

The Hidden Path of Semantic Content within Pragmatic Context: The definite article, "*the*"

Ring Mei-Han Low

Abstract for the International Workshop "Where Semantics Meets Pragmatics"
Michigan State University, July, 11-13, 2003

This study argues that semantic traits can determine the likelihood of a noun occurring in a definite construction in discourse. The research presents a set of corpus analysis and web-search data, illustrating that some noun phrases being "definite" in discourse is not just the result of the pragmatic contexts involved, but also that of their semantic contents.

Many researchers who have studied the definite construction of "*the*" in English have followed a pragmatic, and perhaps anaphorically-orientated approach, emphasizing that the use of the construction is primarily "discourse-driven" and predominantly "anaphoric" in nature. Some approach radiating from this perspective include: the notions of "identifiability", "familiarity" and "givenness" (e.g., Gundel et al 1993, 1996, 2001; Chafe 1976; Prince 1981, 1992); the mechanism of "bridging inferences" (e.g., Haviland and Clark 1974, Clark 1978); the theory of "common ground" (e.g., Clark 1992); the hypothesis of "relevance" (e.g., Matzui, 2000; Sperber and Wilson 1995), and the effect of "accessibility" (e.g., Ariel 1988).

Although some studies point out that certain semantic relationships, e.g., the part-whole relationship, may be used by the hearer to identify definite referents (e.g., Hawkins 1978; Erku and Gundel 1987, Prince 1981), the process has been discussed primarily as a general pragmatic mechanism, rather than one that could be lexically specific to the words being used.

A handful of researchers, however, have introduced a stronger semantic perspective to the issue. Lobner 1985 argues that some nouns are "definite" simply because they are semantically so. He proposes that many definites in discourse are indeed "Semantic Definites". Semantic Definites establish their referents independently from the immediate situation or context of the utterance, but refer to their referent with a semantic property of their own. Fraurud 1996 expresses a similar view. She studied her Finnish corpus the correlation between definite encoding and the ontological classes of the definite referents. She found that definite noun phrases that denote non-human entities are more likely to be "antecedent-less" in discourse, when compared to those denoting human entities. Moreover, many of the antecedent-less definites (i.e., First Mention Definites) are actually semantic "functionals" - to be identified indirectly through their

semantic arguments (e.g., *the nose*, but **a nose*).

In this investigation, I extend this second view of Semantic Definiteness on several counts.

The study begins with new evidence to confirm that non-anaphoric definites are prevalently common in natural discourse. It presents an analysis of 1417 definite noun phrases in a database, that contains two interview transcriptions and twenty articles of various genres: stories, reviews, columns, news. It shows that similar to what Fraurud has found in the Finnish corpus, approximately half of the definite noun phrases of "*the*" in the English corpus are non-anaphoric, First Mention Definites (i.e., do not have an explicit previous mention in the discourse) - 414 NPs out of a total of 876 in the written corpus and 262 NPs out of a total of 541 NPs in the interview transcriptions (excluding possessive expressions, relative clauses, and prepositional phrases). This result adds to the limited but growing amount of corpus evidence provided by other researchers (Fraurud 1996, Gundel 2001, Poesio and Vieira 1998), who recently have also shown that non-anaphoric definites are in fact common in natural contexts; and more research is needed to explore the contents of these definite noun phrases.

Next, the study presents data to support the idea that semantic properties of certain noun phrases, rather than the pragmatic contexts, play a role in speakers' choice of their definite encoding in discourse. I propose that there are two components to this phenomenon of "semantic definiteness": the Modifier-Driven Definites and Head-Noun-Driven Definites.

First, some antecedent-less definite noun phrases are definite, because they consist of modifiers of certain properties. My analysis shows that about thirty percent of the antecedent-less definites in the corpus contain modifiers favoring the use of a definite article with the noun phrase, regardless of the discourse contexts. These modifiers include those that are typically use to express quantities (e.g. *a few of*, *the rest of*), relative positions (e.g., *the other*, *the middle of*, *early*), ordinal concepts (e.g., *first*, *second*), and the status of familiarity (e.g., *famous*, *well-known*). They also include some adjectives that are "contrastive" in their semantic content (e.g., *big*, *small*, *only*).

Second, I found about sixty five percent of the antecedent-less definites consist of head nouns that have specific semantics to them. They denote generic concepts (e.g., *the sun*; 18%), functional concepts (e.g., *the doors*; 29%), or proper names (e.g., *the U.S. Appeal Court*; 18%). This set of findings is consistent with the corpus patterns presented in recent studies by other researchers (Fraurud 1996; Poesio and Vieira 1998). They show that the semantic properties of the head noun can also contribute to its definiteness.

The study further provides direct evidence to demonstrate that some lexical items are indeed more likely to be definite when compared to others, due to the specific traits of

their semantic contents. A sample of 150 nouns was collected from the corpus to enter a “web-search procedure” with the Google search engine, to measure how many web pages on the Internet may contain the phrases, “*a* + N” vs. “*the* + N”. If the semantic traits of a lexical item have little bearing on its being definite or not, all the items would have a frequency ratio of “*the*-N : *a*-N” to be within the baseline variation, i.e., the frequency ratio of “*the*” : “*a*” in natural discourse (from corpora and web-search measures, about 1:1 to 2.5:1). However, some extreme ratios were found. Initial analyses have shown that the more “semantically relational” a noun is, the more likely it would be modified by *the* in various contexts, i.e., to have a higher “*the*-N vs. *a*-N” ratio of frequency (some examples of these words are: *denominator* = 18:1, *bottom* = 13:1, *dough* = 8.5:1). On the other hand, nouns that seem to have a “low-information” content are least likely to be modified by “*the*”, the measures show (e.g., *compliment* = 0.3:1, *gift* = 0.35:1, *foothold* = 0.04:1).

This research adds new evidence to support a semantic view of definiteness. It reveals three relevant phenomena. First, some nouns are likely to be definite regardless of the discourse context, because of their semantic properties of “intrinsic identifiability”. Some items, on the other hands, are “unlikely” to be definite, perhaps because they are semantically “vague” in content. Finally, the semantics of the modifier in a noun phrase frequently play a role in determining the definiteness of the referent, regardless of its discourse status.

The role of semantic content in definiteness, as discussed in this study, is an illustration on how semantic contents can enter the realm of pragmatic mechanisms, to exercise their influence on a choice of a “seemingly” pragmatic construction - the definite encoding of nouns.

Cited References:

- Ariel, Mira. 1988. Referring and accessibility. *Journal of Linguistics* 24: 65-87.
- Chafe L Wallace. 1976. Givenness, Contrastiveness, Definiteness, Subjects and Topics. In *Subject and Topic*, Charles N. Li (Ed). Academic. New York.
- Clark H. 1977. Bridging. In Johnson-Laird and Wason (eds) *Thinking: Readings in Cognitive Science*. John Wiley and Sons NY
- Clark, H. 1992. *Arenas of Language use*. The University of Chicago Press.
- Erku and Gundel 1987 The Pragmatics of Indirect Anaphors. In Jef Verschueren and Marcella Bertuccelli-Papi (eds) *The Pragmatic Perspective: Selected Papers from the 1985 International Pragmatics Conference*. John Benjamins. Amsterdam.
- Fraurud, Kari. 1996. Cognitive Ontology and NP Form. In *Reference and Referent Accessibility*. Thorstein Fretheim and Jeunette K Gundel (Eds).
- Gundel et al 2001. Definite Descriptions and Cognitive Status in English: Why Accommodation Is Unnecessary. *English-Language-and-Linguistics* 2001, 5, 2, Nov, 273-295.
- Gundel K. Jeanette. 1996. Relevance Theory Meets the Givenness Hierarchy. In Fretheim, Thorstein, & Gundel, Jeanette K. [Eds], *Reference and Referent Accessibility*.
- Gundel, Jeanette K; Nancy Hedberg, Ron Zacharski. 1993. Cognitive Status and the form of referring expressions in discourse.

- Haviland,-Susan-E.; Clark,-Herbert-H.. 1974. What's new? Acquiring new information as a process in comprehension. *Journal-of-Verbal-Learning-and-Verbal-Behavior*;1974, 13, 5, Oct, 512-521..
- Lobner 1985. Definites. *Journal-of-Semantics*; 1985, 4, 4, 279-326.
- Matsui,-Tomoko. 1998. Pragmatic Criteria for Reference Assignment: A Relevance-Theoretic Account of the Acceptability of Bridging. *Pragmatics-and-Cognition*; 1998, 6, 1-2, 47-97.
- Prince, Ellen F. 1981. Toward a Taxonomy of Given-New Information. in: P. Cole (ed.), *Radical Pragmatics*, Academic, New York
- Prince, Ellen F.. 1992. The ZPG letter: Subjects, definiteness, and information status.. In: William Mann and Sandra Thompson, eds., *Discourse description: Diverse linguistic analyses of a fund-raising text*, 295-325.
- Sperber, D. and D. Wilson. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.