A Domain-Independent Model for the Automated Interpretation of Nominal Compounds

Cécile Fabre
cfabre@irisa.fr

Abstract

Our project focuses on the calculability of the semantics of English nominal compounds. Our goal is to design a general model, based on domain-independent lexical information, to describe general interpretation mechanisms that can be used in unrestricted texts. We use linguistically motivated rules to retrieve the relation between the nominal constituents. In particular, we demonstrate how the system of relations that describes nominal semantics in J. Pustejovsky's *The Generative Lexicon* allow us to formulate general mechanisms for the interpretation of compounds when the underlying predicate is implicit.

1 Motivation

Our project focuses on the calculability of the semantics of nominal compounds in English. These sequences, such as *system interface* or *drainage pump* are very frequent in technical texts. Compounding is a productive linguistic process which provides a way of forming new denominations and enriching the terminology of a domain by the adjunction of modifiers[1] to existing nouns or nominal sequences. Our purpose is to calculate the semantics of these binominal sequences, that is to formulate the domain-independent principles which allow us to automatically infer the relation between the constituents from their lexical features. Our aim is thus opposite to Ter Stal's (1996) who has designed an efficient model for the interpretation of compounds in a technical domain. In his system, the interpretation is drawn from a rich conceptual definition of the objects of the domain (*ontology-based interpretation*).

The first interest of domain-independent rules is reusability. Several studies have shown that pragmatic knowledge is required to perform compound interpretation (Bauer, 1979; Ryder, 1994), but few useful principles

[1] In a noun + noun sequence, we distinguish between the head of the sequence — generally the rightmost constituent in English — and the modifier. The modifier characterizes the head in some way. In this work, we only focus on non-recursive terms because compounds with three constituents or more raise furthermore the problem of ambiguous bracketing (Stal, 1996; Resnik, 1993).

have been formulated about the kind of conceptual knowledge involved. We want to exhibit general interpretation principles that can be applied in any domain, to guide the identification of semantic patterns in specific ontologies. Secondly, this is a necessity if we aim at an application to unrestricted texts: in our case, we are studying the integration of such semantic information in an information retrieval system that handles unconstrained lexical data[2].

Domain-independent systems (Finin, 1980; Donald, 1982) have been built for the automated interpretation of compounds. Yet, these systems do not focus on the definition of explicit and linguistically motivated rules of interpretation: in MacDonald's system, an interpretation is identified if one constituent can fill a slot of the other. These slots represent any piece of real-world knowledge that contributes to the definition of the semantics of a noun; consequently, it is difficult to control and explain the way interpretations are constructed in this model, since there is no detailed account on the principles that underlie the knowledge base generation. Finin's model includes the definition of rules that are not inferred from the properties of the constituents, namely idiomatic rules or *productive rules*, in opposition to *structural rules* which are based on morpho-semantic principles. In both systems, frequency and probability scores are added to the rules. Such numeric weighing of general semantic rules is hardly defensible in the absence of any reference to a domain. Consequently, our own position is to argue for a clear distinction between two tasks: the determination of the possible interpretations of a compound given the syntactic and semantic properties of the constituents and the selection of the most probable meaning (mainly resulting from a lexicalization process)[3]. This selection involves domain-specific or text-specific knowledge whereas domain-independent mechanisms can contribute to account for the interpretation of compounds in isolation.

Consequently, we have designed a model of interpretation which account for productive patterns of interpretation, independently of any domain. The aim of our research is to define as precisely as possible the border line between what can be regularly described with general linguistic mechanisms, and what has to do with subregular or irregular phenomena which depend on corpus characteristics. This is a crucial issue when dealing with compound semantics because regular semantic patterns (involving relational properties of nominals) and extralinguistic data are mingled.

---

[2]This project is supported by the CNET, National Center of Telecommunications Studies (contract CNET-INRIA no951B030). The information retrieval system is part of the french Minitel service; its characteristics are open-domain and automated indexing based on conceptual representation of texts (Gilloux, Lassalle, and Ombrouck, 1993).

[3]For example, the meaning settled for the compound *whale boat* is: "a boat which is used to hunt whales". Nevertheless, this compound may also be interpreted as "a boat used to transport whales", as in *cattle boat*. The first meaning has been lexicalized, i.e. established from the common use of the compound (Downing, 1977).

Section 2 describes the linguistic and conceptual mechanisms that allow us to link predicative information to nouns and to infer compound interpretation from argument selection principles. Concerning root nouns, we make extensive use of Pustejovsky's principles as described in *The Generative Lexicon*(1995). The second Section focuses on the implementation of these mechanisms in our model and shows that the interpretation is performed in two steps: firstly, the list of predicates that are linked to the constituents is retrieved from their lexical description; secondly, only the predicates that can assign a role to both constituents are retained. Finally, we evaluate the results of our model; two points are emphasized as characteristic of our implementation: the disambiguation of the constituents and the generation of multiple interpretations.

2 Principles for the retrieval of predicative information

Interpreting nominal compounds in our model consists in retrieving the semantic relation between the constituents from their morphological, syntactic and semantic characteristics. The recovery of the missing predicate may be more or less problematical depending on the clues provided by the compound. While it is rather straightforward to calculate the relation in the first example below, the interpretation of the other two seems far less obvious[4]:

sleeping pill =
 sleep(instrument :<u>*pill*</u>)[5]
seasickness pill =
 heal(instrument :<u>*pill*</u>, theme: *seasickness*)
antihistamine pill =
 compose(agent :*antihistamine*, theme :<u>*pill*</u>)

In the first example, the relation is easily grasped since the verbal predicate is made explicit by the presence of a deverbal noun (i.e. a noun morphologically derived from a verb). Compounds with a deverbal constituent are called verbal compounds (Selkirk, 1982). On the contrary, in the second example, the verbal predicate cannot be directly recovered on the basis of regular morpho-semantic operations: the noun *pill* is only semantically linked to the predicate *heal*. The third example shows that a single noun may be linked to several predicates according to the semantic information provided by the other constituent. We therefore need conceptual data in order to decide which underlying relation can be inferred from morphologically simple constituents in root compounds.

[4]These examples are taken from (Bauer, 1979).

2.1 Verbal compounds

Compounds with a deverbal constituent have been thoroughly described in the generative framework (Selkirk, 1982; Lieber, 1983). The predicate is explicit and one constituent saturates an argument of the deverbal constituent. Linguistic principles allow us to calculate the interpretation of nominal compounds of the form N V-suf and V-suf N, where V-suf refers to action deverbals (*coverage*, *parsing*) or agentive deverbals (*parser*): when the deverbal noun is the head of the compound, the non-head may saturate an internal or an adverbial argument of the head:

sentence parsing =
  parse(theme : *sentence*)
sentence parser =
  parse(instrument : *parser*, theme : *sentence*)
hand parsing =
  parse(means: *hand*)

When the deverbal noun is the non-head, it cannot satisfy its internal argument within the compound, so that the head may only saturate an external or an adverbial role:

testing procedures =
  test(instrument: *procedures*)

For a detailed account of these principles, see (Selkirk, 1982; Lieber, 1983; Sébillot, 1993).

This first series of compounding patterns, where semantic interpretation reflects morphological structure, has been considered as the only type of compound which can be described in semantic terms (Selkirk, 1982). In Ter Stal's model (1996), the distinction between root and synthetic compounds (in which the head is an argument-taking nominal, that is a deverbal (Lieber, 1983)) is a crucial linguistic principle: for root compounds, the initial semantic representation mentions an underspecified semantic relation, whereas the morphologically related predicate is retrieved for synthetic compounds.

Our own position is to argue that the same predicate-argument pattern may be applied to deal with other types of compounds, provided that we rely on a richer semantic representation of nominals, when no morpho-syntactic information are available to constrain the semantic interpretation.

2.2 Root compounds

The distinction between deverbal and non deverbal nouns does not cover the distinction between predicative nouns, which govern syntactic arguments, and non predicative nouns. Valency does not only apply to deverbals: a category of morphologically simple nouns has an argument structure (Isabelle, 1984). For example, in the phrase "to make an effort to do something", it is the noun *effort*, and not the support verb *make*, that governs the complement. Properties of predicative nouns can be used to calculate the meaning of compounds when the other constituent satisfies one of the arguments of the predicative noun. In our model, the predicate is given by the association between the predicative noun and its support verb:

> *reform effort* =
> *make-effort*(theme: *reform*)
> *launch opportunity* =
> *have-opportunity*(theme: *launch*)

It is more problematic to assign an interpretation to a compound when the two constituents are non predicative nouns. For the most part, these nouns are concrete nouns. Psycholinguistic studies (Downing, 1977; Ryder, 1994) show that in such case the interpretation consists in the identification of an underlying event structure that can link the referents of the two constituents: conceptual knowledge is required when no explicit linguistic information are available. Psycholinguistic experiments attest that regular relational patterns are activated to interpret new compounds: when native English speakers are asked to give a definition for a new N N presented in isolation, their answers show regular interpretation strategies, though the list of the predicative relations that can be conceived between the two referents of the constituents seems unlimited. Thus, our aim is to define the predicative relations associated to a noun that are typical enough to be implicited in a compound and be recovered without the assistance of contextual information. Previous work related to compound interpretation gives only partial solutions to answer this question: P. Downing (1977) shows that the semantic relationship between the two constituents largely depends on the semantic class of the head noun, but the interpretation is obtained with a certain degree of approximation. T.W. Finin (1980) defines a category of nouns that is typically linked to a functional predicative information: a *role nominal* is tied to a characteristic activity in which it participates, in the same way as an agentive deverbal. But none of these models gives a synthetic view of the principles that govern the semantic association between non predicative nouns; in particular, none of them goes into the problem of the relation between nouns and conceptual knowledge, to account for typical relations between referents of the nouns.

The idea that noun meaning involves event-based description has been emphasized by J. Pustejovsky (1991; 1995) in another context, in particular to account for the phenomenon of type coercion. We propose to apply a crucial component of his *generative lexicon*, the *qualia structure*, to the semantic interpretation of compounds. The key idea that underlies the *qualia structure* is that nouns are implicitly related to predicative information that are projected into linguistic structures such as verb sucategorization ("to begin a book" means "to begin *reading* or *writing* a book"), adjectival modification ("a fast book" means "a book *that can be read* in a short time"), etc. The features of the qualia structure express the various semantic *facets* of the nouns that correspond to the typical relations structuring nominal representation. We support the hypothesis that Pustejovsky's model provides a theoretical framework to deal with implicit predicative information in compounds. Nominal facets that introduce relational information are the *telic* role, that refers to the purpose and function of the referent, the *agentive* role, that concerns the factors involved in its origins, the *constitutive* role, that captures the relation between an object and its constituent parts, and the *formal* role, that distinguishes the object within a larger domain. Each role is illustrated in the following compounds:

> *shirt box* =
> > *contain*(locative : $\underline{\textit{box}}$, theme : *shirt*)
> > (telic role)
>
> *bullet wound* =
> > *cause*(agent : *bullet*, theme : $\underline{\textit{wound}}$)
> > (agentive role)
>
> *iron gate* =
> > *compose*(agent : *iron*, theme : $\underline{\textit{gate}}$)
> > (constitutive role)
>
> *water temperature* =
> > *characterize*(agent : $\underline{\textit{temperature}}$, theme : *water*)
> > (formal role)

Pustejovsky's model gives an understanding of the general principles of interpretation for root compounds; it suggests how the principles of predicate-argument selection can be extended. Unlike deverbals, root nouns are not provided with an argument structure that may be syntactically satisfied; nevertheless the argument structure of the related verb correspond to the distributional properties of the noun. For example, the verb *heal* which corresponds to the telic role of the noun *pill* requires a subject and an object argument; since the noun *pill* refers to its first argument, the position which is left empty (the theme) may be occupied by the first constituent of a compound of the form *N pill*, as in:

> *seasickness pill* =
> > *heal*(instrument :$\underline{\textit{pill}}$, theme :*seasickness*).

Thus, our model is built upon the assumption that the interpretation of verbal and root compounds is based on the same principle of argument selection; however, whereas the recovery of the underlying predicate requires morpho-syntactic rules in the former case, conceptual relations related to the head, i.e. qualia roles, are involved in the latter.

3 Interpretation mechanism

In our model, the first step of the interpretation mechanism consists in retrieving the list of predicates that are potentially associated with the constituents. In the second step, the predicates that cannot assign a role to the other constituent are eliminated via selectional restrictions.

We use a conceptual hierarchy. Firstly, the interpretation of nominal compounds requires a detailed semantic sorting of the nouns to handle selectional constraints. Secondly, a rich semantic classification is needed in order to distribute predicative information, when the same conceptual description applies to all elements of a class. We have chosen WordNet[6] to handle rich but non-specialized conceptual information. We use the upper part of the arborescence which provides generic nominal classes such as INSTRUMENTALITY, PLANT, MATERIAL, COMMUNICATION, etc.

3.1 Lexical description

Predicative information is part of the lexical definition of the nouns. We distinguish between two cases:

1. the predicative information associated with a noun cannot be generalized to a whole class of nouns. It appears directly in the lexical entry of the noun. For example, the predicate *wash* is typically associated with the noun *soap*, via its intrument role. This information cannot be linked to a whole class of nouns (unless we define a class of CLEANSER)[7] and is thus specific to the lexical entry of the noun *soap*.

2. the predicative information may be generalized to a whole class of nouns. It is moved up to the level of the class itself. For example, the noun *box* shares the definition "something that holds things" with any member of the class CONTAINER. The telic role is thus expressed at the level of the class:

   class(CONTAINER →
       telic role =
           contain(locative : CONTAINER, theme : ENTITY))

This states that a CONTAINER refers to the locative role of the underlying verbal predicate *contain*, and that the empty role (the theme) can be satisfied by a constituent of the class ENTITY, as in *shirt box*, or *gas tank*.

Similarly, all members of a class can share the same agentive, constitutive or formal role:

[6]WordNet is a trademark of Princeton University.

[7]The class CLEANSING AGENT, CLEANSER is available in the WordNet hierarchy but we discard it because the conceptual information it conveys is too specific .

class(INJURY →
agentive role =
cause(agent : ENTITY, theme : INJURY))
(shrapnel wound, snake bite)

class(ARTIFACT →
constitutive role =
compose(agent : SUBSTANCE or MATERIAL,
theme : ARTIFACT))
(velvet dress, stone wall)

class(ATTRIBUTE →
formal role =
characterize(agent : ATTRIBUTE, theme : ENTITY))
(fuel shortage, color intensity)

The difficulty lies in determining the predicates that define these attributes. Since there is no morphological evidence of the relation between the noun and the predicate, these predicative information are related to stereotyped conceptual knowledge about the referents of the noun[8]. Pustejovsky, Anick, and Bergler (1993) show that this information can be retrieved from corpus observations, on the basis of the mutual information value between the noun and the verbs with which they co-occur.

Each noun is connected to the conceptual hierarchy through its semantic label. For example, the word *pill* is categorized as MEDECINE and is linked to the superordinates:

MEDECINE → ARTIFACT → OBJECT → ENTITY

A noun may therefore inherit several predicates, as illustrated in Figure 1.

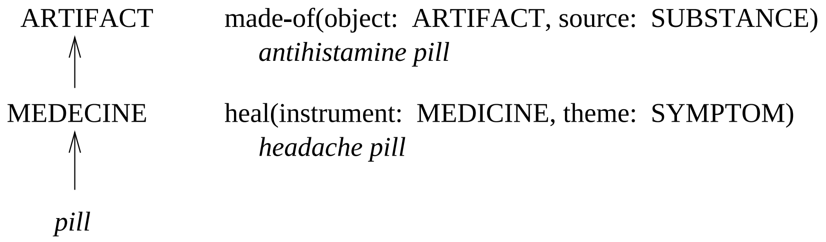

Figure 1: The word *pill* attached to the lexical hierarchy

Deverbals are also linked to the semantic hierarchy and can therefore be related to multiple predicates. For example, the verbal base of the noun

[8] Miller et al. (1990) suggests that such functional attributes of nominal concepts should be added in WordNet by means of pointers from nouns to verbs.

signal illustrates only one relational attribute (the telic role) of the noun:

> *temperature signal* =
> $signal(\text{instrument} : \underline{signal}, \text{theme} : temperature)$

As any member of the class COMMUNICATION, the noun *signal* can also be characterized by its origin (the origin of the communication):

> *satellite signal* =
> $signal(\text{agentive} : satellite, \text{theme} : \underline{signal})$

3.2 Predicate selection and constituent disambiguation

The interpretation consists in deciding which predicates are compatible with both constituents and which are eliminated according to selectional restrictions. The semantic class of the non head enables the choice between several available interpretations. In a domain-independent model, it is not possible to express fine-grained semantic features on the argument positions of the predicates. Consequently, several interpretations are generated when several predicative relations apply. Let us illustrate the interpretation mechanism on the example *fish disease*. Two predicates are retrieved in connection with the head noun *disease*. The first one refers to the telic role:

> $disease \rightarrow$
> $affect(\text{agent} : disease, \text{theme} : \text{BEING or BODY-PART})$

The second predicate refers to the agentive role:

> $disease \rightarrow$
> $cause(\text{agent} : \text{ENTITY}, \text{theme} : disease)$

Our program generates three interpretations for this compound:

> $affect(\text{agent} : disease, \text{theme} : fish_1)$
>
> $cause(\text{agent} : fish_1, \text{theme} : disease)$
>
> $cause(\text{agent} : fish_2, \text{theme} : disease)$

The noun fish has two distinct meanings: the first one refers to the animal (class ANIMAL) and the second to its flesh (class FOOD). Consequently, the telic attribute applies to the first meaning of the non-head (the restriction on the theme role is satisfied). Regarding the agentive attribute, the predicative relation *cause* applies to both meanings, because the restriction on the agent role is very loose. The first interpretation ("a disease that affects fish") and the second one ("a disease that is caused by consuming fish") are equally plausible and only fine pragmatic knowledge can determine which interpretation is the correct one in a given context. The third one ("the disease is caused by fish, but not directly by consuming fish") may be less instantly conceived, but is also valid. These examples illustrate two main characteristics of our model: firstly, it generates any solution that can be derived from the general mechanisms that we have described. Secondly, the identification of a predicate-argument structure and the use of selectional restrictions contributes to disambiguate the constituents, as illustrated by the distinction between the two meanings of *fish*.

4 Evaluation

Our model for the automated interpretation of English nominal compounds has been implemented and tested on a list of 100 sequences in isolation. Since these sequences are deconnected from their original context, the point is not to calculate the most probable meaning but to generate any meaning that could be predicted considering the constituents.

It is difficult to assess the appropriateness of the answers that are produced, since we are dealing with compounds in isolation. Other answers are always conceivable, if we apply less regular principles of semantic associations (Ryder, 1994), so that we cannot compare our results with a closed set of correct answers. Moreover, we cannot set a clear frontier between probable and hardly conceivable interpretations. Having said this, we can estimate our results as follows: 71% of the compounds that we have examined receive acceptable answers, which means that the program generates conceivable interpretations, with a variable degree of plausibility. For example, four interpretations are generated for the compound *cardboard box*:

1. *constitute*(agent : *cardboard*, theme : <u>*box_4, box_5, box_6, box_7*</u>)
   objects made of cardboard (constitutive role)
2. *contain*(locative : <u>*box_7*</u>, theme : *cardboard*)
   box that contains cardboard (telic role)
3. *produce*(agent : <u>*box_3*</u>, theme : *cardboard*)
   plant that produce cardboard (telic role)
4. *measure*(agent : <u>*box_2*</u>, theme : *cardboard*)
   a quantity of cardboard (formal role)

Interpretations 2, 3 and 4 are surely mistaken in a standard context, if we refer to extralinguistic knowledge (*box_3* — a kind of shrub — does not produce cardboard, the way *gum trees* produce gum) or to lexicalization; the compound *cardboard box* has only one usual meaning, namely *constitute*(agent : *cardboard*, theme : <u>*box_7*</u>), where *box_7* refers to the container. Yet, each answer is conceivable because it corresponds to productive semantic patterns and therefore to existing cognitive strategies.

29% of the answers that are generated are incorrect: they miss expected answers (6%) or give no answers at all (23%). These results are satisfactory if we keep in mind that only domain-independent principles are applied and most of all that a few general principles of predicative attachment are formulated. Therefore, we are able to determine the limits of domain-independent mechanisms in analysing uncorrect and missing answers. The two main problems are inappropriate selectional constraints and unpredicted semantic patterns. The mechanisms that we have implemented cover typical interpretation patterns. Yet, supplementary information (mainly extralinguistic) or more complex interpretation schemes are involved in some cases:

- the predicative information may be associated with the non-head: when the head refers to an underspecified event structure, the semantics of the non head is crucial for the determination of the correct predicate:

  malaria program =
  $\quad$ *fight*(instrument : $\underline{program}$, theme : *malaria*)
  crop program =
  $\quad$ *develop*(instrument : $\underline{program}$, theme :*crop*)

  This is an illustration of the notion of co-compositionality (Pustejovsky, 1991). In most cases, it occurs when the head refers to a generic notion of location (*steak house*) or instrumentality (*wood machine*).

- the predicative information is too specific to be associated with the head noun as a typical conceptual information. For example, the noun *fly* is not connected to a typical telic role, yet a specific purpose is expressed in the compound *fruit fly*, i.e. "feed on").

- relations such as subclass (*marathon tour*) or ressemblance (*carpet shark*) are not taken into account in our model, because they cannot be retrieved from structural principles.

- more complex inference principles are needed in order to interpret contextual compounds (Downing, 1977) or compounds in which several terms are implicited (*cocaine babies*). Fine-grained extralinguistic data are involved in the interpretation of such compounds.

Generation of multiple interpretations and unpredicted patterns due to selectional violation or extralinguistic information are thus the two inherent limits of a domain-independent model of interpretation.

5 Conclusion

We have implemented a model that computes the interpretation of nominal compounds from general linguistic and conceptual principles about the predicative properties of nouns. Our main contribution consists in extending the predicate argument-structure for the interpretation of compounds with non predicative constituents, and to define mechanisms to go into the integration of conceptual knowledge in the interpretation rules. This is crucial to deal with a linguistic phenomenon that involves implicit predicative information. Predicative nouns (deverbals or not) and relational attributes of concrete nouns as described by J. Pustejovsky in *The Generative Lexicon* are taken into account to retrieve the relation that underlies English nominal compounds.

Besides, this non-specialized model of interpretation allows us to draw a comparison with nominal sequences across languages, and especially with French sequences of the form "N de N" and "N à N", in which the prepositional link is semantically weak (Fabre and Sébillot, 1994). The distinction between verbal compounds

> *séquençage d'ADN* =
> *séquencer*(theme : *ADN*)

and root compounds

> *four pain* =
> *cook*(instrument : <u>*four*</u>, theme : PAIN))

also holds in French; the relation between root nominal constituents and predicative structures can be described in the same way for N N and N *de* N binominal sequences. Multigeneration is limited in French because the prepositional link and the determiner help to partially disambiguate the interpretation in the second step of the analysis.

The main difficulty lies in the identification of the predicate and its selectional restrictions in the relational attributes. In a domain-independent model, these data are necessarily coarse-grained. Our assumption is that the definition of such mechanisms in a general framework can help the extraction of significant data in the texts (Fabre, 1996), as illustrated by Pustejovsky, Anick, and Bergler (1993).

From an application-oriented perspective, we are currently studying the use of such a semantic representation of binominals in text indexing. The identification of the predicative relation between the constituents is a means to improve the accuracy of the representation of a text content. Semantic analysis based on mere cooccurrence relations is replaced by a structured representation that is used to disambiguate the constituents of the sequence and to increase the generality of the descriptors. We are currently experimenting this hypothesis in the CNET indexing system.

References

Bauer, Laurie. 1979. On the Need for Pragmatics in the Study of Nominal Compounding. *Journal of Pragmatics*, 3:45–50.

Donald, D.B. Mc. 1982. *Understanding Compounds Nouns.* Ph.D. thesis, Carnegie Mellon University.

Downing, Pamela. 1977. On the Creation and Use of English Compound Nouns. *Language*, 4(53):810–842, April.

Fabre, Cécile. 1996. Interpretation of Nominal Compounds: Combining Domain-Independent and Domain-Specific Interpretation. In *Proceedings of Coling, to appear.*

Fabre, Cécile and Pascale Sébillot. 1994. Interprétation sémantique des composés nominaux anglais et français. In *Proceedings of the Workshop on Compound Nouns : Multilingual Aspects of Nominal Composition*, Genève, Suisse.

Finin, Timothy Wilking. 1980. *The Semantic Interpretation of Compound Nominals.* Ph.D. thesis, University of Illinois.

Gilloux, M., E. Lassalle, and J-M. Ombrouck. 1993. Interrogation en langage naturel du Minitel Guide des Services. *Echo des Recherches.*

Isabelle, Pierre. 1984. Another Look at Nominal Compounds. In *Proceedings of Coling*, pages 509–516.

Lieber, Rochelle. 1983. Argument Linking and Compounds in English. *Linguistic Inquiry*, 2(14):251–285.

Miller, George, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller. 1990. Five Papers on WordNet. Technical Report CSL 43, Cognitive Science Laboratory, Princeton University, July.

Pustejovsky, James. 1991. The Generative Lexicon. *Computational Linguistics*, 4(17).

Pustejovsky, James. 1995. *The Generative Lexicon.* MIT Press.

Pustejovsky, James, Peter Anick, and Sabine Bergler. 1993. Lexical Semantic Techniques for Corpus Analysis. *Computational Linguistics*, 19(2):331–358.

Resnik, Philip S. 1993. *Selection and Information : a Class-Based Approach to Lexical Relationships.* Ph.D. thesis, University of Pennsylvania.

Ryder, Mary Ellen. 1994. *Ordered Chaos : the Interpretation of English Noun-Noun Compounds.* University of California Press.

Sébillot, Pascale. 1993. Sémantique des composés anglais : approche générative, limites et applications. In *Proceedings of "Informatique et Langue Naturelle", ILN'93.*

Selkirk, Elisabeth. 1982. The Syntax of Words. *MIT Press.*

Stal, Wilco Ter. 1996. *Automated Interpretation of Nominal Compounds in a Technical Domaine.* Ph.D. thesis, University of Twente, Netherlands.