JOAN BRESNAN

Stanford University & CSLI,

The Center for the Study of Language and Information

A simplifying assumption of linguistic theory: knowledge of language is characterized by *a static, categorical system of grammar*.

A simplifying assumption of linguistic theory: knowledge of language is characterized by *a static, categorical system of grammar*.

—A fruitful idealization, but it ultimately *underestimates human language capacities.*

A simplifying assumption of linguistic theory: knowledge of language is characterized by *a static, categorical system of grammar*.

—A fruitful idealization, but it ultimately *underestimates human language capacities.*

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.

A simplifying assumption of linguistic theory: knowledge of language is characterized by *a static, categorical system of grammar*.

—A fruitful idealization, but it ultimately *underestimates human language capacities.*

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, a strongly contrasting hypothesis: *grammar itself is inherently variable and stochastic in nature, rather than categorical and algebraic.*

- Low-level continuous phonetic variation is sensitive to high-level construction probabilities in speech production (Gahl & Garnsey 2004, Tily et al. 2009, Kuperman & Bresnan 2012)
- Construction probability is not merely a summary statistic for determinative individual factors affecting pronunciation, but has a direct effect in itself (Kuperman & Bresnan 2012)

Illustrations from

Victor Kuperman & Joan Bresnan (2012). The effects of construction probability on word durations during spontaneous incremental sentence production. *Journal of Memory & Language* 66: 588–611.

Kuperman & Bresnan studied the acoustic characteristics of the spontaneous speech production of the English dative alternation (*gave the book to the boy/ the boy the book*) as a function of the probability of the choice between alternating constructions.

Adjusted difference in monosyllabic function word durations at the start of the most and least probable syntactic alternatives

for V NP NP (*John wrote him a letter*): 109 ms.

and for V NP PP (*She gave the book to them*): 61 ms.

Related work:

Gahl, S., & Garnsey, S. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language* 80(4): 748–774.

Jaeger, R. F. (2006). *Redundancy and syntactic reduction in spontaneous speech*. Stanford: Stanford University Linguistics Department Ph.D. dissertation.

Levy, R., & Jaeger, T. (2007). Speakers optimize information density through syntactic reduction. *Proceedings of the twentieth annual conference on neural information processing systems*, pp. 29–37. Vancouver: NIPS.

Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., & Bresnan, J. (2009). Syntactic probabilities affect pronunciation variation in spontaneous speech. *Language and Cognition* 1(2): 147–165.

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

Illustration from

Joan Bresnan (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In *Roots: Linguistics in search of its evidential base, Series: Studies in Generative Grammar*, ed. by Sam Featherston and Wolfgang Sternefeld. Berlin and New York: Mouton de Gruyter, pp. 75–96.

Joan Bresnan & Marilyn Ford (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English, *Language* 86.1: 168–213.

Questionnaires asking participants to rate the naturalness of contextualized alternative dative constructions sampled from telephone conversations.

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



Related work:

Arnold, J., Wasow, T., Losongco, A., & Ginstrom, R. (2000). Heaviness vs. newness: The effects of complexity and information structure on constituent ordering. *Language* 76(1): 28–55.

Ford, M., & Bresnan, J. (2012). "They whispered me the answer" in Australia and the US: A comparative experimental study. In *From Quirky Case to Representing Space: Papers in Honor of Annie Zaenen*, ed. by Tracy Holloway King and Valeria de Paiva. Stanford: CSLI Publications.

Gries, S.T. (2003). Towards a corpus-based identification of prototypical instances of constructions. *Annual Review of Cognitive Linguistics* 1: 1–27.

MacDonald, M.C. (1999). Distributional information in language and acquisition: Three puzzles and a moral. *The emergence of language*, ed. by Brian MacWhinney, 177–96. Mahwah, NJ: Lawrence Erlbaum.

Rosenbach, A. (2003). Aspects of iconicity and economy in the choice between the *s*-genitive and the *of*-genitive in English. *Determinants of grammatical variation in English*, ed.. by Günter Rohdenburg and Britta Mondorf, 379–411. Berlin: Mouton de Gruyter.

Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language* 81(3): 613–44.

Theijssen, D. (2012) *Making Choices. Modeling the English Dative Alternation*. Nijmegen: Radboud University Centre for Language Studies Ph.D. dissertation.

A skeptical question

Paul M. Postal email to J. Bresnan 9/20/08:*

Dear Joan:

So I was looking forward to hearing you when I learned you were giving a talk at NYU this fall. This is then by way of an explanation of why I won't be there though. Since I live way north of the city, it takes me around three hours round trip and \$30 to arrive at NYU and return. Unfortunately, there is simply nothing in your abstract which suggests any connection to my own work or interests. It doesn't seem to have anything to do with syntax at all.

A skeptical question

(continued)

I am simply amazed that someone with your background, knowledge and abilities finds this kind of thing attractive, especially given how little seems to be understood in syntax even now.

But of course it is none of my affair what other people work on or find interesting.

Best wishes,

Paul

What does this have to do with syntax?

Grammar hard and soft

- reports of ungrammaticality have been greatly exaggerated (hard constraints softer than has been thought)
- soft constraints show up hardened elsewhere (soft constraints mirror hard constraints)

Grammar hard and soft

- reports of ungrammaticality have been greatly exaggerated (hard constraints softer than has been thought)
- soft constraints show up hardened elsewhere (soft constraints mirror hard constraints)

Illustrations from

Bresnan, J. 2007. A few lessons from typology. *Linguistic Typology* 11: 297–306.

Bresnan, J. & Nikitina, T. 2009. The gradience of the dative alternation. In *Reality Exploration and Discovery: Pattern Interaction in Language and Life*, edited by Linda Uyechi and Lian Hee Wee, 161–184. Stanford: CSLI Publications.

Green, G. 1971. Some implications of an interaction among constraints. *Chicago Linguistic Society* 7: 85–100.

The *NP Pron constraint is widely cited in English linguistics. Personal pronouns, but not demonstrative or indefinite pronouns, are avoided when following lexical NPs if both are objects:

```
Erteschik-Shir (1979: 452): Collins (1995: 39)
```

John gave it to Mary.

*John gave Mary it.

*Tom gave an aunt them.

```
Tom gave them to an aunt.
```

Kay (1996):

* She gave John it.

I gave John that.

She gave John one.

Harvested from usage (Bresnan & Nikitina 2009):

I don't give children peanut butter until they are 3 years old since it is recommended not to give children it to avoid possible allergies.

You should never give out your address or phone number online and you should never **send someone them** in the mail either.

Please follow these simple rules and teach your children them, however most dogs are friendly.

Second graders finished their underwater scenes and are very proud of these. They could not wait to **show their parents them** and can't wait to bring them home.

Harvested from usage (continued):

Per[c]eptions about God's absence are due to our lack of **showing people him** through our life.

Mega Blast beam: This is kakuri's strongest ki attack only he has what it takes to know how to use it he can **teach people it** but it takes at least 2 years

Reportedly nonalternating ditransitive verbs:

Ted denied Kim the opportunity to march. The brass refused Tony the promotion.

*Ted denied the opportunity to march to Kim. *The brass refused the promotion to Tony.

Green (1971):

Ted gave Joey permission to march, but he denied it to Kim.

The brass gave Martin permission to sit, but they denied it to Tony.

A conflicting constraint *NP Pron (Green 1971):

*The brass gave Martin permission to sit, but they denied Tony it.

*Ted gave Joey permission to march, but he denied Kim it.

The ***NP** Pron constraint overrides the syntactic biases of *deny, refuse*.

- When using or judging linguistic forms of expression, speakers instantaneously resolve multiple simultaneous conflicting constraints from pragmatics, discourse, grammar, and the lexicon.
- Because resolving constraint conflicts is a probabilistic activity of weighing multiple interacting variables, judgments of ungrammaticality are fundamentally based on probabilities, not categorical rules.

Many, many other cases, such as:

Jason Grafmiller. 2013. *The Semantics of Syntactic Choice: An analysis of English emotion verbs*. Stanford: Stanford University Linguistics Department Ph.D. dissertation.

Helge Lødrup. 2007. Norwegian anaphors without visible binders. *Journal of Germanic Linguistics* 19(1): 1–22.

Joanna Nykiel. 2013. Clefts and preposition stranding under sluicing. *Lingua* 123: 74–117.

Strunk, J. and Neal Snider. In press. Extraposition without subjacency?. *Rightward Movement from a Cross-linguistic Perspective*, ed. by Heike Walker, Gert Webelhuth, and Manfred Sailer. John Benjamins.

Grammar hard and soft

- reports of ungrammaticality have been greatly exaggerated (hard constraints softer than has been thought)
- soft constraints show up hardened elsewhere (soft constraints mirror hard constraints)

Illustrations from

Bresnan, J., S. Dingare, & C. Manning. 2001. Soft constraints mirror hard constraints: Voice and person in English and Lummi. In *Proceedings of the LFG '01 Conference, University of Hong Kong. Stanford: CSLI Publications*

Bresnan (2007), Bresnan and Nikitina (2009)

"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."

-Bresnan, Dingare, and Manning (2001)

Person alignment in actives and agentive passives—

SUBJECT \Leftarrow 1^{st} , 2^{nd} personNON-SUBJECT \Leftarrow 3^{rd} person

—a soft constraint in English, a hard constraint in Lummi, Salish (Bresnan, Dingare, and Manning 2001)

Person-driven passives in Lummi (also Picurís):

 $\mathbf{3} \rightarrow \mathbf{3}$: passive optional

xči-t-scə swəy?qə? cə swi?qo?ətknow-tr-3.tr.subjthe manthe boy'The man knows the boy'

xči-t-ŋcəswi?qo?ətəcəswəy?qə?know-tr-passtheboybytheman'The boy is known by the man'

1,2 \rightarrow 3: passive ungrammatical

xči-t=sən/=sx^w cə swəy?qə?

know-TR=1/2.SG.NOM the man

'l/you know the man'

*___ 'The man is known by me/you'

- $3 \rightarrow 1,2$: passive obligatory
- *__ 'The man knows me/you' *xči-t-ŋ=sən/=sx^w ə сə swəy?qə?* know-тк-разs=1/2.sg.ком by the man

'I am/you are known by the man'

Person alignment in ditransitives—

—a soft constraint in English, a hard constraint in Kanuri, Nilo-Saharan; also Bulgarian, Arabic, Georgian (Haspelmath 2004, Bresnan and Nikitina 2009)

Animacy alignment in spoken English dative constructions (Bresnan, Cueni, Nikitina, and Baayen 2007):

after controlling for possible confounds, inanimate recipients are over *five times* as likely to occur in dative PPs as animates in the Switchboard corpus of spontaneous spoken English.

Animacy determines word order *preferences* in English dative constructions, but has near-categorical effects on word order elsewhere:

- Shona and Sesotho (Hawkinson and Hyman 1974, Morolong and Hyman 1977)
- Spoken Eastern Armenian (Polinsky 1996)
- Mayali, Gunwinjguan (Evans 1997)
- many languages (Kittilä 2007a,b)

An example of softening (lowering probabilities) over time: The effect of animacy on the preposing of possessors growing weaker in Canadian English over time

the woman's shadow, the shadow of the building

the building's shadow

Jankowski 2009; see Mark Liberman's language blog "The genitive of lifeless things", October 11, 2009

Some other cases:

Givón, T. (1979) *On Understanding Grammar*. New York: Academic Press.

Haspelmath, Martin (2004) Explaining the Ditransitive Person-Role Constraint: A usage-based approach. *Constructions* 2.

Rosenbach, A. (2008) Animacy and grammatical variation—Findings from English genitive variation. *Lingua* 118(2): 151–71.

other cases (continued):

O'Connor, C., J. Maling, & B. Skarabela (2013) Nominal categories and the expression of possession A cross-linguistic study of probabilistic tendencies and categorical constraints. In *Morphosyntactic categories and the expression of possession*, edited by K. Börjars, D. Denison & A. Scott, pp. 89–122.

See also Bresnan (2007) and Bresnan & Nikitina (2009).

Models of Grammar Plasticity

- stochastic optimality theory
- variable rules
- maximum entropy OT grammars
- random fields
- exemplar theoretic models

Stochastic Optimality Theory



		*A	*B	A!			*B	*A	A!
R.	$cand_1$		*			$cand_1$	*		
	$cand_2$	*!			R\$	$cand_2$		*	

The Gradual Learning Algorithm

If cand₁ is correct, then when cand₂ is produced ...:



Result: Categorical data repel constraints A* and B*.

The Gradual Learning Algorithm

If cand₂ is correct, then when cand₁ is produced ...:



Result: Categorical data cause *A and B* to gradually rerank and then continue spreading apart.

The Gradual Learning Algorithm

If both $cand_1$ and $cand_2$ are correct outputs for the same input, then ...:



Result: Variable data attract/repel constraints *A and *B into an eventual holding pattern that matches the frequency of variation.

Grammar hard and soft



Partial stochastic grammar of Lummi:



A representational basis



Modeling 'style'

Boersma and Hayes (2001:83):

```
selectionPoint_i = rankingValue_i + styleSensitivity_i \cdot Style + noise
```

"Constraints with positive values for styleSensitivity take on higher ranking values in formal speech; constraints with negative values for styleSensitivity take on higher ranking values in casual speech, and constraints with zero values for styleSensitivity are style insensitive."

an entry point for individual variation

Variable rules

Cedergren & Sankoff 1974:*

Speech performances are here considered as statistical samples drawn from a probabilistic language competence. This competence is modeled in conventional generative terms, except that optional rules are assigned application probabilities as functions of the structure of the input strings, possibly depending on the extralinguistic environment as well.

*Henriette Cedergren & David Sankoff. 1974. Variable rules: performance as a statistical reflection of competence. *Language* 50(2): 333–355.

Toward variable constraints

- decompose rules into constraints on possible structures
- map the application probabilities into ranking weights
- add a bit of evaluation noise to the weights during generation
- modify the optimization function from summing to maximizing
- —and you get stochastic OT (Boersma 1998, Boersma & Hayes 2001, Maslova 2007)

The optimization family

You can drop the noise and vary the optimization function, as in maximum entropy OT grammars (Goldwater and Johnson 2003, Gerhard Jaeger 2007)

Random fields – Johnson and Riezler 2003

Illustration from

Jennifer Hay & Joan Bresnan (2006). Spoken syntax: The phonetics of *giving a hand* in New Zealand English. *The Linguistic Review* 23(3): 321–349..

Ongoing New Zealand English sound changes are centralizing vowels:

black widow pronounced by a New Zealander sounds to speakers of American English much like "bleck wuddow"

- /æ/ raised to [æ] or [ε]
- /I/ centralized to [I] or [H]

Hay & Bresnan (2006)

—studied the phonetics of the words *hand* and *give* in a spoken corpus of New Zealand English:

- the vowel in give is more likely to be centralized in the more frequent uses, the abstract uses of give (give me a hand, give her a chance), compared to concrete transfer uses (give us presents, give us a plate full of food).
- the vowel in hand is more likely to be raised in the more frequent use, when the word designates the limb

These findings relate to exemplar theories of grammar that store phonetically detailed instances of constructions.



Peter

 \mathbf{v}

hates

NP

Susan



John

 \mathbf{v}

likes

Rens Bod (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, 23(3): 291–320.

NP

Mary

- Root: the Root operation selects any node of a tree to be the root of the new subtree and erases all nodes except the selected node and the nodes it dominates.
- Frontier: the Frontier operation then chooses a set (possibly empty) of nodes in the new subtree different from its root and erases all subtrees dominated by the chosen nodes.



Figure 5. A subtree of the left tree in Figure 4

Figure 5. A subtree of the left tree in Figure 4

















Figure 9. Analyzing Mary likes Susan by combining subtrees

sentence. Then the probability of the derivation in Figure 9 is the *joint* probability of 3 stochastic events (see Bod 2003a for a linguistic introduction to elementary probability theory):

- selecting the subtree _S[NP _{VP}[_V[likes] NP]] among the subtrees with root label S,
- 2. selecting the subtree NP[Mary] among the subtrees with root label NP,

3. selecting the subtree NP[Susan] among the subtrees with root label NP. The probability of each event can be computed from the frequencies of the occurrences of the subtrees in the corpus. For instance, the probability of event 1 is computed by dividing the number of occurrences of the subtree $_{S}[NP VP[V[likes] NP]]$ by the total number of occurrences of subtrees with root label S: $\frac{1}{20}$.

Multilevel exemplar theory

Another computational theory of multilevel exemplar grammar: Walsh et al. 2010*

Components of the multilevel exemplar model

	Syllable Duration	Grammaticality		
Stimuli	Syllable to be produced	Phrase (in perception)		
Constituents	Segments	Words		
Constituent representation	Acoustics, duration	Word's left/right context		
Similarity of constituents	Sum of similarities of the components of the representation			
Units	Syllables	Phrases		
Unit representation	Sequence of constituents			
Similarity of units	Sum of similarities of the constituents of the units			
Property inferred	Duration of syllable	Grammaticality of novel phrase		

*Michael Walsh, Bernd Möbius, Travis Wade, Hinrich Schütze. 2010. Multilevel exemplar theory. *Cognitive Science* 34: 537–582.

Multilevel exemplar theory



Conclusion

Probabilistic/stochastic grammars and exemplar grammars

- model predictive capacities of language users
- show plasiticity with changing experience
- express the softness of 'hard' constraints in syntax
- can harden 'soft' constraints as a function of experience
- in principle can model individual variation