Corpus and Computational Methods for Usage-Based Chinese Language Learning: Toward a Professional Multilingualism

Zhuo Jing-Schmidt

Abstract Language is a complex functional adaptive system (Beckner et al. in Lang Learn 59:1–26, 2009; Ellis in Usage-based perspectives on second language learning, De Gruyter, Berlin/Boston, pp 49–73, 2015). Language learning is learning to use a complex system Larsen-Freeman (Alternative approaches to second language acquisition, Routledge, London, pp 47–72, 2011). It is a multidimensional task involving social cognitive processes that interact both in time and space MacWhinney (Usage based perspectives on second language learning, De Gruyter, Berlin, pp 19–48, 2015). Language learning in the twenty-first century is "enmeshed in globalization, technologization, and mobility" and hence "emergent, dynamic, unpredictable, open ended, and intersubjectively negotiated" (The Douglas Fir Group in Mod Lang J 100:19–47, 2016: 19). Accordingly, the field of language teaching is more interdisciplinary than ever. It requires not only knowledge of language as a functional system and knowledge of language learning as a human socio-cognitive endeavor but also expanded communication across domains traditionally separated by differences in the methodology of knowledge construction. This chapter focuses on the usage-based model of language and language learning within a broader cognitive theory of linguistic knowledge and its development. The goal is to establish a theoretical relevancy and methodological necessity of corpus and computational methods to the knowledge base for language pedagogy as a practical field. In doing so, this chapter serves to ground the application of such methods as part of a professional multilingualism that informs the learning, teaching, and assessment of Chinese as a second language.

Z. Jing-Schmidt (✉)
University of Oregon, Eugene, USA
e-mail: zjingsch@uoregon.edu

X. Lu and B. Chen (eds.), *Computational and Corpus Approaches to Chinese Language Learning*, Chinese Language Learning Sciences,
https://doi.org/10.1007/978-981-13-3570-9_2

1 Introduction

Language is a complex functional adaptive system (Beckner et al. 2009; Ellis 2015). Language learning, including adult second language learning, involves multidimensional social cognitive processes that interact both in time and space (Larsen-Freeman 2011; MacWhinney 2015). More than ever before, language learning in the twenty-first century is "enmeshed in globalization, technologization, and mobility," and therefore "emergent, dynamic, unpredictable, open ended, and intersubjectively negotiated" (The Douglas Fir Group 2016: 19). Accordingly, the field of language teaching is more interdisciplinary than ever, requiring not only knowledge of language as a functional system and knowledge of language learning as a human socio-cognitive endeavor but also expanded communication across domains traditionally separated by differences in the methodology of knowledge construction. The discussion of knowledge base and methodology of knowledge construction inevitably involves the discussion of the place and role of linguistic theory and second language acquisition (SLA) research in second language teaching. The complexity and the contested nature of that relationship is evident in the discourse about second language teaching in general (Byrnes 2000), and in the discourse about knowledge base in the field of teaching Chinese as a second language (TCSL) in particular (Han 2016; Jing-Schmidt 2015; Jing-Schmidt and Peng 2018; Ke et al. 2015).[1]

Treating *Modern Language Journal* as a site of observing the discursive history of the language teaching profession throughout the twentieth century, Byrnes (2000: 489–490) viewed that history as one of identity construction and negotiation in a community striving for professionalization. That history is a symbiosis between "affective-holistic" "grassroots" movements that come and go, and powerful "top-down" "rational-analytical" theoretical constructs, especially in linguistics, that have themselves undergone major revolutions and paradigm shifts. At the same time, Byrnes observed that the professional discourse on language teaching "has increasingly become as multivoiced as the languages we teach and as multilayered as are the societies within which we practice, powerful unifying, centralizing, and standardizing moves notwithstanding." Warning of the danger of misrepresenting "important aspects of our past identities, both of our own individual identities and of the field as a whole" by characterizing "contingent ways of knowing as deviations, as positions that need to be corrected," she contended that professionalization of the language teaching field "can also mean professional multilingualism" (ibid.: 492).

What has transpired in the larger landscape of discipline formation and identity construction of second language teaching as a field in general is mirrored in the history of TCSL as a developing field. Tracing the evolution of TCSL and the impact of discipline inquiries on this field in the last two decades, Jing-Schmidt and Peng (2018: 64) noted that central to that evolution "are efforts to define and delimit TCSL as an academic discipline, to identify and specify its theoretical foundations and guiding principles" for pedagogical decision-making. These efforts eventually led to a recognition of the interdisciplinary nature of TCSL, and negotiations of the

[1] In this chapter, I use the term "second language" in a broad sense that includes "foreign language".

place of disciplines that contribute to its formation, in particular, that of linguistics. The inchoation of a positive interaction between TCSL and linguistics is evident in the increasing "dialogue and synergy between theoretical linguistics and pedagogical practice, motivated by a desire to professionalize TCSL as well as to test linguistic theories" (ibid.). This positive tendency is enabled by the increasing number of CSL instructors who are trained in linguistics or applied linguistics research, which "fortifies the knowledge structure of the field by adding discipline knowledge to existing practical expertise in language teaching," and "rebalances the discursive power in the field through the authorial voice and disciplinary authority of the teachers-cum-researchers" (ibid.: 65).

While the enthusiastic embracement of the various linguistic theories in the recent history of TCSL may have been a strategy of coping with the uncertainty in a developing field, it is also an unmistakable sign of an "open-minded optimism" about professional growth (Jing-Schmidt and Peng 2018: 72). Despite the mistakes and detours that naturally accompany the multitude of choices, that growth mindset has led to "a powerful hybridity of intellectual resources" that have contributed to the professionalization of TCSL in one way or another (ibid.). Not only is the history of TSCL as a field one that has nurtured professional multilingualism, there is strong indication that the synergy will continue. Tao (2016) presented recent examples of research and practice meeting each other in the middle toward continuing mutual integration, and illustrated how various empirical efforts can contribute to sustained professional multilingualism for TCSL. Ke et al. (2015) provided a model of standardizing and assessing the integration of theoretical knowledge in pedagogical practice in CSL teacher professional development.

These positive developments in the integration of disciplinary inquiry and pedagogy notwithstanding, tension remains in the field with regard to the relationship between research and practice. While factors contributing to the tension may be many, there is one that has been overlooked. That is the misunderstanding that knowledge of how a second language is learned, the central concern of SLA research, is complete and free of controversy, and should be trusted as the basis of authoritative guidelines for L2 instruction. Consequently, suspicions of the applicability of certain SLA findings are perceived as irrational. For example, in a recent call for rationality and corrective intervention in TCSL, Han (2016: 238–239) referred to five "broad facts" about L2 learning, which she characterized as "firmly established," and ten generalizations derived therefrom. She urged the field to hold on to the "guiding principles" based on "the robustness of the empirical facts." What Han failed to acknowledge is that knowledge of L2 learners and L2 learning is far from "firmly established." Consider as an example the construct of fossilization as permanent cessation of learning, which is the cornerstone of the Interlanguage Hypothesis proposed by Selinker (1972), the mother theory that gives rise to the five "facts" referred to in Han (2016).

The Interlanguage Hypothesis was born in the generativist tradition that denies the learnability of language from experience in favor of a modular view of language as acquired through an innate language learning device (LAD). It arose in response to the many complications of SLA that could not be accounted for by the construct of the LAD, which presumably explains child language acquisition. Instead of challenging

the empirical status of the LAD, Selinker (1972) found a way around it by proposing that interlanguage, the linguistic system produced by adult L2 learners while learning the L2, is distinct from both the learners' native language and the target language. As such its development is not governed by LAD. The defining feature of interlanguage is fossilization, initially defined as a global "permanent cessation" of learning (Selinker and Lamendella 1978: 187), later modified to be "permanent local cessation of development" (Han and Odlin 2006: 8). Both definitions dwell on the permanence of fossilization as resulting from maturation-related innate constraints, and attribute interlanguage development to a putative latent psychological structure the central component of which is L1 transfer. Thus, L1 transfer was posited as an underlying cognitive process to explain fossilization while what exactly causes L1 transfer itself was never explained.

The fundamental problem, however, is that the exact cognitive mechanisms will remain inexplicable within a modular theory of learning that denies the role of domain-general cognitive capacities essential to language learning, and the effect of language experience on learning. Insisting on the permanency of fossilization, Selinker and his followers consistently described the condition as impervious to experience and intervention, "no matter what the age of the learner or the amount of instruction he receives in the target language" (Selinker 1972: 229), and "no matter what the input or what the learner does" (Han 2004: 20). These universalistic "no matter what" statements are extraordinary claims that were never backed up by extraordinary evidence, and remain central to the fatalism of fossilization despite accumulating counterevidence. When Han (2004) offered "abundant exposure to input," "adequate motivation to learn," and "plentiful opportunity for communicative practice" as the preconditions of fossilization identification, she was negligent of the empirical duty to provide quantitative data to support the quantitative statements made by her adjectives. Namely, how is "abundant," "adequate," and "plentiful" quantified?

The last two decades saw accumulating evidence of learner variability, domain selectivity, and non-inevitability of fossilization (e.g., Abrahamsson and Hyltenstam 2009; Bongaerts 1999; Byrnes 2012; MacWhinney 2006). But what has transpired in the field is bigger and deeper than the discovery of successful L2 learners whose language development defies the notion of fossilization.[2] Mounting converging evidence supports the consensus that both domain-general cognitive capacities and experience with language shape language development, both L1 and L2. Thus, the very premise of the innatist model is shaken. When this happens, a paradigm shift

[2]The 10-year learning results in the Language Flagship Program, a model of undergraduate advanced foreign language education designed to produce professionals with superior proficiency in a critical target language, demonstrate that attainment of professional proficiency is possible. However, it takes high expectations, realistic goal setting, rigorous student-centered and individualized intervention, immersion experience, and accountability at the institutional level (Nugent and Slater 2016). The proportion of students in the program that reach the professional level are not the 5% outliers Selinker (1972) deemed ignorable as data. For example, in 2015, 39.4% of the students who completed the program reached IRL level 3 in three modalities. See the Flagship Annual Reports Archive for data: https://www.thelanguageflagship.org/content/reports.

ensues: on the horizon is an emergentist view of language (Cook et al. 2006; Ellis 1998; MacWhinney 1999). A central part of this shift is the consolidation of the usage-based model of language and language learning (Behrens 2009; Ellis 2010, 2015; Ellis and Wulff 2015; Ibbotson 2013; Tomasello 2003; Wulff 2010). In essence, taking seriously the transformative role of adaptive learning as a lifelong process, the usage-based model rejects the notion of fossilization in the sense of permanent cessation of learning. This, however, does not mean that the enormous difficulty of L2 learning should be ignored. Within the usage-based model, L1 transfer and other risk factors related to the entrenchment of L1 knowledge can be explained by a combination of domain-general factors such as cue strength, cue competition, salience, prototypicality, etc., that influence selective attention and perceptual learning (Ellis 2006, 2008), and factors pertaining to individual aptitude (Winke 2013). In addition, language learning is sensitive to frequency effects (Arnon and Snider 2010; Behrens and Pfänder 2016; Bybee 2006, 2010; Dąbrowska 2008; Ellis 2002; Ellis and Ferreira-Junior 2009; Reali 2014; Wolter and Gyllstad 2013). Converging with evidence of usage-based language learning is neurophysiological evidence of a structural plasticity of the adult brain in adaptation to environmental enrichments (Belzung and Wigmore 2013; Hofman 2002; Sale 2016). Such converging evidence should give cause for concern to adherents of the notion of the imperviousness of language development to exposure, experience, and intervention.

The example of fossilization as a theoretical construct offers a cautionary tale that conveys the need to distinguish established facts from received wisdom, and to evaluate research by examining its theoretical assumptions and the evidence behind those assumptions. SLA research must be grounded in an empirically tested theory of language and language learning. Only in this way can research properly inform teaching. To talk about adhering to SLA research before examining its underlying theoretical assumptions puts the cart before the horse. Teachers would be ill-advised to "hold on" to implications of research the theoretical assumptions of which fail to stand up to empirical evidence.

Given the ongoing construction of knowledge of SLA processes, and given the continuing emergence of counterevidence against received wisdom, future explorations will benefit from a professional multilingualism that embraces research on usage data and learner data. The remainder of this chapter will discuss the central tenets of the usage-based model of language, its application in and impact on SLA and L2 pedagogy, and why corpus and computational methods as essential tools for usage-based approaches to SLA constitute an important voice in the professional multilingualism that informs pedagogical decision-making in TCSL.

2 The Usage-Based Model of Language

Language is learned in social interaction through shared experience and practice. This intuition, tracing back to the works of Quine (1960), has increasingly consolidated into a model of linguistic representation and language learning generally

referred to as the usage-based theory. As briefly noted in the previous section, this theory views linguistic knowledge as emergent from knowledge of usage and generalizations over usage events in interaction (Barlow and Kemmer 2000; Bybee 2006, 2010; Ellis 2002; Ellis and Wulff 2015; Ellis et al. 2016a; Goldberg 2006; Tomasello 2000, 2003). In first language acquisition, knowledge of usage comes from the child's language experience, which consists of encountering utterances that people in the child's surroundings produce in communicative context. Early language learning begins with high-frequency utterances as unanalyzed wholes, which are associated with particular functions. This is the formulaic speech stage of language learning. Learning at this stage is input-sensitive and exemplar-based, and produces an "inventory of item-based utterance schemas" (Tomasello 2000: 70) paired with specific communicative intentions and functions (Ambridge and Lieven 2011; Ellis 2011; Lieven et al. 1997; Pine et al. 1998; Tomasello 1992, 2000).

However, unanalyzed language experience is insufficient for learning grammar, which is in essence a system of generalizations. As the unanalyzed formulas become entrenched over repeated usage events, incoming utterances are sorted based on their similarity to these high-frequency exemplars. Generalization of schemas takes place in this sorting process, which allows prediction of novel exemplars and the acquisition thereof. The abstraction of schemas may occur at different levels within the hierarchical category of a construction (Barlow and Kemmer 2000; Kapatsinski 2014; Tummers et al. 2005). The more different incoming utterances in the input are, the more schematic and productive the schema becomes (Bybee 2010). In the L1 acquisition of verb-centered constructions, high-frequency exemplars with perceptually salient or prototypical verbs are learned first. As "pathbreakers," these prototypical exemplars lead the consolidation of abstract schemas and constructional meanings (Goldberg 1995; Goldberg et al. 2004). Thus, the early holistic learning of formulaic speech is followed by the abstraction of construction meaning from entrenched prototype meaning, and the generalization of abstract patterns over experienced type variations in the input. In these processes, high token frequency of a prototypical exemplar facilitates the learning of constructional meaning (Goldberg et al. 2004), and high type frequency contributes to the productivity of a pattern (Bybee and Hopper 2001; Goldberg 2006; MacWhinney 1999; Tomasello 2000). Induction or generalization involves unconscious statistical inference over utterances encountered and stored in memory (Bybee 2013; Kapatsinski 2014).

The usage-based view of language and language learning is part of a broader cognitive theory of learning that recognizes the following general cognitive strategies critical to learning:

(1) The cognitive system tracks individual items across usage events and establishes integrated representations of them as episodic memory (Barsalou et al. 1998).
(2) The stored representations of individual items in the form of episodic memory have long-term influence on the categorization and learning of items subsequently encountered (Medin and Schaffer 1978; Medin and Smith 1981; Nosofsky 1988; Smith 1991; Smith and Zarate 1992).

(3) Frequency as a structural property of a category has effects on the learning of a category (Anderson 2000; Ebbinghaus 1913; Rosch and Mervis 1975).

In addition to these general cognitive strategies, prelinguistic social cognitive capacities unique to humankind are indispensable in the acquisition of language. Such a "nonlinguistic infrastructure" (Tomasello 2008: 58) consists of:

(1) Understanding communicative intentions in intersubjectively shared context and reality (Rossano et al. 2015; Schulze and Tomasello 2015; Tomasello 1999, 2003)
(2) Joint attention of child and caregiver as scaffolding for language learning (Bruner 1981; Carpenter et al. 1998, 2002; Tomasello 2003; Tomasello and Farrar 1986).

Contra the words-and-rules theory, which assumes the modular innateness of abstract rules that guide the combination of learned words in the lexicon in developing grammatical competence, the usage-based theory explains the rapid development of child grammatical competence from input by drawing on domain-general learning mechanisms and prelinguistic capacities of social cognition, thereby obviating the need to postulate innate grammatical rules (e.g., Chomsky 2002; Pinker 1999).

3 The Usage-Based Model of SLA and Its Pedagogical Impact

SLA is an interdisciplinary field that focuses on the mechanisms by which a second or foreign language is learned. SLA research developed from child language acquisition research in the 1970s in response to the need to teach English as a second language (ESL) around the world, and has since expanded to the learning and teaching of other second and foreign languages (Kramsch 2000). It subsequently became the research base for language teaching in the United States (Byrnes 1998), though SLA researchers have recognized the role of SLA as lying primarily in informing teacher's practice rather than ensuring competent practice (Ellis 1997; Tragant and Muñoz 2004).

In its earlier years, SLA research was heavily influenced by the generative framework and focused on whether and how Universal Grammar (Chomsky 1965) constrains adult L2 acquisition and interlanguage development (Schachter 1989, 1990; Selinker 1972). In the last two decades in which generative linguistics has been increasingly challenged as a theory of linguistic knowledge, its influence on SLA has been on the wane. At the same time, the usage-based model of language acquisition has resonated with researchers in SLA and increasingly gained ground. There is now accumulating evidence that many of the usage-based mechanisms underlying L1 acquisition operate in L2 language learning:

(1) The learning of L2 grammar is data-driven and depends on the properties of input (Blom et al. 2012; Gass 1997; Madlener 2016). Cue strength in the input influences cue validity (Ellis 2006; MacWhinney 2005a, b, 2015).
(2) The learning of L2 grammar is exemplar-based and prototype-driven, and abstract patterns emerge from the generalization of language experience in communication context (Ellis 2013; Ellis and Ferreira-Junior 2009; Eskildsen 2008; Goldberg and Casenhiser 2008; Larsen-Freeman 2011).
(3) The learning of L2 grammar relies on statistical learning and is sensitive to the frequency of use (Ellis 2002; Eskildsen 2012; Medlener 2016; Rebuschat and Williams 2012; Sockett and Kusyk 2015).

Recognizing the common mechanisms underlying L1 and L2 development, the usage-based model of language unifies accounts of L1 and L2 acquisition under a broader cognitive theory of language learning. Using common concepts and relying on similar explanations allow for a more systematic comparison of the psychology of language acquisition in different settings. However, despite the shared learning strategies and mechanisms, it is a recognized fact that L2 acquisition is in general more effortful and less successful than L1 acquisition. The reason is that there are many ways in which SLA differs from L1 acquisition. Ellis (2002, 2015) points out that SLA distinguishes itself from L1 acquisition in that it draws on adult conceptual knowledge, and qualitatively and quantitatively limited language input in a non-naturalistic environment, and is subject to the influence and interference from preexisting L1 knowledge and the learned attention to both L1 form and L1 function (Ellis 2013). Similarly, MacWhinney (2012, 2015: 23) attributes the reduced success of SLA to the "set of risk factors" faced by adult learners "that reduce the effectiveness of mere exposure to L2 input." These include L1-related factors of entrenchment, negative transfer, parasitism, misconnection, and the social factor of isolation that limit language input. Importantly, although the same learning mechanisms are available to both L1 and L2 learners, SLA is complicated by the interference of learners' existing L1 experience and by their insufficient exposure to, and limited meaningful interaction in, the target language. All of these differences have implications for SLA research and L2 instruction.

The insights into the general learning mechanisms underlying L1 and L2 acquisition, as well as the recognition of L2-specific learning conditions have profound impact on L2 pedagogy. Notably, usage-based SLA research has become the empirical base of language pedagogy in many languages. Pedagogical strategies have been proposed in keeping with the usage-based model of language learning (Barcroft 2013; DeKeyser 2003; Fotos 2002; Jing-Schmidt et al. 2015; Medlener 2016; Skehan 1998; Verspoor and Nguyen 2015). Protective or compensatory interventions have also been developed in L2 pedagogy for the purpose of offsetting the negative effects of L2-specific risk factors, including conscious registration of the input and explicit "noticing" of L2 features and patterns (Schmidt 1990, 1993), explicit or form-focused instruction (Ellis 2001), corrective feedback (Ellis 2009; Lyster and Mori 2006; Lyster and Ranta 1997; Sato and Lyster 2012; Van Beuningen et al. 2012), technology-assisted learning (Peterson 2006, 2010), among many other

methods and resources. In TCSL, attempts to translate the usage-based theory of language into an overall philosophy of teaching and professional development have been made (Jing-Schmidt 2015).

The usage-based model of language has reenergized SLA research and infused empirically derived insights and information into L2 language teaching. In doing so, it has pushed the field toward a higher level of professionalism and professional multilingualism. With the usage-based view of language and language learning becoming ever more vocal in SLA, it is necessary to look at the methodological tools essential to its commitment to the study of how language is learned through experience, and to understand how these tools can be part of the solution to the challenges in usage-based SLA and TCSL, and as such contribute to sustained professional multilingualism essential to the advances of our field.

4 Corpus and Computational Methods as Part of Professional Multilingualism

SLA research has traditionally preferred "experimental and introspective data over the exploration or analysis of corpus data" (Gries 2015: 159). This situation is changing. The usage-based theory of language learning has started to impact the methodology for SLA research. As a result, corpus-based methods are on the forefront of providing useful tools of data analysis in SLA research. The same can be said of computational methods. Although in itself not intellectually or theoretically affiliated with usage-based approaches to language, computational linguistics offers powerful algorithms in processing and modeling language use and learning, and therefore has affinity to the usage-based framework at the methodological level.

Corpus Linguistics investigates "relations between frequency and typicality, and instance and norm" based on a body of naturally occurring discourses or texts (Stubbs 2001: 151). It is a "major methodological paradigm in applied and theoretical linguistics" (Gries 2006: 191). Gries (2015: 195) noted that, despite a general neglect of corpus research in SLA, corpus data have become a "major source of data" in SLA research, "both on their own and in combination with experimental data." This change was enabled by the availability of large-scale language corpora including L2 learner language corpora, as well as advances in computational tools and statistical methods. It occurred in response to the call for data-driven studies of the learning of language (Gries and Ellis 2015). Corpus linguistics offers quantitative methods for exploring L2 language production and development by enabling the examination of frequency of use, frequency of collocation, and error patterns in learner corpora. Corpus research also illuminates how L2 production differs from native language use (in terms of patterns of overuse and underuse) by comparing learner corpora with native language corpora (Jing-Schmidt 2011; Li 2014; Xu 2016; Zhang and Lu 2013), and by using descriptive and inferential statistics to analyze learner data (Gries 2015). Although language use and language experience cannot be reduced to pure

frequency effects (Behrens and Pfänder 2016), learner corpora studies have shown that frequency information extracted by corpus methods sheds important light on many notions central to language learning in SLA research. For example, Jing-Schmidt (2011) compared the uses of zero anaphors in Chinese L2 heritage and non-heritage written corpora and those in a native Chinese corpus, and discovered differential learning patterns between the two learner populations, which indicates the need of a differential instructional approach to addressing varying learning needs. Zhang and Lu (2013) compared numeral classifier uses by L2 Chinese learners and native speakers by examining a longitudinal corpus of L2 written samples from two proficiency levels in contrast to a native writing corpus. They found that L2 development in numeral classifier usage is nonlinear and highly variable for the three dimensions examined—fluency, diversity, and accuracy, which suggests the potential benefits of individualized and collocation-based instructional strategies. The application of corpus-based studies goes beyond SLA research. Frequency-based reference books (e.g., Jiao et al. 2011; Xiao et al. 2009) provide useful tools of strategizing the teaching of lexical and idiomatic expressions in Chinese. Multiple chapters in this volume demonstrate as well the utilization of corpus methods in other areas of TCSL.

Computational linguistics as an interdisciplinary field was born from the synergies of multiple related fields concerned with getting computers to perform human-centered, language-related complex tasks (Huang and Lenders 2005; Jurafsky and Martin 2008). Computational linguistics focuses on developing algorithms and software for processing and modeling speech and language. Just as corpus linguistics emerged and gained power at the auspices of the availability of large-scale language data, so is computational linguistics making great strides in speech and language processing thanks to a "startling increase in computing resources available to the average computer user, thanks to the rise of the Web as a massive source of information, and thanks to the increasing availability of wireless mobile access" (Jurafsky and Martin 2008: 8). In SLA, computational methods allow researchers to probe, test, and refine theories by reliably and efficiently detecting patterns of language use in the vastness of naturally occurring linguistic materials, and modeling and explaining mechanisms of language learning and factors that impact learning. Much of the research has found application in TCSL and is revolutionizing the field by incorporating computer-assisted learning and assessment. For example, Hoshino and Yasuda (2013) developed an automatic system of discriminating Chinese retroflex and dental affricates using VOT measurement algorithm and breathing power measurement algorithm. The system can be applied in automatic speech training of L2 Chinese learners who have difficulty distinguishing and pronouncing those sounds. Hsiao et al. (2016) developed The Chinese Listening and Speaking Diagnosis and Remedial Instruction (CLSDRI) system, which employs computerized diagnostic tests to diagnose L2 Chinese learner errors in listening comprehension and speaking, and delivers remedial instruction materials to learners to assist learning (see Chen and Hsu this volume; Lee et al. this volume). Automatic Essay Scoring technologies have also been introduced to TCSL where the assessment of L2 writing has been a perennial challenge (see Chang and Sung this volume).

Corpus and computational methods intersect. Corpus linguistic methods are computational to the extent that the processing and analysis of data from large corpora rely on digital and computational technology. When used for processing large bodies of computerized language materials, computational methods converge with corpus research. By analyzing natural language data in large quantities, computational and corpus methods provide powerful quantitative analysis inaccessible by intuition and introspection, and can provide a level of objectivity or "a layer of order" in the data "where none was previously suspected" (Stubbs 2001: 169). Such methods also complement experimental psycholinguistic research that can reveal what learners know or what they think they know about language but fails to reveal patterns in large-scale naturally occurring language (Fillmore 1992; Gries 2009; Gries et al. 2005). Based on data from the British National Corpus, Ellis et al. (2016b) analyzed the usage patterns of English verb argument constructions. They employed computational tools to investigate the verb selection preferences of these constructions and mapped out the semantic network structure of the verbs. They also explored factors that influence the learning of these constructions by measuring frequencies, semantic prototypicality and cohesion, as well as cue salience related to polysemy. The findings of Ellis et al. (2016b) strongly indicate that the use, processing, and learning of language are nonarbitrary, and opened up the problem space for future research to investigate the complexity of the interaction among the patterns, and to fine-tune our understanding of that complexity. At the operational level, when aimed directly to inform language learning and teaching, computational and corpus methods provide useful resources for L2 instruction. For example, using computational and corpus tools, Shih and Hsieh (2016) constructed a word dependency profile tool through automatically sketching syntagmatic relations of words in an untagged corpus based on dependency parses. The system provides a useful tool for learners of Chinese to visualize the collocation behavior of words, the knowledge of which is essential to the comprehension and production of linguistic conventions of Chinese.

The usage-based theory of language is still evolving, and the tasks of understanding the complexity of language and explicating mechanisms of language learning, which require increased data recording and increased sophistication in data analysis, are far from accomplished. Take input as an example. Researchers from various theoretical camps agree that input matters in learning a second language (Gass 1997; Piske and Young-Scholten 2008). There is the assumption that input flood, a teaching technique designed to inundate learners with massive input containing a target form, facilitates their intake of that form (Sharwood Smith 1985, 1991). However, as Madlener (2016) pointed out, there is no consensus as to exactly what kind of input structure effectively facilitates learning. Thus, everyone takes input for granted as a magic potion, yet no one is quite sure how to concoct one with a reliable measure of ingredients, and in practice, everyone makes their own from scratch, and hopes for the best. Essentially, there is great arbitrariness in how input is handled in L2 instruction. Madlener (2016) wanted to take a closer look at the structure and distributions of input and their effects on incidental learning. She designed a classroom training paradigm that manipulates input type frequency, token-type ratios, and surface structural similarity in authentic learning conditions to test the exact effects played

by these measures on learners' ability to detect and extend pattern in learning. She found "consistent effects of more fine-grained input features" in terms of token and type frequency distributions, which suggests the inadequacy of a simplistic view of input flood.

Madlener's study highlights several important points. First, it shows how specific and testable hypotheses can be formulated within the usage-based model to test SLA constructs and pedagogical strategies that are taken for granted but poorly understood at a greater granularity. Second, it shows the possibility of developing teaching methods that are theoretically grounded and empirically tested and demonstrates what it takes to get there. Third, it reminds us that a full and fine-grained understanding of the complexity of language and language learning is far from a reality. Lastly, the discrepancy between Madlener's classroom training results and those obtained in Artificial Language Learning (ALL) experiments indicates the need for the triangulation of research methods and more work on the replicability of results in usage-based approaches.

The need to drum up the effort to explore language learning at a deeper level and the criticality of methodological triangulation are being recognized. Ellis (2017: 41–42) identified three future priorities that necessitate the triangulation of evidence from the complementary areas of research in Cognitive, Corpus, and Computational Linguistics, and require more sophisticated corpus and computational methods. These are:

(1) Analyzing the distributional characteristics of linguistic constructions and their meanings in large collections of language that are representative of the language that learners experience
(2) Conducting longitudinal analyses of learner language
(3) Conducting Natural Language Processing (NLP) or computational analyses of the dimensions of language complexity.

These tasks, Ellis emphasized, require increased effort in data recording and increased sophistication in data analysis. Ke (2012, 2018) envisioned an interdisciplinary research agenda for Chinese SLA. He articulated the need for international cooperation in building large-scale corpora with "broad scope of genres, registers, styles, text types, and learner backgrounds" including learner corpora with discourse data as part of that agenda, and stressed the importance of raising the "standard of selection and utilization of statistical analysis procedures" in the field. Similarly, Zhang and Tao (2018) called for more quantitative studies on learner corpora, to be triangulated with other kinds of empirical data for corroboration and validation. Without a doubt, corpus and computational methods, with their proven empirical strengths in detecting and modeling patterns of language use and learning, are essential tools for tackling these tasks of corpora construction, data recording, and analysis. As such they are crucial voices in the professional multilingualism that invigorates SLA research and informs instructed second language learning.

5 Conclusion

As a community of practice, we need to recognize the collective journey that has taken us so far. At the same time, we need to acknowledge the persisting barriers in the continuing professionalization of our field. These include the inaccessibility of SLA research to teachers and the lack of direct pedagogic utility of isolated research findings. These problems were raised two decades ago (Crookes 1997; Ellis 1997; Markee 1997) but continue to hinder the integration of research and practice today. Given these barriers, a call for a rational approach to TCSL must not put the blame and burden firstly and solely on the teaching practitioners, and must reflect on the accessibility and utility of theory and research. More important, as a community of practice, we must evaluate the theoretical and empirical soundness of research before we jump at its pedagogical implications in order for a professional multilingualism to be productive.

Because our knowledge of L2 learning is still incomplete, more robust empirical data are needed. The twenty-first century is witnessing an unprecedented boom of digital technology, the direct impact of which can be readily seen in the methodological sophistication of quantitative research based on large bodies of corpora, including L2 corpora. Corpus and computational methods developed in Corpus Linguistics and Computational Linguistics and NLP are making inroads into the field of SLA to throw light on how language is used and learned, and on the contingencies of usage and learning. The data processing and pattern detecting power of these methods make them indispensable for SLA research. With results from this fecund area of exploration gradually entering TCSL, a field with a history of professional multilingualism, and an eagerness to explore methodological innovations, we will likely see an increased interest in learner data, an increased interest in experimenting with usage-based instruction, and an increased demand for computational technology enabled methods for error detection and correction, as well as assessment of learning results, among many other application potentials.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning, 59*(2), 249–306.

Ambridge, B., & Lieven, E. V. M. (2011). Child language acquisition: Contrasting theoretical approaches. Cambridge: Cambridge University Press.

Anderson, J. R. (2000). *Learning and memory* (2nd ed.). New York: Wiley.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*(1), 67–82.

Barcroft, J. (2013). *Input-based incremental vocabulary instruction for the L2 classroom: Innovative research and practices in second language acquisition and bilingualism*. New York: John Benjamins.

Barlow, M., & Kemmer, S. (2000). *Usage-based models of language*. Stanford, CA: CSLI Publications, Center for the Study of Language and Information.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology, 36*(3), 203–272.
Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system. Position paper. *Language Learning, 59*(S1), 1–26.
Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics, 47*(2), 383–411.
Behrens, H., & Pfänder, S. (2016). *Experience counts: Frequency effects in language*. Berlin: Walter de Gruyter.
Belzung, C., & Wigmore, P. (2013). *Neurogenesis and neural plasticity*. Dordrecht: Springer.
Blom, E., Paradis, J., & Duncan, T. (2012). Effects of input properties, vocabulary size, and L1 on the development of third person singular –s in child L2 English. *Language Learning, 62*(3), 965–994.
Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation. The case of very advanced L2 learners. In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 133–159). Mahwah, NJ: Lawrence Erlbaum.
Bruner, J. (1981). The social context of language acquisition. *Language & Communication, 1*(2), 155–178.
Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711–733.
Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
Bybee, J. (2013). Usage-based theory and exemplar representations of constructions. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 49–69). Oxford: Oxford University Press.
Bybee, J., & Hopper, P. J. (Eds.). (2001). *Frequency and the emergence of linguistic structure*. Amsterdam/Philadelphia: John Benjamins.
Byrnes, H. (1998). *Learning foreign and second languages: Perspectives in research and scholarship*. New York: Modern Language Association of America.
Byrnes, H. (2000). Shaping the discourse of a practice: The role of linguistics and psychology in language teaching and learning. *The Modern Language Journal, 84*(4), 472–494.
Byrnes, H. (2012). Advanced language proficiency. In S. M. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 506–521). London: Routledge.
Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monograph of the Society for Research in Child Development, 63*(4), Serial no. 255.
Carpenter, M., Pennington, B., & Rogers, S. (2002). Interrelations among social-cognitive skills in young children with autism. *Journal of Autism and Developmental Disorders, 32*(2), 91–106.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT Press.
Chomsky, N. (2002). *On nature and language*. Cambridge: Cambridge University Press.
Cook, G., Kasper, G., Ellis, N. C., & Larsen-Freeman, D. (2006). Language emergence. *Applied Linguistics, 27*(4), 554–740.
Crookes, G. (1997). SLA and language pedagogy. A socioeducational perspective. *Studies in Second Language Acquisition*, 19(1), 93–116.
Dąbrowska, E. (2008). The effects of frequency and neighbourhood density on adult speakers' productivity with Polish case inflections: An empirical test of usage-based approaches to morphology. *Journal of Memory and Language, 58*(4), 931–951.
DeKeyser, R. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *Handbook of second language acquisition* (pp. 313–348). Oxford: Blackwell.
Ebbinghaus, H. (1913). *Memory; a contribution to experimental psychology*. New York City: Teachers College, Columbia University.
Ellis, N. C. (1998). Emergentism, connectionism and language learning. *Language Learning, 48*(4), 631–664.
Ellis, N. C. (2008). The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *Modern Language Journal, 92*(2), 232–249.

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. *Georgetown University Round Table on Languages and Linguistics*, 35–47.
Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24*(2), 143–188.
Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164–194.
Ellis, N. C. (2013). Second language acquisition. In G. Trousdale & T. Hoffmann (Eds.), *Oxford handbook of construction grammar* (pp. 365–378). Oxford: Oxford University Press.
Ellis, N. C. (2015). Cognitive and social aspects of learning from usage. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 49–73). Berlin/Boston: De Gruyter.
Ellis, N. C. (2017). Cognition, corpora, and computing: Triangulating research in usage-based language learning. *Language Learning, 67*(S1), 40–65.
Ellis, R. (1997). *SLA research and language teaching*. Oxford: Oxford University Press.
Ellis, R. (Ed.). (2001). *Form-focused instruction and second language learning*. Malden, MA: Blackwell.
Ellis, R. (2009). Correct feedback and teacher development. *L2 Journal, 1,* 3–18.
Ellis, N. C., & Ferreira, F., Jr. (2009). Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal, 93*(3), 370–386.
Ellis, N. C., & Wulff, S. (2015). Usage-based approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed., pp. 75–93). New York: Routledge.
Ellis, N. C., Romer, U., & O'Donnell, M. B. (2016a). Constructions and usage-based approaches to language acquisition. *Language Learning, 66*(S1), 23–44.
Ellis, N. C., Romer, U., & O'Donnell, M. B. (2016b). Computational models of language usage, acquisition, and transmission. *Language Learning, 66*(S1), 241–278.
Eskildsen, S. W. (2008). Constructing another language—Usage-based linguistics in second language acquisition. *Applied Linguistics, 30*(3), 335–357.
Eskildsen, S. W. (2012). L2 Negation constructions at work. *Language Learning, 62*(2), 335–372.
Fillmore, C. J. (1992). "Corpus linguistics" or "Computer-aided armchair linguistics". In I. Svartvik (Ed.), *Directions in corpus linguistics* (pp. 35–60). Berlin: Mouton de Gruyter.
Fotos, S. (2002). Structure-based interactive task for the EFL grammar learner. In E. Hinkel & S. Fotos (Eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 135–154). Mahwah, NJ: Lawrence Erlbaum.
Gass, S. M. (1997). *Input, interaction, and the second language learner*. Mahwah, NJ: Lawrence Erlbaum.
Goldberg, A. E. (1995). *Constructions: A Construction Grammar approach to argument structure*. Chicago: University of Chicago Press.
Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Goldberg, A. E., & Casenhiser, D. (2008). Construction learning and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 197–215). New York: Routledge.
Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics, 15*(3), 289–316.
Gries, S. T. (2006). Introduction. In S. T. Gries & A. Stefanowitsch (Eds.), *Corpora in Cognitive Linguistics: Corpus-based approaches to syntax and lexis* (pp. 1–17). Berlin: de Gruyter.
Gries, S. (2009). What is corpus linguistics? *Language and Linguistics Compass, 3*(5), 1–17.
Gries, S. (2015). Statistical methods in learner corpus research. In G. Gilquin, S. Granger, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 159–181). Cambridge: Cambridge University Press.

Gries, S., & Ellis, N. (2015). Statistical measures for usage-based linguistics. *Language Learning, 65*(S1), 228–255.

Gries, S., Hampe, B., & Schönefeld, D. (2005). Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics, 16*(4), 635–676.

Han, Z.-H. (2004). *Fossilization in adult second language acquisition*. Clevedon, UK: Multilingual Matters.

Han, Z.-H. (2016). Research meets practice: Holding off and holding on. *Chinese as a Second Language, 51*(3), 236–251.

Han, Z.-H., & Odlin, T. (2006). Introduction. In Z.-H. Han & T. Odlin (Eds.), *Studies of fossilization in second language acquisition* (pp. 1–20). Clevedon, UK: Multilingual Matters.

Hofman, M. (2002). Plasticity in the adult brain: From genes to neurotherapy. In *Proceedings of the 22nd International Summer School of Brain Research*. Amsterdam/Boston: Elsevier.

Hoshino, A., & Yasuda, A. (2013). Automatic discrimination of pronunciations of Chinese retroflex and dental affricates. In M. Sun, M. Zhang, D. Lin, & H. Wang (Eds.), *Chinese computational linguistics and natural language processing based on naturally annotated big data* (pp. 303–314). Berlin/Heidelberg: Springer.

Hsiao, H.-S., Chang, C.-S., Lin, C.-Y., Chen, B., Wu, C.-H., & Lin, C.-Y. (2016). The development and evaluation of listening and speaking diagnosis and remedial teaching system. *British Journal of Educational Technology, 47*(2), 372–389.

Huang, C.-R., & Lenders, W. (2005). Computational linguistics and beyond: An introduction. In C.-R. Huang & W. Lenders (Eds.), *Computational linguistics and beyond* (pp. 1–16). Taipei: Academia Sinica.

Ibbotson, P. (2013). The scope of usage-based theory. *Frontiers in Psychology, 4,* 255.

Jiao, L., Kubler, C. C., & Zhang, W. (2011). *500 common Chinese idioms: An annotated frequency dictionary*. London/New York: Routledge.

Jing-Schmidt, Z. (2011). Zero anaphora in higher level Chinese writings across learner backgrounds. *Chinese Teaching in the World, 25*(2), 258–267.

Jing-Schmidt, Z. (2015). The place of linguistics in CSL teaching and teacher education: Toward a usage-based constructionist theoretical orientation. *Journal of Chinese Language Teachers Association, 50*(3), 1–22.

Jing-Schmidt, Z., & Peng, X. (2018). Linguistics theories and teaching Chinese as a second language. In C. Ke (Ed.), *The Routledge handbook of Chinese second language acquisition* (pp. 63–81). London: Routledge.

Jing-Schmidt, Z., Peng, X., & Chen, J.-Y. (2015). From corpus analysis to grammar instruction: Toward a usage-based constructionist approach to constructional stratification. *Journal of Chinese Language Teachers Association, 50*(2), 109–138.

Jurafsky, D., & Martin, J. (2008). *Natural language processing, speech recognition, and computational linguistics* (2nd ed.). Upper Saddle River, NJ: Prentice-Hall.

Kapatsinski, V. (2014). What is a grammar like? A usage-based constructionist perspective. *Linguistic Issues in Language Technology, 11*(1), 1–41.

Ke, C. (2012). Research in second language acquisition of Chinese: Where we are, where we are going. *Journal of the Chinese Language Teachers Association, 47*(3), 43–113.

Ke, C. (2018). Chinese SLA: Introduction and future directions. In C. Ke (Ed.), *The Routledge handbook of Chinese second language acquisition* (pp. 1–8). London: Routledge.

Ke, C., Lu, Y., & Pan, X. (2015). 汉语教师教学技能及二语习得理论知识的评估模式. (Assessing international Chinese language teachers' second language acquisition theoretical foundation and language pedagogy). *《世界汉语教学》* (*Chinese Teaching in the World*), *29*(1), 111–129.

Kramsch, C. (2000). Second language acquisition, applied linguistics, and the teaching of foreign languages. *Modern Language Journal, 84*(3), 311–326.

Larsen-Freeman, D. (2011). A complexity theory approach to second language development/acquisition. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 47–72). London: Routledge.

Li, X. (2014). Variation in subject pronominal expression in L2 Chinese. *Studies in Second Language Acquisition, 36*(1), 39–68.
Lieven, E. V. M., Pine, J. M., & Baldwin, G. (1997). Lexically-based learning and early grammatical development. *Journal of Child Language, 24*, 187–219.
Lyster, R., & Mori, H. (2006). Interactional feedback and instructional counterbalance. *Studies in Second Language Acquisition, 28*(2), 269–300.
Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake: negotiation of form in communicative classrooms. *Studies in Second Language Acquisition, 19*(1), 37–66.
MacWhinney, B. (Ed.). (1999). *The emergence of language*. Mahwah, NJ: Lawrence Erlbaum Associates.
Macwhinney, B. (2005a). The emergence of linguistic form in time. *Connection Science, 17*(3–4), 191–211.
MacWhinney, B. (2005b). Extending the competition model. *International Journal of Bilingualism, 9*(1), 69–84.
MacWhinney, B. (2006). Emergent fossilization. In Z.-H. Han & T. Odlin (Eds.), *Studies of fossilization in second language acquisition* (pp. 134–156). Clevedon, UK: Multilingual Matters.
MacWhinney, B. (2012). The logic of the Unified Model. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 211–227). New York, NY: Routledge.
MacWhinney, B. (2015). Multidimensional SLA. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 19–48). Berlin: De Gruyter.
Madlener, K. (2016). Input optimization: effects of type and token frequency manipulation in instructed second language learning. In H. Behrens & S. Pfänder (Eds.), *Experience counts: Frequency effects in language* (pp. 133–174). Berlin: De Gruyter.
Markee, N. (1997). Second language acquisition research: A resource for changing teachers' professional cultures? *The Modern Language Journal, 81*(1), 80–93.
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*(3), 207–238.
Medin, D. L., & Smith, E. E. (1981). Strategies in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7*(4), 241–253.
Nosofsky, R. (1988). On exemplar-based exemplar representations: Reply to Ennis (1988). *Journal of Experimental Psychology: General, 117*(4), 412–414.
Nugent, M., & Slater, R. (2016). The language flagship: Creating expectations and opportunities for professional-level language learning in undergraduate education. In D. Murphy & K. Evans-Romaine (Eds.), *The U.S. language flagship program: Professional competence in a second language by graduation* (pp. 1–9). Clevedon, UK: Multilingual Matters.
Peterson, M. (2006). Learner interaction management in an avatar and chat-based virtual world. *Computer Assisted Language Learning, 19*(1), 79–103.
Peterson, M. (2010). Computerized games and simulations in computer-assisted language learning: A meta-analysis of research. *Simulation and Gaming: An Interdisciplinary Journal, 41*(1), 72–93.
Pine, J. M., Lieven, E. V. M., & Rowland, C. F. (1998). Comparing different models of the development of the English verb category. *Linguistics, 36*, 807–830.
Pinker, S. (1999). *Words and rules*. New York: Harper Perennial.
Piske, T., & Young-Scholten, M. (2008). *Input matters in SLA*. Clevedon: Channel View Publications.
Quine, W. (1960). *Word and object*. Cambridge, MA: MIT Press.
Reali, F. (2014). Frequency affects object relative clause processing: Some evidence in favor of usage-based accounts. *Language Learning, 64*(3), 685–714.
Rebuschat, P., & Williams, J. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33*(4), 829–856.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605.

Rossano, F., Fiedler, L., Tomasello, M., & Eccles, J. S. (2015). Preschoolers' understanding of the role of communication and cooperation in establishing property rights. *Developmental Psychology, 51*(2), 176–184.
Sale, A. (2016). *Environmental experience and plasticity of the developing brain*. Hoboken, NJ: Wiley/Blackwell.
Sato, M., & Lyster, R. (2012). Peer interaction and corrective feedback for accuracy and fluency development. *Studies in Second Language Acquisition, 34*(4), 591–626.
Schachter, J. (1989). Testing a proposed universal. In S. Gass & J. Schachter (Eds.), *Linguistic perspectives on second language acquisition* (pp. 73–88). Cambridge: Cambridge University Press.
Schachter, J. (1990). On the issue of completeness in second language acquisition. *Second Language Research, 6*(2), 93–124.
Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129–158.
Schmidt, R. (1993). Awareness and second language acquisition. *Annual Review of Applied Linguistics, 13,* 206–226.
Schulze, C., & Tomasello, M. (2015). 18-month-olds comprehend indirect communicative acts. *Cognition, 136,* 91–98.
Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*(2), 209–231.
Selinker, L., & Lamendella, J. (1978). Two perspectives on fossilization in interlanguage learning. *Interlanguage Studies Bulletin, 3*(2), 143–191.
Sharwood Smith, M. (1985). From input to intake: On argumentation in second language acquisition. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 394–403). Rowley, MA: Newbury House.
Sharwood Smith, M. (1991). Speaking to many minds: On the relevance of different types of language information on the L2 learner. *Second Language Research, 7*(2), 118–132.
Shih, M.-H., & Hsieh, S.-K. (2016). Word dependency sketch for Chinese language learning. *Concentric, 42*(1), 45–72.
Skehan, P. (1998). Task-based instruction. *Annual Review of Applied Linguistics, 18,* 268–286.
Smith, E. R. (1991). Illusory correlation in a simulated exemplar-based memory. *Journal of Experimental Social Psychology, 27*(2), 107–123.
Smith, E. R., & Zarate, M. A. (1992). Exemplar-based model of social judgment. *Psychological Review, 99*(1), 3–21.
Sockett, G., & Kusyk, M. (2015). Online informal learning of English: Frequency effects in the uptake of chunks of language from participation in web-based activities. In T. Cadierno & S. W. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 153–177). Berlin: De Gruyter Mouton.
Stubbs, M. (2001). Texts, corpora, and problems of interpretation: A response to Widdowson. *Applied Linguistics, 22*(2), 149–172.
Tao, H. (Ed.). (2016). *Integrating Chinese linguistic research and language teaching and learning*. Amsterdam: John Benjamins.
The Douglas Fir Group. (2016). A transdisciplinary framework for SLA in a multilingual world. *Modern Language Journal, 100*(S1), 19–47.
Tomasello, M. (1992). *First verbs: A case study of early grammatical development*. Cambridge/New York: Cambridge University Press.
Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, MA: Harvard University Press.
Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics, 11*(1–2), 61–82.
Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
Tomasello, M. (2008). *Origins of human communication*. Cambridge, MA: MIT Press.

Tomasello, M., & Farrar, M. J. (1986). Joint attention and early language. *Child Development, 57*(6), 1454–1463.

Tragant, E., & Muñoz, C. (2004). Second Language acquisition and language teaching. *International Journal of English Studies, 4*(1), 197–219.

Tummers, J., Heylen, K., & Geeraerts, D. (2005). Usage-based approaches in Cognitive Linguistics: A technical state of the art. *Corpus Linguistics and Linguistic Theory, 1*(2), 225–261.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing. *Language Learning, 62*(1), 1–41.

Verspoor, M., & Nguyen, T. P. H. (2015). A Dynamic usage-based approach to second language teaching. In T. Cadierno & S. Eskildsen (Eds.), *Usage-based perspectives on second language learning* (pp. 305–328). Berlin: De Gruyter.

Winke, P. (2013). An investigation into second language aptitude for advanced Chinese language learning. *Modern Language Journal, 97*(1), 109–130.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition, 35*(3), 451–482.

Wulff, S. (2010). *Rethinking idiomaticity: A usage-based approach*. London: Continuum.

Xiao, R., Rayson, P., & McEnery, T. (2009). *A frequency dictionary of Mandarin Chinese: Core vocabulary for learners*. London: Routledge.

Xu, Q. (2016). Item-based foreign language learning of give ditransitive constructions: Evidence from corpus research. *System, 63,* 65–76.

Zhang, J., & Lu, X. (2013). Variability in Chinese as a foreign language learners' development of the Chinese numeral classifier system. *Modern Language Journal, 97*(S1), 46–60.

Zhang, J., & Tao, H. (2018). Corpus Linguistics and Chinese SLA. In C. Ke (Ed.), *The Routledge handbook of Chinese second language acquisition* (pp. 48–63). London: Routledge.