Containment, exclusion, and implicativity in natural logic

While semanticists have focused mainly on formal logic as a model of inference and medium of semantic representation, considerations of linguistic typology and human processing of language (Dowty 1994) have repeatedly drawn people towards developing what Lakoff (1970) called *natural logic*, which characterizes valid patterns of natural language inference in terms of surface forms. A computational model of natural logic might have less deductive power than one based on formal logic, yet handle a greater proportion of everyday inferences, by sidestepping the myriad difficulties involved in translating natural language fully and accurately to logical form.

A natural logic approach was the basis of traditional logic (e.g., Aristotle's syllogisms), and was revived in a formal form by van Benthem (1986) and Sánchez Valencia (1991). Their *monotonicity calculus* explains inferences involving semantic containment and inversions of monotonicity, even when nested. But because it lacks any representation of exclusion (as opposed to containment), it fails to explain many simple inferences.

Another model which arguably follows the natural logic tradition (though not presented as such) was developed by Nairn et al. (2006) to explain inversions and nestings of implicative (and factive) predicates. Their *implication projection algorithm* bears some resemblance to the monotonicity calculus, but does not incorporate containment relations or explain interactions between implicatives and monotonicity.

This paper proposes a model of natural logic which generalizes the monotonicity calculus to cover inferences involving exclusion, and (partly) unifies it with Nairn et al.'s model of implicatives. The elements of the model are: an inventory of basic entailment relations, a concept of *projectivity* of entailment, and an inference algorithm based on composing entailments.

Entailment relations. We propose an inventory of seven mutually exclusive *basic entailment relations*, which are defined for expressions of every semantic type via induction. This inventory of relations (1) provides a level of expressivity well-suited for natural logic, by accommodating representations of both containment and exclusion; (2) constitutes a partition, so that every pair of expressions (of like type) maps to exactly one relation; and (3) forms the basis of a Tarskian *relation algebra*, so that relation composition is well-defined.

Projectivity. We categorize semantic functions according to their *projectivity class*, a concept which generalizes both Sánchez Valencia's monotonicity classes and the implication signatures of Nairn et al. The projectivity class of a function f specifies how the entailment relation between $f(x)$ and $f(y)$ depends on the entailment relation between x and y. Projectivity thus allows us to determine the entailments of a compound expression recursively, by propagating entailments upward through a semantic composition tree.

Inference. We describe an inference algorithm based on (1) finding a chain of *atomic edits* which transform the premise into the conclusion, (2) determining the lexical entailment relation generated by each atomic edit, (3) projecting lexical entailment relations upward to find entailment relations across edit steps, and (4) composing entailment relations across the chain of edits.

Though it lacks the deductive power of first-order logic, by incorporating monotonicity, exclusivity, and implicativity, this model succeeds in explaining many everyday patterns of inference. We have implemented a computational model of natural-language inference based on this framework which correctly handles most problems in the FraCaS test suite (Cooper et al. 1996).

Cooper, R. et al. (1996). Using the framework. Technical Report LRE 62-051 D-16, The FraCaS Consortium.

Dowty, D. (1994). The role of negative polarity and concord marking in natural language reasoning. In *Proceedings of SALT-IV*, pp. 114–144.

Lakoff, G. (1970). Linguistics and natural logic. *Synthese 22*, 151–271.

Nairn, R., C. Condoravdi, and L. Karttunen (2006). Computing relative polarity for textual inference. In *Proceedings of ICoS-5*.

Sánchez Valencia, V. (1991). *Studies on Natural Logic and Categorial Grammar*. Ph. D. thesis, University of Amsterdam.

van Benthem, J. (1986). *Essays in logical semantics*. Dordrecht: Reidel.