Is There Evidence for a Probabilistic Language Faculty?

Rens Bod (rb@dcs.st-and.ac.uk)
Cognitive Systems Group
School of Computer Science, University of St Andrews
St. Andrews, KY16 9SX, Scotland, UK

Introduction

One of the foundations of modern linguistics is the maxim of categoricity: language is categorical. Numbers play no role, or, where they do, they are artifacts of nonlinguistic performance factors. Thus, while it is widely recognized that real language can be highly variable, gradient, and rich in continua, many linguists would argue that the competence that underlies such "performance factors" consists of well-defined discrete categories and categorical grammaticality criteria. However, a large number of recent results challenges the idea that linguistic competence is categorical and discrete. It has become increasingly clear that phonological alternations and syntactic well-formedness judgments display properties of continua and show gradient behavior. Moreover, it has been shown that speakers' well-formedness judgments of words and sentences are well predicted by the combined probabilities of their subparts (Bod 1998; Bod et al. 2003). The current paper discusses three central linguistic phenomena, i.e. grammaticality, learnability and universality, that challenge the maxim of categoricity providing evidence for a probabilistic language faculty.

Evidence for a Probabilistic Language Faculty

Grammaticality

A groundswell of recent results indicates that speakers' grammaticality judgments display clear properties of continua (Manning 2003; Crocker & Keller 2005). There is no well-defined distinction between sentences generally regarded as "grammatical" in the literature, and those regarded as ungrammatical. Instead, there is a cline of well-formedness, wherein some constructions are highly preferred, others are used less frequently and some are used not at all. The distinction drawn between grammatical and ungrammatical is often somewhere in the middle of the cline, ruling out those constructions that tend to be less frequent as "ungrammatical". However, nowhere in the cline is there a dramatic drop in frequency; in fact, the cline is often gradual, so that the decision where to draw the distinction is relatively arbitrary.

Learnability

One common argument against a probabilistic language faculty stems from skepticism regarding the mind's ability to acquire and store a complex range of generalizations and frequencies. However, adding probabilities to grammars makes the acquisition problem easier, not harder. Gold (1967) demonstrated that formal languages alone cannot be learned without negative evidence. Moreover, negative evidence is not readily available to children. Together, these two facts are often used as evidence that language is special and largely innate, a line of reasoning known as the "argument from the poverty of the stimulus". However, Horning (1969) demonstrated that, unlike categorical grammars, probabilistic grammars *are* learnable from positive evidence alone. It is certainly not the case that probabilities complicate the learning task. On the contrary, if the language faculty is probabilistic, the learning task is considerably more achievable. This is supported not only by psycholinguistic experiments with language learning of infants (Saffran et al. 1996; Tomasello 2003), but also by recent work in computational linguistics where syntactic structure is learned entirely in a statistical, item-based way by means of distributional regularities (Bod 2006).

Universality

Many syntactic constraints are present in a great many languages, reflecting universal tendencies of the language faculty. They are operative in various degrees in different languages and in some cases are highly grammaticalized. A categorical framework does not enable us to capture the different degrees to which constraints are operative in different languages. By contrast, a probabilistic framework as in Bresnan et al. (2001) does enable us to formally model such situations, capturing both the ways in which languages are similar (operating under similar constraints) and the ways in which they differ (the probabilities associated with those constraints).

Conclusion

Language displays all the hallmarks of a probabilistic system. Grammaticality judgments and linguistic universals are probabilistic and stochastic grammars enhance learning. All evidence points to a probabilistic language faculty.

References

Bod, R. (1998). *Beyond Grammar: An Experience-Based Theory of Language*. Stanford, CSLI Publications.

Bod, R. (2006). An All-Subtrees Approach to Unsupervised Parsing, *Proceedings ACL-COLING'2006.*

Bod, R., Hay, J. & Jannedy, S. (eds.) (2003). *Probabilistic Linguistics*. Cambridge, the MIT Press.

Bresnan, J., Dingare, S. & Manning, C. (2001). Soft constraints mirror hard constraints. In *Proceedings of the LFG01 Conference* (pp. 13-32). Stanford, CSLI Publications.

Crocker, M. & Keller, F. (2005). Probabilistic Grammars as Models of Gradience in Language Processing. In Fanselow, G., et al. (eds.), *Gradience in Grammar: Generative Perspectives*. Oxford University Press.

Gold, E. (1967). Language identification in the limit. *Information and Control* 10, 447-474.

Horning, J. (1969). A study of grammatical inference. PhD thesis, Stanford University.

Manning, C. (2003). Probabilistic Syntax. In Bod, R., Hay, J. & Jannedy, S. (eds.) (2003). *Probabilistic Linguistics*. Cambridge, the MIT Press.

Saffran, J., Newport, E. & Aslin, R. (1996). Statistical learning by 8-month-old infants. *Science* 274, 1926-1928.

Tomasello, M. (2003). *Constructing a Language*. Harvard, Harvard University Press.