

## Soft constraints mirror hard constraints: voice and person in English and Lummi

Joan Bresnan, Shipra Dingare, and Christopher Manning  
(abstract for Lfg2001, February 2001)

The present study demonstrates that 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.

The effects of a hierarchy of person (1st, 2nd > 3rd) on grammar are categorical in some languages, most famously in languages with inverse systems (see Dahlstrom 1984 for an excellent LFG analysis), but also in languages with person restrictions on passivization. In Lummi, for example, the person of the subject argument cannot be lower than the person of a nonsubject argument. If this would happen in the active, passivization is obligatory; if it would happen in the passive, the active is obligatory (Jelinek and Demers 1983). These facts follow from the theory of harmonic alignment in OT: constraints favoring the harmonic association of prominent person (1st, 2nd) with prominent syntactic function (subject) are hypothesized to be present as subhierarchies of the grammars of all languages, but to vary in their effects across languages depending on their interactions with other constraints (Aissen 1999).

Lummi person/role by voice:		
	% Active	% Passive
Ag1,2/Pt3	100	0
Ag3/Pt1,2	0	100
Ag3/Pt3	(opt*)	(opt*)
Ag1,2/Pt1,2	100	0

(\*optional, frequency unknown)

In the present study we show that there is a statistical reflection of these hierarchies in English. The same disharmonic person/argument associations which are avoided categorically in languages like Lummi are found to depress or elevate the frequency of passives relative to actives in the SWITCHBOARD corpus of spoken English.

The percentage of passives in some written genres is very low (Rosenhouse 1988), but unfortunately in genres with higher passivization (such as scientific prose) the occurrence of first and second person pronouns is practically nonexistent. To obtain a full range of persons, we chose the parsed subset of the SWITCHBOARD corpus of spoken English telephone conversations (Godfrey et al. 1992, Marcus et al. 1993). In this corpus the overall rate of occurrences of passivizable transitive verbs with expressed agents is less than 1%. Nevertheless, when broken into the same person/role categories as Lummi, there is a highly significant influence of person/role on voice in the same directions as in Lummi ( $\chi^2 = 95.66, p < 0.0001$ ). In the table below one can see that, compared to the overall rate of passivization of 0.685%, the rate of passivization for 1st and 2nd person Agent combined with 3rd person Patient input is substantially depressed while that for the reverse input is substantially elevated:

English person/role by voice:		
	% Active	% Passive
Ag1,2/Pt3	99.98	0.02
Ag3/Pt1,2	97.2	2.8
Ag3/Pt3	98.6	1.4
Ag1,2/Pt1,2	100	0

We show that the English data can be grammatically analyzed within the stochastic OT framework (Boersma 1998, Boersma and Hayes 2001) in a way which provides a principled and unifying explanation for their relation to the crosslinguistic categorical person effects studied by Aissen (1999).

The results argue against the interpretation of soft constraints as a preference mechanism overlaid on a normal generative grammar such as LFG for purposes of disambiguation (Frank et al. 1998), and favor the more radical interpretation of violable constraints as defining core grammaticality (as in OT-LFG and its stochastic generalizations, Kuhn 2001, Sells, ed., forthcoming, Bresnan and Deo 2000). In line with other work (e.g. Hawkins 1994) these results also raise questions about whether a deep and fundamental division between structure and usage, competence and performance is ultimately well-founded.

## References

- Aissen, Judith. 1999. Markedness and subject choice in Optimality Theory. *NLLT* 17: 673–711.
- Boersma, Paul. 1998. *Functional Phonology. Formalizing the interaction between articulatory and perceptual drives*. Amsterdam: University of Amsterdam Ph.D. thesis.
- Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the gradual learning algorithm. *Linguistic Inquiry* 32: 45–86.
- Bresnan, Joan and Ashwini Deo. 2000. ‘Be’ in the *Survey of English Dialects*: A stochastic OT account. Paper presented at the Symposium on Optimality Theory, English Linguistic Society of Japan, November 18, 2000, Kobe, Japan.
- Dahlstrom, Amy. 1984. *Plains Cree Morphosyntax*. Berkeley: University of California, Berkeley Ph.D. dissertation. Published in 1991. New York: Garland.
- Frank, Anette, Tracy Holloway King, Jonas Kuhn, and John Maxwell. 1998. Optimality Theory style constraint ranking in large-scale LFG grammars. In *Proceedings of the LFG98 Conference, Brisbane, Australia*. Stanford: CSLI Online Publications. Revised version to appear in Peter Sells (ed.).
- Godfrey J., E. Holliman and J. McDaniel. 1992. SWITCHBOARD: Telephone Speech Corpus for Research and Development. *Proceedings of ICASSP-92*, San Francisco. 517-520.
- Hawkins, John A. 1994. *A Performance Theory of Order and Constituency*. Cambridge Cambridge: Cambridge University Press.
- Jelinek, Eliose and Richard Demers. 1983. The agent hierarchy and voice in some Coast Salish languages. *IJAL* 49: 167–85.
- Kuhn, Jonas. 2001. Formal and Computational Aspects of Optimality-theoretic Syntax. Stuttgart: IMS, Universität Stuttgart Ph.D. dissertation draft.
- Marcus, Mitchell, Beatrice Santorini, and May Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn treebank. *Computational Linguistics* 19:313–330.
- Rosenhouse, Judith. 1988. Occurrence of the passive in different types of text in English, Hebrew and Arabic. *Babel* 34: 90–103.
- Sells, Peter (ed). Forthcoming. *Formal and Empirical Issues in Optimality-theoretic Syntax*, Stanford: CSLI Publications.