

SEMANTIC DATA AND THE COMPUTATIONAL SYSTEM

Frederick J. Newmeyer

University of Washington, University of British Columbia, and Simon Fraser
University

Like the other papers at this workshop, mine is concerned with the question of the data relevant to the construction of a syntactic theory. Unlike the others, however, it deals not with the collection of and reliability of the data, but rather with its ontological nature. To be specific, I argue that semantic data (i.e., judgments of paraphrase, ambiguity, scope, nuances of aspect and the nature of events, etc.) is, in general, inappropriate for the construction of a syntactic theory, where 'syntactic theory' is understood in terms of the structure-building operations of the computational system.

Throughout most of its history, what has distinguished generative syntax from virtually all other approaches to grammar is the hypothesis of the autonomy of syntax (AS), namely the idea that the rules (principles, constraints, etc.) that determine the combinatorial possibilities of the formal elements of a language make no reference to constructs from meaning, discourse, or language use. Support for AS comes from the fact that every language has formal generalizations that do not map smoothly onto generalizations from meaning or use. For example, consider subordination markers in Japanese (Yuasa and Francis 2003; Yuasa 2005). Yuasa demonstrates that the markers 'totan', 'toori', 'kuse', and 'ageku' manifest a mismatch between form and meaning. Syntactically they are nouns, but semantically they are predicates. What makes this fact especially interesting is that the class of subordination markers including 'kara' and 'keredo' are syntactically postpositions and semantically predicates, while the class including 'toki' and 'uchi' are syntactically nouns, but semantically arguments. In other words, we do not have a smooth mapping between form and meaning – syntactic notions like 'noun' and 'postposition' in Japanese need to be given a characterization independently of semantic notions like 'predicate' and 'argument'. It is facts like these that argue for a level of purely formal patterning – a level linking to meaning, but not subsidiary to it.

AS has a methodological corollary, namely that semantic *data* should be avoided as well. A quote from *Aspects* affirms AS, both in its theoretical and methodological variants:

For the moment, I see no reason to modify the view, expressed in Chomsky (1957) and elsewhere, that although, obviously, semantic considerations are relevant to the construction of general linguistic theory ... there is, at present, no way to show that semantic considerations

play a role in the choice of the syntactic or phonological component of a grammar or that semantic features (in any significant sense of this term) play a role in the functioning of the syntactic or phonological rules.
(Chomsky 1965: 226)

My purpose in this talk is to reaffirm the correctness of the methodological corollary.

The ‘prohibition’ against semantic data has been based largely on the fact that the nature of the form-meaning interface is one of the most difficult problems in linguistics. The worst thing, therefore, would be to *presuppose* that such evidence is relevant to syntactic theory. Also, by avoiding semantic data, one can use form to get at meaning: ‘In general, as syntactic description becomes deeper, what appear to be semantic questions fall increasingly within its scope...’ (Chomsky 1964: 936). For example, Chomsky (1957) motivated the passive transformation purely on its formal properties. The rough paraphrase relation between actives and passives was *not* one of Chomsky’s motivations. Hence, Chomsky felt comfortable concluding that his analysis *explained* why actives and passives are largely synonymous. And later (Chomsky 1973), a constrained theory of movement rules led to the Specified Subject Condition and Tensed-S Condition, which led to the trace theory of movement rules, which led to surface interpretation of meaning, which led to capturing certain aspects of quantifier scope structurally (May 1977).

In the past 20 years, however, the methodological corollary to AS has been increasingly violated. Routinely now, projections (NegP, TopP, FocP, AspP, etc.) are proposed purely on data pertaining to the meanings or discourse properties of the items involved. Likewise scope differences and differences in event interpretations drive the proposed hierarchical orderings among these projections. The remainder of the talk is devoted to demonstrating the negative consequences of using semantic data within syntactic theory.

I begin with English modal auxiliaries. The structural generalizations have been known for many years:

- (1) a. They occur before all other auxiliaries (*must have gone*; **have must gone*)
- b. They do not occur in sequence (in Standard English) (**might could*)
- c. They take neither infinitive marker nor inflection (**to would*; **woulded*)
- d. They must be followed by non-finite form of the verb or auxiliary (**I must had gone*)
- e. They invert in questions and are followed by the negative morpheme.
- f. All of (1a-e) apply to modals both in their root and epistemic senses:

Much current work (Stowell 2004; Zagana forthcoming-a; b) ignores (1a-f). The goal has become to represent the subtle scopal differences between root and epistemic modals structurally. As I demonstrate, the downside is that (1a-f) become extremely difficult to capture.

I turn next to English derived nominalizations (DNs) (*refusal, height, aggression, etc.*). It has been known since Chomsky (1970) that underived nouns and DNs have identical structures in relevant respects (2a-b). Also, DNs occur in DPs corresponding to base structures, but not to transformationally derived structures (3a-b):

- (2) a. Mary's three boring books about tennis
b. Mary's three unexpected refusals of the offer
- (3) a. Harry was certain to win the prize.
b. *Harry's certainty to win the prize (no Raising within DP)

These profound formal generalizations are all but ignored in a lot of current work. Instead, the goal has become to capture subtle event reading generalizations structurally (see Borer 2003, Alexiadou 2001). (4b) is Borer's derivation of (4a):

- (4) a. Kim's destruction of the vase
b. [NP *-tion*_{NOM} [EP *Kim* [Arg-SPQ *the vase* [VP *destroy*]]]]

I show that the generalizations represented in (2-3) are next to impossible to capture if a VP node underlies the DN *destruction*.

I turn then to a critique of the (crosslinguistic) projection NegP, which is typically posited using exclusively semantic data. As I show, negation can pattern structurally with complement-taking verbs (Tongan); with auxiliaries (Estonian); with derivational affixes (Turkish); with nouns (Evenki); and with adverbs (English). If all negative elements are heads of the NegP projection, then these language-particular patterns are unaccounted for.

A crucial question is whether it is actually the case that 'as syntactic description becomes deeper, what appear to be semantic questions fall increasingly within its scope'. For decades, that seemed to be completely correct. In GB, for example, several principles accounted both for purely syntactic processes and for data about construal as well. The two best known examples:

- (5) a. The binding theory accounted both for constraints on movement and constraints on anaphora.
b. The Empty Category Principle (ECP) accounted for both purely

structural facts (eg *that*-trace, the order of elements in incorporation structures, constraints on the extraction of adjuncts, etc.) and facts about (semantic) scope.

But are (5a-b) (or their current instantiations) correct? Probably not. The locality of movement is said to follow from the nature of Merge, but not the locality of anaphor binding (which applies at LF). In any event, long-distance anaphors (LDA) appear to be the norm in language, not the exception. Within the framework of Optimality Theory, Moon (1995) argued that the thematic role of the antecedent and anaphor in Korean is the major factor determining binding possibilities. The purely structural c-command relation is a distant fourth:

- (6) Ranked constraints in the binding of Korean long distance anaphors:
- a. Thematic Hierarchy Constraint (LDA must be bound by a thematically higher NP)
 - b. Larger Domain Preference Constraint (Given potential antecedents for LDA in different domains, the more distant the domain, the stronger the preference)
 - c. Subject-Orientation Constraint (LDA must be bound by a subject NP)
 - d. C-Command Constraint (LDA must be bound by a c-commanding NP)
 - e. Discourse Binding Constraint (LDA must be bound by a prominent discourse NP if no sentential antecedent is available)

And the ECP is not even expressible, given minimalist assumptions.

I conclude with a brief discussion of why semantic data has become increasingly used by syntacticians. I suggest that it is (unfortunately) due in large part to the fact that Chomsky has never endorsed (or even expressed interest in) a formal semantic theory. The tendency then has been to expand syntax to encompass what is naturally the domain of such a theory.

REFERENCES

- Alexiadou, Artemis. 2001. Functional structure in nominals: Nominalization and ergativity. *Linguistik Aktuell*, vol. 42. Amsterdam: John Benjamins.
- Borer, Hagit. 2003. Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon. *The nature of explanation in linguistic theory*, ed. by John Moore and Maria Polinsky, 31-67. Stanford: CSLI Publications.
- Chomsky, Noam. 1957. Syntactic structures. *Janua Linguarum Series Minor*, vol. 4. The Hague: Mouton.
- . 1964. The logical basis of linguistic theory. *Proceedings of the Ninth International Congress of Linguists*, ed. by Horace G. Lunt, 914-77. The Hague: Mouton.
- . 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

- . 1970. Remarks on nominalization. *Readings in English transformational grammar*, ed. by Roderick Jacobs and Peter Rosenbaum, 184-221. Waltham, MA: Ginn.
- . 1973. Conditions on transformations. A festschrift for Morris Halle, ed. by Steven Anderson and Paul Kiparsky, 232-86. New York: Holt Rinehart & Winston.
- May, Robert. 1977. *The grammar of quantification*. MIT Ph. D. dissertation.
- Moon, Seung Chul. 1995. *An optimality approach to long distance anaphors*. University of Washington Ph. D. dissertation.
- Stowell, Timothy A. 2004. Tense and modals. *The syntax of time*, ed. by Jacqueline Guéron and Jacqueline Lecarme, 621-35. Cambridge, MA: MIT Press.
- Yuasa, Etsuyo. 2005. *Modularity in language: Constructional and categorial mismatch in syntax and semantics*. Berlin: Mouton de Gruyter.
- Yuasa, Etsuyo and Elaine J. Francis. 2003. *Categorial mismatch in a multi-modular theory of grammar. Mismatch: Form-function incongruity and the architecture of grammar*, ed. by Elaine J. Francis and Laura A. Michaelis, 179-227. Stanford: CSLI Publications.
- Zagona, Karen. forthcoming-a. *Phasing in modals: Phasing and the epistemic/root distinction*. ????????, ed. by Jacqueline Guéron and Jacqueline Lecarme, xxx-xxx. Amsterdam: John Benjamins.
- . forthcoming-b. *On the syntactic features of epistemic and root modals*. *Proceedings of the Fifteenth Colloquium on Generative grammar*, ed. by Luis Eguren and Olga Fernandez Soriano, xxx-xxx. Amsterdam: John Benjamins.