



The cognitive reality of morphemes. Evidence from Italian

Chiara Cappellaro¹ · Nina Dumrukcić² · Isabella Fritz¹ · Francesca Franzon³ · Martin Maiden¹

Received: 6 March 2023 / Accepted: 6 December 2023 / Published online: 29 December 2023
© Crown 2023

Abstract

This study reports and discusses the results of a pilot psycholinguistic investigation into the *morphome* – a term created (Aronoff 1994) to indicate systematic relations between form and meaning in morphology which lack synchronic semantic, functional, or phonological determinants and are thereby purely morphological.

Despite a general consensus (cf. Bermúdez-Otero and Luís 2016) on the need to approach the question of the existence and nature of morphomic structures experimentally and interdisciplinarily, there has been no study beyond Nevins, Rodrigues, and Tang (2015), which focused on the morphomic structure in Romance verb morphology identified by Maiden (1992) and labelled (arbitrarily) the ‘L-pattern’ and concluded that in Italian, Spanish and Portuguese this structure is no longer part of native speakers’ grammar.

The present study has replicated, for Italian, the basic experimental design of Nevins *et al.* It has obtained behavioural measurements (from two experiments) including eyetracking measures (from one experiment). All these measurements converge in showing (i) a statistically significant preference for target items that are consistent with the L-/U-pattern distribution and (ii) a faster decision-making process when the L-item was chosen. We conclude that (*pace* Nevins *et al.*) this morphomic structure is part of the internalized grammar of Italian adult speakers.

Keywords Morphome · L-/U-pattern · Lexeme · Inflexional paradigm · Root allomorphy · Suppletion · Psychological reality · Eye tracking

1 Introduction

This study reports and discusses the results of an ‘on-line’ psycholinguistic investigation into the *morphome* – a term created by Aronoff in his seminal work of 1994,

✉ C. Cappellaro
chiara.cappellaro@ling-phil.ox.ac.uk

¹ Research Centre for Romance Linguistics, University of Oxford, Oxford, UK

² University of Cologne, Cologne, Germany

³ Universitat Pompeu Fabra, Barcelona, Spain

Morphology By Itself, to indicate systematic relations between form and meaning in morphology which lack synchronic semantic, functional, or phonological determinants and are thereby purely morphological. We explore a morphomic structure identified in Romance verb morphology by Maiden (1992) and labelled (arbitrarily) the ‘L-pattern’ (with a variant called ‘U-pattern’). This is a pattern of allomorphy such that the first-person singular present indicative and the whole of the present subjunctive share a root allomorph distinct from the rest of the inflexional paradigm, as exemplified in (1) with the Spanish verb *hacer* ‘to do’ and the two alternants of the root *hag-* /ag-/ and *hac-* /aθ-/ . In some varieties (in central Italy and, partially, in Romanian), this distributional pattern also includes third-person plural present indicative (and is labelled the ‘U-pattern’).

- (1) L-pattern distribution in the present indicative and subjunctive of Spanish *hacer* ‘to do’

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>hago</i>	<i>haces</i>	<i>hace</i>	<i>hacemos</i>	<i>hacéis</i>	<i>hacen</i>
PRS.SBJV	<i>haga</i>	<i>hagas</i>	<i>haga</i>	<i>hagamos</i>	<i>hagáis</i>	<i>hagan</i>

At the core of our study lies the broader question of what constitutes evidence for the existence of morphemes both at the level of linguistic analysis and in language production and acquisition – a crucial question in the current ‘morpheme debate’ (cf. the volume edited by Luís & Bermúdez-Otero in 2016, and, more recently, Hecce, 2023:83f.). The notion of the morpheme is highly contentious in linguistic theory for at least two reasons: not everybody accepts the existence of autonomously morphological phenomena and, even among those who do, not everybody agrees on their theoretical implications for the architecture of grammar and the status of morphology as an autonomous linguistic component separate from syntax and phonology (cf. the discussion of the ‘existence claim’ and the ‘morphomic-level claim’ in Bermúdez-Otero & Luís, 2016:309f.)

The compelling evidence for the existence of morphemes until now has been *diachronic* in nature and has come from morphological change in the Romance verb and in particular from what Maiden calls *diachronic coherence* (2018:13f.). This latter is the phenomenon whereby, after morphomic patterns have arisen in a linguistic system – their emergence often being caused by regular sound changes which then have ceased to operate – further morphological innovations affecting any one of the relevant paradigm cells equally affect all the others, ‘in lockstep’. If, in past stages in the history of the Romance languages, speakers were able to generalize and extend an existing distribution pattern to words where it was not present in the first place, such a pattern must have been part of the speakers’ internalized grammatical knowledge: the abstract morphomic pattern must have been ‘psychologically real’.

However, the question whether morphomic structures are an active and stable part of speakers’ linguistic mental representation, whether morphemes are psychologically real in the sense that they can be detected directly and traced through measurable psycholinguistic processes, has remained unanswered. In order to address this question, we here shift the approach from comparative-historical inference to direct psycholinguistic experimentation and observation.

A growing body of research addresses the ability to learn artificially designed verbal paradigm-like patterns, exhibiting different degrees of morphological regularity

(Saldana et al., 2022). As these studies mostly focus on general properties that foster efficiency in language learning, they exploit completely novel lexical items. While establishing a useful method to assess generalization abilities, they do not tap into the internal representation of the specific patterns of an acquired natural language.

The sole attempt hitherto to address experimentally the question of the existence of morphemes is the study by Nevins, Rodrigues, and Tang (2015) focusing on the Romance L/U-pattern. Their experiment involved a fill-in-the-blank production task with pseudo-verbs. Speakers of relevant languages were exposed to a sentence with a pseudo-verb displaying two different invented root allomorphs, one in a non-L/U-pattern cell of the paradigm and another in one of the L/U-pattern cells, and were asked to predict which of these forms would be produced in other, L/U-pattern or non-L/U-pattern cells of the paradigm, as example (2) shows (where *svimi* and *svipa* are the nonce-forms). In the majority of cases, speakers did not distribute the nonce-forms in a way consistent with the L/U-pattern and this led the authors to conclude that the putative morpheme, although it may have existed in the past,¹ is now defunct in all three languages investigated, namely Spanish, Portuguese, and Italian.

- (2) *Tu quando svimi? Spero che tu non svipa troppo tardi. Così ____ anch'io in contemporanea.*
 'When do you *svimi*2SG.IND? I hope you don't *svipa*2SG.SBJV too late. In that case, I'll ____1SG.IND at the same time.
 (Nevins et al., 2015:122)

The study by Nevins et al. (2015) suffers from a number of problems, as will be discussed in detail in §3. For example, it is an 'off-line'² study which does not record important behavioural measures such as reaction times (RTs). Furthermore, no actual measures of the distributions of L/U-patterns guided the study design, inducing possible biases in the stimuli selection and in the interpretation of results. Our study represents a significant step in the direction of a full quantitative, psycholinguistic investigation into the morpheme and it is based on a large-scale corpus analysis and two experiments administered to native speakers of Italian. The first experiment was conducted remotely via *Pavlovia* and recorded only response type distribution and reaction times. This experiment was replicated and extended methodologically with the use of eyetracking in a second experiment carried out in person at SISSA (*Scuola Internazionale Superiore di Studi Avanzati*) laboratories in Trieste (Italy). In addition to response type distribution and reaction times, the second experiment thus recorded eyetracking movements measured by mean fixation proportions and temporal trajectory of gaze for each interest area. The results from both our experiments are consistently different from those obtained by Nevins and associates, and suggest that adult speakers' knowledge of morphomic patterns is alive and well.

¹ '[I]f we had access to a time machine and ran the exact same set of stimuli and task in the experiments above in the sixteenth century (a time at which Maiden's (2005)'s observation about the L-morphome's productivity had held), it is entirely possible that participants would generalize the L-shaped morphome at that moment in time in the experiment, as by hypothesis, the L-morphome was an active principle of the grammar then, generalizable to novel forms at that stage of the language.' (Nevins et al., 2015:147).

²In 'on-line' experiments the ongoing cognitive process is captured in real time, while 'off-line' experiments can only show the outcome of that process.

Section §2 of this paper discusses the emergence of morphomic patterns in the Romance verb and the rich indirect diachronic evidence for their psychological reality. Section §3 shows how experimental evidence can be complementary to diachronic evidence: it begins with a detailed description of Nevins et al. (2015) and proceeds to discuss the rationale behind our own experiment design. Section §4 discusses the large-scale quantitative analysis of Italian verb lexemes showing the L/U-Pattern. Section §5 presents the findings of our first experiment, conducted on-line, and section §6 our second experiment conducted under laboratory conditions. Section §7 presents our overall analysis and conclusions.

2 Morphomic structures in the Romance verb

The present study was conceived as an attempt to test, with rigorous experimental methods, the theoretical implications of work which, hitherto, has been carried out principally through comparative and historical analysis of the morphology of the Romance languages (see particularly Maiden, 2018). Maiden (1992) explored the existence in the history of Italo-Romance and Ibero-Romance languages of ‘irregularity as a determinant of morphological change’, recurrent patterns of allomorphy in verb roots which are the accidental historical result of regular phonological change yet not only persist in the grammar long after the relevant phonological processes are defunct, but also serve as a kind of ‘model’ or ‘template’ for subsequent changes which have no connexion whatever with the original phonological change. That finding was greatly indebted to Malkiel’s notion of ‘deep morphology’ (Malkiel, 1974; Maiden, 2011b:64f.). What was involved was a pattern of alternation originally created by sound change, but long bereft of its original phonological causation, which could be shown to have continued to be an active force in morphological change, since it served as a kind of abstract ‘template’ for subsequent morphological changes. Such phenomena clearly qualified as ‘morphomic’ in the sense of the seminal and tellingly entitled work *Morphology By Itself* by Mark Aronoff (1994), in that they involved patterns of form-distribution which wholly lacked synchronic functional or phonological determinants, whatever their diachronic origin.

Aronoff’s own examples (for example, the so-called Latin ‘third stem’, an allomorph of the verb distributed over a disparate set of environments comprising the past participle, the future participle, the supine, and some derivational categories) are generally synchronic,³ so that a possible objection might be that the observed pattern is simply an inert relic of something which was originally, but is no longer, extra-morphologically motivated; a pattern of which only linguists, but not native speakers, might be aware. In short, are alleged morphemes ever psychologically real?

What a diachronic approach to morphomic patterns can offer is some guarantee of their psychological reality. The diachrony of a number of Romance morphomic structures reveals clear evidence that speakers are capable of abstracting distributional patterns of allomorphy and then using them as models for the distribution of completely new types of alternant (whatever their origin). Among these (see, e.g., Maiden, 2018)

³For arguments that the Latin ‘third stem’ was a diachronically robust morpheme see, e.g., Maiden (2013).

are three prominent, and practically pan-Romance, morphomic patterns involving root-allomorphy, arbitrarily⁴ labelled ‘PYTA’, ‘N-pattern’, and ‘L-pattern’. The first of these is a pattern of allomorphy which in Latin was correlated with perfective aspect and whose phonological form could vary unpredictably from verb to verb: in Romance languages the allomorphs remain, but the original aspectual underpinning has all but collapsed so that, to take the example of modern Portuguese, the remnants of the old perfective allomorph now find themselves distributed over a heterogeneous array of tenses and moods that were all originally united by being perfective but now comprise the disparate set of preterite, imperfect subjunctive, pluperfect indicative, and future subjunctive. The N-pattern comprises the singular and third person forms of the present indicative, the present subjunctive, and the imperative (in opposition to all other parts of the verb paradigm), forms which in (most) Latin verbs, and for purely phonological reasons, bore stress on the lexical root. In early Romance, the presence or absence of stress on a vowel led to sometimes radical differentiation of vowel quality and consequent allomorphy; with the loss of the original phonological processes, the result is a variety of type of vocalic allomorphy distributed over the set of cells mentioned above. The guarantee that these distributional patterns are internalized by speakers is the fact that, repeatedly and across the Romance languages, these patterns display a robustly *coherent* behaviour in the face of disparate morphological innovations. If, for example, the morphomically distributed allomorph is eliminated from one member of the specified set of cells, it will be eliminated from all of them; if some analogical modification (including the introduction of suppletive⁵ allomorphs) affects one of the relevant cells, it will affect all of them in the same way.

The L-pattern (and its variant the U-pattern), which was the focus of Maiden (1992)⁶ and is the focus of this study, developed in the following way. The L-pattern occurs throughout Romance (the U-pattern variant is restricted to parts of Italy, and to a subclass of verbs in Romanian), and arises from two sets of phonological changes (for which see, for example, Lausberg, 1965: §§387-395; 451-78; Loporcario, 2011: 143-48; Maiden, 2011a). The first set, ancestral to all Romance languages, arose principally as a result of palatalization and/or affrication of consonants immediately preceding early Romance yod. The second, of later date and found, at least historically, in all Romance languages except Sardinian, is palatalization and affrication of velar consonants before front vowels. Two things must be emphasized: the first is that these two, chronologically and phonologically distinct, sound changes coincidentally produced (in most places) the *same* distributional pattern of alternation in all the relevant verbs – a pattern opposing the first person singular present indicative and originally all forms of the present subjunctive to the whole of the rest of the inflexional paradigm (in U-pattern distributions, the third person plural present

⁴For discussion of the arbitrary labelling of morphomic patterns, see Maiden (2018:7f.). The L-pattern (and its variant the U-pattern), which will be the focus of the present study, are so called from the resemblance of the layout of the relevant patterns, as they are conventionally set down on paper, to those letters of the alphabet. The term has absolutely no phonological significance.

⁵By ‘suppletion’ we mean here alternants which are so phonologically distinct that they give the appearance of being different lexical words.

⁶Although the labels had not been invented at that stage.

indicative is also included); the second is that the phonological processes in question long ago became defunct and that they produced an extremely disparate array of phonological distinct alternants (involving the creation of novel palatal, affricated, and sometimes lengthened consonants). For a detailed account, see Maiden (2018:84-91). Some representative examples of the phonologically regular paradigmatic effects of these processes may be seen in Portuguese (L-pattern) and old Tuscan (U-pattern):

(3) Paradigmatic effects of yod

a. Portuguese (*nh* = [ɲ], *j* = [ʒ], *ç* = [s])

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>tenho</i> 'have'	<i>tens</i>	<i>tem</i>	<i>temos</i>	<i>tendes</i>	<i>têm</i>
PRS.SBJV	<i>tenha</i>	<i>tenhas</i>	<i>tenha</i>	<i>tenhamos</i>	<i>tenhais</i>	<i>tenham</i>
PRS.IND	<i>veja</i> 'see'	<i>vês</i>	<i>vê</i>	<i>vemos</i>	<i>vedes</i>	<i>vêm</i>
PRS.SBJV	<i>veja</i>	<i>vejas</i>	<i>veja</i>	<i>vejamos</i>	<i>vejaís</i>	<i>vejam</i>
PRS.IND	<i>faço</i> 'do'	<i>fazes</i>	<i>faz</i>	<i>fazemos</i>	<i>fazeis</i>	<i>fazem</i>
PRS.SBJV	<i>faça</i>	<i>faças</i>	<i>faça</i>	<i> façamos</i>	<i>façais</i>	<i>façam</i>
PRS.IND	<i>venho</i> 'come'	<i>vens</i>	<i>vem</i>	<i>vimos</i>	<i>vindes</i>	<i>vêm</i>
PRS.SBJV	<i>venha</i>	<i>venhas</i>	<i>venha</i>	<i>venhamos</i>	<i>venhais</i>	<i>venham</i>
PRS.IND	<i>meço</i> 'measure'	<i>medes</i>	<i>mede</i>	<i>medimos</i>	<i>medis</i>	<i>medem</i>
PRS.SBJV	<i>meça</i>	<i>meças</i>	<i>meça</i>	<i>meçamos</i>	<i>meçais</i>	<i>meçam</i>
PRS.IND	<i>caibo</i> 'fit'	<i>cabes</i>	<i>cabe</i>	<i>cabemos</i>	<i>cabeis</i>	<i>cabem</i>
PRS.SBJV	<i>caiba</i>	<i>caibas</i>	<i>caiba</i>	<i>caibamos</i>	<i>caibais</i>	<i>caibam</i>

b. Old Tuscan (*gli* = [ʎʎ], *gn* = [ɲ], *ggi* = [dʒʒ], *cci* = [tʃʃ])

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>vaglio</i> 'am worth'	<i>vali</i>	<i>vale</i>	<i>valemo</i>	<i>valete</i>	<i>vagliano</i>
PRS.SBJV	<i>vaglia</i>	<i>vagli</i>	<i>vaglia</i>	<i>vagliamo</i>	<i>vagliate</i>	<i>vagliano</i>
PRS.IND	<i>rimagno</i> 'stay'	<i>rimani</i>	<i>rimane</i>	<i>rimanemo</i>	<i>rimanete</i>	<i>rimagnono</i>
PRS.SBJV	<i>rimagna</i>	<i>rimagni</i>	<i>rimagna</i>	<i>rimagnamo</i>	<i>rimagnate</i>	<i>rimagnano</i>
PRS.IND	<i>veggio</i> 'see'	<i>vedi</i>	<i>vede</i>	<i>vedemo</i>	<i>vedete</i>	<i>veggiono</i>
PRS.SBJV	<i>veggia</i>	<i>veggi</i>	<i>veggia</i>	<i>veggiamo</i>	<i>veggiate</i>	<i>veggiano</i>
PRS.IND	<i>piaccio</i> 'please'	<i>piaci</i>	<i>piace</i>	<i>piacemo</i>	<i>piacete</i>	<i>piacciono</i>
PRS.SBJV	<i>piaccia</i>	<i>piacci</i>	<i>piaccia</i>	<i>piacciamo</i>	<i>piacciate</i>	<i>piacciano</i>
PRS.IND	<i>muoio</i> 'die'	<i>muori</i>	<i>muore</i>	<i>morimo</i>	<i>morite</i>	<i>muoiono</i>
PRS.SBJV	<i>muoia</i>	<i>muoi</i>	<i>muoia</i>	<i>moiamo</i>	<i>moiate</i>	<i>muoiano</i>

(4) Paradigmatic effects of palatalization of velars before front vowels

a. Portuguese

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>digo</i>	<i>dizes</i>	<i>diz</i>	<i>dizemos</i>	<i>dizeis</i>	<i>dizem</i>
PRS.SBJV	<i>diga</i>	<i>digas</i>	<i>diga</i>	<i>digamos</i>	<i>digais</i>	<i>digam</i>

b. Old Tuscan (before *i* and *e*, *c* = [tʃ], *g* = [dʒ], *gl* = [ʎʎ], *sc* = [ʃʃ]; *ch* = [k], *ggh* = [ɡɡ])

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>dico</i>	<i>dici</i>	<i>dice</i>	<i>dicemo</i>	<i>dite</i>	<i>dicono</i>
PRS.SBJV	<i>dica</i>	<i>dichi</i>	<i>dica</i>	<i>dicamo</i>	<i>dicate</i>	<i>dicano</i>
PRS.IND	<i>leggo</i> 'read'	<i>leggi</i>	<i>legge</i>	<i>leggemo</i>	<i>leggete</i>	<i>leggono</i>
PRS.SBJV	<i>legga</i>	<i>legghi</i>	<i>legga</i>	<i>leggamo</i>	<i>leggate</i>	<i>leggano</i>

Now there is abundant evidence, from across the history of the Romance languages, that this abstract pattern, long devoid of phonological conditioning and lacking any coherent functional common denominator, are 'templates' for morphological innovation. Extensive examples are given in Maiden (2018:91-122), but we give here some representative cases that involve generalization of velar alternants into verbs where they have no etymological justification, giving rise to novel and unprecedented pat-

terms of alternation whose distribution nonetheless essentially replicates that created by the original sound changes:

(5) (Early) Modern Italian⁷

	1SG	2SG	3SG	3PL
PRS.IND	<i>vengo</i>	<i>vieni</i>	<i>viene</i>	<i>vengono</i>
PRS.SBJV	<i>venga</i>	<i>venga</i>	<i>venga</i>	<i>vengano</i>
PRS.IND	<i>veggo</i>	<i>vedi</i>	<i>vede</i>	<i>veggono</i>
PRS.SBJV	<i>vegga</i>	<i>vegga</i>	<i>vegga</i>	<i>veggano</i>
PRS.IND	<i>valgo</i>	<i>vali</i>	<i>vale</i>	<i>valgono</i>
PRS.SBJV	<i>valga</i>	<i>valga</i>	<i>valga</i>	<i>valgano</i>

(6) Portuguese

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	<i>perco</i> 'lose'	<i>perdes</i>	<i>perde</i>	<i>perdemos</i>	<i>perdeis</i>	<i>perdem</i>
PRS.SBJV	<i>perca</i>	<i>percas</i>	<i>perca</i>	<i>percamos</i>	<i>percais</i>	<i>percam</i>

These innovations can even be a matter of suppletion, as demonstrated by the following pattern in certain dialects of Galician (whose morphological history is closely similar to that of Portuguese in all relevant respects), where two verbs, *colher* and *caber*, both meaning 'fit, be containable' have tended to merge according to the L-pattern (see Maiden, 2018:116f.):

(7) Galician dialects

	1SG	2SG	3SG	1PL	2PL	3PL
PRS.IND	'koʎo	'kaβes	'kaβe	ka'βemos	ka'βeðes	'kaβen
PRS.SBJV	'koʎa	'koʎas	'koʎa	ko'ʎamos	ko'ʎaðes	'koʎan

It is comparative-historical evidence of the kind sketched here that supports the inference that the Romance L-pattern (or U-pattern) is—or was in the past—psychologically real. It is of course a strong inference and not, perforce, the result of direct psycholinguistic experimentation on native speakers of Romance varieties. What would happen if we tested the psychological reality of alleged morphomic patterns on modern Romance speakers? This, as we have said, has actually been done for the L/U-pattern in Nevins et al. (2015) and their finding is disconcerting in suggesting that the L/U-pattern is extinct for modern speakers of Portuguese, Spanish, or Italian. The possibility of finding more about the mental representation of morphomic patterns has therefore seemingly slipped beyond our grasp forever. Evidence from our experiments suggests this is not the case.

3 Experiment design: going beyond Nevins, Rodrigues, and Tang (2015)

The study by Nevins and associates has been the starting point in the design of our own experiments but we have also departed from it in significant directions. The

⁷A complication with the Italian data (but not for Ibero-Romance) is that in most verbs the first and second person plural present subjunctive forms no longer show a U-pattern alternant, which is why those forms are omitted from this table. The reasons for this development, and its relation to the general theory developed here, are explained in Maiden (2012).

following discussion will focus on the Italian part of the experiments by Nevins and associates. In the Italian experiment they had 135 participants who were instructed as follows: “You will be presented with examples of invented verbs, such as *io marbo*, *tu marbi*. Then you will see a sentence with a blank space. Your task is to fill in the blank with the appropriate form of the verb”. It took participants 10 minutes to complete the task but no reaction times were measured. Target stimuli were 15 nonce verbs with root allomorphy based on the alternation of one segment (either vocalic or consonantal); the chosen alternations are not present in Italian and are thus unfamiliar to speakers. Ex.: *io svip-o*, *tu svim-i* (cf. (2)) but also *mupp-o*, *tu mopp-i* in (8) below.⁸ No information given about inflexion class. Stimuli also included 15 non-alternating nonce-verbs as fillers. Stimuli were presented in 10 carrier sentence frames of generic meaning.

- (8) *Io muppo ogni giorno, ma tu moppi soltanto una volta alla settimana. È meglio che anche tu ____ più frequentemente.*
 ‘I *muppo*1SG.IND every day, but you *moppi*2SG.IND only once a week. It’s better if you, too, ____2SG.SBJV more frequently.’
 (Nevins et al., 2015:122)

50% of participants were required to read indicative forms and produce subjunctive forms (labelled ‘Indicative>Subjunctive’); 50% of participants required to read one indicative form and one subjunctive form and produce an indicative form (labelled ‘Subjunctive>Indicative’).

The problems we identified in Nevins et al. (2015), beside the already mentioned methodological limitation of an ‘off-line’ production experiment (cf. footnote 2), concern the method of stimulus selection, in particular the *exclusive* use of non-existent phonological alternations and the decision to ignore differences of inflexion class (‘conjugation’) in the target production. We identified a further potential methodological problem in the way participants were instructed to carry out their required task (see Maiden, 2018:165f.).

3.1 Root allomorphy and phonological alternations in targets

One questionable feature of the design of Nevins et al. (2015) is the exclusive use of pseudo-verbs with root alternants which bear no phonological resemblance to those found in real verbs of the languages investigated. With reference to the Portuguese part of their experiment, the authors claim:

“In order to best test the predictions of whether the L-morpheme is actively employed in structuring the inflectional paradigms for newly-learned verbs (and hence a principle that forms some detectable part of these speakers’ morphological grammar), we created verbs with divergent forms for the 1SG.IND and

⁸The authors however do not consider that the u/o alternation (/u/ vs /ɔ/) does exist in Italian with a different distribution (N-pattern) and is not totally unfamiliar to Italian native speakers. See the verb *udire* ‘to listen’ for example: *odo*.PRS.IND.1SG, *odi*.PRS.IND.2SG, *ode*.PRS.IND.3SG, *udiamo*.PRS.IND.1PL, *udite*.PRS.IND.2PL, *odono*.PRS.IND.3PL; *oda*.PRS.SBJV.1SG; *oda*.PRS.SBJV.2SG; *oda*.PRS.SBJV.3SG; *udiamo*.PRS.SBJV.1PL, *udiate*.PRS.SBJV.2PL, *odano*.PRS.SBJV.3PL..

the 2SG.IND [. . .]. However, we used three novel morphophonological alternations, none of which are extant in Portuguese: p,f, t,s, k,x – but which are part of the phonology of other languages (e.g. spirantization in Hebrew). The motivation for avoiding existing alternations (e.g. those like ouç-/ouv- or dig-/diz-) was to specifically test the claim that ‘L-shapes’, once incorporated into the grammar of the language, form an autonomous kind of paradigm knowledge, independent and above any of the specific phonological forms themselves.” (Nevins et al., 2015:7)

Our view of the role played by word-based analogy in the development and processing of morphomic structures differs fundamentally from that of Nevins et al. (2015) and this difference is reflected in our methodology (choice of target stimuli). First, diachronic evidence (e.g. morphomic patterns even involving defectiveness, where there is no form to base analogy on, see Maiden & O’Neill, 2010) shows that morphomic structures can “form an autonomous kind of paradigm knowledge”. Second, analogy is unpredictable but is triggered by speakers’ ability to detect formal similarities between existing words or structures in their language. The frequency with which these patterns actually occur in the language can be informative of the extent to which speakers are exposed to them. This fact is particularly relevant, as similarities at the level of word forms can be active in the processes of interpretation of morphological structures (Crepaldi et al., 2010; Marelli et al., 2015; Ramscar et al., 2013). The degree of identity or variation between inflected forms, and the regularity in their patterns of variation, therefore need to be quantified, rather than assumed a priori. Our results show that formal similarity is not necessary (but may be sufficient) for morphomic structures to be activated and that when root allomorphy is extended, the abstract pattern of distribution with which the allomorphs are associated can also be extended with all three types of targets. The paradigmatic distributional pattern is totally abstract and not conditioned by any extramorphological factors, even though extension to novel lexemes may be a concrete phonological replication of an existing pattern of alternation.

3.2 Inflexion class (‘conjugation class’)

Besides phonological alternations, Nevins et al. (2015) assume no role for inflexional class in the representation of morphomic structure. They present the following description of how they treated inflexion class in their target selection and claim:

“Conjugation class was discarded as a factor, as the 1st and 2nd person present indicative are not distinct among conjugation classes in Italian. Where distinguished (in the subjunctive), we employed unambiguously 2nd conjugation verbs, to favor the likelihood of L-shaped responses.” (Nevins et al., 2015:21).

This motivation for discarding conjugation class as a factor, as formulated, is ill-founded because the authors overlook the L/U-pattern behaviour in verbs such as *venire* ‘to come’ and *finire* ‘to finish’ where PRS.IND.1SG and PRS.IND.2SG have different forms (*io vengo, tu vieni; io finisco /fi'nisko/, tu finisci /fi'nifi/*). More importantly, the L/U-pattern extends over all of the present indicative and the present subjunctive, and not only their singular forms: in Italian the PRS.IND.2PL form (*-ate* for first-conjugation, *-ete* for second-conjugation and *-ite* for third-conjugation) is an

unambiguous marker of inflexion class and in our experiments it is the PRS.IND.2PL form which is presented to speakers instead of PRS.IND.2SG. Yet again, this a priori assumption in the study by Nevins et al. (2015) may have induced a bias in the experiment design.

3.3 Form-meaning mapping bias in participants' instructions

A further major criticism is that speakers were not explicitly made aware that the variant forms in each trial shared an identical lexical meaning (see also Maiden, 2018:165f.). In the experiment conducted by Nevins et al. (2015), the instructions⁹ did not make it clear to informants that the nonce alternants were lexically identical. Well-known facts about cognitive behaviour, and the tendency to assign different meanings to different forms (e.g. the 'Principle of contrast' proposed by Clark (1993) might have led speakers to assume that the two phonologically alternants were also semantically distinct and therefore not alternants in the inflexional paradigm of the same lexeme. It is also true that in natural languages there are verbal types that present marked differences across inflected forms (see also §4.1.2), and that even suppletive forms are regularly processed by speakers. However, since such verb types are only few, occurring at high token frequency, and generally acquired at a young age (Ramscar et al., 2013), we cannot rule out the possibility that adult participants require a stabler form-to-meaning connexion before inferring that two alternating roots can belong to the same paradigm. Therefore, the possibility of inducing an uncontrollable bias in some participants and not in others calls for a more precise set of instructions, specifying explicitly the semantic unity of the novel verb forms at issue. In both our experiments, participants were made aware in the course of the preliminary explanation stage – in person (experiment 2) or they would read on their computer screen (experiment 1)– that in each trial they would be dealing with only one lexical verb. They were told that:

“For each sentence (first and second screen) you will see one verb with one meaning even if the two forms had slightly different forms such as ‘lepo’ and ‘lemete’. Imagine, for example, that you can translate ‘io lepo’ and ‘voi lemete’ with ‘I walk with my head down’ and ‘you walk with your head down’. In our imaginary world they are both forms of one and the same verb *lemere* meaning ‘to walk with one’s head down.’”¹⁰

4 A quantitative analysis of the distributions

A quantitative analysis of the presence of L/U-patterns in the lexicon is therefore crucial for two aspects. First, it gives a measure of the extent of the phenomenon.

⁹“You will be presented with examples of invented verbs, such as *io marbo, tu marbi*. Then you will see a sentence with a blank space. Your task is to fill in the blank with the appropriate form of the verb”.

¹⁰In the original Italian: “Per ogni frase (prima e seconda schermata) avrai di fronte un unico verbo con un unico significato anche se dovessero esserci forme leggermente diverse come ‘lepo’ e ‘lemete’. Per esempio, immagina di tradurre mentalmente ‘io lepo’ e ‘voi lemete’ con ‘io cammino a testa in giù’ e ‘voi camminate a testa in giù’. Nel nostro mondo immaginario infatti entrambe sono voci del verbo ‘lemere’ ovvero ‘camminare a testa in giù.’”.

This will probabilistically permit to infer the exposition of a speaker to L/U-pattern structures, and to their different subtypes (e.g., for class, for phonological type). As a consequence, it will orient the experimental questions and the experimental design offering quantitative measures to select controlled stimuli. These aspects together will allow us to have a comprehensive view of the morphomic patterns in the Italian verbal systems, and to overcome possible design flaws introduced in previous studies.

4.1 Methods

We scrutinized the lexicon of Italian verbs in order to quantify the diffusion of L/U-shaped morphemes and to assess the possible presence of phonological and morphological constraints ruling their occurrence.¹¹

Verbs were retrieved from Morph-It!, a list of about 500 000 morphologically annotated word forms of Italian (Zanchetta & Baroni, 2005). From the Morph-It! list, we extracted all the verb forms inflected in the PRS.IND.1SG and PRS.IND.3SG, PRS.IND.1PL and PRS.SBJV.1SG. Only verbs occurring in all of these forms in the list were included in our dataset, for a total of 6117 verb lemmas. We transcribed each verb form using a derived form of X-SAMPA¹² providing an unambiguous mapping between each consonant phoneme and a single letter of Latin alphabet. A conversion table of the symbols used is available in the file documentation, while in this article we will use IPA symbols. This transcription allows easily to disambiguate some opaque encodings of Italian orthography, in which the same letter or bigram stands for different phonemes. For example, the letter <c> stands for the palatal affricate [tʃ] when followed by <i> or <e>, and for the velar plosive [k] in all the other cases. For historical reasons, these ambiguities are frequently found in the presence of palatalization; hence, their disambiguation is crucial in order to retrieve morphomic patterns.

We removed the inflexional affix (inflexional ending and thematic vowel) from the inflected form of each verb and classified verbs according to lexical root and presence or absence of L/U-shaped morphemes ('L/U-pattern verbs' or 'non-L/U-pattern verbs' respectively). Verbs were classified as showing the L/U-pattern when the root of PRS.IND.1SG and subjunctive are the same, and these roots are themselves different from PRS.IND.1PL and PRS.SBJV.1SG. All other verbs in which the root is the same in all these inflected forms, are classified as non having the L/U-pattern. We obtained counts of the verbs with morphomic patterns for all conjugations. Within conjugations, we counted the distribution of root allomorphy across the phonological types, and quantified the extension of the various types of allomorphy.

4.2 Results

The distribution of verbs presenting an L/U-pattern across the conjugation is reported in Table 1 and plotted in Fig. 1. Conjugation membership plays a crucial role in the distribution of L/U-pattern verbs within the verb system of Italian.¹³

¹¹The dataset and analysis code used in this study are available at: https://github.com/franfranz/Verb_morphemes_ITA, where they are distributed with the relevant documentation for its reuse.

¹²See <http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>.

¹³For a discussion of the role of inflexion class in the development of root allomorphy in the Romance verb see Maiden (2018:273f.).

Table 1 Distribution of verb types with L/U- and non-L/U-pattern across inflexion classes (raw counts, proportions of total number of lexemes counted)

conjugation		non-L/U-pattern verbs	L/U-pattern verbs	total
first	-are	4805	17	4822
		<i>0.7855</i>	<i>0.0028</i>	<i>0.7883</i>
second	-ere	410	228	638
		<i>0.0670</i>	<i>0.0373</i>	<i>0.1043</i>
	-rre ^a	0	57	57
			<i>0.0093</i>	<i>0.0093</i>
third	-ire	96	504	600
		<i>0.0157</i>	<i>0.0824</i>	<i>0.0981</i>
total		5311	806	6117
		<i>0.8682</i>	<i>0.1318</i>	

^aThis group includes only reflexes of Latin verbs PONERE, TRAHERE and DUCERE (plus derived lexemes) which as a result of process of syncope and assimilation developed an infinitive in -rre: e.g. Italian *porre* ‘to put’, *trarre* ‘to get’ and *condurre* ‘to lead’

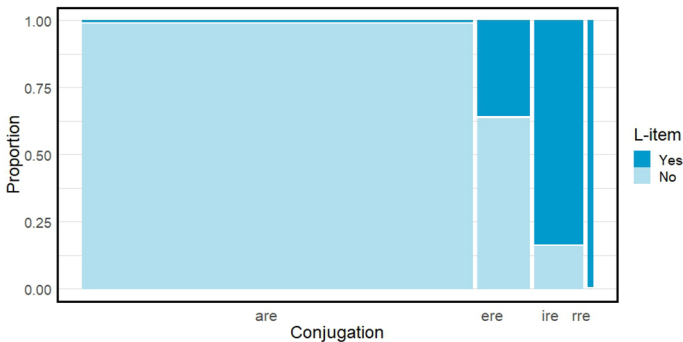


Fig. 1 Spine plot of the distribution of verb types across the conjugations. The width of the columns represents the number of types; the proportion of L/U- and non-L/U-pattern types is represented on the y-axis

The first conjugation is the most numerous class and one almost entirely devoid of and resistant to root allomorphy, a characteristic observed not only for Italian but for Romance more broadly. As Maiden claims:¹⁴

“[T]he first conjugation tends to repel root allomorphy just because first-conjugation roots show vast numbers of verbs that historically lack allomorphy. And, conversely, I suggest that non-first-conjugation verbs are vulnerable to innovative root allomorphy because they inherit large amounts of root allomorphy, much of it of regular phonological origin, and much of it associated with

¹⁴For a discussion of the special diachronic behaviour of first conjugation verbs with regard to root allomorphy see Maiden (2018:277-283).

Table 2 Distribution of Italian verbs with a lexical root (PRS.IND.1SG) ending in *-lg/*

root final <i>-lg/</i>	L/U-pattern <i>-lg/</i> vs <i>-lʎ/</i>	non-L/U-pattern
<i>-ARE</i>	N/A	e.g. <i>divulgare</i> (<i>divulgo</i> ; <i>divulghi</i>); <i>promulgare</i> (<i>promulgo</i> ; <i>promulghi</i>)
<i>-ERE</i>	e.g. <i>cogliere</i> (<i>colgo</i> , <i>cogli</i>); <i>scegliere</i> (<i>scelgo</i> , <i>scegli</i>); <i>togliere</i> (<i>tolgo</i> , <i>togli</i>); <i>sciogliere</i> (<i>sciolgo</i> , <i>sciogli</i>)	N/A

high token-frequency verbs. Speakers have reanalysed a *contingent* association with conjugation class –namely that allomorphy (and especially consonantal allomorphy) is rare in the first conjugation– as an *inherent* characteristic of that class” (Maiden, 2018: 281).

The inherent association with inflexion class becomes even more visible when we consider the shape of verb roots within the Italian lexicon. Data show, for example, a complementary distribution for verbs whose lexical root (PRS.IND.1SG) ends in *-lg/* as illustrated in Table 2. There is no such verb in the non-first conjugation that does not display the L/U-pattern.

This leads us to hypothesize that, if the L/U- morphomic structure exists in speakers’ mental representation, it must be associated with information about inflexion class. That is to say that, information about conjugation may be necessary (but is clearly not sufficient) for the morphomic structure to be activated and extended to novel words.

The corpus analysis clearly shows that root allomorphy is restricted to certain phonological types (see Table 3 and Fig. 2): all such verbs display a PRS.IND.1SG form with a root ending in velar */g, k/* or in */ʎ/* (*volere* which underwent palatalization by yod) and */tʃ/* (*fare, piacere, giacere, nuocere, tacere*, which also underwent palatalization by yod). These data also show that phonological identity of the alternants is insufficient to account for the distribution of the allomorphy, because the phonological alternations are sensitive to inflexion class (as discussed in §3.2), not occurring in the first conjugation.

4.3 Stimuli selection

Ascertaining the role (if any) of phonological substance in the potential activation of a putative morphomic pattern (reflected in speakers’ behaviour in experiments) is a much more complex and subtle process than Nevins et al. (2015) imply and is in fact at the core of our investigation into the morpheme (cf. §3.1). They claim in their conclusions: “A potential future experimental manipulation consistent with the present line of discussion would be one that contrasted wug forms of the type we employed above (e.g. mipo, mifes) with ones that bear high analogical resemblance to existing L-morphomic forms (e.g., say, mengo, mienes). If indeed the amount of L-shaped responses was modulated by the degree to which the wug forms resembled existing memorized forms, such that Natural responses continued to prevail in the former but gave way to L-shaped reponses in the latter, this would constitute potentially strong

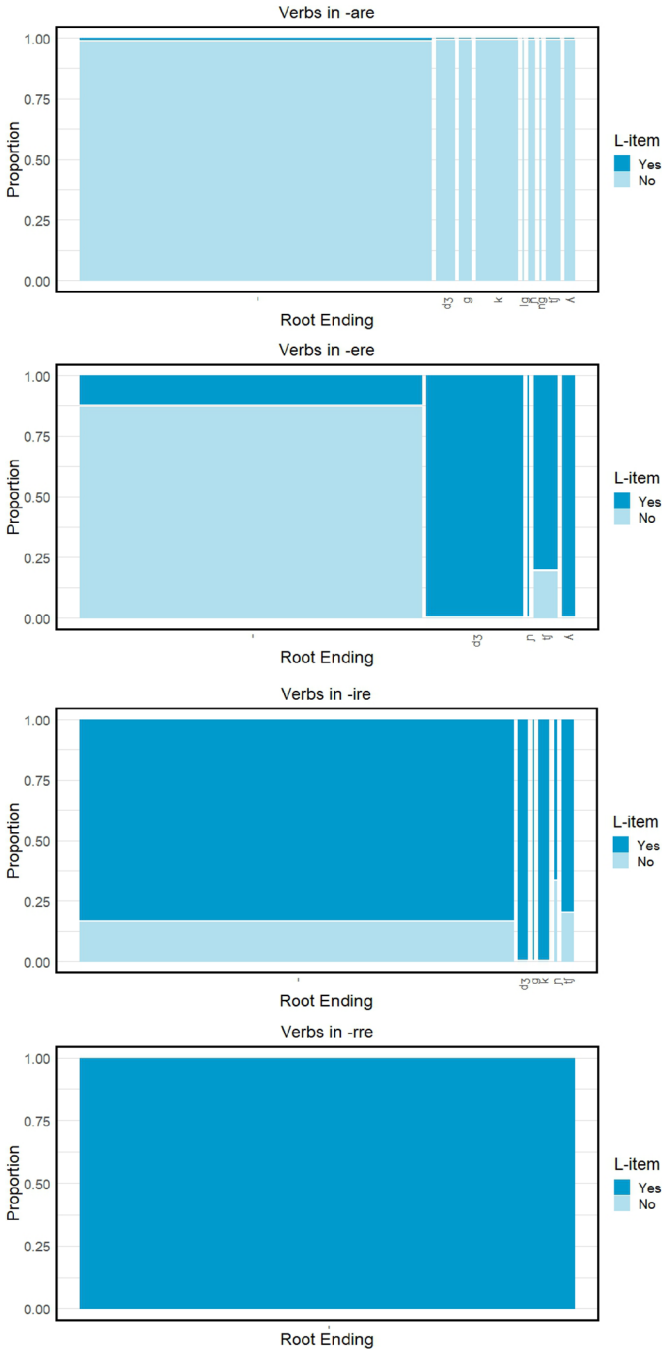


Fig. 2 Spine plots of the distribution of verb types across phonological roots for each conjugation. The width of the columns represents the number of types; the proportion of L/U- and non-L/U-pattern types is represented on the y-axis

Table 3 Distribution of verb types with L/U- and non-L/U-pattern across inflexion classes (raw counts, proportions of total number of lexemes counted)

	1-ARE		2-ERE		2-RRE ^a		3 -IRE	
	non-L/U	L/U	non-L/U	L/U	non-L/U	L/U	non-L/U	L/U
other	3701	2 ^b	393	1 ^c	0	0	86	9 ^d
	0.6050	0.0003	0.0642	0.0002			0.0141	0.0023
/dʒ/	198	0	0	0	0	0	0	0
	0.0324							
/g/	124	0	8	65	0	10	0	3 ^e
	0.0203		0.0013	0.0106		0.0016		0.0005
/ŋg/	3	0	0	39	0	0	0	4 ^f
	0.0005			0.0064				0.0007
/k/	101	0	1	5 ^g	0	0	0	0
	0.0165		0.0002	0.0008				
/ŋg/	15	0	0	69	0	30	0	17 ^h
	0.0025			0.0113		0.0049		0.0028
/p/	64	0	0	0	0	0	1	
	0.0105						0.0002	
/k/	422	0	0	19	0	17	0	13 ⁱ
	0.0690			0.0031		0.0028		0.0021
/tʃ/	154	15 ^j	8	12 ^k	0	0	3	0
	0.0252	0.0025	0.0013	0.0020			0.0005	
/sk/	23	0	0	18	0	0	6	453
	0.0038			0.0029			0.0010	0.0741

^aThis group includes only reflexes of Latin verbs PONERE, TRAHERE and DUCERE (plus derived lexemes) which following syncope acquired an infinitive in -rre: e.g. Italian *porre* 'to put', *trarre* 'to get' and *condurre* 'to lead'

^bThe verbs are *andare*, *riandare*

^cThe verb is *parere*

^dThe verbs are *apparire*, *comparire*, *riapparire*, *ricomparire*, *scomparire*, *trasparire*; *morire*, *premorire*; *riempire*

^e*Fuggire* and derived verbs

^f*Salire* and derived verbs

^g*Volere* and derived verbs

^h*Venire* and derived verbs

ⁱ*Dire* and derived verbs

^j*Fare* and derived verbs

^k*Piacere*, *giacere*, *nuocere*, *tacere* and derived verbs.

confirmation for an extension of the dual-route approach as a model of the synchronic state of L-shaped forms in Romance.” (Nevins et al., 2015:48)

In designing our experiment, therefore, we added two further groups of stimuli to that adopted by Nevins and associates, to introduce three levels of formal similarity

(high, low, zero) and test the hypothesis that speakers' behaviour might be differentiated according to the level of similarity.

1. (highest similarity) pseudo-verbs whose root alternation resembled those found in existing L/U-verbs in Italian, labelled as 'mimicking': e.g. /lɡ/ vs /k/ in nonce forms *io felgo, voi fegliete* on the model of the existing forms *io scelgo, voi scegliete* 'I/you choose'.
2. (low similarity) pseudo-verbs with root alternation that do not resemble those found in existing L/U-verbs in Italian, labelled as 'non-mimicking' but which still share all segments but one (the unfamiliar alternating segment): e.g. /d/ vs /t/ in nonce forms *io lando, voi lantete*. This group corresponds to the type of targets used in Nevins et al. (2015);
3. (no similarity) pseudo-verbs with extreme allomorphy or 'suppletion' and thus totally idiosyncratic root alternation: e.g. *io sido, voi egrete*.

5 Experiment 1

As mentioned in section §3, the design of this experiment was largely modelled on the study by Nevins et al. (2015).

5.1 Methods

Participants: The participants in this first behavioural study were 31 Italian native speakers (mean age=32.2 (SD=9.0)), recruited via Prolific.¹⁵ All participants confirmed that they had received schooling in Italian and were resident in Italy at the time of testing. They all had normal or corrected-to-normal vision and no language- or speech-impairment or dyslexia and they were all compensated for their participation. A questionnaire was administered to elicit participants' linguistic background and, in particular, their additional knowledge of Italo-Romance dialects or other Romance languages.

Stimuli and design: Each trial involved a first carrier sentence with two forms of the same pseudo-verb followed by a second sentence with a blank space or the verb, as illustrated in (9).

(9) Stimuli presentation

carrier sentence	Io verb.PRS.IND.1SG sempre e voi non verb.PRS.IND.2PL mai. 'I always verb and you never verb.'
task sentence	Non voglio che anche Maria ____ sempre. 'I don't want Maria ____ either'
target items	<i>ROOT A verb.PRS.SBJV.3SG</i> <i>ROOT B verb.PRS.SBJV.3SG</i>

Targets had root allomorphy and participants saw both roots in the carrier sentence (L/U-item with root A and the non-L/U-item with root B). The non-L/U-item was

¹⁵<https://www.prolific.com>.

invariably in the PRS.IND.2PL. Verbs used as fillers did not have root allomorphy. One important difference with respect to the design of Nevins et al. (2015) is that we forced participants to *choose* from two fixed options (target items), i.e. one L/U-item and one non-L/U-item, while Nevins and associates asked participants to *produce* a form. As discussed in Sect. 3.1, target pseudo-lexemes had to belong to one of three types (a condition labelled as ‘*Shape*’):

I. Type 1: 40 ‘mimicking’ pseudo-verbs, that is 40 invented verbs with root allomorphy based on familiar phonological alternations that mimic –to different degrees– those found in the present indicative of existing Italian verbs (with a L/U- or N-pattern distribution). In particular, we chose a consonantal type of alternation based on velar /g/ alternating with either /dʒ/ or \emptyset and the cluster /lg/ with /k/ and a vocalic type of alternation based on /e/ vs /je/ and /ɔ/ vs /wɔ/:

- /g/ vs /dʒ/ *brungol/brungete* on the model of *spingol/spingete*
- /lg/ vs /k/ *nalgol/nagliete* on the model of *scelgol/scegliete*
- /g/ vs \emptyset *drelgoldrelete* on the model of *pongolponete*
- /je/ vs /e/ *fievolfefete* on the model of *siedolsedete*
- /wɔ/ vs /ɔ/ *tuodoltodete* on the model of *muoiolmorite*

Moreover, given the pilot nature of our study and in order to allow for further potential lines of enquiry, we included in this group a smaller number of items with some of these phonological distinct alternants distributed in a mirror-image configuration, that is in the opposite direction with respect to the one found in real Italian verbs, e.g. /k/ vs /lg/ *rigliol/rilghete*, \emptyset vs /g/ *plarol/plarghete*, /e/ vs /je/ *deroldierete*, /ɔ/ vs /wɔ/ *gopolguopete*.

I. Type 2: 40 ‘non-mimicking’ pseudo-verbs whose root allomorphy is based on alternations really found in other languages such as Finnish (e.g. /t/ vs /d/ and /p/ vs /v/) Hebrew (e.g. /p/ vs /m/), or Celtic (e.g. /p/ vs /f/) but not in Italian (cf. stimuli in Nevins et al., 2015). In this group, as well, we included a number of items with the phonological distinct alternants in the opposite direction to that found in Finnish, Hebrew, and Celtic, e.g. /t/ vs /d/, /v/ vs /p/, /m/ vs /p/, and /f/ vs /p/.

II. Type 3: 20 ‘suppletive’ pseudo-verbs, that is 20 lexemes with totally idiosyncratic alternating word-forms, such as *arbo*.PRS.IND.1SG / *neschete*.PRS.IND.2PL.

Filler items that contained no stem alternation were used to mask the paradigm and were later removed from the analysis. While mimicking and non-mimicking targets were shown in the first two blocks of the experiment, together with fillers, in a (2:1) proportion, suppletive targets (half of mimicking and non-mimicking targets) with fillers also in the (2:1) proportion were presented alone in block 3 of the experiment.

Participants were equally and randomly exposed to trials where they saw the L/U-item in the PRS.IND.1SG in the carrier sentence and had to choose a PRS.SBJV.3SG form or, viceversa, they saw the L/U-item in the PRS.SBJV.3SG in the carrier sentence and had to choose a form in the PRS.IND.1SG. We replicated Nevins et al. (2015) in this respect (i.e. Indicative>Subjunctive vs Subjunctive>Indicative directions) with a crucial difference: while each participant in the Nevins et al. (2015) study was exposed to only one direction for the whole duration of their experiment, we created

a within-subjects design such that half of the participants were shown list 1 where the top 50% of the items had the Indicative>Subjunctive direction and the bottom 50% had Subjunctive>Indicative randomly, and the other half of the participants saw list 2 with the opposite direction. We also balanced for the position of pseudo-verbs in the carrier sentence (PRS.IND.2PL either in first or second position) and for the position of roots in the second sentence (root A with an L/U-item was in the left Interest Area (IA) in 50% of trials and in the right Interest Area in the other 50%). Examples of the stimuli with the invented ‘mimicking’ verb *fegliere* can be seen in (10) below.

(10) Stimuli examples

1	Indicative>Subjunctive carrier sentence	Pseudo-verbs in the	L/U-pattern Left; non-L-pattern Right Interest Area
	PRS.IND.1SG: <i>felgo</i>	PRS.IND.2PL: <i>fegliete</i>	PRS.SBJV.3SG: <i>felga /feglia</i>
2	Subjunctive>Indicative carrier sentence	Pseudo-verbs in the	L/U-pattern Left; non-L/U-pattern Right Interest Area
	PRS.SBJV.3SG: <i>felga</i>	PRS.IND.2PL: <i>fegliete</i>	PRS.IND.1SG: <i>felgo/feglio</i>
3	Indicative>Subjunctive carrier sentence	Pseudo-verbs in the	Non-L/U-pattern Left; L/U-pattern Right Interest Area
	PRS.IND.1SG: <i>felgo</i>	PRS.IND.2PL: <i>fegliete</i>	PRS.SBJV.3SG: <i>feglia/felga</i>
4	Subjunctive>Indicative carrier sentence	Pseudo-verbs in the	Non-L/U-pattern Left; L/U-pattern Right Interest Area
	PRS.SBJV.3SG: <i>felga</i>	PRS.IND.2PL: <i>fegliete</i>	PRS.IND.1SG: <i>feglio/felgo</i>
5	Indicative>Subjunctive carrier sentence	Pseudo-verbs in the	L/U-pattern Left; non-L/U-pattern Right Interest Area
	PRS.IND.2PL: <i>fegliete</i>	PRS.IND.1SG: <i>felgo</i>	PRS.SBJV.3SG: <i>felga /feglia</i>
6	Subjunctive>Indicative carrier sentence	Pseudo-verbs in the	L/U-pattern Left; non-L/U-pattern Right Interest Area
	PRS.IND.2PL: <i>fegliete</i>	PRS.SBJV.3SG: <i>felga</i>	PRS.IND.1SG: <i>felgo/feglio</i>
7	Indicative>Subjunctive carrier sentence	Pseudo-verbs in the	Non-L/U-pattern Left; L/U-pattern Right Interest Area
	PRS.IND.2PL: <i>fegliete</i>	PRS.IND.1SG: <i>felgo</i>	PRS.SBJV.3SG: <i>feglia/felga</i>
8	Subjunctive>Indicative carrier sentence	Pseudo-verbs in the	Non-L/U-pattern Left; L/U-pattern Right Interest Area
	PRS.IND.2PL: <i>fegliete</i>	PRS.SBJV.3SG: <i>felga</i>	PRS.IND.1SG: <i>feglio/felgo</i>

Targets and fillers were randomly slotted into one of ten sentence frames and counter-balanced across the aforementioned conditions. Black Courier New font was used so that each letter had the same horizontal length. The ten carrier sentences had approximately the same length and had a generic meaning (having highly frequent words or a very common sentence structure for example). The experiment was programmed in PsychoPy3.

Procedure: Participants signed up to the study via Prolific. They received a link to the online survey tool Jisc¹⁶ where they completed a consent form and answered questions regarding their schooling history and language background. At the end of the questionnaire participants were then directed to the online experiment which was hosted on Pavlovia.¹⁷ In the ‘instructions’ section participants were walked through an example trial of the experiment. In particular, as discussed in §3.3, participants’ attention was explicitly drawn to the fact that in each trial they would be dealing with one and the same lexeme regardless of the differences across word-forms.

¹⁶<https://www.jisc.ac.uk>.

¹⁷<https://pavlovia.org>.

(11) Voi bridete sempre e io non brito mai¹⁸

Self-paced reading was used, and the participants pressed the space bar on their keyboard to move. A second sentence appeared where the blank space in the clause required the reader to choose between two options eliciting a third form:

(12) Non è giusto che anche Maria non ____ mai.

They were then shown two potential items, one consistent with the L-pattern (*brita*) while the other was not (*brida*). All the trials were fully randomised without any restrictions, thus every participant had a different order.

(13) brita brida

Participants were instructed to press F if they thought the pseudo-verb on the left fitted the blank space best and J if they wanted to choose the verb presented on the right-hand side of the screen. No feedback on whether they had chosen the ‘correct’ response was provided. Participants were asked to answer as quickly as possible and RT measurement started as soon as the two optional target items appeared on the screen. Trials were presented in a random order across three blocks. The first two blocks included mimicking and non-mimicking trials plus controls whilst the third block included suppletive trials plus controls. The whole experiment including the questionnaire, took about 30 minutes to complete.

5.2 Results

5.2.1 Response type distribution

First we considered the *Response Type* (choice of L/U-item vs a non-L/U-item) across *Shape* (the three types of mimicking, non-mimicking, and suppletive targets). Participants chose an L/U-item significantly more often than a non-L/U-item across all three types of targets. A Pearson’s Chi-squared test showed the values ($\chi^2=350.2$, $df=3$, $p<0.0001$) for the mimicking and non-mimicking conditions collapsed, and ($\chi^2=57.1$, $df=1$, $p<0.0001$) for suppletion. Descriptive statistics are summarized in Table 4 and include the parameter labelled *Direction* (Indicative>Subjunctive or Subjunctive>Indicative). For all the analyses, we calculated a 95% confidence for the population mean, and the range can be found in the respective columns of the tables. The highest proportion of L/U-item responses occurred in the mimicking condition (Indicative>Subjunctive) with the other conditions having only slightly lower means and for all conditions participant chose the L/U-item over 60% of the time.

We next considered whether *Direction* interacted with *Shape*. For that, we fitted binomial generalized linear mixed effects models using the lmer package (Bates et al., 2015). *Shape* and *Direction* plus their interactions were fixed effects with *Subject* and *Item* as random effects which was the random effect structure best fitting the data (based on the model’s AIC). Both fixed factors were dummy coded with ‘Indicative>Subjunctive’ and ‘mimicking’ respectively as baseline conditions. Neither the

¹⁸‘You always *bridete* and I never *brito*.’

Table 4 Experiment 1: response type distribution

RESPONSE DISTRIBUTION	Proportion of L-/U-pattern responses	Standard Deviation	Standard Error	Confidence Interval (95%)
L-/U-pattern <i>mimicking</i> Indicative>Subjunctive	0.71	0.16	0.03	0.06
L-/U-pattern <i>mimicking</i> Subjunctive>Indicative	0.67	0.19	0.03	0.07
L-/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	0.65	0.19	0.03	0.07
L-/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	0.66	0.18	0.03	0.07
L-/U-pattern <i>suppletion</i> Indicative>Subjunctive	0.67	0.25	0.05	0.09
L-/U-pattern <i>suppletion</i> Subjunctive>Indicative	0.63	0.20	0.04	0.07

interaction nor the fixed effects yielded a significant result. Response data are illustrated in Fig. 3, where dots represent the median of L/U-pattern responses for each of the participants.

In step 1 the model told us that there is no overall effect according to whether the target was mimicking, non-mimicking, or suppletive. We then tested whether any of these three types differed significantly from each other and ran pair-wise comparisons of *Shape* conditions to see whether the participants' responses differed across mimicking, non-mimicking, and suppletive items, whilst controlling for the factor *Direction*. Pair-wise comparisons were run using the testInteractions function which is implemented in the phia package (De Rosario-Martínez et al., 2015). None of the pair-wise comparisons reached significance (all $p < 0.1$).

5.2.2 Reaction times (RTs)

For the reaction time data, we excluded one participant who took on average 33 seconds to respond. We also excluded trials exceeding ± 3 SD from each participant's mean which led to the exclusion of 1.8% of the trials (i.e., 53 trials). We then fitted models with the log-transformed RTs as a dependent variable and *Shape*, *Response Type* (choice of L/U-item vs non-L/U-item) and *Direction* as fixed factors. The random effect structure included *Item* and *Subject* as well as random slopes for *Shape* and *Direction* for *Subject* which best fitted the model's random effect structure. Using the same step-wise removal approach as for the response data, we did not find a significant interaction between *Shape*, *Response Type* and *Direction*. However, removing *Direction* from the model yielded a significant effect of this factor with slower reaction times for Indicative>Subjunctive items compared to Subjunctive>Indicative items ($\chi^2 = 10.99$, $df = 1$, $p < 0.0001$). The summary of this model also showed a significant effect of response type ($\beta = 0.06$, $SE = 0.02$, $t = 2.679$, $p = 0.007$), i.e. faster RTs when the L-/U-pattern was chosen. However, the factor *Shape* was not significant in our model. In a last step, we ran pair-wise comparisons of the *Shape* conditions to see

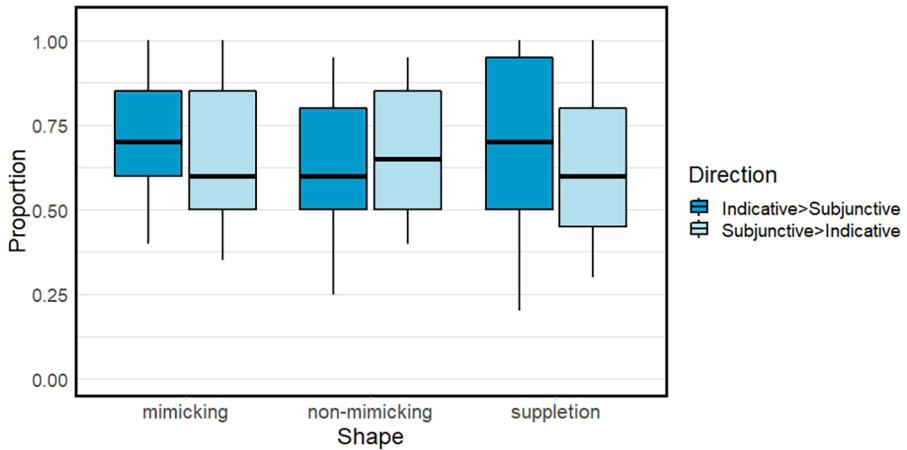


Fig. 3 Proportion of L/U-pattern responses across *Shape* and *Direction*. The line in the middle of the box is the mean and between the upper and lower edges of the box is the interquartile range

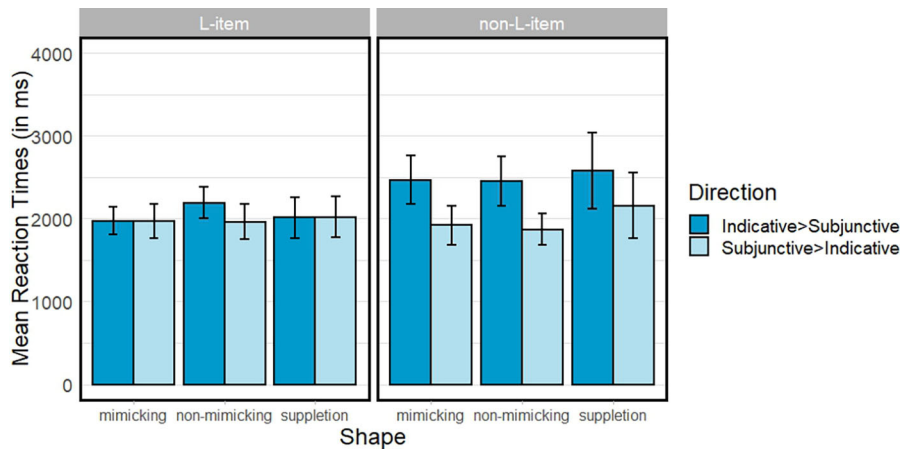


Fig. 4 Mean RTs (in ms) for *Response Type* across *Shape* and *Direction* conditions. The graph shows how long it took the participant to press the key when they selected the L/U-item and when they selected the non-L/U-item and the standard error bars are marked

whether the RTs differed overall across mimicking, non-mimicking, and suppletion. None of the pair-wise comparisons reached significance (see Table 5 and Fig. 4).

5.3 Discussion

The data from experiment 1 showed that participants chose an L/U-item significantly more often than a non-L/U-item across all three *Shape* conditions (i.e., mimicking, non-mimicking, and suppletion). For the mimicking type, participants chose the L/U-item more often but this difference did not reach significance. However, even though we did not find any overall significant differences across target types we believe that

Table 5 Experiment 1: Reaction Times

REACTION TIMES	Mean RTs	Standard Deviation	Standard Error	Confidence Interval (95%)
L-/U-pattern <i>mimicking</i> Indicative>Subjunctive	1970	912	167	341
L-/U-pattern <i>mimicking</i> Subjunctive>Indicative	1971	1146	209	428
Non-L-/U-pattern <i>mimicking</i> Indicative>Subjunctive	2467	1572	292	598
Non-L-/U-pattern <i>mimicking</i> Subjunctive>Indicative	1920	1260	234	479
L-/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	2191	1041	190	389
L-/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	1962	1149	210	429
Non-L-/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	2451	1609	294	601
Non-L-/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	1870	1035	189	386
L-/U-pattern <i>suppletion</i> Indicative>Subjunctive	2010	1358	248	508
L-/U-pattern <i>suppletion</i> Subjunctive>Indicative	2018	1362	249	509
Non-L-/U-pattern <i>suppletion</i> Indicative>Subjunctive	2577	2203	459	953
Non-L-/U-pattern <i>suppletion</i> Subjunctive>Indicative	2158	2116	393	805

the role of *Shape* needs be investigated further for two reasons. A closer look at the phonological make-up of alternants showed some interesting trends especially within the mimicking targets (see Fig. 5), that is when the root-final segment reflects the alternant distribution existing in real verbs, the choice of L/U-item tends to be very high. Consider, for example, the proportion of L/U-items chosen for the /lg/ vs /k/ alternants (e.g. *io molgo, mogliete* on the model of the real Italian verb *scelgo, scegliete* ‘to choose’). If, however, the segments are distributed between root A and root B in the reverse order, that is /lg/ in the non-L/U-item and /k/ in the L/U-item (e.g. *io riglio, voi rilghete*) the proportion of L/U-items chosen is considerably lower, in fact even lower than for non-mimicking targets.

This trend may suggest a weakening effect when an existing phonological alternation is reversed. In future experiments we will add this variable as a separate *Shape* condition quantitatively balanced within the overall design.

As for the reaction times, the overall effect of L-pattern indicates that when participants chose the non-L/U-pattern they took longer to respond. This, as will be discussed in more detail in §6.3, correlates with uncertainty and hesitation prior to

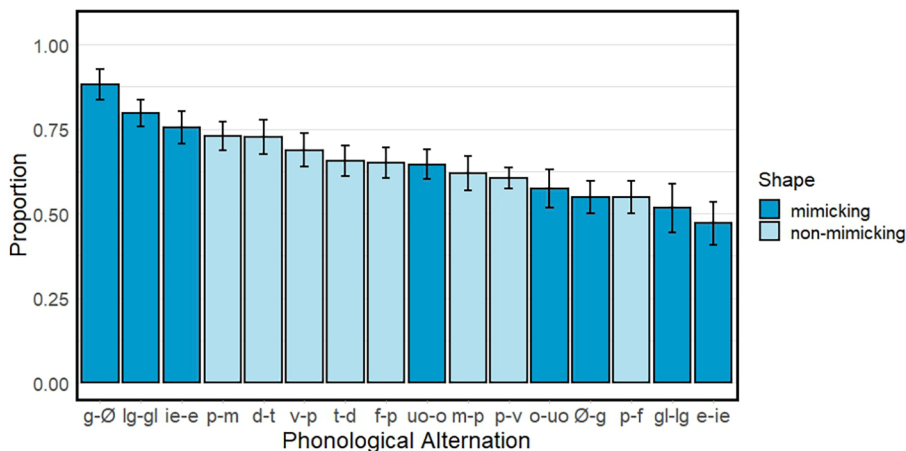


Fig. 5 The proportion of L/U-pattern responses according to *Shape* (subgroups mimicking and non-mimicking) in experiment 1. The colours show which phonological alternations belong to the mimicking subgroup and which belong to the non-mimicking subgroup

them responding. Moreover, reaction times did not differ overall across mimicking, non-mimicking and suppletive items.

6 Experiment 2

The second experiment examined, in addition to response type distribution and reaction times, eyetracking movements measured by mean fixation proportions and temporal trajectory of gaze for each interest area.

The eyetracking methodology was employed in this second experiment with the aim of gaining a deeper insight into the underlying cognitive mechanisms associated with morphomic-pattern processing. It is known from previous eyetracking studies that visual input is obtained during fixations and is largely suppressed during saccades (see Rayner, 1998). Fixation proportions and gaze likelihood over time, obtained from eyetracking data, may indicate hesitation (Prokaeva et al., 2021), preference in a forced-choice task (Shimojo et al., 2003; Simion & Shimojo, 2006; Glaholt & Reingold, 2009), and decision-making in a gap-filled task (McCray & Brunfaut, 2018). Shimojo et al. (2003) examined preference and the decision-making process by analysing eye movements in a two-alternative forced-choice task. Their findings show that just before the participants' response, there was a progressive increase in gaze likelihood towards the stimulus which was chosen. The authors referred to this phenomenon as 'gaze-cascade effect' which modelled the process of choosing one item over the other when they are simultaneously presented and how this is reflected by looking at that item longer before pressing a key. Glaholt & Reingold (2009) replicated the two-alternative forced-choice paradigm by Shimojo et al. (2003) and gaze contingent two-alternative forced-choice task resembling Simion and Shimojo (2006) and they obtained very similar results to those previously reported. Furthermore, the authors also carried out an experiment with an eight-alternative forced-choice task to

examine preference and their findings demonstrated a gaze bias effect for preference. Further studies (Simion and Shimojo 2006, 2007; Morii & Sakagami, 2015; Saito et al., 2017) have provided robust evidence of the gaze bias during decision-making in a forced-choice task.

6.1 Methods

Participants: forty-eight adult readers volunteered to take part in the experiment. All were native speakers of Italian who had normal or corrected-to-normal vision with no language/neurological/hearing disorders. Because handedness might have affected the button presses, five left-handed subjects and one person who did not complete all blocks of the experiment, were excluded, leaving a total of 42 right-handed participants for the analysis ($M_{age}=25$, $SD_{age}=5$, male=11). All participants signed a consent form prior to the experiment and were compensated for their time.¹⁹ The sample size was determined using recommendations by Brysbaert and Stevens (2018).

Stimuli and design: the same stimuli and design were used as in experiment 1.

Apparatus: eye movements were recorded using an SR Research EyeLink Portable Duo that uses a sampling rate of 1000 Hz i.e., records the position of the eye every millisecond. The text was displayed on a white screen and participants were seated approximately 70 cm away from the monitor. The head movements were restricted using a chin rest. Viewing was binocular although the data were collected only from the participants' dominant eye. The study was conducted at SISSA (*Scuola Internazionale Superiore di Studi Avanzati*) in Trieste, Italy.

Procedure: the participants' eye movements were first calibrated and validated using a 9-point grid. The study began with practice trials so that the participants could get used to the task and ask any questions. A timer was included to measure the response rate before the button was pressed from the onset of the two target items. The experiment block had a looping structure starting with an action sequence that prepared and loaded the stimuli including a drift correct to account for "small drifts in the calculation of gaze position that can build up over time" (SR Research Manual 2011, 39). A stable fixation initiated the stimuli presentation followed by recording the position and movement of the eyes while participants are completing the task. Each trial started with a fixation cross being presented for 500 ms before the carrier sentence and the two pseudo-verbs appeared. The task was the same as in experiment 1.

The participants had a button box which they used to select the item they considered to be the best fit for the blank space. No feedback on whether they chose the 'correct' response was provided, but if the participants took longer than 10 seconds, the trial timed out. The trials were presented in random order in three blocks, with short breaks in between followed by recalibration to prevent track loss.

6.2 Results

The initial data cleaning was done using SR Research DataViewer software. After visual inspection, any timed-out responses along with practice trials and filler item trials

¹⁹Ethics approval was granted by CUREC (Central University Research Ethics Committee, University of Oxford) with code: R74844/RE002.

Table 6 Experiment 2: response type distribution

RESPONSE DISTRIBUTION	Proportion of L/U-pattern responses	Standard Deviation	Standard Error	Confidence Interval (95%)
L/U-pattern <i>mimicking</i> Indicative>Subjunctive	0.695	0.110	0.017	0.034
L/U-pattern <i>mimicking</i> Subjunctive>Indicative	0.646	0.151	0.023	0.047
L/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	0.634	0.148	0.022	0.046
L/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	0.600	0.169	0.026	0.053
L/U-pattern <i>suppletion</i> Indicative>Subjunctive	0.631	0.220	0.034	0.069
L/U-pattern <i>suppletion</i> Subjunctive>Indicative	0.564	0.297	0.046	0.093

were removed. Blinks and fixations shorter than 80 ms and longer than 800 ms were also removed (Godfroid, 2020) and the left and right Interest Areas were labelled according to whether or not they contained the L/U-item or not. The independent variables were coded as *Response Type* (choice of L/U-item vs non-L/U-item), *Shape* (mimicking/non-mimicking/suppletive), and *Direction* (Indicative>Subjunctive/Subjunctive>Indicative) as in experiment 1.

6.2.1 Response type distribution

First, the response type data were analysed. The overall response proportion for the L/U-items was 0.63 and 0.37 for the non-L/U-items: therefore the participants' choice of a target item consistent with an L/U-pattern distribution was higher than it would have been due to random chance. To first test whether there was an overall difference in the selection of L/U-item vs non-L/U-item, a Pearson's Chi-squared test with Yates' continuity correction was performed and it was statistically significant ($\chi^2=80.048$, $df=1$, $p<0.0001$). Mimicking targets had the highest proportion of L/U-pattern responses for both directions. When the *Direction* was Indicative>Subjunctive, the non-mimicking targets had a higher proportion of L/U-pattern responses than the suppletive items. Changing the *Direction* to Subjunctive>Indicative also showed a higher proportion of L/U-pattern responses for non-mimicking items than suppletive items (see Table 6 and Fig. 6).

A binomial generalized linear mixed effects model was fitted by maximum likelihood (Laplace Approximation) using the lmer package (Bates et al., 2015). The dependent variable was the proportion of L-Pattern Responses and *Shape* and *Direction* as well as their interactions were added as fixed effects with *Subject* and *Item* as random effects. The fixed factors were dummy coded with 'Indicative>Subjunctive' and 'mimicking' respectively as baseline conditions. A maximal approach was used for the random slopes with *Shape*Direction* for *Subject*, and when the model did not converge, the random slopes were fitted in a step-wise removal approach and compared using an ANOVA function. The model with the lowest AIC (9370) was

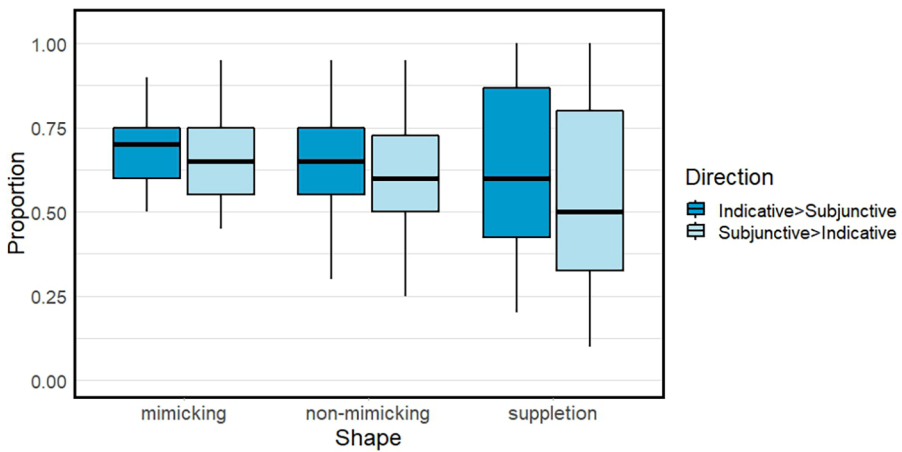


Fig. 6 Proportion of L/U-pattern responses across *Shape* and *Direction*. The line in the middle of the box is the median and between the upper and lower edges of the box is the interquartile range

the one that included *Shape* and *Direction* for the random slopes. Subsequently, the model was assessed for goodness of fit for the fixed effects. Removing *Shape:Direction* interaction did not significantly improve the fit ($\chi^2=0.875$, $df=2$, $p=0.6457$), nor did further removing *Shape* ($\chi^2=2.742$, $df=2$, $p=0.2539$). However, the inclusion of *Direction* as a main effect significantly improved the model fit ($\chi^2=5.9841$, $df=1$, $p=0.01444$) when compared to a base model and it was the model with the lowest AIC (9366). The model indicated there was no overall effect according to *Shape*, however the *Direction* was significant, with the proportion of responses being lower for the Subjunctive>Indicative ($\beta=-0.4046$, $SE=0.1641$, $z(8382)=-2.466$, $p<0.01$). We further tested whether the participants' responses differed significantly across the three *Shape* types and ran a pair-wise comparison using the testInteractions function (De Rosario-Martínez et al., 2015) with *Direction* as a fixed effect but none of the results were statistically significant.

Overall results for response type distribution across experiment 1 and 2 converge. One difference, however, is the impact of *Direction* on the response type which we only found in experiment 2. Interestingly, a further point of convergence regards the behaviour with mimicking items with respect to the alternants phonological make-up: the trend detected in experiment 1 (see §5.3, Fig. 5) is also found in experiment 2 (see Fig. 7). For example, the proportion of L/U-items chosen for the /lg/ vs /k/ alternants when their order corresponds to the one found in real verbs (e.g. *io molgo, mogliete* on the model of *scelgo, scegliete* 'to choose') is very high. If, however, the segments are distributed between the two roots in the reverse order, that is /lg/ in the non-L/U-item and /k/ in the L/U-item (e.g. *io riglio, voi rilghete*) the proportion of L/U-items chosen is considerably lower, in fact even lower than for non-mimicking targets. The full output of the model and how it was fitted can be viewed in the supplementary material (<https://osf.io/pxd59/>).

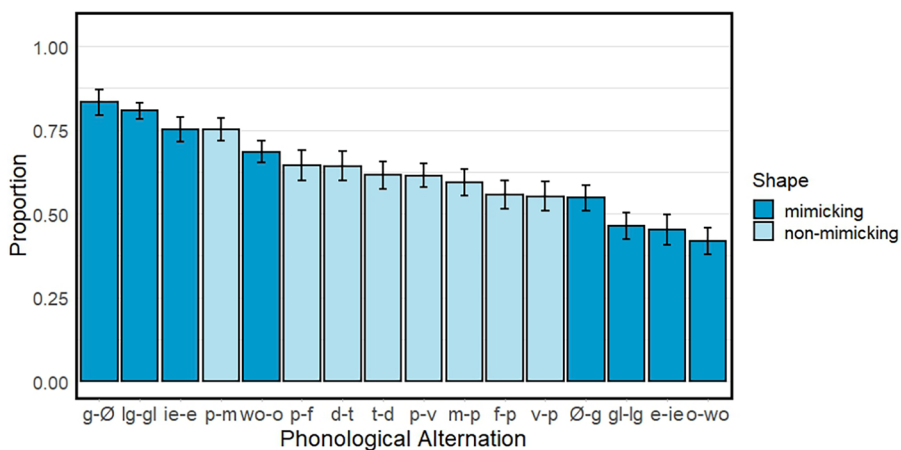


Fig. 7 The proportion of L/U-pattern responses according to *Shape* (subgroups mimicking and non-mimicking) in experiment 1. The colours show which phonological alternations belong to the mimicking subgroup and which belong to the non-mimicking subgroup

6.2.2 Reaction times (RTs)

We then analysed the time it took participants to respond measured from the onset of the two target items to the button press. Overall, mean RTs were shorter when participants chose the L/U-item ($M=2896$ ms, $SD=1362$) as opposed to when they chose the non-L/U-item ($M=3199$ ms, $SD=1474$). The full means and standard deviations across all conditions can be seen in Table 7.

The full mean RTs showed that it took the participants on average longer to press a button when faced with the Indicative>Subjunctive *Direction* as opposed to the Subjunctive>Indicative *Direction*. Moreover, even when grouped according to *Direction*, the non-L/U-pattern items had longer reaction times than the L/U-pattern items (see Fig. 8).

In order to improve skewness, the dependent variable *Reaction Times* variable was log transformed. The baseline conditions were set as ‘L/U-pattern’, ‘Indicative>Subjunctive’ and ‘mimicking’. A maximal mixed effects model was fitted with reaction times as dependent variables, *Response Type*, *Shape*, and *Direction*, and *Shape*Response Type*Direction* interaction as fixed effects with *Subject* and *Item* variables as random effects and *Shape*Response Type*Direction* as random slopes for *Subject*. The same step-wise removal approach was used as for the other measurements and the models were compared using the ANOVA function. The maximal model had the lowest AIC (5236.186) and subsequently the fixed effects were compared using *Shape*Response Type*Direction* as random slopes for *Subject*. When looking at the fixed effects, removing *Shape:Direction* and *Response Type:Direction* interactions did not significantly affect the model fit. Inclusion of *Shape:Response Type* interaction led to a significantly better fit ($\chi^2=12.028$, $df=4$, $p=0.01714$). The best fitting model showed that there was a significant main effect of *Response Type* with the non-L/U-pattern items having longer RTs than the L/U-pattern items ($\beta=0.125$,

Table 7 Experiment 2: difference between mean reaction times (in ms)

REACTION TIMES	Mean RTs	Standard Deviation	Standard Error	Confidence Interval (95%)
L/U-pattern <i>mimicking</i> Indicative>Subjunctive	3043	708	109	220
L/U-pattern <i>mimicking</i> Subjunctive>Indicative	2624	805	124	251
Non-L/U-pattern <i>mimicking</i> Indicative>Subjunctive	3520	988	152	308
Non-L/U-pattern <i>mimicking</i> Subjunctive>Indicative	2859	1030	159	321
L/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	3331	915	141	285
L/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	2643	824	127	257
Non-L/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	3494	1020	157	318
Non-L/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	2913	1408	217	439
L-/U-pattern <i>suppletion</i> Indicative>Subjunctive	3491	1248	192	389
L-/U-pattern <i>suppletion</i> Subjunctive>Indicative	2859	1106	171	345
Non-L/U-pattern <i>suppletion</i> Indicative>Subjunctive	3519	1119	177	358
Non-L/U-pattern <i>suppletion</i> Subjunctive>Indicative	2807	905	147	298

$SE=0.021$, $t(8382)=5.945$, $p<0.001$) and *Direction* whereby the Subjunctive>Indicative condition had shorter RTs than Indicative>Subjunctive ($\beta=-0.199$, $SE=0.019$, $t(8382)=-10.635$, $p<0.001$). The interaction between *Response Type:Direction* was on the verge of significance ($\beta=-0.040$, $SE=0.021$, $t(8382)=-1.933$, $p=0.05$).

Considering *Shape* as a main effect also reached significance, therefore, we also ran a pair-wise comparison for the *Shape* condition with *Response Type* as a fixed effect and found that when the L/U-pattern items were chosen, there was a significant difference in RTs between mimicking and non-mimicking items ($\chi^2=5.901$, $df=1$, $p=0.015$) as well as between mimicking and suppletive items ($\chi^2=6.025$, $df=1$, $p=0.014$). However, the other comparisons of non-mimicking and suppletive items when the L/U-pattern was chosen did not yield significant results. Moreover, when the non-L/U-pattern items were chosen, there was no significant difference in RTs across *Shape*. The interaction between *Shape:Response Type* also yielded significance when comparing non-mimicking non-L/U-pattern items ($\beta=-0.060$, $SE=0.025$, $t(8382)=-2.427$, $p=0.015$) and suppletive non-L/U-pattern items ($\beta=-0.099$, $SE=0.039$, $t(8382)=-2.522$, $p=0.012$) to the baseline condition. The full output of the model and how it was fitted is provided in the supplementary material (<https://osf.io/pxd59/>).

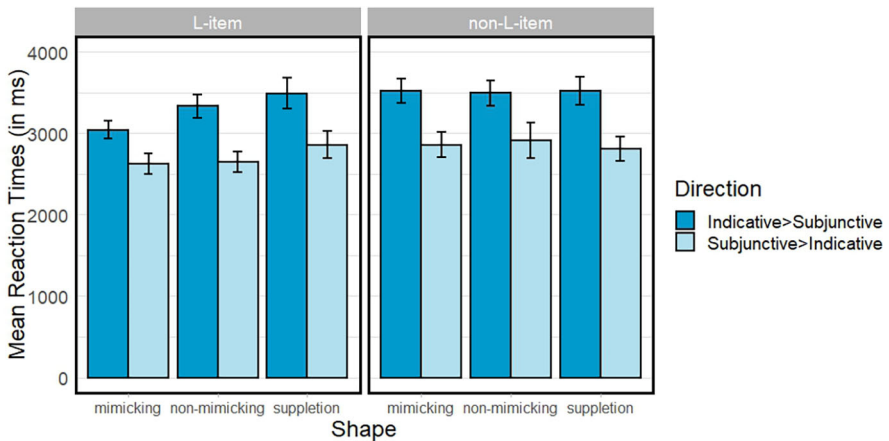


Fig. 8 Mean RTs (in ms) for *Response Type* across *Shape* and *Direction* conditions. The graph shows how long it took the participant to press the key when they selected the L/U-item and when they selected the non-L/U-item and the standard error bars are marked

The comparison between experiment 1 and 2 for reaction times show that there is not full convergence. While in both experiments, reaction times for L-/U- are always shorter than reaction times for non-L-/U-items and *Direction* also impacted the RTs (the Subjunctive>Indicative condition having shorter RTs), in experiment 2 (but not in 1) there is a significant difference across targets *Shape* between mimicking and non-mimicking/suppletive items but only when the L-item is chosen. This further indicates that the role of formal similarity (thus proportional analogy) in pattern activation deserves to be investigated further.

6.2.3 Mean fixation proportion

This eyetracking measure, averaged across trials, is often used to determine whether participants looked longer at a target or a competitor item. Recall that in this experiment both target and competitor items appear on the screen below the second sentence (with a blank to be filled) and that each sits in a coded Interest Area (IA). The Interest Period starts when the participants' pupil is first recorded entering one of the two interest areas. Therefore, the proportions do not add up to a 100% because the participants' gaze might regress back to the sentence containing the blank space while looking for further clues and deciding on which item fits better. We are considering whether participants spent more time on average looking at the interest area containing the L/U-item or the one containing the non-L/U-item throughout each trial. If participants took longer than 10 seconds to decide, the trial timed out and was excluded from the analysis. When scrutinizing the overall fixation proportions between L/U- and non-L/U-items, a Pearson's Chi-squared test with Yates' continuity correction showed a statistically significant difference between the proportion of looks to the two interest areas ($\chi^2=8378$, $df=1$, $p<0.0001$) with the L/U-item receiving a higher mean fixation proportion (0.217) than the non-L/U-item (0.206) as Table 8 shows.

Table 8 Difference between fixation proportions across conditions

FIXATION PROPORTION	Mean fixation proportion	Standard Deviation	Standard Error	Confidence Interval (95%)
L/U-pattern <i>mimicking</i> Indicative>Subjunctive	0.207	0.032	0.005	0.010
L/U-pattern <i>mimicking</i> Subjunctive>Indicative	0.238	0.036	0.006	0.011
Non-L/U-pattern <i>mimicking</i> Indicative>Subjunctive	0.190	0.030	0.005	0.009
Non-L/U-pattern <i>mimicking</i> Subjunctive>Indicative	0.224	0.037	0.006	0.011
L/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	0.203	0.032	0.005	0.010
L/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	0.230	0.041	0.006	0.013
Non-L/U-pattern <i>non-mimicking</i> Indicative>Subjunctive	0.192	0.030	0.005	0.009
Non-L/U-pattern <i>non-mimicking</i> Subjunctive>Indicative	0.222	0.038	0.006	0.012
L/U-pattern <i>suppletion</i> Indicative>Subjunctive	0.193	0.048	0.007	0.015
L/U-pattern <i>suppletion</i> Subjunctive>Indicative	0.216	0.056	0.009	0.017
Non-L/U-pattern <i>suppletion</i> Indicative>Subjunctive	0.190	0.042	0.007	0.013
Non-L/U-pattern <i>suppletion</i> Subjunctive>Indicative	0.212	0.042	0.006	0.013

According to the descriptive statistics, the fixation proportions were highest for the Subjunctive>Indicative direction and within each subgroup sorted according to *Direction*, the L/U-items had generally longer mean fixation proportions than the non-L/U-items, as shown in Fig. 9.

The analysis was carried out by fitting a maximal linear mixed effects model with the mean fixation proportion as the dependent variable, *Response Type*, *Shape*, and *Direction* as fixed effects, and *Subject* and *Item* as random effects due to high variability among the participants and *Response Type*Shape*Direction* as a random slope for *Subject*. When the model did not converge, a stepwise approach was used to remove interactions for random slopes which did not yield significance. However, including *Response Type* as a random slope for *Subject* lead to a better fit ($\chi^2=71.706$, $df=2$, $p<0.0001$) whereas including *Shape* and *Direction* as random slopes did not. Regarding the fixed effects, we started out with a maximal model including *Shape*Direction*Response Type* with *Response Type* as a random slope for *Subject*. Removing all the interactions in a stepwise manner did not reach significance. The model with the lowest AIC (-15316.31) included *Response Type* ($\beta=-0.011$, $SE=0.004$, $t(8382)=-2.490$, $p<0.001$) and *Direction* ($\beta=0.028$, $SE=0.003$,

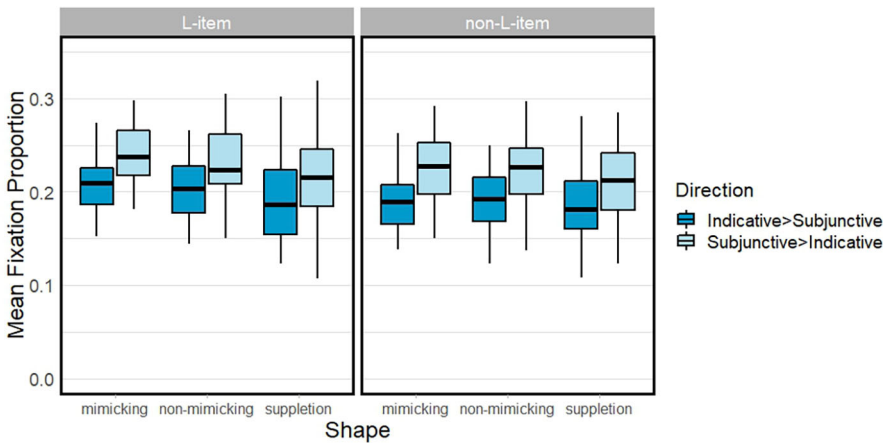


Fig. 9 Mean fixation proportion for the items tagged as *L/U-items* or *non-L/U-items* across *Shape* and *Direction* conditions

$t(8382)=10.622$, $p<0.001$) as main effects which were statistically significant for the dependent variable i.e. mean fixation proportion. The factor *Shape* showed up as significant main effect but only for suppletive items ($\beta=-0.011$, $SE=0.004$, $t(8382)=-2.972$, $p<0.003$). We therefore conducted a pair-wise comparison: using the testInteractions function with *Response Type* as a fixed factor indicated that, for both the *L/U-items* and *non-L/U-items*, there was a significant difference in mean fixation proportion for mimicking and suppletive ($\chi^2=8.834$, $df=1$, $p<0.003$) as well as non-mimicking and suppletive items ($\chi^2=5.750$, $df=1$, $p<0.016$), however the difference between mimicking and non-mimicking was statistically nonsignificant ($\chi^2=0.580$, $df=1$, $p=0.446$). The process of how the model was fitted can be found here (<https://osf.io/pxd59/>) along with the full output.

In order to explore the mean fixation proportions in further detail, we considered how likely the participants were to look at both of the words presented as choices at different points in time over the course of the trial: the data are given in the following section and presented as a temporal trajectory.

6.2.4 Temporal trajectory of gaze likelihood

The temporal trajectory of a trial shows the proportion of time a participant is looking at both interest areas (for *L/U-* and *non-L/U-item*) at given time intervals during the course of the decision-making process. The items were counterbalanced with both the *L/U-* and *non-L/U-item* occurring an equal number of times in the left and the right interest area respectively.

When, as in this case, the script is left-to-right, the reader usually enters from the left side of the screen and we observe that it is not unusual for the left-hand item to receive higher proportions, particularly at the beginning of the trial. However, as shown in Fig. 10, the *L/U-items* received higher proportions of looks with respect to the *non-L/U-items* both when they appeared on the left and when they appeared on the right. This was particularly the case at the end of the trial after the exploration period.

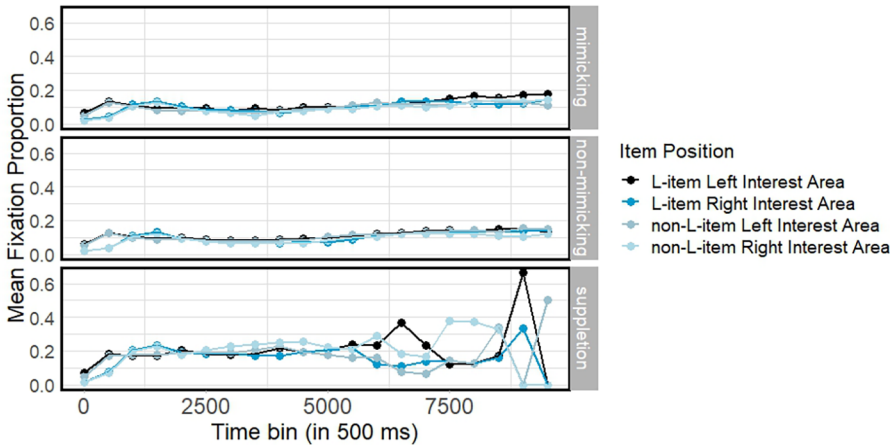


Fig. 10 Temporal Trajectory of Gaze Likelihood for each of the *Shape* conditions since the beginning until the end of the trial in 500 ms increments. The colours represent the proportion of looks when the target L-item was in the left or the right interest area and proportion of looks to the non-L-item in the left and right interest area

The proportion of looks was not determined by position of the items, which was counterbalanced. The mean fixation proportions were aggregated across participants and trials, and binned into time periods of 500 ms.

The interest period is marked from the moment participants start looking at either pseudo-verb interest area and not from the moment they start reading the second sentence. Upon entering the interest areas containing the two choices, the L/U-item received more fixations than the non-L/U-item, and this is most pronounced for the suppletive target type. At the end of the trial, just before the participants make their decision and press the key, the L/U-items receive more fixations in both the mimicking and non-mimicking types, whereas the non-L/U-items receive a higher proportion of looks in the suppletive type. This pattern did not emerge clearly when considering the overall mean fixation proportions because there is a higher proportion of looks to the L/U-items throughout the trial, particularly at 500 ms and 3000 ms. While the mean fixation proportion remains relatively consistent and stable for mimicking and non-mimicking items, for the suppletive items there appears to be a larger discrepancy across different time bins.

6.3 Discussion

The response type data analysis first focused on whether subjects were more inclined to select the target item consistent with a L/U-pattern distribution. A binomial distribution analysis showed that the tendency to choose the L/U-item is higher than chance level and therefore the vast majority of participants, given the choice, preferred an L/U-item over a non-L/U- one. One person had a very low preference for the L/U-item but we chose not to exclude any outliers that would skew the distribution since we provided no feedback for the responses. The participants chose the items that they felt fitted best in the gap purely using their own intuition and uninfluenced

by any knowledge of our theoretical preoccupations. There was a notable difference between the L/U-item being chosen over the non-L/U-item which is contradictory to the results found in Nevins et al. (2015). Our data indicate that the L/U-morphomic pattern is cognitively real and was activated during the experiment.

Note that the feature composition of the three elements in this experiment, namely forms of the first person singular present indicative, third person singular present subjunctive, and second person plural present indicative, is balanced. All three share the same value for tense (present), while two share the same value for number (singular) but have different values for person and mood, and two share the same value for mood (indicative) but have different values for person and number. Any suggestion that shared root allomorphy between a first person singular present indicative stimulus and a third person singular present subjunctive target (or vice versa) could be due to the shared value for number would have to demonstrate independently that identity of form for number was somehow more ‘natural’ (cf. Nevins et al., 2015:108) than identity of form for mood. The evidence of Italian inflexional morphology overall is actually the opposite. Root allomorphy associated with the present subjunctive is generally insensitive to number, and especially so in the third person. There are no verbs in Italian in which the third person plural present subjunctive has a different root allomorph from the third person singular present subjunctive. Our (easily falsifiable) prediction is that our experiment can be repeated with the same results for third person plural present subjunctive targets and stimuli.

Regarding the second question, whether *Shape* affects the response rate, participants selected the L/U-items more often than the non-L/U-items across all three conditions. The descriptive statistics show that participants were more likely to choose the L/U-item when the target was a mimicking type of pseudo-verb, but the differences found across the three types of targets are not statistically significant. The lowest proportion of responses for the L/U-items was for the suppletive items when the direction was Subjunctive>Indicative. We can also observe that for the suppletive targets, there was a broader range of proportions for the responses which means that there were lots of items where the participants opted for the form consistent with a L/U-pattern distribution but also a lot of items where they chose the non-L/U-item. *Direction* as a main effect was significant because the subjects were even more likely manually to select the L/U-item when the direction was Indicative>Subjunctive although in both directions the L/U-items were predominantly chosen. The tendency of the participants to select the L/U-item as opposed to a non-L/U-item did not depend on whether the participants could rely on formal similarities existing between pseudo- and real Italian verbs because they chose the L/U-item more often with all three types of targets (mimicking, non-mimicking and suppletive).

Regarding the reaction time data, our aim was to observe latency when making a decision in this two-alternative forced-choice gap-filled task and whether this was affected by choosing the option consistent with the L/U-morphomic distribution. Overall, the participants took less time to make a decision and press a key when they chose the L/U-items than when they opted for the non-L/U-items. Longer reaction times when choosing the non-L/U-item may suggest greater hesitation, whereas when they chose the L/U-items the participants felt more confident that this was the word which should be inserted into the gap. *Direction* played an important role with the participants taking less time to respond when being prompted with the subjunctive: the

L/U-pattern is in fact more likely to be activated quicker in the Subjunctive>Indicative direction compared to the Indicative>Subjunctive direction. Within the model for experiment 2, *Shape*, *Choice of L/U-item/non-L/U-item*, and *Direction* as main effects all significantly impacted the reaction times, as did the interactions between *Shape:Response Type*, however interactions between all three parameters were statistically nonsignificant.

As regards the correlation (if any) between *Shape* and reaction times, we can observe from the pairwise comparison in experiment 2 that when the L/U-item was chosen, there are significant differences between mimicking and non-mimicking as well as mimicking and suppletive items, but not significant differences between non-mimicking and suppletive items. The reason might be that the suppletive items do not follow any specific stem alternation whereas the other two types do follow our chosen morphophonological patterns. When the non-L/U-items were chosen, there was no significant difference in reaction times across *Shape*. Experiment 1 showed no significant differences in reaction times across *Shape*. This could potentially be due to differences in experiment setup since a behavioural task requires the target words to be presented last in order to measure the reaction time from the onset of target items, whereas in an eye tracking study, we presented the second sentence and target words simultaneously because it is possible to analyse the data from when the eyes enter the target areas. We also wanted to see where in the second sentence the participants' eyes searched for clues to complete the task, however this is beyond the scope of the current paper.

On the basis of the eyetracking data, inferences can be drawn about the underlying cognitive processes at work when subjects encounter pseudo-verbs that display more than one lexical root. As in McCray and Brunfaut (2018) –one of the few studies that used eyetracking for a gap-filled task– mean fixation proportion was considered both for the target interest area containing the L/U-item and the interest area containing the non-L/U-item. This measurement is often used to assess lexical activation in Visual World Paradigm tasks (e.g., Altmann, 2004; Huettig et al., 2011; Godfroid, 2020), however for our purposes this measurement was chosen to see whether the eye movements corresponded to the response type. Moreover, we were interested in hesitations and whether subjects looked longer at the items that were ultimately chosen more frequently.

Participants fixated on the L/U-items longer and the difference between mean fixation proportion in L/U- as opposed to non-L/U-items is significant. Within the full linear mixed effects model, the main cause of the fixation proportions to each interest area seems to be the *Direction* with the Subjunctive>Indicative receiving a greater proportion of looks. A pairwise comparison across *Shape* showed that while there were no differences between mean fixation proportions between mimicking and non-mimicking items, there were significant differences both between mimicking and suppletive as well as non-mimicking and suppletive with the suppletive looks receiving fewer fixation proportions. This further supports the observations that suppletive items were processed differently because they do not follow a regular pattern.

We looked more closely not just at the overall fixation proportions for the two interest areas, but also at 500 ms intervals from seeing the interest areas before making a decision. The interest period began when the participants saw the second sentence

with the blank space and two potential options were shown simultaneously. The temporal trajectory depicts a greater fixation proportion for the L/U-items, particularly at the end of the trial just before the subjects pressed the response key, regardless of whether they appeared in the left-hand or the right-hand interest area.

In their study on predicting preference by looking at fixations, Glaholt & Reinhold (2009) demonstrated robust bias towards the chosen item in gaze duration, in gaze frequency, or in both. The longer mean fixation proportions for the L/U-items is consistent with the response type which shows that participants fixated longer the interest areas that contained the item they ended up selecting. This is the case for the mimicking and non-mimicking items where we see participants first looking at the L/U-item, then going back and forth briefly to consider both items before lingering on the one consistent with the L/U-pattern distribution and pressing the key to make their choice. With the exception of the suppletive type of targets we observe a difference which indicates greater hesitancy. This fact is in line with the initial predictions, because suppletion itself is unpredictable, a fact which makes it more difficult for the subjects to choose the item. Yet even with such irregular type of target stimuli we see a preference for the L/U-items. If one were to argue that the preference for the L/U-item among mimicking items is due merely to transfer and generalization to pseudo-words of morphological structures really existing in Italian verb paradigms, one would not expect this pattern to hold also for the non-mimicking and suppletive targets.

7 Overall findings and conclusions

This study has presented an exploration of the cognitive representation and quantification of morphomic patterns in the verbal morphology of Italian. We found evidence for the psychological reality of morphemes across two separate experiments using distinct methods.

The quantitative study, aimed at assessing the distribution of verbs with and without the L/U-pattern within the inflexional system of Italian, allowed us to explore the relation between the structure of the linguistic signal and the system that processes it, a long standing issue in the search of a comprehensive theory of language in general (Mandelbrot, 1953) and of morphological structures in particular (Franzon et al. 2016, 2020; Franzon & Zanini, 2022; Pescuma et al., 2021). Further large-scale studies will be needed in order to explore this relation in typologically different languages, and assess its relevance for language change in quantitative diachronic studies, thus developing further previous insightful research conducted on limited number of types (Cathcart et al., 2022; Herce, 2022).

The experimental studies were designed to trace the presence of a mental representation of morphomic patterns in Italian native speakers. As a starting point, we assessed the robustness of the experimental findings of Nevins et al. (2015), which claimed that there was no trace of the L/U-morphomic pattern in modern native speakers' internalized grammar and which, thereby, implied that there was no purpose in developing a broad research plan to investigate the Romance evidence for the psychological reality of morphomic structures from an experimental psycholinguistic perspective. We believe we have shown, in contrast, that it is in fact vital to start

Table 9 Comparison of most significant behavioural measures across experiments

Measure	Parameter	Experiment 1	Experiment 2
Response Type	a. Choice of L/U- vs non-L/U-target	L/U-item preferred	L/U-item preferred
	b. <i>Shape</i> : mimicking vs non-mimicking/suppletive	nonsignificant	nonsignificant
	c. <i>Shape</i> : phonological alternation in mimicking targets	reverse order strongly dispreferred	reverse order strongly dispreferred
Reaction Times (RTs)	d. Time necessary to choose L/U- vs non-L/U-target	L/U-item chosen faster	L/U-item chosen faster
	e. <i>Shape</i> : mimicking vs non-mimicking/suppletive	nonsignificant	significant
	f. <i>Shape</i> : phonological alternation in mimicking targets	reverse order strongly dispreferred	reverse order strongly dispreferred

developing a full-scale psycho/neurolinguistic research agenda into the morpheme and that we can do so on the basis of the ‘classic’ and readily accessible Romance data.

Our quantitatively and qualitatively sound findings indicate in fact that Italian native speakers, given our experimental conditions, were able to extend the L/U-pattern that exists in their language to novel words. We reiterate what the most important conditions are: (i) speakers were clearly instructed on the fact that they were dealing with one lexical verb at a time, each having variant roots allomorph; (ii) speakers were forced to choose between two options; and (iii) pseudo-verbs were overtly marked as non-first conjugation.

Given these conditions, and by replicating the basic experimental design of Nevins et al. (2015), we obtained response type and reaction times measures from two experiments, summarised in Table 9 (and eyetracking measures from one experiment). They converge in showing (i) a statistically significant preference for target items that are consistent with the L/U-pattern distribution and (ii) a faster decision-making process when the L/U-item is chosen.

These measures have also indicated some further lines of enquiry. The role of *phonological similarity* (or total lack thereof) between the two root allomorphs of a lexeme, and between pseudo-verbs and real verbs, needs to be investigated further. The presence of a notable number of types with a phonologically well-defined alternance in the L/U-pattern, as revealed by the corpus analysis, suggests that speakers can be frequently exposed to those cues and use them to build an internal representation of the morpheme. In experiment 1, unlike experiment 2, differences in *Shape* did not show a statistical difference in reaction times but the response type results for individual phonological alternations revealed in both experiments a marked difference within the mimicking type of stimuli when the existing alternation was reversed (e.g. /lg/-/k/ vs /k/-/lg/), indicating a potential inhibitory effect of this condition.

However, the fact that the suppletive stimuli were processed similarly to mimicking and non-mimicking items is a strong indication that allomorphy is a sufficient

(although not necessary) trigger for pattern extension. What, in our analysis, is a necessary condition is the overt marking of verbs as belonging to a non-first conjugation class. In order to test this hypothesis, future experiments will need to have conjugation class as a variable with both first and non-first conjugation pseudo-verbs.

What insights were provided by eyetracking measures with regard to the cognitive processes associated with the activation of abstract morphological patterns? We found a statistical difference in mean fixation proportion between L/U- and non-L/U-items: targets containing an L/U-pattern distribution received a higher proportion of looks than non-L/U-items in all three types of targets (mimicking, non-mimicking and suppletive) which is consistent with their response type. This finding further supports the hypothesis that longer fixations correlate with the choice being made and therefore may indicate higher confidence in the participants' decision after considering both options. Even though online eyetracking measures are an index of online processing, they do not allow us to easily disentangle the underlying cognitive processes leading to longer fixations (see e.g. Southwell et al., 2020) and in future research, we intend to employ electroencephalographic (EEG) measures to obtain crucial information to further interpret and complement the eyetracking data, including information on whether morphomic structures are processed in real time in the brain and whether processing is underlain by differential linguistic and cognitive processes.

Acknowledgements We thank of the John Fell Fund, University of Oxford (grant n. 0009686) is gratefully acknowledged as are Prof Davide Crepaldi and his team at SISSA (Trieste, Italy) for their key collaboration in the experimental data collection, in particular for feedback on the experiment design and the use in kind of SISSA Human Laboratory equipment and facilities.

Funding This study has been funded by the John Fell Fund (University of Oxford), Project n. 0009686 (P.I. Prof Martin Maiden; C.I. Dr Chiara Cappellaro).

Declarations

Ethics approval Ethics approval was obtained from CUREC (University of Oxford) REF. R74844/RE001/RE002.

Competing Interests All authors declare that they have no financial or non-financial interests that are directly or indirectly related to the work submitted for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm'. *Cognition*, 93(2), 79–87.

- Aronoff, M. (1994). *Morphology by itself*. Cambridge: MIT Press.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bermúdez-Otero, R., & Luís, A. R. (2016). A view of the morpheme debate. *The Morpheme Debate*, 309–340.
- Brybaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, 1(1), 1–20.
- Cathcart, C., Hecce, B., & Bickel, B. (2022). Decoupling speed of change and long-term preference in language evolution: Insights from Romance verb stem alternations. In *Proceedings of the joint conference on language evolution (JCoLE)*, Kanazawa, Japan, 5 September 2022–8 September 2022. JCoLE. <https://doi.org/10.17617/2.3398549>.
- Clark, E. V. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.
- Crepaldi, D., Rastle, K., Coltheart, M., & Nickels, L. (2010). ‘fall’ primes ‘fall’, but does ‘bell’ prime ‘ball’? Masked priming with irregularly-inflected primes. *Journal of Memory and Language*, 63(1), 83–99.
- De Rosario-Martínez, H., Fox, J., Team, R. C., & De Rosario-Martínez, M. H. (2015). Package ‘phia’. CRAN Repos. Retrieved, 1.
- Franzon, F., Arcara, G., & Zanini, C. (2016). Lexical categories or frequency effects? A feedback from quantitative methods applied to psycholinguistic models in two studies on Italian. In A. Corazza, S. Montemagni, & G. Semeraro (Eds.), *Proceedings of the third Italian conference on computational linguistics CLiC-it 2016* (pp. 152–156). Turin: Accademia University Press.
- Franzon, F., & Zanini, C. (2022). The entropy of morphological systems in natural languages is modulated by functional and semantic properties. *Journal of Quantitative Linguistics*, 1–25.
- Franzon, F., Zanini, C., & Rugani, R. (2020). Cognitive and communicative pressures in the emergence of grammatical structure: A closer look at whether number sense is encoded in privileged ways. *Cognitive Neuropsychology*, 37(5–6), 355–358.
- Glaholt, M. G., Wu, M. C., & Reingold, E. M. (2009). Predicting preference from fixations. *Psychology Journal*, 7(2), 141–158.
- Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. London: Routledge.
- Hecce, B. (2022). Quantifying the importance of morphomic structure, semantic values, and frequency of use in Romance stem alternations. *Linguistics Vanguard*, 8(1), 53–68.
- Hecce, B. (2023). *The typological diversity of morphemes*. Oxford: Oxford University Press.
- Huetig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, 137(2), 151–171.
- Lausberg, H. (1965). *Lingüística románica, vol. 1*. Fonética: Gredos.
- Loporcaro, M. (2011). Phonological processes. In M. Maiden, J. C. Smith, & A. Ledgeway (Eds.), *The Cambridge history of the Romance languages* (pp. 109–154).
- Luís, A., & Bermúdez-Otero, R. (2016). *The morpheme debate*. Oxford: Oxford University Press.
- Maiden, M. (1992). Irregularity as a determinant of morphological change. *Journal of Linguistics*, 28(2), 285–312.
- Maiden, M. (2005). Morphological autonomy and diachrony. *Yearbook of Morphology*, 2004, 137–175.
- Maiden, M. (2011). Morphological persistence. In M. Maiden, J. C. Smith, & A. Ledgeway (Eds.), *The Cambridge history of the Romance languages* (pp. 155–215).
- Maiden, M. (2011). Morphemes and “stress-conditioned allomorphy”. In M. Maiden, J. C. Smith, M. Goldbach, & M.-O. Hinzelin (Eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology* (pp. 36–50).
- Maiden, M. (2012). A paradox? The morphological history of the Romance present subjunctive. In S. Gaglia & M.-O. Hinzelin (Eds.), *Inflection and word formation in Romance languages* (pp. 27–54).
- Maiden, M. (2013). The Latin “third stem” and its Romance descendants. *Diachronica*, 30, 492–530.
- Maiden, M. (2018). *The Romance verb: Morphomic structure and diachrony*. London: Oxford University Press.
- Maiden, M., & O’Neill, P. (2010). On morphomic defectiveness: Evidence from the Romance languages of the Iberian Peninsula. In M. Baerman, G. G. Corbett, & D. Brown (Eds.), *Defective paradigms: Missing forms and what they tell us* (pp. 103–124).
- Maiden, M., Smith, J. C., & Ledgeway, A. (Eds.) (2011). *The Cambridge history of the Romance languages, vol. 1: Structures*. Cambridge: Cambridge University Press.
- Malkiel, Y. (1974). New problems in Romance interfixation (I): The velar insert in the present tense (with an excursus on -zerl-zir verbs). *Romanic Philology*, 27, 04.

- Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory*, 84, 486–502.
- Marelli, M., Amenta, S., & Crepaldi, D. (2015). Semantic transparency in free stems: The effect of orthography-semantics consistency on word recognition. *Quarterly Journal of Experimental Psychology*, 68(8), 1571–1583.
- McCray, G., & Brunfaut, T. (2018). Investigating the construct measured by banked gap-fill items: Evidence from eye-tracking. *Language Testing*, 35(1), 51–73.
- Morii, M., & Sakagami, T. (2015). The effect of gaze-contingent stimulus elimination on preference judgments. *Frontiers in Psychology*, 6, 1351.
- Nevins, A., Rodrigues, C., & Tang, K. (2015). The rise and fall of the L-shaped morpheme: Diachronic and experimental studies. *Probus*, 27(1), 101–155.
- Pescuma, V. N., Zanini, C., Crepaldi, D., & Franzon, F. (2021). Form and function: A study on the distribution of the inflectional endings in Italian nouns and adjectives. *Frontiers in Psychology*, 4422.
- Prokaeva, V., Riekhakaynen, E., & Zubov, V. (2021). Can your eyes tell us why you hesitate? Comparing reading aloud in Russian as L1 and Japanese as L2. In A. Karpov & R. Potapova (Eds.), *Speech and computer: 23rd international conference, SPECOM 2021, proceedings 23*, St. Petersburg, Russia September 27–30, 2021 (pp. 540–552).
- Ramscar, M., Dye, M., & McCauley, S. M. (2013). Error and expectation in language learning: The curious absence of “mouses” in adult speech. *Language* 760–793.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Saito, T., Nouchi, R., Kinjo, H., & Kawashima, R. (2017). Gaze bias in preference judgments by younger and older adults. *Frontiers in Aging Neuroscience*, 285, 9.
- Saldana, C., Herce, B., & Bickel, B. (2022). A naturalness gradient shapes the learnability and cross-linguistic distribution of morphological paradigms. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).
- Shimojo, S., Simion, C., Shimojo, E., & Scheier, C. (2003). Gaze bias both reflects and influences preference. *Nature Neuroscience*, 6(12), 1317–1322.
- Simion, C., & Shimojo, S. (2006). Early interactions between orienting, visual sampling and decision making in facial preference. *Vision Research*, 46(20), 3331–3335.
- Simion, C., & Shimojo, S. (2007). Interrupting the cascade: Orienting contributes to decision making even in the absence of visual stimulation. *Perception & Psychophysics*, 69, 591–595.
- Southwell, R., Gregg, J., Bixler, R., & D’Mello, S. K. (2020). What eye movements reveal about later comprehension of long connected texts. *Cognitive Science*, 44(10), e12905.
- SR Research (2011). SR Research experiment builder user manual, Version 1.10.1630. Ontario: SR Research Ltd. Retrieved September 21, 2022 from <http://sr-research.jp/support/files/a2ab23fd3769ea34af5d83a3429be0e2.pdf>.
- Zanchetta, E., & Baroni, M. (2005). Morph-it! A free corpus-based morphological resource for the Italian language. In *Proceedings of corpus linguistics conference series 2005 (ISSN 1747-9398)* (Vol. 1, pp. 1–12). University of Birmingham.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.