Formalising Syntactic Variability: Context and Issues

Bill Haddican and Bernadette Plunkett

1. Introduction.

Throughout most of the history of generative grammar, intra-speaker variation as a theoretical problem has received relatively little attention. Early generative work explained intra-speaker variability in terms of optionality of syntactic rules, and assumed that the mechanisms governing choice between the application and non-application of the relevant rules were distinct from the computational knowledge of language that the research programme aimed to characterise. Chomsky's early comments on the distinction between probability and grammaticality were influential in this regard. In particular, Chomsky's (1957:15) *locus classicus* contrasting *colorless green ideas sleep furiously* and *furiously sleep ideas green colorless* was intended to show that probability of occurrence must be distinct from grammaticality. From the perspective of optional rules and this modular distinction—that is between probabilistic (usage) knowledge of language and computational knowledge of language—intra-speaker syntactic variation promised relatively little insight for formal theory.

Minimalist work since the early 1990's has generally eschewed the possibility of optionality in grammar. This turn has prompted some syntacticians to take a greater interest in variability in order to explain it without reference to "free variation." Kroch and colleagues' grammar competition framework, developed in the late 80's and early 90's provided a timely and attractive solution to this problem (Kroch 1989, 1994, 2001, Kroch and Taylor 1997, 2000, Pintzuk 1991). In particular, Kroch (1989, 1994) proposed that intra-speaker variation in syntax reflects bi-dialectalism, where the dialects in competition are distinguished by minimal featural differences. In Kroch's classic example of variable verb raising in the history of English, the learner posits two competing variants: one T with an EPP ("strong") feature responsible for the verb-raising variant, and another EPP-less ("weak") T variant corresponding to contemporary English sentences with the verb lower in the structure. On this approach, the learner's response to variable input is to posit competing features responsible for the different surface forms. The learner keeps track of frequencies of use of variants across contexts, and uses this knowledge in production, but this probabilistic competence is extra-computational; the frequencies attaching to different forms play no role in the derivation (Kroch 1989).

Kroch's competing grammars framework remains a touchstone for work in this area. In more recent literature, however, several promising alternative approaches to intra-speaker variation in syntax have reinvigorated debate on abiding problems including the formal status of probabilistic knowledge in syntax, and the nature of parameters (Adger 2006, Yang 2001, Clark 2004, Asudeh 2001, Bresnan & Nikitina 2003, Manning 2003). These issues are addressed in different ways in the articles in this volume.

2. This volume's contribution and perspectives.

The articles gathered here grew out of a conference dedicated to these issues at the University of York in May 2007. [1] The conference brought together current researchers in formal syntax, language variation and acquisition in an effort to stimulate debate on issues including the following:

- Is variability in child language at the root of syntactic change or is variation in the adult grammar a necessary impetus?
- What is the relationship between intra-speaker and inter-speaker variation in syntax?
- What formal and syntactic mechanisms best explain the existence of intra-speaker variation?
- Are frequencies of variants (partially) predictable from the formal properties (e.g. features) involved?
- What if any is the role of parameters in intra-speaker variation?

Each of the papers in this volume takes up a different subset of these issues and addresses them in somewhat different ways. We briefly summarise some of the contributions of these papers below.

Thráinsson's and Tortora & den Dikken's papers are broadly within the Pollock-Kaynean "microparametric" research tradition. (Black & Motapanyane 1996, Cinque & Kayne 2005). Using comparative techniques not altogether dissimilar to those used in sound/paradigm reconstructions in historical linguistics, this line of research aims to infer loci of variation in structure from differences in constraints on surface forms across speakers and/or closely related varieties. Thráinsson's paper defends Bobalijk & Thráinsson's (1998) theory of verb movement linking the availability of verb raising to split agreement and tense (or aspect) morphology (Pollock 1989). To this end, Thráinsson presents evidence from a broad comparison of word order, morphology and extraction facts across Scandinavian dialects suggesting that the availability of verb movement in these languages has a single abstract locus, namely the presence vs. absence of a split IP.

In a similar vein, Tortora & den Dikken focus on variation in patterns of past tense *be* agreement (*was/were* variation) across dialects of English. In particular Tortora & den Dikken propose that cross-dialectal differences in agreement patterns are usefully expressed in terms of differences in subject positions available in each dialect (Henry 1995, Cardinaletti 2004, Rizzi 2006). They propose that Appalachian dialects share with Belfast English the availability of a spec, AgrsP position, where agreement always obtains. Appalachian English, differs from Belfast English however in that it has an additional high (SubjP) non-agreeing position, and lacks a low (spec, TP) non-agreeing position. They argue that this approach accounts for a range of agreement facts as well as differences in subject NPI licensing and pronominal case marking across these varieties.

Adger & Smith's paper develops and extends their previous combinatorial variation model (Adger & Smith 2005, Adger 2006, 2007). The model is based on standard minimalist assumptions (Chomsky 1995) about the architecture of the grammar and the role of lexical features in computation, with two crucial enrichments. The first is Adger's (2006) learning algorithm for the featural content of individual lexical items. In cases of variable input

[1] We are grateful to participants Formal Approaches to Variation in Syntax for excellent papers and discussion. We gratefully acknowledge the financial support of the Department of Language and Linguistic Science of the University of York.

where the learner has evidence that a single morphological feature set may spell out in more than one phonetic guise—e.g. *we were* and *we was*—the algorithm produces a one-to-many mapping of feature bundles to lexical items. The second enrichment is a choice function which randomly selects mappings of features to lexical items in production. These two assumptions have the consequence that differences in frequencies of variants across contexts in usage data are (partially) predictable from the features involved in the relevant paradigm (modulo extralinguistic social constraints on variation): in cases where the algorithm produces a set of mappings where a given phonetic string is mapped to more frequently than a competitor, the two variants will be favoured in production roughly in proportion to the number of times they are mapped to. Adger & Smith's paper in this volume tests the predictions of this approach with data on several features from a corpus of spoken English from the Scottish town of Buckie (Smith 2000).

Nevins & Parrott address the problem of probabilistic constraints on variation in production data from a somewhat different perspective. They propose to adapt the notion of variable rewrite rule (Labov 1969, Cedergren & Sankoff 1974) to impoverishment rules in a Distributed Morphology framework (Halle & Marantz 1993, Noyer 1998). In this framework, morphological rules are not necessarily categorical but rather may apply probabilistically. In particular, Nevins & Parrott propose that impoverishment operations, triggered by universally marked features or combinations of features, may be of this sort. Using data from three dialects of English—Monmouthshire (UK), Buckie (Scotland), and Smith Island, MD (US)—Nevins & Parrott illustrate how their model expresses variable agreement patterns in different types of paradigm leveling.

Papers by Yang and Miller & Schmitt focus on the implications of variable input for acquisition theory, both in reference to Yang's variational learning model (Yang 2001, 2002). Yang's paper extends his earlier work based on mainly morphological evidence to the syntactic domain. He argues that the variation seen in child language production comes from two different sources and these relate to what Chomsky (1981) identified as the Core and Periphery. One type of variation, relating to the Core, involves the child producing forms for which there is no direct evidence in the input. Yang argues that the most plausible explanation here is that the child is instinctively retrieving grammatical options from the genetic endowment i.e. UG. The other type of variation relates to the periphery and relies on children's ability to develop appropriate generalizations from the input; he argues that here the child is conservative: cases of overregularization are found but these are rarely cases where the child over-generalizes an irregular pattern. Although probabilistic information is used in Yang's model, it is not the source of linguistic hypotheses as it would be in a statistical learning model. Yang acknowledges the failure of purely parameter setting models of syntactic acquisition to capture the gradualness of syntactic development by incorporating both experience and general, not-necessarily linguistic, principles of learning. Yang's paper, however, suggests that the notion of parameter must be retained.

Miller & Schmitt report on a study intended to test one of the predictions of Yang's (2002) model, namely that the setting of a given parameter may be retarded by variation which renders evidence for the setting unreliable. Miller & Schmitt test this with data from the acquisition of plural morphology in Chilean Spanish. In this dialect, variable deletion of syllable final /s/ has the consequence that noun phrases are sometimes marked with a word-final [s] and sometimes not. This variation is also socially stratified: /s/-deletion applies at a higher rate in working class communities in Chile and less frequently in middle class communities. Miller & Schmitt's data show that children from working class communities with higher rates of *–s* deletion in production (including child directed speech) perform

lower on comprehension tasks targeting plural *–s*, than do children from middle class communities with lower rates of *–s* deletion in production data. The results provide strong evidence in favour of the claim that variation in input data may delay parameter setting.

3. Outlook.

Several of the papers collected here take novel approaches to the question of how/whether probabilistic constraints in production data should be understood in formal terms. Since the late 1980's, a popular understanding of this issue has been in terms of Kroch's competing grammars framework, discussed above. Again, in Kroch's framework, stochastic knowledge lies in the domain of performance, outside the computational system. Some rather different approaches to this issue have been proposed in recent literature. In their contribution to this volume and elsewhere, Adger & Smith propose that the narrow syntax may strongly shape frequencies of variants in production data. Their approach is to be contrasted with the more radical departure from the assumption that syntactic knowledge is non-probabilistic by Bresnan and colleagues in a stochastic optimality theory framework, where probabilities are built directly into the formalism (Bresnan & Nikitina 2003, Boersma & Hayes 2001, Clark 2004, cf Manning 2003). We see the debate between these opposing views as potentially instructive.

Not represented in the papers gathered here is a new line of research focusing on the formal status of gradient intuitions of well-formedness (Keller 2000, Hayes 2000, Boersma & Hayes 2001, Sorace & Keller 2005, Featherston, 2005). Generative linguists routinely refer to gradience in well-formedness in formal description, and have done so since the beginning of the research programme (cf. Chomsky 1964). Chomsky (1975: 131-2) in fact proposes that the ability to explain gradience in judgment data must be a criterion of an adequate theory of competence. Generative linguistics, however, has been slow to develop theories capable of explaining gradient well-formedness within an approach to the computational component as binary, rather than probabilistic and gradient. As Embick (2008) notes, the fact that judgment data (in controlled and non-controlled conditions) may be gradient does not entail that the computational component of the grammar itself must be gradient. It cannot be excluded that knowledge of gradience (and probabilistic knowledge more generally) is rather the product of interaction between the computational component and some interfacing external system (e.g. pragmatic component). Explicating such a theory of gradience that maintains a distinction between computational knowledge and (probabilistic) usage knowledge has nevertheless proved a persistent problem for the generative research programme.

Recent literature has seen more studies using controlled experimental techniques that take greater account of gradient data (Cowart, 1997, Alexopoulou & Keller 2007, Sprouse 2009). We hope that this experimental turn will stimulate theoretical advances on the problem of gradience.

References

Adger, D., 2006. Combinatorial variability. Journal of Linguistics 42, 503-530.

Adger, D., 2007. Variability and Modularity: A Response to Hudson. Journal of Linguistics 43, 695-700.

Adger, D., Smith, J., 2005. Variation and the minimalist program. In: Cornips, L., Corrigan, K. (Eds.), Syntax and Variation: Reconciling the Biological and the Social. John Benjamins, Amsterdam, pp. 149-178.

Alexopoulou, T., Keller, F., 2007. Locality, Cyclicity and Resumption: At the Interface between the Grammar and the Human Sentence Processor. Language 83, 110-160.

Asudeh. A., 2001. Linking, optionality, and ambiguity in Marathi. In: Peter Sells (Eds.), Formal and empirical issues in optimality-theoretic syntax. Stanford, CA: CSLI Publications.

Black, J., Motapanyane, V. (Eds), 1996. Micro-parametric Syntax and Dialect Variation. John Benjamins, Amsterdam.

Bobaljik, J. D., Thráinsson, H., 1998. Two Heads Aren't Always Better than One. Syntax 1:37-71.

Boersma, P., Hayes, B., 2001. Empirical tests of the Gradual Learning Algorithm. Linguistic Inquiry 32: 45–86.

Bresnan, J.,Nikitina, T., 2003. On the Gradience of the Dative Alternation. Ms. Stanford University.

Cardinaletti, A., 2004. Toward a cartography of subject positions. In: Rizzi, L. (Ed.), The Structure of CP and IP, vol. 2. Oxford University Press, New York, pp. 115-165.

Cedergren H.J., Sankoff, D., 1974. Variable rules: performance as a statistical reflection of competence, Language 50, 333–355.

Chomsky N. 1957. Syntactic Structures. Mouton, The Hague.

Chomsky, N., 1964. Degrees of grammaticalness. In: Fodor, J. A., Katz, J. J. (Eds.), The Structure of Language: Readings in the Philosophy of Language. Prentice-Hall, Englewood Cliffs, NJ, pp. 384–389.

Chomsky, N., 1975. The Logical Structure of Linguistic Theory. Plenum Press, New York.

Chomsky, N., 1981. Lectures on Government and Binding. Foris, Dordrecht.

Chomsky, N., 1995. A Minimalist Program for Linguistic Theory, MIT Press, Cambridge, MA.

Cinque, G., Kayne, R.S. (Eds.), 2005. The Oxford Handbook of Comparative Syntax, Oxford, Oxford University Press.

Clark, B. 2004. A Stochastic Optimality Theory Approach to Syntactic Change. PhD thesis, Stanford University.

Embick, D. 2008. Variation and Morphosyntactic Theory: Competition Fractionated. Language and Linguistics Compass 2(1): 59–78

Featherston, S. 2005. Universals and grammaticality: Wh-constraints in German and English

Halle, M., Marantz, A., 1993. Distributed morphology and the pieces of inflection. In: Keyser, S. J. & Hale, K. (Eds), The view from building 20, MIT Prss, Cambridge, Mass, pp. 111-76.

Henry, A., 1995. Belfast English and Standard English: Dialect Variation and Parameter Setting. Oxford, Oxford University Press.

Keller, F., 2000. Gradience in grammar: experimental and computational aspects of degrees of grammaticality. Ph.D. thesis, University of Edinburgh.
Kroch A., 1989. Reflexes of grammar in patterns of language change. Journal of Language Variation and Change 1, 199-244.
Kroch, A. 1994. Morphosyntactic variation. In: Beals, K. (Ed.), Proceedings of the 30th Annual Meeting of the Chicago Linguistics Society, Chicago: Chicago Linguistic Society.
Kroch, A., 2001. Syntactic change. In: Baltin, M., Collins, C. (Eds.): The Handbook of Contemporary Syntactic Theory. Oxford, Blackwell.
Kroch, A., Taylor, A., 1997. Verb movement in Old and Middle English: Dialect variation and language contact. In: van Kemenade, A., Vincent, N., (Eds.), Parameters of Morphosyntactic Change. Cambridge University Press, Cambridge.
Kroch, A., Taylor, A., 2000. Verb-Object Order in Early Middle English. In: Pintzuk, S., Tsoulas, G., Warner, A. (Eds.), Diachronic Syntax, Models and Mechanisms: Proceedings of DIGS 5, Oxford University Press, Oxford, pp. 132-163.
Labov, W., 1969. Contraction, deletion, and inherent variability of the English copula. Language 45, 715-62.
Manning, C., 2003. Probabilistic syntax. In: Bod, R., Hay, J., Jannedy, S. (Eds.), Probabilistic Linguistics, MIT Press, Cambridge, MA, pp. 289–341.
Noyer, R., 1998. Impoverishment theory and morphosyntactic markedness. In: Lapointe, S. G., Brentari, D. K., Farrell, P. M.. (Eds), Morphology and its relation to phonology and syntax,. CSLI, Stanford, pp. 264-85
Pintzuk, S., 1991. Phrase structures in competition, Routledge, London.
Pollock, J.-Y., 1989. Verb movement, universal grammar and the structure of IP. Linguistic Inquiry 20, 261-305.
Rizzi, L., 2006. On the form of chains: Criterial positions and ECP effects. In: Cheng, L., Corver, N. (Eds.), Wh-movement Moving on, MIT Press, Cambridge, MA, pp. 97-134.
Smith, J., 2000. Synchrony and Diachrony in the Evolution of English: Evidence from Scotland. Ph.D. dissertation, University of York.
Sprouse, Jon. 2009. Revisiting Satiation. Linguistic Inquiry. 40.
Yang, C., 2001. Internal and external forces in Language Change. Language Variation and Change. 12, 231-250.
Yang, C. 2002. Knowledge and Learning in Natural Language, Oxford University Press, Oxford.