Focus Article

On language 'utility': processing complexity and communicative efficiency

T. Florian Jaeger[1]* and Harry Tily[2]

Functionalist typologists have long argued that pressures associated with language usage influence the distribution of grammatical properties across the world's languages. Specifically, grammatical properties may be observed more often across languages because they improve a language's *utility* or decrease its *complexity*. While this approach to the study of typology offers the potential of explaining grammatical patterns in terms of general principles rather than domain-specific constraints, the notions of utility and complexity are more often grounded in intuition than empirical findings. A suitable empirical foundation might be found in the terms of processing preferences: in that case, psycholinguistic measures of complexity are then expected correlate with typological patterns. We summarize half a century of psycholinguistic work on 'processing complexity' in an attempt to make this work accessible to a broader audience: What makes something hard to process for comprehenders, and what determines speakers' preferences in production? We also briefly discuss recently emerging approaches that link preferences in production to communicative efficiency. These approaches can be seen as providing well-defined measures of utility. With these psycholinguistic findings in mind, it is possible to investigate the extent to which language usage is reflected in typological patterns. We close with a summary of paradigms that allow the link between language usage and typology to be studied empirically. *WIREs Cogn Sci* 2011 2 323–335 DOI: 10.1002/wcs.126

INTRODUCTION

Researchers from diverse approaches with a common functional understanding of language have long stressed that the *utility of a form* relative to a human language user's communicative needs is crucial for understanding its ontogeny and typological distribution (see Refs 1–5, papers in Ref 6). Utility means suitability for a certain communicative function, but also suitability for the abilities of the user, implying that human cognitive abilities directly or indirectly constrain the space of possible grammars. Work within functional linguistics[7,8] has aimed to define utility in terms of general cognitive principles (e.g., pressure for iconicity of form and function, or for concise representation of salient/frequent concepts). Uutility has also been related to language processing specifically. Consider the particularly succinct hypothesis stated in Hawkins[9]:

> '*Grammars have conventionalized syntactic structures in proportion to their degree of preference in performance, as evidenced by patterns of election in corpora and by ease of processing in psycholinguistic experiments.*' [Ref 9, p. 3]

In other words, typological distributions should mirror gradient measures of production and comprehension complexity. More easily processed forms are by hypothesis preferred. Under the assumption that it is known what is complex (hard to process), the processing complexity approach makes empirically testable explanations for cross-linguistically observed properties of grammars (Hawkins[9,10] and Christiansen and Chater,[11] among others). However, in order to test the hypothesis that typological distributions reflect processing complexity, an independently

*Correspondence to: fjaeger@bcs.rochester.edu

[1]Brain and Cognitive Sciences, Computer Science, University of Rochester, Rochester, NY, USA

[2]Linguistics, Stanford University, Stanford, CA, USA

DOI: 10.1002/wcs.126

motivated, well defined, and empirically assessable notion of processing difficulty is essential. In practice, there is a considerable gap between the fields of typology and psycholinguistics. Hence, relatively little work on typology has integrated what contemporary work on language processing has to offer. This article is intended as a stepping stone for researchers with an interest in crossing that gap.

We review the most influential accounts and findings from the study of language production and comprehension that bear on the question 'what is complex'? For reasons of space, we limit ourselves to work on *sentence* processing. We also summarize recently (re-)emerging information theoretic accounts of language processing that have their roots in a slightly different perspective on utility, in terms of efficient communication rather than processing.[3] We close with a brief summary of recent approaches to studying the link between language use and grammar.

PROCESSING DIFFICULTY IN COMPREHENSION AND PRODUCTION

To a first approximation, contemporary accounts of sentence processing fall into one of two broad categories. In *memory-based* accounts, processing difficulty arises when the sentence being processed overloads limits on some sort of mental storage system, the nature of which is not necessarily specified further (see Refs 9,12–14, though see Refs 15,16). In *expectation-based* and *constraint satisfaction* accounts, processing difficulty is intimately linked to the probability of processed structures: words and structures that are infrequent or unexpected or for which there are conflicting cues are predicted to be hard to process.

Overtaxed memory has been invoked to explain comprehension difficulty (e.g., slowed reading times, reduced comprehension accuracy), in center-embedded clauses (1b) compared to right-embedded sentences (1a) and object-extracted relative clauses (2b) compared to subject-extracted relative clauses (2a).

1. (a) This is the malt that was eaten by the rat that was killed by the cat.
   (b) This is the malt that the rat that the cat killed ate.
2. (a) The reporter that interviewed the senator is from New York.
   (b) The senator that the reporter interviewed is from New York.

Evidence that processing difficulty may originate in memory resources comes from findings that individual differences in working memory capacity correlate with difficulty[17–20] and from studies showing poorer performance on a secondary memory task, while comprehending more complex structures (see Refs 21,22,23). At the heart of many memory-based accounts is the idea that comprehension difficulty reflects *dependency lengths*—the distance between words that are dependent on each other for interpretation, such as a verb and its object (*interviewed* and *reporter/senator* in (2a/b); see Ref 12 for a review of previous accounts). As the linguistic signal unfolds over time, comprehension proceeds incrementally: words are integrated one by one into a representation of the structure and interpretation of the sentence. This integration requires retrieval of previous material. Processing difficulty at the integration point increases with increasing dependency length.[12,13,16,24–28] Figure 1 illustrates the role of dependency length for examples (1a) and (1b) above (in the upper and lower panel, respectively). The word 'eaten' in the top panel of Figure 1 should be relatively easy to process given that it only requires the integration of one local dependency. In contrast, processing of the word 'ate' in the bottom panel of Figure 1 requires integration of two nonlocal dependencies, which would be predicted to result in relatively high processing cost.

How exactly dependency length is to be measured (e.g., in terms of intervening words, syntactic nodes or phrases, new discourse referents, etc.) is a matter of ongoing debate, though in practice all of these measures are highly correlated.[29,30] Dependency lengths can thus be used to generate word-by-word predictions of comprehension difficulty (or, combined in some way, to yield a per-sentence measure). Hawkins[9,14] proposes that word orders which minimize *constituent recognition domains* (roughly, the shortest string of words within which all putative children of a syntactic constituent can be identified) are processed more efficiently. Dependency and domain-based theories make identical predictions in most cases. We discuss results in terms of the former only because dependencies naturally fit into a fuller memory-based theory.

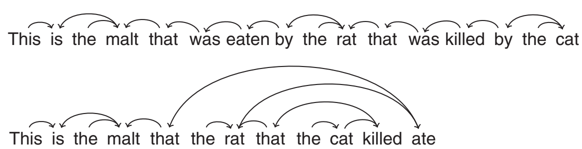

FIGURE 1 | Illustration of sentences with different dependency lengths. The sentence in the top panel contains mostly local dependencies. The sentence in the bottom panel contains several complex nonlocal dependencies.

Dependency length seems to affect production as well. Production research usually assesses difficulty directly in terms of production latencies and fluency, or indirectly by *speaker preference* for one of several meaning-equivalent structures. For instance, (1a) and (1b) have the same meaning, but if speakers typically avoid (1b) in favor of (1a), we can infer that (1b) incurs more processing difficulty in production. The same holds for the so-called dative alternation:

3. (**a**) Give the carrot to the white rabbit. (NP-PP; theme-recipient)

(**b**) Give the white rabbit the carrot. (NP-NP; recipient-theme)

Given structural choices like (3), speakers of English prefer to order the shorter constituent first.[9,10,29,31–37] Consider, for example, the dative alternation in (3) with a shorter recipient (e.g., *the rabbit*) and a longer theme (e.g., *the biggest carrot you can find*). In both orders, the dependency between the verb and the first argument is immediately resolved. The two orders differ, however, in the number of words that need to be processed before the dependency with the second argument can be resolved. By ordering the shorter phrase first, speakers shorten verb-argument dependencies. Since speakers, just like comprehenders, need to incrementally retrieve the words they produce in order to integrate them into a well-formed sentence, the observed preference has been interpreted as evidence for resource accounts.[9,14,31] An alternative explanation is that speakers simply defer the production of more complex phrases, presumably to buy more time to plan them.[29,37] In this case, observed ordering preferences would not necessarily reflect the processing of dependencies, but rather a strategy to deal with the demands of incremental production. While there is independent support for such a strategy (see below), recent evidence suggests that constituent length effects cannot be reduced to the delayed production of complex constituents: Speakers of head-final languages prefer the opposite, long-before-short, order (for Japanese, see Refs 14,38,39; for Korean, see Ref 40). This is consistent with the dependency minimization resource account, since the long-before-short order reduces dependency length if the head follows its arguments.[9,41]

The most articulated proposal for a memory-based architecture for language processing is arguably work by Lewis and colleagues (for an introduction, see Ref 16; for more details, see Ref 15). Lewis and colleagues' model is based on general assumptions about cognitive architecture (e.g., memory, perceptual, and motor processing) that are empirically supported by research on both linguistic and nonlinguistic cognitive abilities (for an overview, see Ref 42). In their model, 'chunks' in memory are bundles of features representing semantic/structural properties of words, phrases, or referents. Chunks are created with high *activation*, which decays over time. Higher activation levels lead to easier retrieval (e.g., at a later verb), capturing the dependency length phenomena described above.

The model predicts that further properties of memory, such as similarity-based interference, should be evidenced in processing difficulty.[16,26,43,44] In line with this prediction, processing of a referent during comprehension is slowed when similar competitors are activated in memory. For example, object-extracted clefts like (4) lead to increased processing difficulty when the two NPs are both names or both professions.[26,45]

(**4**) It was the barber/John who the lawyer/Bill saw in the parking lot.

Interference is also observed when a secondary task requires comprehenders to maintain a separate set of words in memory that are similar to the target (see Refs 22,23,43; see also Refs 44,46,47 for related findings). In Lewis and colleagues' model, retrieval of a target chunk is *content-based,* and hence competitors with similar (feature) content decrease the retrieval efficiency.

Comprehension difficulty is also affected by repeated mention: Targets are retrieved faster when recently mentioned.[48] Lewis and colleagues attribute such effects to repeated retrieval from memory: Every time a target is retrieved, its activation increases. Some targets may also be *inherently* more 'accessible' or easier to retrieve. Evidence from isolated lexical decision and ERP studies suggest that words denoting concrete, animate, imageable concepts are processed faster than words denoting abstract, inanimate, and less imageable concepts.[49–51] Sentences containing NPs higher on the accessibility hierarchy (pronoun < name < definite < indefinite[52]) are processed faster.[53,54] In a framework like that of Lewis and colleagues, this could be modeled as higher resting activation (though Lewis and Vasishth[15] and Lewis et al.[16] do not address this case).

'Conceptual accessibility'[55] also affects production. In a variety of languages, speakers' preferences in word order alternations (e.g., (3) above) are affected by referents' imageability,[55] prototypicality,[56,57] animacy/humanness,[34,58–60] givenness due to previous mention[61–63] and semantic similarity to recently mentioned words.[64,65] Conceptual accessibility affects the linking between referents and grammatical functions, with more accessible referents being linked to higher grammatical functions like the subject. These indirect effects of accessibility (also called alignment effects), influence, for example, construction choice in the passive and dative alternations.[34,55] Independently, accessibility affects word order which also leads to more accessible referents being ordered earlier in a sentence (direct effects of accessibility, a.k.a. 'availability' effects; see Refs 59,60,66–68). Interestingly, this accessible-before-inaccessible tendency seems to hold independent of the language's headedness,[62,68,69] unlike the short-before-long preference of head initial languages which contrasts with the long-before-short preference of head-final languages (for a more exhaustive cross-linguistic summary, see Ref 70). Accessibility effects remain after controlling for constituent complexity.[34,55,59,66,71,72]

Beyond conceptual accessibility, the ease of word form retrieval also seems to affect sentence production, with easier to retrieve forms being ordered earlier cross-linguistically.[61,62] In short, there is strong evidence that inherent and contextually conditioned properties of referents and word forms contribute to processing difficulty in both production and comprehension.

In the broadest possible sense, the effects discussed so far can be considered memory-based: processing difficulty is correlated with ease of retrieval from memory, as modeled by (1) the retrieval target's baseline activation, (2) boosts in activation associated with previous retrievals, (3) activation decay over time, and (4) the activation and similarity of other elements in memory (cf. Refs 15,16).

Processing difficulty also has been related to the degree of *uncertainty* in the input. Although in principle compatible with memory-based accounts, so-called *expectation-based* and *constraint satisfaction* accounts of language processing focus on the allocation of resources rather than resource limitation. Consider, for example, the temporary ambiguity in (5), where the verb form *raced* is strongly biased to be interpreted as a past tense intransitive form rather than the passive participle of a transitive in a reduced subject relative. This leads to increased processing difficulty (slower comprehension times) at the disambiguation point (here *fell*).

(5) The horse raced past the barn fell.

The difficulty of so-called *garden path* sentences like (5) has prompted a substantial amount of research (see Refs 73–76; for a recent summary, see Ref 77). Yet comprehension is usually successful. Noticeable garden paths like (5) are rare, thus implying that listeners have efficient and robust ways to deal with ambiguity.[71,78,79] Comprehenders more or less simultaneously incorporate a variety of linguistic and nonlinguistic cue types, including syntactic, semantic, and pragmatic cues.[80–88] This insight underpins many contemporary accounts of language processing.[1–3,89–94] Some accounts explicitly commit to the hypothesis that word-by-word processing difficulty is determined by expectations based on the linguistic and nonlinguistic cues.[90–94] For a rational listener, comprehension difficulty should be correlated with a word's *surprisal*.[90,95] A word's surprisal is the logarithm of the reciprocal of its probability. In other words, the more *un*expected a word is, the higher its surprisal. A word's probability can be estimated in many different ways and based on different representational assumptions (e.g., *n*-grams, probabilistic phrase structure grammars, construction grammars, topic models, to name just a few). What types of cues language users integrate (i.e., what the relevant representations are) and how multiple cues are integrated into probability estimates is a subject of ongoing research.[72,94,96–98]

Figure 2 shows a garden path sentence annotated with the probability of upcoming words at each point under a simple probabilistic grammar adapted from Ref 90. Words with higher probability are more predictable, and so incur less processing difficulty.

Evidence that surprisal affects incremental processing difficulty comes from self-paced reading experiments and eye-tracking corpora (see Refs 93,99–101; for an overview of expectation-based effects in comprehension, see Refs 93,102). One straightforward way to incorporate probability/surprisal into a memory-based model would be to have the former contribute to baseline activation, which is essentially how connectionist accounts explain probabilistic effects.[84,103]

Probabilities also affect production. For example, disfluencies are more likely with less frequent and less predictable words[104,105] and structures.[106–108] This seems to suggest that in production, like in comprehension, more surprising words and structures are harder to process. Interestingly, the pronunciation of words is not only affected by the probability of *upcoming* words and structures, but also by their own probability: more predictable instances of the same word are on average produced with shorter duration.[108–113] One possible interpretation of this finding is that more predictable words are produced with shorter duration because they are *easier* to produce. Note, however, that the duration of a word is not the same as the time it takes to plan that word (latency). While longer latencies would be expected if less predictable word

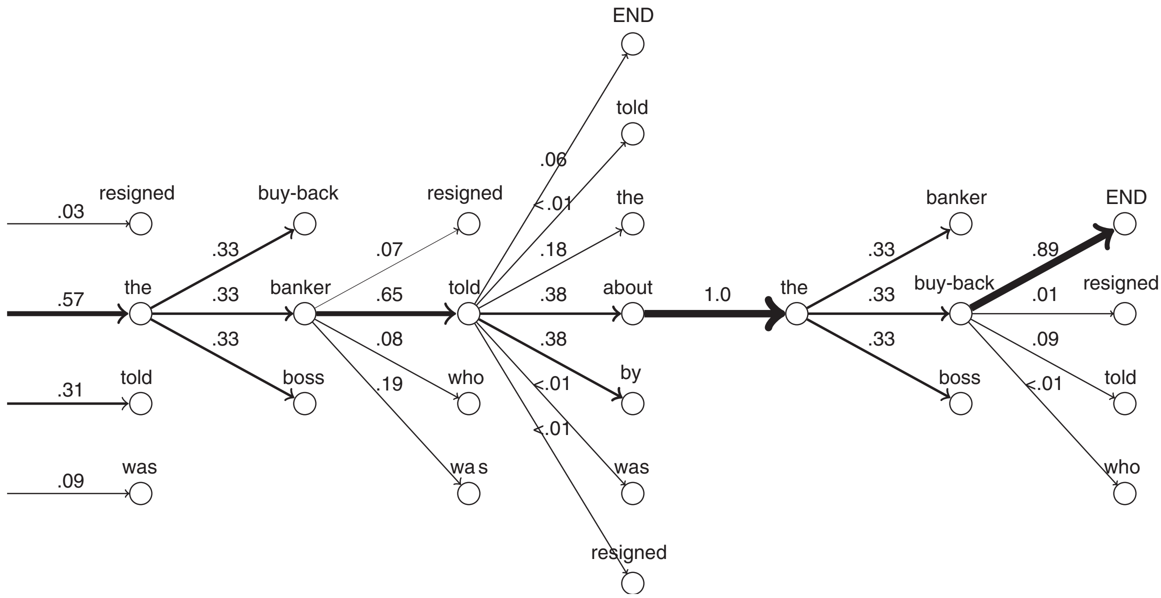

FIGURE 2 | Illustration of the probability of upcoming words under a probabilistic grammar generating the sentence 'The banker told about the buy-back resigned'.

tokens are harder to produce, longer *durations* would only be expected if each segment (phone) in less predictable word tokens is harder to retrieve than the segments in more predictable word tokens. In other words, longer durations for less predictable tokens are only expected if the segments of less predictable word tokens are on average less predictable than the segments of more predictable word tokens. This prediction seems to not have been tested so far. There also are alternative interpretations of the finding that more predictable word tokens are pronounced with less articulatory detail. We briefly discuss them next.

LINKING PROCESSING PREFERENCES TO UTILITY: COMMUNICATIVE EFFICIENCY

One interpretation of the finding that more predictable words are pronounced with shorter duration is that language production is efficient in that it distributes information uniformly. Formally, surprisal is identical with Shannon information[114]: Information(word) $=$ $\log_2[1/p(\text{word})] = \text{Surprisal(word)}$. A word's Shannon information can be understood as a measure of how much new information the word adds to the preceding discourse. Shannon information is a clearly defined and intuitive measure: the more predictable a word is given the preceding context, the less information it adds. If the next word is predictable with absolute certainty, it adds no new information ($\log_2 1/1 = 0$ bits of new information). A uniform distribution of information across the signal has been argued to be theoretically optimal for communication[71,72,109,115–118] and in terms of processing difficulty.[119] In other words, the shorter duration of more predictable instances of words is expected if language production is organized to be efficient. If speakers aim to avoid peaks and troughs in information density (the distribution of information over the linguistic signal and hence over time), instances of words that carry more information should be pronounced with more duration. This is illustrated in Figure 3.

The interpretation of the word duration findings in terms of communicative efficiency presents an intriguing possibility. Recall that functionalist accounts of the typological distribution of grammatical properties are based on the hypothesis that the *utility of a form* relative to a human language user's communicative needs is crucial for understanding its ontogeny and typological distribution. Above we have focused on accounts that interpret utility in terms of *processing difficulty*,[9] rather than in a broader interpretation of utility. This has the advantage that the notion of processing difficulty is much better understood than the notion of utility. The study of processing difficulty has been at the heart of psycholinguistics for over 50 years. Intuitively though, the concept of utility put forward in functionalist work is more closely related to communicative efficiency than to processing difficulty. To the extent that there is empirical support for communicatively efficient production, this would present a promising link

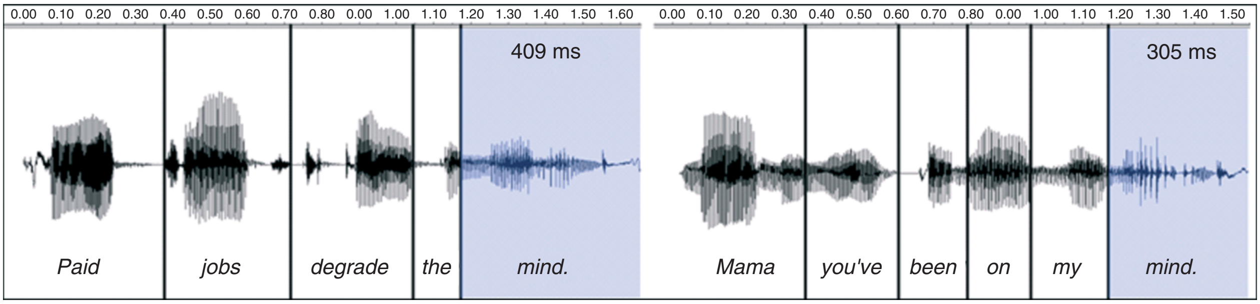

FIGURE 3 | Illustration of the relation between word duration and word predictability (and hence information density) predicted by certain accounts of communicative efficiency.[71,72,109,116–119] The word 'mind' is more predictable (and carries less information) in the right panel compared to the left panel. (Reprinted with permission from Ref 120.)

between recent work in psycholinguistics and functional approaches to typology and language change.

It is now generally accepted that frequency and contextual probability play crucial roles in language change: frequency affects the reduction of a form[121–124] and, hence, the likelihood of it being stored as a single chunk; highly frequent sequences or commonly expressed grammatical relationships are more likely to become grammaticalized (see Refs 122,125–127; for recent discussions, see Refs 128,129). Communicative efficiency effects on production provide a viable account of these findings.[71]

Indeed, recent work has begun to uncover further evidence that communicative efficiency affects speakers' preferences during production. Instances of words that add less information are not only pronounced with shorter duration, they are also pronounced with less articulatory and phonological detail.[110,112,117,118] Within words, more informative segments have more duration and more distinctive centers of gravity[116,118,130] and are less likely to be deleted.[131] Beyond articulation and phonology, there is evidence that morphological choice points in production are affected by information density. Speakers are more likely to reduce more predictable instances of morphologically contractible words (e.g., *could not* vs. *couldn't*[132]). This inverse link between form reduction and information density is also observed for the optional mentioning of function words. For example, speakers prefer the reduced variants of English complement and relative clauses, the more expected such a clause is in its context.[71,72,79,133] Consider the following example of optional relativizer mentioning, where the relativizer 'that' can be mentioned or omitted.

6. (a) That's the painting [they told me about].
   (b) That's the painting [(that) they told me about].

Simplifying for the current purpose, the relative clause onset 'they' in (6a) contains two pieces of information, the presence of a relative clause boundary and the fact that the relative clause onset starts with the word 'they'. In (6b), these two pieces of information are distributed over two words ('that they') rather than one. Thus, speakers should prefer to use variant (6b) over (6a) whenever the relative clause or the onset is unexpected in its context. In other words, speakers should prefer to spread out high information content at the relative clause onset over more words by producing the relativizer 'that'. This is illustrated in Figure 4 (see Ref 71 for more detail).

Even beyond the level of clausal planning, there is evidence that speakers prefer to distribute information uniformly (for an overview, see Refs 115,134–137). In short, communicative efficiency seems to affect speakers' preferences during production (for further discussion, see Ref 71).[a]

INVESTIGATING THE LINK BETWEEN LANGUAGE USAGE AND TYPOLOGY

We have summarized psycholinguistic findings that speak to processing preferences in production and

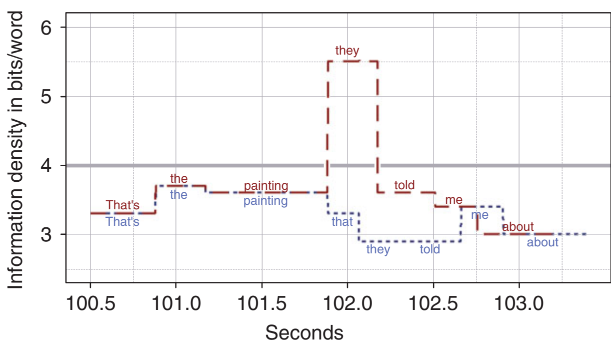

FIGURE 4 | Illustration of the predictions of uniform information density, an account of communicative efficiency, for optional relativizer mentioning in nonsubject-extracted relative clauses.

comprehension. These processing preferences have been linked to mechanisms of language processing and communicative pressures. Next, we briefly discuss a few empirical paradigms that we consider particularly promising in terms of their potential for research on how *processing difficulty* and *communicative efficiency* might contribute to typological patterns.

There are several ways in which processing could come to shape grammar (for a detailed discussion, see Ref 138). In acquisition, less complex forms may be learned preferentially due to higher frequency in the input: there is extensive evidence that input frequency plays a central role in acquisition.[139–141] Higher comprehension complexity may also impede acquisition regardless of input frequency. Similar pressures may occur throughout the lifespan if adult speakers reorganize their usage patterns in response to previous communicative episodes.[142,143] Although relatively little empirical work has directly addressed these questions to date, interdisciplinary researchers have begun to invent new methodologies which will allow us to directly study the processing-grammar link.

Naturally, typological surveys may be conducted with a view to the processing complexity of the forms studied. This is the primary methodology of Hawkins.[9,14] Beyond this, we can also make predictions based on processing theories for the kind of languages we expect to see, and ask whether extant languages tend to look more like these than would be expected by chance. For instance, a body of work stretching back to Zipf[3,144] attempts to explain diverse properties of the lexicons by showing that they are close to theoretically derived models of an optimal communicative system.[145–152] In syntax, less work of this kind has been completed, though Gildea and Temperley[153] show English grammar yields dependency lengths that are remarkably close to the theoretical minimum that could be achieved, thereby supporting the hypothesis that (memory or other processing) pressure to minimize dependency lengths has influenced the evolution of English. Work within the information theoretic accounts of language production mentioned above has shown that information is distributed efficiently across the linguistic signal *given the grammar of the language*.[71,72,79,109,115,116,118,119,137] The information efficiency of existing languages has, however, not yet been compared against theoretically possible grammars.

Within the burgeoning literature on computer simulation of language emergence and change (see Refs 154–156; see also Ref 157), researchers have implemented human processing theories into their artificial agents, observing the languages that emerge as a result. Implementing the processing theories found in Hawkins[10] leads to a distribution of emergent language types similar to the true typological distribution (see Ref 158; see also Refs 11,159). In a closely related paradigm, human subjects invent communication systems, or learn artificial languages and then teach them to the next 'generation' of subjects. The communication systems that emerge display cross-linguistically observed properties of human language.[160–165] Similar studies could certainly test for influences of the specific processing theories described here.

Another paradigm ripe for application to the question at hand is so-called artificial language experimentation, where participants learn new languages designed by the experimenter.[166,167] Experimentation with the comprehension and production of carefully designed artificial languages could prove valuable in separating language-specific processing complexity from the kind of general universal processing principles which may underlie language universals. For example, preliminary evidence suggests that artificial languages with shorter verb-dependent distances are more easily learned.[168]

CONCLUSION

Functionalist linguistics has long held that the observed distribution of grammars across languages of the world can at least in part be accounted for in terms of biases that operate during language use (e.g., during language acquisition or during everyday language use). These biases are assumed to lead to a preference for forms that have higher 'utility'—for example, in that they are less complex or, in other words, easier to process. This raises the question of what it means for a form to be more or less easy to process. We have summarized psycholinguistic work from the last couple of decades on sentence production and comprehension that speaks to this question. The notion of 'utility' is, however, much broader than processing complexity. Language utility can be understood as relative to a human language user's communicative needs. We have briefly summarized recent work in computational psycholinguistics that builds on this idea. Compatible with a long standing claim of functionalist linguistics, this line of work has provided evidence that incremental language production seems to be affected by communicative pressures. While some have started to incorporate psycholinguistic theory and findings into the study of typology (most notably, Refs 9,14,41), research in psycholinguistics, linguistics and, specifically, typology will benefit from further integration.

NOTES

[a]It is arguably an appealing property of the above works that it avoids one of the pitfalls of frequency-based explanations of typological patterns. Simple frequency/linguistic expectation-based accounts of processing difficulty remain relatively weak as predictors of universal typological patterns (comprehension difficulty based on frequency being used to predict cross-linguistic frequency). However, in the work on communicative efficiency discussed here, the relevant probabilities are conditioned on nonlinguistic properties like meaning (e.g., the probability of negation or clausal modification).

ACKNOWLEDGEMENTS

The authors thank Jacqueline Gutman and Camber Hansen-Karr for help with the preparation of the manuscript. This work was partially funded by NSF Grant BCS-0845059 to TFJ.

REFERENCES

1. Givón T. Markedness in grammar: distributional, communicative and cognitive correlates of syntactic structure. *Stud Lang* 1991, 15:335–370.
2. Givón T. On Interpreting Text-Distributional Correlations. Some Methodological Issues. In: Payne DL, ed. *Pragmatics of word order flexibility*. Amsterdam and Philadelphia: John Benjamins Publishing Co; 1992, 305–320.
3. Zipf GK. *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press; 1949.
4. Slobin DI. Cognitive prerequisites for the development of grammar. In: Ferguson CA, Slobin DI, eds. *Studies of Child Language Development*. New York: Holf, Rinehart & Winston; 1973.
5. Hockett C. The origin of speech. *Sci Am* 1960, 203: 88–96.
6. Bates EA, MacWhinney B. Functionalism and the competition model. In: MacWhinney B, Bates EA, eds. *The crosslinguistic study of sentence processing*. Cambridge, England: Cambridge University Press; 1989, 3–73.
7. Givón T. *Syntax*, vol. I–II. Amsterdam & Philadelphia: John Benjamins; 2001.
8. Croft W, Cruse D. *Cognitive Linguistics*. Cambridge, UK: Cambridge University Press; 2004.
9. Hawkins JA. *Efficiency and Complexity in Grammar*. Oxford: Oxford University Press; 2004.
10. Hawkins JA. The relative ordering of prepositional phrases in English: going beyond manner-place-time. *Lang Var Change* 1999, 11:231–266.
11. Christiansen MH, Chater N. Language as shaped by the brain. *Behav Brain Sci* 2008, 31:489–509.
12. Gibson E. Linguistic complexity: locality of syntactic dependencies. *Cognition* 1998, 68:1–76.
13. Gibson E. The dependency locality theory: a distance-based theory of linguistic complexity. In: Miyashita Y, Marantx P, O'Neil W, eds. *Image, Language, Brain*. Cambridge, MA: MIT Press; 2000, 95–112.
14. Hawkins JA. *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press; 1994.
15. Lewis RL, Vasishth S. An activation-based model of sentence processing as skilled memory retrieval. *Cogn Sci* 2005, 29:375–419.
16. Lewis RL, Vasishth S, Van Dyke JA. Computational principles of working memory in sentence comprehension. *Trends Cogn Sci* 2006, 10:447–454.
17. King J, Just MA. Individual differences in syntactic processing: the role of working memory. *J Mem Lang* 1991, 30:580–602.
18. Just M, Carpenter P. A capacity theory of comprehension: individual differences in working memory. *Psychol Rev* 1992, 99:122–149.
19. Perlmutter NJ, MacDonald MC. Individual differences and probabilistic constraints in syntactic ambiguity resolution. *J Mem Lang* 1995, 34:521–542.
20. Daneman M, Merikle PM. Working memory and language comprehension: a meta-analysis. *Psych Bull Rev* 1996, 3:422–433.
21. Wanner E, Maratsos M. An ATN approach to comprehension. In: Halle M, Bresnan J, Miller GA, eds. *Linguistic Theory and Psychological Reality*. Cambridge, MA: MIT Press; 1978, 119–161.
22. Fedorenko E, Gibson E, Rohde D. The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *J Mem Lang* 2006, 54:541–553.
23. Gordon PC, Hendrick R, Levine W. Memory-load interference in syntactic processing. *Psychol Sci* 2002, 425–430.
24. Ford M. A method for obtaining measures of local parsing complexity throughout sentences. *J Verb Learn Verb Behav* 1983, 22:203–218.
25. Grodner DJ, Gibson E. Consequences of the serial nature of linguistic input for sentential complexity. *Cogn Sci* 2005, 29:261–290.

26. Gordon PC, Hendrick R, Johnson M. Memory interference during language processing. *J Exp Psychol Learn Mem Cogn* 2001, 27:1411–1423.
27. McElree B, Foraker S, Dyer L. Memory structures that subserve sentence comprehension. *J Mem Lang* 2003, 48:67–91.
28. Vasishth S, Lewis R. Argument-head distance and processing complexity: explaining both locality and antilocality effects. *Language* 2006, 82:767–794.
29. Wasow T. *Postverbal Behavior*. Stanford: CSLI Publications; 2002.
30. Szmrecsányi BM. On operationalizing syntactic complexity. In: Purnelle G, Fairon C, Dister A, eds. *Le Poids des mots. 7th International Conference on Textual Data Statistical Analysis*, vol. II. Louvain-la-Neuve: Presses universitaires de Louvain; 2004, 1032–1039.
31. Hawkins JA. Why are categories adjacent? *J Ling* 2001, 37:1–34.
32. Arnold JE, Wasow T, Asudeh A, Alrenga P. Avoiding attachment ambiguities: the role of constituent ordering. *J Mem Lang* 2004, 55:55–70.
33. Arnold JE, Wasow T, Losongco T, Ginstrom R. Heaviness vs. newness: the effects of structural complexity and discourse status on constituent ordering. *Language* 2000, 76:28–55.
34. Bresnan J, Cueni A, Nikitina T, Baayen H. Predicting the Dative Alternation. In: Boume G, Kraemer I, Zwarts J, eds. *Cognitive Foundations of Interpretation*. Amsterdam, Netherlands: Royal Academy of Science; 2007, 69–94.
35. Bresnan J, Hay J. Gradient grammar: an effect of animacy of the syntax of *give* in New Zealand and American English. *Lingua* 2007, 118:245–259.
36. Lohse B, Hawkins J, Wasow T. Processing domains in English verb-particle construction. *Language* 2004, 80:238–261.
37. Wasow T. End-weight from the speaker's perspective. *J Psycholinguist Res* 1997, 26:347–362.
38. Yamashita H. Scrambled sentences in Japanese: linguistic properties and motivations for productions. *Text* 2002, 22:597–633.
39. Yamashita H, Chang F. Long before short preference in the production of a head-final language. *Cognition* 2001, 81:B45–B55.
40. Choi HW. Length and order: a corpus study of Korean dative-accusative construction. *Discour Cogn* 2007, 14:207–227.
41. Hawkins JA. Processing typology and why psychologists need to know about it. *New Ideas Psychol* 2007, 25:87–107.
42. Anderson J, Bothell D, Byrne M, Douglass S, Lebiere C, Qin Y. An integrated theory of the mind. *Psychol Rev* 2004, 111:1036–1060.
43. Van Dyke J, McElree B. Retrieval interference in sentence comprehension. *J Mem Lang* 2006, 55: 157–166.
44. Lewis RL. Interference in short-term memory: The magical number two (or three) in sentence processing. *J Psycholinguist Res* 1996, 25:93–115.
45. Gordon PC, Hendrick R, Johnson M. Effects of noun phrase type on sentence complexity. *J Mem Lang* 2004, 51:97–114.
46. Lewis RL, Nakayama M. Syntactic and positional similarity effects in the processing of Japanese embeddings. In: Nakayama M, ed. *Sentence Processing in East Asian Languages*. Stanford, CA: CLSI Publications; 2001, 85–113.
47. Gordon PC, Hendrick R, Johnson M, Lee Y. Similarity-based interference during language comprehension: evidence from eye tracking during reading *J Exp Psychol Learn MemCogn* 2006, 32:1304.
48. McKoon G, Ratcliff R. Priming in item recognition: the organization of propositions in memory for text. *J Verb Learn Verb Behav* 1980, 19:369–386.
49. Schwanenflugel PJ. Why are abstract concepts hard to understand? In: Schwanenflugel PJ, ed. *The Psychology of Word Meanings*. Hillsdale, NJ: Lawrence Erlbaum; 1991, 223–250.
50. Kounios J, Holcomb P. Concreteness effects in semantic processing: ERP evidence supporting dual-coding theory. *J Exp Psychol Learn Mem Cogn* 1994, 20: 804–823.
51. Swaab T, Baynes K, Knight R. Separable effects of priming and imageability on word processing: an ERP study. *Cogn Brain Res* 2002, 15:99–103.
52. Ariel M. Accessibility theory: an overview. In: Sanders T, Schilperoord J, Spooren W, eds. *Text Representation: Linguistic and Psycholinguistic Aspects*. Amsterdam: John Benjamins; 2001, 29–87.
53. Warren T, Gibson E. The influence of referential processing on sentence complexity. *Cognition* 2002, 85: 79–112.
54. Warren T, Gibson E. Effects of NP type in reading cleft sentences in English. *Lang Cogn Process* 2005, 20:751–767.
55. Bock JK, Warren R. Conceptual accessibility and syntactic structure in sentence formulation. *Cognition* 1985, 21:47–67.
56. Kelly MH, Bock JK, Keil FC. Prototypicality in a linguistic context: effects on sentence structure. *J Mem Lang* 1986, 25:59–74.
57. Onishi KH, Murphy GL, Bock JK. Prototypicality in sentence production. *Cogn Psychol* 2008, 56: 103–141.
58. Bock JK, Loebell H, Morey R. From conceptual roles to structural relations: bridging the syntactic cleft. *Psychol Rev* 1992, 99:150–171.

59. Branigan HP, Pickering MJ, Tanaka M. Contributions of animacy to grammatical function assignment and word order during production. *Lingua* 2008, 118: 172–189.
60. Prat-Sala M, Branigan HP. Discourse constraints on syntactic processing in language production: a cross-linguistic study in English and Spanish. *J Mem Lang* 2000, 42:168–182.
61. Bock JK, Irwin DE. Syntactic effects of information availability in sentence production. *J Verb Learn Verb Behav* 1980, 19:467–484.
62. Ferreira VS, Yoshita H. Given-new ordering effects on the production of scrambled sentences in Japanese. *J Psycholinguist Res* 2003, 32:669–692.
63. MacWhinney B, Bates EA. Sentential devices for conveying givenness and newness: a cross-cultural developmental study. *J Verb Learn Verb Behav* 1978, 17: 539–558.
64. Bock JK. Meaning, sound, and syntax: lexical priming in sentence production. *J Exp Psychol Learn Mem Cogn* 1986, 12:575–586.
65. Igoa JM. The relationship between conceptualization and formulation processes in sentence production: some evidence from Spanish. In: Carreiras M, Garcia-Albea JE, Sebastián-Gallés N, eds. *Language Processing in Spanish*. Mahwah, NJ: Lawrence Erlbaum Associates; 1996, 305–351.
66. Ferreira VS, Dell GS. Effect of ambiguity and lexical availability on syntactic and lexical production. *Cogn Psychol* 2000, 40:296–340.
67. Jaeger TF, Wasow T. Processing as a source of accessibility effects on variation. In: Cover RT, Kim Y, eds. *Proceedings of the 31st Annual Meeting of the Berkeley Linguistic Society*. Ann Arbor: Sheridan Books; 2006, 169–180.
68. Kempen G, Harbusch K. A corpus study into word order variation in German subordinate clauses: animacy effects linearization independently of grammatical function assignment. In: Pechmann T, Habel C, eds. *Multidisciplinary Approaches to Language Production*. Berlin, Germany: Mouton De Gruyter; 2004, 173–181.
69. Tanaka MN. Branigan HP, Pickering MJ. Conceptual influences on word order and voice in Japanese sentence production (forthcoming).
70. Jaeger TF, Norcliffe E. The cross-linguistic study of sentence production: state of the art and a call for action. *Linguist Lang Comp* 2009, 3:866–887.
71. Jaeger TF. Redundancy and reduction: speakers manage syntactic information density. *Cogn Psychol* 2010, 61:23–62.
72. Jaeger TF. *Redundancy and Syntactic Reduction in Spontaneous Speech*. Unpublished PhD thesis, Stanford: Stanford University; 2006.
73. Bever T. The cognitive basis for linguistic structures. In: Hayes JR, ed. *Cognition and the Development of Language*. John Wiley & Sons; 1970, 279–362.
74. Kimball J. Seven principles of surface structure parsing in natural language. *Cognition* 1973, 2:15–47.
75. Frazier L, Fodor J. The sausage machine: a new two-stage parsing model. *Cognition* 1978, 6:291–325.
76. Frazier L. Syntactic complexity. In: Dowty DR, Karttunen L, Zwicky AM, eds. *Natural Language Parsing: Psychological, Computational, and Theoretical Perspectives*. Cambridge: Cambridge University Press; 1985, 129–189.
77. Pickering MJ, Van Gompel RPG. Syntactic parsing. In: Gaskell G, ed. *Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press; 2006.
78. Ferreira VS. Ambiguity, accessibility, and a division of labor for communicative success. *Psychol Learn Motiv* 2008, 49:209–246.
79. Jaeger TF, Corpus-based research on language production: information density and reducible subject relatives (submitted).
80. Altmann GTM, Kamide Y. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 1999, 73:247–264.
81. Altmann GTM, Kamide Y. Discourse-mediation of the mapping between language and the visual world: eye movements and mental representation. *Cognition* 2009, 111:55–71.
82. Garnsey SM, Perlmutter NJ, Meyers E, Lotocky MA. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *J Mem Lang* 1997, 37:58–93.
83. Kamide Y, Altmann GTM, Haywood SL. The time-course of prediction in incremental sentence processing: evidence from anticipatory eye movements. *J Mem Lang* 2003, 49:133–156.
84. MacDonald MC, Pearlmutter NJ, Seidenberg MS. Lexical nature of syntactic ambiguity resolution. *Psychol Rev* 1994, 191:676–703.
85. Ni W. Sidestepping garden paths: assessing the contributions of syntax, semantics and plausibility in resolving ambiguities. *Lang Cogn Process* 1996, 11:283–334.
86. Staub A, Clifton CJ. Syntactic prediction in language comprehension: evidence from either... or. *J Exp Psychol Learn Mem Cogn* 2006, 32:435–436.
87. Trueswell JC, Tanenhaus MK, Kello C. Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *J Exp Psychol Learn Mem Cogn* 1993, 19:528–553.
88. Trueswell JC, Tanenhaus MK, Garnsey SM. Semantic influence on syntactic processing: use of thematic information in syntactic disambiguation. *J Mem Lang* 1994, 33:265–312.

89. Trueswell J, Tanenhaus M. Toward a lexical framework of constraint-based syntactic ambiguity resolution. *Perspect Sent Process* 1994, 155–179.

90. Hale J. A probabilistic Earley parser as a psycholinguistic model. *North American Chapter of the Association for Computational Linguistics (NAACL)*. Morristown, NJ: Association for Computational Linguistics; 2001, 1–8.

91. Hale J. The information conveyed by words in sentences. *J Psycholinguist Res* 2003, 32:101–123.

92. Jurafsky D. A probabilistic model of lexical and syntactic access and disambiguation. *Cogn Sci* 1996, 20: 137–194.

93. Levy R. Expectation-based syntactic comprehension. *Cognition* 2008, 106:1126–1177.

94. Narayanan S, Jurafsky D. A Bayesian model predicts human parse preference and reading times in sentence processing. *Advances in Neural Information. Processing Systems (NIPS) 14: Proceedings of the 2002 Conference*.

95. Levy R. *Probabilistic Models of Word Order and Syntactic Discontinuity*. Stanford University; 2005.

96. Frank SL. Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In: Taatgen NA, van Rijn H, eds. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society; 2009, 1139–1144.

97. Hale J. Uncertainty about the rest of the sentence. *Cogn Sci Multidiscip J* 2006, 30:643–672.

98. Levy R, Bicknell K, Slattery T, Rayner K. Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proc Natl Acad Sci* 2009, 106:21086–21090.

99. Demberg V, Keller F. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 2008, 109:193–210.

100. Boston M, Hale J, Kliegl R, Patil U, Vasishth S. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *J Eye Movement Res* 2008, 2:1–12.

101. McDonald SA, Shillcock RC. Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vis Res* 2003, 43: 1735–1751.

102. Jurafsky D. Probabilistic modeling in psycholinguistics: linguistic comprehension and production. In: Bod R, Hay J, Jannedy S, eds. *Probabilistic Llinguistics*. Cambridge, MA: MIT Press; 2003, 39–95.

103. McClelland J, Rumelhart D. An interactive activation model of the effect of context on language learning (part 1). *Psychol Rev* 1981, 88:375–401.

104. Shriberg E, Stolcke A. Word predictability after hesitations: a corpus-based study. *Proceedings of ICSLP 96*; 1996.

105. Schnadt MJ, Corley M. The influence of lexical, conceptual and planning based factors on disfluency production. *28th Annual Conference of the Cognitive Science Society*. Vancouver, BC; 2006.

106. Cook SW, Jaeger TF, Tanenhaus MK. Producing less preferred structures: more gestures, less fluency. In: Taatgen NA, Rijn H. v., eds. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society; 2009, 62–67.

107. Cook SW, Jaeger TF, Tanenhaus MK. Producing less preferred structures: what happens when speakers take the road less traveled (submitted).

108. Tily H, Gahl S, Arnon I, Snider N, Kothari A, Bresnan J. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Lang Cogn* 2009, 1: 147–165.

109. Aylett MP, Turk A. The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang Speech* 2004, 47: 31–56.

110. Bell A, Jurafsky D, Fosler-Lussier E, Girand C, Gregory M, Gildea D. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *J Acoust Soc Am* 2003, 113:1001–1024.

111. Bell A, Brenier JM, Gregory M, Girand C, Jurafsky D. Predictability effects on durations of content and function words in conversational English. *J Mem Lang* 2009, 60:92–111.

112. Gahl S, Garnsey SM. Knowledge of grammar, knowledge of usage: syntactic probabilities affect pronunciation variation. *Language* 2004, 80:748–775.

113. Gahl S, Garnsey SM, Fisher C, Matzen L. "That sounds unlikely": phonetic cues to garden paths. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*; 2006.

114. Shannon C. A mathematical theory of communications. *Bell Syst Tech J* 1948, 27:623–656.

115. Genzel D. Charniak E. Entropy rate constancy in text. *Proceedings of ACL-2002*. Philadelphia; 2002, 199–206.

116. van Son R, Pols LCW. How efficient is speech? *Proceedings of the Institute of Phonetic Sciences*, vol. 25, Amsterdam; 2003, 171–184.

117. Aylett MP, Turk A. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *J Acoust Soc Am* 2006, 119:3048.

118. van Son R, van Santen JPH. Duration and spectral balance of intervocalic consonants: a case for effcient communication. *Speech Commun* 2005, 47:464–484.

119. Levy R, Jaeger TF. Speakers optimize information density through syntactic reduction. In: Schlökopf B, Platt J, Hoffman T, eds. *Advances in Neural Information*

Processing Systems (NIPS). vol. 19 Cambridge, MA: MIT Press; 2007, 849–856.

120. Jaeger TF, Kidd C. *Toward a Unified Model of Redundancy Avoidance and Strategic Lengthening*. Paper presented at the 21st Annual CUNY CUNY Conference on Human Sentence Processing, Chapel Hill, NC; 2008.
121. Bybee JL. *Morphology: A Study of the Relation between Form and Meaning*. Amsterdam: John Benjamins; 1985.
122. Bybee JL. Mechanisms of change in grammaticization: The role of frequency. *The Handbook of Historical Linguistics*. Oxford: Blackwell; 2003, 602–623.
123. Bybee JL. *Frequency of Use and the Organization of Language*. New York: Oxford University Press; 2007.
124. Bybee JL, Hopper P. Introduction to frequency and the emergence of linguistic structure. In: Bybee JL, Hopper P, eds. *Frequency and the emergence of linguistic structure*. Amsterdam and Philadelphia: John Benjamins Publishing Company; 2001, 1–24.
125. Hopper PJ, Traugott EC. *Grammaticalization*. Cambridge: Cambridge University Press; 1993.
126. Bybee J, Hopper P. 2001. Introduction to frequency and the emergence of linguistic structure. *Frequency and the Emergence of Linguistic Structure*, 1–24.
127. Croft W. Lexical rules vs. constructions: A false dichotomy. In: Cuyckens H, Berg T, Dirven R, Panther K, eds. *Motivation in Language: Studies in Honour of Guenter Radden, Studies in the Theory and History of Linguistic Science, Series 4*. Amsterdam: J. Benjamins; 2003, 49–68.
128. Haspelmath M. Creating economical morphosyntactic patterns in language change. In: Good J, ed. *Language universals and language change*. Oxford: Oxford University Press; 2008, 185–214.
129. Haspelmath M. Frequency vs. iconicity in explaining grammatical asymmetries. *Cogn Linguist* 2008, 19: 1–33.
130. van Son R, Pols LCW. How efficient is speech? In: *Proceedings of the Institute of Phonetic Sciences*. Amsterdam; 2003, 25:171–184.
131. Cohen Priva U. Using information content to predict phone deletion. In: Abner N, Bishop J, eds. *Proceedings of the 27th West Coast Conference on Formal Linguistics*. Somerville, MA: Cascadilla; 2008, 90–98.
132. Frank A, Jaeger TF. Speaking Rationally: Uniform Information Density as an Optimal Strategy for Language Production. *The 30th Annual Meeting of the Cognitive Science Society (CogSci08);* 2008, 933–938.
133. Wasow T, Jaeger TF, Orr D. Lexical variation in relativizer frequency. In: Wiese H, Simon H, eds. *Proceedings of the Workshop on Expecting the Unexpected: Exceptions in Grammar at the 27th Annual Meeting of the German Linguistic Association*. Germany: University of Cologne; 2006.
134. Genzel D. Charniak. E. Variation of entropy and parse trees of sentences as a function of the sentence number *Proceedings of the Conference on Empirical Methods in Natural Language Processing. Sapporo*; 2003, 65–72.
135. Gomez Gallo C, Jaeger TF, Smyth R. Incremental Syntactic Planning across Clauses. *Proceedings of the 30th Annual Meeting of the Cognitive Science Society (CogSci08)*; 2008, 845–850.
136. Qian T, Jaeger TF. Evidence for Efficient Language Production in Chinese. *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*; 2009.
137. Qian T, Jaeger TF. Entropy profiles in language: a cross-linguistic investigation (submitted).
138. Bates EA, MacWhinney B. Functionalist approaches to grammar. In: In WE, Gleitman LR, eds. *Language Acquisition: The State of the Art*. Cambridge: Cambridge University Press; 1982, 173–218.
139. Tomasello M. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press; 2003.
140. Lieven E, Behrens H, Speares J, Tomasello M. Early syntactic creativity: a usage-based approach. *J Child Lang* 2003, 30:333–370.
141. Diessel H. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas Psychol* 2007, 25:104–123.
142. Wells J, Christiansen M, Race D, Acheson D, MacDonald M. Experience and sentence processing: statistical learning and relative clause comprehension. *Cogn Psychol* 2009, 58:250–271.
143. Rosenbach A, Jäger G. Priming and unidirectional language change. *Theor Linguist* 2008, 34:85–113.
144. Zipf GK. *The Psychobiology of Language: An Introduction to Dynamic Philology*. Boston, MA: Houghton Mifflin; 1935.
145. Ferrer i Cancho R, Solé RV. Least effort and the origins of scaling in human language. *Proc Natl Acad Sci* 2003, 100:788.
146. Ferrer i Cancho R, Riordan O, Bollobas B. The consequences of Zipf's law for syntax and symbolic reference. *Proc R Soc Lond B* 2005, 272:561–565.
147. Gasser M. The origins of arbitrariness in language. *Annual Conference of the Cognitive Science Society* 2004, 26:1–6.
148. Piantadosi ST, Tily HJ, Gibson E. *The Communicative Lexicon Hypothesis*. Paper presented at the 22nd Annual CUNY Conference on Human Sentence Processing; 2009.
149. Piantadosi ST, Tily HJ, Gibson E. The communicative function of ambiguity in language (submitted).
150. Graff P, Jaeger TF. *Locality and Feature Specificity in OCP Effects: Evidence from Aymara, Dutch, and Javanese*. Paper presented at the Proceedings of the

Main Session of the 45th Meeting of the Chicago Linguistic Society, Chicago, IL; 2009.

151. Kuperman V, Ernestus M, Baayen R. Frequency distribution of uniphones, diphones, and triphones in spontaneous speech. *The J Acoust Soc Am* 2008, 124:3897–3908.

152. Ferrer i Cancho R. Zipf's law from a communicative phase transition. *Eur Phys J B* 2005, 47:449–457.

153. Gildea D, Temperley D. *Optimizing Grammars for Minimum Dependency Length*. Paper presented at the Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics, Prague, Czech Republic; 2007.

154. Perfors A. Simulated evolution of language: a review of the field. *J Artif Soc Social Simul* 2002, 5. Available from: http://jass.soc.surrey.ac.uk/5/1/4/.html.

155. Christiansen MH, Kirby S. Language evolution: consensus and controversies. *Trends Cogn Sci* 2003, 7: 300–307.

156. Steels L. How to do experiments in artificial language and why. In: Cangelosi A, Smith A, Smith J, eds. *Proceedings of the 6th International Conference on The Evolution of Language (EVOLANG6)*. London: World Scientific Publishing; 2006, 323.

157. Van Everbroeck E. Language type frequency and learnability from a connectionist perspective. *Linguist Typol* 2003, 7:1–50.

158. Kirby S. *Function, Selection, and Innateness: The Emergence of Language Universals*. USA: Oxford University Press; 1999.

159. Christiansen MH, Devlin JT. Recursive inconsistencies are hard to learn: a connectionist perspective on universal word order correlations. *Nineteenth Annual Conference of the Cognitive Science Society*. Stanford University: Lawrence Erlbaum Associates; 1997, 113.

160. Galantucci B. An experimental study of the emergence of human communication systems. *Cogn Sci* 2005, 29:737–767.

161. Griffiths T, Christian B, Kalish M. Using category structures to test iterated learning as a method for identifying inductive biases. *Cogn Sci* 2008, 32:68–107.

162. Fay N, Garrod S, Roberts L. The fitness and functionality of culturally evolved communication systems. *Phil Trans B* 2008, 363:3553.

163. Reali F, Griffiths T. The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition* 2009, 111: 317–328.

164. Kirby S, Cornish H, Smith K. Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc Natl Acad Sci* 2008, 105:10681–10686.

165. Goldin-Meadow S, So WC, Ozyurek A, Mylander C. How speakers of different languages represent events nonverbally. *Proc Natl Acad Sci* 2008, 105: 9163–9168.

166. Saffran J, Aslin R, Newport E. Statistical learning by 8-month-old infants. *Science* 1996, 274.

167. Wonnacott E, Newport E, Tanenhaus M. Acquiring and processing verb argument structure: distributional learning in a miniature language. *Cogn Psychol* 2008, 56:165–209.

168. Christiansen MH. Using artificial language learning to study language evolution: exploring the emergence of word order universals. In: Dessalles JL, Ghadakpour L, eds. *The Evolution of Language: 3rd International Conference*. Paris, France: Ecole Nationale Superieure des Telecommunications; 2000, 45–48.

FURTHER READING

For readers interested in more in-depth overviews of work on language processing, we recommend the following survey articles.

Syntactic production: Bock JK, Levelt WJM. Language Production: Grammatical Encoding. In: Gernsbacher MA, ed. *Handbook of Psycholinguistics*. San Diego: Academic Press; 1994, 945–984.

Word production: Griffin ZM, Ferreira VS. Properties of spoken language production. In: Gernsbacher MA, Traxler M, Eds., *Handbook of Psycholinguistics*. San Diego: Elsevier Academic Press; 2006, 21–59.

The role of sentence processing in language acquisition: Trueswell JC, Gleitman LR Learning to parse and its implications for language acquisition. In: Gaskell MG, ed. *Handbook of Psycholinguistics* Oxford: Oxford University Press; 2007, 635–656.

