

Where does the meaning of compounds and possessives come from? A contrastive view

Anders Søgaard
Center for Language Technology
Njalsgade 80
DK-2300 Copenhagen
anders@cst.dk

July 1, 2005

Abstract

The literature on the semantics of (non-argumental) compound nouns and possessives is rich on so-called pragmatic proposals, which argue that any novel compound or possessive is semantically underspecified, and that its more specific meaning is only contextually determined, i.e. the Variable \mathfrak{R} Hypothesis. It has sometimes been argued that there is no positive evidence against pragmatic proposals (Lieberman and Sproat, 1992), but in this paper, it is demonstrated why this is not true. In fact, there is a clear-cut test for saying whether meaning is semantically or pragmatically determined, namely a test for *semi-productivity*, i.e. non-syntactically and non-pragmatically motivated unacceptability. The test is used to show that while one may reasonably argue that possessives are semantically underspecified in the majority of languages, this is not the case for compound nouns. It is also demonstrated how this becomes much clearer, once a contrastive view is adopted.

1 The Variable \mathfrak{R} Hypothesis

The Variable \mathfrak{R} Hypothesis was formulated by Allen (1970), but authors have proposed similar ideas long before her. In this paper, we discuss the application of this hypothesis to the semantics of compound nouns and possessives. The literature is rich on representative examples of such applications, but only two of those are explicitly discussed here. They are discussed in rather general terms, however, and our investigation will have consequences for most, if not all, pragmatic proposals in the field. The first work, which is discussed here, is Lauri Bauer's theory of compound semantics, which he presented in short form in Bauer (1979) (see also Bauer, 1978). A representative application of the hypothesis to the domain of possessives is Barbara Partee's earlier work, e.g. Partee (1997 [1983]); her later work on possessives is not discussed here (see e.g. Partee and Borschev, 1998).

Lauri Bauer (1979) posits an abstract proverb (\mathfrak{R}) that is pragmatically determined to capture the (ontological) relation between the referents of the

two constituents of binominal compound nouns. His terminology is slightly misleading, since he talks about this in grammatical terms, but there is no doubt that he is in fact talking about an inferred relation:

there is a connection between lexeme A and lexeme B in a compound of the form AB such as can be predicted by the speaker/hearer partially on the basis of her knowledge of the semantic make-up of the lexemes involved and partially on the basis of other pragmatic factors (Bauer, 1979:46)

It is clear from the article that by “semantic make-up” Bauer means lexically encoded ontological information. Consequently, under a standard grammar design with a minimal lexicon, his principle amounts to saying that the meaning of a compound is predicted on the basis of ontological knowledge. He identifies the relevant categories of ontological knowledge as (i) knowledge of the world, (ii) beliefs about the structure of the world, (iii) knowledge of the artefacts connected with the society and (iv) knowledge of the qualities associated with a particular entity in the society.

As inconsistent as this classification seems, it is evident how Bauer’s principle amounts to saying that the possible readings of a compound are determined by our ontological knowledge, and that the actual readings are those consistent with the context.¹ Consequently, if the interpretation of compound nouns can be explained on ontological grounds, this is evidence in favor of Bauer’s application of the Variable \mathfrak{R} Hypothesis; if not, there is reason to reject it.

Barbara Partee (1983) presents a similar analysis of non-argumental possessives (if the possessum phrase is headed by a relational noun, the inherent relation is added to the set of possible possessive relations). Consider the proposed logical form for the genitival possessive:

$$(1) \quad \llbracket \textit{Shakespeare}'s \rrbracket = \lambda P_1. \iota x_1. P_1(x_1) \wedge \mathfrak{R}(\textit{Shakespeare}, x_1)$$

The ι -operator is for Russellian definite descriptions. The form is not supposed to indicate that the book, which is referred to, is the only book of relevance in the present context, but that it is the only book in \mathfrak{R} -relation to $\llbracket \textit{Shakespeare} \rrbracket$. Some problems arise with this logical form. General problems with the ι -operator are discussed e.g. in Goard (2003).

2 The semantics-pragmatics distinction

By hypothesis, meaning is definable by Herbrand models, i.e. models in which ground terms are interpreted as themselves. More formally, the meaning of the utterance σ is defined by the set of models $\mathfrak{M}_\sigma \models \phi(\sigma)_i \cup \Sigma \cup \Lambda$, where $\phi(\sigma)_i$ is some logical translation of σ , i.e. a hypothesis about the semantic contribution, Σ defines the structure of the semiotic situation in which σ was uttered, i.e. the location, participants, etc., and Λ is the extra-linguistic ontology, or what is more commonly referred to as background knowledge.

¹Consistency is probably not the only pragmatic parameter of relevance to the interpretation of compound nouns, i.e. the relevant parameters probably include minimality and informativity (see below), but this has no consequence for our general discussion.

\mathfrak{M}_σ is a set of Herbrand models, which may be mutually inconsistent (if σ is ambiguous such that there exists a set of mutually inconsistent analyses $\{\phi(\sigma)_1, \dots, \phi(\sigma)_i, \dots, \phi(\sigma)_n\}$). The preferred reading is determined by various pragmatic principles, incl. minimality and informativity. Call the set of pragmatic principles \mathcal{P} . Minimality and informativity can be formally stated:

Minimality Suppose two theories Δ_i and Δ_j (obtained from two sets of hypotheses Υ_i and Υ_j , Σ and Λ) are satisfied by two models \mathfrak{M}_i and \mathfrak{M}_j , i.e. $\mathfrak{M}_i \models \Delta_i$ and $\mathfrak{M}_j \models \Delta_j$. Then \mathfrak{M}_i is said to be the minimal model if $|\mathfrak{M}_i|_\iota \leq |\mathfrak{M}_j|_\iota$ (where $|x|_\iota$ is a function from structures into the integers, which can be paraphrased as “the number of individuals in x is ...”) and vice versa. If $|\mathfrak{M}_i|_\iota = |\mathfrak{M}_j|_\iota$, then \mathfrak{M}_i is said to be the minimal model if $|\mathfrak{M}_i|_\rho \leq |\mathfrak{M}_j|_\rho$ (where $|x|_\rho$ is a function from structures into the integers, which can be paraphrased as “the number of relations in x is ...”) and vice versa. The binary relation symbol “ \preceq ” denotes minimality, i.e. if \mathfrak{M}_i is more minimal than \mathfrak{M}_j , this is written $\mathfrak{M}_i \preceq \mathfrak{M}_j$.

Imagine a discourse δ consisting of two utterances σ_1 and σ_2 . The background knowledge of the discourse agents at the time of the first utterance is some ontology Λ and the semiotic space Σ , i.e. knowledge about the setting of δ and its participants. When σ_2 is uttered, the ontology is substituted with the minimal model \mathfrak{M}_{σ_1} satisfying the logical form of σ_2 ($\phi(\sigma_2)$) and $\Lambda \cup \Sigma$. Informativity is now straight-forwardly defined:

Informativity If $\mathfrak{M}_{\sigma_1} \models \phi(\sigma_2)$, σ_2 is a non-informative utterance.

Say there is some (possibly infinite) structure \mathcal{C} which is the possible interpretations of all possible linguistic utterances to the human mind. The contribution of semantics can then be defined as \mathcal{S} such that $\mathcal{S} \subseteq \mathcal{C} \wedge \mathcal{S} \subseteq \Lambda^{-1}$, i.e. all that which is not ontologically motivated.

On safe and formal grounds, we are now set to investigate the semantics of compound nouns and possessives. It is easy to formulate exactly a question that can be used to evaluate Bauer’s and Partee’s proposals: *Is the specification of \mathfrak{R} inferable from Λ and Σ by means of \mathcal{P} ?*

Is there a short-cut to finding an answer to this question? One reasonable method seems to be to check whether the relevant acceptability data can be explained on ontological grounds. If certain syntactically wellformed combinations of constituents are unacceptable, and the unacceptability is without ontological motivation, then their unacceptability must come from semantics. Consequently, the meaning of compound nouns is partially determined by constructional specification of \mathfrak{R} . This, of course, violates the Variable \mathfrak{R} Hypothesis.

3 Where does the meaning of compounds and possessives come from?

In the literature, certain unacceptable combinations of nouns have been claimed to be non-idiosyncratic, i.e. rule-governed:

- English does not allow human-denoting modifiers in standard endocentric constructions (Copestake and Lascarides, 1997)
- Danish does not allow instrumental modifiers in standard endocentric contexts (Søgaard, 2004)
- Estonian does not allow a compound AB to be interpreted as “a B which looks like an A” (Hiramatsu et al., 2000)

This is evidenced by the following data:

- (2) (a) *butcher knife (‘a knife to be used by a butcher’)
 (b) knife butcher (‘a butcher who uses a knife’)
 (c) butcher sculpture (‘a sculpture that looks like a butcher’)
- (3) (a) slagterkniv (‘a knife to be used by a butcher’)
 (b) *knivslagter (‘a butcher who uses a knife’)
 (c) slagterskulptur (‘a sculpture that looks like a butcher’)
- (4) (a) lihunikunuga (‘a knife to be used by a butcher’)
 (b) unugalihunik (‘a butcher who uses a knife’)
 (c) *lihunikskulptur (‘a sculpture that looks like a butcher’)

If it is accepted that these semi-productivities are in fact non-idiosyncratic, this is clear evidence against the application of the Variable \mathfrak{R} Hypothesis to compound semantics. Is similar data found for possessives?

Apparently not. Intuitions are vague, but the closest I have come to interesting data is that (i) the predicative possessive in Danish seems more restrictive than in Italian, and (ii) the various possessive constructions of Norwegian Bokmål reportedly are more restrictive than, say, the Japanese *no*-construction and the Danish genitival possessive.² The data, however, are fragile, and serious quantitative and qualitative studies would be needed to validate these hypotheses.

Qualia-based proposals Both in the literature on compound nouns and in the literature on possessives, a number of alternative semantic theories have centered around the application of qualia structure; e.g., respectively, Johnston and Busa (1999) and Vikner and Jensen (2001). The notion of qualia structure was introduced as part of the generative lexicon enterprise initiated in Pustejovsky (1991). Qualia-based proposals seem to have some success in explaining semi-productive compound formation. Søgaard (2004) proposes the following principles to account for the phenomena presented above:

- In English, there are no binominal compounds $\alpha\beta$ in which the index of α is linked to Θ_{Agent} of β ’s agentive or telic quale.

²While *dette digt er Shakespeares* (‘this poem is Shakespeare’s’) can only mean ‘this poem belongs to/is read by Shakespeare’ to a speaker of Danish, the Italian equivalent, i.e. *questa poema è di Shakespeare*, is ambiguous between that reading and an agentive one (‘this poem was written by Shakespeare’). Similarly, *Sheekusupia no hon* can reportedly be taken to mean ‘the book written by Shakespeare’, ‘the book Shakespeare reads’ and ‘the book about Shakespeare’ in Japanese (Kikuchi and Sirai, 2003), while only the first two readings are licensed in Norwegian Bokmål.

- In Danish, the index of α is never linked to Θ_{Instr} of β 's agentive or telic quale if the index of β is linked to Θ_{Agent} of that quale.
- English and Danish have a fifth quale for contour. Either Estonian doesn't have such a quale, or it is just unavailable in compound formation.

The comparative view tells us what kind of semi-productivities to look for. In fact, most qualia-based accounts of compound semantics are motivated by cross-linguistic comparison (see also Bassac and Bouillon, 2001). Often cited data include Romance compound nouns, in which prepositional linking elements often reflect qualia selection (though see Paggio and Ørsnes, 1993, for some discussion):

(5) coltello di ghiaccio ('knife made of ice')

(6) coltello da ghiaccio ('knife for cutting ice')

I know only of one language in which structural elements disambiguate possessives, namely Yucatec Maya. Some data of interest is presented in Lehmann (1998). In Yucatec Maya, classifiers determine the meaning of the possessives:

(7) *in w-o'ch hàanal*
 my CL-eat food
 'my food (the food which I eat)'

(8) *in mehen hàanal*
 my CL-make food
 'my food (the food which I made)'

While (7) and (8) cannot be seen as evidence against Partee's application of the Variable \mathfrak{R} Hypothesis, they provide interesting heuristics for identifying possibly relevant grammaticalizations. In the case of compound nouns, the comparative view historically led to the discovery that the application of the Variable \mathfrak{R} Hypothesis to the semantics of compound nouns is inadequate. In other words, constructional semantics is identified and established on language-internal grounds, but comparing cross-linguistic data suggests where to look for the evidence.

On the basis of these findings, how is the semantics of compound nouns and possessives properly implemented for a language like English?

On grounds of *parallelism*, one might argue that if compound nouns employ lexicalized qualia structures, it is reasonable for the compositional semantics of possessives to make use of these too. In a grammar based on type hierarchies, the encoding of qualia structure and the compositional semantics of compound nouns can be partially reused when the grammar is extended to cover the various possessive constructions. In other words, it is possible to let the semantics of compound nouns and possessives be specific instances of more general linguistic types.

It may still be possible to prefer a pragmatic account of possessives on grounds of *efficiency*, while disambiguating compound nouns in the compositional semantics, since underspecified predicates minimize the number of output analyses drastically. In addition, this split approach can be preferred on grounds of *conservativity*, i.e. usually it is not recommendable to posit structure that is not evidently needed.

4 Conclusion

It was demonstrated how the contrastive perspective is a fruitful one in the study of the semantics-pragmatics interface. While the contrastive perspective is a first methodological step in identifying possible semi-productive formation rules, the crucial evidence is intralingual. In the case of compound nouns, such formation rules can be identified, while there is no clear evidence for similar rules for possessives. This asymmetry raises interesting problems for grammar engineering.

5 References

- Allen, M. 1970. Morphological investigations. Ph.D. thesis, University of Connecticut.
- Bassac, C. & P. Bouillon. 2001. The telic relationship in compounds in French and Turkish. 1st International Workshop on Generative Approaches to the Lexicon. Geneva, Switzerland.
- Bauer, L. 1978. The grammar of nominal compounding. Odense: Odense University Press.
- Bauer, L. 1979. On the need for pragmatics in the study of nominal compounding. *Journal of Pragmatics* 3: 45-50.
- Copestake, A. & A. Lascarides. 1997. Integrating symbolic and statistical representations: the lexicon-pragmatics interface. Proceedings of ACL-EACL 1997. Madrid, Spain.
- Goard, J. 2003. Definiteness and English prenominal possessives. First International Workshop on Current Research in the Semantics-Pragmatics Interface. Michigan, USA.
- Hiramatsu, K. et al. 2000. On musical hand chairs and linguistic swing. Proceedings of the 24th Boston University Conference on Language Development. Somerville, USA.
- Johnston, M. & F. Busa. 1999. Qualia structure and the compositional interpretation of compounds. In E. Viegas (ed.), *Breadth and depth of semantic lexicons*. Dordrecht: Kluwer.
- Kikuchi, R. & H. Sirai. 2003. Analysis and interpretation of Japanese postposition 'no'. Workshop on the linguistic dimensions of prepositions and their use in computational linguistics formalisms and applications, ACL-SIGSEM. Toulouse, France.
- Lehmann, C. 1998. Possession in Yucatec Maya. Unterschleißheim: LINCOM.
- Liberman, M. & R. Sproat. 1992. The stress and structure of modified noun phrases in English. In A. Szabolcsi and I. Sag (ed.), *Lexical matters*. Stanford: CSLI.
- Paggio, P. & B. Ørsnes. 1993. Automatic translation of nominal compounds. *Rivista di Linguistica* 5: 129-156.
- Partee, B. 1997 [1983]. Genitives, a case study. J. v. Benthem and A. t. Meulen (eds.), *The handbook of logic and language*. Dordrecht: Elsevier.

- Partee, B. & V. Borschev. 1998. Integrating lexical and formal semantics: genitives, relational nouns and type-shifting. R. Cooper and Th. Gamkrelidze (eds.), *Proceedings of the 2nd Tbilisi Symposium on Language, Logic and Computation*. Tbilisi, Georgia.
- Pustejovsky, J. 1991. The generative lexicon. *Computational Linguistics* 17: 409-441.
- Søgaard, A. 2004. A compound matrix. *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar*. Leuven, Belgium.
- Vikner, C. & P. A. Jensen. 2002. A semantic analysis of the English genitive. *Studia Linguistica* 56: 191-226.