

# **Towards a pragmatic model of cognitive onomasiology**

Stefan Grondelaers and Dirk Geeraerts

## **Abstract**

We argue for an onomasiological reorientation of cognitive lexical semantics: the existing focus on semasiological phenomena like prototypicality should be supplemented with a focus on lexical selection and the factors governing the same. We show what kind of methodology may be used to effectuate such a shift, and we present a conceptual framework for distinguishing between the relevant phenomena. We further argue that the onomasiological perspective inevitably encompasses pragmatic and sociolinguistic variation: the proper development of cognitive onomasiology implies the elaboration of a full-fledged sociolexicological theory and methodology.

*Keywords:* lexical selection, lexicology, onomasiology, pragmatics, salience, semasiology, sociolexicology, sociolinguistics, variation.

## **1. Introduction**

Cognitive Linguistics (as it is associated with the ideas of Ron Langacker, George Lakoff, Len Talmy, and others) is to a large extent characterized by a research strategy that extrapolates findings from lexical semantics to grammar at large: grammatical constructions are assumed to exhibit the same semantic features (such as prototypical structuring) as lexical items. This shift from the lexicon to the grammar of constructions does not imply, however, that the cognitive linguistic theory of the lexicon has in any way reached a final stage. In this paper, we would like to present one specific

way of taking cognitively inspired lexicology one step further, viz., by extending it into the field of onomasiological research.

Our aim is twofold: on the empirical level, we will present the distinction between a *static, structural*, and a *dynamic, pragmatic* interpretation of onomasiology, and we will define the basic factors at work in a dynamic conception of onomasiology. It will be shown that the choice of a term for a referent in a given context is determined simultaneously by (i) the semasiological salience of that term, (ii) its onomasiological salience, and (iii) contextual factors of a sociolinguistic nature. On the methodological level – which is predominant in this paper – we will demonstrate that the actual implementation of a dynamic model of onomasiology (or of lexical choice) requires a more integrated type of lexical studies. A major impediment to the latter is the fact that, due to a lack of corpus-based research, (i) the onomasiological perspective in Cognitive Linguistics is underdeveloped, and (ii) contextual factors have so far received little theoretical attention in cognitively inspired work. Research into variational phenomena of a contextual kind is mostly restricted to sociolinguistics, traditionally a field with its own methodology and quantitative techniques. We will, however, demonstrate with a selected number of case studies that a true quantitative sociolinguistics of the lexicon, or sociolexicology, is indispensable in order to obtain a reasonably comprehensive model of lexical selection. We believe, in addition, that this goal can never be achieved without thorough empirical substantiation. It is our intention to demonstrate that the three components of the model (semasiological salience, onomasiological salience, contextual factors) – and not just the first one – are amenable to the same type of quantitative analysis, viz., observation of non-elicited language use. A unified methodology, we will also show, is a pivotal step in bridging the traditional gap between lexicology and sociolinguistics.

This paper is organized as follows. The second section presents the basic principles of pragmatic onomasiology. Sections 3 and 4 focus on empirical data substantiating the import of semasiological salience and onomasiological salience, respectively. Section 5 introduces the contextual perspective and sociolexicology, and sections 6 and 7 discuss sample studies which illustrate contextual influence on conceptual onomasiological

variation and formal onomasiological variation. Section 8, finally, is devoted to conclusions and theoretical consequences.

## **2. Towards the basic principles of cognitive onomasiology**

The distinction between onomasiology and semasiology is a traditional one in Continental structural semantics, but it has hardly found its way into the canonical English terminology of linguistics; it does not appear, for instance, in most standard textbooks of linguistics. The following quotation from the Swiss Romanist Kurt Baldinger provides a good definition of this distinction: “Semasiology ... considers the isolated word and the way its meanings are manifested, while onomasiology looks at the designations of a particular concept, that is, at a multiplicity of expressions which form a whole” (1980: 278). The distinction between semasiology and onomasiology, then, equals the distinction between meaning and naming: semasiology takes its starting-point in the word as a form, and charts the meanings that the word can occur with; onomasiology takes its starting-point in a concept, and investigates by which different expressions the concept can be designated, or named.

While Baldinger’s definition of onomasiology suggests that “the designations of a particular concept” and “a multiplicity of expressions which form a whole” are equivalent, we submit that they point to two different onomasiological subfields. On the one hand, studying “a multiplicity of expressions which form a whole” is the domain of the traditional, structuralist conception of onomasiology, i.e., the study of semantically related expressions (as in lexical field theory, or the study of the lexicon as a relational network of words interconnected by links of hyponymy, antonymy, synonymy, etc.). On the other hand, studying “the designations of a particular concept” paves the way for a contextualized, pragmatic conception of onomasiology, which focuses on the actual choices made for a particular name as a designation of a particular referent.

This distinction between a structural and a pragmatic interpretation of onomasiology can be defined even more precisely by equating it with a distinction between an investigation of *structure* and an investigation of *use*. On the structural interpretation, onomasiology deals with sets of re-

lated expressions, and basically asks the question: What are the relations among the alternative expressions? On the pragmatic interpretation, onomasiology examines the actual choices made from among a set of related expressions, and basically asks the question: What factors determine the choice for one or the other alternative? It is precisely with the latter issue that this paper will be concerned.

A systematic treatment of pragmatic onomasiology, however, presupposes one more distinction, namely that between what may roughly be described as the *qualitative* versus the *quantitative* aspects of linguistic semantic structure. Let us first consider how this distinction operates in semasiology. The qualitative study of semasiological structure looks at the following questions: What meanings does a word have, and how are they semantically related? The outcome is an investigation into polysemy, and the relationships of metonymy, metaphor, etc. that hold between the various readings of an item. A quantitative study of semasiological structure, on the other hand, involves the question whether all the readings of an item carry the same structural weight. The outcome, obviously, is an investigation into prototypicality effects of various kinds. Obviously, the qualitative perspective is much more traditional in semasiological lexicology than the quantitative one, which has been taken up systematically only recently, with the development of prototype theory.

The distinction between qualitative and quantitative aspects of semasiological structure has an onomasiological counterpart. The qualitative question then takes the following form: What kinds of (semantic) relations hold between the lexical items in a lexicon (or a subset of the lexicon)? The outcome, clearly, is an investigation into various kinds of lexical structuring: field relationships, taxonomies, and lexical relations like antonymy, hyponymy, etc. Research into the quantitative aspects of onomasiology revolves around the following questions: Are some lexical categories cognitively more salient than others, that is, are there any differences in the probability that one category rather than another will be chosen for designating things out there in the world? Are certain lexical categories more obvious names than others? Again, this type of quantitative research is fairly new, and does as yet only exist in an embryonic form. The best-known model to date is Berlin and Kay's basic level model (1969), which

claims that the generic level in a taxonomy constitutes the default “basic level” for categorization. (This hypothesis will be nuanced below.)

The overall picture of the field of lexicology that emerges from the foregoing is presented in Table 1. This table serves as a background for the next important step in our discussion, where we will make clear that the quantitative aspects of semasiological structure and onomasiological structure are among the crucial factors determining lexical selection.<sup>1</sup>

Table 1. The field of lexicology.

	Semasiology	Structural onomasiology	Pragmatic onomasiology
<i>Qualitative aspects:</i> elements & relations	senses + semantic links (metaphor, metonymy, etc.)	lexical items + lexical relations (fields, taxonomy, etc.)	factors determining lexical selection
<i>Quantitative aspects:</i> differential weights	salience phenomena, core vs. periphery	cognitive salience of categories	

The factors determining the selection of a name for a referent, or more generally, the principles of a pragmatic onomasiology, are in fact the following:

- (a) the degree of prototypicality of the referent with regard to the semasiological structure of the category (*semasiological salience*);
- (b) the onomasiological entrenchment of the category labeled by the name (*onomasiological salience*);
- (c) contextual features of a classical sociolinguistic (geographical, stylistic, etc.) nature.

1. The overall sketch of the field of lexicology that is presented in Table 1 is elaborated in Geeraerts (2002a). The historical evolution of lexicology (which may be seen as a gradual expansion of the field as charted in Table 1) is discussed in Geeraerts (2002b).

In what follows, empirical results pertaining to (a)–(c) will be presented. Just like the model of lexicology presented here, these results are treated in more detail in Geeraerts, Grondelaers and Bakema (1994), Grondelaers and Geeraerts (1998), and Geeraerts, Grondelaers and Speelman (1999).

### **3. The influence of semasiological salience on naming**

Semasiological salience, or prototypicality, in a lexical category can be determined in a rather straightforward manner. Indeed, prototypical status is usually assigned to that category member whose configuration of defining features is not only attested the most (i.e., has the highest frequency) in a sufficiently large database, but also whose feature configuration combines the highest number of dominant features. As such, prototypicality shows a correlation between the intensional and the extensional aspects of semasiological structure. This is, to be sure, a classical conception of prototypicality, as it was first introduced by Eleanor Rosch. Its impact may be established by examples such as Figure 1.

Figure 1 charts the semasiological range of the Dutch clothing term *colbert* ‘formal jacket’, such as it is used in a number of fashion magazines and women’s magazines. The database on which the analysis is based consists of 9,000 records in which an attested clothing term is coupled with a componential description of its referent. These featural descriptions are set up on the basis of the pictures

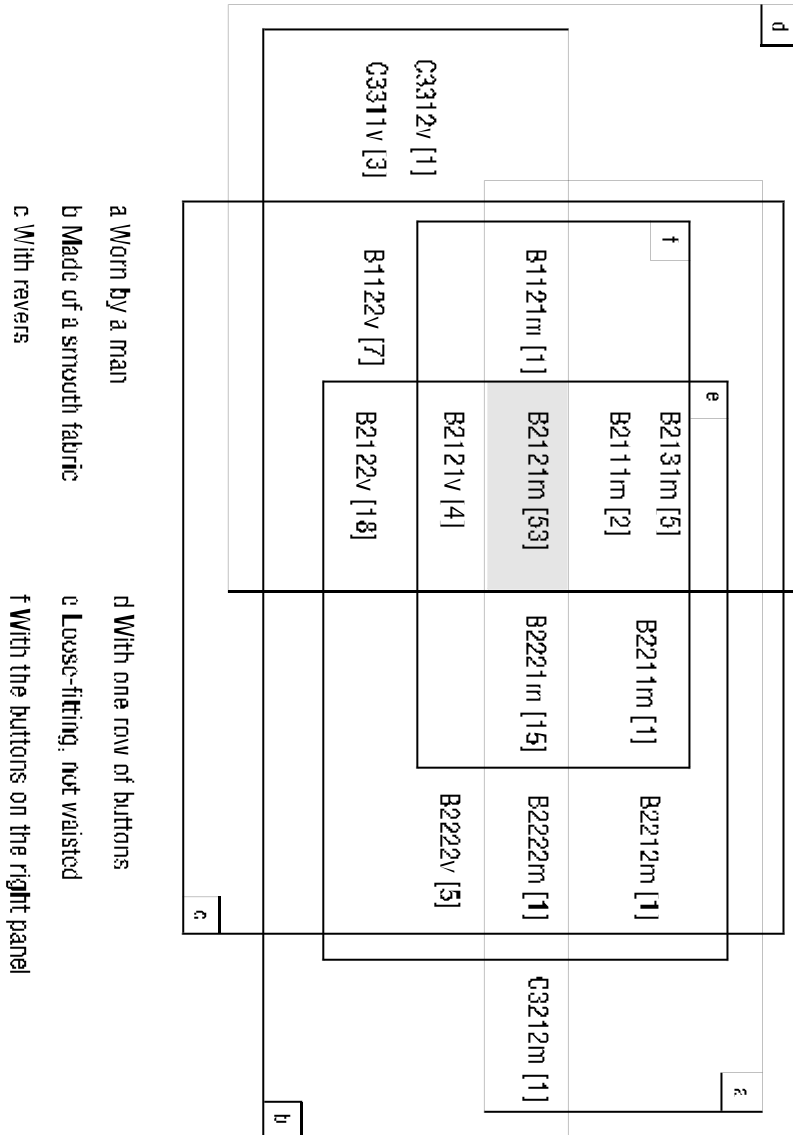


Figure 1. The semasiological range of the Dutch clothing term *colbert*.

and drawings the clothing terms occur with in the magazines, whereby each feature configuration denotes a distinct type of garment. Figure 1 shows various feature configurations (such as B2121m) which are each

labeled *colbert*. The capitals “B” and “C” in the configurations respectively refer to formal jackets with revers and knitted cardigan-like jackets, while “m” and “v” indicate the sex of the wearer. The feature breakup is as follows: the first digit after the capital indicates the jacket’s width, with values “1” (waisted) and “2” (loose-fitting); the second digit specifies whether the jacket has a single-breasted (“1”) or a double-breasted (“2”) button fastening; the third digit identifies the fabric the jacket is made of: thin (“1”), thick but smooth (“2”), or thick and coarse (“3”); the last digit indicates whether the buttons are located on the right (“1”) or left front panel (“2”).

The numbers between square brackets following the feature configurations indicate the absolute frequency with which *colbert* occurs in the corpus as the name for the type of garment represented by that particular feature configuration. Now, the various configurations in the structure of *colbert* may be divided into subsets on the basis of the features that they share. The feature configurations in Figure 1 allow a classification in terms of subsets “a” to “f”; each of these subsets shares a specific value on the features discussed in the previous paragraph. Prototypicality shows up in that the type of referent with the highest frequency (B2121m,  $n = 53$ ) is at the same time the type of referent that combines all the dominant features “a”–“f”, whereas the less frequently attested referents lack one or more of these features. Graphically, this highest frequency referent is situated in the area where all of the subsets “a”–“f” overlap (indicated by shading).

The kind of prototypicality structure of lexical items such as illustrated in Figure 1 will have an influence on the selection of lexical items if the following hypothesis holds:

*A referent is more readily named by a lexical item if it is a salient member of the category denoted by that item.*

That is to say, there is a positive correlation between the prototypical referent of a word (i.e., a referent’s centrality in the semasiological structure of a word), and the probability of that word being selected – rather than alternative words – to name/label the referent. The latter may be expressed by an *onomasiological cue validity* measure: a referent (or a set of referents) in the extension of a lexical category has a high cue validity

with regard to that category if the category is readily chosen as a name for designating the referent. Onomasiological cue validity, then, can be measured by computing the ratio between the number of times that a lexical item is chosen as a name for a particular (set of) referent(s), and the total corpus frequency of that (set of) referent(s).

Table 2. Correlation between prototypicality (in terms of frequency) and onomasiological cue validity for the item *colbert*.

Set of configurations	Cumulative semasiological frequency per set	Cumulative onomasiological cue validity per set
B2121m	53	72.60
B2121v, B2122v, B2131m, B2111m	82	33.60
B2121m, B2121v, B2122v, B2131m, B2111m, B1121m, B1122v, B2212m, B2211m, B2221m, B2222m, B2222v	113	29.89
B2121m, B2121v, B2122v, B2131m, B2111m, B1121m, B1122v, B2212m, B2211m, B2221m, B2222m, B2222v, C3312v, C3311v, C3212m	118	24.48

Table 2 presents four central subsets of the category COLBERT in Figure 1; these subsets were singled out by starting from the maximally overlapping area in Figure 1 (B2121m) (the most prototypical case), and gradually enlarging this core area towards less central cases. The more the semasiological core of the category is enlarged, the fewer features the items in the subsets have in common, and the more their relative frequency diminishes. It is crucial for a correct understanding of Table 2 that the semasiological frequencies in the second column and the cue validities in the third relate to cumulated sets of configurations. The cue validity value 33.6, for instance, relates to the set constituted by the configurations B2121m, B2121v, B2122v, B2131m, and B2111m. The rightmost column in Table

2 shows that our hypothesis concerning the correlation of semasiological salience and onomasiological cue validity is indeed correct for the category COLBERT as charted in Figure 1: the cue validity diminishes (from 72.6 to 24.48) as the subset under consideration is extended towards the periphery of the category (for more examples, see Geeraerts, Grondelaers & Bakema 1994: 156 ff.).

#### **4. The influence of onomasiological salience on naming**

Let us now turn our attention to the onomasiological domain. If a particular category A (named by A) is more firmly anchored in the user's knowledge than category B, i.e., if the members of a category A are more readily identified as members of A than as members of B, then category A can be said to be more *entrenched* (or onomasiologically salient) than category B. In a corpus-based approach, categorial entrenchment can be operationalized by computing the ratio between the corpus frequency of a category name (and its synonyms) and the corpus frequency of the members of the category. This is, to be sure, a cue validity measure just like the onomasiological cue validity measure computed above. At this point, however, it is applied to a lexical category as a whole (within a corpus), rather than to individual members or subsets (within a category): a lexical category is highly entrenched if its extension as a whole has a high cue validity with regard to the name(s) of the category.

Note that Berlin and Kay's basic level hypothesis mentioned earlier (cf. section 2) stipulates that the entrenchment of individual categories is determined by their taxonomical level: all categories belonging to the basic level should roughly have the same (high) entrenchment, if the basic level does indeed embody naming preferences. The empirical results in Geeraerts, Grondelaers and Bakema (1994), however, indicate that this is not the complete story. Major differences of entrenchment among co-hyponyms establish that entrenchment is a matter of individual categories rather than taxonomical levels.

The onomasiological hypothesis involving entrenchment may now be formulated as follows:

*A referent is more readily named by a lexical item if that item denotes a highly entrenched lexical category.*

As such, if in a particular corpus of language use, a particular referent is named by two lexical items, A and B, then there is a positive correlation between the respective entrenchment of the two categories A and B and the choice for either A or B as a name for the referents. Consider, by way of example, the clothing terms *topje* ‘top’ and *blouse* ‘blouse’; the referential ranges of these terms overlap because informal, sleeveless blouses may be called *blouse* as well as *topje*. The absolute frequency with which *topje* is used for referents that lie in the intersection of both categories (i.e., that could be called either *topje* or *blouse*) is 11, whereas 26 of the garments in the intersection of both categories are actually called *blouse*. Now, the entrenchment of *topje* is 29.62; that of *blouse* is 61.52. As predicted by the hypothesis, the actual selection of either *topje* or *blouse* for referents in the overlapping area correlates highly with the entrenchment values of these categories, in that the item with the highest entrenchment is indeed more readily chosen as a name for the referents in the intersection of both items. Mathematically, the ratio of, on the one hand, the ratio of absolute frequencies ( $11/26 = 0.423$ ) and, on the other hand, the ratio of entrenchment values ( $29.62/61.52 = 0.481$ ) is as high as 0.879 (for more examples, see Geeraerts, Grondelaers & Bakema 1994: 169 ff.).

## 5. Introducing the contextual perspective

Whether a lexical category is chosen as a name for a referent is not only determined by the degree of prototypicality of the referent with respect to the semasiological structure of the category or the onomasiological entrenchment of that category: semasiological and onomasiological factors themselves correlate with numerous types of contextual factors, such as the sociological and geographical properties of the participants in the

communicative interaction, or characteristics of the speech situation itself. In integrating contextual values, we have in fact ventured from lexicology into sociolinguistics, a discipline traditionally considered to have its own subject matter and empirical methodology. We will demonstrate by means of two sample studies that this orientation to the field of sociolinguistics is a necessary one as far as lexical selection is concerned, viz., that a true *sociolexicology* is indispensable in order to obtain a workable model of lexical selection.

The fact that lexicology and sociolinguistics are not evident partners is due, we feel, to two factors. First and foremost, sociolinguistics rarely concerns itself with lexical variation. Sociolinguists who examine the relationship between social factors and language preferably select linguistic factors they can measure in a reliable way; as a result, they predominantly focus on easily quantifiable phonological or morphological variables such as “absence or presence of postvocalic /r/” (which features prominently in, e.g., Labov’s classic (1966) investigation into the stratification of English in New York and Trudgill’s (1974) study on the social differentiation of English in Norwich). Since the sociolinguistic study of lexical variation is hindered by the collective interference of several types of – fairly unmanageable – semantic variation, lexical variables are but occasionally included in traditional sociolinguistic research.

An indispensable tool for any linguist – whether sociolinguist or lexicologist – who wishes to undertake the hazardous enterprise of studying contextual differentiation in lexical variation is a theoretical framework in which distinct types of semantic and non-semantic variation are minutely identified, distinguished, and disentangled. In Geeraerts, Grondelaers and Bakema (1994), the preliminaries to such a model are presented, and considerable attention is devoted to a distinction which is central to the appreciation of pragmatic onomasiology, viz., the distinction between “conceptual” (or “semantic”) and “formal” (non-semantic) onomasiological variation. Whereas conceptual onomasiological variation involves the choice of different conceptual categories for a referent, formal onomasiological variation merely involves the use of different names for the same conceptual category. The names *jeans* and *trousers* for denim leisure wear trousers – to give an example – constitute an instance of conceptual name variation, for they represent different categories; *jeans* and *blue*

*jeans*, however, represent no more than different (but synonymous) names for the same category.<sup>2</sup>

If sociolinguists are interested in lexical variation, they invariably – and unsurprisingly – focus on non-semantic variation of the formal onomasiological type (as illustrated in section 7). In Geeraerts, Grondelaers and Bakema (1994), however, we have established that all the types of semantic variation we have hitherto identified are also – and to a large extent – influenced by contextual factors. For a well-known example of geographical variation in the semasiological realm, consider the case of *vest* (1994: 105 ff.). In Netherlandic Dutch, *vest* is used exclusively to denote knitted, cardigan-like garments, whereas in Belgian Dutch, *vest* also refers to formal jackets (in substandard Belgian Dutch, *vest* is in fact the principal term to refer to formal jackets). Contextual differences also determine the relative preference of groups of speakers for a certain taxonomical level, in the sense that an expert is more likely to use specific, technical terms than the layman (1994: 146 ff.; a more elaborate example of contextual influence on conceptual onomasiological variation is presented in the following section). Sociolexicology is the practical consequence of the idea that all semantic and non-semantic components of our model of lexical selection – and not just one – require sociolinguistic attention, if we want the model to work.

---

2. Although Cognitive Linguistics upholds the idea that different words designate different categories, we assume that categories can be isolated from the words which are used to designate them. In this paper, the concept “category” basically receives a referential definition: if words have the same referential range, they designate identical categories in our view. To this procedure, one could object that referential identity need not imply conceptual identity, since denotationally synonymous words may have different stylistic or sociolinguistic values. Rather, however, than a priori demanding that the sociolinguistic and stylistic value of denotationally synonymous words be identical, we exclusively rely on denotational synonymy as far as establishing categories is concerned, because in the model of lexical selection we advocate, stylistic variation – of the type discussed in section 6 – and sociolinguistic variation – as in section 7 – have to be determined empirically on the basis of text corpora. Stylistic and sociolinguistic values of categories, in other words, are dependent variables in our investigation, rather than pre-empirically assumed categorial features (cf. also Geeraerts, Grondelaers & Bakema 1994: 178 ff.).

Second, lexicology and sociolinguistics differ not only with respect to subject matter, but also with respect to empirical methodology. Labov (1972) lists three major kinds of linguistic method: linguistic analysis may be based on introspection, on the elicitation of data with questionnaires and experiments, and on the observation of non-elicited language use. Although the bulk of the sociolinguistic research into lexical variation in Dutch is based on elicitation rather than observation (cf. Geeraerts, Grondelaers & Speelman 1999: 32, 52–53, for an overview and exceptions), it is our intention in the remainder of this paper to demonstrate with two case studies that sociolinguistic variation in the lexical domain can be studied just as efficiently and rigorously as semasiological or onomasiological variation by concentrating on non-elicited written language data. As such, our methodology is in line with the Anglo-American sociolinguistic tradition, where observation of non-elicited language material is customary – notable cases in point are, again, Labov (1966) and Trudgill (1974). This possibility of a common method constitutes a major bond between the two linguistic disciplines sociolinguistics and lexicology, which are traditionally thought to be very different.

Before we embark on the case studies, two points need to be addressed. First, it should be noted that the concept of sociolexicology is not entirely new (although the term “sociolexicology” may be new). By insisting on a quantitative sociolinguistics of the lexicon, we are in fact resuming the thread of lexicological work by William Labov (1973, 1978). The second point relates to the choice of independent and dependent variables in quantitative sociolexicology. With respect to the selection of *independent* variables, it should be pointed out that our corpus-based method and sociolinguistic elicitation differ greatly concerning the basis on which contextual factors are identified. When elicitation is involved, sociolinguistic factors (geography, social stratification, expertise, etc.) can be defined in terms of the characteristics of the speakers in the samples that are compared. In sociolexicology, on the other hand, the comparanda are (necessarily) identified in terms of intended audience, the underlying rationale being that authors of texts intended for a regional or professional subgroup will tailor their word use to that subgroup. The *dependent* variable in sociolexicological corpus-based research differs as to the type of onomasiological variation concerned. When the choice of different conceptual

categories – i.e., conceptual variation – is studied, contextual variation is measured in terms of entrenchment values; when different names for the same category are at stake – i.e., formal variation – the dependent variable of the sociolexicological investigation will be defined as the onomasiological profile of a concept, viz., the set of synonyms designating the concept, differentiated by their relative frequency.

In what follows, we will first present an example of contextual influence on conceptual onomasiological variation. Our account of how euphemistic considerations structure the designation of the disease cancer will also focus on how one can identify and disentangle competing contextual factors. The second case study not only concentrates on formal onomasiological variation, it also introduces the diachronic perspective. The investigation of onomasiological change in Netherlandic Dutch and Belgian Dutch presented in section 7 demonstrates that lexical choice is subject to significant change, even when no semantic change is involved.

## 6. Contextual influence on conceptual onomasiological variation

In this section, we deal with the contextual values of different lexical categories used to designate the disease cancer. The basic question here is how avoidance strategies influence the choice of cancer designations (an elaborate answer to this question is found in Grondelaers & Geeraerts 1998). More particularly, we are interested in finding out how the emotive value of, on the one hand, generic or specific cancer terms such as *cancer* or *breast cancer* and, on the other, vague terms such as *disease* or *illness* influence lexical choice; it is indeed to be expected that in some contexts the vaguer terms will be preferred for euphemistic reasons. To that effect, we have investigated a CD-ROM text corpus – which consists of the 1991–1994 volumes of the Belgian weekly *Knack*, and the 1994 edition of the Dutch quality newspaper *De Volkskrant*. In particular, we looked for quantitative support for the hypothesis that vague terms for cancer are favored in (i) non-scientific contexts, viz., articles which do not

or not primarily report on medical topics, and in (ii) personalized contexts, i.e., contexts in which the effects of cancer on individual victims are portrayed (vs. “generic” contexts, in which cancer is referred to in general). The dependent variable in this study is lexical specificity in the naming of cancer, and is quantified as the ratio between the frequency of hyperonymous designations for cancer (such as *disease* or *illness*), and the overall frequency with which the disease cancer is mentioned (by means of various lexical items) in the sources. Again, this is a sort of entrenchment measure, albeit one that is applied to taxonomical level rather than to referents or to individual lexical categories.

Table 3 contains the hyperonym ratios in the naming of cancer in the Belgian weekly *Knack* and the Dutch newspaper *De Volkskrant*. This table distinguishes, vertically, between personalized contexts and generic contexts, and horizontally between medical and non-medical texts. As predicted, average hyperonym ratios are indeed significantly higher in non-medical contexts ( $0.647 > 0.126$ ) and personalized contexts ( $0.837 > 0.147$ ), which seems to confirm the hypothesis outlined in the previous paragraph.

Table 3. Hyperonym ratios in the naming of cancer in *Knack* 1991–1994 and *De Volkskrant* 1994.

	-medical	+medical	Total
+personal	0.878 (36/41)	0 (0/2)	0.837 (36/43)
-personal	0.296 (8/27)	0.127 (26/204)	0.147 (34/231)
Total	0.647 (44/68)	0.126 (26/206)	0.255 (70/274)

A methodological problem which complicates the identification of avoidance factors – and contextual factors in general – is the fact that the same variational pattern may be caused by more than one factor. In our cancer-example, the predominance of vague designations in non-medical contexts could just as well be due to an upward shift of the taxonomical basic level as a result of the irrelevance of medical detail in articles which are not pri-

marily concerned with scientific progress (cf. Dougherty 1978: 77, but see also Cruse 1977: 163).

Table 4. Absolute and relative frequency of different taxonomical ranks in the naming of specific types of cancer in *Knack* 1991–1994 and *De Volkskrant* 1994.

	-medical	+medical
Unique beginner	22 (61.12%)	23 (13.45%)
Generic	2 (5.5%)	16 (9.36%)
Specific	12 (33.33%)	132 (77.19%)

Table 4, however, constitutes additional evidence in favor of a taboo-related explanation of the asymmetrical distribution of hyperonymy in Table 3. Table 4 charts absolute and relative frequencies of different taxonomical ranks in the designation of specific types of cancer such as breast cancer or lung cancer. On the vertical axis, unique beginners such as *disease* or *disorder* are contrasted with the generic item *cancer* and specific terms like *breast cancer* or *lung cancer*. Now, if the absence of technicality engendered an increased use of hyperonymy in non-medical contexts, this increase would affect both the generic level and the unique beginner level in roughly the same way; in other words, unlike the medical context, the non-medical context would show an increase of hyperonymy, but this increase would be comparable on both the generic and the unique beginner level. Table 4, however, shows that the generic level and the unique beginner level are affected differently, in that increased hyperonymy almost exclusively affects the unique beginner level; there is, in other words, a tendency to “jump over” the generic level. This avoidance of the generic term can best be explained in light of the fact that the generic term *cancer* is still specific enough to cause offense; it would be difficult to explain as the result of decreased technicality alone.

## 7. Contextual influence on formal onomasiological variation

This section contains some results from a systematic study of geographical variation between Netherlandic and Belgian Dutch which was undertaken at the University of Leuven (a more detailed report can be found in Geeraerts, Grondelaers & Speelman 1999). The reader who is not familiar with the linguistic situation in the Dutch speaking region, which comprises Belgian Flanders in the south and The Netherlands in the north, should keep in mind that in the Renaissance, linguistic standardization was well on its way in many European countries. In Flanders, however, the ongoing standardization was blocked as a result of the political separation of Flanders from the Netherlands. Instead of developing its own linguistic standard, Flanders increasingly used French as a supraregional standard language. In the nineteenth century, this frenchification was at its height in all public domains. Although Dutch was reinstated as the official language in Flanders during the reunification of the Southern and the Northern Netherlands (1814–1830), and standard Dutch emerged as the supraregional Flemish language in the 1930s, the standardization of Belgian Dutch is not completely finished up to this day. An issue which occupies students of Dutch in this respect concerns the question whether – and to what extent – Belgian Dutch orients itself to Netherlandic Dutch in its development. Is there convergence between Belgian and Netherlandic Dutch in the sense that Belgian Dutch takes Netherlandic Dutch as a norm, or is there divergence, in the sense that Belgian Dutch is developing its own separate standard?

The existing literature lists conflicting tendencies (see Geeraerts, Grondelaers & Speelman 1999, for an overview). On the one hand, there is the traditional expectation of convergence, due to a growing standardization in Belgian Dutch along with a normative orientation towards Netherlandic Dutch. On the other hand, research into the development of pronunciation differences clearly shows actual divergence (Van de Velde 1996).

To settle matters in the lexical domain, we have embarked on a study which is innovative in regard to its empirical basis and its analytical methodology. Empirically, our study is based on a 40,000-record database of lexical items by means of which 30 concepts from two lexical fields – 15 concepts from the field of clothing and 15 concepts from the field of foot-

ball – are designated. This corpus is structured to accommodate geographical variation – Belgian and Netherlandic sources are compared – as well as diachronic variation. The diachronic variable is implemented by comparing data from 1950, 1970, and 1990. Recall that in this investigation of formal onomasiological variation, the starting-point is the onomasiological profile of a concept in a language sample, i.e., the set of alternative terms (each with its individual frequency) for a concept attested in a source or in a collection of sources (for instance, the Belgian magazines). Recall also that semantic variation is factored out by restricting the analysis to synonymous designations of the same conceptual category. For instance, only the synonyms *legging*, *leggings*, and *caleçon* ‘legging’ are included in the onomasiological profile of the category LEGGING, but not the hyperonym *broek* ‘trousers’ or the hyponym *stretchlegging* ‘legging made of elastic fabric’.

Analytically, convergence has been quantified in terms of the degree of uniformity between the onomasiological profiles in two language samples, whereby convergence equates with increasing uniformity, and divergence equates with decreasing uniformity. Maximal uniformity between two samples obtains when the distribution of synonyms in the onomasiological profiles in the two samples is identical, either because only one name is used for a concept in the two varieties (as in Table 5), or because several names are used with identical relative frequencies (as in Table 6).

Table 5. Identical onomasiological profiles with 1 item.

CONCEPT X	Var A	Var B
<i>item 1</i>	56 = 100%	20 = 100%

Operationally defined as the sum of the minimal frequency per item over the compared varieties, this uniformity measure (to be called  $\text{MIN}_{\text{item}}$ ) yields maximal uniformity for the examples in Tables 5 and 6: since the frequency per item over varieties is identical in these examples, the sum cannot be anything but 100% in Table 5 and  $\text{MIN}_{\text{item}1}(50;50)+\text{MIN}_{\text{item}2}(35;35)+\text{MIN}_{\text{item}3}(15;15) = 100\%$  in Table 6.

Table 6. Identical onomasiological profiles with 3 items.

CONCEPT X	Var A	Var B
<i>item 1</i>	100 = 50 %	20 = 50 %
<i>item 2</i>	70 = 35 %	14 = 35 %
<i>item 3</i>	30 = 15 %	6 = 15 %

Table 7. Differing onomasiological profiles.

CONCEPT X	Var A	Var B
<i>item 1</i>	56 = 80 %	20 = 50 %
<i>item 2</i>	7 = 10 %	14 = 35 %
<i>item 3</i>	7 = 10 %	6 = 15 %

In the – more realistic – situation illustrated in Table 7, the uniformity between Var A and Var B is the sum of  $\text{MIN}_{\text{item}1}(80;50)$ ,  $\text{MIN}_{\text{item}2}(10;35)$  and  $\text{MIN}_{\text{item}3}(10;15)$ , which is 70%.

The degree of uniformity between varieties A and B as a whole is quantified as the average uniformity of the concepts on which the comparison between A and B is based. In this type of global uniformity measure, the 15 concepts in the fields of clothing and football carry the same weight. Should we wish, however, to take into account the real frequency with which a concept appears in our material, we do not average the uniformity per concept, but sum the degrees of uniformity per concept weighted with the relative frequency of the concept within the total set of concepts. This weighted procedure is preferable from the pragmatic point of view that significantly different onomasiological profiles may cause communicative inconvenience where high frequency concepts are concerned, whereas a low degree of uniformity will not normally lead to communicative difficulties when low frequency concepts are involved.

Let us apply this weighted procedure to the field of clothing in Figure 2, and the field of football in Figure 3. In these diagrams, B stands for Belgian Dutch, N for Netherlandic Dutch, and 50, 70, and 90 for respectively the 1950, 1970, and 1990 material. The figure on the left vertical axis of the diagrams reflects the evolution of uniformity in the Belgian material,

whereas the figure on the right vertical axis represents the same evolution in the Netherlandic material. The figures on the horizontal axes denote the evolution from 1950 to 1990 of the uniformity between Belgian and Netherlandic Dutch.

The data in Figures 2 and 3 exhibit a gradual – albeit modest – increase of uniformity – and hence convergence – between Belgian and Netherlandic Dutch, as represented by the lexical fields of football (69.84 < 74.59 < 81.70) and clothing (66.22 < 72.04 < 77.08). In addition, the lower figures on the left vertical axis (56.95 < 63.54 and 69.54 < 75.20) indicate that linguistic change is more significant in Belgian Dutch, which is indicative – in the context of convergence – of a normative orientation of Belgian Dutch towards Netherlandic Dutch.

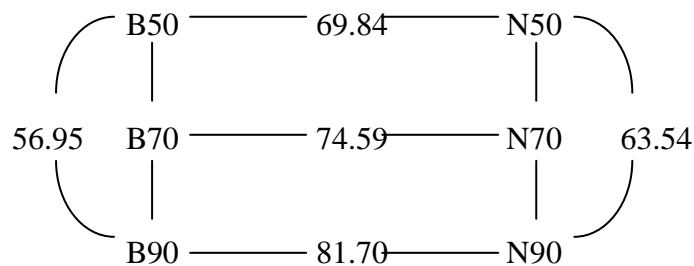


Figure 2. Uniformity in the field of clothing terms.

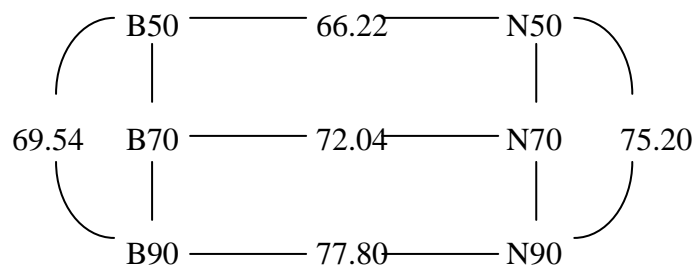


Figure 3. Uniformity in the field of football terms.

There is a lot more, of course, to the linguistic situation in Belgium and The Netherlands than this limited case study could possibly show (again, see Geeraerts, Grondelaers & Speelman 1999, for a more elaborate over-

view). Rather, however, than showing definite results, it was our intention in this sample study to demonstrate that sociolinguistic and especially language-geographical diversity manifests itself at its best where variation cannot be reduced to semantic differences.

## 8. Conclusion

Let us briefly go over the observations in this paper. As far as lexical choice is concerned, we have established that the choice of a lexical item as the name for a particular referent is determined by the degree of prototypicality of the referent with regard to the semasiological structure of the category, by the onomasiological entrenchment of the category represented by the name, and by contextual features which interact with these principles. Of these three components, only semasiological salience has enjoyed some theoretical attention in mainstream Cognitive Linguistics, though predominantly in the shape of prototypicality effects, never in the context of lexical selection. The importance of the onomasiological perspective has mostly been neglected, and interfering contextual factors are – incorrectly – assumed to be outside the scope of Cognitive Linguistics.

In order to give these components the attention they deserve, Cognitive Linguistics is in need of a more integrated conception of lexical analysis. This entails, first and foremost, a theoretical framework in which distinct types of semantic and non-semantic variation are minutely identified, distinguished, and disentangled. In earlier work, we have provided the preliminaries to such a model. This paper contains an example of semasiological and onomasiological variation, but it concentrates chiefly on the treatment of contextual, sociolinguistic variation; in this respect, it advocates the integration into Cognitive Lexicology of a genuine sociolinguistics of the lexicon, or sociolexicology. An integrated model of lexical studies presupposes, second, a unified methodological framework. We have demonstrated with a number of case studies that all the varieties of semantic, non-semantic, and contextual variation discussed in this paper are amenable to the same type of quantitative analysis, viz., the observation of non-elicited language data. The demonstration, surely, has been based on a single example in each case, but more examples may be found in Geer-

aerts, Grondelaers and Bakema (1994), Geer-aerts (1997), Grondelaers and Geeraerts (1998), and Geeraerts, Grondelaers and Speelman (1999).<sup>3</sup> It was our special intention in this paper to show that contextual, sociolinguistic factors can fruitfully be studied with non-elicited corpus data. This possibility of a common quantitative approach constitutes a major link between two linguistic disciplines traditionally thought to be so different as sociolinguistics and lexicology.

The importance of the extrapolation of lexicology towards a use-oriented conception of onomasiology, as suggested here, lies in yet another field: the extrapolation from lexicology to grammar that is typical of Cognitive Linguistics might be effectuated here as well. The focus then shifts from lexical selection to the choice of a grammatical construction. The key concept is that of motivation: just like lexical items, constructions can be both semasiologically and onomasiologically motivated. Semasiologically, a construction is motivated if the idea to be expressed fits in the semantic range of the construction. Onomasiologically, the choice of a particular construction is motivated if it is plausible, given the joint effect of prototypicality and entrenchment. The next step to take, obviously, will be to put this model of constructional motivation to the test: can the determination of constructional selection by prototypicality effects and entrenchment be established with the same kind of quantitative methods that we illustrated here for the lexicon?

---

3. Further refinements of the approach described here may be found in the following articles: Geeraerts (2001a, 2001b), Geeraerts and Grondelaers (2000), Grondelaers, Speelman and Geeraerts (2002), Grondelaers, Geeraerts, Speelman and Tummers (2001). In all of these papers, specific attention is paid to the formal onomasiological type of lexical variation and to the quantitative methods that may be used to study it. Geeraerts (2000) offers a typology of the different types of lexical salience that issue from the variational approach presented here. An elaboration in a slightly different direction is Geeraerts (1999), where our preference for an empirical, data-based approach is defended in contrast with an introspection-based conception of lexical studies.

## References

- Baldinger, Kurt  
1980 *Semantic Theory*. Oxford: Basil Blackwell.
- Berlin, Brent and Paul Kay  
1969 *Basic Color Terms: Their Universality and Evolution*. Berkeley and Los Angeles: University of California Press.
- Cruse, D. Alan  
1977 The pragmatics of lexical specificity. *Journal of Linguistics* 13: 153–164.
- Dougherty, Janet  
1978 Salience and relativity in classification. *American Ethnologist* 5: 66–80.
- Geeraerts, Dirk  
1997 *Diachronic Prototype Semantics. A Contribution to Historical Lexicology*. Oxford: Clarendon Press.  
1999 Idealist and empiricist tendencies in Cognitive Linguistics. In: Theo Janssen and Gisela Redeker (eds.), *Cognitive Linguistics: Foundations, Scope, and Methodology*, 163–194. Berlin: Mouton de Gruyter.  
2000 Salience phenomena in the lexicon. In: Liliana Albertazzi (ed.), *Meaning and Cognition*, 125–136. Amsterdam: John Benjamins.  
2001 a Everyday language in the media: The case of Belgian Dutch soap series. In: Matthias Kammerer, Klaus-Peter Konerding, Andrea Lehr, Angelika Storrer, Caja Thimm and Werner Wolski (eds.), *Sprache im Alltag: Beiträge zu neuen Perspektiven in der Linguistik*, 281–291. Berlin: Mouton de Gruyter.  
2001 b On measuring lexical convergence. In: Augusto Soares da Silva (ed.), *Linguagem e Cognição: A Perspectiva da Linguística Cognitiva*, 51–61. Braga: Associação Portuguesa de Linguística.  
2002 a The scope of diachronic onomasiology. In: Vilmos Ágel, Andreas Gardt, Ulrike Hass-Zumkehr and Thorsten Roelcke (eds.), *Das Wort: Seine strukturelle und kulturelle Dimension*, 29–44. Tübingen: Max Niemeyer.  
2002 b The theoretical and descriptive development of lexical semantics. In: Leila Behrens and Dietmar Zaefferer (eds.), *The Lexicon in Focus: Competition and Convergence in Current Lexicology*, 23–42. Frankfurt: Peter Lang.
- Geeraerts, Dirk and Stefan Grondelaers  
2000 Purism and fashion: French influence on Belgian and Netherlandic Dutch. *Belgian Journal of Linguistics* 13 (1999): 53–68.

- Geeraerts, Dirk, Stefan Grondelaers and Peter Bakema  
1994 *The Structure of Lexical Variation: Meaning, Naming, and Context*. (Cognitive Linguistics Research 5.) Berlin: Mouton de Gruyter.
- Geeraerts, Dirk, Stefan Grondelaers and Dirk Speelman  
1999 *Convergentie en divergentie in de Nederlandse woordenschat: Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: P.J. Meertensinstituut.
- Grondelaers, Stefan and Dirk Geeraerts  
1998 Vagueness as a euphemistic strategy. In: Angeliki Athanasiadou and Elzbieta Tabakowska (eds.), *Speaking of Emotions: Conceptualization and Expression* (Cognitive Linguistics Research 10), 357–374. Berlin: Mouton de Gruyter.
- Grondelaers, Stefan, Dirk Speelman and Dirk Geeraerts  
2002 Regressing on ‘er’: Statistical analysis of texts and language variation. In: Anne Morin and Pascale Sébillot (eds.), *6ièmes Journées internationales d'Analyse statistique des Données Textuelles - 6th International Conference on Textual Data Statistical Analysis*, 335–346. Rennes: Institut National de Recherche en Informatique et en Automatique.
- Grondelaers, Stefan, Dirk Geeraerts, Dirk Speelman and José Tummers  
2001 Lexical standardisation in internet conversations: Comparing Belgium and The Netherlands. In: Josep M. Fontana, Louise McNally, M. Teresa Turell and Enric Vallduví (eds.), *Proceedings of the First International Conference on Language Variation in Europe*, 90–100. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada, Unitat de Investigació de Variació Lingüística.
- Labov, William  
1966 *The Social Stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.  
1972 Some principles of linguistic methodology. *Language in Society* 1: 97–120.  
1973 The boundaries of words and their meanings. In: Charles-James Bailey and Roger Shuy (eds.), *New Ways of Analyzing Variation in English*, 340–373. Washington, DC: Georgetown University Press.

- 1978            Denotational structure. In: Donka Farkas, Wesley Jacobsen and Karol Todrys (eds.), *Papers from the Parasession on the Lexicon*, 220–260. Chicago: Chicago Linguistics Society.
- Trudgill, Peter  
1974            *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Van de Velde, Hans  
1996            *Variatie en Verandering in het Gesproken Standaard-Nederlands (1935–1993)*. Ph.D. dissertation, Katholieke Universiteit Nijmegen.