Nick C. Ellis (University of Michigan, Ann Arbor)

Frequency in language learning and language change

The contributions to this volume from a cognitive and psycholinguistic perspective

The beginnings of cognitive linguistics made much of categorization, imagery, and metaphor, but less of psycholinguistics, quantitative modeling, or corpus linguistics. Despite a "cognitive commitment to make one's account of human language accord with what is generally known about the mind and the brain" (Lakoff 1990: 40), much of the psychology of early cognitive linguistics relied upon introspection; there was little by way of corpus analyses; statistical and computational modeling featured hardly at all. This is ironic given that most of psychological processing involves unconscious processing (Nisbett and Wilson 1977; Reber 1993) and that learning involves statistical inference of systematicities in the problem space (McClelland and Rumelhart 1988; Rebuschat and Williams 2012). This is not to deny the successes of cognitive/usage- based linguistics: indeed it is currently the best game in town.

Later, Sanders (1997) introduced psycholinguistics and the study of discourse and corpora to cognitive linguistics, while Tomasello (1998) did the reverse, promoting cognitive linguistics to the post-Skinnerian and subsequently post-Chomskian psychological world as "The New Psychology of Language." Today, there is much cross- influence in the study of child language acquisition where usage-based approaches are common (Tomasello 2003). Nevertheless, one can still see divisions between cognitive linguistics, corpus linguistics, psycholinguistics, the psychology of learning, connectionism, NLP, computer modeling, language change, bilingualism research, applied linguistics, sociolinguistics, and the study of language change, despite the fact that we are all fundamentally study-ing the effects of language usage (on language perception, learning, knowledge, processing, structure, and change).

However, the articles in the current volume are different. They are impres-sively cutting-edge inter-disciplinary. They start from the phenomena of frequency effects in language, and then try to understand the relevant processes and outcomes. This approach has been successful before (Bod, Hay, and Jannedy 2003; Bybee and Hopper 2001; Ellis 2002). But one difference here is that, as with pidgin and creolization, with the next generation learning the language, the effects are more powerful and systemic. Research training group GRK DFG 1624/1

"Frequency effects in language: Frequency as a determinant in usage-based models of language change, language processing and language acquisition", following the broader goals of the Freiburg Institute for Advanced Studies, set out to bring together interdisciplinary expertise relevant to these phenomena, to train cohorts of graduate students in the relevant techniques, to engender discussion across disciplines, thus to encourage peer tutoring and constructive criticism. The results are impressive. The particular strength of the Freiburg group are its command of a broad range of methods, incorporating experimental and analytic procedures from corpus-linguistics, sociolinguistics, psycholinguistics and cognitive psychology, and the fact that the empirical data investigated are drawn from a wide range of European languages.

In what follows, I will not repeat the interdisciplinary background as described in the fine introductory overview chapter (Pfänder and Behrens, this volume), but instead I will react briefly to each of the focused empirical chapters, considering what they teach us and trying to gather the next priorities for research.

Diessel, a pioneer of frequency effects in construction grammar acquisition (Diessel 2007), provides a critical review of frequency and lexical specificity in grammar. He argues that the 'bare' exemplar model is not sufficient to explain how lexemes and constructions are related, and that there are other cognitive mechanisms, notably automatization, that create lexical effects in grammar. Language users' knowledge about constructions includes a great deal of lexical information: the lexical biases of the ditransitive (and other verb-argument constructions), for example, are not entirely predictable from semantic and pragmatic criteria but involve a degree of conventionalization. People 'know' the lexical patterns of the ditransitive and other verb-argument constructions from their linguistic experience and this knowledge affects their behavior in sentence comprehension and production. Thus a construction is both paradigmatically and syntagmatically defined. This view has been independently expressed from within corpus linguistics (Gries and Stefanowitsch 2004; Sinclair 1996, 2004), phraseological research (Ellis 2012a; Ellis, O'Donnell, and Römer 2014), and probabilistic data-oriented parsing (Bod 2009; Bod et al. 2003). Frequency effects occur between and across all levels of the form-function structure of constructions (Goldberg 2006) as "thought-sound" (de Saussure 1916): there is chunking and automatization within "sound" (Ellis 1996), within "thought" (Rumelhart, Smolensky, McClelland, and Hinton 1988), and, with varying degrees of contingency, across "thought-sound" (MacWhinney, Bates, and Kliegl 1984). Associative learning theory assesses this using measures of contingency (Shanks 2010), corpus linguistics does the same using collostructional analysis (Gries 2012). In language learning, we compute associations within and across modalities (Ellis 2015); learning constructions is a function of form type- and token- frequency distribution, meaning prototypical-

ity, form-function contingency, and form collocational chunks (Ellis 2012b; Ellis and Ferreira-Junior 2009a,b). In language processing, as Diessel demonstrates, it is the satisfaction of all of these constraints which provides the most likely interpretation.

Hakimov uses corpus analyses and statistical modeling to investigate whether the marking of plural code-mixed nouns retains the plural morphology of their German source language or receives plural markers from the Russian matrix language. The choice of the language for plural marking on the code-mixed nouns was found to be determined by frequency factors in both the source and matrix languages: (1) The frequencies of the plural forms of the inserted lexical items in the source language influence marking so that lexical items commonly used as plurals tend to be selected as a whole and become inserted into the matrix clause retaining their German plural markers. However, for lexical items where the singular representation is stronger, only the stem of the lexical item is inserted and it then receives the plural marker from the matrix language. (2) The structural requirements imposed on these items by the matrix language are also influential: (a) German lexical items with an accented vowel in the stem-final position cannot take Russian inflectional suffixes directly as the Russian declensional system depends on stems with consonants in the final position. If the embedded German noun has an accented vowel in the stem-final position, either a compromise strategy is employed, such as the use of epenthetic consonants, or German plural forms are produced. (b) When the matrix structure projects an oblique case on the slot in which a German lexical item is inserted, the tendency is towards using Russian inflections, fusing plural and case.

Language processing involves the competition of many cues and their associations and the satisfaction of these multiple constraints (MacDonald and Seidenberg 2006; MacWhinney 2008; Pearl and Goldwater, in press). Usage-based theory ultimately aims to provide a cognitively real account of language processing and organization, allowing for the integration of multiple kinds of factors – those based on discreet categories, such as the presence or the absence of a vowel in the stem-final position, and factors distinguished by gradience, such as frequency of use. Hakimov demonstrates the interaction of these factors as probabilistic tendencies.

Schneider investigates speech hesitations from a usage-based perspective, analyzing the placement of hesitations in prepositional phrase contexts in the Switchboard NXT corpus of American English. She shows that cohesive sequences that are used frequently are less likely to be interrupted by hesitations than low frequency word pairs, and that more complex measures of the associations between words not only predict hesitation behavior, but model semantic unity as well.

Semantic coherence is reflected in the statistical associations between words: the closer a word pair is to representing a single semantic concept, the more the words in the pair attract each other. This means that an increase in semantic cohesiveness comes with a greater likelihood of co- occurrence. Constructions are both syntagmatically and paradigmatically associated. The more we think of a concept, the more we activate routinized chunks for expressing it; the more we use these chunks, the greater their association. Sinclair came to this phraseological view of language from his corpus-based analyses of text, but paradoxically so, since he would only trust the text – given a spoken corpus he would transcribe it and analyze the transcription (Sinclair and Mauranen 2006). Schneider demonstrates that physical speech has still more evidence of these routine effects of language processing and usage on the chunking of language structure. Her triangulation of different corpus-based measures of association against measures of psychological processing exemplifies an important new trend which shows the true collaborative potential of corpus-based and psycholinguistic approaches to usage (Ellis, Simpson-Vlach, and Maynard 2008; Gries and Divjak 2012; Gries and Wulff 2009; Jaeger 2010; Simpson-Vlach and Ellis 2010; Smith and Levy 2013).

MADLENER investigated the effects of type and token frequency in instructed Second Language Learning involving a two-week training study of adult L2 learners of German learning the sein *be* + present participle construction presented in structured input floods in meaning-focused listening comprehension training. The study is important in that it tests the generalizability of prior findings from laboratory-based artificial language learning experiments which proposed that learning would be affected by type- token frequency distributions in the input (Bybee 2008; Ellis 2008b; Goldberg 2008). She found that increased surface similarity due to reduced type frequency of the target construction in the input floods (9 compared to 25 or 50 different target types during training) resulted in significantly better target availability for learners and enhanced their pattern generalization ability. Furthermore, learners exposed to skewed input with 25 types substantially outperformed all other learners, whilst skewed input with a low number of types was non-optimal. Madlener argues that skewed input provides an optimal balance of surface similarity and type variation. When high-frequency prototypical exemplars occur often and in varying contexts, they are registered and mapped to their relevant communicative meaning, and numerous subsequent encounters lead them to be increasingly well entrenched as prefabs in the learner's mind. These then act as anchors for the entrenchment of the more abstract pattern in the learner's emerging interlanguage system. Being salient and strongly entrenched, the central exemplars serve as easily retrievable models for analogical pattern extension from early on. Subsequent additional type variability at the crucial slot in the more peripheral low-frequency exem-

plars then triggers the formation of an abstract, item-general schema. As the learner regularly encounters unfamiliar, low-frequency exemplars in the input, there is then generalization to further unfamiliar cases. Madlener considers the implications of these findings for instructed adult L2 acquisition involving incidental grammar learning from meaning-based input processing and how textbooks for German as a second language could provide input floods in a more principled way in order to optimize incidental grammar learning: moderately skewed input with substantial type variation should be the preferred choice for structured input floods. It is a long way from the laboratory experiment to the language classroom, too far an extrapolation to assume. Usage-based theory has important consequences for effective instruction (Littlemore 2009; Tyler 2012). Madlener's work itself serves as a prototypical exemplar in the steps necessary for ecological validation, generalization and development of theory and practice.

ROSEMEYER, like Madlener, also considers the effects of type and token frequency, but here in processes of language change where a grammatical construction decreases in usage. His diachronic analysis of Spanish compound tense auxiliary selection (*haber* + PtcP vs. *ser* + PtcP) argues that the replacement of *ser* with *haber* was affected by (a) the salience of the use of *haber* + PtcP in usage contexts previously associated with the use of *ser* + PtcP, and (b) the token frequency of the *ser* + PtcP syntagms formed from specific verbs. His statistical modeling of these frequency effects in language change suggests both a type frequency effect (the prototypicality of use of *ser* + PtcP with telic change of state predicates drove their replacement by *haber* + PtcP), as well as a protecting entrenchment token frequency effect (*ser* + PtcP syntagms formed from frequent verbs were more protected from the ongoing change, with the skewed frequency distribution in this verb population then leading to further changes in auxiliary selection rules). Rosemeyer's work illustrates very clearly that a vague, under-specified acknowledgment of frequency effects is not sufficient. Modeling the effects of frequency in language change presupposes a very explicit model of the distribution of the function range, i.e., the type frequency of the disappearing construction in earlier stages of the grammatical change (Bybee 2010). The study of actualization then focuses upon changes in the type frequency of the construction, the study of conservation investigates differences in the speed with which syntagms instantiating that construction decline in token frequency. Rosemeyer's work exemplifies how different statistical methods can illuminate these separate processes.

SCHWARZ considers a "sequential" kind of frequency effect: The impact of recency on phonological variation. Recency or priming effects state that in a situation where two variants alternate, a variant is more likely to appear if it has been preceeded by the same variant shortly before. Schwarz investigates this

phenomenom with regard to phonological variation between standard and non-standard phonological vowel variants in South-West German Alemannic dialects. Drawing on a large data-base of spontaneous dialect speech from the area, he shows that a standard vowel realization is more likely to occur if preceded by a standard realization as opposed to a non-standard realization. The same applies for non-standard variants. In addition, he shows that recency effects are stronger for the variant that is relatively less frequent: The less frequent the dialect variant is relatively to the standard variant, the stronger its recency/priming effect. According to Schwarz, the term "recency" refers to "the tendency of a speaker to repeat identical phonological items within a speaking sequence, with the tendency being stronger the less time between two utterances has passed. Thus, recency can be understood as a sequential frequency effect of identical phonological items. This means that recency cannot be understood as frequency in a strict sense, i.e. how often a certain (phonological) linguistic event occurs. Rather, it relates to the time span that lies between two linguistic events and the cohesion between them". As Schwarz stresses in his contribution, recency does not explain why an innovative phonological item is taken into the repertoire of a dialect speaker, but does account for its spread.

Just as recency, salience may be an important factor which needs to be taken into account if we investigate frequency.

Contemporary learning theory holds that learning is driven by prediction errors. We learn more from the surprise that comes when our predictions are incorrect than when our predictions are confirmed (A. Clark 2013; Rescorla and Wagner 1972; Rumelhart, Hinton, and Williams 1986; Wills 2009). In addition, there is increasing evidence for surprisal-driven language processing and acquisition (Dell and Chang 2013; Demberg and Keller 2008; Jaeger and Snider 2013; Pickering and Garrod 2013; Smith and Levy 2013). Adaptation allows comprehenders' expectations to converge towards the actual statistics of the linguistic input (Jaeger and Snider 2013). Fine and Jaeger (2013) show this adaptation can take place very rapidly and can be tagged to markedly salient contexts (such as a laboratory setting), indeed readers can adjust their linguistic expectations in light of recent experience such that (a) previously difficult structures become easier to process, and, even more strikingly, (b) previously easy to process structures come to incur a processing cost. Warker and Dell (2006) show how the statistics of the linguistic input becomes incorporated into production as a mini-grammar, hierarchically subsumed within the main population parameters, with its own local parametric interactions tied to a particularly salient context (again, a laboratory setting) (see also Dell and Chang 2013).

However, if implicit learning from surprisal / prediction error was all that was needed for language acquisition, given enough input, we would all learn

second languages and be able to mimic second dialects just as well as we learn our native ones. Typically we don't. Some second language learners can spend their lives immersed in the target language, and yet despite this, not all of it 'goes in', and L2A is typically less successful than L1A (Han and Odlin 2006; Schmidt 1984). This is Corder's distinction between input, the available target language, and intake, that subset of input that actually gets in and which the learner utilizes in some way (Corder 1967). Associative L2 learning from naturalistic usage typically falls far short of a native-like endstate, often stabilizing at a 'Basic Variety' of interlanguage which, although sufficient for everyday communicative purposes, predominantly comprises just nouns, verbs and adverbs, with little or no functional inflection and with closed-class items, in particular determiners, subordinating elements, and prepositions, being rare, if present at all (Klein 1998). Nativelike second language accent is perhaps the most difficult challenge (Abrahamsson and Hyltenstam 2009). Given that these learners have a large enough sample of input to statistically learn from, what is it that limits their intake?

Typically the aspects of a second language that are less readily acquired are those that are less physically salient in the input or consequential in processing (Goldschneider and DeKeyser 2001). The fragile features of L2 acquisition are those which, however available, fall short of intake because of one of the factors of salience, interference, overshadowing, blocking, contingency, cue competition, or perceptual learning, all shaped by the L1 (Ellis 2006b). Salient items or features are attended to, are more likely to be perceived, and more likely to enter into subsequent cognitive processing and learning. Salience can independently be physically and psychologically determined:

- *Physical salience*: Salience arises in sensory data from contrasts between items and their context. Loud noises, bright lights, and moving stimuli capture our attention. These stimuli deliver intense signals in the psychophysics of our data-driven perception. Stimuli with unique features compared to their neighbors (Os in a field of Ts, a red poppy in a field of yellow), "pop out" from the scene, but in a shared feature context do not (Os among Qs). These are aspects of bottom-up processing.
- *Psychological salience*: Emotional, cognitive, and *motivational* factors affect the salience of stimuli too. A loved one stands out from the crowd, as does a stimulus with weighty associations ($500000.0 vs. $0.000005), or one which matches a motivational state (a meal when hungry but not when full).

These are separable constructs from statistical surprisal.

Salience is also relevant to the social context. Learners can clearly pick up that different language statistics hold in a psychology lab than on the street. But what other aspects of social variation (like the social background, age, or

gender of the speaker) are salient enough to be taggable? There are an infinite number of associations that might be made to any language event: the gender of the speaker, their age, their mood, their eye color, their tie, the stain on their tie...?, the national context, the state, the town, the building, the room, the wall color...? Physical salience is one part, so too might be the consequences of not learning the association, consequences which are defined in the social context. I will return to this theme in my conclusion. Physical salience, psychological salience, attention and learned attention are important factors in the associative learning of language that are additional to frequency and contingency (Ellis 2006b, 2008d; Schmidt 2001).

RUETTE, EHRET and SZMRECSANYI apply usage-based linguistics and state-of-the-art aggregation methodologies to explore methods for quantifying the lexical distance between different varieties of English. Different varieties use different labels to encode the same concept (*trunk* vs. *boot*, *fender* vs. *bumper*, *soda* vs. *pop*, etc.). The more numerous such variables, the greater the lexical distance between the varieties. Ruette et al. then compare different varieties of written English aggregating across these variables to measure distance. When aggregating across variables, Ruette et al. quite ingeniously apply weighting functions which manipulate the impact of particular lexical variables. They determine if different types of frequency weighting have an effect on the lectal distances calculated , including (i) boosting low-frequency concepts, (ii) boosting high-frequency concepts, and (iii) no frequency weighting at all. Despite their allowance for contrasts to emphasize salience or frequency, they failed to obtain any such effect, and they discuss two types of reasons for the absence of frequency effects in lexical sociolectometry. The first is in terms of methodology: the qualitative configuration of aggregate lexical distances is very similar no matter which weighting scheme is used because the underlying distance matrices are very highly correlated (indeed generally, they claim, aggregation studies tend to not be "overwhelmed by noise"). Their second hypothesis is that studies of frequency effects typically center upon grammar, and they speculate that perhaps entrenchment, neuromotor patterns, and similar patterns play a less decisive role in the structure of the lexicon. Their subsequent consideration that concept and/or synonym frequencies exhibit an extreme sensitivity to lectal dimensions, such as register, leads them to discount this speculation. Frequency effects pervade all aspects of language and usage, including the lexicon (Diependale, Lemhöfer, and Brysbaert 2013; Ellis 2002). A third alternative is that frequency, salience, and surprisal all play a part. The power law of practice summarizes how the greater the frequency of experience, the greater the expertise, with the incremental effects of experience eventually diminishing and reaching asymptote. The Rescorla-Wagner formula summarizing the determinants of learning show how salience

and surprisal play their roles on each and every trial of experience (Ellis 2006a,b). The fact that frequency, surprisal, and salience all feature in learning suggests that we should include all three factors in our models and that we should study longitudinal development too – they will play different roles at different points in learning. The study by Ruette et al. has presented important and original questions concerning usage, variability, and variety, and it provides important developments in quantitative lexical sociolectometry.

These are fine papers. They all extend our knowledge of the effects of frequency of usage in language. They prompt one or two other interdisciplinary associations too.

Both in language variation and change, but not less so in language acquisition of children and adults, we continually discuss the nature of abstraction in language and pit against each other the exemplar-based vs. abstract prototype, and the richness of the exemplar details and their associations vs. more connectionist learning mechanisms which tune the features but throw away the exemplars. Are there units of language that are abstracted from language exemplars so that they have an independent status as units? There is much work to be done here. In the interim, I am persuaded by accounts that acknowledge both (Pierrehumbert 2006). But the empirical work should not take place within cognitive or corpus linguistics alone. The issue relates across cognition and is informed by categorization theory (Rips, Smith, and Medin 2012), memory and neuropsychology relating to the dissociations between episodic vs. abstract memory systems (Nadel and Moscovitch 1998), implicit vs. explicit learning (Ellis 1994; Rebuschat 2015), dual memory systems and avoidance of catastrophic interference (McClelland 1998), as well as the consolidation of new episodic memories and their integration into a representational system (Davis and Gaskell 2009; Gaskell et al. 2012).

Much of the work of frequency tuning is in terms of automatic implicit learning or tallying (Ellis 2002). But the limitations of second-language learning show us that this tallying is by no means guaranteed. Effects of salience and learned attention entail that, broadly, it is not until a representation has been noticed and consolidated, that the strength of that representation can thereafter be tuned implicitly during subsequent processing (Ellis 2006a,b). There is thus a strong role for consciousness in language learning (Ellis 2005).

Both conscious and unconscious processes affect the dance of dialogue in the way in which native conversation partners align perspectives and means of linguistic expression (Garrod and Pickering 2004; Pickering and Garrod 2004). The predictive error signals that we receive, notice, and learn from are socially negotiated. Conversation partners scaffold and co-construct meanings (H. H. Clark 1996). Socially scaffolded 'noticing' (Schmidt 1990) solves Quine's

problem of 'referential indeterminacy' and builds so much more. The dynamics of language learning are inextricably linked to the dynamics of consciousness, in neural activity and in the social world as well (Frith and Frith 2010). Input to the associative network is gated by consciousness, and consciousness is co-constructed in social interaction (Ellis 2005; C. Frith 2010).

Language and language learning take place in social usage, involving action, reaction, collaborative interaction, intersubjectivity, and mutually assisted performance (Lantolf and Thorne 2006). Speech, speakers, identity, and social relationships are inseparable (Lee et al. 2009; Norton 1997; Tarone 2007).

Activity theory emphasizes how individual learning is an emergent, holistic property of a dynamic system comprising many influences, both social and individual, as well as contextual. Action provides a context within which the individual and society, mental functioning and sociocultural context can be understood as interrelated moments which all play a role in what we call experience (Wertsch 1998).

The associative learning network is culturally gated. Tomasello's constructionist approach to language combines with his research in comparative primate cognition, the unique place of social cooperation in humans, and the Vygotskian intelligence hypothesis whereby regular participation in cooperative, cultural interactions during ontogeny leads children to construct uniquely powerful forms of perspectival cognitive representation including language itself (Moll and Tomasello 2007). The last 40 years have seen huge progress in research into social cognition. Within social cognitive neuroscience there is now a rich understanding of (a) the role of implicit and explicit knowledge in social cognition (Frith and Frith 2008), of (b) consciousness and meta- cognition for social interaction (C. Frith 2010), and (c) of the brain mechanisms involved in these processes (Frith and Frith 2010).

How then should we research language usage, change and acquisition from here on? If experience in terms of language learning is in the social cognitive linguistic moment, we need to capture all these moments, so that we can objectively study them. We need large, dense longitudinal corpora of language use, complete with audio, video, transcriptions and multiple layers of annotation, fit to be shared in open archives. We need these in a sufficiently dense mass so that we can chart learners' usage history and their development (Behrens 2008, 2009; Tomasello and Stahl 2004).

We need them in sufficient detail that we can get down to the fine detail of CA analyses of the moment (Kasper and Wagner 2011; Markee 2008). MacWhinney has long been working towards these ends, first with CHILDES (MacWhinney 1991), then with Talkbank (MacWhinney 2007). These projects have developed variously CLAN tools for computer analyses of large bodies of data, right down

to, in collaborations with Wagner, tools for a fine-grained CABank (MacWhinney and Wagner 2010). With these types of data, we can study the cognitive alongside the social. This way, the future lies. We need much more, especially for SLA. Big, dense, longitudinal datasets that we can all observe. It is up to the field to contribute to these open archives.

So we indeed started with frequency; expanded frequency to consider type and token frequencies of forms, functions and their contingencies, quasi-regular domains, and neighborhood densities. In a next step we made a point for the well known trilogy of "frequency, recency, and context" as the driving forces of rational information organization for optimal retrieval (Anderson 1991; Anderson and Schooler 2000) and added salience and/or surprisal. Finally, we also incorporated transitional probabilities, entrenchment, exemplars and representation.

Cognition, consciousness, experience, embodiment, brain, self, and human interaction, society, culture, and history are all inextricably intertwined in rich, complex, and dynamic ways in language usage. We need theoretical integrative theories. We require additional perspectives on dynamic interactions between levels, perspectives provided by approaches such as Complex Adaptive Systems (Ellis and Larsen-Freeman 2009; Beckner et al. 2009), Dynamic Systems Theory (de Bot, Lowie, and Verspoor 2007; Ellis 2008a; Spivey 2006), and Emergentism (Ellis 1998; Ellis and Larsen-Freeman 2006; MacWhinney 1999).

References

Abrahamsson, Niclas and Kenneth Hyltenstam 2009: Age of onset and nativelikeness in a second language: Listener perception versus linguistic scrutiny. *Language Learning* 59: 249–306.

Anderson, John R. 1991: The adaptive nature of human categorization. *Psychological Review* 98(3): 409–429.

Anderson, John R. and Lael J. Schooler 2000: The adaptive nature of memory. In: Endel Tulving and Fergus I. M. Craik (eds.), *The Oxford Handbook of Memory*, 557–570. London: Oxford University Press.

Beckner, Clay, Richard Blythe, Joan L. Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann 2009: Language is a complex adaptive system. Position paper. *Language Learning* 59 (Supplement 1): 1–26.

Behrens, Heike 2009a: Usage-based and emergentist approaches to language acquisition. *Linguistics* 47: 383–411.

Behrens, Heike (ed.) 2008: *Corpora in Language Acquisition Research: Finding Structure in Data*. Amsterdam/Philadelphia: Benjamins.

Bod, Rens 2009: From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science* 33: 752–793.

Bod, Rens, Jennifer Hay, and Stefanie Jannedy (eds.) 2003: *Probabilistic Linguistics*. Cambridge, MA: MIT Press.

Bybee, Joan L. 2008: Usage-based grammar and second language acquisition. In: Peter Robinson and Nick C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, 216–236. London: Routledge.

Bybee, Joan L. 2010: *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.

Bybee, Joan L. and Paul J. Hopper (eds.): 2001 *Frequency and the Emergence of Linguistic Structure*. Amsterdam/Philadelphia: Benjamins.

Clark, Andy 2013: Whatever next? Predictive brains, situated agents, and the future of Cognitive Science. *Behavioral and Brain Sciences* 36(3): 181–204.

Clark, Herbert H. 1996: *Using Language*. Cambridge: Cambridge University Press.

Corder, Stephen Pit 1967: The significance of learners' errors. *International Review of Applied Linguistics* 5: 161–169.

Davis, Matthew H. and M. Gareth Gaskell 2009: A complementary systems account of word learning: Neural and behavioural evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364: 3773–3800.

de Bot, Kees, Wander Lowie, and Marjolijn Verspoor 2007: A dynamic systems theory to second language acquisition. *Bilingualism: Language and Cognition* 10: 7–21.

de Saussure, Ferdinand 1916: *Cours de linguistique générale* (Roy Harris, Trans.). London: Duckworth.

Dell, Gary S. and Franklin Chang 2013: The P-Chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: X-X. Demberg, Vera and Frank Keller 2008: Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109: 193–210.

Diependale, Kevin, Kristin Lemhöfer, and Marc Brysbaert 2013: The word-frequency effect in first- and second-language acquisition: A lexical entrenchment effect. *Quarterly Journal of Experimental Psychology* 66: 843–863.

Diessel, Holger 2007: Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25: 108–127.

Ellis, Nick C. (ed.) 1994: *Implicit and Explicit Learning of Languages*. San Diego, CA: Academic Press.

Ellis, Nick C. 1996: Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition* 18(1): 91–126.

Ellis, Nick C. 1998: Emergentism, connectionism and language learning. *Language Learning* 48(4): 631–664.

Ellis, Nick 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, Nick C. 2005: At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition* 27: 305–352.

Ellis, Nick C. 2006a: Language acquisition as rational contingency learning. *Applied Linguistics* 27(1): 1–24.

Ellis, Nick C. 2006b: Selective attention and transfer phenomena in SLA: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics* 27(2): 1–31.

Ellis, Nick C. 2008a: The dynamics of second language emergence: Cycles of language use, language change, and language acquisition. *Modern Language Journal* 41(3): 232–249.

Ellis, Nick. C. 2008b: Optimizing the input: Frequency and sampling in usage-based and form-focussed learning. In: Michael H. Long and Catherine J. Doughty (eds.), *Handbook of Language Teaching*, 139–158. Oxford: Blackwell.

Ellis, Nick C. 2008c: The psycholinguistics of the Interaction Hypothesis. In: Alison Mackey and Charlene Polio (eds.), *Multiple Perspectives on Interaction in SLA: Second Language Research in Honor of Susan M. Gass*, 11–40. New York: Routledge.

Ellis, Nick C. 2008d: Usage-based and form-focused language acquisition: The associative learning of constructions, learned-attention, and the limited L2 endstate. In: Peter Robinson and Nick C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, 372–405. London: Routledge.

Ellis, Nick C. 2012a: Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics* 32: 17–44.

Ellis, Nick C. 2012b: Second language acquisition. In: Graham Trousdale and Thomas Hoffmann (eds.), *Oxford Handbook of Construction Grammar*, 365–378. Oxford: Oxford University Press.

Ellis, Nick C. 2015: Implicit AND explicit learning of language. In: Peter Rebuschat (ed.), *Implicit and Explicit Learning of Language,* 1–24. Amsterdam/Philadelphia: Benjamins.

Ellis, Nick C. and Fernando Ferreira-Junior 2009a: Construction learning as a function of frequency, frequency distribution, and function. *Modern Language Journal* 93: 370–385.

Ellis, Nick C. and Fernando Ferreira-Junior 2009b: Constructions and their acquisition: Islands and the distinctiveness of their occupancy. *Annual Review of Cognitive Linguistics* 7: 111–139.

Ellis, Nick C. and Diane Larsen-Freeman 2006: Language emergence: Implications for Applied Linguistics (Introduction to the Special Issue). *Applied Linguistics* 27(4): 558–589.

Ellis, Nick C. and Diane Larsen-Freeman (eds.) 2009: *Language as a Complex Adaptive System*. Boston, MA: Wiley-Blackwell.

Ellis, Nick C., Matthew Brook O'Donnell, and Ute Römer 2014: The processing of verb-argument constructions is sensitive to form, function, frequency, contingency, and prototypicality. *Cognitive Linguistics* 25(1): 55–98.

Ellis, Nick C., Rita Simpson-Vlach, and Carson Maynard 2008: Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly* 42(3): 375–396.

Fine, Alex B. and T. Florian Jaeger 2013: Syntactic priming in language comprehension allows linguistic expectations to converge on the statistics of the input. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society (CogSci13)*. Austin TX: Cognitive Science Society.

Frith, Chris D. and Uta Frith 2008: Implicit and explicit processes in social cognition. *Neuron* 60(3): 503–510.

Frith, Uta and Chris D. Frith 2010: The social brain: Allowing humans to boldly go where no other species has been. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365: 165–175.

Frith, Chris D. 2010: What is consciousness for? *Pragmatics and Cognition* 18(3): 497–551.

Garrod, Simon and Martin J. Pickering 2004: Why is conversation so easy? *Trends in Cognitive Sciences* 8(1): 8–11.

Gaskell, Gareth, Jill Warker, Rebecca Frost, James Guest, Reza Snowdon, and Abbi Stackhouse 2012: Sleep facilitates acquisition of implicit phonotactic constraints in speech production. *Journal of Sleep Research* 21: 249.

Goldberg, Adele E. 2006: *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2008: The language of constructions. In: P. Robinson and N. C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*. London: Routledge.

Goldschneider, Jennifer M. and Robert M. DeKeyser 2001: Explaining the 'natural order of L2 morpheme acquisition' in English: A meta-analysis of multiple determinants. *Language Learning* 51(1): 1–50.

Gries, Stefan Th. 2012: Frequencies, probabilities, association measures in usage- /exemplar-based linguistics: Some necessary clarifications. *Studies in Language* 36(3): 477–510.

Gries, Stefan Th. and Anatol Stefanowitsch 2004: Extending collostructional analysis: A corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9: 97–129.

Gries, Stefan Th. and Dagmar S. Divjak (eds.) 2012: *Frequency Effects in Cognitive Linguistics. Vol. 1: Statistical Effects in Learnability*, *Processing and Change*. Berlin/Boston: de Gruyter.

Gries, Stefan Th. and Stefanie Wulff 2009: Psycholinguistic and corpus linguistic evidence for L2 constructions. *Annual Review of Cognitive Linguistics* 7: 164–187.

Han, Zhao-Hong and Terence Odlin (eds.) 2006: *Studies of Fossilization in Second Language Acquisition*. Clevedon: Multilingual Matters.

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

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

Kasper, Gabriele and Johannes Wagner 2011: A conversation-analytic approach to second language acquisition. In: Dwight Atkinson (ed.), *Alternative Approaches to Second Language Acquisition*, 117–142. Oxford: Routledge.

Klein, Wolfgang 1998: The contribution of second language acquisition research. *Language Learning* 48: 527–550.

Lakoff, George 1990: The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* 1(1): 39–74.

Lantolf, James P. and Steven L. Thorne 2006: *Sociocultural Theory and the Genesis of Second Language Development*. Oxford: Oxford University Press.

Lee, Namhee, Lisa Mikesell, Anna Dina L. Joaquin, Andrea W. Mates, and John H. Schumann 2009: *The Interactional Instinct: The Evolution and Acquisition of Language*. New York: Oxford University Press.

Littlemore, Jeanette 2009: *Applying Cognitive Linguistics to Second Language Learning and Teaching*. Basingstoke/New York: Palgrave Macmillan.

MacDonald, Maryellen C. and Mark S. Seidenberg 2006: Constraint satisfaction accounts of lexical and sentence comprehension. In: Matthew J. Traxler and Morton A. Gernsbacher (eds.), *Handbook of Psycholinguistics*, 2nd edition, 581–611. London: Elsevier.

MacWhinney, Brian J. 1991: *The CHILDES Project: Tools for Analyzing Talk*. Mahwah, NJ: Erlbaum.

MacWhinney, Brian J. (ed.) 1999: *The Emergence of Language*. Mahwah, NJ: Erlbaum.

MacWhinney, Brian J. 2007: The TalkBank Project. In: Joan C. Beal, Karen P. Corrigan, and Hermann L. Moisl (eds.), *Creating and Digitizing Language Corpora: Synchronic Databases*, Vol. 1, 163–180. Houndmills: Palgrave-Macmillan.

MacWhinney, Brian J. 2008: A unified model. In: Peter Robinson and Nick C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition*, 341–371. New York: Routledge.

MacWhinney, Brian J. and Johannes Wagner 2010: Transcribing, searching and data sharing: The CLAN software and the TalkBank data repository. *Gesprächsforschung* 11: 154–173.

MacWhinney, Brian J., Elizabeth Bates, and Reinhold Kliegl 1984: Cue validity and sentence interpretation in English, German, and Italian. *Journal of Verbal Learning and Verbal Behavior* 23(2): 127–150.

Markee, Numa P. 2008: Toward a learning behavior tracking methodology for CA-for-SLA. *Applied Linguistics* 29: 404–427.

McClelland, James L. 1998: Complementary learning systems in the brain: A connectionist approach to explicit and implicit cognition and memory. *Annals of the New York Academy of Sciences* 843: 153–169.

McClelland, James L. and David E. Rumelhart 1988: *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs, and Exercises*. Cambridge, MA: MIT Press.

Moll, Henrike and Michael Tomasello 2007: Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences* 362: 639–648.

Nadel, Lynn and Morris Moscovitch 1998: Hippocampal contributions to cortical plasticity. *Neuropharmacology* 37: 431–439.

Nisbett, Richard E. and Timothy DeCamp Wilson 1977: Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84(3): 231–259.

Norton, Bonny 1997: Language, identity, and the ownership of English. *TESOL Quarterly* 31: 409–430.

Pearl, Lisa and Sharon Goldwater (in press): Statistical learning, inductive bias, and Bayesian inference in language acquisition. In: Jeffrey Lidz, William Snyder, and Joe Pater (eds.), *Oxford Handbook of Developmental Linguistics*. Oxford: Oxford University Press.

Pickering, Martin J. and Simon Garrod 2004: Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences* 27: 169–190.

Pickering, Martin J. and Simon Garrod 2013: An integrated theory of language production and comprehension. *Behavioral and Brain Sciences* 36(4): 329–347.

Pierrehumbert, Janet 2006: The next toolkit. *Journal of Phonetics* 34: 516–530.

Reber, Arthur S. 1993: *Implicit Learning and Tacit Knowledge: An Essay on the Cognitive Unconscious*. Oxford: Oxford University Press.

Rebuschat, Patrick (ed.) 2015: *Implicit and Explicit Learning of Language*. Amsterdam/ Philadelphia: Benjamins.

Rebuschat, Patrick and John N. Williams (eds.) 2012: *Statistical Learning and Language Acquisition*. Berlin/Boston: de Gruyter.

Rescorla, Robert A. and Allan R. Wagner 1972: A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In: Abraham H. Black and William F. Prokasy (eds.), *Classical Conditioning II: Current Theory and Research*, 64- 99. New York: Appleton-Century-Crofts.

Rips, Lance J., Edward E. Smith, and Douglas L. Medin 2012: Concepts and categories: Memory, meaning, and metaphysics. In: Keith J. Holyoack and Robert G. Morrison (eds.), *The Oxford Handbook of Thinking and Reasoning*, 177–209. Oxford: Oxford University Press.

Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams 1986: Learning representations by back-propagating errors. *Nature* 323: 533–536.

Rumelhart, David E., Paul Smolensky, Jay L. McClelland, and Geoffrey Hinton 1988: Schemata and sequential thought processes in PDP models. In: Allan M. Collins and Edward E. Smith (eds.) *Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence*, 224–249. San Mateo, CA: Morgan Kaufmann.

Sanders, Ted 1997: Psycholinguistics and the discourse level: Challenges for cognitive linguistics. *Cognitive Linguistics* 8: 243–265.

Schmidt, Richard 1984: The strengths and limitations of acquisition: A case study of an untutored language learner. *Language, Learning, and Communication* 3: 1–16.

Schmidt, Richard 1990: The role of consciousness in second language learning. *Applied Linguistics* 11: 129–158.

Schmidt, Richard 2001: Attention. In: Peter Robinson (ed.), *Cognition and Second Language Instruction*, 3–32. Cambridge: Cambridge University Press.

Shanks, David R. 2010: Learning: From association to cognition. *Annual Review of Psychology* 61: 273–301.

Simpson-Vlach, Rita and Nick C. Ellis 2010: An Academic Formulas List (AFL). *Applied Linguistics* 31: 487–512.

Sinclair, John M. 1996: The search for units of meaning. *Textus* 9: 75–106.

Sinclair, John M. 2004: *Trust the Text: Language, Corpus and Discourse*. London: Routledge.

Sinclair, John M. and Anna Mauranen 2006: *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam/Philadelphia: Benjamins.

Smith, Nathaniel J. and Roger Levy 2013: The effect of word predictability on reading time is logarithmic. *Cognition* 128: 302–319.

Spivey, Michael 2006: *The Continuity of Mind*. Oxford: Oxford University Press.

Tarone, Elaine 2007: Sociolinguistic approaches to second language acquisition research 1997–2007. *The Modern Language Journal* 91(1): 837–848.

Tomasello, Michael (ed.) 1998: *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*. Mahwah, NJ: Erlbaum.

Tomasello, Michael 2003: *Constructing a Language: A Usage-based Theory of Language Acquisition*. Boston, MA: Harvard University Press.

Tomasello, Michael 2008: *The Origins of Human Communication*. Cambridge, MA: MIT Press.

Tomasello, Michael and Daniel Stahl 2004: Sampling children's spontaneous speech: How much is enough? *Journal of Child Language* 31: 101–121.

Tyler, Andrea 2012: *Cognitive Linguistics and Second Language Learning: Theoretical Basics and Experimental Evidence*. Abingdon: Routledge.

Warker, Jill A. and Gary S. Dell 2006: Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32: 387–398.

Wertsch, James V. 1998: *Mind as Action*. Oxford: Oxford University Press.

Wills, Andy J. 2009: Prediction errors and attention in the presence and absence of feedback. *Current Directions in Psychological Science* 18: 95-100.