NN COMPOUNDS IN ITALIAN: MODELLING CATEGORY INDUCTION AND ANALOGICAL EXTENSION

Marco Baroni[*] Emiliano Guevara[**] Vito Pirrelli[***]

1. INTRODUCTION

Word compounding, a phenomenon common to all human languages, constitutes an object of interest for many fields in the domain of cognitive science. Both theoretical linguists and typologists – since Pāṇini's *Aṣṭādhyāyī* (*circa* 520 BC) – have striven to provide classificatory and explanatory accounts of the common traits of compounds across languages, as well as of their language-specific peculiarities. For cognitive psychologists and psycholinguists, the focus of inquiry has mainly been on crucial aspects of concept-combination in compound formation and interpretation. More recently, computational linguists have emphasized the critical role played by compounds – as well as by other multi-word expressions – in many Natural Language Processing tasks such as text segmentation, POS-tagging and parsing.

Each of these different approaches has generated a considerable body of dedicated literature, theoretical insights and practical methods. However, each sub-field has remained to a great extent isolated from all the others, at times giving rise to a proliferation of overlapping experimental tasks and theoretical accounts of closely-related phenomena. We would like to suggest that a unified analysis of compounding is possible and desirable. Needless to say, any attempt in this direction has to face methodological issues of considerable complexity, but the perspective of obtaining an integrated approach to compound words that is at the same time theoretically sound, cognitively reliable and quantitatively verifiable will – should the attempt prove correct – be worth the effort.

---

[*] CIMeC, Università di Trento, Italy.
[**] Dipartimento di Lingue, Università di Bologna, Italy.
[***] Istituto di Linguistica Computazionale, Pisa, Italy.

We gratefully acknowledge the support of the Italian Ministry of University and Research. The research described here has been partially funded by the *CompoNet* national project (PRIN 2005) coordinated by S. Scalise. We would like to thank our *CompoNet* collaborators, as well as Alessandro Lenci, who contributed to the creation of the word space model we used in our experiments.

LINGUE E LINGUAGGIO VI.2 (2007) 263–290

In this article, we report a series of experiments inspired by recent work in psycholinguistics and theoretical linguistics, carried out by applying computational/quantitative techniques. Unlike most of the existing literature focusing on compounding in English and other Germanic languages, we concentrate our attention on Italian NN (noun+noun) compounds. The most apparent difference between Italian NN compounds and more familiar English-like compounds is that the former are left-headed: i.e. the morpho-syntactic and semantic head (H) of the whole construction is the first component of the NN pair, while the second constituent acts as a modifier (M); cf.:

(1) *capo-stazione* (m.) 'lit. head+station, station master'
  a. *capo* (m.) = H
  b. *stazione* (f.) = M.

Italian NN compounds thus allow us to assess the influence of the order of nouns in compound processing, and to control our evidence against that of languages, like English, where the HM order is reversed. Moreover, our preliminary corpus-based survey of Italian compound types carried out in the framework of the Italian CompoNet Project (Baroni *et al.*, in press) bears witness to a wide variety of attested patterns and to a growing number of different types for each pattern. Such a wealth of data, somewhat unexpected in a Romance language, can be interpreted as marking a turning point in the morphological productivity of Italian compounds, which show signs of increasingly more liberal usage. We believe that, by inspecting distributional patterns of usage of a prospectively compositional system, we can gain new and sharper insights into the linguistic and cognitive processes underlying both lexical and conceptual composition.

The paper is structured as follows: in Section 2 we present an overview of the cognitive literature on concept-combination in compounding; in Section 3 we provide a brief overview of Bisetto & Scalises's (2005) classificatory scheme for compounds in the world's languages; Section 4 contains a sketchy outline of the novel model for compound interpretation we intend to propose and put to an empirical test; in Section 5 we report a clustering experiment testing if and how the important distinction between the categories of relational and attributive compounds can emerge with no supervision from corpus-based data; we then report an experiment with human subjects (Section 6) which models new compound generation as an analogical process based on frequently attested compound schemata.

Taken together, the unsupervised simulation and the analogical model provide an account of how different types of compound arise and are extended in Italian. In turn, this evidence lends support to the compound typology proposed both in the cognitive literature and by Bisetto & Scalise (2005).

2. THE COGNITIVE ACCOUNT

From a psychological and cognitive perspective, compounds are used in communicative contexts as tools for concept combination. The combination refers to a category that differs in some way (being a somewhat more specific instantiation) from the category conveyed by the compound head. For example, an *apartment dog* is a kind of dog, whose size and behaviour make it suitable for living in apartments. The communicative purposes that are served by compounding are manifold and may range from the need for designating significantly new categories that have important and enduring traits, to the need for referring back, on-line, to a previous referent in the discourse context or to an otherwise pragmatically available entity (as in Downing's famous example of *apple-juice seat*, cf. Downing 1977: 818).

No matter how many different purposes compounding may serve, since Wisniewski's original survey (Wisniewski 1996, 1997) there is substantial agreement that people combine concepts in three basic ways:

- In *relational interpretations*, people assume that there exists a relation between the referents of the two concepts being combined: for example an *apartment dog* is a dog *living* in an apartment, and an *apple-juice seat* is a seat *in front of which* a glass of apple-juice is placed.
- In *property* or *attributive interpretations*, on the other hand, people assume that one or more properties of the dependant or modifier (non-head) constituent are attributed in some way to the head concept: a *ball fish* is a fish whose shape resembles the shape of a ball, rather than – say – a fish playing with a ball or dragging a ball with its fins.
- A third, less frequent, type of interpretation assumed by the cognitive literature involves *hybridization*, whereby the compounded concept is understood as a cross (or a conjunction) between its two constituent concepts as in the case of *musician painter* or *robin canary*.

There is a striking parallelism between these interpretative categories and the tripartite classification of compounding recently proposed by linguists (here reviewed in Section 3). Such a remarkable convergence of independent lines of scientific inquiry bears witness to the solidity of our current understanding of the kinds of output yielded by concept composition. We provisionally conclude that, in its simplest instantiation, interpreting a novel compound requires two interlocked processing steps:

i. identification of the head of the compound,

ii. interpretation of the contextually appropriate *processing link* between the head and its modifier, be it an argument relation, a property transfer or a conceptual hybridization.

However, while these two basic steps are uncontroversially assumed by most scholars, there is considerable disagreement on the logical and algorithmic relationship between them. We take this to be the most central issue in the current cognitive debate on compounding; it is to this issue that we turn now.

2.1 Algorithmic Models

A traditional assumption (Cohen & Murphy 1984; Murphy 1988) is that interpretation of the appropriate processing link (step ii. above) follows, more or less automatically, from identification of the compound head (step i.). Once the head is identified, information coming from its modifier is transferred into the conceptual schema provided by the head.

In relational compounds, transfer consists in saturating one of the argument slots in the sub-categorization frame of the head. For example, in a compound such as *concept combination*, the noun head *combination* provides a grid of argument slots: the combining argument and the combined one. The modifier *concept* may fill in the latter slot in the grid as it complies with the *semantic profile* required by the head for that slot, thereby being understood as the internal argument of *combination*.

In the case of attributive compounds, information transfer amounts to filling in, with some relevant semantic features, one or more available semantic dimensions (e.g. SHAPE, COLOUR, SIZE, WEIGHT, etc.) making up the multi-dimensional conceptual structure associated with the head. Accordingly, in *ball fish*, the semantic dimension SHAPE is made contextually available through the conceptual schema associated with the head *fish*, for the relevant information conveyed by *ball* (SHAPE = round) to be assigned appropriately.

This approach, also known in the literature as the *schema approach* to compound understanding, is radically head-driven. It is the head that provides the backbone for concept combination, the modifier's contribution being limited to the role of fleshing out the head's schema through feature merger or saturation of functional arguments. A corollary of this assumption is that the conceptual representations of both head and modifier are accessed uniformly, independently of contextual requirements. Such representations are conceived of as invariant semantic bricks, assembled together in a complex structure according to well-established building rules. The picture also presupposes a rather sharp subdivision of labour between lexicon and grammar, where the former is conceptualized as a list of fundamentally arbitrary pairs of stem forms and concept representations, and grammar is seen as a generative symbolic system, combining lexical items into abstract formal structures. However simple and elegant, the approach has been the target of extensive criticism on several grounds.

One principled line of criticism stems from the observation that the schema approach relies on a strict notion of *syntactic compositionality* (Jackendoff 1997). According to this notion:

i. the conceptual representation of a complex construction only contains the conceptual representations of the lexical units making up the construction;

ii. principles of semantic composition are uniquely determined by the way lexical units are syntactically combined in complex constructions.

In fact, both these assumptions fail to account for the interpretation of a simple compound like *rocket fuel*. Here, it would be impossible to understand the relation linking *fuel* to *rocket* unless we conceptualize *fuel* as a substance whose proper function is to propel something else through combustion. It is such an internally articulated conceptual structure that explains the potential of *fuel* for combining with *rocket* in a syntactically inert context like a meaningful compound.

Wisniewski (1997) endorses the need for richer conceptual representations in compound interpretation which he dubs *scenarios*. Moreover, in his *dual process* theory, he builds on the schema approach by proposing that attributive interpretations require a preliminary step of *concept alignment*, whereby structural similarities and differences between the corresponding concepts are computed, possibly followed by a step of *concept construal*, which allows combinations to refer to something other than typical members of the head category (e.g. *stone lion*).

In a more radical departure from the schema approach, some scholars have questioned the assumption that access and activation of the whole constituent concepts must precede their integration. For example, Springer and Murphy (1992) compared the time taken to verify a property that is true of a noun (e.g. *celery is green*) versus a property that is true of a phrase headed by the same noun (e.g. *boiled celery is green*). Somewhat paradoxically, their findings show that phrase properties are verified more quickly than noun properties, suggesting that emergent features of the combined concept are activated more quickly than the features of the component concepts. As a way out of this apparent paradox, Maguire *et al.* (2007) entertain the hypothesis that people detect the presence of a compound structure *before* its constituent parts have been interpreted. Following this suggestion, compound interpretation may be understood as a process of selective access of the conceptual information associated with its constituent nouns, where selectional requirements are imposed by the compound construction as a whole. Such construction, far from being the end-result of general rules of concept combination, is the basic prior to the dynamic integration of constructionally-profiled concepts.

A similar insight runs through the work on concept combination and compounding by Shoben (1991), Gagné and Shoben (1997), Gagné (2001), Gagné and Spalding (2006), who suggest an approach to compound interpretation based on thematic relations. Following a well-established linguistic tradition (Downing 1977; Levi 1978), the approach assumes that nouns are combined by determining a fairly general relation holding between them and that there is a relatively small set of such relations (one or two dozens). In developing this view, Shoben and colleagues further assume that a noun's combinatorial history influences the interpretation of a novel phrase involving that noun. That is, people use the distributional knowledge of how nouns have previously combined to interpret a novel combination. For example, when *mountain* is used as a modifier, it typically instantiates a locative relation (e.g., *mountain resort*, *mountain goat*) and is only rarely involved in other types of relations (e.g., *mountain range*). As a result, people tend to interpret a novel combination such as *mountain fish* by using their knowledge that *mountain* is typically used as a locative modifier. Moreover, Gagné and Shoben (1997) show that the time to judge the meaningfulness of a novel compound is a function of the relative frequency of the thematic relations associated with the modifier noun, rather than with the head noun.

In a similar vein, Estes & Glucksberg (2000) claim that the type of property that is transferred from the non head constituent to the head in the interpretation of attributive compounds does not depend on how similar the two corresponding conceptual representations are. Rather, it is determined by the interaction between the relevant defining dimensions in the head's conceptual structure and the salient features in the modifier's conceptual representation. For example, the interpretation of *feather luggage* as 'light luggage' is the result of *light* being a very salient property in the definition of *feather*, and of WEIGHT being a relevant dimension in the conceptual structure of *luggage*. Similarly, Costello and Keane (1997, 2001), in their constraint satisfaction model of concept combination, argue that properties that are diagnostic of a modifier concept are attributed to the head concept. For example, a *cactus fish* is more likely to be interpreted, they say, as a 'prickly fish' than as a 'green fish', because 'prickly' is more diagnostic than is 'green' of cactus.

From a strictly computational standpoint, all such constraint-based approaches make the intriguing suggestion that complex concept interpretation and production are not carried out on the basis of one-off inferential steps (involving both analogical alignment and concept integration, as suggested by Wisniewski). Rather, they make use of pre-compiled, more or less schematized information, which is memorized in the mental lexicon and applied probabilistically in the on-line interpretation of novel constructions.

In our view, this suggestion makes eminent sense for two basic reasons. In the first place, it is in keeping with recent findings concerning the paradigmatic organization of compounds in the speaker's mental lexicon. De Jong *et al.* (2002) report that the interpretation of a Dutch compound of the A+B type like *watermolen* ('water mill') is facilitated by the cardinality of the family of compounds of the X+B type (e.g. *windmolen*, *koffiemolen* etc., where X varies and B = *molen*). Moreover, such a facilitatory effect does not correlate with the frequency of B in other constructions than compounds. If, on the one hand, this evidence lends support to the view that compounds are not memorized as a whole, on the other hand it seems to suggest that they are separately stored as abstract combinatorial structures rather than being processed on-line and understood on the basis of their independent constituents.

Secondly, although it may be the case that one-off compounds are understood through an on-line context-based inferential process akin to analogy and metaphor/metonymy comprehension (Gentner 1983, Glucksberg & Keysar 1990, Glucksberg *et al.* 1997), common or garden compounds are likely to be processed more routinely. Surely, not all of language understanding should necessarily be computationally simple. However, we argue that compounding is a too basic morphological tool for concept composition to be compatible with highly unconstrained and potentially combinatorial processing strategies.

3. THE THEORETICAL LINGUISTICS ACCOUNT

From the point of view of theoretical Linguistics and linguistic typology, the treatment of compound words has focused on the description and classification of the various compound types attested in the world's languages. It is assumed that each different compound type is formed on the basis of a different grammatical/functional process.

In the history of language studies, a wealth of different classificatory schemes have been proposed to account for the variety of compounds in the world's languages. Simplifying an otherwise fairly complex state of affairs, we may say that every time a linguist deals with compounding phenomena, some sort of classificatory sketch is proposed, adopted or slightly modified to describe and organize the data. Needless to say, many of the proposed classifications are contradictory with one another, although some features are common to many of them.

There are linguists who are skeptical of the very possibility of classifying compounds cross-linguistically. For example, Bloomfield (1933: 233–235) expresses this skepticism very clearly, although he acknowledges that some sort of classification may be "useful":

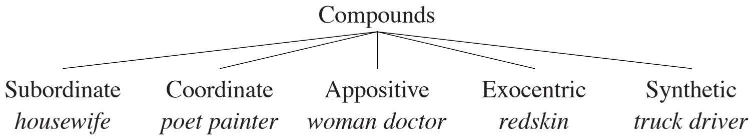
(2) "The description and classification of the forms which the structure of the language leads us to describe as compound words, will depend upon the characteristic features of this language. Linguists often make the mistake of taking for granted the universal existence of whatever types of compound words are current in their own language. It is true that the main types of compound

> words in various languages are somewhat similar, but [. . .] the details, and especially the restrictions, vary in different languages [. . .] preventing our setting up any scheme of classification that would fit all languages, but two lines of classification are often useful. One of these two lines concerns the *relation of the members*. [. . .] The other [. . .] concerns the *relation of the compound as a whole to its members*."

3.1 Heterogeneous or Flat Schemes

All the most recently used and widespread classificatory schemes for compounding (whether tacitly or explicitly adopted, cf. Spencer 1991, Bauer 2001, Olsen 2001, Haspelmath 2002, among many others) assume a primarily flat arrangement of the different compound types in the world's languages, which can be summarised graphically as follows:

(3)



Such a classification is well established in the traditional literature and often taken for granted without criticism. In its basic form, it is nothing else than an extension of Pāṇini's treatment of Sanskrit compounding. However, such a classificatory scheme is a mere nomenclature of types defined on the basis of heterogeneous criteria (i.e. *grammatical relation between the constituents* for Subordinate, Coordinate and Appositive compounds, *presence vs. absence of a lexical head* for Exocentric compounds, *concomitant compounding and affixation* for Synthetic compounds[1]).

A further problem arises when one considers that these classificatory criteria are combinable: for instance, a subordinate compound can be either endocentric (*housewife*) or exocentric (*pickpocket*). All in all, it should be observed that inconsistent and imcomplete classificatory schemes of this kind have been the rule rather than the exception in the literature.

In addition to this, the field has adopted a remarkably unsystematic terminology, which has contributed to creating an even greater confusion. For example, subordinate compounds are often called *determinative compounds*, *root compounds* or *primary compounds*; appositional compounds are sometimes called *attributive*; coordinate compounds are often called *copulative*, *aggregative*, etc.; exocentric compounds are very frequently called *bahuvrihi*

---

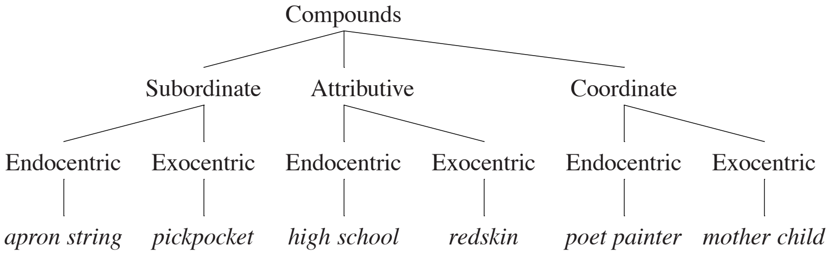
[1] Synthetic compounds like *truck driver* (also called secondary compounds or verbal-nexus compounds) are formed by a deverbal nominal head, and by a first constituent that is interpreted as the internal argument of the head's verbal base.

or *possessive compounds*. Surely, different terms in each of these clusters are not perfect synonyms: each terminological choice, however, usually goes together with a number of theoretical assumptions which are seldom spelled out in detail.

3.2 Hierarchical Schemes

An interesting effort to bring order into compound classification is made by Bisetto & Scalise (2005), who argue that the classification of compounds should comply with a set of homogeneous criteria. In the first place, they identify three macro-types in compounding, each characterised by a different *grammatical relation between the constituents*, i.e. subordination, attribution and coordination. In turn, each of these macro-types is subject to a further subdivision by virtue of a second criterion: the presence or absence of a lexical head. In this way, they further subdivide subordinate, coordinate and attributive types into the *endocentric* and *exocentric* sub-types, as shown below:

(4)



Bisetto & Scalise (2005) argue that the scheme in (4) is able to account for the major compound types in the world's languages. We shall not pursue this issue further here. However, Bisetto & Scalise's proposal is the first explicit attempt to modify a long tradition of unsystematicity in the treatment of compounding in theoretical Linguistics.

In what follows, we will adopt Bisetto & Scalise's proposal as a starting point in the study of Italian NN compounds, for two main reasons.

Firstly, there seems to be no other classificatory scheme that is explicitly and systematically argued for in the literature and that proposes a system that is both simple and hierarchically organized.

Secondly, although the arguments underpinning this proposal are purely theoretical and totally independent of the psychological literature, the three macro-types in Bisetto & Scalise's tripartite system appear to perfectly match the interpretative categories individuated by cognitive studies (cf. Section 2 above). In particular, focusing on the specific phenomenon that is addressed in our present work, we can establish the following correspondences:

- subordinate compounds show *relational interpretations*
- attributive compounds are characterised by *property interpretations*
- coordinate compounds entail conceptual *hybridization*.

4. TOWARDS AN INTEGRATED ACCOUNT OF COMPOUND INTERPRETATION: THE LIS MODEL

Compound interpretation requires integration of the conceptual representations associated with head and modifier. Both from a cognitive and a linguistic perspective, it is tempting to conceive of such an integration as a straightforward information transfer from head to modifier. According to this hypothesis, probably the logically simplest one, conceptual representations of the compound constituents are first independently accessed and then integrated on the basis of the selectional requirements profiled by the head. In section 2, however, we overviewed some reasons to believe that such a strictly compositional view is fundamentally flawed. Access to conceptual representations is considerably more dynamic and context-sensitive, so that the whole construction appears to prompt a process of selective activation of contextually-relevant semantic properties.

As an alternative to a strictly compositional view of compound interpretation, it has been suggested (Wisniewski 1996, Wisniewski 1997) that compounds are processed on the basis of contextually-conditioned alignment and construal processes, akin to analogy and metaphor understanding. We object that, although this may hold for one-off anaphoric and pragmatic compounds, other more constrained processing strategies must apply in processing ordinary compounds, especially those that are linguistically interpretable in the absence of a pragmatically available situation. Recent work by Wisniewski & Clancy (2004), based on a large number of novel combinations (744) in naturally occurring discourse contexts from magazine and newspapers articles, shows that a very large percentage of these contexts (86%) did not mention both the modifier and head noun of the novel combination prior to its occurrence. These results suggest that novel compounds are often coined that are believed to be readily understood by the reader with no explicit contextual indication of how their meaning relation should be understood.

As already pointed out in Section 2, there is converging evidence that the type of knowledge required for context-free interpretation of novel compounds is mostly based upon previous exposure to similar word combinations in other contexts (Shoben and Gagné 1997, Gagné 2002, Gagné and Spalding 2007). We suggest (following De Jong *et al.* 2002) that the mental lexicon is the place where distributional and interpretive knowledge of this kind is stored, based on the prior availability of a family of lexically-related compounds, all sharing what we may call a constant *lexicalized interpretation schema* (LIS). By assuming that the long-term lexical store has a superpositional nature, we hypothesize that a LIS is an abstract constructional pattern (in Goldberg's 2006 sense of a pairing of formal and semantic relations) shared by all members of the same compound family. Schemata of this kind are *lexicalized*, since we assume that a constant lexical constituent B (hereafter referred to as the schema *pivot*) provides the main clue to the intended interpretive relation. However, unlike Shoben and colleagues, we conjecture that B is not to be identified with the compound modifier M in *all* schemata. Rather, the pivot role in compound interpretation can be played by either the head H or M depending on the compound type. In clearly relational compounds such as *sugar box*, it is the X BOX schema (and not a putative SUGAR X schema) that conveys most constraints on compound interpretation. By contrast, an attributive compound like *feather luggage* is interpreted on the basis of a FEATHER X schema.

There are two measurable consequences of this view. First, a crucial assumption of our account is that emergence of a given lexicalized schema in the lexicon should follow from recurrent distributional properties of the members of a schema-sharing family. This means that in relational compounds heads should recur more systematically than modifiers both in terms of token frequency (i.e., the number of times members of a compound family are found in real contexts of use) and type frequency (i.e., the size of the family). The prediction for attributive compounds is that the reversed distributional pattern obtains. The second crucial consequence is that if we generate a new compound of an X + B (or B + X) type by replacing the pivot B, the resulting compound will be more difficult to interpret than the one resulting from replacement of the non pivot constituent. For example, a newly-created compound like *wing luggage*, where *wing* replaces *feather* in *feather luggage* should be more difficult to understand than, e.g., *feather trolley*, where the non pivot is replaced. The prediction is based upon the observation that by replacing B in a X + B schema we are, in fact, destroying the integrity of the lexicalized schema licensing the interpretation of the compound as a whole.

To sum up, we are inclined to assume the existence of both head-driven and modifier-driven LIS, whose activation depends on whether the compound is of a relational or attributive type. Note that, although the LIS model counteracts the role of strict compositionality in compound interpretation, it is a far cry from claiming that compounds are stored as unanalyzed lexical wholes. In fact, the emergence of an X + B (or B + X) schema requires that the variable X be instantiated by a variety of different nouns, thus yielding different compounds. This means that each X-filling constituent makes an *independent* contribution to the meaning of the whole construction, i.e., that it behaves *compositionally*. Nonetheless, the fact that the whole interpretation is crucially conditioned to the constant lexical constituent B has the implication that the interpretation is only *weakly compositional*. Such a mixed status of compound interpretation schemata, halfway between truly compositional processes and lexicalized patterns, points to a dynamic conception of the interaction between grammar and lexicon, an interaction that is hard to reconcile with the classical grammar-lexicon decoupling.

Finally, the experimental observation (made by Gagné and Shoben first, and recently confirmed by Storms and Wisniewski 2005 for Indonesian) that a LIS is more highly activated by the modifier noun than by the head may be due to processing factors such as the higher degree of lexical entrenchment or specialization of compound modifiers with respect to heads, recently observed for Italian by Baroni *et al.* (in press), and confirmed by the data we are about to report. In this respect, observe that the semantic features contributed by the modifier to an attributive compound are often idiosyncratically salient properties of the corresponding concept, such as the "prickly" contribution of *cactus* to *cactus fish*. As we will see below, this makes modifiers of attributive compounds harder to substitute with semantic neighbours than heads of relational compounds.

Were our hypothesis confirmed, we would obtain two theoretically desirable results: on the one hand, we account for the distinction between relational and property compounds by grounding it on different interpretation processes. On the other hand, we acknowledge the underlying one-route nature of compound interpretation, as, in both cases, we avail ourselves of a single processing mechanism: the lexicalized interpretation schema.

To test the hypothesis that prior availability of lexically-related compound schemata profoundly influences our interpretation of novel combinations, we undertook a careful analysis of the real distributions of compounds. It is moot that the relation frequencies used by Gagné and Shoben reflect such distributions. Recall that they examined all arbitrary combinations of 91 heads and 91 modifiers from a list derived from Levi's (1978) book, and determined which were sensible. Counts were eventually based on the outstanding 3,239 sensible combinations, that were classified by the authors in 15 categories. We contend that the relation frequencies derived from such an analysis do not mirror emergent distributional patterns in the speaker's mental lexicon (a similar point is also made by Storms & Wisniewski 2005). To have a better sense of these patterns, we should look into combinations that speakers themselves produce and read/hear in everyday discourse settings. This concern motivates our decision to look into a very large Web-based corpus to elicit reliable distributional patterns.

While here we presented a general theoretical approach to compounding as concept combination, it should be clear that the experiments we are about to present only pertain to some aspects of this model.

5. CATEGORY INDUCTION: THE EMERGENCE OF THE RELATIONAL VS. ATTRIBUTIVE DISTINCTION

Our first empirical investigation explores the issue of whether distributional cues in the corpus are robust enough to let the classes of relational and attributive compounds emerge without supervision (we ignore hybrid or coordinative compounds, that have been shown to be rare in our data-base by Baroni *et al.*, in press). We focus on the most frequent compounds, given that these are the ones that speakers will hear more often (and probably earlier), and thus they should serve as a basis for generalization. In the experiment we report in the next section we will then test how our model can generalize from frequent compounds to newly generated (and thus rarer) compounds on the basis of the LIS extracted from distributional patterns in frequent data.

5.1 Data Preparation

Data were extracted from *itWaC* (Baroni & Ueyama 2006), a corpus of crawled web-pages containing approximately 2 billion tokens, with POS and lemma information. Using regular expressions defined on POS sequences, we identified N+N strings that occurred in plausible "connected speech" contexts (as opposed to lists, titles, etc.). We analyzed the 398 candidate N+N pairs with frequency of at least 1,000 occurrences. Among these, we found 24 attributive compounds and 95 relational compounds, that constitute the basis for the analysis reported here.

We then identified a set of distributional scores that might cue compound type differences (see Baroni *et al.* in press, for a more detailed analysis of a richer set of cues).

Two cues pertain to the level of specialization of H and M: **H N** measures the proportion of times H occurs immediately *before* another noun in the corpus, divided by the total number of occurrences of H; **N M** measures the proportion of times M occurs immediately *after* another noun, divided by the total number of occurrences of M. Although of course not all N+N sequences in the corpus are compounds, and thus these cues are potentially noisy, they should give us an indication of the propensity of H and M to function as heads vs. modifiers. We expect relational compounds to have "prototypical" heads, and attributive compounds to have "prototypical" modifiers (see discussion in section 4 above).

One cue, **H M COS**, pertains to the semantic similarity between H and M, the idea being that relational H and M will tend to be semantically more related (since M must fill a slot in the semantic representation of H), whereas often there is no relation between attributive H and M (since M tends to be used metaphorically, or it has a very generic, impoverished meaning). Semantic similarity is measured automatically by the cosine distance between H and M in the word space described in section 6.1 below.

The next cue, **H del M**, quantifies the tendency of H and M to appear in a context where the presence of a relation between the two elements is syntactically realized by the preposition *del* ('of the').[2] The idea is that this will characterize relational compounds only. In the attributives, the modifier functions essentially as an adjective, and thus no relation between entities is predicated (cf. *zebra pot* $\neq$ *pot of the zebra*). The propensity of H and M to be connected by *del* is measured by a log-likelihood ratio score based on the comparison between the observed frequency of co-occurrence of H and M in the *N del N* frame and the expected frequency in case of independence between H and M (Evert 2004).

The last two cues, **H del N** and **N del M**, measure to what extent H and M are relational nouns, in terms of proportion of occurrences in the *H del N* vs. *N del H* frames. At least for heads, the idea is that relational compounds should, for sure, be headed by highly relational nouns, and these will often occur with the semantically empty preposition *del*. For modifiers, the interpretation is less clear, but it is possible that there are nouns that tend to be typical "slot fillers" in a number of situations.

5.2 Clustering

In Table 1, we report the median scores of the various cues for the two compound classes, and whether the difference is significant. The two classes differ significantly (and in the expected direction) in all scores except **H M COS** (the latter might play a more significant role when coordinatives are also taken into account). Are the differences we see sufficient to let the distinction between classes emerge in an unsupervised setting?

Visual inspection of plots showed that no single cue came even close to separating the two classes by itself, so the cues should be combined. Moreover, since most cues are significantly correlated to each other, and we are interested in dimensionality reduction for ease of data plotting and exploration, we decided to run a Principal Component Analysis on the compound-by-cue matrix, and to look at the location of compounds in Principal Component space.[3]

---

[2] More precisely, *del* is a fusion of the preposition *di* and the article *il*, and it has several inflectional variants, all included in our counts (*del*, *della*, *dei*, etc.). The fused preposition works better than bare *di* as a relational marker because the latter has also attributive usages in which it means something like 'having the function of', e.g., *problema di base* 'base problem'.

[3] In non-mathematical terms, Principal Component Analysis attempts to identify "hidden" variables that account for more variance than any single observed variable, and that might be amenable to a more abstract interpretation. Observed variables will be more or less correlated to these hidden components (the "loadings" illustrated in the second panel of Figure 1), and might be seen as superficial cues to the corresponding more abstract factors (e.g., **H del M** is one of the observed variables that is strongly associated with the first principal component, that can be interpreted in abstract terms as "degree of syntactic realization"). The first few

cue	*REL median*	*ATT median*	*sig*
H N	6%	1%	++
N M	1%	5%	++
H M COS	0.06	0.07	-
H del M	470	-0.04	++
H del	8%	3%	+
del M	11%	5%	++

TABLE 1: Median relational and attribute values for the target distributional cues; significance of differences measured by Mann-Whitney test with significance levels $++ : p < .001; + : p < .01; - : p >= 0.05$.

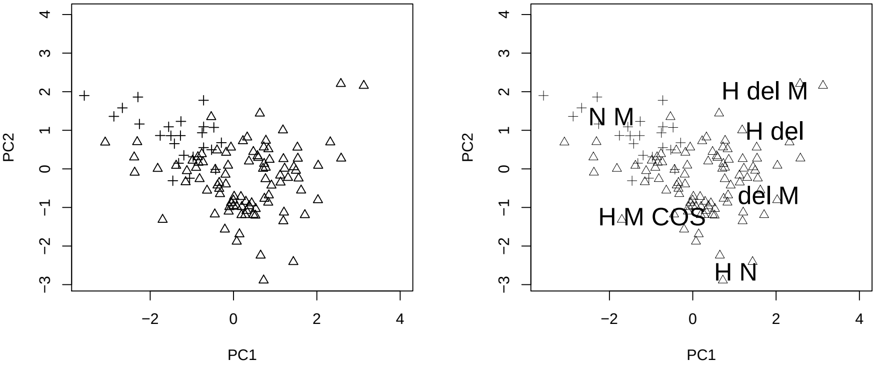


FIGURE 1: Relational (triangles) and attributive compounds (circles) on the PC1 and PC2 dimensions; right plot adds (re-scaled) loadings of the original variables on PC1 and PC2.

The left panel of Figure 1 shows compounds in the space spanned by the first two principal components (accounting for about 50% of the total variance). The distinction between relational and attributive compounds emerge rather clearly, with the attributives characterized by low PC1 and high PC2 values.

The right panel of Figure 1 shows (rescaled) loadings of the original independent variables on PC1 and PC2, as a cue to the interpretation of these components. The first component is strongly positively associated with **H del M**, **H del** and **del M**, thus expressing the general tendency of relational heads and modifiers to occur in explicit relational constructions. The component has an equally strong negative association with **N M**, which expresses the degree of specialization of M as a compound modifier. The variable most strongly

---

components typically account for a larger proportion of the variance, and they are the easiest ones to interpret.

(negatively) associated with PC2 is **H N**, i.e., the degree of specialization of H as a compound head. "Good" relational compounds have both a highly specialized head and a tendency for both head and modifier to occur in explicit syntactic relational constructions, but they can also get a high score along one of these two dimensions only.

We also experimented with the *k-means* algorithm looking for 2 clusters in the space defined by the first 4 principal components (accounting for 85% of the overall variance).[4] The algorithm found a "pure" cluster containing 75 relationals only and a mixed cluster with all 24 attributives and 20 relationals. This is better than chance, but obviously far from a perfect result. In future research, we intend to explore more powerful clustering approaches (the *k-means* solution finds a linear separator and fails to follow the non-linear profile of the true attributive/relational separation), and study the nature of the less separable points in more detail.

As a first step in this direction, we looked for systematic trends among the relational compounds that have a value below -1 on PC1 and above -1 on PC2 (the 8 triangles on the left, just below the attributive clump in the left panel of Figure 1). These 8 outliers are: *associazione ambientalisti* ('environmentalists' association'), *stazione radio* ('radio station'), *conferenza stampa* ('press conference'), *comunicato stampa* ('press announcement'), *ufficio stampa* ('press office'), *pausa pranzo* ('lunch break'), *casa vacanze* ('vacation house') and *monte ore* ('amount of working hours').

Out of these 8 compounds, the problem with *associazione ambientalisti* is rather plain: the modifier is homographic with an adjective, and tagging problems in the corpus lead the statistics for the noun reading to be infected by the adjectival distribution.

Stazione radio, *conferenza stampa*, *comunicato stampa* and *ufficio stampa* are also not too worrying, since both *radio* and *stampa* are very common modifiers with special distributional properties: it is not even clear that, in this context, they should be treated as true nouns. In any case, these are certainly not prototypical relational compounds.

These leaves us with 3 unexplained outliers: *pausa pranzo*, *casa vacanze* and *monte ore*. Further research should analyze these cases in more detail. In particular, we hypothesize that adding further syntactic cues (based on more prepositional connectors besides *del*) might allow us to better distinguish these cases from the attributives.

Despite these problems, our current results show how simple distributional cues could lead speakers to postulate distinct classes of attributive and relational compounds.

---

[4] Briefly, the *k-means* algorithm starts by choosing *k* random cluster centers; it then iteratively assigns items to the cluster they are nearest to and re-computes cluster centers until it reaches convergence at a solution that maximizes intra-cluster cohesion or a similar criterion.

6. ANALOGICAL EXTENSION: LICENSING NEW COMPOUNDS

The experiment in the previous section suggests that a speaker could discover the distinction between frequent relational and attributive compounds on the basis of distributional cues that point to the highly relational characters of relational heads and modifiers and to the tendency of nouns to specialize as (relational) heads or (attributive) modifiers.

In this section, we model novel compound generation as an analogical process that takes the LIS abstracted from high-frequency compounds as sources, creates new compounds by substituting the head, modifier or both, and evaluates the acceptability of the newly created compounds on the basis of constraints that depend on whether the source (and thus the target) is an attributive or relational compound. The idea that generating novel compounds is essentially an analogical process based on (heads and modifiers of) high-frequency compounds is supported by the analysis of Baroni *et al.* (in press), who found a strong tendency for the same heads and modifiers to be repeatedly used within the sample of compounds from all frequency ranges they analyzed.

We propose the following constraints on compound generation. For relational compounds, the head is the pivot element, and it has a relational semantic representation that calls for the modifier as a slot filler. Thus, in a new, analogically formed relational compound, it is more important to preserve the head than the modifier. However, under the assumption that semantic neighbours of relational nouns are also relational (*group*: *team*, *gang*, *club*...), we also predict that, if the head is replaced by a semantically similar word, the new compound should still in general be acceptable. Given the "slot filling" nature of the head-modifier relation, and under the assumption that semantic neighbours of a slot filler will also be appropriate slot fillers (e.g. *city center*, *village center*, *settlement center*), we predict that replacing the modifier with a neighbour will lead to an acceptable compound. Finally, replacing the head and/or the modifier of a relational compound with a random noun should break the slot-filling relation and lead to ill-formedness.

The pivotal component of an attributive compound is the modifier, that has often a rather generic, semantically impoverished meaning, potentially compatible with many heads (you can have a *zebra pot* as well as a *zebra train*, a *zebra table*, etc.). Thus, as long as the modifier of an attributive compound is not changed, replacement of the head with a neighbour or with a random noun should lead to an acceptable compound. On the other hand, given the often idiosyncratic nature of the attributive properties of a noun (either because they depend on salient features of an entity, like in the case of *zebra*, or because the relevant nouns developed a lexically specialized meaning in their modifier function, as with *pilot, base*, etc.), replacing the modifier of an attributive compound with another noun should always lead to an ill-formed compound, even if the replacement is a semantic neighbour of the modifier.

The predictions we make are summarized in Table 2. These will only surface as general trends, since various stages in the data preparation procedure may introduce noise. In particular, not all neighbours generated by our corpus-based semantic model will actually be neighbours of the appropriate kind. Furthermore, the random substitution process might occasionally select nouns that are contextually appropriate by chance. Finally, while in general attributive modifiers should be quite free to combine, there are of course subtler constraints that we are not modeling here (the head of *zebra N* can be virtually any concrete noun, but it is hard to make sense of a *zebra motivation* or a *zebra sadness* in the attributive reading).

head change	*mod change*	*REL*	*ATT*
none	none	+	+
none	neighbour	+	-
none	random	-	-
neighbour	none	+	+
neighbour	neighbour	+	-
random	none	-	+
random	random	-	-

TABLE 2: Predicted acceptability of newly generated compounds by change pattern and compound type

6.1 Data Preparation

We randomly selected 10 relational and 10 attributive compounds from the frequent compound list described in section 5.1 above. In selecting the compounds, we applied the constraints that no head and no modifier could be repeated across compounds, and both head and modifier had to be in the word space model we used to extract neighbours. As a result, 10 is the maximum number of attributive compounds we could extract while respecting these constraints.

In order to generate nominal neighbours for the source heads and modifiers, we relied on a word space model (Sahlgren 2006) constructed from the *la Repubblica* corpus with the Infomap toolkit.[5] The model is based on patterns of co-occurrence of content words with other content words in a text window of 5 tokens. Semantic similarity is measured by the cosine distance between vectors representing the words in a reduced dimensionality space (135 dimensions) derived by SVD from the full co-occurrence matrix.

---

[5] See `http://repubblica.sslmit.unibo.it` for information about the *La Repubblica corpus*, a corpus of newspaper Italian of almost 400 million tokens. On Infomap, see `http://infomap-nlp.sourceforge.net`.

The model is of course quite noisy. A central problem is the intrusion of target neighbours whose similarity to the source heads and modifiers cannot be expressed in terms of an ISA (or taxonomic) relation (for example, *flessibilità* 'flexibility' is one of the nearest semantic neighbours of *lavoro* 'job', however the relation of 'flexibility' to 'job' is of type `property`, and not `synonymy/co-hyponymy`). Polysemy is another source of undesired neighbours: for example, *pallonetto* 'lob kick' is a co-hyponym of *corner* in the realm of football kicks, but not in the sense found in our source compound (*angolo cottura* 'cooking corner').

Despite these problems, we chose a corpus-driven semantic model for two reasons. First, finding semantic neighbours is an extremely difficult task to carry out by hand in a completely objective manner. Second, our ultimate goal is to produce a fully automated corpus-based simulation of how speakers learn and extend compounds; thus, it makes sense to focus on an automated model of lexical semantics as well.

For each source head and modifier, we extracted the 3 nearest nominal neighbours from the model just described. An ANOVA indicates that neither component (head vs. modifier) nor compound type (relational vs. attributive) have a significant effect on cosine similarity between source noun and neighbour; the interaction is also not significant.

For each extracted neighbour, we randomly select a noun in the itWaC corpus with comparable frequency (log frequency difference of 1 or less).[6]

For each source compound, we construct 3 quadruples by pairing a neighbour of the head and its frequency-matched random counterpart with a neighbour of the modifier and its frequency matched random counterpart. Each quadruple is used to generate 6 new compounds by combining neighbour and random heads and modifiers, in turn, with the source components (source H + neighbour M, neighbour H + source M, neighbour H + neighbour M, source H + random M, random H + source M, random H + random H). Thus, from each source we generate 18 new candidate compounds.

In total, we generate $18 \times 20 = 360$ candidate compounds, to which we add the 20 source compounds for a total of 380 stimuli used in the experiment reported below.

6.2 Experimental Set-up

Following a recent and promising trend in psychological and psycholinguistic experimental design (cf., among others, Reips 2002, Reips and Musch 2002, Joinson *et al.* 2007, and the references therein), we developed a web-based

---

[6] Since we pick automatically selected nearest neighbours of heads and modifiers, we cannot control for differences in frequency between and within sources and neighbours. Head and modifier frequencies are, however, among the factors we take into account in the logistic regression model discussed below.

experimental task to test human subjects' reactions to our set of analogically generated compounds.

Our web-experiment consists of a series of server-based scripts written in PHP and Javascript, accessible by means of a standard web-browser.

6.2.1 Participants

Fifty one adult university students voluntarily participated in an acceptability rating test of analogically generated Italian NN compounds. No student participated in more than one experimental session of the present study, and all were native speakers of Italian.

The experiment was carried out in four different sessions taking place at the Universities of Bologna, Padova, Pavia and Trento between March 2007 and May 2007. The participants were not rewarded for taking part in the experiment.

6.2.2 Materials and Procedure

Participants had to log onto a password-secured dedicated website from one of the workstations in the university computer labs.[7] Before the beginning of the experimental session, an introductory presentation of the experiment materials and modality was given by an instructor.

Each participant was presented with 180 test-items, each displayed on a browser window. The web pages were designed to be maximally neutral (white background, test item centered in a highly contrasting, easily readable and large *sans serif* black font).

At the top of the page, in a smaller gray font, the experimental task/question was displayed: *Trovi ben formato questo composto?* ('Do you find this compound acceptable/well-formed?'). Underneath each test item there were three large gray clickable buttons respectively labeled *sì* ('yes'), *forse* ('maybe, don't know') and *no* for the participant's response. When the mouse pointer hovered above the buttons, they changed from the neutral state to an active state signaled by a change in color (respectively, green, blue and red, cf. Figure 2).

The experimental session was divided into two phases. First, all participants were presented with a "training set" of the same 10 items (not considered in the analysis). After the training set was over, the real experimental set was presented (overall 170 test items in random order), including 100 randomly selected stimuli out of the 380 analogically generated NN compounds and 70

---

[7] Although one of the main advantages of web-based experiments is their potentially unlimited replicability under uncontrolled conditions and their being open to a virtually enormous number of unknown participants (who are usually invited to take part in the experiment by means of messages to mailing lists or by web advertising), we decided to distribute our acceptability test under controlled conditions in this first trial. We are planning to open the experiment and advertise it in the near future.

Trovi ben formato questo composto?

sala stampa

Evento 3 di 180

FIGURE 2: Screenshot of a test item being displayed while the mouse pointer hovers on the *si* button (triggering the colour replacement from the neutral state – gray – to active – green –).

foils (comprising grammatical and ungrammatical coordinative and argumental NN compounds – which are not the object of the present study – as well as grammatical and ungrammatical VN compounds).

Along with the response given, for each test item also the reaction time was recorded in two different ways: first, using the server's UNIX timestamps when the page was being served and when the response was received by the script, and second, using a client-based Javascript applet.

As with any web-based experiment, we experienced some problems with respect to the client-side software and to the internet connection. This led to some subjects not being able to complete their full experimental set, or to some reaction times being unreliable. However, all subject responses were kept for the analysis because there were no obvious outliers in response patterns and time of response. Also partial responses were kept, since the experimental design was made to focus on each single stimulus and not on each single subject.

In the post-experimental analysis, subject *sì* responses were interpreted as "good" compounds, and *no* responses as "bad" compounds. *Forse* ("don't know") responses were treated alternatively as "good" or "bad" compounds in two different evaluation settings. Only the former setting is reported in the following sections.

6.3 Results

For ease of analysis, we merged the “good” and “don’t know” responses into an “acceptable” category, since we noticed, qualitatively, a tendency of subjects to be conservative. However, patterns and results similar to the ones reported here also emerge from an analysis in which the “don’t know” responses are merged with the “bad” ones.

Table 3 reports the results in terms of proportion of acceptable compounds by change pattern and compound type.

head change	*mod change*	*REL*	*ATT*
none	none	95%	82%
none	neighbour	70%	39%
none	random	36%	33%
neighbour	none	50%	52%
neighbour	neighbour	32%	28%
random	none	16%	38%
random	random	15%	16%

TABLE 3: Proportion of *acceptable* responses to newly generated compounds by change pattern and compound type

First, we observe that relational and attributive compounds trigger clearly different response patterns. This brings further support to the hypothesis that this distinction should play an important role in the analysis of compounds.

Second, we notice the slightly worrying fact that speakers showed a clear preference for the source relational compounds over the attributive compounds. Given that we slected high frequency compounds (above 1,000 occurrences in itWaC), we would have expected all our source compounds to be considered acceptable. On closer inspection, we find anomalies such as the presence of attributive compound *funzione obiettivo* (lit. ‘target function’), apparently a technical administrative term, that is found completely unacceptable by most subjects. This points to the over-representation of administrative text in itWaC. The logistic regression discussed below takes the difference in baseline relational/attributive acceptability into account, and we leave to further study the issue of whether this asymmetry is systematic (e.g., because attributives are more context-dependent) or due to outliers such as *funzione obiettivo*.

Most of our predictions (refer again to Table 2) are borne out by the data, albeit noisily so. Replacing the modifier with a neighbour has a very strong negative effect on attributives, and a much weaker one on relational compounds (an effect we expected because of the noise inherent in the word space model that provided the neighbours). On the other hand, replacing the modifier with a random noun has a negative effect on the relationals as well, since it disrupts the semantic coherence between head and modifier. Replacing the head with a random noun while preserving the modifier produces the predicted asymmetry in response, causing a much stronger negative effect on relationals than on attributives (although the overall effect is negative, given the unconstrained nature of the inserted random noun). This asymmetry disappears when the head is replaced with a neighbour. In the case of relationals, the neighbour is likely to be another relational noun, with good chances of licensing the source modifier as an appropriate slot filler.

We predicted a much stronger negative impact on attributives than relationals when both head and modifier are replaced by semantic neighbours. Although we do see an asymmetry in the predicted direction, the difference is small. The low acceptability scores for relationals can perhaps be explained by the compounded noise due to replacing both the head and the modifier with an automatically selected neighbour. In the attributive case, however, we would have expected an acceptability level as low as the one we see in the `random random` condition, and this is plainly not the case. Clearly, we are missing some factors here, and we must look for them in future work.

While the data in Table 3 shed some light on the results, they leave several questions unanswered. How does the base asymmetry between relational and attributive acceptability influence the other statistics? Are the interactions described above statistically significant? Could the patterns be explained by uncontrolled subject, item or frequency effects?

In order to deal with these issues, we analyzed the data with a logistic regression, a statistical modeling approach appropriate for binary responses (such as acceptable/not acceptable) that predicts the probability of a positive response. In particular, we are interested in modeling the probability of having an “acceptable” response as a function of compound type (`relational` vs. `attributive`), change pattern (the seven levels of Table 3) and their interaction.

However, we must be aware that there are some other factors that might affect our probabilistic model: in particular, the specific compounds we used, their heads and modifiers, the frequency of such heads and modifiers,[8] the subjects who participated in the experiment.

We introduce (log) head and modifier frequency as numerical factors in the logistic regression model. Following recent trends in statistical modeling of subject and item factors (see, e.g., Baayen in press, ch. 7, and references there), we treat subjects, compounds, heads and modifiers as *random effects* (all other factors above, by contrast, are called *fixed effects*).[9]

---

[8] We do not consider the frequency of the whole compound because the large majority of generated compounds does not occur in the corpus at all.

[9] Random effects can be seen as adjustments to the intercept that are assigned to each subject (or compound, head, modifier) to provide a better fit by taking inter-subject(/compound/head/modifier) differences into account. Overfitting is avoided by using as adjustment term for each subject (/compound/head/modifier) a weighted sum of an estimate

A systematic exploration of increasingly more complex regression models led us to accept the most complex model we considered, since it provides a significant improvement in goodness of fit when compared to simpler models. Thus, our model includes head and modifier frequency, compound type (`relational` vs. `attributive`), change pattern (the seven levels of Table 3) and the type/pattern interaction as fixed effects; and subject, compound, head and modifier as random effects.

When using `attributive` as the base type level and `neighbour neighbour` as the base change pattern level (because, *a priori*, it should constitute a middle-ground in terms of acceptability between the two extremes of `none none` and `random random`), the best fitted model includes the *significant* fixed effects reported in table Table 4. This model has excellent goodness-of-fit: Somers' D_{xy}, a rank correlation computed as described in Baayen (in press: 305), has a value of 81% (although, like in the case discussed by Baayen on p. 283, most of the variance is accounted for by the random and main fixed effects).

factor	*coefficient (std. err)*	*p*
`log head freq`	0.42 (0.06)	++
`log mod freq`	0.14 (0.07)	+
`neighbour none`	1.3 (0.42)	++
`none none`	2.74 (0.6)	++
`REL` *by* `random none`	-1.95 (0.63)	++
`REL` *by* `none neighbour`	1.14 (0.57)	+

TABLE 4: Fixed effects (and interactions) with coefficient significantly different from 0 in logistic model of acceptability responses; significance measured by z-test with significance levels $++ : p. < 01$; $+ : p < .05$.

Not surprisingly, head and modifier frequencies have a positive effect on acceptability, and the estimated coefficient for the source compounds (condition: `none none`) is also significantly higher than 0, i.e., source compounds are more acceptable than the ones we generate. The small but significantly positive effect in the `neighbour none` condition is in accordance with our model, since relational compounds will often be acceptable in this context because a relational head can be replaced by a semantic neighbour, and attributives will be acceptable because the modifier is not changed.

Looking at interactions, we observe that replacing the head with a random noun while preserving the source modifier (`random none`) has a significantly worse effect for relationals than for attributives. *Vice versa*, replacing the modifier with a semantic neighbour while keeping the head unchanged has a sig-

---

based on the data from the subject only and an estimate based on all subjects, that provides a form of smoothing known in the statistical literature as *shrinkage*. For more on these models, see, e.g., Gelman and Hill (2007).

nificantly better (less worse) effect for relationals than for attributives. Peering back at Table 2, we see that the only differential effect that we predicted and that is missing is the one for the `neighbour neighbour` condition (since we are using `neighbour neighbour` as our base change condition, this effect would show up in terms of significance of the `relational` type condition, rather than as an interaction). Overall, the logistic regression model confirms the results of our qualitative analysis of Table 3.

To summarize, while the results are noisy, due to the imperfect nature of both neighbour and random replacements, they confirm most of the predictions about analogically generated compounds that are made by our model. Specifically, they show that the relational and attributive classes have different patterns of acceptability of new compounds. Moreover, acceptability of new relationals will depend on semantic properties of the head (that its semantic neighbours might also possess) and on the semantic relation between head and modifier. Acceptability of new attributives will depend instead on the preservation of the specific modifiers of existing attributives, and less on general semantic constraints.

7. CONCLUSION

Corpus analysis of a representative sample of Italian NN compounds seems to lend support to an interpretive model whereby distributional patterns attested in high-frequency NN compounds allow Italian speakers to discriminate between a class of head-driven relational compounds and a class of modifier-driven attributive compounds. The former class is characterized both by purely lexical constraints (the tendency of relational head nouns to occur as compound heads) and by more general cues to the relational nature of the compound (such as the presence of a head in relational syntactic constructions). The second class, by contrast, is mostly characterized by a tendency of the corresponding modifier nouns to specialize in being used as compound modifiers. All in all, the picture seems to accord well with the theoretical LIS model we sketched in Section 4. Experimental evidence of acceptability judgments by human subjects largely confirms the prediction, made by our model, that analogical processes of new compound formation are particularly sensitive to lexically specific features of relational and attributive heads and modifiers.

Future work should be aimed at tackling a range of empirical issues. In particular, we would like to understand to which extent more and more refined cues can improve the induction of the relational and attributive classes (and add coordinative/hybrid compounds to the task). Further improvements in analogical modeling should come from a more constrained semantic space model, that would allow us to distinguish between different types of semantic neighbours.

More in general, the approach to compounding we sketched here naturally suggests a number of different corpus-based and behavioural experiments we would like to undertake. These include more detailed computational modeling of how properties are transferred to compounds in property ascription, and how LIS structures are generated and extended. From a psycholinguistic perspective, it will be interesting to verify whether structural differences in compounding are reflected by time latencies in lexical decision and priming experiments.

Although much remains to be done, a tighter and tighter integration of insights, methods and explorative tools coming from as different scientific fields as cognitive science, theoretical Linguistics, corpus-based Natural Language Processing and Psycholinguistics, strikes us today as an extremely promising way ahead.

REFERENCES

Baayen, H. (in press). *Analyzing Linguistic Data: A practical introduction to statistics*. Cambridge: Cambridge University Press.

Baroni, M., E. Guevara and V. Pirrelli. (in press). *Sulla tipologia dei composti N+N in italiano*. In *Atti del 40esimo Congresso della SLI*. Roma: Bulzoni.

Baroni, M. and M. Ueyama. (2006). *Building general- and special-purpose corpora by Web crawling*. In *Proceedings of the NIJL International Symposium on Linguistic Corpora*, 31-40.

Bauer, L. (2001). Compounding. In Haspelmath, M. (ed.) *Language Typology and Language Universals*. Berlin: Mouton de Gruyter.

Bisetto, A. and Scalise, S. (2005). *The classification of compounds*. «Lingue e Linguaggio», IV(2), 319–332.

Bloomfield, L. (1933). «Language». New York: Holt.

Cohen, B. and Murphy, G. L. (1984). *Models of concepts*. «Cognitive Science», 8(1), 27–58.

Costello, F. J., and Keane, M. T. (1997). *Polysemy in conceptual combination: Testing the constraint theory of combination*. In *Nineteenth Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Costello, F. J. and Keane, M. T. (2001). *Testing two theories of conceptual combination: Alignment versus diagnosticity in the comprehension and production of combined concepts*. «Journal of Experimental Psychology: Learning, Memory and Cognition», 27(1), 255–271.

de Jong, N. H., Feldman, L. B., Schreuder, R., Pastizzo, M., and Baayen, R. H. (2002). *The processing and representation of Dutch and English compounds: Peripheral morphological and central orthographic effects*. «Brain and Language», 81(1-3), 555–567.

Downing, P. (1977). *On the creation and use of english compound nouns*. «Language», 53(4), 810–842.

Estes, Z. and Glucksberg, S. (2000). *Interactive property attribution in concept combination*. «Memory & Cognition», 28 (1), 28–34.

S. Evert. 2004. *The statistics of word cooccurrences: Word pairs and collocations*. Ph.D. thesis, University of Stuttgart.

Gagné, C. L. and Shoben, E. J. (1997). *Influence of thematic relations on the comprehension of modifier-noun compounds*. «Journal of Experimental Psychology: Learning, Memory and Cognition», 23(1), 71–87.

Gagné, C. L. (2002). *Lexical and Relational Influences on the Processing of Novel Compounds*. «Brain and Language», 81, 723–735.

Gagné, C. L. and Spalding, T. L. (2006). *Conceptual combinations: Implications for the Mental Lexicon*. In Libben, G. and Jarema, G. (eds.), *The Representations and Processing of Compound Words 2006*. Oxford: Oxford University Press.

Gelman, A. and Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Gentner, D. (1983). *Structure-mapping: A theoretical framework for analogy*. «Cognitive Science», 7(2), 155–170.

Glucksberg, S. and Keysar, B. (1990). *Understanding Metaphorical Comparisons: Beyond Similarity*. «Psychological Review», 97 (I), 3–18.

Glucksberg, S., McGlone, M. S., and Manfredi, D. (1997). *Property attribution in metaphor comprehension*. «Journal of Memory and Language», 36(1), 50–67.

Goldberg, A. (2006). *Constructions at Work: the nature of generalization in language*. Oxford: Oxford University Press.

Haspelmath, M. (2002). *Understanding Morphology*. London: Arnold.

Jackendoff, R. (1997). *Twistin' the night away*. «Language», 73(3), 534–559.

Joinson, A. N., McKenna, K., Postmes, T., and Reips, U.-D. (eds.)(2007). *The Oxford Handbook of Internet Psychology*. Oxford: Oxford University Press.

Levi, J. N. (1978). *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.

Maguire, P., Maguire, R. and Cater A. H. (2007). *In Search of the Frog's Tail: Investigating the Time Course of Conceptual Knowledge Activation*. In *Proceedings of the second European Cognitive Science Conference*, Delphi, Greece.

Murphy, G. L. (1988). *Comprehending complex concepts*. «Cognitive Science», 12(4), 529–562.

Olsen, S. (2001). *Copulative compounds: a closer look at the distinction between morphology and syntax*. In Booij, G. and van Marle, J. (eds.), *Yearbook of Morphology 2000*. Dordrecht: Kluwer Academic Publishers.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Reips, U.-D. (2002). *Standards for Internet-based experimenting*. «Experimental Psychology», 49, 243–256.

Reips, U-D. and Musch, J. (eds.)(2002). *Special Issue: Internet-Based Psychological Experimenting*. «Experimental Psychology», Vol. 49, No. 4.

M. Sahlgren. 2006 *The Word-Space Model: Using distributional analysis to represent syntagmatic and paradigmatic relations between words in high-dimensional vector spaces*. Ph.D. thesis, Stockholm University.

Shoben, E. (1991). *Predicating and nonpredicating combinations*. In Schwanenflugal, P. J. (ed.), *The Psychology of Word Meanings 1991*. Hillsdale, NJ: Erlbaum.

Spencer, A. (1991). *Morphological Theory*. Oxford: Blackwell.

Springer, K. and Murphy, G. (1992). *Feature availability in conceptual combination*. «Psychological Science», 3, 111–117.

Storms, G. and Wisniewski, E. J. (2005). *Does the order of head noun and modifier explain response times in conceptual combination?* «Memory and Cognition», 33(5), 852–861.

Wisniewski, E. J. (1996). *Construal and similarity in conceptual combination*. «Journal of Memory and Language», 35(3), 434–453.

Wisniewski, E. J. (1997). *When concepts combine*. «Psychonomic Bulletin & Review», 4, 167–183.

Wisniewski, E. J. (2000). *Similarity, alignment, and conceptual combination: Comment on Estes and Glucksberg* «Memory & Cognition», 28 (1), 35–38.

Wisniewski, E. J., and Clancy, E. J. (2004). *You don't need a weatherman to know which way the wind blows: The role of discourse context in conceptual combination*. Unpublished manuscript.

SUMMARY: Dopo un inquadramento teorico del problema della composizione nelle scienze cognitive e in linguistica, presentiamo una serie di esperimenti sui composti nominali in italiano che mettono alla prova l'ipotesi che ci sia una distinzione fondamentale tra composti relazionali (legittimati da proprietà della testa) e composti attributivi (legittimati dal modificatore). Un'analisi computazionale basata su un corpus conferma che tale distinzione può in linea di principio venire indotta da dati di tipo distribuzionale. Inoltre, dati sperimentali mostrano che il modello è in grado di predire almeno in parte l'accettabilità di nuovi composti formati cambiando testa e modificatore di composti esistenti.