pronoun. This is shown in (25).

(25) dass derjenige Professor einige Studenten besucht hat, der/die … war/waren that that professor some students visited has who was/were *masc.sing plural masc. sing/plur sing/plural*|

The relative pronoun *die* is not compatible with the feature specification [masculine singular]. For sentences with masculine forms, it is therefore already the first-word of the relative clause (the relative pronoun) which signals to the reader that the relative clause must be associated with the subject. Thus, sentences with *derjenige* provide maximal help for finding the correct head noun of the relative clause.

16 Method

Participants and Procedure Sixteen students of the University of Jena participated for course credits or payment.

Materials The materials for Experiment 4 was identical to the materials of Experiment 3 with one exception. The factor focus particle was replaced by the factor Gender with the two conditions “feminine” (*diejenige*, taken from Experiment 3) and “masculine” (*derjenige*). Table 3 shows a complete stimulus set for Experiment 4.

Table 3 A complete stimulus set of Experiment 4

Feminine	
Object RC	Der Direktor wunderte sich darüber, dass diejenige Professorin The director wondered himself about that that professor einige Studenten besucht hat, die letzte Woche krank waren some students visited has who last week sick were "The director was surprised that that professor visited some students who were sick last week"
Subject RC	Der Direktor wunderte sich darüber, dass diejenige Professorin The director wondered himself about that that professor einige Studenten besucht hat, die letzte Woche krank war some students visited has who last week sick was "The director was surprised that that professor who was sick last week visited some students"
Masculine	
Object RC	Der Direktor wunderte sich darüber, dass derjenige Professor The director wondered himself about that that professor einige Studenten besucht hat, die letzte Woche krank waren some students visited has who last week sick were "The director was surprised that that professor visited some students who were sick last week"
Subject RC	Der Direktor wunderte sich darüber, dass derjenige Professor The director wondered himself about that that professor einige Studenten besucht hat, der letzte Woche krank war some students visited has who last week sick was "The director was surprised that that professor who was sick last week visited some students"

17 Results

The results for Experiment 4 are shown in Fig. 4. Two-way ANOVAs revealed that both main effects were significant whereas their interaction was not.

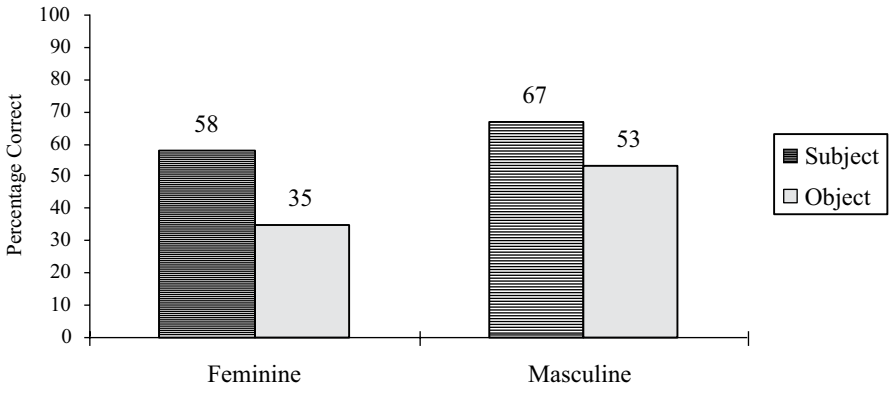


Fig. 4 Percentages of correct responses in Experiment 4

18 Discussion

As expected, sentences with a masculine subject were judged better than sentences with a feminine subject. The only difference between these two types of sentences is that a relative clause with a feminine relative pronoun can also be associated with the object as long as the clause-final verb with its number information has not been read. For morphological reasons, this is not possible with a masculine relative pronoun. Thus, finding the correct association site for an extraposed relative clause is more difficult if an NP intervenes, that is, a potential relative clause host in terms of morpho-syntactic features. However, as also shown by the results of Experiment 4, extraposed relative clauses pose parsing problems even in the absence of any morpho-syntactic ambiguity when they have to be associated with the subject across an intervening object.

19 Implicit Prosody and Relative Clause Extraposition

The experiments reported in this chapter have investigated how readers process extraposed relative clauses in German. The most important findings can be summarized as follows. First of all, sentences with extraposed relative clauses were difficult to comprehend when the extraposed relative clause had to be construed with the subject across an intervening object. Processing difficulties were greatest for locally ambiguous sentences in which the subject did not provide any hint as to the existence of an upcoming extraposed relative clause, that is, sentences with a definite NP that was not preceded by a focus particle. When a focus particle was introduced before the subject, sentences became easier to process, although there were still signs of substantial difficulties. Finally, when the subject was introduced by a determiner that usually requires a relative clause, processing was eased further, but was still on a somewhat poor level.

For sentences in which the extraposed relative clause was associated with the object, processing was easy as long as there were no indications that relative clauses should be associated with the subject. When the subject was preceded by a focus particle, performance for sentences with object-modifying relative clauses was reduced, but still better than for sentences with subject-modifying relative clauses. Only when the subject started with the determiner *diejenige* or *derjenige* were sentences with a subject-modifying relative clause judged better than sentences with an object-modifying relative clause.

In sum, the results of the four experiments on relative clause extraposition presented in this chapter provide further evidence for the IPH of Fodor (2002). In particular, the results show that the assignment of sentence stress does not only constrain garden-path recovery, as captured by the PCR of Bader (1994), but that the assignment of sentence stress also affects decisions made during first-pass parsing, as claimed by the IPH. By directing sentence accent onto the following determiner, a focus particle caused readers to associate the extraposed relative clause with the

subject during the first-pass. This resulted in a better performance for sentences with an extraposed subject-modifying relative clause and, crucially, to a worse performance for the alternative structure with an object-modifying relative clause. The latter finding clearly shows that the insertion of a focus particle did not just ease reanalysis, but that it affected the initial parse of the extraposed relative clause.

What are the implications of the results presented here for the more general question of what factors govern whether a relative clause is extraposed or not? So far, the only determinant of relative clause extraposition that we have considered was the material intervening between the NP containing the relative clause and the clause final verb. In (26), this material is abbreviated as *(XP)**.

(26) a. C° … [NP [relative clause]] (XP)* V°

b. C° … [NP] (XP)* V° [relative clause]

As can be seen in (26), extraposition of the relative clause is advantageous because it shortens the distance between the head NP and the clause-final verb. The advantage of extraposition is the greater, the longer the relative clause is. The drawback of extraposition is that the relative clause and its head noun get separated from each other. The distance crossed by an extraposed relative clause includes (XP)* and the clause-final verb(s). Thus, the cost of extraposition increases with increasing complexity of (XP)*.

When distance is measured simply in number of words, relative clause length and extraposition distance should be equally important. Both corpus and experimental data show, however that relative clause length is much less important than extraposition distance for deciding whether to extrapose or not. This is not compatible with theories of weight in which dependency relations of different types trade-off in terms of word-based distance (Hawkins 2004; Temperley 2007)[3], but it is in agreement with the results presented in this chapter.

In the experiments presented here, extraposition had to cross an average distance of about five words. The relative clauses had a mean length of about six words. Thus, there is even a slight advantage for extraposition in terms of distance measured in words. The experimental results show, however, that the disadvantage brought about by extraposition was in no way offset by the advantage that results from making the dependency between head noun and verb shorter by moving the relative clause to the end of the sentence. The comprehension difficulties observed in the experiments reported above, thus clearly show that extraposition distance is more important than relative clause length.

With regard to the question of what factors govern the decision to extrapose or not, the current results provide evidence against the assumption that extraposition is just a matter of weight. The length of the two dependencies that are involved in extraposition—the dependency between head noun and relative pronoun and the dependency between head noun and clause-final verb—was not affected by the exper-

[3] Note that this issue does not apply to Hawkins' theory as presented in Hawkins (1994). Although number of words also play a role in this theory (in the definition of Constituent Recognition Domains), the relationship between order preferences and word-based distance is not as direct as in the theories mentioned in the text.

imental manipulations because all experimental manipulations involved material preceding the head noun (the presence or absence of a focus particle, the particular type of determiner). The experimental effects observed in this chapter are thus independent of weight. Because the experimental manipulations varied whether the subject NP, which hosted the relative clause, was focused or not, the experimental results can be taken as evidence in favor of Shannon's (1992) information structural account of extraposition. As predicted by this account, extraposition from the subject across an intervening object was easier to process when the subject was explicitly marked as a focus.

The experimental results are also compatible with a prosodic account along the lines of Féry (Extraposition and Prosodic Monsters in German). Because focusing of the subject implies defocusing of the object, extraposition is no longer blocked by an intervening accented object noun when the subject is focused. Since intervening accented nouns are the major obstacle to extraposition in the prosodic account, the finding that extraposition from a focused subject is easier than extraposition from a non-focused subject is also compatible with a prosodic theory of extraposition.

Even if the present results cannot decide between an information-structural and a prosodic approach to extraposition, they still show that information structure and/or prosody have a crucial impact on extraposition. This is not to deny that weight also has an important place in a comprehensive theory of extraposition. Corpus studies and experiments (Hawkins 1994; Konieczny 2000; Uszkoreit et al. 1998) have found that both extraposition distance and relative clause length affect the placement of relative clauses. Thus, as shown by Arnold et al. (2000) for heavy-NP shift in English, it seems most likely that relative clause extraposition in German is subject to different types of constraints.

Given the substantial processing difficulties observed for relative clauses extraposed from the subject across the object—even in unambiguous cases—one may wonder why sentences in which a relative clause is extraposed across intervening material that includes more than just the clause-final verb(s), are still produced with some regularity.

In the literature on sentence complexity, two major reasons have been proposed to make the computation of a dependency between two items X and Y difficult. The first one is the amount of referential processing that goes on between X and Y (Gibson 2000). The second is the interference that can result if an item similar to X intervenes between X and Y (van Dyke and Johns 2012). If neither of these two main reasons for sentence complexity applies, extraposition can be easy even across intervening material.

As the two representative examples in (27) and (28) show, extraposition across material that is less complex than an object and less similar to the head NP of the relative clause is indeed easy.

(27) … dass der Handwerker **bereits hier** angerufen hat, der die Fenster reparieren soll.
that the craftsman already here called has that the windows repair shall
"… that the craftsman that has to repair the windows already called."

(28) … dass kein Handwerker **zur Verfügung** stand, der die Fenster reparieren kann.
that no craftsman for availability stood that the windows repair can
"… that no craftsman that can repair the windows was available."

In both (27) and (28), the extra material in front of the clause-final verb consists of two words, as in all experimental sentences investigated in Experiments 1–4. Despite the equal number of words, extraposition seems to be much easier in these examples. In (27), only adverbials intervene. In (28), there is a noun, but this noun is non-referential and part of a quasi-idiomatic expression (*zur Verfügung stehen*, lit. "to stand at availability"—"to be available"). Thus, computing the dependency between the head noun and the relative clause is easy because there are no intervening referential expressions and no potential alternative relative clause hosts which could cause interference.

In sum, the comprehension results for extraposed relative clauses reported in this chapter provide further evidence that the main determinant of extraposition is the material that has to be crossed by extraposition. The results also indicate that the specific type of the intervening material matters, not just its length in number of words. An intervening object NP in particular makes the association of an extraposed relative clause with the subject difficult, even if the object is excluded as a potential attachment site for morphological reasons.

20 Conclusion

According to the IPH of Fodor (2002), the implicit prosody that is an integral part of the reading process can influence the parser's decisions during first-pass parsing. While effects of this sort have already been shown for prosodic phrasing, the research reported in this chapter addressed the question of whether the same also holds for the assignment of sentence stress. Can the stress pattern assigned to a sentence during reading also influence how syntactic ambiguities are resolved during first-pass parsing? For the case of relative clause extraposition, this chapter provides positive evidence with regard to this question. Experiments showed that the association site that the parser chooses for an ambiguous extraposed relative clause is affected by the implicit prosody assigned before encountering the relative clause. With default prosody, readers almost always associated the relative clause with the most recent NP, the object in the sentences under investigation. Introducing a focus particle in front of the more distant subject NP enhanced the probability that readers associate the extraposed relative clause with the subject. Performance was nevertheless still poor for sentences with subject-modifying relative clauses, but, as shown by further experiments, processing of such sentences remained difficult even after all ambiguity had been removed.

References

Arnold, J. E., Ginstrom, R., Losongco, A., & Wasow, T. (2000). Heaviness vs.newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76,* 28–53.
Bader, M. (1994). *Syntax und Prosodie beim Lesen.* Stuttgart: Universität Stuttgart.

Bader, M. (1998). Prosodic influences on reading syntactically ambiguous sentences. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in sentence processing* (pp. 1–46). Dordrecht: Kluwer.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalysis of garden-path sentences. *Journal of Memory and Language, 30*(6), 725–745.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. *Proceedings of the North East Linguistic Society, 32,* 113–132.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain. Papers from the first Mind Articulation Project Symposium* (pp. 95–126). Cambridge: MIT Press.

Gibson, E., Pearlmutter, N., Caseco-Gonzales, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59,* 23–59.

Gildea, D., & Temperley, D. (2010). Do grammars minimize dependency length? *Cognitive Science, 34,* 286–310.

Gorrell, P. (1995). *Syntax and parsing*. Cambridge: Cambridge University Press.

Haider, H. (2010). *The syntax of German*. Cambridge: Cambridge University Press.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hopf, J.-M., Bayer, J., Bader, M., & Meng, M. (1998). Event related brain potentials and case information in syntactic ambiguities. *Journal of Cognitive Neuroscience, 10,* 264–280.

Just, M. A., Carpenter, P. A., & Wolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111,* 228–238.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29,* 627–645.

Nespor, M., & Vogel, I. (1989). On clashes and lapses. *Phonology, 6,* 69–116.

Pickering, M. J., & van Gompel, Roger P. G. (2006). Syntactic parsing. In M. Traxler & M. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 455–503). New York: Academic Press.

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: The University of Chicago Press.

Schafer, A., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and cognitive processes, 11,* 135–163.

Shannon, T. F. (1992). Toward an adequate characterization of relative clause extraposition. In I. Rauch, G. F. Carr, & R. L. Kyes (Eds.), *On Germanic linguistics: issues and methods* (pp. 253–281). Berlin: de Gruyter.

Speer, S. R., Kjelgaard, M. M., & Dobroth, K. M. (1996). The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research, 25,* 249–271.

Stolterfoht, B., Friederici, A. D., Alter, K., & Steube, A. (2007). Processing focus structure and implicit prosody during reading: Differential ERP effects. *Cognition, 104,* 565–590.

Temperley, D. (2007). Minimization of dependency length in written English. *Cognition, 105,* 300–333.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition, 1*(2), 147–165.

Uszkoreit, H., Brants, T., Duchier, D., Krenn, B., Konieczny, L., Oepen, S., & Skut, W. (1998). Studien zur performanzorientierten Linguistik Aspekte der Relativsatzextraposition im Deutschen. *Kognitionswissenschaft, 7,* 129–133.

van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass, 6*(4), 193–211.

Wagner, M., & Watson, D. G. (2010). Experimental and theoretical advances in prosody: A review. *Language and cognitive processes, 25*(7–9), 905–945.

Warner, J., & Glass, A. L. (1987). Context and distance-to-disambiguation effects in ambiguity resolution: Evidence from grammaticalicity judgements of garden path sentences. *Journal of Memory and Language, 26*(6), 714–738.

Warren, P., Grabe, E., & Nolan, F. (1995). Prosody, phonology, and parsing in closure ambiguities. *Language and cognitive processes, 10,* 457–486.

Watt, S. M., & Murray, W. S. (1996). Prosodic form and parsing commitments. *Journal of Psycholinguistic Research, 25,* 291–318.

Webelhuth, G., Sailer, M., & Walker, H. (Eds.). (2013). *Rightward movement from a cross-linguistic perspective*. Amsterdam: Benjamins.

Prominence in Relative Clause Attachment: Evidence from Prosodic Priming

Sun-Ah Jun and Jason Bishop

Abstract This chapter presents two experiments utilizing prosodic adaptations of the structural priming paradigm. In each experiment, the goal was to explore the relation between the location of a prosodic boundary and the preferred parsing of a relative clause (RC) with ambiguous attachment to a preceding head noun. In Experiment 1, using read materials, ambiguous target sentences were preceded by prime sentences with RCs of different length: long, medium, and short. RC length was hypothesized to influence the location of an implicit prosodic boundary in the primes. However, no effect for this RC-length manipulation was found. In Experiment 2, the location of a boundary was manipulated in overt (spoken) prime sentences. For these auditorily-presented primes, the location of a prosodic boundary was found to influence attachment preference for targets. Interestingly, the effect was in the opposite direction as predicted: In the configuration NP1 NP2 RC, a boundary after NP2 resulted in *more* NP2 attachments. We propose that in the experimental materials, which contained equivalent accents on the two noun phrases (NPs), the boundary after NP2 leads to the accent on NP2 being interpreted as the nuclear pitch accent. Consequently, that accent was perceived as being more prominent than the accent on NP1, thus attracting RC attachment. The results suggest a close relationship between prosodic phrasing and prosodic prominence in English, and demonstrate a role for both in sentence processing.

Keywords Explicit prosody · Prosodic priming · Working memory · Autistic traits · Autism Spectrum Quotient (AQ) · Head prominence · Edge prominence

S.-A. Jun (✉)
Department of Linguistics, University of California, Los Angeles, Box 951543, Los Angeles, CA 90095-1543, USA
e-mail: jun@humnet.ucla.edu

J. Bishop
Linguistics Program, City University of New York (College of Staten Island and The Graduate Center), New York, NY, USA
e-mail: jbishop@gc.cuny.edu

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,* Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_12

1 Introduction: The Implicit Prosody Hypotheses and Sentence Processing

In a sentence such as (1), it is ambiguous whether the relative clause (RC) modifies NP1 *the servant* (high attachment) or NP2 *the actress* (low attachment):

(1) *Someone shot the servant of the actress who was on the balcony.*

Although attachment possibilities are, strictly speaking, ambiguous in such cases, it is now well known that readers show preferences towards either high or low attachment in silent reading tasks. Furthermore, the direction of this preference is in part predictable based on native language, although the divisions defy reasonable typological distinctions. For example, while native English speakers prefer low attachment (Frazier and Clifton 1996; Ehrlich et al. 1999), Dutch and German speakers favor high attachment (Hemforth et al. 1998; Brysbaert and Mitchell 1996); and while Spanish speakers prefer high attachment (Cuetos and Mitchell 1988; Carreiras and Clifton 1993), Romanian speakers prefer low attachment (Ehrlich et al. 1999). It is also known that attachment preference can vary within a language as a function of experimental task (e.g., Augurzky 2006; Sekerina et al. 2004), and especially as a function of phonological weight (which we describe in more detail below).

In the present study, we are interested in a particular and very influential theory about the source of these preferences, namely the Implicit Prosody Hypotheses (IPH) proposed by Fodor (1998, 2002). While acknowledging that some individuals are claimed to lack subvocal prosody in reading, Fodor (2002) describes the IPH as in (2), which is accompanied by the two assumptions in (3):

(2) **The IPH (Implicit Prosody Hypothesis):**
In silent reading, a default prosodic contour is projected onto the stimulus, and it may influence syntactic ambiguity resolution. Other things being equal, the parser favors the syntactic analysis associated with the most natural (default) prosodic contour for that sentence.

(3) **Working Assumptions**:
a. The comprehender will be more likely to postulate a large syntactic boundary at the location of a large prosodic boundary.
b. The implicit prosody projected onto the sentence in reading will be identical to the explicit (i.e., overtly spoken) prosody for that sentence in a comparable context.

Of particular interest in the present study is the assumption in (3b), which pertains to the accessibility of implicit prosody to empirical investigation. As we discuss below, recent studies have produced findings that are problematic for this assumption, leading to the question of how implicit prosody can be studied, and therefore also how the IPH can be tested. In the rest of this chapter, we pursue this issue as follows. In Sect. 2, we discuss recent studies of implicit prosody and the problems associated with equating it to explicit prosody. Then, in Sects. 3 and 4, we present two novel experiments, using adapted versions of the structural priming paradigm, which are intended to circumvent these problems. Finally, in Sect. 5, we discuss our results and their implications for the relation between prosody and sentence processing.

2 Testing the Implicit Prosody Hypothesis

2.1 Support

A useful starting point for assessing the claims of the IPH is to consider one of its most compelling applications: the phenomenon of length effects on attachment bias in silent reading. An early observation in the study of attachment preference was that simple constituent length was positively correlated with a preference for high attachment (e.g., Fernández and Bradley 1999; Quinn et al. 2000; Lovric et al. 2001; Fernández 2003; Jun 2003a). In speech it is understood that, other things being equal, the edges of longer constituents tend to require more phrase boundaries (e.g., Selkirk 2000; Nespor and Vogel 1986; Jun 1996, 2003a, b), and so length effects on attachment have a straightforwardly prosodic explanation. According to the IPH, therefore, it needs only to be assumed that this prosodic "chunking" of a long RC occurs implicitly during reading, and, according to (3a), a resulting implicit boundary before an RC prompts high attachment of that RC.

Further evidence that the IPH is a necessary component to an explanation of length effects comes from the findings of Swets et al. (2007). In their study, with Dutch and English native speakers, subjects were asked to read sentences such as (1), above, where the RC either occurred or did not occur on the same printed line as the rest of the sentence. Although English (unlike Dutch) is a low-attachment language, the authors found that readers of both languages preferred high attachment when the RC was read on a separate line. Additionally, Swets and colleagues found that a preference for high attachment was present in individuals with lower working memory capacity (indicated by a latent variable based on measures of both spatial and verbal working memory). The IPH is able to handle both of these findings neatly and in the same way: The insertion of an implicit prosodic boundary drives attachment as per (3a). In the first case, it need only be assumed that the implicit prosodic juncture is cued by the visual juncture; in the second case (and as pointed out by Swets and colleagues), it can be assumed that individuals with lower working memory are more likely to insert an implicit boundary before the RC so that the sentence is chunked into smaller, more manageable processing units. This boundary, in turn, encourages closure at that location (i.e., high attachment).

2.2 Problems

Patterns of attachment preference related to length, which occur within language, thus seem to require reference to implicit prosody, and in this way motivate the IPH's explanation for the cross-linguistic asymmetries as well. According to the IPH (Fodor 2002, p. 123), languages in which native speakers exhibit an overall high attachment bias do so because they are, by default, inserting an implicit boundary between the RC and the adjacent head noun, but not between the two head nouns. In languages where speakers exhibit an overall low-attachment bias, however, it is

assumed that this default implicit boundary occurs instead between the two head nouns and not between the RC and the adjacent head noun (thus grouping the lower NP together with the RC). According to (3b), we therefore expect to observe this very same cross-linguistic asymmetry in the placement of explicit boundaries in speakers' productions of the same sentences. In fact, production studies have been carried out for both Japanese and Korean languages with the syntactic configuration of RC preceding the head nouns (Jun and Koike 2003 for Japanese; Jun and Kim 2004; Jun 2007 for Korean). Because these languages are both high-attachment languages, it is assumed they are assigned the implicit prosodic structure (RC)//(NP's NP), and in both cases, the strongest prosodic boundaries were indeed found to be directly after the RC, supporting the IPH.

In more recent work, however, the analogous question has been asked for English, which, as described above, is a language with a low-attachment bias. In English, where the head nouns are ordered as NP1 and NP2 and precede the RC, the IPH predicts the largest implicit—and under assumption (3a) also explicit—prosodic boundary to occur between the two nouns (NP1)//(NP2 RC) in the same sort of out-of-the-blue readings. However, this was found not to be the case by Bergmann and Ito (2007), Bergmann et al. (2008), or Jun (2010; see also Jun and Shilman 2008), the latter of which was a large-scale production study. Instead, these authors all report native English speakers to place the largest prosodic boundary in the sentence directly before RC—just as the speakers of Japanese and Korean did. Furthermore, even when sentences contained grammatical and sematic information favoring low attachment (and speakers were given a chance to read the sentence carefully before reading it aloud), this late-occurring boundary was nonetheless the preferred pronunciation. Thus, although initially promising, production evidence now suggests that speakers, in out-of-the-blue readings, prefer a large prosodic break between the RC and the head noun regardless of attachment differences—and so fails to support the IPH's explanation for cross-linguistic differences in attachment biases.

2.3 Accessing Implicit Prosody

The length effects discussed above make a strong case for the existence of implicit prosody, and its influence on sentence processing. The production findings just described, however, indicate that, if implicit prosody functions as Fodor and colleagues claimed, the use of explicit prosody may not be a reliable way to investigate it. Though unfortunate from the perspective of methodological convenience (and in conflict with Fodor's assumption 3b), this does not necessarily constitute strong evidence against the basic insight of the IPH. Indeed, in discussion of her findings for English, Jun (2010) reviews evidence showing that, if the goal is to obtain speech that maximally encodes syntactic or semantic structure, the standard method of eliciting prosody from reading aloud may be inadequate. Instead, speakers may produce only a very basic "surface" prosody, in which fluency is the speaker's primary goal. Such "performance prosody" may differ significantly from prosody

produced in spontaneous speech, intentionally disambiguating speech, or, most pertinent here, the internal speech generated during silent reading. While further production research is needed to confirm whether a (NP1 NP2)//(RC) (or (RC)//(NP2 NP1)) phrasing is in fact the universally preferred out-of-the-blue production, there is sufficient reason to doubt that reading aloud is a well-suited way to determine the form of implicit prosody.

However, if the "surface" prosody obtained in reading aloud is insufficient for investigating "deeper" implicit prosody, how might we then approach this problem? Jun (2010) offers a number of possibilities that involve either encouraging readers to encode structure in their productions, or that utilize online processing methodologies that circumvent the production issue altogether. One of the possibilities that is not mentioned in Jun (2010), but which may be effective, is to instead influence implicit prosody more directly, such as through structural priming, and observe outcomes on sentence comprehension. The structural priming paradigm (e.g., Bock 1986), well known in syntactic processing literature, exploits the tendency of speakers and listeners to "reuse" recently encountered syntactic structures (see Pickering and Ferreira 2008 for a recent review). However, can prosodic structure also be primed?

This matter has been investigated in recent research, but we feel that the evidence is somewhat unclear at this point. Tooley et al. (2013) present a study with native English-speaking subjects who heard sentences such as "The dog that pawed the door needed to be let out." that contained either (a) no boundary, (b) a boundary in a "dispreferred"/marked location before the object, (c) a boundary in a "preferred" location after the object, or (d) a boundary in both locations. In one experiment, subjects were to listen to sentences with one of these prosodic structures, and repeat it back, i.e., to produce the same sentence overtly after listening. When this was the task, subjects tended to repeat the sentence with a boundary in the same location, suggesting there may have been some priming that took place (the authors argue that this is less well explained by simple phonetic repetition than abstract structure priming, and present evidence in support of their interpretation). However, in subsequent experiments in which subjects heard sentences and then reproduced a novel (but similar) one, the prosodic structure was not found to carry over to the novel sentence. This finding was interpreted by the authors as indicating that the abstract representation of prosodic structure can be primed (just as syntactic representations can be primed), but tends to be weaker, not lasting as long as syntactic priming has been found to. Possibly, according to Tooley and colleagues, this is due to the fact that prosody is subject to a larger number of constraints than is generally assumed to be the case for syntax.

As discussed earlier, however, the act of producing sentences from reading can often result in prosody that is more focused on fluency than on the encoding of structure. It is thus possible that the priming effects found by Tooley et al. (2013) were obscured by this fact. The question we wish to ask here is whether priming effects can be observed in implicit prosody, which, in principle, should be less affected by the difficulties involved in eliciting overt productions. In the first experiment presented below, we attempt to exploit the reliable relation between

boundary placement and constituent length (discussed in Sect. 2.1) to explore whether implicit prosody can prime implicit prosody, and whether we can observe its effects on the parsing of ambiguous RCs. In a second experiment, we attempt a slightly different and more direct approach, using explicit prosody to prime implicit prosody. In both cases, the goal is to examine whether RC attachment in English can be influenced by factors that we can attribute only to the location of a prosodic boundary, and whether the result is as predicted by the IPH.

Another factor we considered is the role played by what are sometimes collectively referred to as "cognitive processing styles." These include the aspects of information processing such as working memory capacity and certain personality traits. As discussed in Sect. 2.1, verbal working memory is known to influence attachment bias specifically, and so there was an obvious motivation for collecting this information about participants in our experiments. A second and less common measure of processing style involve "autistic" traits, which, being less commonly studied, we describe in further detail before proceeding.

Autistic traits are behaviors and patterns of information processing associated with a clinical diagnosis with Autism Spectrum Disorder. However, such traits—for example, non-holistic attentional focus, lack of social engagement, and poor communication skills—are known to occur to varying degrees in the neurotypical population as well. These traits are measured in nonclinical individuals using the Autism Spectrum Quotient (AQ; Baron-Cohen et al. 2001), a nondiagnostic, self-administered questionnaire (requiring agree/disagree responses) that divides autistic traits into five separate dimensions pertaining to *social skills, attention to detail, attention switching abilities, communication skills,* and *imagination*. Studies have shown that the AQ, which is scored such that higher scores indicate more autistic traits, has a high level of cross-cultural validity (Wakabayashi et al. 2006; Hoekstra et al. 2008; Ruta et al. 2011; Sonié et al. 2012), although there may be some variation related to culture on the imagination and attention switching subscales (Freeth et al. 2013).

Of primary relevance to the phenomena of interest here is the communication subscale (henceforth AQ-Comm), which contains items such as "I know how to tell if someone listening to me is getting bored." and "Other people frequently tell me that what I've said is impolite, even though I think it is polite." Intuitively, compared with the other subscales, AQ-Comm items relate most to "Theory of Mind" (Premack and Woodruff 1978), or, roughly, the ability to attribute and understand the thoughts and intention of others. Consistent with this, high AQ-Comm scores (indicating poorer, more autistic-like communication skills) are known to be negatively correlated with the use of pragmatic inference in sentence processing in both online and offline tasks (Nieuwland et al. 2010; Xiang et al. 2013; see also Xiang et al. 2011). Crucial to the present study, AQ-Comm has been shown to predict the online interpretation of prosody by Bishop (2012a; see also Bishop 2013) who found individuals with high AQ-Comm to exhibit weaker sensitivity to prosodic prominence in a cross-modal semantic priming task. In particular, although native English speakers are known to prefer high relative prominence on the object in subject–verb–object (SVO) constructions when semantic focus is narrowly on that object (e.g., Bishop 2012b; Breen et al. 2010; among others), this was not replicated

for high AQ-Comm individuals, who actually showed the reverse preference. While we lack a detailed understanding of how these individual differences influence prosodic perception and sentence processing, we wished to control for such influences here, since they are highly relevant to our task. Therefore, in both of the experiments we present below, participants completed standard measures of both autistic traits and verbal working memory.

3 Experiment 1: Implicit-to-Implicit Prosodic Priming

As described in Sect. 2.1, the effect of RC length on attachment preference has a straightforwardly prosodic explanation: the longer the RC, the more likely a reader is to insert an implicit prosodic boundary before the RC, in turn increasing the probability of a high-attachment parsing. In Experiment 1, we explore whether reading a prime sentence with a long RC, predicted to be more likely to contain a prosodic boundary before that RC, induces a boundary before the RC in a subsequently presented novel sentence. If implicit prosody can be primed in this way, the IPH predicts that its effects should be observable in the comprehension of the novel sentence. In the present case, the priming of a boundary should result in the increased likelihood of a high-attachment parsing of the RC.

3.1 Methods

Stimuli

Sentences, to be used as primes and targets, were designed for a reading experiment. Sixteen sentences containing RCs of medium (6–7 syllables) length were first selected as targets, and were intended to lack any grammatical (e.g., agreement) or semantic information that would favor high or low attachment. To this end, these targets, such as "Someone shot the servant of the actress that was on the balcony." were based on sentences used in previous studies (Frazier 1990; Frazier and Clifton 1996; Felser et al. 2003; Dussias 2003). The prime sentences, to be presented and read immediately before the target sentences, were based on 15 sentences structurally similar to the targets. These basic sentences were used to create 30 total prime sentences, each of different RC length: short (2–4 syllables), medium (6–7 syllables), and long (9–12 syllables). Each of the 15 basic sentences had two versions, each in a different length condition. The full list of prime and target sentences can be found in Appendix A.

In addition to the experimental target and prime sentences, 24 filler targets and 30 filler primes were also designed. In order to reduce the difficulty of the task, 18 filler targets contained either no RC or an RC with one head noun. The remaining filler targets contained RCs and two head nouns, but of these only three had poten-

tial attachment ambiguity. The primes designed for filler trials, however, were the same as the primes used on experimental trials, in that they all had an RC with two head nouns and had the same three RC length conditions. Each RC length condition included ten primes, but only four of them had potential ambiguity for RC attachment and the rest had an attachment bias either toward NP1 or NP2.

Procedure

A MATLAB script was used to present participants with the primes, targets, and the question that elicited their attachment decisions about the targets. Participants controlled the presentation of the sentences by reading at their own pace. Once a sentence appeared on the computer screen, the participant was to read the sentence silently, and then push a key, which removed that sentence and brought up the next. For each experimental trial, the MATLAB script selected one of the 18 target sentences and 3 prime sentences, all from the same RC-length condition, and presented them randomly. The decision to present three primes on each trial was made in an attempt to induce stronger priming effects (cf. Tooley et al. 2013). The participant proceeded through these three sentences, then, finally the target, in the manner described above. Following the participant's key press after the target sentence, however, a question rather than a new sentence appeared. That question appeared below the target, and asked the subject the standard RC-attachment decision, presenting the two possible NPs as the options "A" and "B." Whether the high-attachment response (i.e., whether the first NP) appeared on the left as "A" or on the right as "B" was counterbalanced. The participant's response was then collected and the next trial began.

Filler trials proceeded in a similar manner, with the following exceptions. First, as described above, the filler targets did not always contain RCs. Second, participants were, at random, required to answer a question about filler primes; this was done to prevent participants from knowing exactly which sentence in a trial (i.e., every fourth sentence presented to them on test trials) would be the one requiring an attachment decision. Participants carried on through all experimental and filler trials (randomized for each participant), and the assignment of a test item to a particular RC-length priming condition was counterbalanced across subjects.

Following the main reading and comprehension task, all participants completed the AQ and an automated version (Unsworth et al. 2005) of Daneman and Carpenter's (1980) reading span task, a widely used measure of verbal working memory capacity. The entire experiment lasted approximately 50 min.

Participants

Participants were 102 (68 female, 34 male) native speakers of American English. They were students at the University of California, Los Angeles (UCLA) and received either monetary compensation or course credit. All participants confirmed that they were never diagnosed with a communication disorder, and all had normal or corrected vision.

3.2 Results

We analyzed attachment decisions that were given by subjects within two standard deviations of the group's mean response time. Mixed-effects logistic regression (using the *glmer* function in the *lme4* package (Bates et al. 2013) of *R* (R Development Core Team 2013)) was used to model the probability of participants' "high-attachment" responses, with participant and item modeled as random effects. In addition to the experimental manipulation (length of the RC in the prime sentences), the fixed effects also included several stimulus- and participant-based factors. Stimulus variables included the length (in syllables) of the NPs and the RC in the target sentences, the order of the presentation of answers (i.e., NP1 appears on left or right), and experimental trial. Participant-level variables included gender (self-identified), reading span score (henceforth RSPAN), and AQ-Comm scores (calculated as a four-point Likert-scale; see Baron-Cohen et al. 2001). A preliminary model contained all of these predictors, and two-way interactions between prime condition and each of the other predictors. Then, predictors with a p value larger than 0.1 were removed if this did not result in a significant decrement to the fit of the model as determined by a likelihood-ratio test (e.g., Baayen 2008). The simplest, best fitting model was the one retained.

The results of the final model are shown in Table 1. There was a nonsignificant tendency for participants to give fewer high-attachment responses when the question ordered the NPs in the opposite order that they appeared in the sentence. Although there was no effect for AQ-Comm (as indicated by its high p value in the first round of modeling), RSPAN had a significant main effect with a negative coefficient value, indicating that participants with lower RSPANs were more likely to give high-attachment responses overall. This effect, consistent with Swets and colleagues' (2007) findings—and also consistent with the Implicit Prosody Hypothesis—is shown in Fig. 1. There was, however, no effect for priming condition; attachment preferences did not differ significantly depending on the length of the RC in the prime sentences. Indeed, as can be seen in the model's coefficients, there was a trend in the opposite direction in the Long RC condition, inconsistent with the IPH. That is, relative to the high-attachment response rate following the primes containing medium-length RCs (our control condition), participants were numerically less likely to give a high-attachment response after reading primes with long RCs.

Table 1 Estimates, standard errors, z and p values for Experiment 1. Positive estimates indicate the amount of increase in log-odds relative to the Intercept. For each categorical predictor, the change from the intercept is for the value given in parentheses

	β	SE (β)	z	p
(Intercept)	0.47			
Order (NP1 = left)	0.622	0.377	1.65	0.099
RSPAN	–0.025	0.011	–2.30	**0.022**
Prime length (long)	–0.247	0.152	–1.63	0.104
Prime length (short)	0.082	0.149	0.55	0.580

RSPAN reading span score, *NP* noun phrase

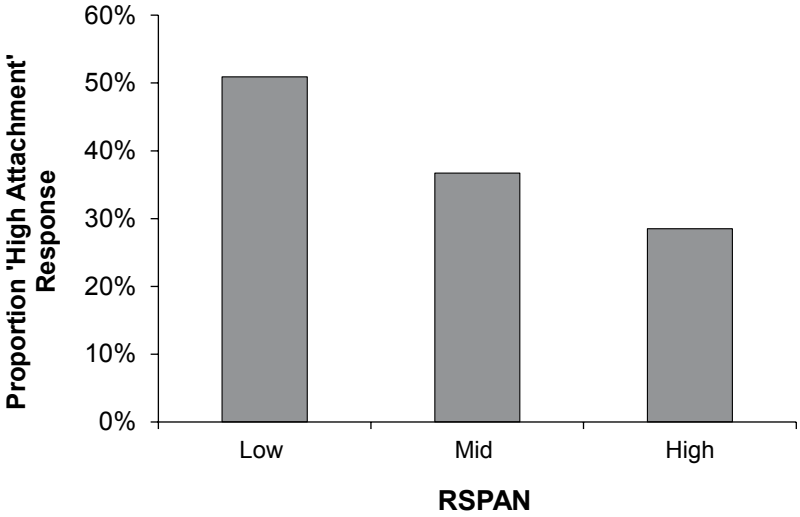


Fig. 1 Proportion of high-attachment responses as a function of participants' reading span score. The three levels refer to the group distribution, "Mid" being those subjects scoring within 1 standard deviation (SD) of the mean, the "Low" and "High" below or above 1 SD.

3.3 Discussion

The goal of Experiment 1 was to test whether prime sentences with longer RCs, which should be more likely than sentences with shorter RCs to contain an implicit prosodic boundary, influenced how participants parsed the RC in novel target sentences. If the implicit boundary were present in the prime sentences, and carried over to the novel sentence, we predicted this would result in a greater probability of the RC's being attached high in the target. However, we are faced with a null result from the experiment with respect to our manipulation (the only significant finding being the main effect for RSPAN, a replication of Swets et al.'s (2007) result); indeed, there was a trend in the opposite direction than predicted, with targets tending towards high attachment less often following long primes.

While it is possible that no priming was found in Experiment 1 because no implicit prosodic boundaries were in fact generated in the primes, we find this possibility unconvincing, for the reasons laid out in Sect. 2.1. Instead, it is more likely that priming did not occur either because (a) prosodic structure priming is inherently weak, as suggested by Tooley et al. (2013) or (b) there was something about our task which obscured or further weakened priming that would otherwise have been observable. We believe that both of these factors may be responsible.

Our intent was to manipulate what has been the most reliable predictor of both prosodic boundary placement and high-attachment preferences, namely RC length. However, it may be the case that the 9–12-syllable length for the long-RC condition

was excessively long, making the modest 6–7-syllable-long RCs in the target sentences seems short in comparison. It may thus be that participants in the reading task, after reading a prime in the long-RC condition, treated targets as if they were short RCs, thus projecting no boundary before the RC. If this were the case, we would expect to see an effect in the opposite direction. In fact, this is what was found, although the trend was insignificant ($p = 0.10$).

Additionally, since Tooley et al. (2013) found prosodic priming to be weak (possibly too weak to be observed in truly novel sentences at all), it may be that a simple reading task does not result in sufficiently salient prosodic structures. In Experiment 2, we attempted to circumvent both of the possible pitfalls of the task used in Experiment 1 by presenting subjects with auditory primes—i.e., explicit prosody. The goal in Experiment 2 was therefore to test whether hearing a sentence with a prosodic boundary in a certain location influenced the comprehension of silently read sentences, and if the predictions of the IPH are useful in understanding any patterns.

4 Experiment 2: Explicit-to-Implicit Prosodic Priming

4.1 Methods

Stimuli

The basic design of sentence materials for Experiment 2 was similar to those for Experiment 1. For Experiment 2, targets were 18 sentences intended to lack any kind of biasing information, grammatical, or otherwise, and were again directly used or were modified version of sentences that have been normed in previous studies. Unlike the target sentences used in Experiment 1, however, these 18 target sentences contained RCs that were of one of two different lengths: shorter (3–5 syllables) or medium length (6–7 syllables).

The purpose of the primes for Experiment 2 was to deliver different explicit prosodic structures to participants, and so auditory prime sentences were created. The 16 prime sentences were, as other sentences in this study, based on those used in previous studies, and contained RCs (short, 3–5 syllables) with ambiguous attachment to a preceding NP. An example of one such prime was “The chef couldn’t find the lid of the pan that was clean” (see Appendix B for the full list); three versions of such primes were produced and recorded by a native speaker of English from California with extensive training in intonational phonology; each version was intended to manipulate the location of a prosodic boundary as shown schematically in (4). Example f0 contours for a prime in each of the three conditions are shown in Fig. 2. The control condition contained no large prosodic boundary—i.e., the prosodic juncture between each pair of words in the sentence was equal, i.e., by having a break index 1 in English ToBI (Beckman and Hirschberg 1994).

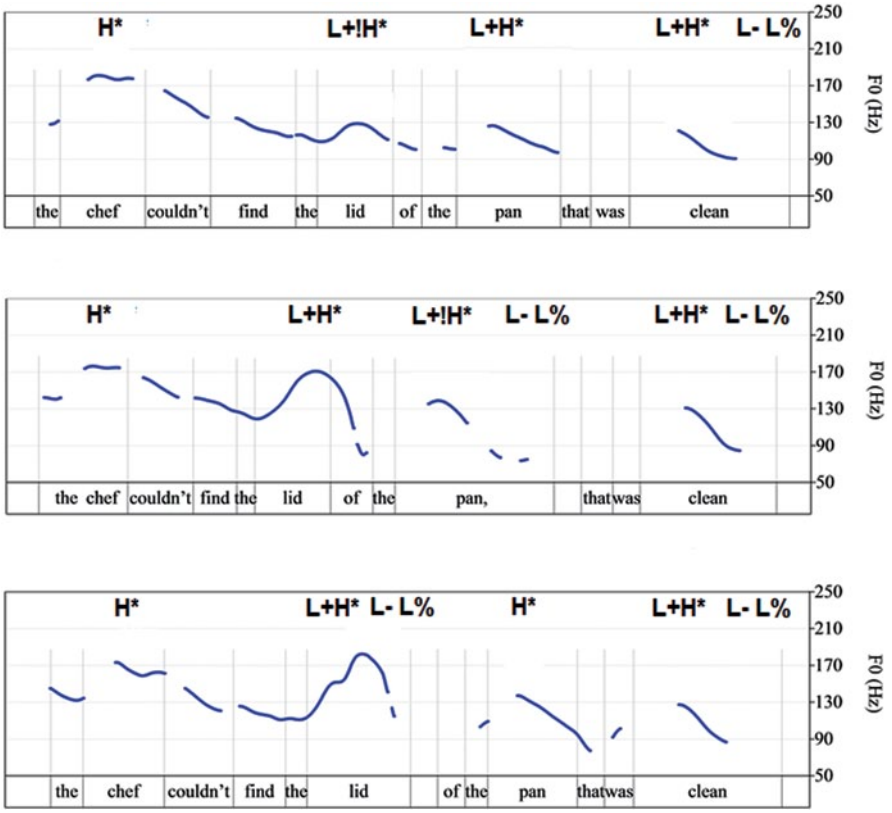


Fig. 2 Example of an auditory prime for Experiment 2, in each of the experimental conditions: no boundary/control (*top*), late boundary (*middle*), and early boundary (*bottom*). Accent status (accented/unaccented) was held constant across boundary conditions

(4) (a) boundary condition (control)
(…NP1 NP2 RC)
(b) Late boundary condition
(…NP1 NP2)//(RC)
(c) Early boundary condition
(…NP1)//(NP2 RC)

The boundaries used were the largest in the intonational phonological model for English, namely the Intonational Phrase (Pierrehumbert 1980; Beckman and Pierrehumbert 1986) and were marked by low f0 targets in each case (the L-phrase accent and L% boundary tone in the English ToBI conventions; Beckman and Hirschberg 1994). Additionally, accent status was held constant across the conditions such that pitch accents of the type (L+)H* (downstepped on NP2 in the late boundary condition) occurred on each of the two NPs, and one single-pitch accent occurred in the RC. Filler

trials included sentences with RCs that had unambiguous attachment (i.e., there was only one head noun, or there was only one semantically plausible head noun), as for Experiment 1; filler primes, however, were designed to manipulate the location of an intonational phrase boundary in sentences lacking any RC. For example, filler primes were of the form "Jackie telephoned Paul between lunch and dinner." which (on different trials) contained either no boundary, a boundary after the subject, or a boundary after the direct object.

Procedure

A MATLAB script was used to present participants with the prime sentences, target sentences, and the questions that elicited their attachment decisions about the targets. Each trial proceeded as follows. First, three (different) prime sentences from the same boundary condition were presented, one after another with a short (0.5 s) interval between them. Then, following the offset of the third prime, a target sentence appeared on the screen. Finally, after 2.5 s, the standard attachment question appeared below the target sentence, and both the target and question remained on the screen until the participant selected an answer, which then began the next trial. Following this auditory priming task, participants completed the RSPAN task and AQ questionnaire; Experiment 2 took approximately 40 min to complete.

Participants

Participants were 120 (74 females, 46 males) native speakers of American English. They were undergraduate students at UCLA and received either monetary compensation or course credit. None had participated in Experiment 1, and all confirmed they had normal hearing, were never diagnosed with a communication disorder, and all had normal or corrected vision.

4.2 Results

The outcome variable "high-attachment response" was modeled as in Experiment 1, using the same factors, except that (a) the prime-related factor was boundary location rather than prime RC length, and (b) length of the RC in targets was a factor in Experiment 2 (Both boundary location and RC length were permitted to enter into three-way interactions with RSPAN and AQ scores in our models). The resulting model, following the procedures from Experiment 1, included the factors shown in Table 2.

Results showed the following; first, there was a main effect of RC length (in the target); consistent with previous work, target sentences containing shorter RCs were associated with fewer high-attachment responses ($p < 0.01$). There was also a robust main effect for the prosodic manipulation, i.e., the location of the prosodic break in

Table 2 Estimates, standard errors, *z* and *p* values for Experiment 2. Positive estimates indicate the amount of increase in log-odds relative to the Intercept. For each categorical predictor, the change from the intercept is for the value given in parentheses

	β	*SE* (*β*)	*z*	*p*
(Intercept)	0.164			
RC length (short)	–0.889	0.335	–2.651	**0.008**
Boundary (early)	–0.702	0.583	–1.203	0.229
Boundary (late)	–1.686	0.584	–2.886	**0.003**
AQ-Comm	0.033	0.036	–0.902	0.367
Boundary (early) × AQ-Comm	0.028	0.030	0.919	0.358
Boundary (late) × AQ-Comm	0.085	0.030	2.860	**0.004**

RC relative clause, *AQ* autism spectrum quotient

the prime sentences. The effect, however, was in the opposite direction predicted by the IPH: Overall, targets following primes with a late prosodic boundary were associated with a lower rate of high-attachment responses by participants (relative to the control condition; $p<0.01$). Finally, although the AQ was not a significant predictor as in Experiment 1, the best-fitting model for Experiment 2 included a two-way interaction between AQ-Comm and boundary location, indicating that the effect just described was modulated by AQ-Comm scores. In particular, the probability of participants giving a high-attachment response to targets following late boundary primes (i.e., the pattern predicted by the IPH) was directly related to AQ-Comm scores. That is, the pattern predicted by the IPH was present, but limited to those subjects with more autistic-like communication skills. This pattern can be seen in Fig. 3. Also apparent in the figure is a (nonsignificant) trend in the direction of high AQ individuals being less likely to interpret the RC as attaching high in a target if that target was read after hearing primes with a break between the two NPs (i.e., early boundary) compared to the control prime. Note that RSPAN is not included in the best-fitting model because it was not a significant predictor of attachment. Further, it did not correlate with AQ-Comm ($r=-0.02$, $p>0.33$).

4.3 Discussion

The purpose of Experiment 2 was to test for the influence of prosodic boundaries in auditory prime sentences on the parsing of RCs in silently read sentences. The results indicate that there was such an influence, but the details are highly surprising. First, as a group, after hearing RC-sentences with a prosodic boundary after the second of two NPs, participants were actually more likely to attach the RC low to NP2 in an analogous target sentence; this priming effect is exactly the opposite of what we predicted based on the IPH. However, the second finding was that a subset of participants was more likely to attach the RC high after hearing sentences with late-occurring prosodic boundaries (also showing a trend towards attaching low after hearing sentences with an early occurring prosodic boundary), in line with our IPH-based predictions. What is surprising about the second finding is that these

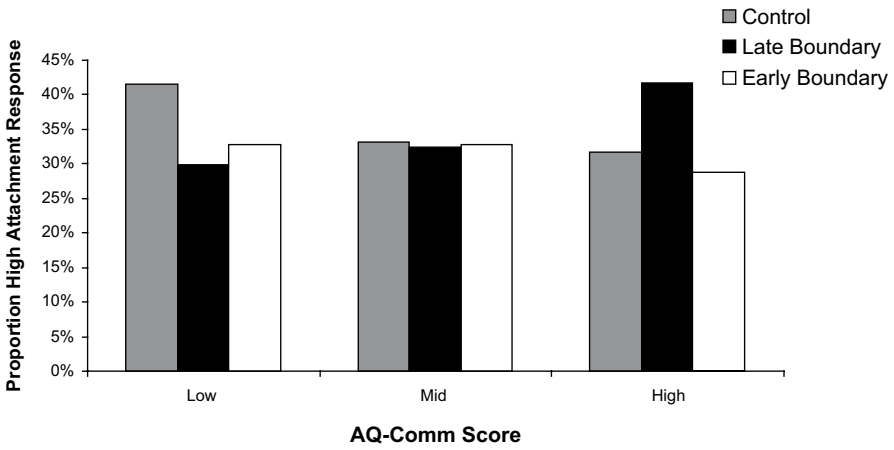


Fig. 3 Proportion of high-attachment responses for each boundary condition, shown for three groups of subjects according to their AQ-Comm score. The three levels refer to the group distribution: "Mid" is for those subjects scoring within 1 SD of the mean, and the "Low" and "High" levels are below or above 1 SD (the "mid" condition represents approximately 50% of the participants, and the other two groups approximately 25% each). Higher AQ-Comm scores reflect more prominent autistic traits in the communication dimension (the control condition in the figure represents primes without any prosodic break). *SD* standard deviation, *AQ* Autism spectrum quotient

were the participants with the most autistic-like communication skills, usually associated with lower sensitivity to prosody. The results are therefore quite puzzling from the perspective of the hypotheses we were testing.

We first address the primary finding, i.e., the main effect showing late boundaries to prime low attachment in targets, for which there are two basic logically possible interpretations. The first is that the prosodic boundary separating NP2 and the RC had the direct effect on syntactic parsing of cueing structural proximity, i.e., the lack of syntactic juncture. This interpretation is in such clear contradiction to previous research on the prosody–syntax relation as to render it untenable, and so we do not consider it further here. The second possibility, however, is that there was a cue relevant to attachment disambiguation concomitant with the location of prosodic boundaries in our design; we believe that such a correlate of our intended manipulation was in fact responsible, and that it was prosodic prominence.

Prosodic prominence, i.e., phrase-level accentuation, has been found previously to influence RC attachment decisions by Shafer et al. (1996; see also Lee and Watson 2011). In their study, using a traditional off line attachment decision task similar to ours, the authors presented participants with auditory sentences (as targets, not primes, as in our experiment) with a single prosodic boundary after NP2. In one experiment, it was shown that if only one of NP1 or NP2 contained a pitch accent, a strong bias was observed towards parsing the accented NP as the head of the RC. In a second experiment, holding accent location constant on NP2, the authors manipulated accent type, comparing a nuclear H* with a phonetically (and perceptually; see Turnbull et al. 2014) more prominent nuclear L+H*, and

found that NP2 with L+H* triggered more RC attachment. Thus, increasing the prominence of NP2 through either accent status (i.e., [± accent]) or accent type increased also its likelihood of being parsed as the head of the RC.

In our Experiment 2, the manipulation of prominence was an inevitable result of manipulating boundary location. Recall that in the auditory primes, accent status for NP1 and NP2 was held constant, as was the fact that a single-pitch accent occurred within the RC across phrasing conditions. This being the case, the location of boundaries determines which of the accent on the two NPs is nuclear and which is prenuclear, a distinction in structural prominence (Beckman 1986; Beckman and Edwards 1994), which is illustrated schematically in (5), with nuclear accents in bold:

(5) (a) T* **T*** **T*** (…NP1 NP2)//(RC) *Late boundary, stronger NP2 prominence*

(b) **T*** T* **T*** (…NP1)//(NP2 RC) *Early boundary, weaker NP2 prominence*

What must be taken away from (5) is that structural prominence favors attachment to exactly the opposite head noun as does phrasing (given the IPH's working assumption (3a), noted in Sect. 1). It therefore seems, we conclude, that for the majority of listeners in our Experiment 2, prominence trumped boundary location in cueing attachment. This understanding of the main effect for phrasing in Experiment 2 also allows for a more intuitive interpretation of our finding that phrasing is a better predictor for individuals with high AQ-Comm scores—previously equated to poor pragmatic processing (Nieuwland et al. 2010; Xiang et al. 2013). That is, our finding is not because these individuals were especially sensitive to phrasing for the purposes of syntactic parsing, but that they were *less* sensitive to prominence than the rest of the group. This is more in line with Bishop's (2012b, 2013) findings in cross-modal priming, described at the end of Sect. 2.3. While we speculate that high AQ-Comm may therefore be associated with poor use of prosodic prominence, we do not suggest that our results should be interpreted as suggesting these individuals have an enhanced or even typical sensitivity to phrasing, a matter which we comment on further below.

5 General Discussion

The purpose of the present study was to explore a basic prediction of the Implicit Prosody Hypothesis about how prosody influences syntactic disambiguation in silent reading. According to Fodor's (2002) conceptualization, a large prosodic boundary, implicitly projected onto a sentence, should cue a large syntactic boundary at that location. In two experiments, we attempted to test this relationship between phrasing and parsing; although we did not generate a statistically significant finding in Experiment 1, in which we attempted to use (length-induced) implicit prosodic boundaries, we did find that explicit prosodic boundaries in primes in Experiment 2 influence parsing. Surprisingly, however, for the majority of our subjects, the pat-

tern was in the opposite direction from what the IPH predicts. Nonetheless, having found prosody to be a significant predictor of ambiguity resolution in RC attachment, we do not believe that the results of that experiment invalidate the fundamental proposal embodied by the IPH. Rather, it suggests that it needs to be revised so as to take into account a wider range of prosodic structure; as discussed in the previous section, we believe we can appeal to patterns of prosodic prominence to explain the patterns of syntactic parsing observed in Experiment 2.

On this point, it must be emphasized that Fodor and her colleagues did not fail to consider the relevance of prosodic prominence in crafting their proposal about implicit prosody's role in sentence processing. Rather, they selected a methodological approach that sought, perhaps temporarily, to limit the investigation to phrasing, since phrasing would seem to be least confounded with semantic and information structural factors. What we believe our findings indicate is that this approach may simply not be tenable, not only because we know that prominence influences parsing but also because prominence and phrasing are closely linked aspects of prosodic structure.

It thus follows that the emphasis on the relation between syntax and phrasing to the exclusion of prominence—which we understand to be quite standard—is likely to be misguided. Having found the behavioral outcome of our phrasing manipulation in Experiment 2 to be better explained by the prominence contrast that accompanied it, we wonder how well prominence might account for previous findings as well. Certainly we see a plausible explanation for the results of Jun (2010), one of the original motivations for the present study. As discussed in Sect. 1, Jun found speakers of English, a low-attachment language, to prefer placing a boundary before NP2, the pattern predicted by the IPH for high-attachment languages. Jun reports that both NP1 and NP2 were always pitch accented, thus suggesting that NP2s, being phrase final, were nuclear accented, and therefore placed in a structurally prominent position by her speakers. This is in fact what we would expect from a low-attachment language if prominence (explicit or implicit) rather than phrasing were central. Similarly, this understanding would also help us make sense of the Japanese and Korean data from the earlier studies. In those studies (Jun and Koike 2003; Jun and Kim 2004; Jun 2007), prosodic boundaries were so crucial to predicting native speakers' high-attachment preferences (in a way predicted by the IPH) because these languages are edge-marking languages—unlike English, which is a head-marking language. In the prosodic typology proposed in Jun (2005, 2014), English, as a lexical stress language, marks prominence by pitch accenting, i.e., marking the *heads* of prosodic constituents, but Korean and Japanese encode prominence in boundaries, i.e., by tonally marking *edges* of prosodic constituents and/or by positioning a prominent word at the beginning of a prosodic constituent. This happens because Korean, lacking lexical prosody, lacks a *head* of a word; in Japanese, some but not all words have a lexically assigned pitch accent, but word edges are reliably marked by a rising tone (Beckman and Pierrehumbert 1986; Pierrehumbert and Beckman 1988; Warner et al. 2010). Thus, sensitivity to boundaries for syntactic parsing is most apparent in languages with edge marking, but not necessarily head-marking, prosody.

In summary, then, if we revise Fodor's claim about the relation between prosody and parsing to include a wider range of prosodic structure—*and we allow the dominant aspect of that structure to be fixed typologically*—the basic proposal of the IPH finds support in the results of our study.

We also wish to make some brief comments on our findings related to autistic traits—a conspicuous and surprising part of the study. As described in our discussion of Experiment 2's results, we have concluded that the high AQ-Comm individuals may have been largely ignoring prominence patterns, and this is the primary way that they differed from other subjects. However, this would leave open the possibility that this sample of participants therefore used phrasing in the way that the IPH predicts, i.e., prosodic juncture was used to posit syntactic juncture, and, unlike other participants, they were simply not "distracted" by prominence. While we leave open this possibility, we believe a less-positive scenario might also explain their responses. This is one in which the parsers of high-AQ individuals, rather than productively using prosodic juncture to posit a syntactic boundary, were simply disrupted by the juncture—prompting closure at that location. While the distinction between this disruption-prompted closure and the effect of prosodic boundaries already assumed by the IPH may seem subtle (see Ferreira and Karimi, this volume, for insightful discussion of this matter), something along these lines would be necessary if it were found that the same individuals do not have the same boundary-attachment correspondences in their productions. We leave this question open for further research.

Finally, while the results of our structural priming study demonstrate a significant relation between prosodic structure and attachment resolution, we have until now left unspecified the details regarding the mechanism responsible. In particular, it is not yet clear whether the structure that was primed was in fact prosodic or syntactic, and in fact it cannot be teased apart in our Experiment 2. This is because the auditory primes had, in principle, two opportunities to influence participants' comprehension of the target sentences. For example, participants may have first assigned a syntactic structure to the primes, based on their overt prosody, and then reassigned this syntactic structure—but not the prosodic structure—to the target sentence. In this case, the prosody–syntax relationship is still confirmed, but prosody's influence took place at the parsing of primes, not the parsing of targets (and thus was syntactic priming in the usual sense). On the other hand, a scenario is also possible whereby listeners retained the prosodic structure from the primes, reusing that structure for the implicit prosody projected onto the silently read target sentence, and it was at this point that the prosody had its impact on ambiguity resolution for the target.

It is also possible, and in our opinion probable, that both types of priming took place. For the moment, we must be satisfied to have shown that prosody, at some point in the process, influenced attachment resolution systematically. In other work, however, we are currently attempting to tease the two possibilities apart. While this represents work in progress, we are optimistic that prosodic priming will in fact be demonstrable using a paradigm like the one used here, which, unlike Tooley et al.'s (2013) approach, does not involve the additional complications associated with eliciting speakers' productions.

6 Conclusion

In conclusion, this study utilized novel prosodic adaptations of the structural priming paradigm. Our primary finding comes from Experiment 2, where it was demonstrated that the location of a prosodic boundary in auditory primes influenced the attachment of the RC in silently read target sentences. The correlation between boundary location and attachment, however, was in the exact opposite direction predicted by the Implicit Prosody Hypothesis for most of our subjects. Our proposed interpretation of these results relies on a more holistic consideration of prosodic structure, one in which structural prominence is a key factor, as well as a typological difference in prominence marking across languages. Our results, therefore, have crucial implications for future work on Fodor's influential Implicit Prosody Hypothesis, and indeed future work on the relationship between prosody and the processing of syntactic structure in general.

7 Appendix A: Experimental Stimuli for Experiment 1

7.1 Target Sentences

1. Linda wrote to the managers of the assistants that are late all the time.
2. My friend met the aide of the detective that was fired yesterday.
3. Nobody noticed the bodyguard of the actor that was talking on the phone.
4. The reporter interviewed the son of the colonel that had a car accident.
5. The woman knew the photographer of the singer that was reading a book.
6. Patricia saw the teachers of the students that were in class today.
7. Rob talked to the coach of the gymnast that was sick on Saturday.
8. Charlie met the interpreter of the ambassador that was eating dinner.
9. Roxanne read the review of the play that was written by John's friend.
10. The receptionist greeted the clients of the lawyers that were chatting loudly.
11. Jane wrote a story about the uncle of the milkman that was a gentleman.
12. Julia had spoken to the secretary of the doctor that was on vacation.
13. The journalist talked to the daughter of the hostage that was about to leave.
14. Lisa couldn't find the refills of the pens that were in the bottom drawer.
15. Someone shot the servant of the actress that was on the balcony.
16. The dog bit the mother of the teacher that lived in the South of France.

7.2 Prime Sentences

Short Primes

S1. The nurse called in the sister of the hostess that got hurt.
S2. Everybody ignored the stepfather of the monk that had a beard.
S3. The drunk man hit the brother of the neighbor that was yelling.
S4. Lucy admired the hallways of the apartments that were painted.
S5. The chef couldn't find the lid of the pan that was clean.
S6. Ivana hated the father of the delegate that smokes.
S7. Andy ate with the cousin of the dentist that was divorced.
S8. I was talking with the niece of the midwife that lost weight.
S9. The children followed the aunt of the girl that wore a skirt.
S10. The thief took the key of the trunk that was outside.

Medium Primes

M1. Everybody ignored the stepfather of the monk that had a silky white beard.
M2. The driver talked to the guides of the tourists that were angry at the bird.
M3. Ivana hated the father of the delegate that always smokes cigarettes.
M4. Andrew had dinner with the nephew of the butler that loved his former job.
M5. The children followed the aunt of the girl that wore a yellow skirt.
M6. Peter met the uncle of the guest that was a famous chef.
M7. Andy ate with the cousin of the dentist that got divorced last April.
M8. Laura consoled the grandson of the general that lost his right arm and leg.
M9. The drunk man hit the brother of the neighbor that was yelling at the dog.
M10. Mary replaced the wire of the amplifier that got damaged last week.

Long Primes

L1. Peter met the uncle of the guest that was the most famous chef in Los Angeles.
L2. The nurse called in the sister of the hostess that got hurt in a terrible boat accident.
L3. The driver talked to the guides of the tourists that were angry at the restaurant owner.
L4. Lucy admired the hallways of the apartments that were painted light blue and lavender.
L5. The chef couldn't find the lid of the pan that was cleaned after the party in the evening.
L6. Laura consoled the grandson of the general that lost his leg during the Iraq war.
L7. Andrew had dinner with the nephew of the butler that loved his former job in Beverly Hills.
L8. I was talking with the niece of the midwife that lost weight before Maria's wedding.
L9. The thief took the key of the trunk that was outside of the bedroom next to the door.
L10. Mary replaced the wire of the amplifier that has been damaged since last Halloween.

8 Appendix B: Experimental Stimuli for Experiment 2

8.1 Target Sentences

3–4 syllables
1. My friend met the aide of the detective that was fired.
2. Linda wrote to the managers of the assistants that were late.
3. Jamie had inspected the monitor of the computer that was stolen.
4. Rob talked to the coach of the gymnast that was sick.
5. Patricia saw the teachers of the students that were in class.
6. The plumber adjusted the pipe of the sink that was cracked.
7. Charlie met the interpreter of the ambassador that was eating.
8. The dog bit the mother of the teacher that lived in France.
9. The receptionist greeted the clients of the lawyers that were chatting.

6–7 syllables
1. Jane wrote a story about the uncle of the milkman that was a gentleman.
2. Someone shot the servant of the actress that was on the balcony.
3. Roxanne read the review of the play that was written by John's friend.
4. The woman knew the photographer of the singer that was reading a book.
5. The reporter interviewed the son of the colonel that had a car accident.
6. Nobody noticed the bodyguard of the actor that was talking on the phone.
7. Julia had spoken to the secretary of the doctor that was on vacation.
8. The journalist talked to the daughter of the hostage that was about to leave.
9. Lisa couldn't find the refills of the pens that were in the bottom drawer.

8.2 Prime Sentences

1. Peter met the uncle of the guest that was a boxer.
2. The nurse called in the sister of the hostess that hurt herself.
3. Everybody ignored the stepfather of the monk that had a beard.
4. Linda helped to carry the baby of the lady that was upset.
5. The drunk man hit the brother of the neighbor that was yelling.
6. The driver talked to the guides of the tourists that were angry.
7. Lucy admired the hallways of the apartments that were painted.
8. The chef couldn't find the lid of the pan that was clean.
9. Ivana hated the father of the delegate that was smoking.
10. Laura consoled the grandson of the general that lost his leg.
11. Andy ate with the cousin of the dentist that was divorced.
12. Andrew had dinner with the nephew of the butler that loved his job.
13. I was talking with the niece of the midwife that lost her ring.
14. The children followed the aunt of the girl that wore a skirt.
15. The thief took the key of the trunk that was outside.
16. Mary replaced the wire of the amplifier that was damaged.

References

Augurzky, P. (2006). *Attaching relative clauses in German: The role of implicit and explicit prosody in sentence processing* (Vol. 77). Leipzig: MPI Series in Human Cognitive and Brain Sciences.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males, females, scientists and mathematicians. *Journal of Autism & Developmental Disorders, 31,* 5–17.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). Package lme4. Version 1.0-5 (10/25/2013). http://lme4.r-forge.r-project.org/. Accessed Jan 2014.

Beckman, M. (1986). *Stress and non-stress accent* (*Netherlands Phonetic Archives 7*). Dordrecht: Foris.

Beckman, M., & Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. A. Keating (Ed.), *Phonological structure and phonetic form: Papers in laboratory phonology III* (pp. 7–33). Cambridge: Cambridge University Press.

Beckman, M., & Hirschberg, J. (1994). The ToBI annotation conventions. Columbus: Ms. The Ohio State University.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook, 3,* 255–309.

Bergmann, A., & Ito, K. (2007). Attachment of ambiguous RCs: A production study. Talk given at the 13th annual conference on architectures and mechanisms for language processing (AMLaP), Turku, Finland, 24–27 Aug 2007.

Bergmann, A., Armstrong, M., & Maday, K. (2008). Relative clause attachment in English and Spanish: A production study. *Proceedings of Speech Prosody* 2008, Campinas, Brazil.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18,* 355–387.

Bishop, J. (2012a). Focus, prosody, and individual differences in "autistic" traits: Evidence from cross-modal semantic priming. *UCLA Working Papers in Phonetics,* 111*,* 1–26.

Bishop, J. (2012b). Information structural expectations in the perception of prosodic prominence. In G. Elordieta & P. Prieto (Eds.), *Prosody and meaning* (*interface explorations*). Berlin: Mouton de Gruyter.

Bishop, J. (2013). *Prenuclear accentuation: Phonetics, phonology, and information structure*. PhD dissertation, UCLA.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18,* 355–387.

Breen, M., Fedorenko, E., Wagner, M., & Gibson, E. (2010). Acoustic correlates of information structure. *Language and Cognitive Processes, 25,* 1044–1098.

Brysbaert, M., & Mitchell, D. C. (1996). Modifier attachment in sentence parsing: Evidence from Dutch. *Quarterly Journal of Experimental Psychology, 49A*(3), 664–695.

Carreiras, M., & Clifton, C. Jr. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech, 36,* 353–372.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition, 30,* 73–105.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Dussias, P. E. (2003). Syntactic ambiguity resolution in second language learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition, 25,* 529–557.

Ehrlich, K., Fernández, E. M., Fodor, J. D., Stenshoel, E., & Vinereanu, M. (1999). Low attachment of relative clauses: New data from Swedish, Norwegian, and Romanian. Poster presented at the 12th annual CUNY conference on human sentence processing, New York, 18–20 March.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11,* 127–163.

Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins Publishers.

Fernández, E. M., & Bradley, D. (1999). Length effects in the attachment of relative clauses in English. Poster presented at the 12th annual CUNY conference on human sentence processing, New York.

Fodor, J. D. (1998). Learning to parse. *Journal of Psycholinguistic Research, 27*(2), 285–319.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. *NELS, 32,* 113–132.

Frazier, L. (1990). Parsing modifiers: Special purpose routines in the human sentence processing mechanism? In D. A. Balota, G. G. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 303–330). Hillsdale: Lawrence Erlbaum.

Frazier, L., & Clifton, C. (1996). *Construal*. Cambridge: MIT Press.

Freeth, M., Sheppard, E., Ramachandran, R., & Milne, E. (2013). A cross-cultural comparison of autistic traits in the UK, India and Malaysia. *Journal of Autism and Developmental Disorders*. [On-line version ahead of print publication doi:http://dx.doi.org/10.1007/s10803-013-1808-9].

Hemforth, B., Konieczny, L., Scheepers, C., & Strube, G. (1998). Syntactic ambiguity resolution in German. In D. Hillert (Ed.), *Syntax and semantics: A cross-linguistic perspective* (pp. 293–312). San Diego: Academic.

Hoekstra, R., Bartels, M., Cath, D., & Boomsma, D. (2008). Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders, 38,* 1555–1566.

Jun, S.-A. (1996) *The phonetics and phonology of Korean prosody: Intonational phonology and prosodic structure*. New York: Garland.

Jun, S.-A. (2003a). Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research, 32*(2), 219–249.

Jun, S.-A. (2003b). The effect of phrase length and speech rate on prosodic phrasing. *Proceedings of the International Congress of Phonetic Sciences, XV,* 483–486.

Jun, S.-A. (2005). Prosodic typology. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 430–458). Oxford: Oxford University Press.

Jun, S.-A. (2007) The intermediate phrase in Korean intonation: Evidence from sentence processing. In C. Gussenhoven & T. Riad (Eds.), *Tones and tunes: Studies in word and sentence prosody* (pp. 143–167). Berlin: Mouton de Gruyter.

Jun, S.-A. (2010). The implicit prosody hypothesis and overt prosody in English. *Language and Cognitive Processes, 25*(7), 1201–1233.

Jun, S.-A. (2014). Prsodic typology: By prominence type, word prosody, and macro-rhythm. In S.-A. Jun (Ed.), *Prosodic typology II: The phonology of intonation and phrasing* (pp. 520–539). Oxford: Oxford University Press.

Jun, S.-A., & Kim, S. (2004). Default phrasing and attachment preferences in Korean. *Proceedings of Interspeech-ICSLP*, Jeju, Korea.

Jun, S.-A., & Koike, C. (2003). Default prosody and RC attachment in Japanese. Talk given at the 13th Japanese-Korean Linguistics Conference, Tucson, AZ. [Published in *Japanese-Korean Linguistics* 3, 41–53, CSLI, Stanford, in 2008].

Jun, S.-A., & Shilman, M. (2008). Default phrasing and English relative clause attachment data. *Proceedings of Speech Prosody* 2008, Campinas, Brazil.

Lee, E.-K., & Watson, D. G. (2011). Effects of pitch accents in attachment ambiguity resolution. *Language and Cognitive Processes, 26*(2), 262–297.

Lovric, N., Bradley, D., & Fodor, J. D. (2001). Silent prosody resolves syntactic ambiguities: Evidence from Croatian. Presented at the SUNY/CUNY/NYU Conference, Stonybrook.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nieuwland, M., Ditman, T., & Kuperberg, G. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language, 63,* 324–346.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*(3), 427–459.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. PhD dissertation, MIT.

Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. Cambridge: MIT Press.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 4,* 515–526.

Quinn, D., Abdelghany, H., & Fodor, J. D. (2000). More evidence of implicit prosody in silent reading: French, English and Arabic relative clauses. Poster presented at 13th CUNY Conference on Human Sentence Processing.

R Development Core Team. (2013). *R: A language and environment for statistical computing* (*Version 3.0.2*). Vienna: R Foundation for Statistical Computing. http://www.r-project.org. Accessed Jan 2014.

Ruta, L., Mazzone, D., Mazzone, L., Wheelwright, S., & Baron-Cohen, S. (2011). The autism-spectrum quotient: Italian version: A cross-cultural confirmation of the broader autism phenotype. *Journal of Autism and Developmental Disorders, 42,* 625–633.

Schafer, A. J., Carter, J., Clifton, C., & Frazier, L. (1996). Focus in relative clause construal. *Language and Cognitive Processes, 11,* 135–163.

Sekerina, I. A., Fernández, E. M., & Petrova, K. A. (2004). Relative clause attachment in Bulgarian. In O. Arnaudova, W. Browne, M. L. Rivero, & D. Stojanović (Eds.), *The proceedings of the 12th annual workshop on formal approaches to Slavic linguistics*. The Ottawa meeting 2003 (pp. 375–394). Ann Arbor: Michigan Slavic Publications.

Selkirk, E. (2000). The interactions of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and experiment*. Dordrecht: Kluwer Academic.

Sonié, S., Kassai, B., Pirat, E., Bain, P., Robinson, J., Gomot, M., Barthélémy, C., Charvet, D., Rochet, T., Tatou, M., Assouline, B., Cabrol, S., Chabane, N., Arnaud, V., Faure, P., & Manificat, S. (2012). The French version of the autism-spectrum quotient in adolescents: A cross-cultural validation study. *Journal of Autism and Developmental Disorders,* 1–6 (online version accessed: doi:10.1007/s10803-012-1663-0).

Swets, B., Demset, T., Hambrick, D., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136*(1), 64–81.

Tooley, K., Konopka, A. E., & Watson, D. (2013). Can intonational phrase structure be primed (like syntactic structure)? *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Nov 4. [Epub ahead of print. doi:10.1037/a0034900].

Turnbull, R., Royer, A., Ito, K., & Speer, S. (2014). Prominence perception in and out of context. *Proceedings of Speech Prosody 2014,* 1164–1168.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37,* 498–505.

Wakabayashi, A., Baron-Cohen, S., Wheelwright, S., & Tojo, Y. (2006). The autism-spectrum quotient (AQ) in Japan: A cross-cultural comparison. *Journal of Autism and Developmental Disorders, 36,* 263–270.

Warner, N., Otake, T., & Arai, T. (2010). Intonational structure as a word boundary cue in Japanese. *Language and Speech, 53,* 107–131.

Xiang, M., Grove, J., & Giannakidou, A. (2011). Interference "licensing" of NPIs: Pragmatic reasoning and individual differences. Poster presented at the 24th CUNY Conference on Human Sentence Processing, Stanford University.

Xiang, M., Grove, J., & Giannakidou, A. (2013). Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology, 4,* 708. doi:10.3389/fpsyg.2013.00708.

The Interplay of Visual and Prosodic Information in the Attachment Preferences of Semantically Shallow Relative Clauses

Eva M. Fernández and Irina A. Sekerina

Abstract Many studies have investigated the attachment of relative clauses (RCs) modifying complex noun phrases (NPs). Cross-language differences in how ambiguous RCs are interpreted have been attributed to a number of factors, among which lexical semantics and prosody seem to play a special role. We report data from an experiment conducted in English using semantically shallow sentences that describe geometric shapes. The spoken sentences contained the ambiguity of interest and were paired with visual displays that contained two scenes. In the disambiguating conditions, only one of the scenes was compatible with the attachment of the RC as high or low. In the ambiguous condition, either scene could be chosen. Sentences were presented to participants with one of two prosodic contours: compatible with high attachment (phrasal break before the RC) or compatible with low attachment (phrasal break after the head noun in the complex NP). Participants' interpretation preferences were assessed via their choice of the scene which disambiguated the interpretation of the RC; we additionally recorded participants' eye movements as they performed the task. We discuss the interplay of prosodic and visual disambiguation in determining the attachment preferences of semantically shallow RCs.

Keywords Relative clause attachment ambiguity · Visual disambiguation · Prosodic disambiguation · Eye tracking · English

E. M. Fernández (✉)
Queens College, City University of New York, Flushing, NY 11367, USA
e-mail: eva.fernandez@qc.cuny.edu

I. A. Sekerina
College of Staten Island, City University of New York, Staten Island, NY 10314, USA
e-mail: irina.sekerina@csi.cuny.edu

E. M. Fernández · I. A. Sekerina
Graduate Center, City University of New York, New York, NY 10016, USA

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_13

1 Introduction

Models of the human sentence processing mechanism aim to explain how the structure of a sentence is built given a linearly ordered string of words with possibly multiple structural configurations. Putting aside how preferred initial syntactic attachments are evaluated, if at all, against competitors (is the parser serial or parallel?), a central question in sentence processing research is what kind of initial ("online") information guides the parser in the phases of processing that lead to an ultimate ("offline") interpretation of a globally ambiguous string. In this investigation, we use the relative clause (RC) attachment construction to examine the interplay of explicit prosody and visual displays in determining the time course of attachment decisions when the materials are semantically shallow. We describe our materials as "semantically shallow" because the spoken target sentences are accompanied by visual displays of simple geometric shapes designed to eliminate plausibility confounds that exist in other studies that typically use lexical, and in particular animate noun phrases (NPs); the visual displays, we assume, make it unnecessary to engage in deep semantic processing of the accompanying sentences.

The RC attachment ambiguity involves attachment of a RC inside either a local or a nonlocal constituent, as in the following example:

(1) Someone shot the maid of the actress who was on the balcony.

Here, the parser should prefer—by a locality principle like late closure (Frazier and Fodor 1978) or recency (Gibson et al. 1996)—to attach the RC low to the local noun *the actress,* and not to the nonlocal noun *the maid* (high attachment). The empirical record for how this construction is processed, however, documents many deviations from this predicted preference for low attachment, with effects modulated by many variables. These variables are related to the aspects of the materials, including the type of complex NP (De Vincenzi and Job 1993; Shaked 2009; Traxler et al. 1998), the restrictiveness of the RC (Carreiras 1992), and the length of the RC (Fernández 2003; Shaked 2009). Variation in attachment preference can also be linked to variation in the participants derived from factors, including proficiency and language history (Dussias and Sagarra 2007; Fernández 2003), working memory (Ferreira and Karimi, this volume; Swets et al. 2007), and other individual differences (Jun and Bishop, this volume).

A productive line of investigation has examined the role of prosody, both explicit and implicit, in the processing of the RC attachment ambiguity (Augurzky 2006; Fernández 2007; Shaked 2009; Stoyneshka et al. 2010; Teira and Igoa 2007). The experiment reported in this chapter is a promising pilot study whose novelty is that it examines the moment-by-moment time course of processing by the use of the visual world eye-tracking paradigm, with materials whose shallow semantics eliminate potential plausibility confounds that arise with animate NPs. The data also provide some insights about the time course of the use of prosody to determine attachment, which contribute toward the development of models of the human sentence processing mechanism.

We begin with an overview of existing evidence on how the RC attachment ambiguity is processed, with a specific focus on research that examines the role of prosody in the interpretation of this construction. We then present data from an eye-tracking experiment conducted in English. We conclude the chapter with a discussion of our results.

2 Background

A parsing preference for local attachment with the RC attachment construction could be stipulated to be the parser's preference for an initial attachment for all languages and all variants of the structure. This stipulation is not without experimental antecedent (De Vincenzi and Job 1993; Fernández 2003; Maia et al. 2007; Traxler et al. 1998). It permits framing the problem of variation in RC attachment as a problem of determining what type of information in the stimulus results in a reanalysis from the local (N2) to the nonlocal (N1) attachment.

One type of information that modulates attachment preference, at least offline, is contained in the explicit prosody of an utterance. The use of explicit prosody cues for syntactic disambiguation has been the topic of many investigations using the RC attachment construction. These studies have established that the explicit prosody of an utterance containing the RC attachment ambiguity can bias interpretation. Specifically, a phrasal break after N1 and before the prepositional phrase in the complex NP (an intonational contour which we will call "early break prosody") strongly encourages low attachment, while a phrasal break after N2 and before the RC ("late break prosody") encourages high attachment (Augurzky 2006; Fernández 2007; Shaked 2009; Stoyneshka et al. 2010; Teira and Igoa 2007). Many investigations of the role of explicit prosody in RC attachment were sparked by the influential proposal by Fodor (1998, 2002) that even implicit prosody could guide syntactic processing, formulated as the *implicit prosody hypothesis* (Fodor 2002). A number of investigations have suggested that implicit prosody can guide attachment decisions (Fernández 2007; Jun and Bishop, this volume), though perhaps not in the earliest stages of processing (Augurzky 2006). Discussing the role of implicit prosody in the resolution of syntactic ambiguities is beyond the scope of this chapter, though it is amply covered elsewhere in this volume (Bader, this volume; Breen, this volume; Jun and Bishop, this volume; Speer and Foltz, this volume; Wasow et al., this volume).

The overwhelming majority of studies of the RC attachment ambiguity has used written materials like the sentence in (1) as isolated (zero-context) sentences presented to participants in writing. A handful of studies have employed the auditory modality, requiring participants to listen to spoken ambiguous sentences, and choose a preferred interpretation. These studies predominantly have demonstrated that for languages as different from each other as English (Fernández 2007), German (Augurzky 2006), Spanish (Teira and Igoa 2007), Bulgarian (Stoyneshka et al. 2010), and Hebrew (Shaked 2009), explicit prosody exerts a disambiguating effect,

asymmetric in the same way regardless of the language: A prosodic break after N1—which itself is rather rare (Fernández 2005; Gryllia and Kügler 2010; Shaked 2009)—affects ambiguity resolution strongly, whereas a break after N2 results in a weaker or even absent effect. Materials based on sentences like (1) have almost universally been semantically deep and, as such, their interpretation could be influenced by semantic/pragmatic properties that have not been controlled experimentally. At best, these properties (for instance, actresses could be more likely to be on balconies than maids) contribute noise to the data, though at worst they could result in materials that are inadvertently biased toward one or another attachment (Fernández 2003, Chap. 4). This is an important shortcoming of existing studies, including the auditory ones. Studying the RC attachment ambiguity in both auditory format and with materials that do not present a danger of uncontrolled semantic/pragmatic properties is precisely what is needed if we want to go beyond the convenience samples of college students and explore an entire range of factors that affect such preferences (one exception to studies using convenience samples is Felser et al. (2003), a self-paced listening study of RC attachment in school-aged children).

The visual world eye-tracking paradigm makes it possible to avoid using semantically deep sentences and written stimuli. We are aware of only one study employing this paradigm to study RC attachment, January and Trueswell (2007). In this study, participants viewed visual displays with six geometric shapes (diamonds, squares, and circles) arranged in a 3×4 grid. Three of the six shapes had another shape embedded in them (e.g., a circle inside a triangle). The task was to act out auditory instructions like "Click on the square above the circle that has the triangle." Three different configurations of the six shapes created three visual display conditions: disambiguated low attachment, disambiguated high attachment, and ambiguous. No clear preference emerged in interpretation of the ambiguous displays, but RTs were faster in those of them that were disambiguated toward high attachment. Crucially, participants were sensitive to the visual context because the first-gaze duration predicted the ultimate interpretation. Participants also were faster to direct their gaze at the target shape they would eventually click on with ambiguous than with disambiguated displays.

The study we report below also used materials with shallow semantics, asking participants to name the colors of elements in visual displays involving geometric shapes, where attachment was either disambiguated or kept ambiguous. This was achieved *visually* by placing the referent of the RC (i.e., the umbrella in (2)) in either only one picture depicting the triangle with a tip (disambiguated) or in both pictures (ambiguous), as illustrated in Fig. 1 below. In contrast to January and Trueswell (2007), we additionally manipulated explicit prosody. This allowed us to examine the moment-by-moment preferences of our participants as they were processing ambiguous RCs with prosody that encouraged high or low attachment. Our materials were short narratives that consisted of two preamble sentences followed by a question containing the complex NP with the RC that was morphosyntactically ambiguous:

(2) What color is the tip of the triangle that has an umbrella in the middle?

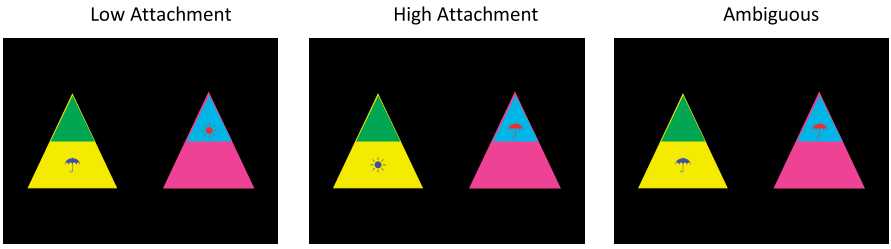


Fig. 1 Sample experimental item, presented with the sentences in (4) in three conditions: disambiguated low attachment (*left* panel), disambiguated high attachment (*middle*), and ambiguous (*right* panel)

Accompanied by a visual display where an umbrella appeared in the middle of a triangle or in a triangle's tip, these "semantically shallow" sentences lend themselves to simple visualization using basic icons and shapes, all highly recognizable and with high lexical frequency. Sentences like these have reduced inherent pragmatic and lexical biases, as they all describe the co-occurring visual scenes with equal plausibility. Such visual displays and their accompanying sentences can be used with language learners, including pre-literate children, since the technique does not require reading and involves a relatively reduced vocabulary that can be introduced in a training session, if necessary.

The experiment reported below, conducted in English, was a follow-up to our two-part study in Bulgarian (Sekerina et al. 2008). The first part was a traditional offline written questionnaire that used semantically deep materials (similar to the sentence in (1)), and there we found a high attachment preference for ambiguous RCs (59%). The second part (Experiment 2) was an auditory experiment in which we switched to visually presented semantically shallow materials identical to those used in the experiment in English reported below, and paired them with auditory sentences (3), the Bulgarian equivalent of the English sentence in (2). The procedure was exactly as for the experiment described below, only without eye tracking.

(3) Kakâv cvjat e vârxât na triâgâlnika, v kojto e narisuvan čadâr?
What color is the tip of the triangle in which is drawn umbrella
"What color is the tip of the triangle that has an umbrella in the middle?"

We found a ceiling effect for accuracy for the visually disambiguated low attachment sentences (98%), but for the high attachment sentences, accuracy was only 64%. That is, for a third of the latter sentences, participants named the color of the triangle itself, instead of the color of the tip, a response that we will refer to below as "whole-object" answers. Moreover, for the visually ambiguous sentences, attachment preference was the opposite from the offline written experiment (63% low attachment). To explain this language-internal shift, we argued that attaching low is not only universal but is also less computationally demanding than attaching high. The color-naming task employed in the auditory experiment with semantically shallow materials (geometric shapes) limits the resources of the processor in a way that

a questionnaire does not. Taxed computational resources limit the likelihood that the parser will deviate from its initial (and universal) preference for low attachment.

The auditory experiment in Bulgarian was an offline study, and its results fit well with the RC attachment literature in which there is widespread agreement with respect to attachment preferences when the method is "offline". Sometimes, however, discrepancies emerge between offline and online methods (Fernández 2003; Maia et al. 2007), suggesting that not all studies or methods are tapping the same level of processing. A dissociation between offline and online performance is well documented in first-language acquisition studies with children whose interpretation of various construction can vary dramatically depending on the task employed (Bergmann et al. 2012; O'Grady et al. 2010; Sekerina et al. 2004). Therefore, for this investigation, we chose a method that allowed us not only to collect offline end-of-sentence responses (for examining the ultimate preferred interpretation) but also to capture aspects of the time course of processing leading up to that ultimate interpretation. The fine-grained time course information derived from eye movements allows us to discover any dissociations between end-of-sentence response and ongoing processing preferences, while offering the opportunity to examine how participants arrive at the preferred interpretation in resolving the attachment of the RC.

3 Semantically Shallow Sentences with Visual Displays and Prosodic Disambiguation

In this experiment, we presented semantically shallow sentences recorded with prosody that encouraged either high or low attachment and accompanied by visual displays. The visual displays either matched the interpretation suggested by the prosody (high or low) or were ambiguous.

3.1 Method

Participants Undergraduate students ($N=21$, 11 women; mean age 21.5), recruited from the Department of Psychology research participation pool at Rutgers University, volunteered to participate in exchange for credit. All were native speakers of American English and naïve with respect to the goals of the experiment.

Design and Materials The experiment had 9 experimental and 21 filler items. The experimental materials each had were three visual display variants (low attachment, high attachment, ambiguous), and were presented with one of two kinds of prosody (early break, late break) in a design where disambiguated visual displays (low attachment and high attachment) were presented with appropriate corresponding prosody (early break and late break, respectively). Ambiguous visual displays were

presented with either early break or late break prosody, both of which were equally compatible.

Each item consisted of a picture (one of the panels in Fig. 1) paired with a set of three spoken sentences like the triplet in (4).

(4) a. Here is a pink triangle and a yellow triangle.
b. They have different color tips.
c. What color is the tip of the triangle that has an umbrella in the middle?

The visual stimuli were designed to avoid lexical/pragmatic biases that the two nouns in the complex NP might exert on the RC in terms of its attachment preference. To this end, we combined a number of abstract geometric shapes (crosses, rectangles, ovals, arrows, stars, circles, etc.) with icons (suns, umbrellas, hearts, butterflies, etc.) and background patterns (stripes, polka dots, etc.), to create composite figures whose parts were of different colors. In Fig. 1, in all three conditions, the triangle on the left is yellow and has a green tip, and the triangle on the right is pink and has a blue tip. What varies is the placement of the umbrella referred to by the RC: in the middle of the triangle (low attachment interpretation) or in the tip of the triangle (high attachment interpretation).

The audio materials for the experimental items all followed the pattern in the example in (4). The preamble sentences (a) and (b) set the stage for the question bearing the construction of interest (c) which was presented as a question. In the experimental materials, a complex NP (always of the form N1 of N2) was followed by an RC introduced by the complementizer *that,* the RC was morphosyntactically ambiguous and could permissibly modify N1 (*tip*) or N2 (*triangle*).

For each trial, the visual displays (each of the panels in Fig. 1) offered one (low attachment, high attachment) or two (ambiguous) visual depictions of the sentence. The question bearing the construction of interest was about the color of N1. The correct answer for the low-attachment condition illustrated in Fig. 1 is “green”; the correct answer for the high-attachment condition is “blue”. For the ambiguous condition, the participant’s answer would indicate the interpretation of the attachment of the RC: “green” indicated a low-attachment interpretation (the umbrella is in the triangle with a green tip), “blue” a high-attachment interpretation (the umbrella is in the triangle’s blue tip).

The visual stimuli comprise the visual display factor with three conditions: Disambiguated high attachment, disambiguated low attachment, and ambiguous attachment. A second factor manipulated in the experiment was prosody that was either compatible with low attachment (early break) or high attachment (late break).

The audio stimuli were recorded by a female native speaker of English trained to produce the preambles and the target question with the intended prosodic contours naturally and in a register that would be appropriate for children (to permit for future reuse of these materials with children). For the target question, the speaker phrased the materials by adding pauses at predetermined locations in the sentences: before the preposition in the complex NP for the early break versions, and before the complementizer of the RC for the late break versions. Waveforms and pitch tracks for one of the stimulus pairs are presented in Fig. 2. As the figure indicates,

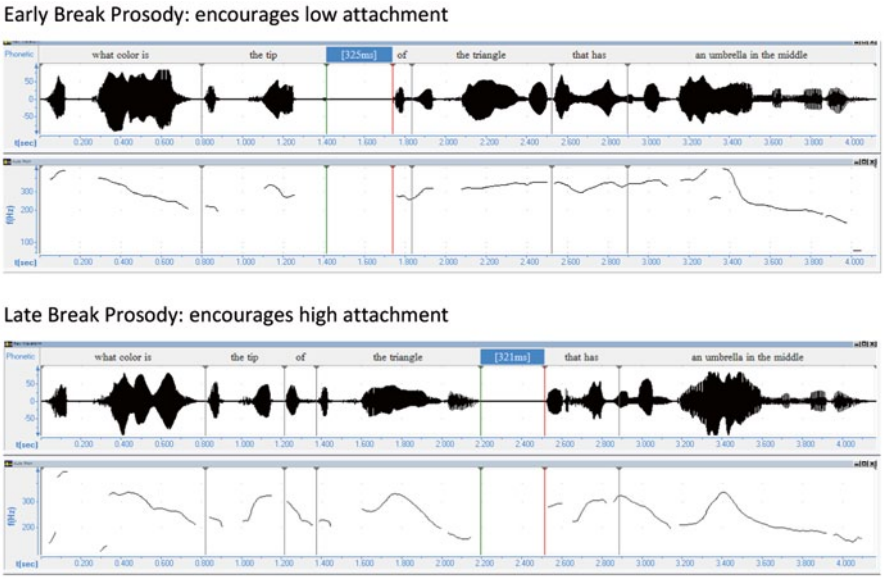


Fig. 2 Waveform and pitch track for target sentence in the sample experimental item in (4) and Fig. 1. The *top* panel displays the variant of the sentence with early break prosody (facilitating low attachment); the *bottom* panel displays late break prosody (facilitating high attachment). The highlighted regions indicate the location of the phrasal break and its duration

the pauses in each of the members of the pair are very similar in duration (325 and 321 ms, respectively). The phrasal break in each is accompanied by pitch reset and pre-boundary lengthening (the duration of *the tip* is 616 ms when it precedes a break, compared to 395 ms; the duration of *the triangle* is 820 ms preceding the break, compared to 689 ms).

The nine experimental items were rotated through three visual display conditions with three items per condition—disambiguated low attachment (Fig. 1a), disambiguated high attachment (Fig. 1b), and ambiguous (Fig. 1c). However, we could not use a standard fully crossed 3×2 design in which visual display (high, low, ambiguous) was crossed with prosody (early break, late break), since that would have produced two conditions with an unnatural pairing of high-attachment visual displays with early break prosody and low-attachment visual displays with late break prosody. In our materials, the prosody systematically matched the visual display for two of the three conditions: High-attachment visual stimuli were always presented with late break prosody, and low-attachment visual stimuli were always presented with early break prosody. For the nine ambiguous visual displays distributed across three versions of the experiment in a between-participants design, six were paired with early break prosody (facilitating low-attachment interpretations) and three with late break prosody (facilitating high-attachment interpretations).

To preclude participants from discovering the purpose of the experiment, the number of experimental trials was kept to a minimum, with 9 experimental and

Fig. 3 Example of a filler item, presented with the sentences in (5)

21 filler items in each list, of which 18 were interspersed pseudorandomly among the experimental trials, and three served as list protectors at the beginning of each version. The fillers were designed to resemble experimental items, with pictures containing elements in different colors. An example, presented in Fig. 3, contained a striped triangle with a blue base and a polka-dotted triangle with a green base. This visual stimulus was paired with a set of three sentences like those in (5).

(5) a. Here is a polka-dotted triangle and a striped triangle.
b. Their bases are different colors.
c. What color is the base of the polka-dotted triangle?

The questions in filler items did not contain a complex NP of the type employed in the experimental items. The correct response was always unambiguous: In this example, the base of the polka-dotted triangle is green.

Three versions of the experiment rotated the disambiguated experimental items through the three visual display conditions in a Latin square design. For the prosody manipulation, two ambiguous items were always presented with the late break prosody and one with the early break prosody. Participants were pseudorandomly assigned to one of the three versions.

Procedure Participants were seated in front of a 19-in. monitor attached to a desktop personal computer (PC). The visual stimuli were presented using a PowerPoint file: Visuals appeared on the monitor, and auditory stimuli played simultaneously through speakers at a comfortable audio level. Participants' task was to gaze at the displays, while listening to the sentences, and to answer the question by naming a color. Participants' oral responses were recorded manually by a research assistant and confirmed later from the audio record for each participant.

The complete sequence was as follows (using Fig. 1 and example (4), to illustrate). First, a black screen with fixation point (a small square with a pink border) at the top and centered was displayed on the monitor. After 500 ms, the pink triangle (without the tip) appeared on the right side of the monitor followed by the yellow triangle with an umbrella in the middle, on the left. The event was synchronized with the auditory presentation of the first sentence (a) in the triplet in (4). After a 500-ms pause, the blue tip with an umbrella appeared inside the pink triangle in conjunction with the second sentence (b). Finally, after another 500-ms pause, the experimental question (c) was played. Participants had 5000 ms to produce an answer by naming the color of the appropriate part of one of the pictures in the display.

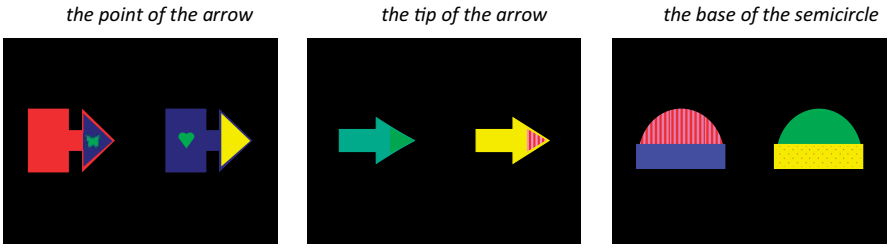


Fig. 4 High -attachment visual displays that induced whole-object answers

The monitor was connected to a remote free-viewing tabletop ISCAN eye-tracking system that allowed for participants' head movement (for technical details, see Sekerina et al. 2004, pp. 135–136). Eye movements were sampled at a rate of 30 times/s and were recorded on a digital SONY DSR-30 videotape recorder. Prior to the start of the experiment, each participant underwent a short calibration procedure. The experiment itself lasted approximately 20 min.

Data Treatment and Analysis We tabulated and analyzed three types of data: (i) color-naming accuracy (for disambiguated displays) or attachment preference (for ambiguous displays), (ii) color-naming times, and (iii) eye movements to the high-attachment panel or the low-attachment panel during the trial. There were no missing trials.

First, color-naming responses were established from the audio record for each participant. There were no missing color-naming responses in the dataset. However, for the experimental items, a complication emerged in answers to the target question with disambiguated high-attachment displays, including the display in Fig. 1b. (The same complication was observed in the auditory second part of the Bulgarian experiment discussed above (Sekerina et al. 2008; so it is not particular to English). The expected answer was the color of the part of N1 (e.g., for Fig. 1b, *blue,* referring to the tip of the pink triangle that contains the umbrella). In 17 trials out of 63 (27%), participants selected the correct panel (the triangle on the right), but answered with the color of the whole object, instead of the color of the part (*pink* instead of *blue*). We refer to these responses as "whole-object" answers. Approximately half of the visual displays (including the item illustrated in Fig. 1) elicited whole-object answers. Moreover, half the participants contributed at least one whole-object answer. Additional examples of the experimental items that resulted in whole-object answers are shown in Fig. 4, along with their corresponding target questions in (6).

(6) a. What color is the point of the arrow that's got a butterfly inside it?
Correct answer: blue Whole object answer: red
b. What color is the tip of the arrow that has vertical stripes?
Correct answer: pink Whole object answer: yellow
c. What color is the base of the semicircle that's covered with little dots?
Correct answer: yellow Whole object answer: green

In the Bulgarian study, the whole-object answers also accounted for 30% of responses. We believe that the explanation for this behavior is not linguistic, but rather

is related to working memory, which is very susceptible to interference. In this case, the color of the triangle (N2) interferes with naming of the color of its tip (N1), a cause that has been proposed to explain difficulties in processing of filler–gap dependencies (Van Dyke and McElree 2006). Note, however, that whole-object answers do not bear on RC ambiguity resolution: They reflect working memory constraints. Participants simply lose track of which component of the picture should be named, but they do choose the correct panel (left or right). Following this logic, the whole-object answers were tabulated as indicating the corresponding attachment resolution.

Second, naming times were assessed from the eye-movement video protocols. A ± 2 standard deviation cutoff was applied to the naming times, affecting 4.8 % of the data. In addition, for disambiguated visual displays, naming times for trials answered incorrectly were excluded from the dataset, affecting 3.7 % of the data. In total, 8.5 % of the naming times were excluded from the analyses by the standard deviation cutoff and error criterion.

Third, eye movements were extracted from the videotape using a SONY DSR-30 videotape recorder with frame-by-frame control and synchronized video and audio. For each trial, four categories were coded: looks to the left panel (the yellow triangle), looks to the right panel (the pink triangle), looks to the central fixation square, and track loss. Track loss (3.1 %) and looks to the fixation square (13.5 %) constituted a small proportion of total looks and were removed from the eye-movement analyses. Therefore, fixations to the left and right panels in the visual display were in complementary distribution to each other. The analyses reported below are based on proportions of looks to the left and right panels in separate time windows, or regions of interest (ROIs). Each trial was segmented into six ROIs that were defined relative to the onset of the five phrases of the experimental question:

(7) ROI1	ROI2	ROI3	ROI4	ROI5	ROI6
0–666 ms	667–1300 ms	1301–1866 ms	1867–2466 ms	2467–3200 ms	3201–5000 ms
What color is	*the* N1 *of*	*the* N2	*that* V	NP	[silence until naming]
What color is	*the tip of*	*the triangle*	*that has*	*an umbrella in the middle?*	

Note that for both disambiguated and ambiguous visual displays, the sentence's lexical content up to the middle of ROI5 is compatible with both panels in the visual display. Selecting the correct panel is possible only once the RC has been heard. With disambiguated visual displays, fixations to the correct panel could only emerge in ROI5 or later, unless prosody is used to process the syntactic ambiguity. In the ambiguous condition, in contrast, both panels of the visual display are compatible with the interpretation of the sentence (i.e., low attachment preference (the left panel of Fig. 1c) or high attachment preference (the right panel of Fig. 1c), leaving prosody as the sole cue for choosing one panel over the other.

The dependent variable used for the analyses of the eye-movement data was fixation proportions to the correct panel (with disambiguated displays) or the prosody-compatible panel (with ambiguous displays) out of all fixations made during a given region of the sentence.

4 Results

Accuracy and Naming Times Participants' accuracy in answering the questions by naming the appropriate color for the 21 fillers was near ceiling, 98.6% (only six errors in 441 trials). We take this to mean that the participants were attentive to the task and found it easy to perform.

Figure 5 presents color-naming responses for the experimental materials: percent accuracy for disambiguated displays (left panel), percent high attachment preference for ambiguous displays as revealed by color named (middle panel), and response times (right panel).

For materials with disambiguated visual displays (left panel of Fig. 5), the prosody of the utterance supported the visual disambiguation: Late break prosody for high-attachment displays and early break prosody for low-attachment displays. There was no reliable difference in accuracy between high-attachment (92.1%) and low-attachment (96.8%) materials ($F_1(1,20)=1.30$, $p=0.267$, $SS=233.36$, $MSe=178.91$; $F_2(1,8)=1.03$, $p=0.339$, $SS=107.56$, $MSe=103.93$). High-attachment materials had higher reaction times, 4299 ms, than low-attachment materials, 3994 ms (left side of the right panel of Fig. 5), a difference reliable only in the participant-based analysis ($F_1(1,20)=8.50$, $p=0.009$, $SS=979{,}509$, $MSe=115{,}248$; $F_2(1,8)=1.96$, $p=0.199$, $SS=285{,}652$, $MSe=145{,}444$). This suggests that naming the correct color based on the visual disambiguation (and supported by prosody) was equal for high and low disambiguation, with a minor advantage for low-attachment displays which only emerged in participant-based response times.

For materials with ambiguous displays, the only difference between the two conditions was the prosody: prosody encouraging high attachment (late break) and prosody encouraging low attachment (early break). The color-naming responses are plotted here as percent high attachment preference (middle panel of Fig. 5), assuming that chance is at 50%. We analyzed participant- and item-based means by calculating *t*-tests for a single mean to test for difference from chance set at 50%. Ma-

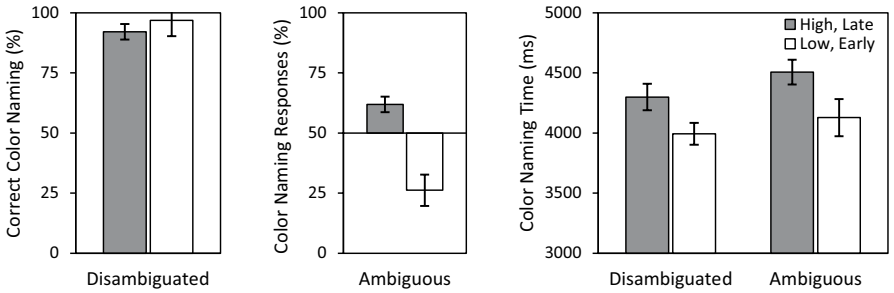


Fig. 5 Responses (accuracy for disambiguated displays and high attachment preference for ambiguous displays, where chance is assumed to be at 50%) and naming times for disambiguated and ambiguous displays, as a function of prosody (late break, early break). Means and standard errors are from participant-based data

terials presented with late break prosody were not significantly above chance (62% high-attachment responses, t_1 (6) = 1.05, $p > 0.30$, $t_2 < 1$), confirming the asymmetric effect of prosody already established in the literature (Augurzky 2006; Fernández 2007; Shaked 2009; Sekerina et al. 2008; Stoyneshka et al. 2010; Teira and Igoa 2007). Materials presented with early break prosody were below chance, marginally in the item-based analyses (26% high-attachment responses, t_1 (13) = −3.33, $p = 0.005$; t_2 (5) = −2.10, $p = 0.089$). For the reaction times (right side of the right panel of Fig. 5), comparison of the two means was performed with *t*-tests for samples with unequal variances. The observed difference between reaction times for responses that favored high attachment preference (4506 ms) and those that favored low attachment preference (4128 ms) was not significant (t_1 (18) = 1.54, $p = 0.141$; t_2 (6) = 1.47, $p = 0.191$). Together with the color-naming responses, this suggests that the prosody that encourages low attachment is a better cue than the prosody that encourages high attachment.

Fine-Grained Eye-Movement Analyses The eye-movement data are presented separately for disambiguated displays (Fig. 6) and ambiguous displays (Fig. 7).

The graphs in Fig. 6 represent the way participants allocated visual attention across the six ROIs in the sentences with disambiguated displays, as they shifted between the two panels (Fig. 1a, b). Analyses of variance were performed on participant- and item-based means for proportions of looks to the panel. The omnibus analysis of variance (ANOVA)-crossed attachment (high, low) × region (ROI1–ROI6)×display (correct, mismatching) as within-participant and within-item factors.[1] The middle panel of Fig. 6 summarizes this interaction: Attachment is indicated by bar color, each pair of bars represents a region, and the *y*-axis represents looks to correct panel. A positive score represents preference to fixate to the high-attachment panel, while a negative score represents a preference to fixate to the low-attachment panel.

The three-way interaction was significant ($F1$ (5100) = 11.6, $p < 0.001$, $SS = 3.23$, $MSe = 0.06$; $F2$ (5,30) = 10.2, $p < .001$, $SS = 1.21$, $MSe = 0.02$); planned comparisons performed for each ROI crossed the factors attachment × display.

ROI1–ROI2 ROI1 introduces the *wh*-question ("What color is"), but bears no disambiguating information, so it is not surprising that no effects emerge in this region (attachment × display interaction: F_1, $F_2 \leq 1$; main effects: $p > 0.05$). ROI2 includes N1 and the preposition in the complex NP ("the tip of"). Early break prosody is signaled within ROI2, but it is too early for making use of this information (attachment × display interaction: F_1, $F_2 \leq 1$; main effects: $p > 0.05$).

ROI3 For both disambiguating displays, there were more looks to the item representing the head of the complex NP with an object depicted in it (in our example, more looks to *the tip* with either a sun or an umbrella in it). Hence, the attachment

[1] One of our reviewers suggested an analysis of the eye-tracking data using log-gaze probability ratios (Arai et al. 2007). We performed this alternative analysis with data for both disambiguated and ambiguous displays. The resulting patterns were identical to those reported here with non-transformed data.

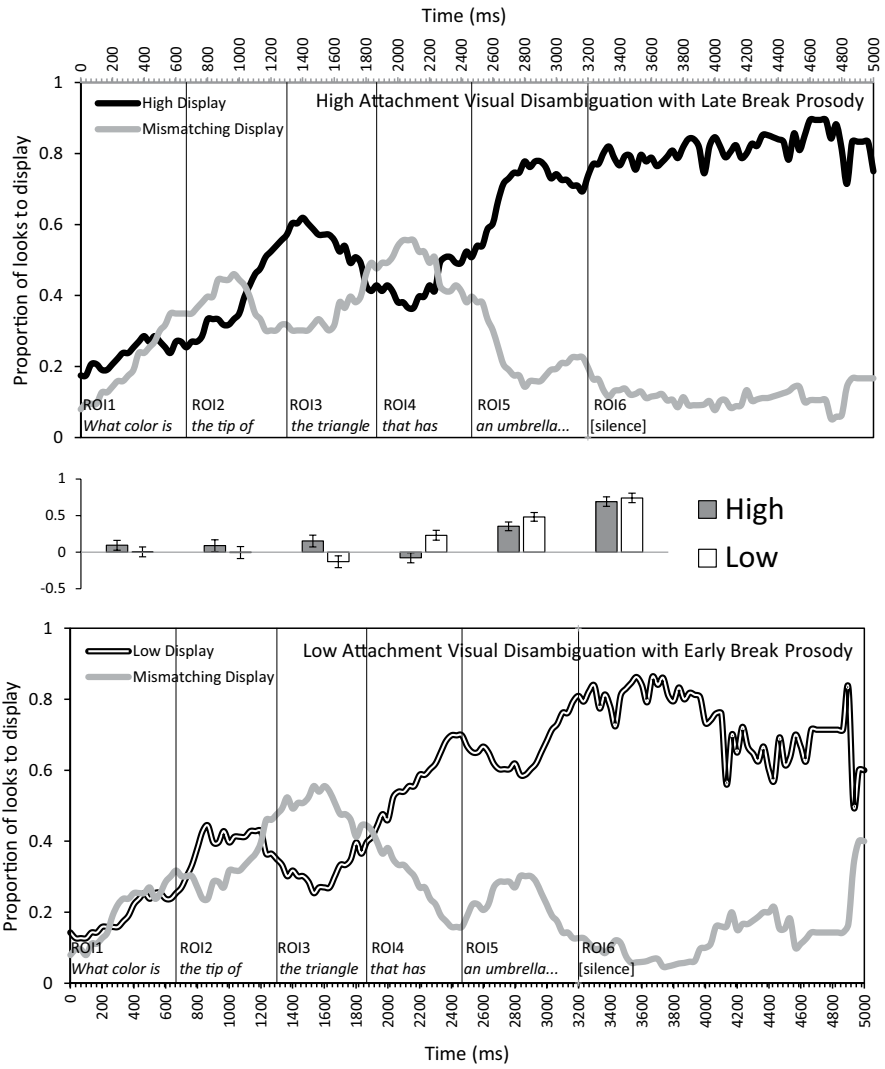


Fig. 6 Proportions of looks over time with disambiguated visual displays for high-attachment display and late break prosody materials (*top* panel) and low-attachment display and early break prosody materials (*bottom* panel). Looks to the correct panel are in *solid black* (*high* attachment) or *hollow black* (*low* attachment); looks to the mismatching (competitor panel) are in *gray*. The *bar* graphs in the middle panel are the participant-based means and standard errors of preference to look at the correct panel for each of the six regions of interest (ROI1–ROI6) for materials disambiguated high (*gray*) or low (*white*)

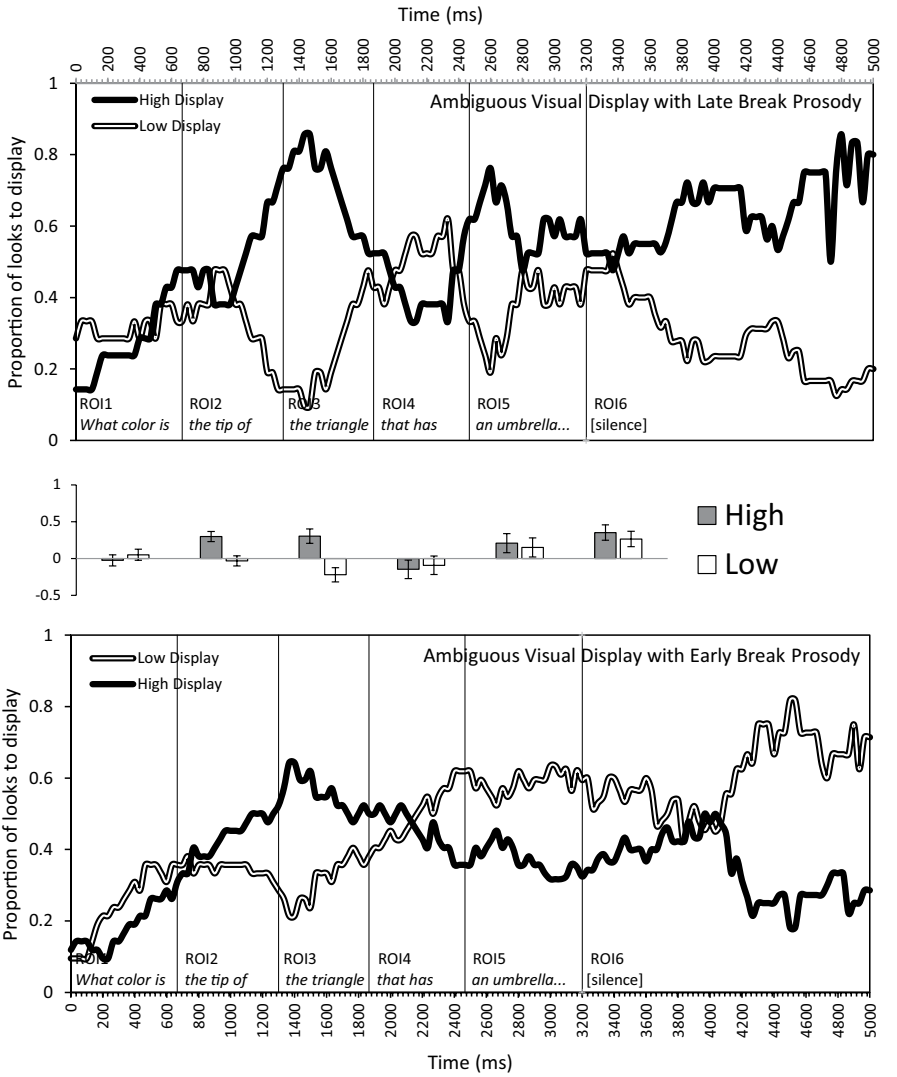


Fig. 7 Proportions of looks over time with ambiguous visual displays for late break prosody materials (*top* panel) and early break prosody materials (*bottom* panel). Looks to the high-attachment panel are in *solid black,* looks to the low-attachment panel are in *hollow black*. The *bar* graphs are the participant-based means and standard errors of looks to the panel that were matched with the prosody of the utterance for each of the six regions of interest (ROI1–ROI6), for late break prosody (*gray*) or early break prosody (*white*)

× display interaction was significant in ROI3 ($F_1(1,20)=8.59$, $p=0.008$, $SS=0.42$, $MSe=0.05$; $F_2(1,6)=6.49$, $p=0.044$, $SS=0.13$, $MSe=0.02$); main effects were not significant ($p>0.15$). Paired comparisons confirm that looks to the correct display, high attachment vs. low attachment, were significantly different (t_1 (20)=3.30, $p=0.004$; t_2 (6)=3.53, $p=0.012$). We attribute this effect to perceptual bias in visual scanning: Participants are more likely to look for the N1, *the tip,* as soon as they hear it, and out of the two tips, they prefer to fixate their gaze on the one that has an object depicted in it, be it the sun (Fig. 1a) or the umbrella (Fig. 1b).

ROI4 In this region, the effect of prosody in the materials emerges clearly, and is supported by the analyses (significant interaction, $F_1(1,20)=4.24$, $p=0.053$, $SS=0.12$, $MSe=0.09$; $F_2(1,6)=5.69$, $p=0.054$, $SS=0.29$, $MSe=0.05$; no significant main effects, $p>0.25$). The critical observation is that no preference is apparent yet with high-attachment materials: Looks to the correct high-attachment panel are not significantly different from looks to the mismatching panel (t_1 (20)=0.63, $p=0.535$; t_2 (6)=1.47, $p=0.191$). In contrast, with low-attachment materials, a preference for the correct display is reliably established in this region: Looks to the correct low-attachment panel are different from looks to the mismatching panel (t_1 (20)=3.37, $p=0.003$; t_2 (6)=2.12, $p=0.079$). This is not surprising if we consider that the prosody that encourages low attachment bears its crucial signal, i.e., an early phrasal break between N1 and N2 (at the beginning of ROI3), earlier in the time course of the utterance than the prosody that encourages high attachment, i.e., a late phrasal break after N2 (at the beginning of ROI4), so looks to the correct panel in the high-attachment disambiguated conditions can be established even before the disambiguating information in the RC is heard (ROI5).

ROI5–ROI6 In these regions, the preference for the correct display is solidified. The interaction observed in ROI4 becomes a trend in ROI5 only in the participant-based analyses ($F_1(1,20)=4.22$, $p=0.053$, $SS=0.09$, $MSe=0.02$; $F_2(1,6)=0.08$, $p=0.783$, $SS=0.01$, $MSe=0.04$). However, in ROI5, the main effect of looks to the correct display is highly significant ($F_1(1,20)=75.3$, $p<0.001$, $SS=3.64$, $MSe=0.05$; $F_2(1,6)=19.3$, $p=0.004$, $SS=0.96$, $MSe=0.05$). (For ROI5, the effect of disambiguation is not reliable, F_1, $F_2<1$.) In ROI6, the interaction and the main effect of disambiguation are not significant (F_1, $F_2<1$), but the main effect of looks to correct display is, as the graphs in Fig. 6 clearly illustrate, highly significant ($F_1(1,20)=193$, $p<0.001$, $SS=10.73$, $MSe=0.06$; $F_2(1,6)=173$, $p<0.001$, $SS=3.56$, $MSe=0.02$).

We now turn to the ambiguous displays, whose data are presented in Fig. 7. Recall that ambiguous displays were manipulated between participants, and there was an unequal number of items: Six ambiguous items were presented with the late break prosody and three with the early break prosody. We, therefore, analyzed the data separately for late break prosody (top panel in Fig. 7) and early break prosody (bottom panel in Fig. 7), in a design-crossing region (ROI1–ROI6) and display (matched, mismatched with the prosody) as within-participant and within-item factors.

For late break prosody materials (top panel in Fig. 7), the target dictated by the prosody is the panel that disambiguates toward high attachment. In the eye-tracking

record, this preference emerged only as a trend, rather late and very weak as the competing panel received attention through the very end. The region × display interaction was not significant ($F_1 = 2.13$, $p = 0.095$, $SS = 0.49$, $MSe = 0.05$; $F_2(5,10) = 1.35$, $p = 0.319$, $SS = 0.29$, $MSe = 0.04$); the main effect of display was also not significant ($F_1 < 1$; $F_2(1,2) = 8.47$, $p = 0.101$, $SS = 0.21$, $MSe = 0.03$). The main effect of region was significant ($F_1(5,25) = 46.5$, $p < 0.001$, $SS = 0.66$, $MSe = 0.01$; $F_2(5,10) = 16.8$, $p < 0.001$, $SS = 0.14$, $MSe = 0.01$), reflecting the fact that the proportion of looks to the high- or low-attachment panel was lower in ROI1 than all other ROIs. At ROI3, there are more looks to the panel that was disambiguated toward high attachment—the same perceptual bias to look at the part of the display with an object inside it observed with disambiguated visual displays—but the effect is only significant in the participant-based analysis (t_1 (5) = 2.75, $p = 0.040$; t_2 (2) = 1.50, $p = 0.273$). All other paired comparisons between looks to the high- or low-attachment panel are not significant ($p > 0.05$). Interpretation of these null effects warrants caution. The trends here might have emerged more clearly if the sample were larger, but this piece of the design has low item power.

For early break prosody materials (bottom panel in Fig. 7), the target dictated by prosody is the panel that disambiguates toward low attachment. In the eye-tracking record, this preference emerges late, but somewhat more clearly than with late break prosody. The region × display interaction was significant for early prosody data ($F_1(5,65) = 2.96$, $p = 0.02$, $SS = 1.06$, $MSe = 0.07$; $F_2(5,25) = 2.13$, $p = 0.095$, $SS = 0.49$, $MSe = 0.05$); the main effect of region was also significant ($F_1(5,65) = 21.9$, $p < 0.001$, $SS = 1.55$, $MSe = 0.01$; $F_2(5,25) = 46.5$, $p < 0.001$, $SS = 0.66$, $MSe = 0.01$); and the main effect of display was not significant (F_1, $F_2 < 1$). We conducted paired comparisons of looks to the high or low panel by ROI.

There was no preference for either display at ROI1 and ROI2 ($p > 0.50$). At ROI3, we observe the bias reported above for looks to the N1 element with an object inside it (t_1 (13) = 2.28, $p = 0.040$; t_2 (5) = 2.75, $p = 0.041$). There was no preference for either display at ROI4 or ROI5 ($p > 0.20$). The preference for the low-attachment panel emerges (marginally by items) in the final ROI, ROI6 (t_1 (13) = 2.52, $p = 0.025$; t_2 (5) = 1.53, $p = 0.186$). Again, low item power may have contributed to the weak effect of prosody on RC attachment in the sentences with ambiguous displays.

5 Discussion

The three types of data reported in the preceding section offer a multidimensional picture of the way the RC attachment ambiguity is processed in English, and how visual and prosodic sources of information influence interpretation.

First, this paradigm differs from other methods used to study prosody and RC attachment, in that the target is produced with interrogative (*wh*-question) intonation ("What color is the tip of the triangle that…?"). Perhaps interrogative intonation was behind the "whole-object" answers produced by our participants. This issue could be explored using materials requiring declarative intonation (e.g., "Click on the tip

of the triangle that…"). At the same time, the interrogatives could be more felicitous materials from the perspective of the discourse structure of the experimental trial: It is more natural to answer a question asked about a visual display than it is to listen to a sentence out of the blue and then answer a comprehension question.

Second, color-naming responses and response times suggested a minor advantage for low-attachment interpretations. Although color-naming responses were equally accurate when the disambiguation was visual and the matching prosody was present (disambiguated items), naming times were faster for materials visually disambiguated low, similar to data reported by January and Trueswell (2007). When the materials were ambiguous, early break prosody resulted in low-attachment interpretations significantly greater than chance and faster naming times, whereas late break prosody did not result in high-attachment interpretations significantly greater than chance. This asymmetry (early break prosody as a better signal for low attachment than late break prosody for high attachment) aligns with earlier findings for this construction in both English (Fernández 2007) and other languages (Sekerina et al. 2008; Shaked 2009; Stoyneshka et al. 2010; Teira and Igoa 2007). These findings suggest that prosody can have a facilitatory effect in terms of processing speed in resolution of RC attachment ambiguity.

Finally, the eye-tracking record offers insights about the time course of processing of the construction online. With both disambiguated and ambiguous visual displays, an effect emerged in ROI3, containing N2, with more looks to the panel containing an object in the part referred to by N1 (the tip with either the sun or the umbrella inside it); this was the high-attachment panel with high attachment and ambiguous materials and even the mismatching panel with low-attachment materials. We attribute this effect to a perceptual bias to look at N1 soon after it is heard, and of the two N1s available, the one with an object in it or with some property like stripes or polka dots is the one that is more perceptually salient (Huetting et al. 2011; Maas and Russo 2003). Perhaps this effect would disappear with a different design, like one using panel position to disambiguate (e.g., "What color is the tip of the triangle that is in the left panel?").

With materials disambiguated visually (as well as prosodically), we observed a significant effect of prosody for both high- and low-attachment materials. More looks were launched by the participants to the correct panel as early as the beginning of the RC, i.e., *that has*… (ROI4) for materials disambiguated low, but later in ROI5 for materials disambiguated high. This time course difference in the effect of prosody can only be attributed to differences between early and late break prosody as the sentence is disambiguated only in ROI5. The prosody that encourages low attachment—signaled by a phrasal break between N1 and the prepositional phrase (*the tip* | *of the triangle that*…)—is a more reliable cue to interpretation than the prosody that encourages high attachment—signaled by a phrasal break after N2 (*the tip of the triangle* | *that*…). Our eye-tracking data indicate that early break prosody is used to interpret RC attachments well before the content of the RC is available, which is a finding of theoretical importance, with implications for models of the human sentence processing mechanism. How exactly early break prosody influences interpretations online is a matter that will require further empirical work. Perhaps

the early break intonational contour is a non-default contour for this construction in English, and, therefore, a more salient cue for determining syntactic structure. Alternatively, early break prosody might simply be a more reliable indicator of attachment because it occurs earlier in the time course of the sentence, and, therefore, offers a durational advantage. Our dataset does not distinguish between these two alternative explanations, so this is a matter ripe for future research.

For materials that had ambiguous visual displays, with prosody as the only cue for possible disambiguation, our data suggest that prosody is used late in the time course of processing. A preference for the display that matched the prosody of the utterances emerged only tenuously in the ambiguous display data. We must interpret this result with caution, however, because item power was very low for this piece of the design, particularly for the materials with late break prosody. The low power in our ambiguous materials was something we tolerated from the outset, because this experiment was designed as a proof of concept. We would have had a more robust dataset if the ambiguous materials had included a full and within-participants manipulation of the prosody. However, this would have made the testing much longer and substantially more tedious. In addition to demonstrating that it is possible to use visual contexts with auditory materials to disambiguate RC attachments, we were interested in producing data with adults that we could then compare to data collected from children.

6 Conclusion

We have reported an investigation of the RC attachment ambiguity, using visual displays to disambiguate attachment and examining the role that overt prosody has in influencing attachment decisions. Indeed, visual displays are effective ways of inducing high versus low interpretations, particularly with early break prosody (which encourages low attachment) which seems to offer a more robust cue for attachment than late break prosody. The observed patterns correspond well with existing findings on how English speakers process this construction, and on how explicit prosody influences interpretations. Our data contribute to the empirical record by offering an informative look at the time course of processing RC attachments online. Prosody is used early, when the visual display cooperates, and early break (low-attachment) prosody facilitates reaching the interpretation earlier than late break (high-attachment) prosody. This difference between the two intonational contours used in our materials might be due to early break prosody having some sort of a phonological advantage (perhaps because it is a clearer signal of the underlying structure) over late break prosody. An alternative explanation is that the early break prosody has a purely durational advantage: By virtue of occurring earlier in the utterance, the early break affords the listener more time to use it as a signal of the syntactic structure.

In addition, this investigation contributes to the set of procedures that can be used to study RC attachment. It is a procedure that is quite user-friendly, so it can be used not only with adults but also with children as well as with low-proficiency

speakers and illiterate speakers. While it is not a procedure without complications, it is a paradigm that offers a flexible way to test hypotheses about the types of variables that can influence ambiguity resolution. Our focus was prosody, but the technique can be adapted to examine the time course of use of morphological information ("What color is the tip of the triangles that have…?"), for example. Finally, in its use of semantically shallow materials, with direct "real world" correlates, the technique avoids noise contributed by uncontrolled semantic/pragmatic biases in the materials.

References

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology, 54*(3), 218–250. doi:10.1016/j.cogpsych.2006.07.001.

Augurzky, P. (2006). *Attaching relative clauses in German: The role of implicit and explicit prosody in sentence processing*. Leipzig: Universität Leipzig.

Bergmann, C., Paulus, M., & Fikkert, P. (2012). Preschoolers' comprehension of pronouns and reflexives: The impact of the task. *Journal of Child Language, 39*(4), 777–803. doi:10.1017/S0305000911000298.

Carreiras, M. (1992). Estrategias de análisis sintáctico en el procesamiento de frases: Cierre temprano versus cierre último. *Cognitiva, 4,* 3–27.

De Vincenzi, M., & Job, R. (1993). Some observations on the universality of the late-closure strategy. *Journal of Psycholinguistic Research, 22,* 189–206.

Dussias, P. E., & Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition, 10*(01), 101. doi:10.1017/S1366728906002847.

Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition, 11*(3), 127–163.

Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins.

Fernández, E. M. (2005). The prosody produced by Spanish-English bilinguals: A preliminary investigation and implications for sentence processing. *Revista Da Abralin, 4*(1), 109–141.

Fernández, E. M. (2007). How might a rapid serial visual presentation of text affect the prosody projected implicitly during silent reading? In E. M. Fernández (Ed.), *Conferências do V Congresso Internacional da Associaçao Brasiliera de Lingüistica*, Vol. 5, pp. 117–154.

Ferreira, F., & Karim, H. (this volume). Prosody and performance in language production. In L. Frazier & E. Gibson (Eds.), *Explicit and implicit prosody in sentence processing*.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27*(2), 285–319.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In Proceedings of the SPEECH PROSODY 2002 Conference. Aix En Provence, France, April 2002.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

Gibson, E., Pearlmutter, N., Canseco-Gonzalez, E., & Hickok, G. (1996). Recency preference in the human sentence processing mechanism. *Cognition, 59*(1), 23–59. doi:10.1016/0010-0277(95)00687-7.

Gryllia, S., & Kügler, F. (2010). What does prosody tell us about relative clause attachment in German? *Speech Prosody 2010,* 100927:1–4. http://speechprosody2010.illinois.edu/papers/100927.pdf. Accessed: 24 Jan 2015.

Huetting, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica, 137,* 138–150.

January, D., & Trueswell, J. C. (2007). Relative clause attachment ambiguities in the visual world. Poster presented at the 20th Annual CUNY Conference on Human Sentence Processing. La Jolla, CA.

Maas, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science, 14*(4), 296–301.

Maia, M., Fernández, E. M., Costa, A., & Lourenço-Gomes, M. do C. (2007). Early and late preferences in relative clause attachment in Spanish and Portuguese. *Journal of Portuguese Linguistics, 5-2/6-1,* 227–250.

O'Grady, W., Suguzi, T., & Yoshinaga, N. (2010). Quantifier spreading: Evidence from Japanese. *Language Learning and Development, 6*(2), 116–125. doi:10.1080/15475440903352799.

Sekerina, I. A., Stromswold, K., & Hestvik, A. (2004). How do adults and children process referentially ambiguous pronouns? *Journal of Child Language, 31*(1), 123–152. doi:10.1017/S0305000903005890.

Sekerina, I. A., Fernández, E. M., & Petrova, K. (2008). Присъединяване на структурно многоэначни подчинени изречения в Българския еэик ('Processing of structurally ambiguous relative clauses in Bulgarian'). In S. Comati (Ed.), *Bulgaristica—Studia et Argument: Festschrift für Ruselina Nitsolova zum 65. Geburtstag* (pp. 328–336). Munich: Verlag Otto Sagner.

Shaked, A. (2009). *Attachment ambiguities in Hebrew complex nominals: Prosody and parsing*. New York: Graduate Center.

Stoyneshka, I., Fodor, J. D., & Fernández, E. M. (2010). Phoneme restoration methods for investigating prosodic influences on syntactic processing. *Language and Cognitive Processes, 25*(7), 1265–1293. doi:10.1080/01690961003661192.

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution. *Journal of Experimental Psychology: General, 136,* 64–81.

Teira, C., & Igoa, J. M. (2007). The prosody-syntax relationship in sentence processing. *Anuario de Psicología/The UB Journal of Psychology, 38*(1), 45–69.

Traxler, M. J., Pickering, M. J., & Clifton, C. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*(4), 558–592. doi:dx.doi.org/10.1006/jmla.1998.2600.

Van Dyke, J. A., & McElree, B. (2006). Retrieval interference in sentence comprehension. *Journal of Memory and Language, 55*(2), 157–166. doi:10.1016/j.jml.2006.03.007.

The Implicit Prosody of Corrective Contrast Primes Appropriately Intonated Probes (for Some Readers)

Shari R. Speer and Anouschka Foltz

Abstract Two visual-to-auditory cross-modal priming experiments looked for evidence of a link between the implicit prosodic contour readers generated during silent reading and the explicit prosodic contour of a subsequently presented auditory probe word. Pairs of text sentences that contained corrective contrasts (e.g., Jacquelyn didn't pass the test. Belinda passed the test) were immediately followed by probes pronounced with pitch accent patterns consistent (*BELINDA*) or inconsistent (*BELINDA*) with the corrective contrast in the read text. Participants were grouped according to individual differences in their pitch accent production while reading aloud in an independent task. Pitch accent production patterns were shown to correlate with the performance in the cross-modal task, providing initial evidence about the content of the auditory image produced as inner speech during silent reading.

Keywords Individual differences · Contrastive pitch accent · Implicit prosody · Inner speech · Cross-modal priming

1 Introduction

The implicit prosody hypothesis (IPH; Fodor 1998, 2002) proposes that a "default" prosodic contour is projected onto sentences during silent reading, and predicts a relationship between the projected contour and the final interpretation of the text. The IPH was initially proposed to explain the effect of constituent length on attachment decisions made after reading ambiguous relative clause (RC) sentences, such as *Someone shot the servant of the actress who was on the balcony (drink-*

S. R. Speer (✉)
Department of Linguistics, The Ohio State University, Columbus, OH, USA
e-mail: Speer.21@osu.edu

A. Foltz
School of Linguistics and English, Language, Bangor University, Bangor, Gwynedd, UK
e-mail: a.foltz@bangor.ac.uk

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_14

ing tea). Such off line judgments show that readers show a stronger preference for low attachment of a short ambiguous RC than of a longer one (e.g., Fernández and Bradley 1999). According to the IPH, a longer sentence-final RC induces a prosodic phrase break before the RC, setting it off from the two preceding noun phrases. Numerous experiments examining readers' syntactic judgments for ambiguous sentences in many languages (e.g., Fernández and Bradley 1999 (Spanish); Hirose 1999 (Japanese); Wijnen 2004 (Dutch); Vasishth et al. 2004 (Hindi), among others) have supported the IPH. Online studies of sentence reading have also shown indirect evidence for implicit prosody effects, such as shorter reading times for sentences with final syntactic interpretations that were consistent with their assumed default prosodic phrasing, as compared to those inconsistent with that phrasing (e.g., Fernández 2003).

Although explanations of implicit prosody effects associate them with the presence of an auditory image, or "inner speech" in the mind of the reader, there is no direct evidence tying such an image to the measured off line judgment and reading time effects. Such evidence would be valuable, as it would increase our understanding of the source of implicit prosodic contours within the language-processing system, and might lend support to claims about the nature and location of the specific prosodic features posited (such as accents and phrasing), on which implicit prosody-based arguments rest. There is some evidence that an auditory image generated during silent reading can affect subsequent processing of related auditory material. Abramson (2007) asked participants to listen to a pair of male and female interlocutors who each uttered one statement and one question. Participants then read silently a set of sentences that began with either "He said" or "She said" and were punctuated to indicate either a question or a statement. They were told the text items had been spoken by the interlocutors previously heard. After a 5-min delay, they gave auditory lexical decisions to words that had been final in the silently read sentences, pronounced with either rising or falling final intonation by a male or a female speaker. Lexical decision times were facilitated when sex, intonation, or both matched the text presentation of the words, as compared to non-matching items. However, in this design, it is difficult to determine whether the auditory lexical decisions were primed by the initial prosody of the interlocutor's intonation, the implicit prosody generated during the reading of the text sentences, or some combination of these factors. In addition, effects due to the evoked memory for the voice of a particular speaker that is not the reader may be different from those generated during normal reading.

Most studies on implicit prosody, however, provide only indirect information about such "inner speech." For example, several investigators have experimentally compared the overt prosodic phrasing readers produce when reading a sentence aloud to their implicit prosody (as determined by their preferred syntactic parse of the sentence during silent reading). These studies have shown mixed results. For example, Jun and Kim (2004; see also Hwang and Schafer 2009, and Jun and Koike 2003, for similar experiments with Japanese) conducted production experiments with Korean relative clauses, recording readers who either skimmed a text and then read it aloud, or read aloud without skimming. Participants then completed an off-

line syntactic judgment task for the same sentences. Results showed that judged syntactic preferences and Korean Tone and Break Indices (K-ToBI)-annotated overt prosodic phrasing patterns were consistent with one another—about a two-third preference for high-attachment readings and a corresponding phrasal break location about two thirds of the time, with skimmers showing a stronger high-attachment preference. Here, the correlation between off line preference and overtly produced prosodic phrasing supports the predictions of the IPH. However, similar work on English has not consistently supported IPH predictions. The most frequently produced prosodic phrasing pattern for ambiguous RC sentences in English read-aloud studies is one where the RC is set off in its own phrase (NP1 NP2) (RC), a pattern consistent with a high-attachment preference, when English is generally considered to "prefer" low attachment (Jun 2010). Bergmann and Ito (2007, 2009) asked participants to read ambiguous RC sentences aloud and manipulated the length of NP1 and of the RC (e.g., *The (defense) lawyer of the mayor who smokes (like a chimney) impressed the guest*). After reading aloud, participants answered a comprehension question indicating attachment, and overwhelmingly gave low-attachment interpretations. In contrast, prosodic boundaries were produced much more frequently at NP2 than NP1—the pattern expected for high-attachment interpretations. In addition, longer NP1s generated more prosodic boundaries at NP1, but longer RCs did not systematically generate more prosodic boundaries at NP2. This pattern of results is inconsistent with the IPH explanation of the low-attachment preference for English. Follow-up studies (Bergmann et al. 2008; Foltz et al. 2011) had participants either (a) read aloud an ambiguous RC sentence and then give a syntactic judgment or (b) silently read the sentence, then give a syntactic judgment, and then read the sentence aloud. Only for (a) were prosodic patterns correlated with readers' interpretations. In another experiment, a separate group of participants listened to the read sentences and judged the speakers' intended syntax. Only the prosodic patterns of the (a) sentences modulated participants' interpretations. Although these findings are inconsistent with the notion of a predictable default prosodic contour, it is not possible to know whether parsing preferences in reading were due to "inner speech" contours, because the silent and overt prosodies were necessarily produced on different trials.

Jun (2010) has argued that it might not be possible to assess implicit prosody by comparing readers' overt read-aloud prosody to the judgments they make while or after silent reading. The prosody of read speech is easily recognizable as such, and differs substantially from spontaneous speech and "laboratory speech" (speech created for use in experiments), with shorter constituent phrases (so more pitch accents and breaks; Howell and Kadi-Hanifi 1991). In both laboratory speech and spontaneous speech, the speaker begins with an intended message-level meaning that contributes to the generation of a corresponding prosodic structure. In contrast, reading aloud involves recovering a meaning provided by the writer, often a bit at a time as the words are recognized and produced. Thus, the overt prosody/message mapping may be shallow and based on minimal constraints as compared to that for implicit prosody, which readers may generate at any point during text comprehension. In addition, goals in reading aloud may differ from goals for silent reading,

with the former more focused on articulatory processes, such as preventing word-level pronunciation errors, and the latter more on message comprehension. Jun's (2010) production results showed that readers probably produced prosodic breaks at NP2 in long RC sentences before they had comprehended the lexical content of the RC, thus producing an infelicitous break pattern for the syntactic structure. This suggests that length constraints rather than meaning controlled the prosody produced. Some of these readers also produced hesitation lengthening later in the same utterances, suggesting the operation of a prosodic repair process based on self-monitoring as the overt prosody unfolds. These issues could easily result in stark differences between overt and implicit prosodies for read text, making it difficult to learn about the prosodic phrasal structure and accent patterns available during silent reading on the basis of productions gathered during reading aloud.

Individual differences may also contribute to inconsistent findings when comparing implicit to overt prosody, because readers may differ from one another and from themselves on repeated trials in the prosody they assign to a particular utterance. The fact that speakers produce a variety of prosodic structures for any given lexico-syntactic sequence is well established for the overt prosody of spontaneous and quasi-spontaneous speech. For example, 13 participants in a game task were restricted to the use of particular syntactic frames, but could insert lexical items to convey their intended meaning. ToBI annotations of pitch accents and phrasal tones showed that when pronouncing a prepositional phrase (PP) attachment ambiguity, they used 62 different prosodies for 78 productions of the high-attached version, and 87 different prosodies for 101 productions of the low-attached version (Schafer et al. 2000; Speer et al. 2011). Similarly, when ten uninstructed speakers in a separate study gave instructions to an interlocutor in a tree-decoration task, they produced 580 adjective–noun sequences in contrastive contexts (e.g., *blue ball* preceded by *green ball*). Annotation showed 223 prosodic patterns across these utterances (Ito and Speer 2008). There is some evidence that readers may differ in the implicit prosody they assign to text sentences as well. Swets et al. (2007) showed differences in attachment preferences for ambiguous relative clause sentences depending on the working memory capacity of the reader. When NP1 NP2 RC sentences were presented in their entirety, participants with less working memory capacity showed a stronger high-attachment preference than participants with greater capacity. The authors suggested that low-capacity participants had generated an implicit prosodic break between NP2 and the RC, while high-capacity participants were able to process the sentence as an integrated unit with a low-attachment final parse.

The role of implicit prosody in silent reading can also be seen in effects due to lexical stress, the pattern of strong and weak syllables associated with individual words in languages like English, Dutch, and German. Findings here are more consistent than those for the syntax-implicit prosody correspondence summarized above. The number of stressed syllables in a word has been demonstrated to affect reading time, such that a word with two stressed syllables (e.g., *RAdiAtion*) takes longer to read than one with the same number of syllables, but only one stressed syllable (e.g., *inTENsity*); this effect was interpreted to indicate that the time to prepare the implicit pronunciation of a word depends on the number of stressed

syllables it contains, as the overall number of syllables in a word did not itself correlate with reading time (Ashby and Clifton 2005). Additional evidence for the influence of the phonological representation of words on processing during silent reading comes from studies of stress-alternating verb–noun homographs. Breen and Clifton (2011, 2013) found longer reading times when the sentence context necessitated a revision of the syntactic category that included a revision of the lexical stress pattern than when it did not (e.g., revision of the noun *ABstract* to the verb *abSTRACT* was more costly than revision of the noun *rePORT* to the verb *rePORT*). In a separate experiment, they used limerick contexts to manipulate metrical expectations for the lexical stress pattern of a phrase-final homograph. For example, a consistent context for the verb *present* was (SMALL CAPS indicate metrically strong syllables): *There* ONCE *was a* CLEVer *young* GENT// *who* HAD *a nice* TALK *to* preSENT, while an inconsistent context was: *There* ONCE *was a* PENniless PEASant// *who* WENT *to his* MAster *to* preSENT. Evidence from eye tracking showed that when the metrically predicted location for lexical stress was inconsistent with the syntactic category of the phrase-final word, reading times were longer than when the metrical context was consistent. This difference was not shown for phrase-final noun homographs in metrically biasing contexts—that is, a processing penalty was found for revising from a strong–weak (SW) to a weak–strong (WS) lexical stress pattern, but not for revising from a WS to an SW pattern. These studies clearly implicate a conflict between implicit prosody and the final interpretation of a silently read sentence as the source of a processing deficit, and do so more convincingly than the previously discussed evidence from relative clause processing. However, there are two possible sources of the conflicting implicit prosody. On the one hand, sentence level metrical structure (implemented in spoken sentences by the alignment of pitch accents with particular lexically stressed syllables) was manipulated to create consistent and inconsistent contexts for the critical words. This would suggest that the effects were due to a sentence-level implicit prosodic representation. On the other hand, the critical words' lexical stress patterns were also either consistent or inconsistent with the syntactic role of the critical words. Thus it is possible that the source of the conflict between implicit prosody and the final interpretation of the sentences was at the lexical level, due to the necessity of reaccessing the properly stressed phonological form from the lexicon. But longer processing times could also have been due to the necessity of reassigning the location of sentence-level implicit pitch accents, or to some combination of these processes.

In an effort to find more direct evidence of the nature of the *sentence-level* auditory image present during silent reading, in the work presented here we conduct cross-modal priming experiments. In particular, we attempt to use the implicit prosody of a read text to prime an appropriately intonated probe word. Readers were induced to generate an implicit pitch accent pattern with paired text sentences that contained a corrective contrast, e.g., *Jacqueline didn't pass the test. Belinda passed the test.* (cf. Bock and Mazzella 1983, who used stimuli like *ARNOLD didn't fix the radio. DORIS fixed the radio*). Corrective contrasts were in either subject or verb position. The participant's task was to read the sentence pair and then respond to the auditory probe by indicating whether it had been the initial word in the second

sentence. The prosody of the probe words was manipulated to contain either a high-rising contrastive pitch accent (L+H*), or no pitch accent. When the corrective contrast was in subject position, readers should be induced to generate an implicit L+H* accent on the initial word in the second sentence, and show shorter response times for a spoken probe with an L+H* accent. But when the corrective contrast was in verb position, readers should be induced to generate a prosodic contour with an unaccented initial word in the second sentence, and show shorter response times to an unaccented probe. Bock and Mazzella (1983) showed faster sentence comprehension times when new information was accented and repeated information was not, as compared to the reversed pattern of accentuation. In experiment 1, we present results from this cross-modal priming task. In experiment 2, we group readers based on the prosody they used when reading aloud to see whether overt reading style can predict the pattern of response times in the cross-modal priming task. Note that we do not assume that read-aloud speech versions of sentences should have the same prosody as implicit prosody versions. Instead, we look at people's reading styles as an indicator of how their inner speech might sound.

We have employed a cross-modal priming paradigm in previous studies (Bergmann and Speer 2007a, b) to prime an appropriately intonated probe word by implicit prosody generated during silent reading. While the results from these studies were quite tentative, they have allowed us to pilot the methodology and develop a paradigm that may advance our understanding of the intonational composition of implicit prosody. In particular, we are using short sentences in contrasting pairs that will have been read silently and completely understood when the auditory probe is presented. In addition, the sentence location of the target word, whose implicit prosody is meant to prime an appropriately intonated auditory probe, is predetermined and consistent across trials. Our previous studies differed from this approach in that they combined self-paced reading with the presentation of auditory probes at varying, unpredictable sentence locations. We believe that our new approach has several advantages: It allows participants to silently read the complete sentence pairs in a natural reading situation and without the interruptions of the reading flow that are associated with button-presses in a self-paced reading task. This allows participants to assign an implicit prosodic pattern that is similar to what we would expect in natural reading tasks. In addition, a self-paced reading paradigm may have induced sentence-medial prosodic boundaries at button-press locations. A predetermined and consistent target word location may reduce the variability in response times because the participants' attention is drawn to the prime word, and the auditory probe is expected at a consistent time. In contrast, in previous studies response times may have been affected by varying levels of expectations about when an auditory probe might occur.

In addition to the above methodological difference, this study also focuses on a different aspect of sentence level prosody: Whereas our previous studies using cross-modal priming focused on edge tones, here, we shift our attention to pitch accents. This has the advantage of reducing possible individual differences in implicit prosody while still investigating a sentence-level prosodic phenomenon. The contrastive accent versus no accent manipulation in the short, simple sentence pairs

that we used is a prosodic feature that is assigned at the discourse or sentence level and cannot be simply associated with a particular lexical item as lexical stress is. The pitch accent manipulation is also in some sense tonally simpler and more predictable than phrasal boundaries in English. Prosodic phrasal boundaries that mark syntactic constituency in English potentially have variable amounts of final lengthening and many tonal shapes (in the ToBI system, intermediate phrase accents may be H, !H, or L, and boundary tones at intonation phrase breaks may also be either H or L, with the combinations creating many possible end contours). In contrast, repeated unaccented words have stressed syllables and words carrying corrective contrastive marking should have a high-rising contrastive pitch accent (L+H*). Some evidence for the frequency of L+H* used to mark contrast (although not corrective contrast) can be found in an analysis of spontaneous speech from a task in which speakers gave directions that included contrasting adjective-noun pairs such as "…hang a red ball. Now hang a green ball…." Contrast-bearing adjectives (*green* in the example) bore an L+H* pitch accent on 53 % of trials, while this accent appeared on only 3 % of adjectives in comparable but non-contrastive contexts such as "…hang a red ball. Now hang a green drum…" (Ito and Speer 2006). Thus, in the current study, the number of possible appropriate prosodies that might allow for the probe word to be primed by implicit prosody is reduced.

2 Experiment 1

In this experiment, we present data from a cross-modal priming task. Participants silently read pairs of sentences that contained a corrective contrast either in subject or verb position. They then responded *yes* or *no* to an auditory probe, saying whether it had been the first word in the second sentence. The pitch accent pattern of the probe word either matched or did not match the implicit prosody presumably generated during reading. If this procedure taps into the auditory image that people generate during silent reading, we predict faster response times for matching auditory probes than for mismatching ones.

3 Methods

3.1 Participants

Sixty-eight adult native English speakers participated in the study. Data from an additional two people were excluded due to too many missing values within one experimental condition. Participants were undergraduate students at The Ohio State University who received course credit for their participation.

Table 1 Experimental conditions with examples: CAPS indicate an auditory probe with an L+H* accent and SMALL CAPS indicate no accent

		Visual sentence pair type	
		Subject contrast	Verb contrast
Auditory probe	Match	Sentence pair: *Jacquelyn didn't pass the test.* *Belinda passed the test.* Auditory probe: BELINDA	Sentence pair: *Belinda didn't fail the test.* *Belinda passed the test.* Auditory probe: BELINDA
	Mismatch	Sentence pair: *Jacquelyn didn't pass the test.* *Belinda passed the test.* Auditory probe: BELINDA	Sentence pair: *Belinda didn't fail the test.* *Belinda passed the test.* Auditory probe: BELINDA

3.2 Materials

The materials for each trial of the experiment consisted of a sentence pair and an auditory probe. Each sentence pair started with a negated statement, such as *Belinda didn't fail the test*, followed by a sentence that used lexical contrast to elaborate on which part of the first sentence was being negated, for example, *Belinda passed the test*. Here, it is the failing that is being negated and *failing* is contrasted with *passing*. Each auditory probe presented a word produced either with a rising corrective contrastive accent (L+H*) or no accent. The sentence pairs and auditory probes were combined in a 2 × 2 design to create the four experimental conditions shown in Table 1.

There were two sentence pair contrast conditions: Either the second sentence presented a correction of the subject (subject contrast) or the verb (verb contrast) of the first sentence. We will call the second sentence in each pair the target sentence (e.g., *Belinda passed the test*) and the first word of this sentence the target word (e.g., *Belinda*). When participants read the subject contrast sentence pairs silently, a felicitous implicit prosody would locate a corrective contrastive accent on the subject of the second sentence, i.e., on the stressed syllable of the target word (*BELINDA*). In contrast, when participants read the verb contrast sentence pairs silently, no implicit accent should be assigned to the target word (*BELINDA*). In this case, the verb of the second sentence (*PASSED*) should receive an implicit corrective contrastive accent. The subject noun in verb contrast sentences should receive no accent for two reasons: first, it was mentioned in the immediately preceding sentence, and speakers generally refrain from accenting repeated, "old" information (cf. Bolinger 1961, 1986; Chafe 1974, 1976; Terken and Hirschberg 1994) and second, being at the beginning of the sentence and adjacent to a contrastively-accented word makes it a likely target for deaccenting (Hirschberg 2008).

There were also two auditory probe conditions: Auditory probes were recordings of the target word (*Belinda*) produced either with a prosodic contour that matched or that did not match the implicit prosody appropriate for the target word. For subject

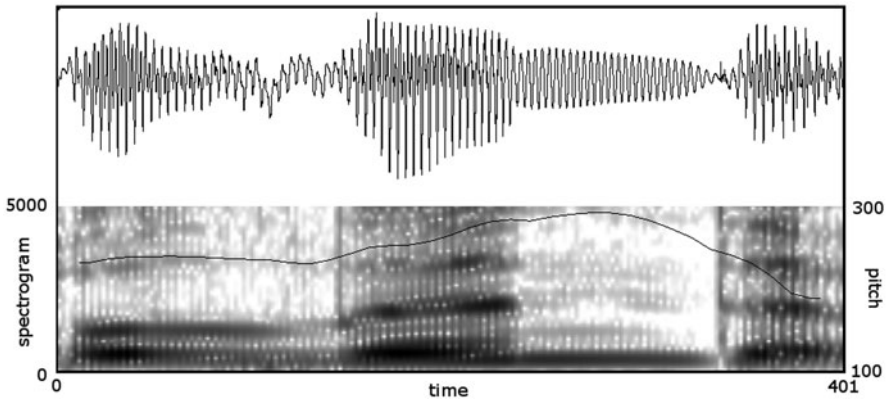


Fig. 1 Spectrogram and fundamental frequency contour for the auditory probe *BELINDA*, produced with a L+H* accent on the syllable with primary stress. The *x*-axis shows time in milliseconds, the left-hand and right-hand *y*-axes show pitch measured in Hertz

contrast sentence pairs, an auditory probe with a L+H* accent was considered to be a match, whereas an auditory probe with no accent was considered to be a mismatch. The reverse was the case for verb contrast sentence pairs: Here, an auditory probe with no accent was considered to be a match, and an auditory probe with a L+H* accent was considered to be a mismatch.

We created four experimental lists. Each list included 32 experimental items, eight in each of the four conditions, and 32 filler items. Experimental items were rotated across lists in a Latin square. Filler trials differed from experimental trials in two ways: Filler sentence pairs contrasted either in the sentences' objects or in prepositional phrases and filler auditory probes were recordings of words other than the target word (either another word in the sentence pair or a word that was phonetically similar to the target word).

The auditory probes were created as follows: A female native English speaker with phonetic training recorded the target sentence of each experimental sentence pair with two different prosodies, that is, either with a L+H* accent on the subject noun and deaccentuation on subsequent elements (e.g., *BELINDA passed the test*, where CAPS indicate an L+H*) or with no accent on the subject noun, a L+H* accent on the verb, and deaccentuation on subsequent elements (e.g., *belinda PASSED the test*, where CAPS indicate an L+H* and small caps indicate no accent). The first word of each sentence was then extracted, so that there were two auditory probes for each target sentence: one produced with an L+H* accent (BELINDA, see Fig. 1) and one produced with no accent (belinda, see Fig. 2). All experimental target sentences started with a three-syllable proper name with main stress on the second syllable (*Belinda*) and were thus long enough to produce clear prosodic patterns that remained even after the target word was excised from the rest of the sentence. Auditory probes with no accent differed reliably from auditory probes with a L+H* accent both in duration (average: 401 ms (standard deviation (*SD*) = 72) for

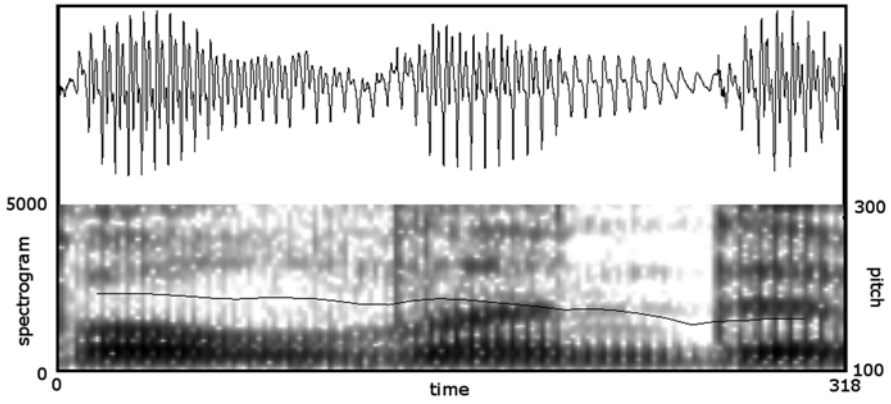


Fig. 2 Spectrogram and fundamental frequency contour for the auditory probe *BELINDA*, produced with no accent. The *x*-axis shows time in milliseconds, the left-hand and right-hand *y*-axes show pitch measured in Hertz

no accent versus 483 ms (SD = 67) for L+H*, paired *t*-test: $t = -12.4943$, $p < 0.001$) and in pitch maxima of the stressed syllable (average: 182 Hz ($SD = 15$) for no accent versus 289 Hz ($SD = 11$) for L+H*, paired *t*-test: $t = -32.1541$, $p < 0.001$). In addition, visual inspection of the fundamental frequency contours showed a high peak pattern toward the end of the vowel of the medial stressed syllable for L+H*-accented probes, and a relatively flat, downward-sloping contour across all three syllables for probes with no accent. The same speaker also recorded the target sentence for each filler sentence pair, either as it appeared in the experiment or with a phonetically similar proper noun in subject position. From these productions, words that were not target words (for example, the verb, object, noun of a prepositional phrase, or a word that phonetically resembles the target word) were extracted.

3.3 Procedure

Before the start of the experiment, participants received written and oral instructions for the experimental task, performed a practice session, and had the opportunity to ask questions. Then the experiment began. Participants were seated in front of a computer screen wearing noise-cancelling headphones (used to reduce distracting sound from other participants who participated in the same room at the same time). On each trial of the experiment, participants first silently read a sentence pair. Each sentence was displayed on a separate line to highlight the contrast between the first and second sentence. After reading the pair they pressed a button, upon which the text disappeared and an auditory probe was presented. To ensure that participants were reading the sentences at normal reading speed, the auditory probe was automatically presented after 3 s if participants had not pushed the button by then. Participants were instructed to respond to the auditory probes by deciding as

fast as they could whether or not the word they heard was the first word of the second sentence, i.e., the target word. To help them with this decision, the target word remained on the screen during the presentation of the auditory probe. After each decision, participants received feedback as to whether or not they had responded correctly. If participants had not responded after 2 s, they received a warning that their response was too slow. Following each response to an auditory probe, participants answered a comprehension question about the sentence pair. There was no time limit for responding to the comprehension question, and participants again received feedback as to whether or not they had responded correctly. After the experiment, participants completed a language background questionnaire.

4 Results

We tested whether participants responded faster to matching than to mismatching auditory probes. Such a result would suggest that participants are generating the expected implicit prosodic contours as they are reading silently and that our task taps into implicit prosody. Incorrect responses (i.e., responding "no" during a target trial, 2.2 % of the data points) and failures to respond within the 2-s limit (3 % of the data points) were excluded from the data set. Incorrect responses to the auditory probes occurred in fewer than 4 % of all target trials; accuracy for verb contrasts (proportion correct 0.98) showed a small but statistically significant advantage compared to that for subject contrasts (proportion correct 0.96) ($t = 2.69$, p<.05), but match and mismatch conditions did not differ ($t < 1$).

Response times under 200 ms and those that were two *SD*s above or below the mean for a given item and participant were also excluded from the data set. Altogether, 82 responses (8.3 %) were excluded from the data set. The response times across the four conditions are shown in Fig. 3. We ran mixed-effects models (cf. Jaeger 2008) with response time as the dependent variable and subjects and items as simultaneous random effects. We added probe type (match vs. mismatch), sentence contrast (subject contrast vs. verb contrast), and the probe type x sentence contrast interaction as fixed effects to the initial model. Probe type is the fixed effect of interest for our research question.

Redundant fixed effects were removed from the initial model until the model was minimally optimized. Random slopes were added if they improved model fit (Barr 2013; Barr et al. 2013). The final model included sentence contrast and the probe type x sentence contrast interaction as fixed effects. Response times were shorter for probes following verb contrast sentence pairs than following subject contrast sentence pairs (*estimate* = 9.345, $t = 2.168$, $p < 0.05$). To explore the reliable probe type x sentence contrast interaction (*estimate* = −12.998, $t = -3.438$, $p < 0.001$), we fit separate mixed-effects models for subject contrast sentence pairs and verb contrast sentence pairs. Each model had response time as the dependent variable, subjects and items as simultaneous random effects, and probe type (match vs. mismatch) as fixed effect. Random slopes were added if they improved model fit. The

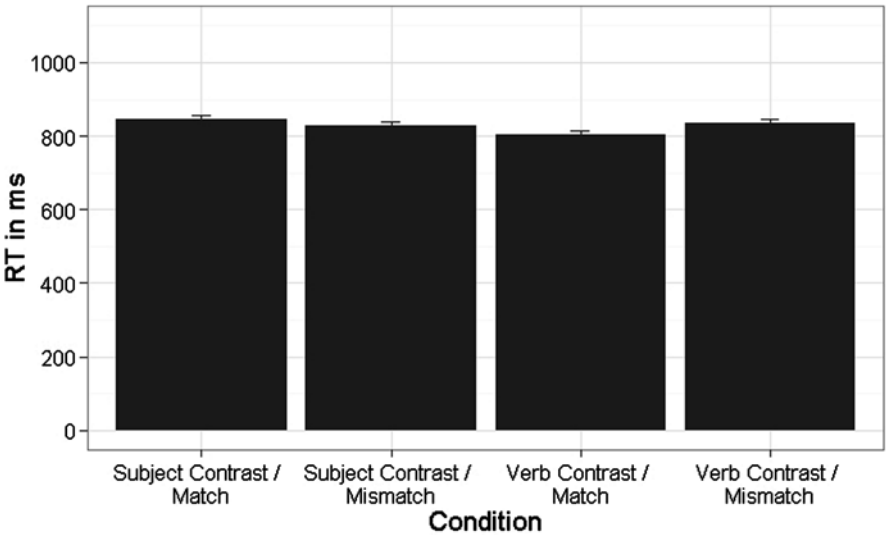


Fig. 3 Experiment 1. Response times for auditory probes in the four critical conditions

results from these models reveal that participants responded marginally faster to mismatching probes than to matching ones for the subject contrast sentence pairs (*estimate* = −8.957, $t = -1.663$, $p = 0.097$), but that they responded reliably faster to matching probes compared to mismatching probes for the verb contrast sentence pairs (*estimate* = 16.949, $t = 3.185$, $p < 0.01$).

In sum, participants responded faster to simpler no accent probes compared to more complex L+H* probes. In contrast to our expectations, probe type (match vs. mismatch) had no effect. To understand why this was the case, we had a closer look at individual participants' data. Individual participants' response patterns were grouped into four categories, based on how they responded to the different kinds of auditory probes: Those who showed shorter response times for matches than for mismatches (shorter matches), those who showed shorter response times for mismatches than for matches (shorter mismatches), those who showed shorter response times for no accent probes compared to L+H* probes (shorter no accent), and those who showed shorter response times for L+H* probes compared to no accent probes (shorter L+H*). Forty-one percent of participants (28 out of 68) fell into the shorter no accent category, followed by shorter matches (28%, 19 out of 68), shorter mismatches (21%, 13 out of 68), and shorter L+H* (12%, 8 out of 68). Thus, the two most common response patterns were shorter no accent and shorter matches. That is, we found that an unexpectedly large percentage of participants showed a processing advantage for the no accent probe conditions. This is also reflected in the reliable probe type x sentence contrast interaction. The expected response pattern (shorter matches) was only the second most frequent pattern. One possibility for why shorter matches was not the most frequent pattern is that there are individual

differences in participants' implicit prosody, such that only a portion of the participants responded to the auditory probes the way we expected. We explore this possibility in experiment 2.

5 Discussion

In experiment 1, we tested whether we could use a cross-modal priming paradigm to tap into the implicit prosodic contours that people generate when they read silently. We assumed that participants would generate an implicit L+H* accent on a correctively contrasted subject and no implicit accent on a repeated subject. If the prosodic contour of the auditory probe word was more similar to the implicit prosody image generated for the target word, as in the match conditions, we predicted that participants would show a processing advantage when deciding that the probe word was the same as the one at the beginning of the read target sentence. In contrast to our predictions, however, we found no effect of probe type, such that matching probes did not elicit shorter response times than mismatching probes. What we did find were reliable effects of sentence contrast, with faster responses to probes following verb contrasts than to probes following subject contrasts, and of the probe type x sentence contrast interaction, with faster responses to no accent probes compared to L+H* probes. Participants may have responded more quickly to probes following verb contrasts than to probes following subject contrasts because it is easier to confirm that the auditory probe word is the same as the target word if the target word is a repetition from the first sentence than if the target word contrasts with the first word of the first sentence. In other words, it may be easier to confirm that the auditory probe was a recording of the proper name shown twice in the written sentences than to confirm that the auditory probe was a recording of one of the two different proper names shown in the written sentences, in particular, of the proper name shown in the second written sentence. Participants may have responded more quickly to probes with no accent compared to probes with a L+H* accent because probes with no accent may require less involved processing than probes with a more complex bi-tonal pitch accent (L+H*). Similar effects have been found for edge tones: Participants in Bergmann and Speer (2007b) showed shorter response times to probes without an edge tone and probes with only a phrase accent than to probes with a boundary tone. Thus, participants were faster to respond when the probe was prosodically less complex. Alternatively, participants may have responded more quickly to probes with no accent compared to probes with a L+H* accent because no accent probes were reliably shorter in duration than L+H*. Such shorter duration may have allowed for faster recognition and thus faster response times.

An exploratory post-hoc analysis of individual participants' response patterns revealed that a large percentage of participants responded faster to no accent probes compared to L+H* probes. What we expected, however, was that a large percentage of participants would respond faster to matching probes compared to mismatch-

ing ones. How can we explain then that so many participants showed a different pattern of responses? As mentioned above, one possibility is that participants simply responded faster to shorter compared to longer probes or simple compared to more complex pitch accents. However, another possibility is that there are individual differences in participants' reading styles, and thus in the implicit prosodies they generate. In particular, it is possible that a sizeable portion of our participants read (both silently and aloud) in a rather monotone way and produced few L+H* accents during reading. If this is the case, these participants may frequently have generated implicit accent patterns unlike those instantiated in our L+H* probes. Thus, for these participants no accent probes may have always been more similar to their internal read speech, regardless of sentence contrast. It would then not be surprising to find a reliable probe type x sentence contrast interaction with faster responses for no accent compared to L+H* probes rather than a reliable effect of probe type (match vs. mismatch) in this study. In experiment 2, we explore the idea that the unexpected results of this experiment were due to such individual differences.

6 Experiment 2

In this experiment, participants read aloud a brief text after performing the same task as in experiment 1. We added a reading aloud task since the results from experiment 1 led us to hypothesize that participants differed in the implicit prosody they generated during silent reading and that these differences may have affected response times. The reading aloud task allows us to group people based on measurable prosodic phenomena. In particular, we can group people based on whether or not they produce L+H* accents when reading aloud. If a related mechanism is involved for generating implicit prosody during silent reading and overt prosody during reading aloud, participants who produce L+H* accents when reading aloud should be more likely to generate implicit L+H* accents during a silent reading task. And if our procedure taps into the auditory image generated during silent reading, it is these participants who should respond more quickly to matching auditory probes than to mismatching ones.

7 Methods

7.1 Participants

Participants were visitors to the "language pod" exhibition laboratory at the Center for Science and Industry (COSI), in Columbus, Ohio. They ranged in age from 15 to 60 years, and had educational backgrounds ranging from incomplete high school diplomas to advanced degrees. All were native speakers of Midwestern American

English, had normal hearing and normal or corrected-to-normal vision. Data from 27 participants were included in the study. All included participants reported hearing an auditory image of read words during silent reading. Data from an additional eight people were excluded due to too many missing values within a single experimental condition.

7.2 Materials

The materials were the same as those of experiment 1. In addition, we used the following text passage, adapted from a 2012 *New York Times* article, for the reading aloud task:

> The more automobile design has changed, the more it has remained the same. A century after the Model T Ford debuted, the vast majority of the cars on the road still feature steel bodies, chassis suspended on four wheels and four-stroke internal combustion engines. Not that would-be revolutionaries haven't tried to "improve" the automobile with a host of innovations: Bodies made of carbon-fiber. Bodies fitted with wings. Bodies that float on water. Three-wheelers. Six-wheelers. Steam engines. Jet engines. For a while, there was even talk of nuclear power. But designers don't control how cars are built, manufacturers control how cars are built.

7.3 Procedure

The procedure was the same as in experiment 1, with one addition: After performing the experiment, participants were recorded reading aloud the above text passage. No connection between the two reading tasks was mentioned to participants.

7.4 Annotation

A coder with extensive training ToBI-annotated (Beckman et al. 2005) the following relevant sentences from the text passage: *Not that would-be revolutionaries haven't tried to "improve" the automobile with a host of innovations* and *But designers don't control how cars are built, manufacturers control how cars are built*. These sentences were chosen because they contained words that participants most frequently pronounced with emphasis or contrast. In particular, *improve* may have received a L+H* accent because it was visually highlighted with quotes, and *manufacturers* may have received a L+H* accent because it is set in contrast to *designers*. Based on these annotations, participants were divided into three groups: L+H* users, H* users, and monotone readers. Twelve participants fell into the L+H* user group. Nine of these participants produced a L+H* on *improve*. Eight of these participants produced a L+!H* and three a L+H* on *manufacturers*. These participants also deaccented repeated material. For example, all of the 12 L+H* users

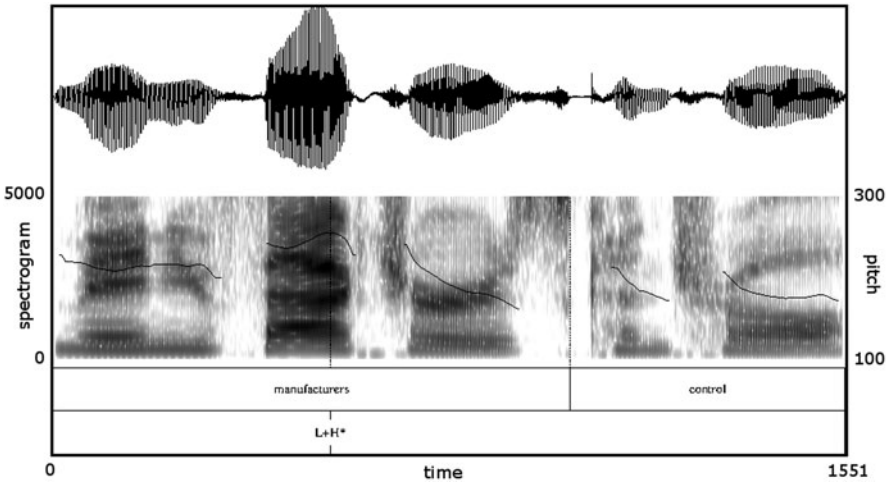


Fig. 4 Spectrogram and pitch track for *manufacturers control*, produced by a L+H* user. *Manufacturers* receives a L+H* accent on the syllable with primary stress and *control* is deaccented. The *x*-axis shows time in milliseconds, the left-hand and right-hand *y*-axes show formants and fundamental frequency, respectively, both measured in Hertz

deaccented the repeated verb *control* that followed *manufacturers*. To illustrate this, Fig. 4 shows the spectrogram and pitch track of *manufacturers control* produced by an example L+H* user. Twelve further participants fell into the H* user group. Seven of these participants produced a H* and one additional participant a H+!H* on *improve*. Nine of these participants produced a H* on *manufacturers*. These participants did not deaccent the repeated verb *control*. Instead, most of them produced *control* with a H* or !H* accent. This is illustrated in Fig. 5, which shows the spectrogram and pitch track of *manufacturers control* produced by a H* user. Three participants were considered monotone: They produced no accents or a L* accent on *improve* and *manufacturers*.

8 Results

We tested whether the L+H* users identified above responded faster to matching than to mismatching auditory probes. Notice that this group L+H*-accented a correctively contrasted noun and did not accent the repeated verb in the read-aloud passage. We thus expect that the auditory probes that we call matches are indeed matches for this group of people, both for the subject contrast sentences, where the matching auditory probe has a L+H* accent, and the verb contrast sentences, where the matching auditory probe has no accent. To test whether the response pattern that we expect for L+H* users is particular to this group, we also analyzed the data from the H* users. We therefore added the factor participant group (L+H* users vs. H*

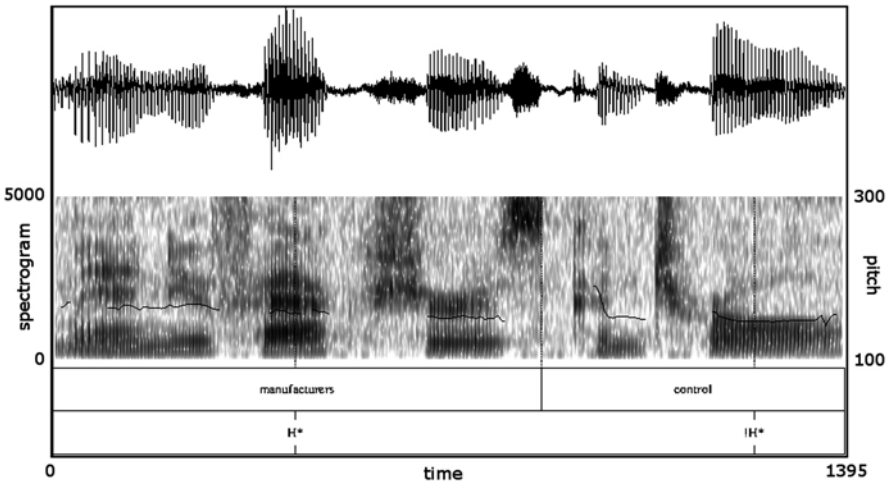


Fig. 5 Spectrogram and pitch track for *manufacturers control*, produced by a H* user. *Manufacturers* receives a H* accent and *control* a !H* accent on the syllable with primary stress. The *x*-axis shows time in milliseconds, the left-hand and right-hand *y*-axes show formants and fundamental frequency, respectively, both measured in Hertz

users) to the analyses reported below. We expect that the H* users do not show a processing advantage for matching compared to mismatching probes. An analysis of data from the monotone readers is not possible due to the small sample size.

Again, failures to respond, incorrect responses, response times under 200 ms, and response times that were two *SD*s above or below the norm for a given item were excluded from the data set. Incorrect responses to the auditory probes occurred in fewer than 10 % of target trials for both the L+H* and H* subject groups in all four conditions, and did not differ significantly among them (all $ts < 1.2$).

The response times across the four conditions are shown in Fig. 6 for the L+H* users and in Fig. 7 for the H* users. We again ran mixed-effects models with response time as the dependent variable and subjects and items as simultaneous random effects. We added probe type (match vs. mismatch), sentence contrast (subject contrast vs. verb contrast), participant group (L+H* users vs. H* users), and all interactions as fixed effects to the initial model. Here, the probe type x participant group interaction is the fixed effect of interest for our research question.

Redundant fixed effects were removed from the initial model until the model was minimally optimized. Random slopes were added if they improved model fit. The final model included sentence contrast (subject contrast vs. verb contrast) and the probe type x participant group interaction as fixed effects. As in experiment 1, participants were faster to respond to probes following verb contrast sentence pairs than following subject contrast sentence pairs ($estimate = -27.877$, $t = -4.148$, $p < 0.001$). To explore the reliable probe type x participant group interaction ($estimate = 17.677$, $t = 2.563$, $p < 0.05$), we fit separate mixed-effects models for L+H* users and H* users. Each model had response time as the dependent variable, sub-

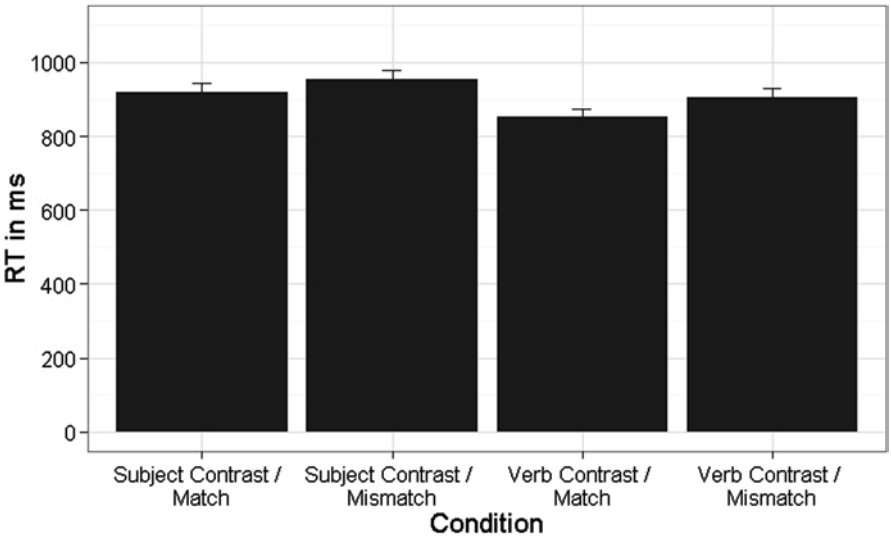


Fig. 6 Experiment 2. Response times for auditory probes for the 12 L+H* users in the four critical conditions

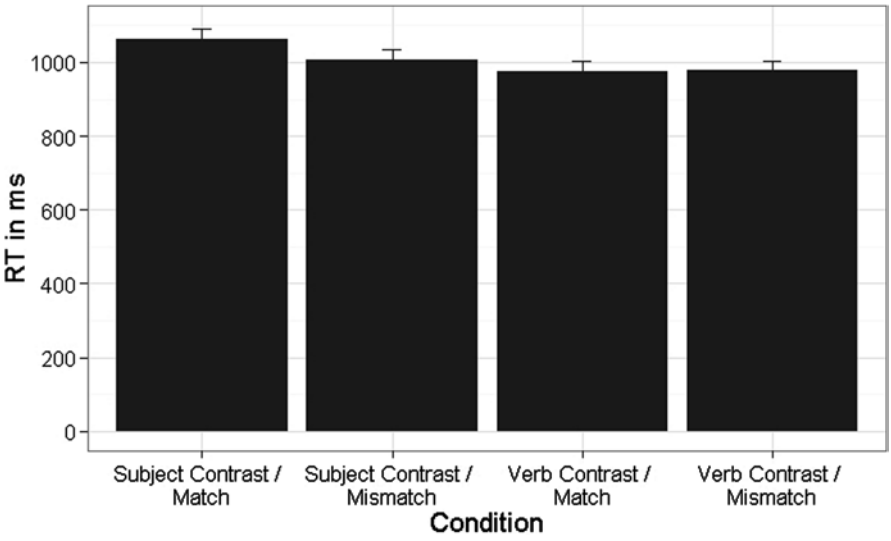


Fig. 7 Experiment 2. Response times for auditory probes for 12 H* users in the four critical conditions

jects and items as simultaneous random effects, and probe type (match vs. mismatch) as fixed effect. Random slopes were added if they improved model fit. The results from these models reveal that, importantly, L+H* users responded reliably faster to matching probes than to mismatching probes (*estimate* = 25.428, $t = 2.954$,

$p < 0.01$). In contrast, H* users responded equally fast to matching and mismatching probes (*estimate* = −8.551, $t = -0.808$, $p = 0.42$).

Thus, those participants who produced L+H* accents on correctively contrasted material and no accent on repeated material when reading aloud showed the expected response pattern: A processing advantage for L+H* probes following subject contrast text sentence pairs, and a processing advantage for no accent probes following verb contrast text sentence pairs. This suggests that these readers may have generated L+H* accents and deaccentuation during silent reading, leading to shorter response times to probes that matched the generated implicit prosody than to probes that did not match. The H* users differed from the L+H* users in their response to the auditory probes: Whereas the L+H* user group's response times were shorter for matches than for mismatches, the H* user group showed no effect of probe type. Thus, as expected, only the L+H* user group showed faster responses for matches than for mismatches.

9 Discussion

In this experiment, we looked for evidence of the implicit prosody readers were induced to generate for sentences with corrective contrasts by testing whether differences in the prosody participants used when reading aloud co-occurred with differences in their pattern of responding in our cross-modal priming task. We found that participants who produced L+H* accents and deaccentuation when reading aloud responded faster to matching auditory probes than to mismatching auditory probes. Since a reliable effect of probe type (match vs. mismatch) relies on participants generating implicit L+H* accents on contrasting items and no accent on repeated items in the experimental task, the data from this experiment provide some evidence that participants who produced measureable L+H* accents and deaccentuation in the read-aloud task also generated implicit L+H* accents on contrast items and no accent on repeated items during the silent reading task. In contrast, participants who produced H* accents on contrast items during the read-aloud task showed no reliable differences in responding to the probe type (match vs. mismatch) manipulation. The absence of a reliable three-way interaction also suggests that, unlike the participant group in experiment 1, the H* users did not respond reliably faster to no accent probes compared to L+H* probes (even though, as shown in Fig. 7, their responses were numerically faster for no accent than for L+H* probes). Thus, the H* user group seemed to show no sensitivity to the auditory probe manipulation at all. One possible explanation for this result is that neither the L+H* nor the no accent auditory probes matched the implicit prosody that the H* users generated during the silent reading task. Instead, H* probes might have been needed in order to constitute "matches" for this group of participants. If so, since the experiment did not contain any target auditory probes with a H* accent, neither probe was primed by the implicit prosody of the silently read text for this group.

We draw two conclusions from this experiment. First, the results suggest that there is some similarity between the prosody produced while reading aloud and that produced while reading silently: It was only the participants who produced measurable L+H* accents on the corrective contrast word and deaccentuation on repeated words when reading aloud that showed a priming response for matching auditory probes in the cross-modal task. Participants who produced H* and !H* accents when reading corrective contrast material aloud showed no sensitivity to the differences between auditory probes. Thus, the L+H* user and the H* user groups behaved differently from each other in both tasks. In addition, both groups behaved consistently across the two tasks, i.e., their read-aloud prosody predicted their performance in cross-modal priming in the silent reading task. This suggests that participants' propensity to use L+H* accents for corrective contrasts and deaccentuation for repeated items was similar for reading aloud and for reading silently. This is a potentially important result. Since we can neither hear nor phonetically measure implicit prosody, we can't know how readers' implicit prosody "sounds," nor can we know if the auditory image generated during reading is the same for every reader. The results from this experiment do give us some insight into how we might assess the "sound" of readers' implicit prosody. In particular, the results suggest that there may not be a one-to-one match between overt and implicit prosody, but that more general characteristics of speech read aloud are also found in silently read speech. One of these characteristics is one's propensity to produce salient L+H* accents for corrective contrast words and no accent for repeated words. The second conclusion that we draw is that the current cross-modal priming paradigm in combination with a read-aloud diagnostic task may be well suited to study the sound of implicit prosody.

10 General Discussion

This paper presents findings from two visual-to-auditory cross-modal priming experiments designed to investigate whether the implicit prosody generated during silent reading can prime an appropriately intonated auditory probe. Our results indicate a qualified "yes" to this question: We found evidence that, for speakers who prosodically marked corrective contrasts and orthographically marked words with a salient rising pitch accent (L+H*) followed by a deaccented region in oral reading, an appropriately intonated probe word could be primed by a corrective contrast in preceding silently read text.

While our first experiment showed no effect of whether the prosody of the probe was consistent with the location of the corrective contrast in the visual sentence pair, it did show effects of the corrective contrast location, with shorter participant responses for verb contrasts, which involved repetition of the sentence-initial target word. In addition, experiment 1 showed longer response times for auditory probes with L+H* accents than for those with no accent. This overall pattern of responding was repeated in experiment 2 for the H* subject group, who showed

significantly longer response times for verb contrast trials, and numerically longer times for L+H* probes within contrast types. The L+H* participant group in experiment 2 provides initial evidence that a well-known prosodic pattern, that for corrective contrast, can be evoked by a sentence pair presented in text and used to prime a subsequent auditory pitch accent pattern. These results are consistent with previous findings that silently read statements and questions can speed processing of subsequently presented auditory words with falling or rising intonation, respectively (Abramson 2007). However, we needed to resort to an overt reading task to provide information about the implicit prosody we might expect to induce from individual readers. This suggests that individual differences may obscure results from the priming paradigm, especially when there are multiple potential grammatical prosodies available for a particular read text.

Such individual differences may be the reason why studies involving word-stress manipulations have so far yielded more consistent results than studies involving sentence-level prosodic phrasing regarding both the existence of an implicit prosodic contour generated during silent reading and information about what this implicit prosody may sound like. While there are some words whose stress pattern is affected by the sentential context (e.g., *He's sixTEEN* vs. *SIXteen candles*) or is subject to individual differences (e.g., the noun *address* can be pronounced with stress either on the first or the second syllable), English stress is a word-level phenomenon, i.e., stress is a property of individual lexical items. As such, there is little room for individual differences when it comes to the stress patterns of particular lexical items. In contrast, there are numerous felicitous pitch accent patterns for the sentence contrasts that readers experienced in the silent reading task in this study. Even though Bock and Mazzella (1983) found a processing advantage for the pattern that we hypothesized to be readers' most common implicit prosodic contour (with an L+H* accent on corrective contrasts and no accent on repeated material), the prosodies produced in our overt reading task suggested that there may be two approximately equally common patterns that readers generated during silent reading, along with a far less likely "monotone" pattern. That is, we found both the hypothesized contour and one with a H* on corrective contrast words and a !H* on repeated material. Interestingly, the rate of L+H*/ no accent versus H*/ !H* production for corrective contrast across participants in our production study seems comparable to that found previously for the spontaneous production of contrastive adjective sequences (Ito and Speer 2006). Any study investigating sentence-level implicit prosody will likely have to deal with such individual differences. Thus, it may not be possible to study sentence-level implicit prosody without recourse to participants' overt prosody. The advantage of the visual-to-auditory cross-modal priming paradigm presented here is that a very brief and simple diagnostic allows grouping participants based on certain prosodic phenomena, so that implicit prosody can then be studied without comparison of implicit and overt prosody on a sentence-by-sentence basis. Indeed, our results suggest that such comparisons, as have been done in previous work (Bergmann and Ito 2009; Foltz et al. 2011; Hwang and Schafer 2009), may not be useful: rather, measurable general characteristics and tendencies found in overt prosody from reading aloud can be used to predict

implicit prosodic behavior and group participants based on these predictions to then see if the cross-modal priming paradigm may be used to confirm the predictions. However, since the current experiments differ from the previous work in that they test cross-sentential pitch accent patterns rather than implicit prosodic phrasing, this conclusion may be premature.

References

Abramson, M. (2007). The written voice: Implicit memory effects of voice characteristics following silent reading and auditory presentation. *Perceptual and Motor Skills, 105,* 1171–1186.

Ashby, J., & Clifton, C. Jr., (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96,* B89–B100.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4,* 1–2.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68,* 255–278.

Beckman, M. E., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–54). Oxford: Oxford University Press.

Bergmann, A., & Ito, K. (2007). *Attachment of ambiguous RCs: A production study*. Talk given at the 13th annual conference on Architectures and Mechanisms for Language Processing (AMLaP), Turku, Finland.

Bergmann, A., & Ito, K. (2009). *Production and comprehension of interpretation-driven versus input-driven speech*. Poster presented at 22nd annual CUNY conference on human sentence processing, Davis, CA.

Bergmann, A., & Speer, S. R. (2007a). *On priming by implicit prosody*. Poster presented at the 20th annual CUNY conference on human sentence processing, La Jolla, CA.

Bergmann, A., & Speer, S. R. (2007b). *More on priming by implicit prosody*. Poster presented at the 13th annual conference on Architectures and Mechanisms for Language Processing (AMLaP), Turku, Finland.

Bergmann, A., Ito, K., & Maday, K. (2008). *Order effects in production and comprehension of prosodic boundaries*. Talk given at TIE3: The third TIE conference on tone and intonation, Lisbon, Portugal.

Bock, J. K., & Mazzella, J. R. (1983). Intonational marking of given and new information: Some consequences for comprehension. *Memory and Cognition, 11,* 64–76.

Bolinger, D. (1961). Contrastive accent and contrastive stress. *Language, 37,* 83–96.

Bolinger, D. (1986). *Intonation and its parts*. London: Edward Arnold.

Breen, M., & Clifton, C. Jr. (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language, 64,* 153–170.

Breen, M., & Clifton, C. Jr. (2013). Stress matters revisited: A display change experiment. *Quarterly Journal of Experimental Psychology, 66*(10), 1896–1909.

Chafe, W. (1974). Language and consciousness. *Language, 50,* 111–133.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and points of view. In C. Li (Ed.), *Subject and topic* (pp. 25–56). New York: Academic.

Fernández, E. M. (2003). *Bilingual sentence processing: Relative clause attachment in English and Spanish*. Amsterdam: John Benjamins.

Fernández, E. M., & Bradley, D. (1999). *Length effects in the attachment of relative clauses in English*. Poster presented at the 12th annual CUNY conference on human sentence processing, New York.

Fodor, J. D. (1998). Learning to parse. *Journal of Psycholinguistic Research, 27*(2), 285–319.

Fodor, J. D. (2002). Prosodic disambiguation in silent reading. *NELS, 32,* 113–132.
Foltz, A., Maday, K., & Ito, K. (2011). Order effects in production and comprehension of prosodic boundaries. In S. Frota, G. Elordieta, & P. Prieto (Eds.), *Prosodic categories: Production, perception and comprehension* (pp. 39–68). Dordrecht: Springer.
Hirose, Y. (1999). *Resolving reanalysis ambiguity in Japanese relative clauses*. Unpublished doctoral dissertation, CUNY Graduate Center, New York.
Hirschberg, J. (2008). Pragmatics and intonation. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics*. Oxford: Blackwell Publishing Ltd.
Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech. *Speech Communication, 10,* 163–169.
Hwang, H., & Schafer, A. J. (2009). Constituent length affects prosody and processing for a dative NP ambiguity in Korean. *Journal of Psycholinguistic Research, 38,* 151–175.
Ito, K., & Speer, S. R. (2006). Using interactive tasks to elicit natural dialogue. In P. Augurzky & D. Lenertova (Eds.), *Methods in empirical prosody research*. Berlin: Mouton de Gruyter.
Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58,* 541–573.
Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.
Jun, S.-A. (2010). The implicit prosody hypothesis and overt prosody in English. *Language and Cognitive Processes, 25*(7), 1201–1233.
Jun, S.-A., & Kim, S. (2004). *Default phrasing and attachment preferences in Korean*. Proceedings of INTERSPEECH-ICSLP, Jeju, Korea.
Jun, S.-A., & Koike, C. (2003). *Default prosody and RC attachment in Japanese*. Talk given at the13th Japanese-Korean Linguistics Conference, Tucson, AZ. [Published in *Japanese-Korean Linguistics 3*, 41_53, CSLI, Stanford, in 2008].
Schafer, A. J., Speer, S. R., Warren, P., & White, D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research, 29,* 169–182.
Speer, S. R., Warren, P., & Schafer, A. J. (2011). Situationally independent prosody. *Laboratory Phonology, 2*(1), 35–98
Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General, 136,* 64–81.
Terken, J., & Hirschberg, J. (1994). Deaccentuation of words representing given information: Contributions of persistence of grammatical function and surface position. *Language and Speech, 37,* 125–145.
Vasishth, S., Agnihotri, R. K., Fernández, E. M., & Bhatt, R. (2004). *Noun modification preferences in Hindi*. In The Proceedings of Seminar on Construction of Knowledge, Vidya Bhawan Education Resource Centre, Udaipur, India, pp. 160–171.
Wijnen, F. (2004). The implicit prosody of jabberwocky and the relative clause attachment riddle. In H. Quené & V. van Heuven (Eds.), On speech and language. *Studies for Sieb G. Nooteboom* (pp. 169–178). Utrecht: Landelijke Onderzoeksschool Taalwetenschap. (LOT Occasional Series 2).

Inner Voice Experiences During Processing of Direct and Indirect Speech

Bo Yao and Christoph Scheepers

Abstract In this chapter, we review recent research concerned with "inner voice" experiences during silent reading of direct speech (e.g., *Mary said, "This dress is beautiful!"*) and indirect speech (e.g., *Mary said that the dress was beautiful*). Converging findings from speech analysis, brain imaging, and eye tracking indicate that readers spontaneously engage in mental simulations of audible-speech like representations during silent reading of direct speech, and to a much lesser extent during silent reading of indirect speech. This "simulated" implicit prosody is highly correlated with the overt prosody generated during actual speaking. We then compare this "simulated" implicit prosody with the sort of "default" implicit prosody that is commonly discussed in relation to syntactic ambiguity resolution. We hope our discussion will motivate new interdisciplinary research into prosodic processing during reading which could potentially unify the two phenomena within a single theoretical framework.

Keywords Implicit prosody · Inner voice experience · Direct speech · Indirect speech · Reading · Mental simulation · Embodied cognition · fMRI · Eye tracking

1 Overview

In this chapter, we review a new body of research on language processing, focussing particularly on the distinction between direct speech (e.g., *Mary said, "This dress is absolutely beautiful!"*) and indirect speech (e.g., *Mary said that the dress was absolutely beautiful*).

First, we will discuss an important pragmatic distinction between the two reporting styles and highlight the consequences of this distinction for prosodic processing.

B. Yao (✉)
School of Psychological Sciences, University of Manchester, Manchester, UK
e-mail: Bo.Yao@manchester.ac.uk

C. Scheepers
Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, UK
e-mail: Christoph.Scheepers@glasgow.ac.uk

L. Frazier, E. Gibson (eds.), *Explicit and Implicit Prosody in Sentence Processing,*
Studies in Theoretical Psycholinguistics 46, DOI 10.1007/978-3-319-12961-7_15

While direct speech provides vivid *demonstrations* of the reported speech act (informing recipients about *how* something was said by another speaker), indirect speech is more descriptive of *what* was said by the reported speaker. This is clearly reflected in differential prosodic contours for the two reporting styles during speaking: Direct speech is typically delivered with a more variable and expressive prosody, whereas indirect speech tends to be used in combination with a more neutral and less expressive prosody.

Next, we will introduce recent evidence in support of an "inner voice" during language comprehension, especially during silent reading of direct speech quotations. We present and discuss a coherent stream of research using a wide range of methods, including speech analysis, functional magnetic resonance imaging (fMRI), and eye tracking. The findings are discussed in relation to overt (or "explicit") prosodic characteristics that are likely to be observed when direct and indirect speech are used in spoken utterances (such as during oral reading). Indeed, the research we review here makes a convincing case for the hypothesis that recipients spontaneously activate voice-related mental representations during silent reading, and that such an "inner voice" is particularly pronounced when reading direct speech quotations (and much less so for indirect speech). The corresponding brain activation patterns, as well as correlations between silent and oral reading data, furthermore suggest that this "inner voice" during silent reading is related to the suprasegmental and temporal characteristics of actual speech. For ease of comparison, we shall dub this phenomenon of an "inner voice" (particularly during silent reading of direct speech) *simulated implicit prosody* (SIP) to distinguish it from *default implicit prosody* (DIP) that is commonly discussed in relation to syntactic ambiguity resolution.

In the final part of this chapter, we will attempt to specify the relation between SIP and DIP. Based on the existing empirical data and our own theoretical conclusions, we will discuss the similarities and discrepancies between the two not necessarily mutually exclusive terms. We hope that our discussion will motivate a new surge of interdisciplinary research that will not only extend our knowledge of prosodic processes during reading, but could potentially unify the two phenomena in a single theoretical framework.

2 Direct and Indirect Speech: Pragmatic and (Explicit) Prosodic Differences

In everyday language use, prosody carries rich information not only about the structure and pragmatic function of an utterance but also about the source of the utterance (e.g., the speaker and their emotional state). When reporting speech (as in quotations), prosody is a key feature that differentiates direct speech (1) from indirect speech (2).

(1) *Mary said, "This dress is absolutely beautiful!"*
(2) *Mary said that the dress was absolutely beautiful.*

Direct speech is often a literal quotation of what the original speaker said. Indirect speech, by contrast, involves more of a summary or paraphrase of what the original speaker said. The quoted utterance in direct speech is usually treated as an independent prosodic unit and is typically marked with a phonetic pitch reset (i.e., resetting vocal pitch to a higher level in order to continue speaking). In contrast, an indirect speech utterance is usually embedded in a complement clause and not prosodically distinguished from the matrix clause.

While there are semantic and syntactic differences between direct and indirect speech (e.g., Banfield 1973, 1982; Li 1986; Partee 1973; Wierzbicka 1974), linguists have also recognized the "theatrical" nature of direct speech, meaning that it tends to carry more vivid paralinguistic information than indirect speech during communication (Li 1986; Tannen 1986, 1989; Wierzbicka 1974). As first conceptualized by Clark and Gerrig (1990), an important pragmatic function of direct speech is to provide *demonstrations* of the reported speech act. Demonstrations enable others to directly *experience* the things depicted. For example, to demonstrate the action of taking a photograph, one may take an imaginary camera to one's eyes and click the imaginary shutter. Direct speech is often used to demonstrate *how* something was said by another speaker. As Clark and Gerrig (1990) argue, direct speech is an important stylistic device for enlivening stories. It provides vivid demonstrations of the reported speech act, thereby enabling the addressee to experience what it would be like to see, hear, or feel what the original speaker did in saying something. Consider example (1): when the reporter quotes *Mary,* he/she may depict *Mary's* voice (e.g., high-pitch, squeaky), her accent (e.g., southern, northern), her emotional state (e.g., excitement), and/or Mary's supposed facial expressions and gestures while making the utterance, so as to demonstrate *how* Mary said those words. Indirect speech, on the other hand, typically provides a mere *description* of what was said, without depicting paralinguistic information surrounding the reported speech act. In terms of prosody, this pragmatic distinction might become manifest in more dramatized and expressive vocal modulations for direct speech as compared to indirect speech, with the latter being generally reported in a more neutral tone.

Indeed, our own research suggests that in an oral reading task, direct speech tends to be interpreted in a more vivid fashion than indirect speech (Yao 2011, experiment 3). In this exploratory study, we examined whether individuals would spontaneously adjust their voices to "act out" the contextually implied emotional state of the reported speaker when reading aloud *direct speech* or meaning-equivalent *indirect speech* text passages. It is well established that a speaker's emotional arousal is reliably reflected in modulations of vocal pitch (fundamental frequency, F_0) during speaking (Banse and Scherer 1996). If direct speech reporting is associated with demonstrations of the reported speech act, it should display a pitch profile that represents the reported speaker's emotional state. In contrast, indirect speech reporting is likely to be characterised by a pitch profile that is emotionally detached from the original source. To test this idea, we prepared short fictitious stories containing direct or indirect speech utterances. Critically, between-items we manipulated the emotional arousal level of the reported speaker (the main protagonist in the

story) by using introductory contexts implying "high", "medium", or "low" arousal of the quoted speaker (see below for examples; the different arousal levels were verified in a separate rating study).

Examples from Yao (2011, experiment 3):

(3) [HIGH AROUSAL] *Millionaire Joseph was addicted to betting on horses. Tipped by a so-called 'insider', he recently placed an enormous bet, but shockingly, the horse had lost.*

[DIRECT SPEECH] *Angry with his informant, Joseph shouted furiously on the phone: "Where did your bloody information come from!? That was a huge amount of money—almost one million pounds!"*

[INDIRECT SPEECH] Angry with his informant, Joseph shouted furiously on the phone, asking where the information had come from, because that was a huge amount of money—almost one million pounds.

(4) [MEDIUM AROUSAL] *Britney is a student at the University of Glasgow. After a heavy snow in the afternoon, she was complaining to her boyfriend James about the weather on their way home.*

[DIRECT SPEECH] *Her voice sounded very grumpy and unpleasant: "I really hate the winter! It's always dark and the roads are too slippery."*

[INDIRECT SPEECH] *Her voice sounded very grumpy and unpleasant, saying that she really hated the winter because it's always dark and the roads are too slippery.*

(5) [LOW AROUSAL] *Smith was working in a small antiques shop down the local high street. Today, a middle-aged posh lady with thick glasses came into the shop.*

[DIRECT SPEECH] *She looked around and said in a nonchalant tone: "You may be surprised to learn that I'm a world-renowned collector of rare memorabilia of White-eared Pheasant."*

[INDIRECT SPEECH] *She looked around and said, in a nonchalant tone, that he might be surprised to learn that she was a world renowned collector of rare memorabilia of White-eared Pheasant.*

Participants were instructed to read these stories aloud as naturally and fluently as possible. Each participant read each story only once, and importantly, the instructions did not explicitly encourage participants to vocally "act out" the stories. We recorded and analysed pitch contours and other characteristics of participants' speech during reading. Overall, we observed significantly larger variation of F_0 during oral reading of direct speech as opposed to indirect speech (mean SD for F_0 over time:12.63 [direct speech] vs. 8.68 [indirect speech], paired-sample $ts > 11$, $ps < 0.001$). In line with their hypothesised *demonstration* pragmatics, direct speech quotations appeared to have been orally interpreted in more varied, fluctuating pitch profiles than indirect speech utterances. More importantly, when reading direct speech aloud, readers' mean F_0 increased as a function of the contextually implied emotional arousal of the quoted speaker, with more arousal leading to a steady increase in F_0. In contrast, no such linear trend was observed during oral reading of indirect speech (Fig. 1). The data confirmed that readers spontaneously adjust their voices in accordance with the contextually implied emotional arousal of the quoted speaker. This was the case particularly for oral reading of direct speech, but not (or considerably less so) for oral reading of indirect speech. These findings highlight the distinctive prosodic profiles of direct and indirect speech in speaking.

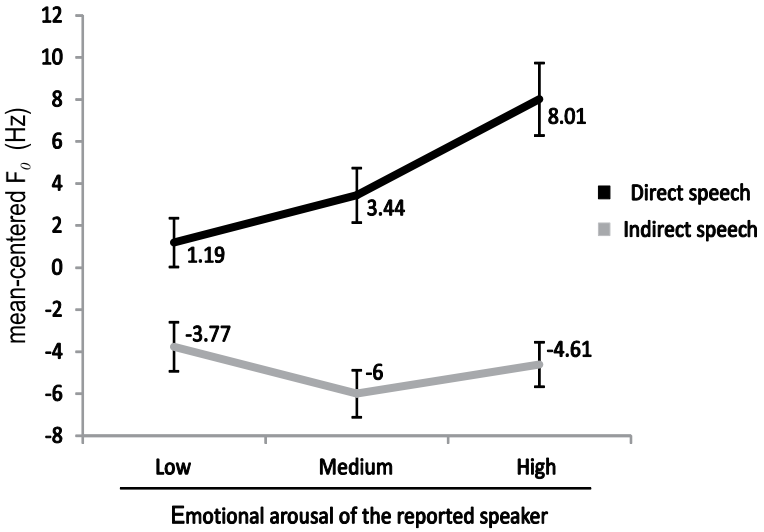


Fig. 1 Reporting style × emotional arousal interaction in oral reading (Yao 2011, experiment 3). The *numbers* indicate the condition means (in mean-centred F_0 to remove systematic gender differences in pitch). The error bars represent the 95% confidence intervals for the means per condition

3 Direct and Indirect Speech During Silent Reading and "Prosodically Impoverished" Listening

Prosodic features of direct and indirect speech are easily measurable in *spoken* language. Here we are going to review evidence suggesting that perceivers also differentiate between the two reporting styles during *written* language processing. To illustrate this, Yao et al. (2011) explored how direct and indirect speech utterances are processed in the brain during silent reading of text where no auditory stimulation is present. Inspired by the common intuition of hearing an "inner voice" during silent reading of quotations, they speculated that the brain might take direct speech as a cue to activate "audible-speech"-like mental representations, even during silent reading of text. Recent embodied cognition theories (e.g., Barsalou 1999, 2008) lend theoretical support to this conjecture. Such theories propose that language processing is grounded in mental simulations (or re-enactments) of sensory and motor experiences that have been acquired through individuals' interaction with the environment and their internal states. Under such a premise, accumulated experiences with how direct versus indirect speech are typically reported in spoken language could form the basis for differential mental simulations during written-language processing. In other words, silent reading of direct speech would be grounded in mental simulations of vivid vocal depictions whereas silent reading of indirect speech would be grounded in simulations of voices that are more neutral. The brain may therefore be more prone to activate "audible-speech"-like representations during silent reading of direct speech than of indirect speech.

To test this hypothesis, Yao and colleagues combined fMRI and eye tracking to measure neural activity within the auditory cortex. The fMRI technique captures changes in oxygen consumption in local blood flow, which in turn estimates the degrees of neural activity within certain brain areas in vivo (Ogawa et al. 1990a, b; Ogawa and Lee 1990). Using this technique, neuroscientists have established that certain areas in the auditory cortex, i.e., those along the upper bank of the superior temporal sulcus (STS), are selectively sensitive to "bottom-up" auditory stimulation by human voices (Belin et al. 2000). These areas, labelled temporal voice areas (TVAs), provided Yao et al. (2011) with clearly defined functional hot spots for locating activations of voice-related representations during silent reading. In their experiment, participants' individual TVAs were identified in a voice localizer task in which audio clips of nonvocal sounds (e.g., telephone ringing) were compared to vocal sounds generated from speech (e.g., vowels) and nonspeech (e.g., laughing) utterances (Belin et al. 2000). Participants' TVAs were thus localizable via the contrast of their brain responses to vocal sounds versus nonvocal sounds. Before this functional voice-localizer task, Yao and colleagues measured neural activity (in the same participants) during silent reading of direct versus indirect speech text passages, as shown in the following example:

Examples from Yao et al. 2011:

(6) *PhD student Ella was summoned to her supervisor Jim's office to give a report on her current progress. Ella asked for an extension but Jim looked concerned.*
[DIRECT SPEECH] He said: "Hmm, we really need those data in by next month for that conference."
[INDIRECT SPEECH] *He said that they really needed those data in by next month for that conference.*

Importantly, the reported speech utterances in both conditions were kept equivalent in terms of linguistic content within each story (see underscored sentences in the above example); this was to rule out potential confounding factors between conditions. In the magnetic resonance imaging (MRI) scanner, these stories were visually presented to participants in a sentence-by-sentence fashion and for a fixed duration. Participants were instructed to silently read these stories for comprehension while their eye movements and brain activity were simultaneously monitored. Yao and colleagues observed that during silent reading of the critical speech utterances (determined via eye tracking), direct speech was associated with greater neural activity across multiple brain areas than indirect speech. The enhanced activity was distributed not only in the right auditory cortex but also in bilateral occipital lobes (associated with visual processing), superior parietal lobules, and precuneus (associated with visuo-spatial imagery, episodic memory retrieval, and self-processing). Such an activation pattern seemed to suggest an enriched multisensory mental simulation process for direct speech, which is consistent with Clark and Gerrig's (1990) hypothesis of direct speech as demonstration. Critically, reading of direct speech quotations (compared to meaning-equivalent indirect speech utterances) elicited significantly higher neural activity along the right STS (rSTS) areas which were clearly part of the TVAs identified in the voice-localizer task. This was the first direct indication that silent reading of direct speech is more strongly associated with

"top-down" simulations of voice-related sensory experiences. Interestingly, compared to a baseline without linguistic stimulation, even indirect speech elicited some activation in those TVAs, but to a considerably lesser extent than direct speech.

Similar kinds of "inner voice" experiences were also observed during silent reading of direct speech in German (Brück et al. 2014). The authors' primary aim in that study was to investigate the neural correlates in processing emotional voice signals described in written texts (e.g., *Als sie sprach, klang ihre Stimme sanft und kehlig und mit einem italienischen Akzent behaftet—When she spoke, her voice sounded smooth and throaty and beset with an Italian accent*). Although not central to their research question, they also explored how direct speech reporting might modulate TVA activation during silent reading. This was possible because one third of their stimuli actually comprised direct speech quotations (e.g., *"Das ist nicht zu ertragen", sprach die Fürstin leise mit zitterender Stimme—"This is unbearable", said the baroness quietly with a quivering voice*). As expected, Brück et al. (2014) observed significantly higher activations of the right TVAs during silent reading of direct speech quotations as opposed to the other types of descriptions without quotations. Although this finding was established "post-hoc", it largely agrees with Yao et al.'s (2011) results, confirming that direct speech is likely to activate speech-(or voice-)related sensory experiences "top down", i.e., without acoustic stimulation.

One objection might be that the direct versus indirect speech materials used in Yao et al. (2011) sometimes differed in grammatical tense (present vs. past), syntactic structure (coordination vs. subordination), the use of pronouns (e.g., first vs. third person), or the use of emotion-signalling punctuation ("!" vs. "."). It is therefore conceivable that the observed differences between direct and indirect speech may be evoked by these extraneous differences, rather than the reporting styles "*per se*". However, Yao et al.'s (2011) additional reading performance analyses revealed no clear differences in either reading time (204 vs. 203 ms/word) or comprehension accuracy (83 vs. 82 %) between the direct and indirect speech conditions. More importantly, Yao et al. (2011) could show that the critical fMRI effect did not disappear when only a subset of items (34 out of 90) was considered, in which the direct and indirect speech conditions could be regarded as equivalent in terms of grammar and punctuation. With respect to the locus of the fMRI effect, the right-lateralized STS activation pattern hardly overlaps with activation patterns observed during processing of present versus past (D'Argembeau et al. 2008), syntax (e.g., Friederici et al.2000a, b), perspective (Vogeley and Fink 2003), or modality-independent emotions (Peelen et al. 2010). Taken together, it appears that an enhanced "inner voice" sensory experience during silent reading of direct speech remains the best explanation of Yao et al.'s (2011) data.

But how does this sensory experience relate to prosody? In fact, the *prosodic* nature of such "inner voices" was illuminated in a follow-up fMRI study by Yao et al. (2012). In Yao et al.'s (2011) study, there was no acoustic stimulation as a reference alongside the silent reading task (except for the functional localizer procedure). It was hence difficult to specify what types of acoustic representations may constitute the "inner voice" experiences during silent reading of direct speech. Interestingly, however, the acoustic processing literature indicates that the right auditory cortex

areas appear to be specialised in processing slow-pitch modulations, including speech melody (Scott et al. 2000), musical melody (Patterson et al. 2002; Zatorre et al. 1994, 2002), and emotional prosody (Mitchell et al. 2003; Wildgruber et al. 2005). Thus, the specifically right-lateralized activation pattern observed in Yao et al. (2011) might be taken to suggest a suprasegmental prosodic nature of the "inner voices" experiences in silent reading of direct speech.

To verify this conjecture, Yao et al. (2012) sought to examine the neural correlates of "top-down" suprasegmental prosodic processing during *auditory* comprehension of reported speech. If these neural correlates show substantial overlap with the differential brain activation regions found in silent reading (Brück et al. 2014; Yao et al. 2011), this would lend support to the hypothesis that the latter may be of a suprasegmental prosodic nature. To this end, Yao and colleagues prepared audio recordings of the same short stories as in Yao et al. (2011). Crucially, both the direct and indirect speech utterances in these recordings were deliberately spoken in a *monotone* which is usually more felicitous for indirect rather than direct speech. The following is an example story:

(7) *Luke and his friends were watching a movie at the cinema. Luke wasn't particularly keen on romantic comedies, and he was complaining a lot after the film.*
[DIRECT SPEECH] He said: "God, that movie was terrible! I've never been so bored in my life."
[INDIRECT SPEECH] *He said that the movie was terrible and that he had never been so bored in his life.*

This example story describes *Luke*'s terrible experience with a boring film. Normally, one would expect Luke to sound rather impatient and moany (e.g., "GOD, that movie was t-EEE-rible!"[1]), depicting how much *Luke* regretted watching the film. In stark contrast, the direct speech quotation was actually spoken in a steady tone which sounded emotionally detached (perhaps even sarcastic), and did not fit into the overall context (recordings can be found at: http://www.psy.gla.ac.uk/~boy/fMRI/samplerecordings/). Acoustically, this *monotone* manipulation preserved (sub)-segmental acoustic information (e.g., the phonological representations of words) but severely curtailed rich suprasegmental prosodic information (e.g., varied intonation patterns) that is typically expected of direct speech quotations. Yao et al. (2012) hypothesized that the brain may actively compensate for monotonously spoken direct speech by "filling in" suprasegmental prosodic information (i.e., expressive prosody) that is missing from the actual input. Such "filling in" processes should be reflected in increased brain activity within the TVAs. Comprehension of monotonous indirect speech utterances, however, is unlikely to involve such processes. Unlike its direct speech counterpart, indirect speech is typically spoken in a more neutral, less varied prosody (e.g., Yao 2011, described earlier). Thus, the brain does not need to compensate for monotonously spoken indirect speech utterances.

Using fMRI, Yao et al. (2012) measured participants' brain activity when they were listening to the monotonously spoken stories illustrated above. The

[1] The capitalization and repetition of letters represent emphases in intensity and length.

participants' individual TVAs were determined using the same voice localizer task as before (Belin et al. 2000). Neural activity within the TVAs was determined while listening to the critical direct speech or indirect speech utterances (underscored sentences in the above example). As expected, it was found that monotonously spoken direct speech elicited significantly higher brain activations within the right TVAs than monotonously spoken indirect speech. Most intriguingly, the increased activations for direct speech were located in virtually the same brain areas (i.e., the posterior, middle, and anterior parts of the right STS) as those previously observed in silent reading of direct versus indirect speech (see Fig. 2).

However, it remained unclear whether these differential brain activations indeed reflected enhanced "top-down" prosodic processing when listening to monotonous direct speech, or whether they were merely evoked "bottom-up" by differential acoustic characteristics of direct versus indirect speech utterances. To address this question, three variables (or "parametric modulators") were specified to potentially account for these increased rSTS activations. These were (a) the *acoustics* of the recordings (i.e., parameters such as pitch, intensity, duration, etc.), (b) the subjectively perceived *vividness* of the speech utterances without

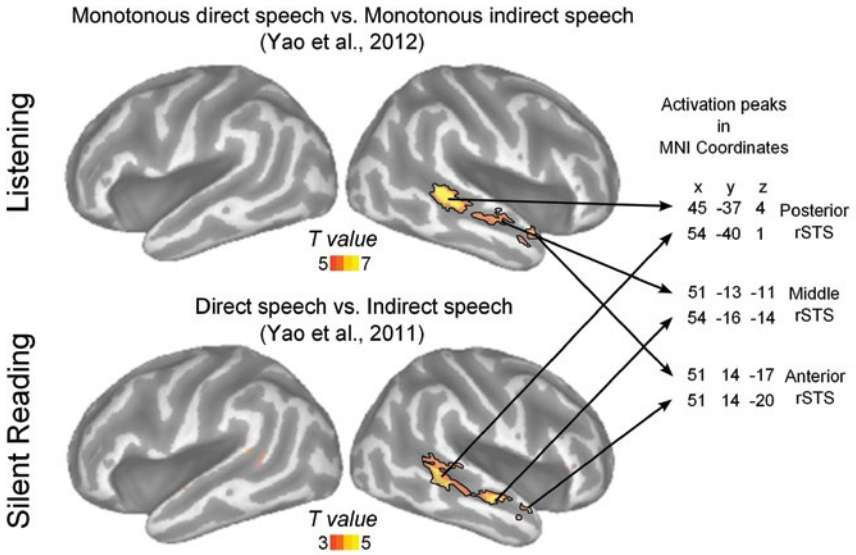


Fig. 2 Consistent findings between the two fMRI studies (only the effects within the TVAs are shown). The *top* panel shows the contrast between the monotonous direct speech and the monotonous indirect speech conditions during listening (Yao et al. 2012). The *bottom* panel shows the contrast between direct speech and indirect speech during silent reading (Yao et al. 2011). The *arrows* point to the peak voxel coordinates (in MNI space) in the activation clusters. The peak voxels were paired with their anatomical counterparts between the two studies. The thresholds for the two contrasts were adjusted to better illustrate the activation clusters

context (established via ratings), and (c) the *contextual congruency* of the speech utterances within the given story contexts (i.e., to what extent the speech utterances were perceived as congruent with a given context or not—again, this variable was established via ratings). *Acoustics* (a) and *vividness* (b) were taken as *objectives,* respectively *subjective* measures of the acoustic differences between the direct and indirect speech conditions without considering the context that the quotations were embedded in. Both modulators were expected to account for differences in "bottom-up" acoustic processing. In contrast, *contextual congruency* (c) was taken to index the degree of mismatch between the actual (monotonous) speech input and perceivers' "top-down" *expectation* for expressive speech prosody in the given story context. *Contextual congruency* was expected to explain differential "top-down" prosodic processing between conditions. It was found that the increased rSTS activations in monotonous direct speech (relative to monotonous indirect speech) was in fact most reliably explained by *contextual congruency* but not by the *acoustics* or *vividness* of the speech utterances. The analyses confirmed that the rSTS activation pattern indeed reflected "top-down" prosodic processing when listening to monotonously spoken direct speech utterances, as if the brain was actively trying to "fill in" prosodic information that was missing from the actual speech input. By reconciling the findings of the two fMRI studies, we conjecture that the "inner voices" observed during silent reading of direct speech may also involve suprasegmental prosodic information similar to that in auditory processing.

The prosodic nature of such an "inner voice" during silent reading of direct speech was further demonstrated behaviourally by Yao and Scheepers (2011). They examined whether the speech-related representations activated during silent reading of direct speech could be characterized in time (or speed). Time is an important dimension of prosody. It determines the rhythm, stresses (e.g., length of articulation), and global dynamics of speech. If prosodic representations were activated during silent reading of direct speech, they should reflect the implied speaking rate of the quoted speech. A potential behavioural consequence of this is that readers may adjust their reading rates in accordance with how fast a quoted speaker would speak in a given context.

Previous research by Alexander and Nygaard (2008) has already suggested that reading speed may be influenced by auditory imagery. In their study, they first familiarized participants with audio recordings of voices from either fast or slow speakers. In subsequent reading sessions, they told participants to imagine those speakers as authors of the materials given for reading. They observed that during both oral and silent reading, participants were faster to read the presented text materials when they were told that the author of the text was a previously introduced "fast speaker". Alexander and Nygaard's findings demonstrated that explicitly *encouraged* auditory imagery of an author's speaking rate had an influence on how fast one would read written text that was supposedly produced by that author. This, in turn, suggests that "inner voices" during silent reading of direct speech—a more *spontaneous* form of auditory imagery—could equally interact with reading behavior.

To test this idea, Yao and Scheepers (2011) prepared short stories containing direct and indirect speech utterances (see below for an example). Each story started with a narrative vignette which set up either a fast-speaking (i.e., where the speaker was likely to speak very quickly) or a slow-speaking scenario. The scenario led to a reported speech utterance that employed either direct speech or indirect speech. The story was then concluded by an additional sentence. Crucially, the critical speech sentences (e.g., the underscored sentences in the example) were identical between the fast-speaking and slow-speaking stories and were largely equivalent between direct speech and indirect speech conditions. Thus, differences in reading rate could not plausibly be attributed to differential wording across conditions.

(8) [FAST-SPEAKING] *It was a typical British day, rainy and gloomy. Sixteen-year-old pianist Bobby was going to play in the quarter-finals of a local talent competition. He was extremely nervous before his performance.*

[DIRECT SPEECH] *His mother encouraged him but he was all shaking and said: "No! I can't do it! This is the end of the journey because it is unlikely that I will make it this time."*

[INDIRECT SPEECH] *His mother encouraged him but he was all shaking and said that he couldn't do it and that it was the end of the journey because it was unlikely that he would make it this time.*

His mother tried to calm him down, saying that it's not the winning that counts, but the taking part.

(9) [SLOW-SPEAKING] *It was a typical British day, rainy and gloomy. At Glasgow Royal Infirmary, an old man was dying, and too weak to sit up. His family members were sitting around the bed, feeling sad. He wanted to say something, so his daughter placed a cushion under his head.*

[DIRECT SPEECH] *Slowly, he looked around and said: "I'm grateful you're all here. This is the end of the journey because it is unlikely that I will make it this time."*

[INDIRECT SPEECH] *Slowly, he looked around and said that he was grateful for their coming and that it was the end of the journey because it was unlikely that he would make it this time.*

Then he closed his eyes and everyone burst into tears.

Yao and Scheepers (2011) tested these materials in both oral and silent reading. In the oral reading task, participants were instructed to read aloud the stories in one go and as naturally and fluently as possible. Importantly, participants were not explicitly told to act out the reported speaker's voice during reading. Oral reading rates during the critical quotation passages were measured in syllables per second. A different group of participants were given the stories for silent reading while their eye movements were continuously monitored. Participants in the silent reading task were told to read the stories carefully for comprehension, and their reading rates were indexed by go-pass reading times (in milliseconds) on the critical direct or indirect speech sentences. In line with the predictions, it was found that in both oral and silent reading, participants spontaneously adjusted their reading rates to the contextually implied speech rate of the quoted speaker, but only when reading direct speech quotations and not when reading indirect speech passages. Most interestingly, Yao and Scheepers (2011) observed a high by-item correlation ($r = 0.56$, after accounting for effects of stimulus length) of reading rates across the two reading

tasks. This suggests a strong temporal relation between "explicit prosody" (oral reading) and "implicit prosody" (silent reading) for the processing of both direct and indirect speech utterances.

In a more recent eye-tracking study, Stites et al. (2013) showed very similar effects during silent reading of direct speech, but again, not during silent reading of indirect speech. Interestingly, they found that these effects can be triggered by a single adverb (e.g., *John walked into the room and said* "energetically" *vs.* "nonchalantly"…) before the critical quotation passages. That is, direct quotations that were described as being said "quickly" were read faster than those described as being said "slowly".

In summary, the research on direct versus indirect speech has provided neuroimaging and behavioural evidence of "top-down" prosodic processes during language comprehension, in particular during silent reading of direct speech quotations. For the prosodic representations that are mentally simulated during silent reading of direct speech, we will use the term SIP to distinguish it from DIP that we shall discuss later. SIP appears to be primarily processed along the rSTS areas of the auditory cortex which are part of the TVAs (Belin et al. 2000). One important aspect of SIP is reflected in the close relationship between modulations of speaking rate (oral reading) and modulations of reading rate (silent reading) on the same language materials. In a broader context, these findings support the demonstration theory of direct speech (Clark and Gerrig 1990) from the perspective of language comprehension, highlighting the fact that direct speech is intrinsically more expressive than its indirect speech counterpart. The findings also extend embodied theories of language comprehension in several respects. First, the reviewed evidence for implicit prosody during silent reading (presumably in the form of mental simulations of actual speech, or at least involving speech-related mental representations) extends embodied theories to the auditory perceptual domain at the sentence/discourse level, which so far has received limited attention in the literature (previous research has mostly focused on sound-related words, see Kiefer et al. 2008 for example). Second, while most empirical research on embodied language comprehension focuses on the grounding of the linguistic meaning in perception and action, the research reviewed here involves differences in language pragmatics (direct speech as demonstration; indirect speech as description) and the consequences of such differences for processing semantically comparable reporting styles. In verbal communication, direct speech usually coincides with vivid demonstrations of the reported speech act whereas indirect speech is reported in a less vivid fashion. The present research shows that this vividness distinction is also reflected in how language is processed, and that direct speech is more likely to evoke mental simulations of voices or voice-related representations than indirect speech. Third, the reviewed fMRI research revealed that the posterior, middle, and anterior parts of the rSTS are potentially involved in mental simulations of suprasegmental prosodic representations. These data would motivate more sophisticated research on the neural mechanisms of implicit prosody and the neural configurations of the TVAs of the auditory cortex in general.

4 Open Questions and the Relation Between "Simulated" and "Default" Implicit Prosody

Many questions remain as to the detailed nature, mechanisms and functions of SIP. One interesting avenue for future research might be to probe its characteristics in other dimensions such as pitch, accent, and speaker identity.

Other questions relate to the durability of SIP representations. The studies above have mostly employed *online* methods (such as eye-tracking and fMRI) that probed into the ongoing processing of reported speech. In contrast, studies using *offline* methods such as probe-reaction *after* reading of quotations, appeared to be less sensitive in detecting differences between direct and indirect speech processing (Eerland et al. 2013; Yao 2011, Chap. 4). Given the temporal correlation between SIP and explicit prosody (Yao and Scheepers 2011), one might infer that effects related to SIP are relatively short-lived. More sophisticated testing is therefore needed to characterise the temporal properties of SIP in greater precision, specifying its onset, saturation, and offset.

Further questions for future research concern the function of SIP during silent reading of quotations. For example, is SIP beneficial to reading and memory? Given that aspects of SIP were shown to influence reading speed (Stites et al. 2013; Yao and Scheepers 2011), it appears worthwhile to further explore its role in eye movement control during silent reading.

While the research on direct and indirect speech is interesting in its own right, one interesting question arises as to how SIP during silent reading (particularly of direct speech) would inform the *implicit prosody hypothesis* (IPH) for silent reading (e.g., Fodor 1998, 2002; Quinn et al. 2000). The IPH assumes that a DIP is projected during silent reading of text, with potential consequences for syntactic processing. Such a default prosodic contour is very similar to the usual "explicit" prosodic contour for actual speech: It implements pauses, emphases, etc., thereby suggesting a prosodic grouping of a sentence during silent reading. These "implicit" prosodic groups appear to influence the syntactic parsing of a sentence, and may even determine its preferred interpretation in the face of syntactic ambiguity. Evidence for DIP processing during silent reading is provided, for example, by a rich body of research on relative clause (RC) attachment. Consider the English sentence "*Someone shot the servant of the actress who was on the balcony*", which is ambiguous as to whether the RC "*who was on the balcony*" should be attached *high* to the complex noun phrase "*the servant of the actress*" or low to the simpler and more recent noun phrase "*the actress*". It has been shown that native speakers of English tend to prefer a low attachment interpretation when silently reading a sentences such as the one quoted above (e.g., Carreiras and Clifton 1993, 1999). By contrast, speakers of other languages such as Spanish (Carreiras and Clifton 1993, 1999), French (Zagar et al. 1997), and German (e.g., Hemforth et al. 1998) prefer a high attachment interpretation for equivalent structures. The IPH provides a promising explanation for such RC attachment biases in different languages. When no other disambiguation cues (e.g., gender agreement, case marking, or semantic constraints) are available,

DIP contours (which may differ across languages) provide structural information that aids syntactic ambiguity resolution. This claim has been *indirectly* supported by research on the effect of explicit prosody on RC attachment disambiguation. For example, Quinn et al. (2000) asked participants to read and interpret ambiguous RC sentences silently and then read the sentences again aloud. They analyzed the F_0 (fundamental frequency) values of N1 and N2 in sentences disambiguated for high/low attachment. They found that pitch accents (i.e., peaks in F_0) on the critical noun phrases (NPs) were related to preferred RC attachment. That is, in an NP1–NP2–RC structure, pitch accents on NP1 were more strongly associated with high attachment, whereas pitch accents on NP2 were more strongly associated with low attachment of the RC. They suggested that in silent reading, RC attachment may be disambiguated by the prominence relations of the NPs and RC that are marked by the purported implicit default prosody. Other prosodic factors such as prosodic breaks or pauses have also been found to influence RC attachment interpretations in speech. It has been established that a prosodic break before an RC generally prompts high attachment of the RC (e.g., Clifton et al. 2002; Lovrić et al. 2000, 2001; Maynell 1999). In a silent reading study, Lovrić et al. (2001) manipulated the duration of NP1 and NP2 in order to trigger implicit prosodic breaks at different locations of an NP1–NP2–RC structure. They found that the lengthening of NP1 (prompting a prosodic break before NP2) resulted in a low attachment preference; the lengthening of NP2 before a long RC (prompting a prosodic break between NP2 and RC) increased probability of high attachment interpretations. Such correlations between DIP breaks and RC attachment preferences also lend support to the IPH.

Although DIP has been established behaviourally in different languages (e.g., Koizumi 2009; Shafran 2011; Shaked 2009), the cognitive and neural mechanisms underlying the projection of DIP remain largely unknown. By its very nature, DIP is not easy to manipulate or to measure, and it has yet to offer a comprehensive explanation for crosslinguistic variation in RC attachment. We believe that theories such as the IPH could potentially benefit from systematic analyses of what we called SIP during silent reading of direct (vs. indirect) speech.

In the following, we will discuss potential relations between DIP (as primarily revealed in research on ambiguity resolution) and SIP (as discussed in the context of reported speech processing). One possibility is that DIP and SIP are two instantiations of the same cognitive process, involving largely the same mental representations. This seems plausible because both refer to prosodic representations that are generated "internally", i.e., without external auditory stimulation. Research has shown that (at least aspects of) DIP and SIP are correlated with explicit prosody during actual speech (e.g., Lovrić et al. 2000, 2001; Yao and Scheepers 2011). This might indicate that DIP and SIP share the same sensory grounding. Moreover, it is evident that the SIP activated (particularly) during direct speech processing may be an enhanced form of DIP which is activated during indirect speech processing and/or the processing of materials that do not involve reported speech. In fact, Yao et al.'s (2011) fMRI study on silent reading of direct versus indirect speech indicated that *both* direct *and* indirect speech processing lead to increased rSTS activation compared to a baseline condition where only a fixation cross was presented (no

reading). This additional observation suggests that even silent reading of indirect speech may not be completely "silent" in that it also involves some form of implicit prosodic processing, although to a much lesser extent when compared to silent reading of direct speech. It therefore appears plausible to speculate that SIP during silent reading of direct speech may be a special, enriched form of the more generic prosody (DIP) assumed by the IPH. One way to test the relations between DIP and SIP might be to embed ambiguous RC structures in direct speech quotations, and examine whether RC attachment preferences during silent reading are in some way "*enhanced*" compared to RC attachment in isolated sentences or sentences introduced as indirect quotes. For example, one could test whether the NP1-NP2-RC structure in *When asked by the police, she said, "Someone shot the servant of the actress who was on the balcony"* would result in a stronger low attachment preference in English than when it is not in direct quotes. If we observed such interaction between RC disambiguation and reporting style, it would add weight to the hypothesis that DIP and SIP share aspects of the same mental representation.

In addition, DIP and SIP both interact with language processing and it seems that a common function of them is to facilitate comprehension. It is well established that DIP can help resolve syntactic ambiguity during silent reading by providing prosodic cues to the configurational interpretation of linguistic structure when other cues (e.g., syntactic or semantic) are not available (Fodor 2002). However, RC attachment is by no means the only processing domain where implicit prosody becomes relevant. For example, a recent eye-tracking study by Ashby and Clifton (2005) examined the effects of lexical stress on eye movements during silent reading. Participants read sentences containing words with a single stressed syllable or words with two stressed syllables. With other factors controlled, it was found that two-syllable words took longer to read compared to one-syllable words. The findings are in line with the IPH, suggesting that a prosodic contour is routinely constructed during silent reading, affecting not only sentence-level processing but also lexical access.

In a similar vein, SIP during silent reading of direct speech also appears to be beneficial to language processing. The notion of SIP essentially refers to the addition, or enhancement, of another sensory (i.e., auditory) layer during silent reading, which is particularly noticeable in direct speech processing. This layer enriches the mental representations of direct speech in many respects, including the emotional states of the quoted speakers, speech pragmatics, speech styles, and so on. For example, consider the following two sentences:

(10) *Mary said with excitement, "This dress is absolutely beautiful!"*
(*This dress is ABSOLUTELY BEAUUU-tiful*)

(11) *Mary said with excitement that the dress was absolutely beautiful.*
(*The dress was absolutely beautiful*)

The sentences in parentheses illustrate how the speech utterances in (10) and (11) may be interpreted prosodically during silent reading. The capital letters in (10) represent a hypothetical increase in pitch and volume (accents), and the repetition of the letter *U* represents the lengthening of the vowel/ju:/ in *beautiful*. Semantically, both sentences describe that *Mary* found a dress very beautiful. In (10), however, the more "dramatic" prosodic contour adds a sensory layer that allows

the brain to *perceptually experience* the excitement in speech. This additional sensory information creates an enriched representation of the emotional state of the quoted speaker, causing (10) to be more accessible and engaging. In contrast, although (11) characterizes the emotionality of the speaker semantically, the more generic, default prosodic contour in (11) does not reinforce this representation. As a result, (11) is likely to be perceived as being more distant and emotionally disconnected.

The perceptually enriched representation of direct speech might explain why direct speech appears to be associated with deeper processing than indirect speech (Bohan et al. 2008; Eerland et al. 2013). Bohan et al. (2008), for example, visually presented participants with a direct or an indirect speech sentence like the following:

(12) *John said, "I needed some nine-inch nails so I went to B&Q".*
(13) *John said he needed some nine-inch nails so he went to B&Q.*

Immediately after the initial presentation, they showed the same sentence again, and asked participants whether or not this sentence was different from the one that had just been shown. In half of the trials, the second sentence was indeed exactly the same as the first sentence. In the other half of the trials, however, the second sentence presentation involved a very subtle text change within the critical quotation passage (e.g., replacing the verb "*went*" with a close semantic relative such as "*walked*"). Bohan et al. (2008) found that such subtle verb exchanges were reliably more detectable when they occurred within a direct speech rather than an indirect speech text passage, suggesting deeper processing (or enhanced verbatim memory) of direct speech. Eerland et al. (2013) later extended these findings to cases where the text changes were not restricted to verbs. Both studies consistently showed a memory advantage for direct speech as compared to indirect speech. These findings support the idea that covert prosody enhances the representations of direct speech. However, the link between such a memory advantage and SIP is yet to be established.

While DIP and SIP appear to be highly comparable from a phenomenological and functional perspective, it is equally conceivable that they actually entail two distinctive cognitive processes. In fact, a rather complex picture emerges as to the potential mechanisms underlying DIP and SIP. By definition, DIP is routinely generated and projected during silent reading. It can be viewed as a regular prosodic channel which informs the configurational interpretation of language when disambiguating cues from other channels (e.g., syntax, semantics) are not available. DIP has been shown to be informed by a default prosodic contour (i.e., phonology) of a given language, as well as surface visual features such as punctuation (e.g., Steinhauer and Friederici 2001; Steinhauer 2003), phrase length (e.g., Lovrić et al. 2001), or line breaks (e.g., Koizumi 2009). In contrast, SIP appears to be highly dependent on linguistic context and pragmatics (Stites et al. 2013; Yao et al. 2012; Yao and Scheepers 2011), and operates at a deeper, semantic level in a "predictive" manner. In line with embodied theories (Barsalou 1999, 2008), SIP is the speech experience that is *mentally simulated* during comprehension of (particularly) direct speech, as part of a more vivid mental representation of the latter. Mental simula-

tions not only re-enact sensory, motor, and introspective experiences for representing language that is currently being processed; more importantly, they also place the perceiver in the simulated situations, thereby producing continual predictions about events likely to be described, actions likely to take place and introspections likely to result in the incoming language stimuli (Barsalou 2009). As evidence for the predictive aspect of SIP, the findings by Yao et al. (2012) showed that when direct speech quotations are spoken in a context-inappropriate monotone, the perceiver's brain automatically "talks over" such boring quotes by actively projecting context-appropriate prosodic structure that is missing from the input. It appears that during listening, SIP can serve as a top-down predictor of actual speech.

The similarities and differences between DIP and SIP may be reconciled in partially overlapping processing models for the two phenomena. Considering their comparable correlations with explicit prosody, it seems plausible to conjecture that DIP and SIP share a common neural network for representing prosodic contours. However, their potentially distinctive cognitive origins (projection of default prosodic contours on the one hand vs. perceptual simulation of voice and speech on the other) may be reflected in differential engagement of brain regions within this common network and/or engagement of additional brain regions that modulate this network. Only future research can tell the exact differences and commonalities between DIP (as reflected in research on ambiguity resolution) and SIP (as revealed by differences in processing direct versus indirect speech).

5 Conclusions

In this chapter, we have examined the mental representations of direct speech (e.g., *Mary said, "This dress is absolutely beautiful!"*) versus indirect speech (e.g., *Mary said that the dress was absolutely beautiful*). We showed that the brain is more likely to generate enriched suprasegmental prosodic representations of the reported speaker during comprehension of direct speech as opposed to meaning-equivalent indirect speech. We dubbed this specific "inner voice" phenomenon SIP. We have presented consistent neuroimaging evidence showing that SIP is primarily processed at the posterior, middle, and anterior areas of the rSTS of the auditory cortex—also parts of the TVAs (Belin et al. 2000). One aspect of SIP becomes evident in processing rates for direct speech quotations, as reflected in modulations of explicit speaking rates during oral reading as well as in eye movements during silent reading. The findings provide empirical support for the theory of direct speech as demonstration (Clark and Gerrig 1990) and embodied theories of language comprehension (e.g., Barsalou 1999, 2008).

What are the implications of these findings for the IPH? The IPH proposes that a default prosodic contour is generated internally and projected onto visual texts during silent reading. We have termed this kind of projected information DIP. DIP provides prosodic cues (e.g., emphases, prosodic breaks) that benefit configurational interpretations of ambiguous language structures (e.g., relative clause attachment)

when other types of cues (e.g., syntactic, semantic) are not available. By their nature, DIP and SIP are both internally generated prosodic representations without external auditory stimulation, and are correlated with prosody in actual speech. Moreover, DIP and SIP both appear to be beneficial to language processing, although in their own ways. While DIP aids in structural interpretation, SIP perceptually enriches the mental representation of language, resulting in deeper processing of (or enhanced verbatim memory for) direct speech compared to indirect speech. With respect to the mechanisms of DIP and SIP, we recognize that they may be derived from distinctive cognitive processes. Based on the existing evidence, we conjecture that DIP operates relatively independently at a surface level of linguistic representation, routinely informing structural interpretations of language. In comparison, SIP appears to be a mentally simulated sensation of voice that is highly dependent on semantic and pragmatic context. We attempt to reconcile the similarities and discrepancies between DIP and SIP by conjecturing partially overlapping processing networks for these two phenomena.

Although research on SIP in silent reading of direct speech is still in its infancy, it complements the research on DIP by providing a potential platform to address how implicit prosody may operate at the neural, cognitive, and behavioural level. By investigating the similarities and discrepancies between DIP and SIP, future research has the potential to venture beyond simple demonstrations of these phenomena by seeking the evidence necessary to develop explicit mechanistic models of the two processes. An interdisciplinary approach would be very useful in pursuing this ambition. For example, a combination of eye tracking with fMRI and electroencephalography (EEG) or with magnetoencephalography (MEG) would allow us to delineate the neural circuitry underlying DIP and SIP processing in high spatiotemporal precision during real-time silent reading. This could illuminate where DIP and SIP originate from and whether they indeed converge into overlapping neural circuits, resulting in comparable prosodic sensations. The precise neural dynamics and parameters provided would lay the biological and empirical foundation for cognitive modelling of DIP and SIP, leading to more sophisticated theories in both domains.

References

Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology-Human Perception and Performance, 34*(2), 446–459. doi:10.1037/0096-1523.34.2.446.

Ashby, J., & Clifton, C. (2005). The prosodic property of lexical stress affects eye movements during silent reading. *Cognition, 96*(3), B89–B100. doi:10.1016/j.cognition.2004.12.006.

Banfield, A. (1973). Narrative style and grammar of direct and indirect speech. *Foundations of Language, 81*(4), 1–39.

Banfield, A. (1982). *Unspeakable sentences: Narration and representation in the language of fiction*. Boston: Routledge.

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614–636. doi:10.1037/0022-3514.70.3.614.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*(4), 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639.

Barsalou, L. W. (2009). Simulation, situated conceptualization, and prediction. *Philosophical Transactions of the Royal Society B-Biological Sciences, 364*(1521), 1281–1289. doi:10.1098/rstb.2008.0319.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309–312. doi:10.1038/35002078.

Bohan, J., Sanford, A. J., Cochrane, S., & Sanford, A. J. S. (2008). *Direct and indirect speech modulates depth of processing*. Poster presented at the 14th Annual conference on architectures and mechanisms for language processing (AMLaP), Cambridge, UK.

Brück, C., Kreifelts, B., Gößling-Arnold, C., Wertheimer, J., & Wildgruber, D. (2014). Inner voices: The cerebral representation of emotional voice cues described in literary texts. *Social Cognitive and Affective Neuroscience*. doi:10.1093/scan/nst180.

Carreiras, M., & Clifton, C. (1993). Relative clause interpretation preferences in Spanish and English. *Language and Speech, 36*(4), 353–372. doi:10.1177/002383099303600401.

Carreiras, M., & Clifton, C. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition, 27*(5), 826–833. doi:10.3758/BF03198535.

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language, 66*(4), 764–805.

Clifton, C., Carlson, K., & Frazier, L. (2002). Informative prosodic boundaries. *Language and Speech, 45*(2), 87–114. doi:10.1177/00238309020450020101.

D'Argembeau, A., Feyers, D., Majerus, S., Collette, F., Van der Linden, M., Maquet, P., & Salmon, E. (2008). Self-reflection across time: Cortical midline structures differentiate between present and past selves. *Social Cognitive and Affective Neuroscience, 3*(3), 244–252. doi:10.1093/scan/nsn020.

Eerland, A., Engelen, J. A. A., & Zwaan, R. A. (2013). The influence of direct and indirect speech on mental representations. *PLoS ONE, 8*(6), e65480. doi:10.1371/journal.pone.0065480.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27*(2), 285–319. doi:10.1023/A:1023258301588.

Fodor, J. D. (2002). *Prosodic disambiguation in silent reading*. In *PROCEEDINGS-NELS* (*Vol. 1*, pp. 113–132).

Friederici, A. D., Meyer, M., & von Cramon, D. Y. (2000a). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language, 74*(2), 289–300. doi:10.1006/brln.2000.2313.

Friederici, A. D., Wang, Y. H., Herrmann, C. S., Maess, B., & Oertel, U. (2000b). Localization of early syntactic processes in frontal and temporal cortical areas: A magnetoencephalographic study. *Human Brain Mapping, 11*(1), 1–11.

Hemforth, B., Konieczny, L., Scheepers, C., & Strube, G. (1998). Syntactic ambiguity resolution in German. In D. Hillert (Ed.), *Sentence processing: A crosslinguistic perspective—syntax and semantics* (*Vol. 31*, pp. 293–312). San Diego: Academic.

Kiefer, M., Sim, E. J., Herrnberger, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience, 28*(47), 12224–12230. doi:10.1523/jneurosci.3579-08.2008.

Koizumi, Y. (2009). *Processing the not-because ambiguity in English: The role of pragmatics and prosody*. Dissertation, City University of New York.

Li, C. N. (1986). Direct speech and indirect speech: A functional study. In F. Coulmas (Ed.), *Direct and indirect speech* (pp. 29–45). Berlin: Mouton de Gruyter.

Lovrić, N., Bradley, D., & Fodor, J. D. (2000). *RC attachment in Croatian with and without preposition*. Poster presented at the 6th Annual Conference on architectures and mechanisms for language processing (AMLaP), Leiden.

Lovrić, N., Bradley, D., & Fodor, J. D. (2001). *Silent prosody resolves syntactic ambiguities: Evidence from Croatian*. Paper presented at the 2nd SUNY/CUNY/NYU Conference, Stonybrook, NY.

Maynell, L. A. (1999). *Effect of pitch accent placement on resolving relative clause ambiguity in English*. Poster presented at the 12th Annual CUNY Conference on human sentence processing, New York.

Mitchell, R. L. C., Elliott, R., Barry, M., Cruttenden, A., & Woodruff, P. W. R. (2003). The neural response to emotional prosody, as revealed by functional magnetic resonance imaging. *Neuropsychologia, 41*(10), 1410–1421. doi:10.1016/s0028-3932(03)00017-4.

Ogawa, S., & Lee, T. M. (1990). Magnetic-resonance-imaging of blood-vessels at high fields: In vivo and in vitro measurements and image stimulation. *Magnetic Resonance in Medicine, 16*(1), 9–18. doi:10.1002/mrm.1910160103.

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990a). Brain magnetic-resonance-imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America, 87*(24), 9868–9872. doi:10.1073/pnas.87.24.9868.

Ogawa, S., Lee, T. M., Nayak, A. S., & Glynn, P. (1990b). Oxygenation-sensitive contrast in magnetic-resonance image of rodent brain at high magnetic-fields. *Magnetic Resonance in Medicine, 14*(1), 68–78. doi:10.1002/mrm.1910140108.

Partee, B. (1973). The syntax and semantics of quotation. In S. R. Anderson & P. Kiparsky (Eds.), *A festschrift for Morris Halle* (pp. 410–418). New York: Holt, Reinhart and Winston.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron, 36*(4), 767–776. doi:10.1016/s0896-6273(02)01060-7.

Peelen, M. V., Atkinson, A. P., & Vuilleumier, P. (2010). Supramodal representations of perceived emotions in the human brain. *Journal of Neuroscience, 30*(30), 10127–10134. doi:10.1523/jneurosci.2161-10.2010.

Quinn, D., Abdelghany, H., & Fodor, J. D. (2000). *More evidence of implicit prosody in reading: French and Arabic relative clauses*. Poster presented at the 13th Annual CUNY Conference on human sentence processing, La Jolla, CA.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123*(12), 2400–2406. doi:10.1093/brain/123.12.2400.

Shafran, R. W. (2011). *Prosody and parsing in a double PP construction in Hebrew*. Dissertation, City University of New York.

Shaked, A. (2009). *Attachment ambiguities in Hebrew complex nominals: Prosody and parsing*. Dissertation, City University of New York.

Steinhauer, K. (2003). Electrophysiological correlates of prosody and punctuation. *Brain and Language, 86*(1), 142–164. doi:10.1016/S0093-934x(02)00542-4

Steinhauer, K., & Friederici, A. D. (2001). Prosodic boundaries, comma rules, and brain responses: the closure positive shift in ERPs as a universal marker for prosodic phrasing in listeners and readers. *Journal of Psycholinguistic Research, 30*(3), 267–295.

Stites, M. C., Luke, S. G., & Christianson, K. (2013). The psychologist said quickly, "Dialogue descriptions modulate reading speed!" *Memory & Cognition, 41*(1), 137–151. doi:10.3758/s13421-012-0248-7.

Tannen, D. (1986). Introducing constructed dialogue in Greek and American conversational and literary narrative. In F. Coulmas (Ed.), *Direct and indirect speech* (pp. 311–332). Berlin: Mouton de Gruyter.

Tannen, D. (1989). "Oh talking voice that is so sweet": Constructing dialogue in conversation. In D. Tannen (Ed.), *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge: Cambridge University Press.

Vogeley, K., & Fink, G. R. (2003). Neural correlates of the first-person-perspective. *Trends in Cognitive Sciences, 7*(1), 38–42. doi:10.1016/S1364-6613(02)00003-7.

Wierzbicka, A. (1974). The semantics of direct and indirect discourse. *Research on Language & Social Interaction, 7*(3–4), 267–307.

Wildgruber, D., Riecker, A., Hertrich, I., Erb, M., Grodd, W., Ethofer, T., & Ackermann, H. (2005). Identification of emotional intonation evaluated by fMRI. *Neuroimage, 24*(4), 1233–1241. doi:10.1016/j.neuroimage.2004.10.034.

Yao, B. (2011). *Mental simulations in comprehension of direct versus indirect quotations*. PhD thesis, University of Glasgow.

Yao, B., & Scheepers, C. (2011). Contextual modulation of reading rate for direct versus indirect speech quotations. *Cognition, 121*(3), 447–453. doi:10.1016/j.cognition.2011.08.007.

Yao, B., Belin, P., & Scheepers, C. (2011). Silent reading of direct versus indirect speech activates voice-selective areas in the auditory cortex. *Journal of Cognitive Neuroscience, 23*(10), 3146–3152. doi:10.1162/jocn_a_00022.

Yao, B., Belin, P., & Scheepers, C. (2012). Brain "talks over" boring quotes: Top-down activation of voice-selective areas while listening to monotonous direct speech quotations. *NeuroImage, 60*(3), 1832–1842. doi:10.1016/j.neuroimage.2012.01.111.

Zagar, D., Pynte, J., & Rativeau IV, S. (1997). Evidence for early closure attachment on first pass reading times in French. *The Quarterly Journal of Experimental Psychology, 50*(2), 421–438. doi:10.1080/713755715.

Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience, 14*(4), 1908–1919.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences, 6*(1), 37–46. doi:10.1016/s1364-6613(00)01816-7.