

Linguistics: The Garden and the Bush*

Joan Bresnan

I want to thank the ACL for the Lifetime Achievement Award of 2016. I am deeply honored, and I share this honor with the outstanding collaborators and students I have been lucky to have over my lifetime.

The title of my talk describes two fields of linguistics, which differ in their approaches to data and analysis and in their fundamental concepts. What I call **the garden** is traditional linguistics, including generative grammar. In the garden linguists primarily analyze what I call “cultivated” data—that is, data elicited or introspected by the linguist—and form qualitative generalizations expressed in symbolic representations such as syntactic trees and prosodic phrases. What I am calling **the bush** could also be called “the wilderness.” In the bush they collect “wild data,” spontaneously produced by speakers, and form quantitative generalizations based on concepts such as conditional probability and information content.

- The garden:
 - qualitative generalizations
 - cultivated data
 - syntactic trees, prosodic phrases
- The bush:
 - quantitative generalizations
 - wild data
 - conditional probability, information content

My talk will recount the path I have taken from the linguistic garden into the bush.

1. MIT Years: Into the Garden

I came into linguistics with a bachelor’s degree in philosophy from Reed College, and after a couple of false starts I ended up in grad school at MIT studying formal grammar in the Department of Linguistics and Philosophy with Chomsky and Halle in the heyday of generative grammar. At MIT, Chomsky was my doctoral advisor, and my mentor was Morris Halle, who ran the Department at that time.

* This article contains the text of my acceptance speech for the ACL Lifetime Achievement Award in 2016. It uses with permission some material from Bresnan (2011). I prepared the plots of wild data with R (Sarkar 2008; R Core Team 2015). E-mail: Bresnan@stanford.edu.

The exciting goal was to infer the nature of the mind's capacity for language from the structure of human language, *viewed as a purely combinatorial set of formal patterns*, like the formulas of symbolic logic. It was apparently exciting even to Fred Jelinek as an MIT doctoral student in information theory ten years before me (Jelinek 2009). In his Lifetime Award speech he recounts how as a grad student he attended some of Chomsky's lectures with his wife, got the "crazy notion" that he should switch from information theory to linguistics, and went as far as discussing it with Chomsky when his advisor Fano got wind of it and said he had to complete his Ph.D. in Information Theory. He had no choice. The rest is history.¹

The MIT epistemology held that *the structure of language could not be learned inductively from what we hear; it had to be deduced from innate, universal cognitive structures specific to human language*. This approach had methodological advantages for a philosophical linguist: First, a limitless profusion of data in our own minds came from our intuitions about sentences that we had never heard before; second, a sustained and messy relationship to the world of facts and data was not required; and third, (with the proper training) scientific research could conveniently be done from an armchair using introspection.

2. Psychological (Un)reality

I got my Ph.D. from MIT in 1972 and taught briefly at Stanford and at UMass, Amherst, before joining the MIT faculty in 1975 as an Associate Professor of Linguistics. Very early on in my career as a linguist I had become aware of discrepancies between the MIT transformational grammar models and the findings of psycholinguists. For example, the theory that more highly transformed syntactic structures would require more complex processing during language comprehension and development did not work.

With a year off on a Guggenheim fellowship (1975–1976), I began to think about designing a more psychologically realistic system of transformational grammar that made much less use of syntactic transformations in favor of an enriched lexicon and pragmatics. The occasion was a 1975 symposium jointly sponsored by MIT and AT&T to assess the past and future impact of telecommunications technology on society, in celebration of the centennial of the invention of the telephone. What did I know about any of this? Absolutely nothing. I was invited to participate by Morris Halle. From Harvard Psychology, George Miller invited Eric Wanner, Mike Maratsos, and Ron Kaplan.

Ron Kaplan and I developed our common interests in relating formal grammar to computational psycholinguistics, and we began to collaborate. In 1977 we each taught courses at the IV International Summer School in Computational and Mathematical Linguistics, organized by Antonio Zampoli at the Scuola Normale Superiore, Pisa. In 1978 Kaplan visited MIT and we taught a joint graduate course in computational psycholinguistics. From 1978 to 1983, I consulted at the Computer Science Laboratory, Xerox Corporation Palo Alto Research Center (1978–1980) and the Cognitive and Instructional Sciences Group, Xerox PARC (1981–1983).

¹ Fred Jelinek, a pioneer in the application of statistical methods to automatic speech recognition and machine translation, became notorious among linguists for his (possibly apocryphal) comment: "Every time I fire a linguist, the performance of the speech recognizer goes up."

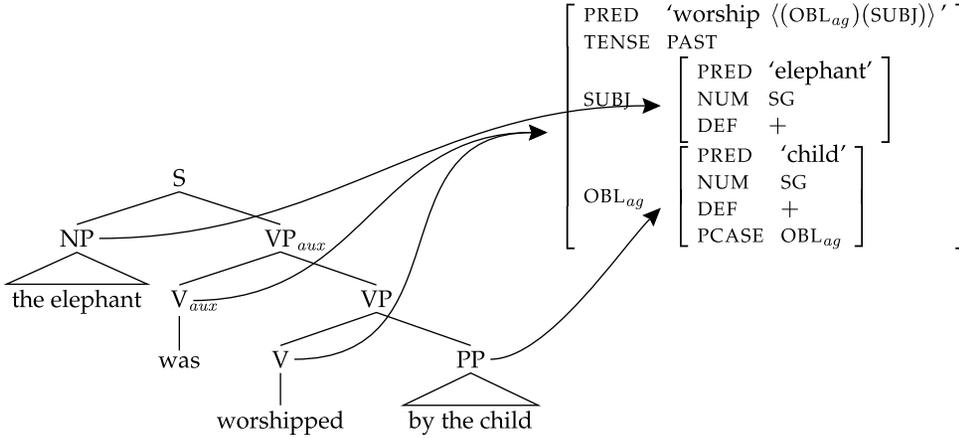


Figure 1 Lexical-Functional Grammar is a dual model providing a surfacy c(onstituent)-structure (on the left) directly mapped to an abstract f(unctional)-structure (on the right).

3. Lexical-Functional Grammar

During the 1978 fall semester at MIT we developed the LFG formalism (Kaplan and Bresnan 1982; Dalrymple et al., 1995). **Lexical-functional grammar** was a hybrid of augmented recursive transition networks (Woods 1970; Kaplan 1972)—used for computational psycholinguistic modeling of relative clause comprehension (Wanner and Maratsos 1978)—and my “realistic” transformational grammars, which offloaded a huge amount of grammatical encoding from syntactic transformations to the lexicon and pragmatics (Bresnan 1978) (see Figure 1).

As often noted, the LFG functional structures can be directly mapped to dependency graphs (Mel’cuk 1988; Carroll, Briscoe, and Sanfilippo 1998; King et al. 2003; Sagae, MacWhinney, and Lavie 2004; de Marneffe and Manning 2008) (see Figure 2). Some early statistical NLP parsers such as the Stanford Parser were dual-structure models like LFG with dependency graphs labeled by grammatical functions replacing LFG f-structures (de Marneffe and Manning 2008).

A key idea in LFG is that both active and passive argument structures are lexically stored (or created by bounded lexical rules) (see Figure 3). Independent evidence for lexical storage is that passive verbs undergo lexical rules of word-formation (Bresnan 1982b; Bresnan et al. 2015); surface features of passive SUBJ(ects) are retained (e.g., in tag questions). All relation-changing transformations are re-analyzed lexically in this way: passive, dative, raising, *there*-insertion, etc., etc.

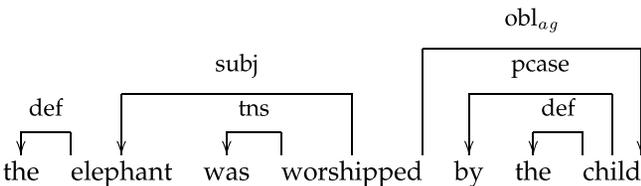


Figure 2 Dependency graph corresponding to the f-structure in Figure 1.

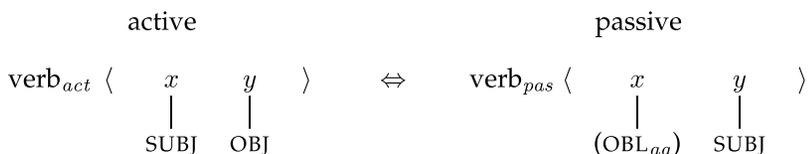


Figure 3
Active and passive lexical forms in LFG.

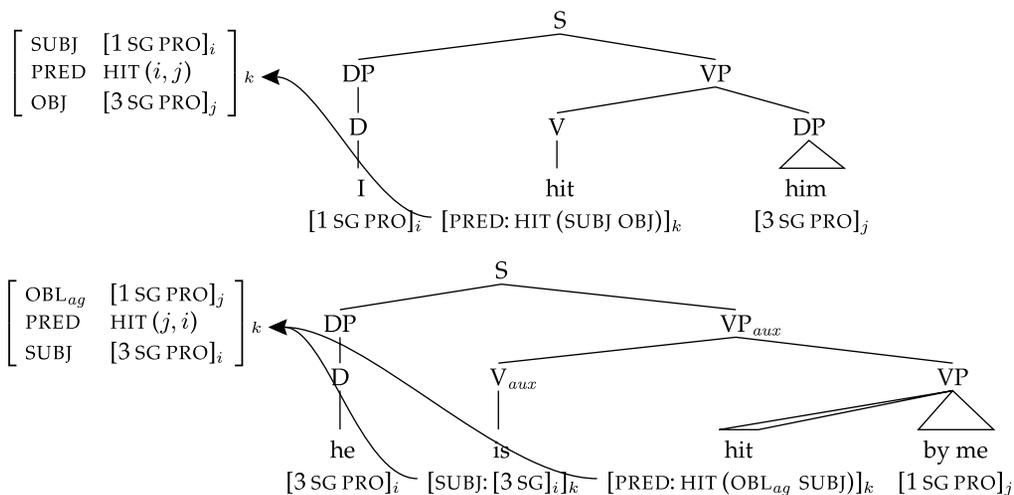


Figure 4
C-structure to f-structure mappings of active and passive sentences.

Figure 4 shows how the lexical features of active and passive terminal strings are mapped into the appropriate predicate–argument relations. The respective subjects of the upper and lower c-structure trees are first and third person singular pronouns, which give rise to the f-structures indexed *i*. The lexical forms are, respectively, active and passive, mapping each subject f-structure to the appropriate argument role of the verb *hit*.

We soon involved a highly productive group of young researchers in linguistics and psychology (initially represented in Bresnan, 1982a). The original group included Steve Pinker as a young postdoc from Harvard, Marilyn Ford as an MIT postdoc from Australia, Jane Grimshaw then teaching at Brandeis, and doctoral students at MIT and Harvard: Lori Levin, K. P. Mohanan, Carol Neidle, Avery Andrews, and Annie Zaenen.

LFG’s declarative, non-procedural design made it easily embeddable in what we then considered more realistic theories of the dynamics of sentence production (Ford 1982), comprehension (Ford, Bresnan, and Kaplan 1982; Ford 1983), and language development (Pinker 1984).

4. Stanford and Becoming an Africanist

In 1982 I took a sabbatical leave from MIT at the Center for the Study of Behavioral Sciences at Stanford, where I also spent time at Xerox PARC nearby.

At Stanford, the Center for the Study of Language and Information (CSLI) was being launched with a grant from the System Development Corporation. I decided to stay in California by joining the Stanford Linguistics Department and CSLI the following year, half-time. From 1983–1992, I worked the other half of my time as a member of the Research Staff, Intelligent Systems Laboratory, Xerox Corporation Palo Alto Research Center, which John Seely Brown headed during that time. LFG was considered useful in some computational linguistics applications such as MT.

In 1984 a Malaŵian linguist on a Fulbright Fellowship came to Stanford as a visitor. I had corresponded with him a few years earlier when I was a young faculty member at MIT and he had received his Ph.D. from the University of London. We both had lexical syntactic inclinations, and he had evidence from the Bantu language Chicheŵa, one of the major languages of Malaŵi. His name was Sam Mchombo.

Sam Mchombo arrived at Stanford just as CSLI was forming, and we began to collaborate on the problems of analyzing the (sooo cool) linguistic properties of Chicheŵa in the LFG framework: Chicheŵa has 18 genders (but not masculine/feminine), tone morphemes, pronouns incorporated into verb morphology, relation changes all expressed by verb stem suffixation (which undergo derivational morphology), configurational discourse functions, ...!

One of Sam’s ideas was that the Chicheŵa object marker, prefixed to the verb stem, functions syntactically as a full-blooded object pronoun. In LFG, the formal analysis is simple (see Figure 5). The analysis has rich empirical motivation in phonology, morphology, syntax, discourse, and language change (Bresnan and Mchombo 1987, 1995). Similar analyses have been explored in many disparate languages (e.g., Austin and Bresnan 1996; Bresnan et al. 2015). This and other work by many colleagues and students on a wide variety of languages helped to establish LFG as a flexible and well-developed linguistic theory useful to typologists and field linguists.

I wrote National Science Foundation grant proposals for successive projects with Sam and other Bantuists including Katherine Demuth and Lioba Moshi, and in the summer of 1986 did field work in Tanzania. In Tanzania I took time out to celebrate my birthday by hiking up Mt. Kilimanjaro with a group of young physicians from Europe doing volunteer work in Africa.

There I was, out of the armchair and onto the hard-packed dirt floor of a thatched home in a village on the slopes of Mt. Kilimanjaro, being served the eyeball of a freshly slaughtered young goat, which I was told was a particular honor normally reserved for men. ... *But I was still in the garden of linguistics:*

- The garden:
 - qualitative generalizations

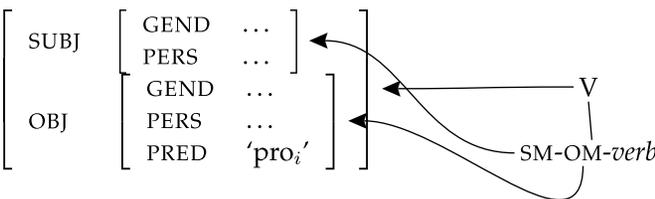


Figure 5 Analysis of the verb-incorporated object pronoun in Chicheŵa.

- cultivated data
- syntactic trees, prosodic phrases

Two intellectual shocks caused me eventually to leave the garden and completely change my research paradigm.

5. The Shock of Constraint Conflict

The first shock was the discovery that universal *principles of grammar may be inconsistent and conflict with each other*. The expressions of a language are not those that perfectly satisfy a set of true and universal constraints or rules, but are those that may violate some constraints in order to satisfy other more important constraints, optimizing constraint satisfaction. This insight came into linguistics from outside the field, from neural network approaches to cognition (Prince and Smolensky 1997). Yet as my former student Jane Grimshaw pointed out, we can see traces of it everywhere, even in corners of English syntax that had seemed exception-ridden.

In response to these ideas, I began in the mid-1990s to work out how to do optimality-theoretic (OT) syntax using LFG as the representational basis (e.g., Bresnan 2000). OT-style constraint ranking in large-scale LFG grammars was adopted in standard LFG parsing systems for ambiguity management (Frank et al. 1998; Kaplan et al. 2004; King et al. 2004). And Jonas Kuhn (2001, 2003) solved general computational problems of generation and parsing for OT syntax with LFG representations.

Ranked or weighted constraints on f-structures can capture unusual morphosyntactic generalizations. For example, suppose that *the grammatical function hierarchy has to align with the person hierarchy* (Aissen 1999):

$$\begin{array}{lcl} \text{SUBJ} & \prec & \text{non-SUBJ} \\ 1^{st}, 2^{nd} \text{ person} & \prec & 3^{rd} \text{ person} \end{array} \quad (1)$$

Imposing the person-alignment constraints on f-structures either requires or prohibits passivization, depending on who does what to whom. Compare Example (2) and Figure 6.

passive optional:	<i>He hits him</i>	<i>He is hit by him</i>	(2)
passive obligatory:	<i>*He hit me</i>	<i>I am hit by him</i>	
active obligatory:	<i>I hit him</i>	<i>*He is hit by me</i>	

An event in which the first person hits someone referred to in the third person must be described in the active voice of *hit*, not the passive, because the passive subject would be higher on the function hierarchy than the non-SUBJ but lower on the person hierarchy, contrary to the aligned hierarchies in Example (1). But if the third person hits the first person, the passive is obligatory, because the active voice sentence would violate the alignment of hierarchies. There are person-driven passives like this in Lummi (Salish, British Columbia), Picuris (Tanoan, New Mexico), and Nootka (Wakashan, British Columbia)—all unrelated languages.

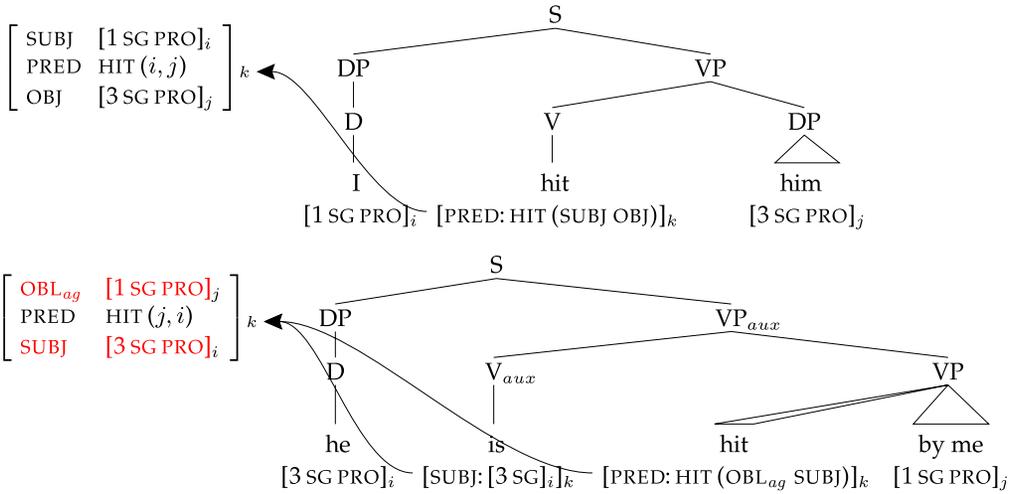


Figure 6
 Illustration of a person-driven passive. F-structure constraints filter out disharmonic alignments of person with grammatical function (in red).

6. Grammars Hard and Soft

My former student Chris Manning (who wrote his Linguistics doctoral dissertation at Stanford under my supervision) joined the Stanford faculty as Assistant Professor of Computer Science and Linguistics in 1999, and I began to attend his lectures and meet with him to discuss research. In studies of English using the Switchboard corpus, we found that English has soft, statistical shadows of hard person constraints in other languages: For example, it has person-driven active/passive alternations (Bresnan, Dingare, and Manning 2001), and person-driven dative alternations (Bresnan and Nikitina 2009). As Bresnan, Dingare, and Manning (2001) observe, “The same categorical phenomena which are attributed to hard grammatical constraints in some languages continue to show up as statistical preferences in other languages, motivating a grammatical model that can account for soft constraints.”

Examples of such models include stochastic OT (Bresnan, Deo, and Sharma 2007; Maslova 2007; Bresnan and Nikitina 2009), maximum-entropy OT (Goldwater and Johnson 2003; Gerhard Jaeger 2007), random fields (Johnson and Riezler 2003), data-oriented parsing (Bod and Kaplan 2003; Bod 2006), and other exemplar-based theories of grammar (Hay and Bresnan 2006; Walsh et al. 2010).

7. Into the Bush

In addition to Chris Manning, another person who helped me go into the bush was Harald Baayen. I first met Harald while attending a 2003 LSA workshop on Probability Theory in Linguistics. The presenters included Harald, Janet Pierrehumbert (with her student Jen Hay), Chris Manning, and others. As I watched the presenters give graphic visualizations of quantitative data showing dynamic linguistic phenomena, I thought, “I want to do that!”

My golden opportunity was the English dative alternation. Many English ditransitive verbs appear in alternative dative constructions, with the recipient realized

Downloaded from http://direct.mit.edu/col/article-pdf/14/2/4/599/1807741/col_a_00260.pdf by guest on 08 August 2022

as a **dative PP**—a prepositional *to*-phrase—or as the first of two noun phrases—the **dative NP**. Spontaneously produced alternations occur:

- a. “You don’t know how difficult it is to find something which will please everybody—especially the men.”
 “Why not just **give them cheques?**” I asked.
 “You can’t **give cheques to people**. It would be insulting.”
- b. “You **carrying a doughnut to your aunt** again this morning?” J.C. sneered. Shelton nodded and turned his attention to a tiny TV where “Hawaii Five-O” flickered out into the darkness of the little booth.
 “Looks like you **carry her some breakfast** every morning.”
- (3)

Which of these alternative constructions is used depends on multiple and often conflicting syntactic, informational, and semantic properties.

Bresnan et al. (2007) collected and manually annotated 2,360 instances of dative NP or PP constructions from the (unparsed) Switchboard Corpus (Godfrey et al. 1992) and another 905 from the Treebank *Wall Street Journal* corpus, using the manually parsed Treebank corpora (Marcus et al. 1993) to discover the set of dative verbs. With Harald Baayen’s help we fit a series of generalized linear and generalized linear mixed-effect models to the data. I read several books on statistical modeling that Harald recommended to me; I learned R, the programming language and environment for computational statistics (R Core Team 2015), which he also highly recommended (stressing that his name is *R*. Harald Baayen). Then I was able to show that under bootstrapping of speaker clusters and cross-validation, the models were highly accurate.

Figure 7 illustrates a quantitative generalization that emerged from the models: The paired parameter estimates for the recipient and theme have opposite signs. I dubbed this phenomenon **quantitative harmonic alignment** after the qualitative harmonic alignment in syntax studied by Aissen (1999) and others in the framework of Optimality Theory. Figure 8 provides a qualitative schematic depiction of the quantitative phenomena. The hierarchies of discourse accessibility, animacy, definiteness, pronominality, and weight are aligned with the initial/final syntactic positions of the postverbal arguments across constructions. In LFG, the linear order follows from alignment with the hierarchy of grammatical functions in *f*-structure (Bresnan and Nikitina 2009).

Interestingly, there are hard animacy constraints on dative as well as genitive alternations in some languages (Rosenbach 2005; Bresnan 2007a; Rosenbach 2008), and even hard weight constraints (O’Connor, Maling, and Skarabela 2009). The facts suggest that these constraints play a role in syntactic typology and should not be brushed off as external to grammar and out of bounds to the theoretical linguist.

In principle, these quantitative harmonic alignment effects could be formulated as constraints on LFG *f*-structures. Annie Zaenen saw how this kind of work might be useful for paraphrase analysis to improve generation, and she got us involved with Mark Steedman and colleagues at Edinburgh on an animacy annotation project (Zaenen et al. 2004).

8. Data Shock

The second intellectual shock that pushed me further into the bush was realizing that *in the garden, we had been relying all along on inconsistent binary grammaticality judgments*

$$\text{logit}[\text{Probability}(\text{Response} = 1)] = \mathbf{X}\beta$$

where $X\hat{\beta} =$

$$\begin{aligned} &0.95 - 1.34\{c\} + 0.53\{f\} - 3.90\{p\} + 0.96\{t\} \\ &\quad +0.99\{\text{accessibility of recipient} = \text{nongiven}\} \\ &\quad -1.1\{\text{accessibility of theme} = \text{nongiven}\} \\ &+1.2\{\text{pronominality of recipient} = \text{nonpronoun}\} \\ &\quad -1.2\{\text{pronominality of theme} = \text{nonpronoun}\} \\ &\quad +0.85\{\text{definiteness of recipient} = \text{indefinite}\} \\ &\quad -1.4\{\text{definiteness of theme} = \text{indefinite}\} \\ &\quad +2.5\{\text{animacy of recipient} = \text{inanimate}\} \\ &\quad -1.2 \cdot \text{length difference (log scale)} < 0 \text{ [len(rec) > len(th)]} \\ &\quad -1.2 \cdot \text{length difference (log scale)} > 0 \text{ [len(rec) < len(th)]} \\ &\quad +0.48\{\text{person of rec} = \text{nonlocal}\} \\ &\quad -0.03\{\text{number of recipient} = \text{plural}\} \\ &\quad +0.5\{\text{number of theme} = \text{plural}\} \\ &\quad -0.46\{\text{concreteness of theme} = \text{nonconcrete}\} \\ &\quad -1.1\{\text{parallelism} = 1\} \end{aligned}$$

and $\{c\} = 1$ if subject is in semantic group c , 0 otherwise; likewise for f, p, t .

Figure 7

Quantitative harmonic alignment in a logistic regression model of the dative alternation: Paired parameter estimates for recipient and theme have opposite signs; positive signs (highlighted in red) favor dative PP construction, negative (highlighted in blue) favor dative NP.

that can be manipulated by changing the probabilities of the contexts, and we had vastly underestimated the human language capacity.

At the suggestion of Jeff Ellman, I used the dative corpus model to measure the predictive power of English language users (Bresnan 2007b). Inspired by Anette Rosenbach’s (2003) beautiful experiment on the genitive alternation, and with her advice, I made questionnaires asking participants to rate the naturalness of contextualized alternative dative constructions sampled from our dative data set, by allocating 100 points between the alternatives (see Figure 9).

In one set of task instructions I asked participants to rate the choices in accordance with their own intuitions; in another I asked them to guess what the original speaker actually said in the discourse excerpt, and to rate their confidence in their guess in the same way, splitting 100 points between the alternatives. The findings were similar: As the log odds of a PP dative construction increased, the ratings of each participant showed a linear increase as well. *The participants could tell which dative construction the original speaker was going to use, and their own ratings matched the corpus probabilities* (see Figure 10). This finding has been replicated across speakers of other varieties of English (Bresnan and Ford 2010).

discourse given < not given
 animate < inanimate
 definite < indefinite
 pronoun < non-pronoun
 less weight < more weight

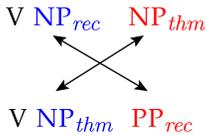


Figure 8

Qualitative view of quantitative harmonic alignment: The construction is chosen that best aligns property hierarchies with the initial/final syntactic positions of the arguments across constructions. Properties in red or blue favor syntactic positions in red or blue, respectively.

Speaker:

About twenty-five, twenty-six years ago, my brother-in-law showed up in my front yard pulling a trailer. And in this trailer he had a pony, which I didn't know he was bringing. And so over the weekend I had to go out and find some wood and put up some kind of a structure to house that pony,

- (1) *because he brought the pony to my children.*
- (2) *because he brought my children the pony.*

Figure 9

Item from experiment in Bresnan (2007a).

A second experiment reported in Bresnan (2007b) used linguistic manipulations that raise or lower probability to see whether they influence grammaticality judgments. Certain semantic classes of verbs reported by linguists to be ungrammatical in the double object construction are nevertheless found in actual usage (Bresnan et al. 2007). For example, *whisper* is reported to be ungrammatical in the double object construction, but Internet queries yield *whisper me the answer*, along with *whisper the password to the fat lady*. The double object context with the pronoun recipient is more harmonically aligned and far more probable. The reportedly ungrammatical examples constructed by linguists tend to utilize the far less probable positionings of argument types, like *whisper the fat lady the answer*. In the experiment I found that *participants rated the reportedly ungrammatical constructions in the more probable contexts higher than the reportedly grammatical constructions in the less probable contexts*.

I began to find that throughout the grammar, linguistic reports of ungrammaticality had been greatly exaggerated (Bresnan 2007a). Our linguistic intuitions of what is ungrammatical may merely reflect our implicit knowledge of what is highly improbable (see also Manning 2003).

9. Probabilistic Syntactic Knowledge

The simplifying assumption in the garden of linguistic theory has been that speakers' knowledge of their language is characterized by a static, categorical system of grammar.

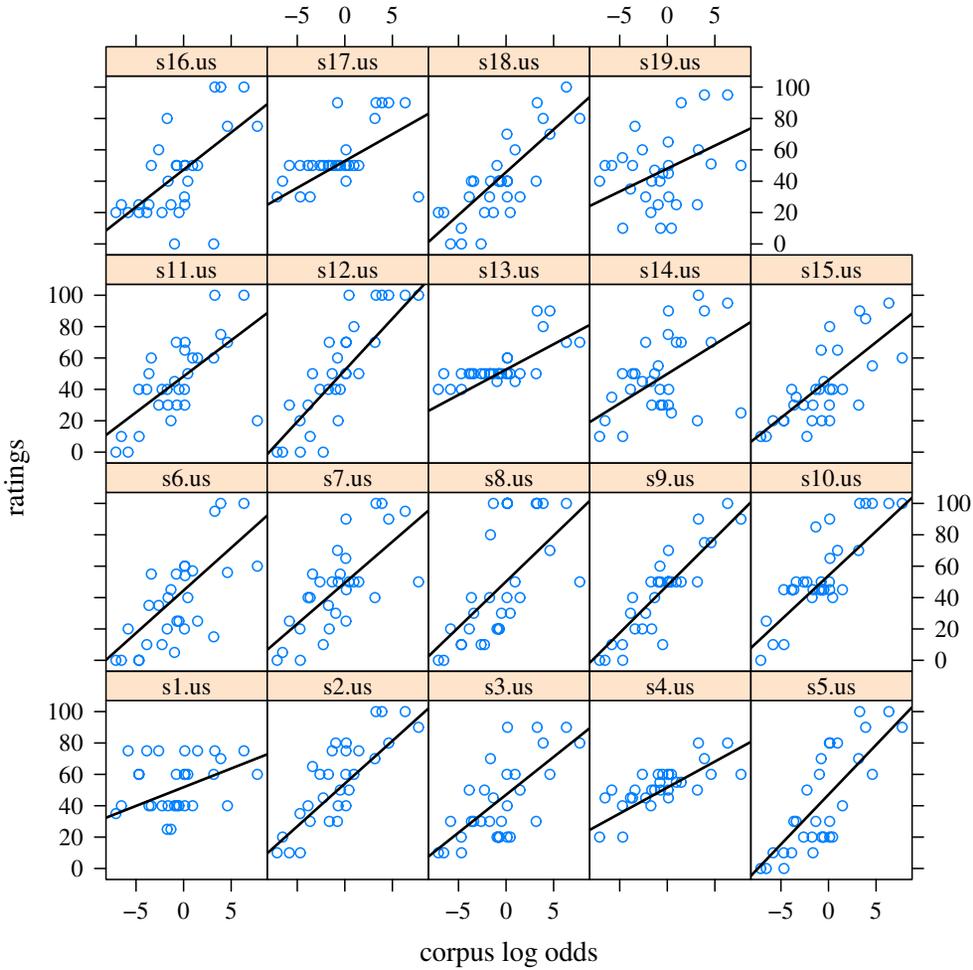


Figure 10 Plots showing correlations between the Switchboard corpus model log odds of the dative PP and individual participant’s ratings of the naturalness of the dative PP in context.

Although this has been a fruitful idealization, I came to see that it ultimately underestimates human language capacities. My own research showed that language users can match the probabilities of linguistic features of the environment and they have powerful predictive capabilities that enable them to anticipate the variable linguistic choices of others. Therefore, the working hypothesis I have adopted contrasts strongly with that of the garden: Grammar itself is inherently variable and probabilistic in nature, rather than categorical and algebraic.

With collaborators, I began to look for implicit knowledge of syntactic probabilities in various areas of linguistics—in phonetic reflexes of construction frequencies and probabilities (Hay and Bresnan 2006; Tily et al. 2009; Kuperman and Bresnan 2012); in language development (de Marneffe et al. 2012; Van den Bosch and Bresnan 2015); in language change (Wolk et al. 2013; Szmrecsányi et al. 2014); and in comparative syntactic variation across varieties of English (Bresnan and Ford 2010; Ford and Bresnan 2015).

Downloaded from http://direct.mit.edu/col/article-pdf/142/4/599/1807741/col_a_00260.pdf by guest on 08 August 2022

10. Deeper into the Bush

My research program is taking me even deeper into the bush, as I will now illustrate with several visualizations of wild data.

In (tensed) verb contraction, the verbs *is*, *are*, *am*, *has*, *have*, *will*, *would* lose all but their final segments, orthographically represented as 's, 're, 'm, 's, 've, 'll, 'd, and form a unit with the immediately preceding word, called the **host**. In Example (4) the host is bolded:

*all this **blood's** pouring out the side of my head* (4)

Previous corpus studies point to informativeness (or closely related concepts) as an important predictor of verb contraction (Frank and Jaeger 2008; Bresnan and Spencer 2012; Barth and Kapatsinski 2014). Informativeness derives from predictability: More predictable events are less informative, and can even be redundant (Shannon 1948). The less predictable the host–verb combination, the more informative it is, and the less likely to contract.

In a corpus, the information content of a word *B* given the context *C* is the average negative log conditional probability of *B* over all instances of *C* (Piantadosi, Tily, and Gibson 2011). We define the **informativity of a host-verb bigram *B*** as in Equation (5):

$$-\frac{1}{N} \sum_{i=1}^N \log P(B = \langle \text{host}, \text{verb} \rangle | C = \text{next}_i) \quad (5)$$

where *B* is the bigram consisting of a host and a verb in either contracted or uncontracted form (for example, *blood's* or *blood is*), *N* is the total frequency of *B*, and *next_i* is the following context word for the *i*th occurrence of *B* in the corpus.

Figure 11 shows the strong inverse relation between verb contraction and informativity of host–verb bigrams for the verb *is/s* in data from the Canterbury Corpus of New Zealand English (Gordon et al. 2004), which I collected with Jen Hay in the summer and fall of 2015.² Jen and I discussed implications of the informativity of verb contraction. As vocabulary richness grows, local word combinations become less predictable. Average predictability across contexts is what makes a host–verb combination more or less informative. Hence, *increases in vocabulary richness would lead to increased informativity of host–verb combinations, potentially causing dynamic changes in verb contractions over time*. These implications led me to ask whether children decrease their use of verb contractions in periods of increasing vocabulary richness during language development—a question never before asked, as far as I could tell from the literature.

It was not difficult to answer my question. I had already collected verb contraction data from longitudinal corpora in CHILDES (MacWhinney 2000).³ The CHILDES corpora are an invaluable resource. They have been morphologically analyzed,

2 The NZ data consist of 11,719 instances of variable *is* contraction from 412 speakers, collected with the assistance of Vicky Watson, and *n*-grams from the entire Canterbury Corpus (*n* = 1,087,113 words).

3 The CHILDES data were collected in the summer of 2015 with Arto Anttila and the assistance of Gwynn Lyons from the corpora described by Brown (1973), Clark (1978), Demetras (1986), Kuczaj (1977), Sachs (1983), and Suppes (1974). The data include all full and contracted instances of the verb *is* (*n* = 56,088) by children (19,888 instances) and their parents (26,592) as well as others present.

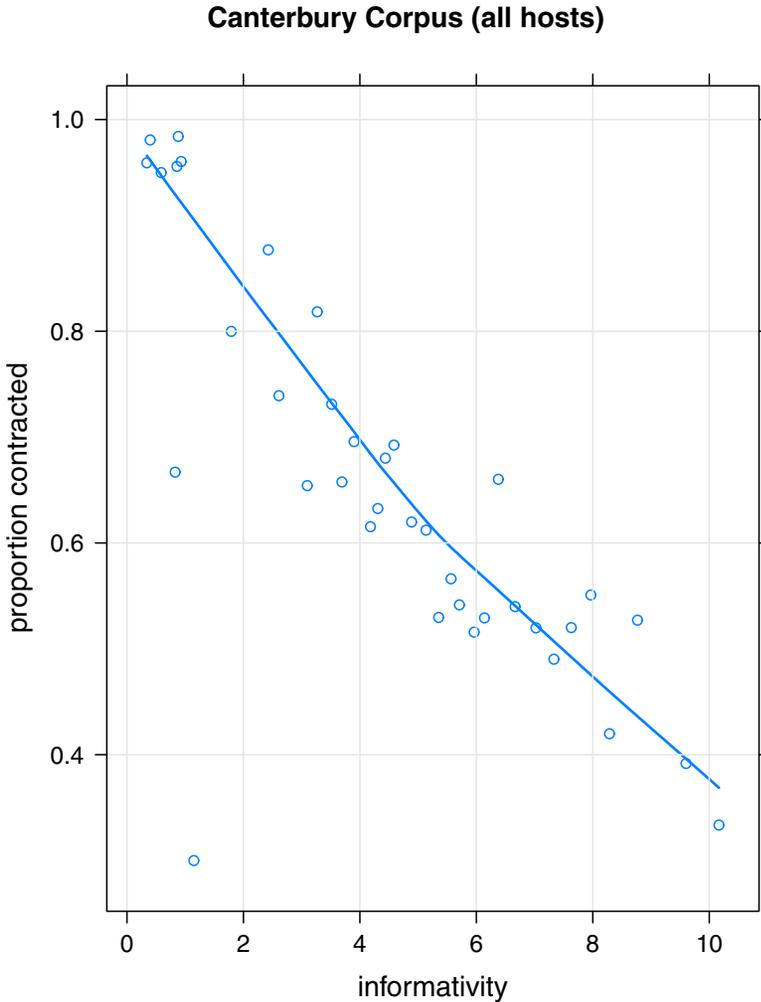


Figure 11
Proportions of *is* contractions per informativity bin in the Canterbury Corpus (both lexical and pronoun hosts).

automatically parsed, and manually checked. As a matter of fact, the syntactic parses use dependency graphs derived from LFG functional structure relations (Sagae, MacWhinney, and Lavie 2004; Sagae, Lavie, and MacWhinney 2005). Computational tools are provided, including the CLAN VOCD tool (MacWhinney 2015), which calculates vocabulary richness based on a sophisticated algorithm for averaging morpheme or lemma counts (lemmas being the distinct words disregarding inflections).

I used the VOCD tool to calculate the lemma diversity of all the language produced by each child at each recording session in the longitudinal corpora we selected. The panel plots in Figure 12 show increasing vocabulary richness as age increases between about 20 and 60 months, with the exception of one child (“Adam”), whose vocabulary richness was already high at the time of sampling and shows a dip midway in the recordings.

From the relations among verb contraction, informativity, and vocabulary richness we would expect children’s subject–verb contractions to decrease as their vocabulary

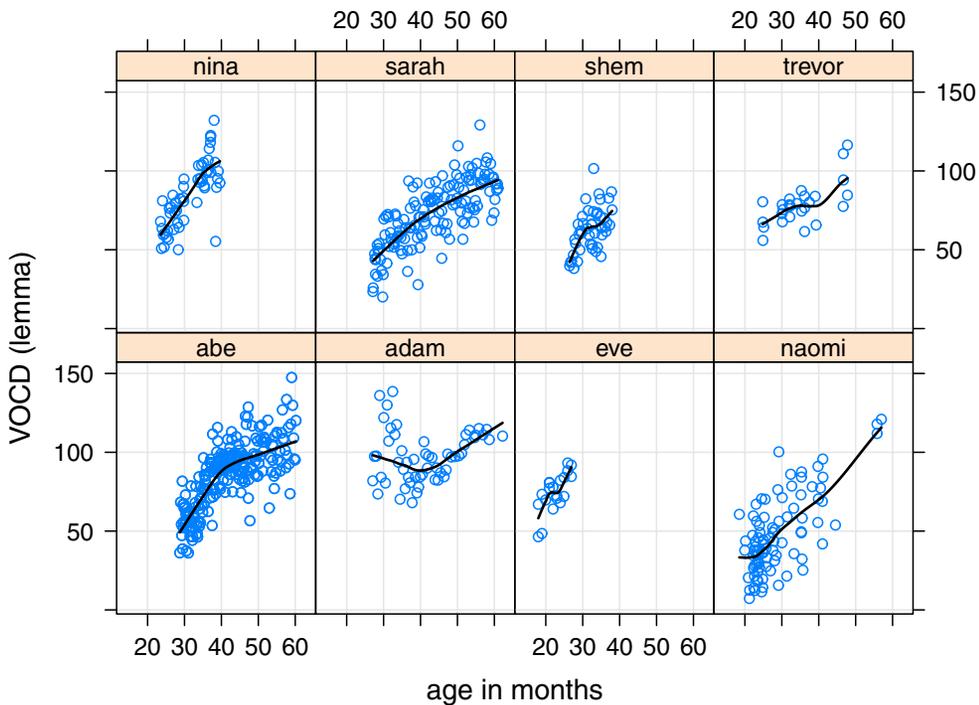


Figure 12
 Vocabulary richness (VOCD lemma diversity) by age by child in selected CHILDES corpora.

richness increases. Figure 13 shows that this expectation may be met: Overall, the children’s contractions are tending to decrease, and only one child shows an increase in contraction during the period from 20 to 60 months: Adam, the child whose vocabulary richness shows the dip in Figure 12.

How do the dynamics of contraction in the children’s data compare to their parents’ contractions in conversational interactions during the same periods? Presumably the parents themselves would not be experiencing rapid vocabulary growth during this period, so they would not show a decline in verb contractions for that reason. Figure 14 shows the aggregated *is* contraction data by children vs. parents for all host-*is*’s bigrams and for the frequent bigrams *what is*’s and *Mommy is*’s. Strikingly, the children in the aggregate show declines in the proportion of contractions while their parents’ proportions remain constant.

These data visualizations appear to support a major cognitive role for implicit knowledge of syntactic probability. But they also raise many, many questions for further research—a good indicator of their potential fertility. This will be the focus of my next research project.

I am aware that this use of information theory characterizes just one small prosodic island in English syntax, the contractible host–verb sequence. It does make me wonder whether the entirety of hierarchical syntactic structure could somehow reflect a vast topography of overlapping peaks and valleys of informativeness, requiring parallel computations on multiple scales—something that one of *you* perhaps could figure out how to do.

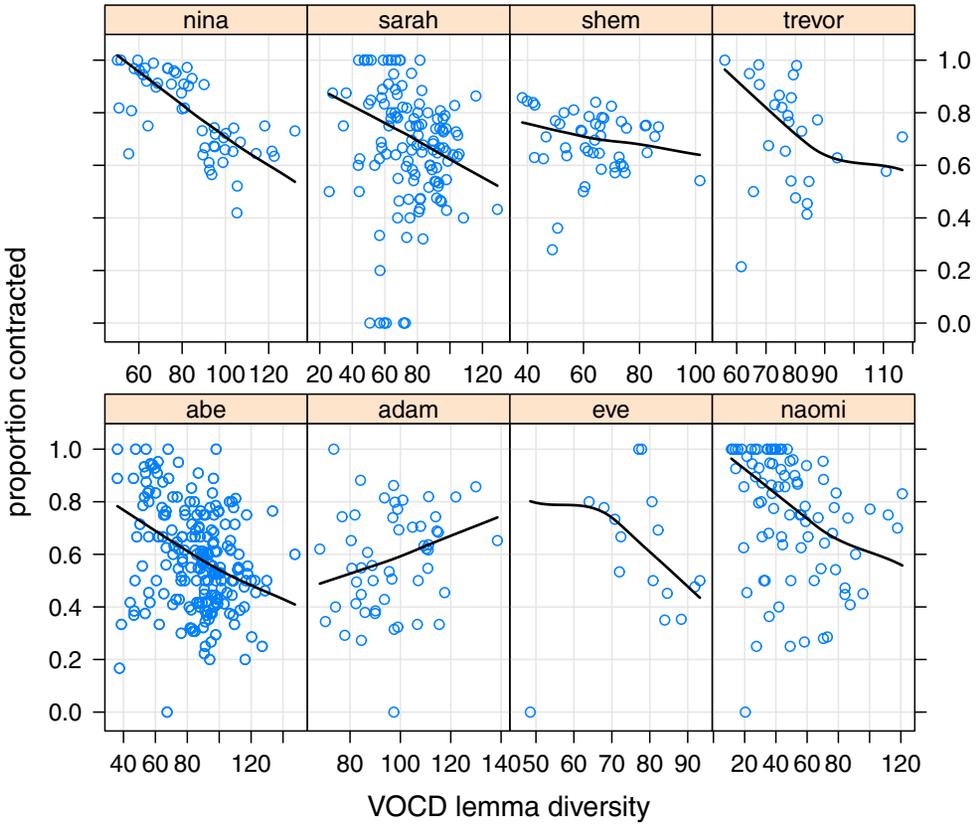


Figure 13
 Each child’s proportion of *is* contractions during the recorded sessions measuring vocabulary richness (VOCD lemma diversity).

11. Looking Forward

My work now has essentially the same exciting goal as when I started my study of linguistics at MIT: to infer the nature of the mind’s capacity for language from the structure of human language. But now I know that linguistic structure is quantitative as well as qualitative, and I can use methods that I have been learning in the bush —

- The bush:
 - quantitative generalizations
 - wild data
 - conditional probability, information content

What I hope to see going forward are increasingly powerful applications of computational linguistic theory, techniques, and resources to deepen our understanding of human language and cognition.

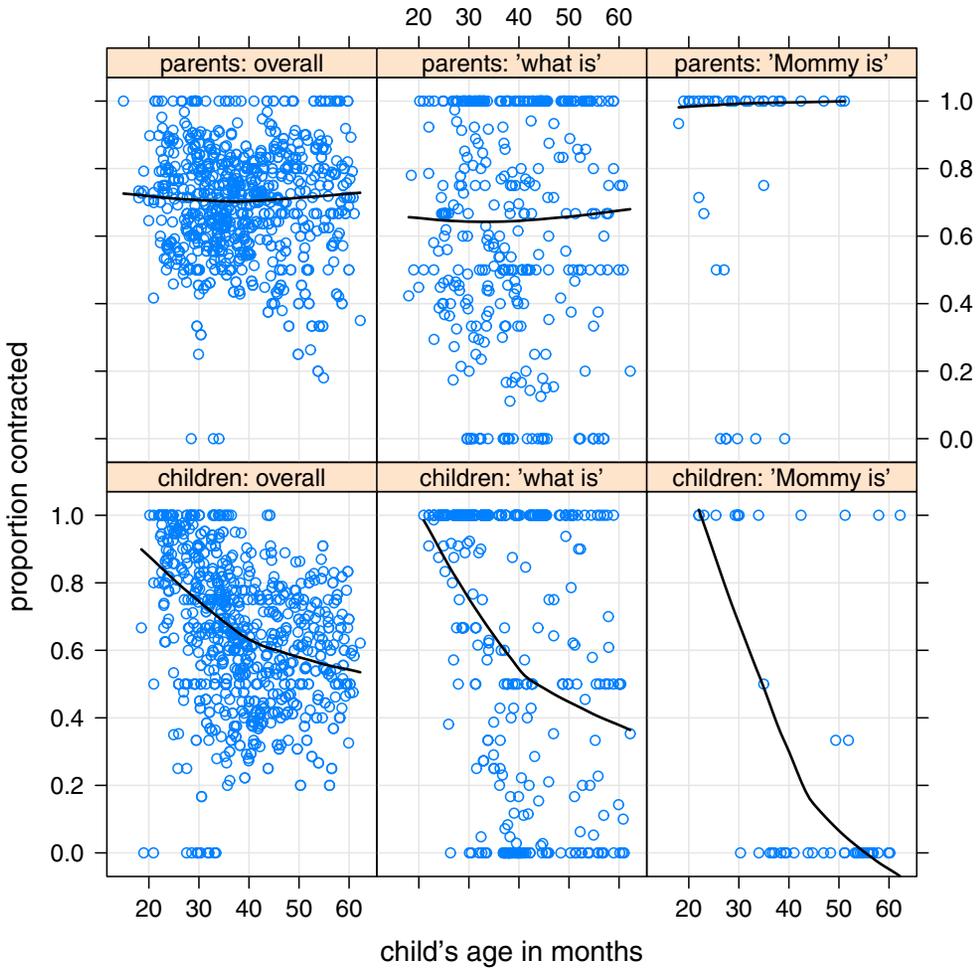


Figure 14
 Proportion of *is* contractions by child's age of recording in eight CHILDES corpora.

References

Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *Natural Language & Linguistic Theory*, 17(4):673–711.

Austin, Peter and Joan Bresnan. 1996. Nonconfigurality in Australian aboriginal languages. *Natural Language & Linguistic Theory*, 14:215–68.

Barth, Danielle and Vsevolod Kapatsinski. 2014. A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of *am*, *are* and *is*. *Corpus Linguistics and Linguistic Theory*. doi: 10.1515/cllt-2014-0022.

Bresnan, Joan. 1978. A realistic transformational grammar. In Morris Halle, Joan Bresnan, and George A. Miller, editors, *Linguistic Theory and Psychological Reality*. The MIT Press, Cambridge, MA, pages 1–59.

Bresnan, Joan, editor. 1982a. *The Mental Representation of Grammatical Relations*. The MIT Press, Cambridge, MA.

Bresnan, Joan. 1982b. The passive in lexical theory. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. The MIT Press, pages 3–86.

Bresnan, Joan. 2000. Optimal syntax. In Joost Dekkers, Frank van der Leeuw, and Jeroen van de Weijer, editors, *Optimality Theory: Phonology, Syntax, and Acquisition*. Oxford University Press, pages 334–385.

Bresnan, Joan. 2007a. A few lessons from typology. *Linguistic Typology*, 11(1):297–306.

- Bresnan, Joan. 2007b. Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In Sam Featherston and Wolfgang Sternefeld, editors, *Roots: Linguistics in Search of its Evidential Base*. Mouton de Gruyter Berlin, pages 75–96.
- Bresnan, Joan. 2011. Linguistic uncertainty and the knowledge of knowledge. In Roger Porter and Robert Reynolds, editors, *Thinking Reed. Centennial Essays by Graduates of Reed College*. Reed College, Portland, OR, pages 69–75.
- Bresnan, Joan, Ash Asudeh, Ida Toivonen, and Stephen Wechsler. 2015. *Lexical-functional Syntax, 2nd Edition*. John Wiley & Sons.
- Bresnan, Joan, Anna Cueni, Tatiana Nikitina, and R. Harald Baayen. 2007. Predicting the dative alternation. In G. Boume, I. Kraemer, and J. Zwarts, editors, *Cognitive Foundations of Interpretation*. Royal Netherlands Academy of Science, Amsterdam, pages 69–94.
- Bresnan, Joan, Ashwini Deo, and Devyani Sharma. 2007. Typology in variation: A probabilistic approach to *be* and *n't* in the Survey of English Dialects. *English Language and Linguistics*, 11(2):301–346.
- Bresnan, Joan, Shipra Dingare, and Christopher D. Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of the LFG01 Conference*, pages 13–32, Stanford, CA.
- Bresnan, Joan and Marilyn Ford. 2010. Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*, 86(1):168–213.
- Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chicheŵa. *Language*, 63(4):741–82.
- Bresnan, Joan and Samuel A. Mchombo. 1995. The lexical integrity principle: Evidence from Bantu. *Natural Language & Linguistic Theory*, 13:181–254.
- Bresnan, Joan and Tatiana Nikitina. 2009. The gradience of the dative alternation. In Linda Uyechi and Lian Hee Wee, editors, *Reality Exploration and Discovery: Pattern Interaction in Language and Life*. CSLI Publications, pages 161–184.
- Bresnan, Joan and Jessica Spencer. 2012. Frequency effects in spoken syntax: *Have* and *be* contraction. Paper presented at conference on New Ways of Analyzing Syntactic Variation, Radboud University Nijmegen, 15–17 November 2012.
- Brown, Roger. 1973. *A First Language: The Early Stages*. Harvard University Press.
- Carroll, John, Ted Briscoe, and Antonio Sanfilippo. 1998. Parser evaluation: A survey and a new proposal. In *Proceedings of the 1st International Conference on Language Resources and Evaluation*, pages 447–454, Granada.
- Clark, Eve V. 1978. Awareness of language: Some evidence from what children say and do. In *The Child's Conception of Language*. Springer, pages 17–43.
- Dalrymple, Mary, Ronald M. Kaplan, John T. Maxwell III, and Annie Zaenen, editors. 1995. *Formal Issues in Lexical-Functional Grammar*. CSLI Publications, Stanford, CA.
- De Marneffe, Marie-Catherine, Scott Grimm, Inbal Arnon, Susannah Kirby, and Joan Bresnan. 2012. A statistical model of the grammatical choices in child production of dative sentences. *Language and Cognitive Processes*, 27(1):25–61.
- De Marneffe, Marie-Catherine and Christopher D. Manning. 2008. The Stanford typed dependencies representation. In *COLING 2008: Proceedings of the Workshop on Cross-Framework and Cross-Domain Parser Evaluation*, pages 1–8, Manchester.
- Demetras, Martha Jo-Ann. 1986. Working parents' conversational responses to their two-year-old sons. Working paper, The University of Arizona.
- Ford, Marilyn. 1982. Sentence planning units: Implications for the speaker's representation of meaningful relations underlying sentences. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, pages 797–827.
- Ford, Marilyn. 1983. A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, 22(2):203–218.
- Ford, Marilyn and Joan Bresnan. 2015. Generating data as a proxy for unavailable corpus data: The contextualized sentence completion task. *Corpus Linguistics and Linguistic Theory*, 11(1):187–224.
- Ford, Marilyn, Joan Bresnan, and Ronald M. Kaplan. 1982. A competence-based theory of syntactic closure. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, pages 727–796.
- Frank, Anette, Tracy H. King, Jonas Kuhn, and John T. Maxwell III. 1998. Optimality Theory style constraint ranking in large-scale LFG grammars. In *Proceedings of the LFG98 Conference*, Brisbane.

- Frank, Austin and T. Florian Jaeger. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)*, pages 939–944, Washington, DC.
- Godfrey, John J., Edward C. Holliman, and Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *IEEE International Conference on Acoustics, Speech, and Signal Processing, 1992. ICASSP-92*, volume 1, pages 517–520, San Francisco, CA.
- Gordon, Elizabeth, Lyle Campbell, Jennifer Hay, Margaret MacLagan, Andrea Sudbury, and Peter Trudgill. 2004. *New Zealand English: Its Origins and Evolution*. Cambridge University Press.
- Hay, Jennifer and Joan Bresnan. 2006. Spoken syntax: The phonetics of *giving a hand* in New Zealand English. *The Linguistic Review*, 23(3):321–349.
- Jäger, Gerhard. 2007. Maximum entropy models and stochastic Optimality Theory. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Christopher Manning, editors, *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*. CSLI Publications, pages 467–479.
- Jelinek, Frederick. 2009. The dawn of statistical ASR and MT. *Computational Linguistics*, 35(4):483–494.
- Johnson, Mark and Stefan Riezler. 2002. Statistical models of syntax learning and use. *Cognitive Science*, 26(3):239–253.
- Kaplan, Ronald M. 1972. Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence*, 3:77–100.
- Kaplan, Ronald M. and Joan Bresnan. 1982. Lexical-functional Grammar: A formal system for grammatical representation. In Joan Bresnan, editor, *The Mental Representation of Grammatical Relations*. MIT Press, pages 173–281.
- Kaplan, Ronald M, Stefan Riezler, Tracy H. King, John T. Maxwell III, Alexander Vasserman, and Richard Crouch. 2004. Speed and accuracy in shallow and deep stochastic parsing. Technical report, Defense Technical Information Center Document.
- King, Tracy Holloway, Richard Crouch, Stefan Riezler, Mary Dalrymple, and Ronald M. Kaplan. 2003. The PARC 700 dependency bank. In *Proceedings of the EACL03: 4th International Workshop on Linguistically Interpreted Corpora (LINC-03)*, pages 1–8, Budapest.
- King, Tracy Holloway, Stefanie Dipper, Anette Frank, Jonas Kuhn, and John T. Maxwell III. 2004. Ambiguity management in grammar writing. *Research on Language and Computation*, 2(2):259–280.
- Kuczaj, Stan A. 1977. The acquisition of regular and irregular past tense forms. *Journal of Verbal Learning and Verbal Behavior*, 16(5):589–600.
- Kuhn, Jonas. 2001. Generation and parsing in Optimality-theoretic syntax — Issues in the formalization of OT-LFG. In Peter Sells, editor, *Formal and Empirical Issues in Optimality-theoretic Syntax*. CSLI Publications, pages 313–366.
- Kuhn, Jonas. 2003. *Optimality-theoretic Syntax: A Declarative Approach*. CSLI Publications.
- Kuperman, Victor and Joan Bresnan. 2012. The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory and Language*, 66(4):588–611.
- MacWhinney, Brian. 2000. *The CHILDES project: The database*, volume 2. Psychology Press.
- MacWhinney, Brian. 2015. The CHILDES project tools for analyzing talk—electronic edition, Part 2: The CLAN programs.
- Manning, Christopher D. 2003. Probabilistic syntax. In Stefanie Jannedy, Jennifer Hay, and Rens Bod, editors, *Probabilistic Linguistics*. MIT Press, pages 289–341.
- Marcus, Mitchell P., Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.
- Maslova, Elena. 2007. Stochastic OT as a model of constraint interaction. In Annie Zaenen, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Christopher Manning, editors, *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*. CSLI Publications, pages 513–528.
- Mel’cuk, Igor Aleksandrovic. 1988. *Dependency Syntax: Theory and Practice*. State University Press of New York.
- O’Connor, Catherine, Joan Maling, and Barbra Skarabela. 2013. Nominal categories and the expression of possession. In Kersti Börjars, David Denison, and Alan Scott, editors, *Morphosyntactic Categories and the Expression of Possession*. John Benjamins, pages 89–121.
- Piantadosi, Steven T., Harry Tily, and Edward Gibson. 2011. Word lengths are

- optimized for efficient communication. *Proceedings of the National Academy of Sciences, U.S.A.* 108(9):3526–3529.
- Pinker, Steven. 1984. *Language Learnability and Language Development*. Harvard University Press.
- Prince, Alan and Paul Smolensky. 1997. Optimality: From neural networks to universal grammar. *Science*, 275(5306): 1604–1610.
- R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Rosenbach, Anette. 2003. Aspects of iconicity and economy in the choice between the s-genitive and the of-genitive in English. In Günter Rohdenburg and Britta Mondorf, editors, *Determinants of Grammatical Variation in English*. Walter de Gruyter & Co., pages 379–412.
- Rosenbach, Anette. 2005. Animacy versus weight as determinants of grammatical variation in English. *Language*, 81(3):613–644.
- Rosenbach, Anette. 2008. Animacy and grammatical variation—Findings from English genitive variation. *Lingua*, 118(2):151–171.
- Sachs, Jacqueline. 1983. Talking about the there and then: The emergence of displaced reference in parent-child discourse. In K. E. Nelson, editor, *Children's Language*, volume 4, pages 1–284. Erlbaum, Hillsdale, NJ.
- Sagae, Kenji, Alon Lavie, and Brian MacWhinney. 2005. Automatic measurement of syntactic development in child language. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics*, pages 197–204, Ann Arbor, MI.
- Sagae, Kenji, Brian MacWhinney, and Alon Lavie. 2004. Automatic parsing of parental verbal input. *Behavior Research Methods, Instruments, & Computers*, 36(1):113–126.
- Sarkar, Deepayan. 2008. *Lattice: Multivariate Data Visualization with R*. Springer, New York.
- Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Report.
- Suppes, Patrick. 1974. The semantics of children's language. *American Psychologist*, 29(2):103.
- Szmrecsányi, Benedikt, Anette Rosenbach, Joan Bresnan, and Christoph Wolk. 2014. Culturally conditioned language change? A multi-variate analysis of genitive constructions in ARCHER. In Marianne Hundt, editor, *Late Modern English Syntax*. Cambridge University Press, pages 133–152.
- Tily, Harry, Susanne Gahl, Inbal Arnon, Neal Snider, Anubha Kothari, and Joan Bresnan. 2009. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition*, 1(2):147–165.
- Van den Bosch, Antal and Joan Bresnan. 2015. Modeling dative alternations of individual children. In *Proceedings of the Sixth Workshop on Cognitive Aspects of Computational Language Learning*, pages 103–112, Lisbon.
- Wanner, Eric and Michael Maratsos. 1978. An ATN approach to comprehension. In Morris Halle, Joan Bresnan, and George A. Miller, editors, *Linguistic Theory and Psychological Reality*. MIT Press, Cambridge, MA, pages 119–161.
- Wolk, Christoph, Joan Bresnan, Anette Rosenbach, and Benedikt Szmrecsányi. 2013. Dative and genitive variability in Late Modern English: Exploring cross-constructural variation and change. *Diachronica*, 30(3):382–419.
- Woods, William A. 1970. Transition network grammars for natural language analysis. *Communications of the ACM*, 13(10): 591–606.
- Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor, and Tom Wasow. 2004. Animacy encoding in English: Why and how. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, pages 118–125, Barcelona.
- Zaenen, Annie, Jane Simpson, Tracy Holloway King, Jane Grimshaw, Joan Maling, and Christopher Manning, editors. 2007. *Architectures, Rules, and Preferences: Variations on Themes by Joan W. Bresnan*. CSLI Publications, Stanford, CA.