

John Benjamins Publishing Company



This is a contribution from *Languages in Contrast* 11:1

© 2011. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Prefixes in contrast

Towards a meaning-based contrastive methodology for lexical morphology*

Marie-Aude Lefer and Bruno Cartoni

Université catholique de Louvain (Belgium) / Université de Genève (Switzerland)

This paper proposes a meaning-based contrastive methodology for the study of prefixation in English, French and Italian which is easily adaptable to other languages and word-formation processes. Our discussion centres on some of the central methodological and theoretical issues involved in contrastive lexical morphology, an area which, to date, has largely remained under-researched. Precise defining criteria for derivative (and prefix) status are presented in order to decide what counts as a derivative (or as a prefix) and what does not. Emphasis is also put on a fined-grained semantic *tertium comparationis* elaborated for the cross-linguistic investigation of lexical morphology and based on a six-tiered semantic categorisation, *viz.* location, evaluation, negation, quantity, modality, and inchoativity, most of which are further divided into finer subcategories. This macro-approach makes it possible to draw important generalisations about the use of word-formation devices across languages.

Keywords: contrastive lexical morphology, methodology, *tertium comparationis*, prefixation, English/French/Italian

1. Introduction

1.1 Background and aims of the present paper

This paper deals with the major methodological and theoretical issues raised by studies of contrastive lexical morphology. It stems from two large-scale contrastive projects on prefixation: a contrastive study of Italian and French prefixation (Cartoni, 2008a) and a corpus-based contrastive study of English and French prefixation (Lefer, 2009), which were carried out from different perspectives and with different objectives.

The first project aimed to develop bilingual lexeme-formation rules for a machine translation (MT) system with a view to solving lexical incompleteness (i.e. automatically translating constructed neologisms that were not listed in the lexica of such MT systems). Practically speaking, a prototype system was built to translate prefixed neologisms and subsequently evaluated on the basis of large electronic corpora. The second project consisted in a corpus-based study of prefixation in English and French writing. It aimed to provide empirically-based descriptions of English and French prefixation, thereby revisiting largely introspection-based claims made in the contrastive literature (see e.g. Vinay and Darbelnet, 1958/1995: 67; Van Hoof, 1989: 17–18; Van Roey, 1990: 79–80). More importantly, it aimed to assess the role that genre variation plays on the frequency and productivity of derivational prefixes (cf. Baayen, 1994; Plag *et al.*, 1999; Grabar *et al.*, 2006).

These two projects led us to tackle a number of central methodological and theoretical issues related to contrastive lexical morphology in general and to the contrastive study of prefixation in particular. The major aim of this paper is to outline these issues and to suggest an adequate and consistent contrastive methodology for contrastive word-formation.

1.2 Contrastive lexical morphology: State-of-the-art survey

Early contrastive studies were mainly devoted to phonology and grammar, to the detriment of other fields of contrastive research such as lexicology (James, 1980: 29; Fisiak, 1983: 13; Ringbom, 1994: 738). The reason for this is that

(...) the closed systems of grammar and phonology lend themselves better to systematic CA [contrastive analysis] than the more elusive areas of lexis and culture, but the general absence of contrastive lexical and cultural studies also reflected where the emphasis lay in linguistics those days. (Ringbom, 1994: 738)

However, the advent of computerised corpora has revolutionised the scope of contrastive linguistics, making the study of lexis much more accessible (see e.g. Granger, 2003). Although it can safely be argued that contrastive lexicology has now emerged as a field of study in its own right (see e.g. Altenberg and Granger, 2002), contrastive lexical morphology, i.e. the study of the processes that are used to coin (new) words (e.g. derivation or compounding), still remains the *parent pauvre* of contrastive studies compared to other fields of contrastive research. This relative neglect is testified, among other things, by the paucity of word-formation studies presented at the five *International Contrastive Linguistics Conferences*.

Traditional (i.e. pre-corpus) contrastive grammars such as Vinay and Darbelnet (1958/1995) and Malblanc (1966) provide seemingly useful information

on some of the major cross-linguistic similarities and differences between word-formation systems. Vinay and Darbelnet (1958/1995:67), for example, state that French derivation and compounding are less productive than in English, providing a handful of carefully-chosen examples to prove their point. However, the comments made on word-formation in these grammars remain largely superficial, incomplete or anecdotal. Moreover, although these grammars undeniably represent a mine of useful information for contrastivists, they are mainly introspection-based, or example-based, and sorely lack a sound empirical foundation. Finally, they tend to focus on the end results of word-formation processes, i.e. attested derivatives and compounds, rather than examining the processes as such. These observations lead us to conclude that the claims made on word-formation in contrastive grammars may not be completely reliable and should be put to the test (for example on the basis of corpus analyses).

Lefer (2009) examined approximately 70 studies in contrastive lexical morphology published between 1950 and 2009, thus covering a period of nearly 60 years. The studies used for the survey were selected irrespective of the theoretical background and methodology adopted or languages investigated. The survey reveals that compounding and suffixation have been the main focus of attention, while prefixation, conversion and other word-formation processes such as blending, clipping, abbreviation, etc. have been relatively neglected. It also appears that the majority of studies in contrastive lexical morphology are limited in scope (cf. Kastovsky, 1990). They usually examine one affix or a limited set of affixes across languages (e.g. Mettinger, 1990; Andor, 2005), one particular process such as nominalisation (e.g. Dahech, 2007) or a given category of compounds (e.g. Rufener, 1971). In other words, thorough and data-intensive contrastive studies of morphological systems as wholes (e.g. prefixation or suffixation across two or more languages) have not yet been carried out, at least to our knowledge. Moreover, the survey shows that very few studies discuss contrastive notions such as the concept of the *tertium comparationis* or set up coherent contrastive frameworks for the analysis of lexical morphology across languages, which seriously undermines the scientific rigour of the field. This paper aims to fill that gap in the field.

1.3 Contrastive methodology

Contrastive analysis has traditionally been said to involve three major steps: (1) description, where each of the observed languages is described at the appropriate level and within the same model; (2) juxtaposition, which involves determining what is to be compared with what, i.e. deciding what platform of reference or *tertium comparationis* constitutes the starting point of the contrastive analysis, and (3) comparison or the contrastive analysis proper (James, 1980). Even though

the methodological steps of contrastive analysis are widely discussed by contrastivists such as James (1980), Krzeszowski (1990) or Chesterman (1998), they are seldom explicitly used in contrastive analyses, let alone in the field of contrastive lexical morphology. One of the main reasons for this is that the theoretical and methodological issues of contrastive linguistics have been relatively neglected in contrastive research, where the emphasis has tended to fall on actual practice and applications (Krzeszowski, 1990: 1; see also Fisiak, 1980).

The focus of this paper is primarily on steps (1) and (2), *viz.* description and juxtaposition, and on the methodological and theoretical issues they raise. Step (3), the comparison proper, is only touched upon in passing, because this step largely depends on the specific aims of the contrastive study. Section 2 is devoted to the description stage, *i.e.* the stage which outlines the scope of the study, the definition of the elements to be compared across languages, the data, and the choice of theoretical background. Section 3 then examines the juxtaposition stage by illustrating the many difficulties linked to the task of comparing morphological items across languages and presenting a proposal for a semantic *tertium comparationis*. Section 4 briefly discusses the third step, *i.e.* the comparison proper, in order to illustrate the types of studies that can be based on the methodological framework presented here. Section 5 contains some concluding remarks.

2. Step 1: Description

2.1 Objects to be compared: On the defining criteria used in word-formation studies

2.1.1 *General considerations*

The first stage of any contrastive analysis involves describing the elements to be compared across languages. This task is far from straightforward in the field of word-formation, as there is yet no general consensus on what counts as a derivative (or as a prefix or suffix) and what does not. This lack of consensus is readily apparent, for example, in reference books on English word-formation and in English grammars. First, it is striking to note that the defining criteria used to decide whether a word can be considered as a derivative or not are rarely explicitly stated (see *e.g.* Plag, 2003; Carstairs-McCarthy, 2002; Adams, 2001; Bauer, 1983; Biber *et al.*, 1999; Greenbaum, 1996 and Quirk *et al.*, 1985, which fail to provide any explicit list of defining criteria). In addition, these defining criteria, whether implicitly or explicitly adopted, are not systematically considered as valid across the literature. For instance, while the criterion of formal analysability is widely accepted, semantic transparency (*i.e.* a word can be considered as derived if its

meaning can be inferred on the basis of the meaning conveyed by the base and by the affix it contains) is rejected as a defining criterion for derivative status in Bassac (2004) and in Tournier (1985), while it is adopted in Marchand (1969) and in Bauer and Huddleston (2002). Diverging definitions can also be noted in the reference literature on Italian and French word-formation.

As stated in Section 1.2, most studies in contrastive lexical morphology are limited in scope in that they focus on one aspect of lexical morphology, such as one affix or one set of affixes. This partly explains the lack of consistent contrastive framework for lexical morphology as the restricted view adopted by the few researchers who have ventured into contrastive morphology has blurred the most fundamental issues in the field, such as the need to provide a clear set of defining criteria with which to decide what constitutes a derivative and an affix and what does not. In this paper, we wish to argue that morphological processes have to be considered together as a system (the macro-approach) rather than as individual and separate phenomena (the micro- or affix-by-affix approach).

From a contrastive perspective, it is highly important to precisely delimit, define and describe the elements to be compared across languages as the comparability of the data, and hence the validity and reliability of the results, depends to a great extent on this description stage. In other words, we have to make sure that we are comparing like with like across languages.

2.1.2 *Defining criteria for derivative status*

One of the first issues that has to be tackled in a contrastive study of derivation concerns the definition of 'derived word' (or derivative). As already mentioned in Section 2.1.1, there are diverging viewpoints in the monolingual literature on English, French and Italian word-formation on this fuzzy notion. More precisely, the definition of the notion of derivative depends on the adoption or rejection of a wide range of defining criteria such as formal analysability (e.g. 'un-happy'), phonetic transparency (e.g. 'business' vs. 'happiness'), semantic transparency (e.g. 'mortal' vs. 'accidental'), productivity (e.g. 'ity' vs. 'ness' or 'in' vs. 'un') and recurrence (i.e. the existence of a derivational family, e.g. 'happiness', 'greatness', 'faithfulness', 'carelessness', etc.), among others. This is illustrated in Figure 1. It is therefore crucial to consider these various defining criteria, to adopt or reject them, to decide from the start what is to be considered as a derivative and what as a simplex (i.e. a non-morphologically constructed word), and then to ensure that the criteria adopted make it possible to disambiguate borderline cases. This first step ensures the reliability of the results of the contrastive study, as the same criteria are applied in the languages compared.¹

As can be seen from Figure 1, the inventories of the derivational affixes of the languages investigated also play a role in determining derivative status (if a given

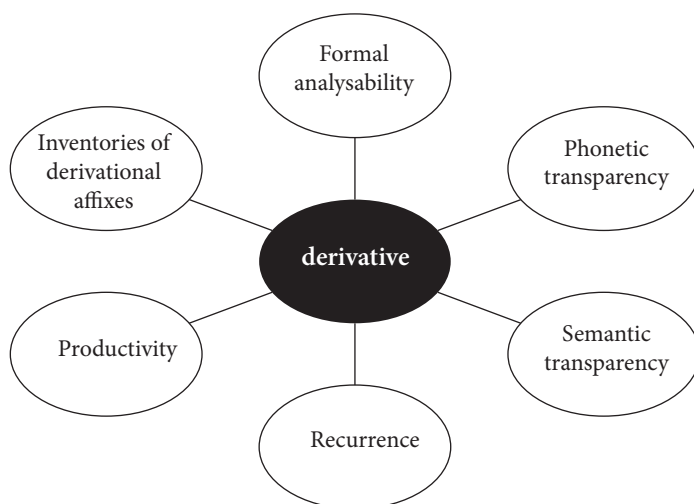


Figure 1. Defining criteria for derivative status (adapted from Lefer, 2009: 76).

word part is considered as an affix, the word that contains it can be classified as a derivative). This issue is addressed in Section 2.1.3.

2.1.3 *Inventories of derivational affixes*

Comprehensive inventories of derivational affixes are difficult to draw up, mainly because of the lack of clear-cut boundaries between derivation and compounding (see e.g. Bauer, 2005). This is a particularly contentious issue for prefixation, the main focus of this paper.²

The category of prefixes is closely related to two other types of lexical element: prepositions and combining forms (see e.g. Warren, 1990; Lehrer, 1998; Fradin, 2000; Bauer, 2005).³ Prepositions are obviously very close to prefixes in many languages, some of them sharing the same form and approximately the same meaning(s) as the corresponding prefix (e.g. Fr. *sous*, It. *sotto* and Eng. ‘under’). However, from a synchronic perspective, prefixes and prepositions have to be clearly distinguished, notably in view of the diverging semantic features they may display. For example, as mentioned by Dal (2003) in relation to French *contre* (which displays both prepositional and prefixal uses), the preposition *contre* mainly conveys an adversative and a locative meaning, while the prefix *contre* tends to coin lexemes with the meaning of ‘a reply to the base’ (as in *contre-espionage*, *contre-amiral*).

The other difficulty in describing the category of prefixes is due to the blurred boundaries between prefixes and combining forms (CFs). Amongst CFs (which are all by definition bound), a distinction can be made between Latin and Greek CFs that are still used in word-formation today (such as ‘paleo’, ‘graph’, etc.) and

Table 1. Inventory of French prefixes in three reference works.**Huot (2001) — 7 prefixes***a/ad, co/con, é/ex, en/em/in, dé/dés/dis, r(e)/r and in***Béchade (1992) — 57 prefixes***a, ab, ad, anté, anti, après, archi, arrière, avant, bien, bis, circon, cis, com, contre, di, dis, dys, en, en, épi, ex, extra, hémi, hyper, hypo, in, infra, inter, intra, juxta, mal, més, méta, mi, non, outre, para, péné, per, péri, plus, post, pré, pro, re, rétro, sans, semi, sous, sub, super, supra, sur, sus, trans, tri, ultra, vice***Thiele (1987) — 72 prefixes***a, a(n), amphi, ana, anté, anti, après, arch(i), arrière, avant, bien, circon, circum, cis, co, contre, contra, dé, dia, demi, dis, ecto, é, en¹, en², endo, entre, épi, eu, ex, exo, extra, hémi, hors, hyper, hypo, in, infra, inter, intra, intro, juxta, mal, més, méta, mi, non, outre, para, per, pén, péri, post, pré, presque, pro, quasi, re, rétro, sans, semi, sous, sub, super, supra, sus, sur, syn, tout, trans, ultra, vice*

CFs that are truncated or suppletive forms of lexemes (such as ‘eco’, ‘bio’, etc. in ‘eco-drive’ or ‘bio-food’). Etymologically, CFs can progressively lose their link with the lexeme they originally come from and become real prefixes.

These tight links with two other types of word-formation elements account for the difficulty of establishing comprehensive inventories of derivational affixes across languages. In addition, it is important to note that reference books and grammars are strikingly inconsistent in their classification of word-formation elements. Table 1 illustrates three divergent literature-based comprehensive inventories of French prefixes presented in Huot (2001), Béchade (1992) and Thiele (1987). They contain 7, 57 and 72 prefixes respectively.

One way of dealing with the fuzziness of this category is to opt for a dynamic definition of prefix. According to this view (inspired by prototypical approaches such as Montermini, 2002; Amiot, 2004), elements can gradually become prefixes through the process of grammaticalisation, i.e. the process by which lexical items come to serve grammatical functions (see Hopper and Traugott, 2003). In other words, some elements are considered more or less prefix-like than others on a ‘prefixity’ scale. This view allows the researcher to decide where prefixation stops and where compounding starts on the prefixity continuum. The main advantage of this approach in a contrastive perspective is that the same limits can be used on the continuum to determine which elements count as prefixes in the languages compared, and which do not.

An additional issue involved in setting up inventories of prefixes is related to the notion of productivity (see e.g. Plag, 1999; Bauer, 2001). When comparing two languages, one has to decide whether to focus on all word-formation elements or only on the productive ones (as is often the case). Even though some qualitative

views on productivity state that productivity is a yes or no question (i.e. a process is either productive or completely unproductive), most morphologists recognise the gradable nature of the concept. Empirical studies (experimental or corpus-based) have to be performed to tease apart productive affixes from unproductive ones. A number of corpus-based productivity measures have been developed over the last twenty years to quantitatively assess derivational productivity (see Baayen's studies, e.g. Baayen and Lieber, 1991). This is not straightforward, however, as comparable corpora of sufficient size that could be used to carry out such measures cross-linguistically are rather scarce (for example, Baayen and collaborators rely on corpora of tens of millions of words, whilst most multilingual corpora are much smaller).

A number of options are available to those who wish to solve this productivity issue. The most obvious solution is to disregard productivity as a selection criterion and to take all affixes into account, whether productive or unproductive. Comprehensive inventories can be set up on the basis of the reference literature on word-formation in the languages investigated. If, by contrast, one wishes to focus on productive affixes only, using existing morphological descriptions is an interesting option (notably because the productivity of affixes is generally assessed in the reference literature), although this option is not without problems, especially because scholars may adopt different definitions of productivity or present intuition-based statements that lack empirical grounding. For example, when literature-based inventories of Italian — supposedly — productive prefixes were used to investigate a large set of unknown words,⁴ there were only 7 different prefixed words beginning in *omni* (e.g. *onnilaterale*) and 21 in *oltre* (e.g. *oltrealpe*) in the Repubblica corpus, both very low type frequencies compared to the prefixed words beginning in *anti* (2,997 types, such as *anticoagulazione*, *antirapinatore*) or *ri* (1421 types, such as *riinvestigazione*, *riassortimento*). This kind of empirical study, based on type frequencies as first rough indicators of productivity (see Baayen and Renouf, 1996), helps broadly assess the productivity of derivational affixes, thus delimiting the inventory of productive affixes and qualifying a number of statements made in the literature.

2.2 Data

The methodological stage of description also entails assembling the data to be compared across languages. Different sources of data can be used in contrastive lexical morphology, such as comparable or translation corpora, (bilingual) dictionaries, and psycholinguistic experiments. These sources of data do not display the same granularity however and are not equal in terms of their ability to compare word-formation processes cross-linguistically.

The survey of studies in contrastive lexical morphology (see Section 1.2) reveals most of the studies involved to be introspection- or dictionary-based. However, closer examination of lexicographical data reveals that derivational affixes and derivatives are not well-represented in bilingual dictionaries and that much remains to be done to adequately include word-formation in dictionaries (see e.g. Prcic, 1999; ten Hacken *et al.*, 2006; Cartoni, 2008b; Lefer, 2009).

The survey also shows that very few studies rely on computerised corpora as the main source of data (notable exceptions are Mauroux and Paillard, 2006; Maniez, 2001; Cvilikaitė, 2005). In corpus-based contrastive research, a distinction is traditionally made between comparable corpora (i.e. original texts in two or more languages matched by criteria such as register, genre, domain, time of publication, size, etc.) and translation corpora (i.e. original texts and their translations into one or several languages). These two types of corpora display complementary advantages in contrastive lexical morphology. Unlike translation corpora, comparable corpora are devoid of translation effects such as translationese, translation norms and universals or outright translation mistakes. They therefore constitute a reliable source of data for measuring productivity, for example, as they show no traces of source language influence. Another major advantage of comparable corpora is that they make it possible to assess the role played by register and genre in the use of lexical morphology across languages (cf. Baayen, 1994; Plag *et al.*, 1999; Grabar *et al.*, 2006; Lefer, 2009), which is not possible if one relies solely on introspection, experimental data or dictionaries. On the other hand, translation corpora make it possible to examine how a particular meaning, such as the meaning conveyed by a set of derivational affixes (e.g. location, negation, quantity, etc.), is expressed in the target language (e.g. by means of morphological, grammatical or lexical means), thereby establishing links between prefixation, suffixation, compounding, lexis and grammar. Finally, corpora (whether comparable or translation) can be used with a view to revisiting the introspection-based claims made in the pre-corpus contrastive literature. As pointed out by Granger (1996:42), “scholars who have adopted a largely impressionistic, non-statistical contrastive approach have a tendency to overgeneralize.” These overgeneralized statements can be qualified by adopting a corpus-based approach to lexical morphology.

Computerised corpora are arguably the most effective type of data in contrastive morphological studies. If the researcher aims to study neologisms, however, corpora may not be suitable as (1) for some language pairs, corpora are not yet available and (2) comparable corpora, if they exist, are not large enough to enable the collection of sufficient data across languages. One solution is to use the Web as Corpus (WaC). The use of the textual resources found on the Internet is very promising in lexical morphology, especially because of its easy and free accessibility. However, even though web-based textual resources are almost ideal for neology

studies, the uncontrolled nature of web data requires a number of methodological precautions (e.g. the distinction between native and non-native language is blurred) (see Hathout *et al.*, 2009; Hundt *et al.*, 2007).

2.3 Theoretical approaches

In the field of contrastive lexical morphology, the choice of linguistic model is also an important issue. In this respect, James (1980) states that “the minimum requirement of ‘parallel description’ is that the two languages be described through the same model of description” (see also Krzeszowski, 1990: 35, 107–108). If the same data from the observed languages are described in different models or theoretical backgrounds, the descriptions will most certainly highlight different facets of the data and be reflections of the use of two models or theories rather than true linguistic contrasts (see e.g. James, 1980: 63–64).

It is important to note here that morphology has a hectic history and is characterised, like the majority of linguistic fields, by the co-existence of a large number of theoretical approaches. As a result, it is often difficult to find morphological descriptions of two or more languages within the same theoretical background. Simplifying somewhat, a broad distinction can be made between the morpheme-based approach and the lexeme- (or word-) based approach to word-formation, in addition to other theories, such as natural morphology (for overviews, see e.g. Kerleroux, 2006; Štekauer and Lieber, 2005). Needless to say, cross-linguistic descriptions of lexical morphology need to be comparable from a theoretical perspective.

3. Step 2: Juxtaposition

3.1 On the notion of *tertium comparationis*

The key issue of the second stage, i.e. juxtaposition, is to determine on what basis elements across languages can be compared (James, 1980: 65), i.e. a *tertium comparationis* (TC) or common platform of reference has to be established, against which similarities and differences between the elements compared can be identified.

The issue of the types of TC suitable for contrastive lexical morphology has not often been raised in the contrastive literature. In this respect, Krzeszowski (1990: 74–75) states that

[c]ontrastive studies of word-formation are better off if they are based on some conceptual framework. Such an analysis could involve a comparison of various means employed in the derivation of *nomina actions*, *nomina agentis*, *nomina loci*,

and also of adjectives of intensity, inclination, possibility, ability, or of verbs of process, causation, instrument, and so on (...). As a matter of fact, any aspect of the meaning can serve as a basis for cross-linguistic comparisons.

This is the line taken in this paper as well, i.e. the adoption of a semantic TC is promoted (see also Hüning, 2009), so that it is possible, for instance, to investigate how the concept of reiterativity or negativity is morphologically expressed across languages.

It should be noted, however, that other types of TC are possible — and probably necessary — in contrastive research on lexical morphology, such as translation equivalence (i.e. the contrastive study of translationally equivalent affixes; e.g. Eng. ‘un’ and Fr. *in* and *dé*) or formal equivalence (i.e. the contrastive study of how affix A is used across languages, such as Eng. and Fr. suffix *able*). In addition, one can investigate the morphological ways in which, for example, verbs are formed out of nouns (i.e. a contrastive study of denominal verbalisation). Whatever the choice made by the researcher, it is of paramount importance to state explicitly the type of TC opted for.

3.2 A semantic *tertium comparationis*

The semantic *tertium comparationis* presented in this section, which was largely inspired by Szymanek (1988), was originally developed for Italian and French prefixation in Cartoni (2008a) and aimed to be applicable to other languages as well. It was successfully adapted to English prefixation in Lefer’s (2009) corpus-based study of English and French prefixation. This adaptation step showed that Cartoni’s TC is exhaustive, fine-grained and readily applicable to corpus data. The semantic TC offers a general framework for the cross-linguistic comparison of prefixation systems as wholes. Further research is needed to determine whether it can be extended to other word-formation processes (such as compounding or suffixation) and to non-morphological phenomena (such as multiword expressions and syntactic constructions).

Practically speaking, the TC was set up by relying on various semantic descriptions of Italian and French prefixation (see Montermini, 2002; Iacobini, 2004; Amiot and Montermini, 2009). A distinction is made between six major semantic categories: location, evaluation, negation, quantity, modality, and inchoativity. The categories are further divided into subcategories. For example, location is divided into space and time, and within spatial location, a distinction is further made between different positions (in front of, behind, beside, etc.). Figure 2 schematically represents the TC.

Once the TC has been set up, the individual affixes (gathered for each of the compared languages according to the criteria delimited in Step 1), can be classified

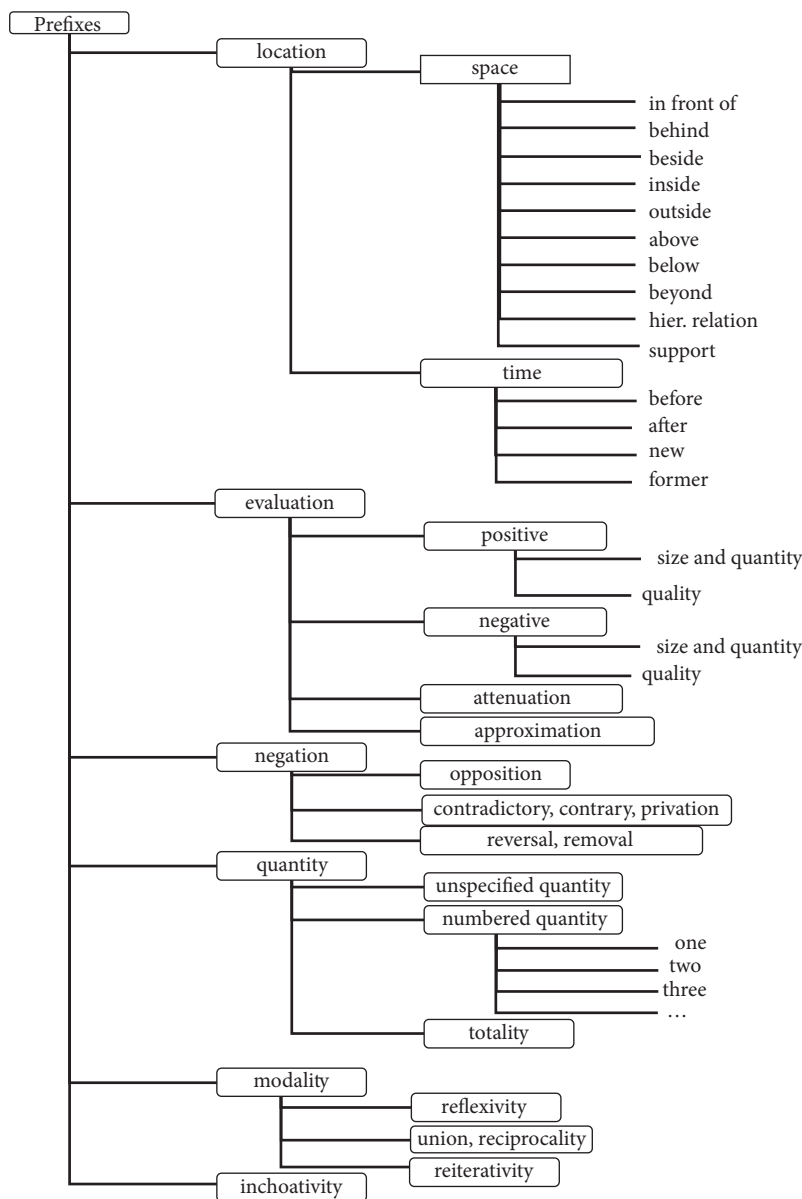


Figure 2. Semantic tertium comparationis.

into the various categories and subcategories. Needless to say, some polysemous or homographic prefixes can be classified into more than one subcategory (It. *sotto*, for example, is both evaluative and locative; Eng. ‘un’ both expresses negation and reversal). Concrete examples of this TC classification are provided in Table 2 (prefixes expressing reversal and removal) and Table 3 (inchoative prefixes).

Table 2. Prefixes expressing reversal and removal in English, French and Italian.

Negation						
	English		French		Italian	
reversal and re- moval	<i>de, dis, un</i>	demonetise, disenfran- chise, unseal	<i>dé, dis</i>	déboiser/ désarmer, disjoindre	<i>de, s, dis</i>	defogliare, smacchiare, disossare

Table 3. Inchoative prefixes in English, French and Italian.

Inchoativity						
English		French		Italian		
<i>be, en</i>	beribbon, entomb	<i>a, é, en</i>	affaiblir, écourter, endurcir	<i>in, s</i>	impresiosire, in- robustire, smagrire, spezzare	

As illustrated in Table 2 and Table 3, the TC gives an overall picture of the number of prefixes involved in each subcategory in the compared languages. Its main advantage is that it allows the researcher to consistently contrast semantic groupings of prefixes cross-linguistically and thereby avoid basing contrastive studies on a (more or less) arbitrarily chosen set of prefixes. The TC fulfils its goal by providing a reliable, language-neutral platform of reference against which cross-linguistic similarities and differences can be observed (see Section 4).

Interestingly, the TC presented here could be further developed by identifying semantic similarities between prefixes, suffixes and compounds (e.g. inchoative prefixes and suffixes, such as 'en' and 'ise', negative prefixes and suffixes, such as 'un' and 'less', number prefixes and compound parts, such as 'uni' and 'one') and between word-formation and grammar, syntax or the lexicon (e.g. inchoative prefixes and causative verbs).

4. Step 3: The comparison proper

The main core of this paper has been devoted to the methodological and theoretical issues of contrastive lexical morphology, i.e. describing the object of study, collecting comparable data and juxtaposing them on a language-neutral platform, viz. a semantic TC. This TC, however, is only a methodological means towards contrasting word-formation systems across languages.

The third step of the contrastive method (i.e. the comparison proper) largely depends on the objectives of the contrastive study. For example, the TC can be used as a basis for the quantitative study of the productivity of prefixation across

languages. Rather than comparing the productivity of individual prefixes across languages (which may prove to be problematic, for example when a given affix A in language X has no equivalent in language Y), the TC makes it possible to contrast semantic groupings of prefixes, thereby allowing interesting generalisations to be made. It was found on the basis of this TC, for example, that negative prefixation (contradictory, contrary and privation) tends to be more productive (in terms of types) in English writing than in French (mainly because of negative ‘un’), while French prefixal reversal and removal (mainly *dé*) is more productive than its English counterpart (despite the existence of two stocks of reversative prefixes in English: one Romance, ‘de’ and ‘dis’, and the other Germanic, ‘un’). Interestingly, this finding indicates that the equation ‘more affixes = higher productivity’ does not hold. Moreover, it shows that the adoption of broad semantic categories (such as negation, which is made up of the subcategories contradiction, contrary, privation on the one hand and reversal and removal on the other) can be misleading and obscure interesting cross-linguistic contrasts. The TC has also brought to light a number of cross-linguistically valid genre variation patterns, such as the higher productivity of prefixes expressing opposition and support in press editorials or the higher productivity of quantitative prefixes in research articles (see Lefer, 2009 for more detail). The TC presented in Section 3 has also been used to implement morphological translation rules to automatically translate morphologically constructed neologisms (see Cartoni, 2008a).

5. Concluding remarks

Contrastive lexical morphology is a relatively neglected research field. This article has sought to present and discuss some of its central methodological issues, with a view to laying a rigorous foundation for what is a highly promising field. Above all, it would seem crucial to adopt precise defining criteria in order to decide what counts as a derivative and what does not and highly advantageous to use a fine-grained semantic *tertium comparationis*. Obviously, the top-down approach suggested in this paper has some limitations (in the sense that language-specific patterns may be overlooked because of the broad classificatory system it relies on). However, we hope to have shown that this macro-approach provides a methodology for contrastive lexical morphology that takes into account morphological systems as wholes and can bring to light fascinating generalisations on the use of lexical morphology in authentic language.

Corpus-based contrastive word-formation studies are useful for a wide range of applied purposes, such as bilingual lexicography, second language learning and teaching and machine translation. Even though the possible applications of the

field are numerous, its contrastive methodology and theoretical issues remain to be clarified. This article represents a first attempt in this direction. We hope that the framework outlined here will pave the way for a series of follow-up exploratory studies.

Notes

* The first author gratefully acknowledges the financial support of the Belgian National Fund for Scientific Research (F.R.S. — FNRS). We wish to thank Sylviane Granger (Université catholique de Louvain) and two anonymous reviewers for their insightful comments on the first draft of this paper.

1. Incidentally, studies that deal with new formations are faced with the issue of defining the notion of neologism, of which many different definitions exist in the literature. Whatever the defining criteria adopted in contrastive studies (e.g. a word is considered as a neologism if it is not found in a reference lexicon or dictionary), it is important to bear in mind that comparable criteria have to be adopted in the languages investigated and that comparable reference tools have to be used, which may be difficult to achieve (e.g. how can the researcher ensure that lexicographical data from monolingual dictionaries in two or more languages are comparable *per se*?).
2. In this paper, we do not address the issue of distinguishing between suffixes and free forms (e.g. Eng. 'type' as in 'workshop-type', 'yoga-type', 'wristwatch-type'; see e.g. Renouf and Baayen, 1998).
3. We use the term 'combining form' as a hyperonym for the elements that are used in clippings and neoclassical formations (see Warren, 1990), such as 'eco' (clipped from 'ecological') in 'eco-drive' and 'geo' and 'logy' in 'geology'.
4. The set of unknown words was obtained by applying a large reference lexicon of Italian Mmorph (Bouillon *et al.*, 1998) to a large newspaper corpus — La Repubblica, 380 m. occurrences (see Baroni *et al.*, 2004).

References

- Adams, V. 2001. *Complex Words in English*. Harlow: Pearson Education / Longman.
- Altenberg, B. and Granger, S. (eds). 2002. *Lexis in Contrast. Corpus-based Approaches*. Amsterdam/Philadelphia: John Benjamins.
- Amiot, D. 2004. "Préfixes ou prépositions? Le cas de *sur*(-), *sans*(-), *contre*(-) et les autres". *Lexique* 16:67–83.
- Amiot, D. and Montermini, F. 2009. "Affixes et mots grammaticaux". In *Aperçus de morphologie du français*, B. Fradin, F. Kerleroux and M. Plénat (eds), 127–141. Saint-Denis: Presses Universitaires de Vincennes.
- Andor, J. 2005. "A Lexical Semantic-Pragmatic Analysis of the Meaning Potentials of Amplifying Prefixes in English and Hungarian. A Corpus-based Case Study of Near Synonymy".

- Proceedings from The Corpus Linguistics Conference Series*, 1(1). Available at <http://www.corpus.bham.ac.uk/PCLC/> [last accessed 12 January 2010].
- Baayen, R.H. 1994. "Derivational productivity and text typology". *Journal of Quantitative Linguistics* 1:16–34.
- Baayen, R.H. and Lieber, R. 1991. "Productivity and English derivation: a corpus-based study". *Linguistics* 29(4):801–844.
- Baayen, R.H. and Renouf, A. 1996. "Chronicling the Times: Productive Lexical Innovations in an English Newspaper". *Language* 72(1):69–96.
- Baroni, M., Bernardini, S., Comastri, F., Piccioni, L., Volpi, A., Aston, G. and Mazzoleni, M. 2004. "Introducing the "La Repubblica" corpus: A large, annotated, TEI(XML)-compliant corpus of newspaper Italian". *Proceedings of LREC 2004*, Lisbon. 1771–1774.
- Bassac, C. 2004. *Principes de morphologie anglaise*. Bordeaux: Presses Universitaires de Bordeaux.
- Bauer, L. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bauer, L. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
- Bauer, L. 2005. "The borderline between derivation and compounding". In *Morphology and its demarcations*, W.U. Dressler, D. Kastovsky, O.E. Pfeiffer and F. Rainer (eds), 97–108. Amsterdam/Philadelphia: John Benjamins.
- Bauer, L. and Huddleston, R. 2002. "Lexical word-formation". In *The Cambridge Grammar of the English Language*, R. Huddleston and G.K. Pullum (eds). Cambridge: Cambridge University Press.
- Béchade, H.D. 1992. *Phonétique et morphologie du français moderne et contemporain*. Paris: Presses Universitaires de France.
- Biber D., Johansson S., Leech G., Conrad S. and Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Bouillon, P., Lehmann, S., Manzi, S. and Petitpierre, D. 1998. "Développement de lexiques à grande échelle". *Proceedings of Colloque des journées LTT de TUNIS*, Tunis. 71–80.
- Carstairs-McCarthy, A. 1994. "Morphological Universals". In *The Encyclopedia of Language and Linguistics*, R.E. Asher (ed), 2553–2557. Oxford: Pergamon Press.
- Carstairs-McCarthy, A. 2002. *An Introduction to English Morphology. Words and their Structure*. Edinburgh: Edinburgh University Press.
- Cartoni, B. 2008a. *De l'incomplétude lexicale en traduction automatique: vers une approche morphosémantique multilingue*. PhD Thesis, ISSCO/TIM ETI, Université de Genève.
- Cartoni, B. 2008b. "La place de la morphologie constructionnelle dans les dictionnaires bilingues: étude de cas". *Proceedings of Euralex 2008*, Barcelone, Spain, IULA: 813–820.
- Chesterman, A. 1998. *Contrastive Functional Analysis*. Amsterdam/Philadelphia: John Benjamins.
- Croft, W. 2003. *Typology and universals*. Cambridge: Cambridge University Press.
- Civilikaitė, J. 2007. "The cases of non-equivalence between English and Lithuanian: a corpus-based solution". *Kalbotyra* 57(3):46–55.
- Dahech, D. 2007. "Les nominalisations déverbales en français et en arabe: étude exploratoire des critères de classement". *Les Cahiers du CRISCO* 23:128–138. Available at <http://www.crisco.unicaen.fr/IMG/pdf/cahiersCRISCO23.pdf> [last accessed 12 January 2010].
- Dal, G. 2003. "Arguments pour un préfixe contre-". *Recherches linguistiques* 26:172–201.
- Fisiak, J. 1980. "Introduction". In *Theoretical Issues in Contrastive Linguistics*, J. Fisiak (ed), 1–7. Amsterdam/Philadelphia: John Benjamins.
- Fisiak, J. 1983. "Present trends in contrastive linguistics". In *Cross-language analysis and second language acquisition 1*, K. Sajavaara (ed), 9–38. Jyväskylä: University of Jyväskylä.

- Fradin, B. 2000. "Combining forms, blends and related phenomena". In *Extragrammatical and Marginal Morphology*, U. Dolsescha and A. Thornton (eds), 11–59. Munich: Lincom Europa.
- Gaeta, L. 2003. "Word Formation and Typology: Which Language Universals?". *Proceedings of the Fourth Mediterranean Morphology Meeting (MMM4)*, Catania. 157–168.
- Goddard, C. and Wierzbicka, A. (eds). 2002. *Meaning and Universal Grammar — Theory and Empirical Findings*. Amsterdam/Philadelphia: John Benjamins.
- Grabar, N., Tribout, D., Dal, G., Fradin, B., Hathout, N., Lignon, S., Namer, F., Plancq, C., Yvon, F. and Zweigenbaum, P. 2006. "Productivité quantitative des suffixations par *-ité* et *-able* dans un corpus journalistique moderne". *Proceedings of TALN 2006*, Leuven, 10–13 April 2006.
- Granger, S. 1996. "From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora". In *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*, K. Aijmer, B. Altenberg and M. Johansson (eds), 37–51. Lund: Lund University Press.
- Granger, S. 2003. "The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies". In *Corpus-based Approaches to Contrastive Linguistics and Translation Studies*, S. Granger, J. Lerot and S. Petch-Tyson (eds), 17–29. Amsterdam/New York: Rodopi.
- Greenbaum, S. 1996. *The Oxford English Grammar*. Oxford: Oxford University Press.
- Greenberg, J.H. 1966. "Some universals of grammar with particular reference to the order of meaningful elements". In *Universals of Language*, J.H. Greenberg (ed). Cambridge, MA: MIT Press.
- Hacken, P. ten, Abel, A. and Knapp, J. 2006. "Word Formation in an Electronic Learners' Dictionary: Eldit". *International Journal of Lexicography* 19(3):243–256.
- Hathout, N., Namer, F., Plénat, M. and Tanguy, L. 2009. "La collecte et l'utilisation des données en morphologie". In *Aperçus de morphologie du français*, B. Fradin, F. Kerleroux and M. Plénat (eds), 267–287. Saint-Denis: Presses Universitaires de Vincennes.
- Hopper, P. and Traugott, E. 2003. *Grammaticalization*. Cambridge: Cambridge University Press.
- Hundt, M., Nesselhauf, N. and Biewer, C. (eds). 2007. *Corpus Linguistics and the Web*. Amsterdam: Rodopi.
- Hüning, M. 2009. "Semantic niches and analogy in word formation. Evidence from contrastive linguistics". *Languages in Contrast* 9(2):183–201.
- Huot, H. 2001. *Morphologie, Forme et sens des mots français*. Paris: Armand Colin.
- Iacobini, C. 2004. "I prefissi". In *La formazione delle parole in italiano*, M. Grossmann and F. Rainer (eds), 99–163. Tübingen: Niemeyer.
- James, C. 1980. *Contrastive analysis*. Harlow: Longman.
- Kastovsky, D. 1990. "Word-formation and translation". *Meta* 35(1):45–49.
- Kerleroux, F. 2006. "Les théories morphologiques à la fin du XXe siècle". In *History of the Language Sciences. An International Handbook on the Evolution of the Study of Language from the Beginnings to the Present*, S. Auroux, E.F.K. Koerner, H.-J. Niederehe and K. Versteegh (eds), 2313–2324. Berlin/New York: De Gruyter.
- Krzyszowski, T.P. 1990. *Contrasting Languages. The Scope of Contrastive Linguistics*. Berlin/New York: Mouton de Gruyter.
- Lefer, M-A. 2009. "Exploring lexical morphology across languages. A corpus-based study of prefixation in English and French writing". PhD thesis, Université catholique de Louvain.

- Lehrer, A. 1998. "Scapes, Holics and Thons: The semantics of combining forms". *American Speech* 73(1):3–28.
- Malblanc, A. 1966. *Stylistique comparée du français et de l'allemand. Essai de représentation linguistique comparée et Étude de traduction*. Paris: Didier.
- Maniez, F. 2001. "La traduction du nom adjectival en anglais médical". *Meta* 46(1):56–67.
- Marchand, H. 1969. *The Categories and Types of Present-Day English Word-Formation. A Synchronic-Diachronic Approach*. München: C.H. Beck'sche Verlagsbuchhandlung.
- Mauroux, S. and Paillard, M. 2006. "Affixes Dérivationnels Causatifs: Contrastes Français-Anglais". In *Causalité et Contrastivité. Études de Corpus*, H. Chuquet and M. Paillard (eds), 20–33. Rennes: Presses Universitaires de Rennes.
- Mettinger, A. 1990. "Comparing the incomparable? English adjectives in *-able* and their rendering in Modern Chinese". In *Further Insights into Contrastive Analysis*, J. Fisiak (ed). Amsterdam/Philadelphia: John Benjamins.
- Montermini, F. 2002. "Le système préfixal en italien contemporain". PhD thesis, Université de Paris X-Nanterre/Università degli Studi di Bologna.
- Plag, I. 1999. *Morphological Productivity. Structural Constraints in English Derivation*. Berlin/New York: Mouton de Gruyter.
- Plag, I. 2003. *Word-Formation in English*. Cambridge: Cambridge University Press.
- Plag, I., Dalton-Puffer, C. and Baayen, R.H. 1999. "Morphological productivity across speech and writing". *English Language and Linguistics* 3(2):209–228.
- Prcic, T. 1999. "The treatment of affixes in the "big four" EFL dictionaries". *International Journal of Lexicography* 12(4):263–279.
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Renouf, A. and Baayen, R.H. 1998. "Aviating among the hapax legomena: Morphological grammaticalisation in current British newspaper English". In *Explorations in Corpus Linguistics*, A. Renouf (ed), 181–189. Amsterdam/Atlanta: Rodopi.
- Ringbom, H. 1994. "Contrastive Analysis". *The Encyclopedia of Language and Linguistics*, R.E. Asher (ed), 737–742. Oxford: Pergamon Press.
- Rufener, J. 1971. *Studies in the motivation of English and German compounds*. Zürich: Juris Druck/Verlag.
- Štekauer, P. and Lieber, R. (eds). 2005. *Handbook of Word-Formation*. Dordrecht: Springer.
- Szymanek, B. 1988. *Categories and Categorization in Morphology*. Lublin: RW-KUL.
- Thiele, J. 1987. *La Formation des Mots en Français Moderne*. Translation and adaptation by A. Clas. Montréal: Les Presses de l'Université de Montréal.
- Tournier, J. 1985. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris: Champion.
- Van Hoof, H. 1989. *Traduire L'Anglais. Théorie et Pratique*. Paris/Louvain-la-Neuve: Duculot.
- Van Roey, J. 1990. *French-English Contrastive Lexicology. An Introduction*. Louvain-la-Neuve: Peeters.
- Vinay, J.-P. and Darbelnet, J. 1958/1995. *Comparative Stylistics of French and English: a methodology for translation*. Translated and edited by J. C. Sager, M.-J. Hamel. Amsterdam/Philadelphia: John Benjamins.
- Warren, B. 1990. "The importance of combining forms". In *Contemporary Morphology*, W.U. Dressler, H.C. Luschützky, O.E. Pfeiffer and J.R. Rennison (eds), 111–132. Berlin/New York: Mouton de Gruyter.

Author's address

Marie-Aude Lefer
Centre for English Corpus Linguistics
Université catholique de Louvain
Collège Erasme
Place Blaise Pascal 1
1348 Louvain-la-Neuve
Belgium
marie-aude.lefer@uclouvain.be