Word Formation versus Etymology in Electronic Dictionaries

Pius ten Hacken[*] & Dorota Smyk[×]

[*]Abt. Geisteswiss. Informatik
UniversitŠt Basel
Petersgraben 51
CH-4051 Basel
pius.tenhacken@unibas.ch

[×]Englisches Seminar
UniversitŠt Basel
Nadelberg 6
CH-4051 Basel
dorota.smyk@unibas.ch

Abstract
The problem of distinguishing word formation relationships in the lexicon from etymological relationships arises because of the interaction of different processes for the extension of the lexicon, in particular word formation and borrowing. In an electronic dictionary in which word formation is taken as the organizing principle for the description of the lexicon, the distinction between word formation and etymology is of central importance. A procedure is described which aims to achieve intersubjective and cross-linguistic consistency in decisions about this borderline. This procedure is primarily based on synchronic relationships. When borrowed items can be related by morphological processes in the borrowing language, however, this analysis is preferred. It has been applied successfully in the development of Word Manager dictionaries for English and Italian.

1 Introduction

It is well-known that the lexicon of a language cannot be described exhaustively, even in principle, because it is in continuous development. Whereas in artificial or dead languages the lexicon can be considered as a closed class, in the description of the lexicon of a living language the mechanisms underlying its extension should be taken into account. In describing these mechanisms we generally operate under the assumption that what happened in the past is a reliable basis for the prediction of what may happen in the future.

New words are usually not created arbitrarily, but emerge because there is a need to name a new concept. We use *concept* here in the technical sense of whatever is deemed salient enough to be the meaning of a word. There are three main types of processes involved in the naming of new concepts:

- *Semantic change.* In the course of time the meaning of lexemes develops. There are two important axes of semantic change. On the one hand we find extension or generalization of the meaning. An example is *novice*, originally "one who has entered a religious house", now more common in the sense of "an inexperienced person". On the other hand we find narrowing or specialization of the meaning. An example is *computer*, which used to be the agentive noun to the verb *compute*, but is now almost exclusively used for the device we write electronic dictionaries for. Minor types involve euphemisms, clichŽs, etc.

- *Word formation.* New words are created by the application of word formation processes to existing words. Major types of word formation are derivation, conversion, and compounding. Examples for theseprocesses are the formation of the noun *dreamer* from

the verb *dream*, of the verb *house* from the corresponding noun, and of the compound *dream house*. Minor types include abbreviation, backformation, clipping, etc.

- *Borrowing*. New words are created by taking over words from other languages. In English, major source languages are French, Latin, and Greek, but Crystal [1995:126f.] mentions 120 source languages for borrowing into English, including *hammock* from Spanish, *lager* from German, *kamikaze* from Japanese, *horde* from Polish etc.

Word formation differs from the other two in the type of predictions which can be made on the basis of a historical description of the lexicon. Only in the case of word formation can the processes involved be described both as classes of items exhibiting them and as a set of instructions to be performed in order to create new words. Therefore it is an advantage for an electronic dictionary to include a set of word formation rules complementing the description of attested words.

Inthe Word Manager system, word formation is taken as the organizing principle of the lexicon. This means that word formation rules are not just added as a complement to the lexicon, but every word is specified in terms of its word formation relationships to other words. This design highlights a number of problems which could remain unnoticed in other contexts. An appropriate description of the word formation system of a language is complicated by the interaction of word formation with the other types of processes for the creation of new words, in particular with borrowing. In order to be able to retrieve decisions and ensure consistency in the specification of entries in Word Manager, an explicit lexicographic policy is required. Formulated in coding guidelines with a wide range of exemplification, this policy is used both as a basis for the instruction of lexicographers and as documentation of the resulting resources.

2 Word Manager

Word Manager (WM) is a system for reusable morphological dictionaries. Its general architecture and design objectives are described by ten Hacken & Domenig [1996]. In the view of the lexicon underlying WM, information is divided into two types, morphological information on the one hand and syntactic and semantic information on the other.

In the central position between the two types of information we find the lexeme. Corresponding largely to the term introduced by Matthews [1974], the lexeme in WM generalizes over inflectional word forms as far as morphology is concerned. The lexeme for the verb *drink* includes the word forms *drink, drinking, drank*, etc. Simultaneously the lexeme generalizes over syntactic and semantic specifications which may diverge in different senses of the word. The verbal lexeme *drink* can be specialized as intransitive or transitive and in the latter case the object may be a liquid or, in the expression *drink someone under the table*, a person. The two types of information about which the lexeme constitutes a generalization are independent of each other in the sense that a verb such as *drink* has the same inflection whether it is intransitive or transitive, and whether the object is a liquid or a person.

This view of the organization of lexical information contrasts with the underlying model in many other approaches, where the lexicon includes all information on individual words and

222

is opposed to the rule component. In WM the entire mapping between lexemeand word forms is covered. This includes not only lexical entries, but also morphological rules. For a discussion of this opposition and a motivation of this aspect of the design of WM, cf. [Ten Hacken 1999].

In the development of WM resources, two stages can be distinguished. In the first stage, the morphological rule system of a language, including its inflection and word formation rules, is specified in a morphological rule database. In the second stage, lexical entries are classified in terms of these rules to constitute a morphological dictionary database. This morphological dictionary is the basis for a wide range of practical applications. As morphological rules are available as a backbone of the dictionary, such an application is not a lexical component consisting of a list of entries with information about them, but rather an independently running component dedicated to performing a specific task. Pedrazzini & ten Hacken [1998] describe how finite-state transducers for specific tasks (e.g. lemmatization or morphological analysis) can be derived from a WM database. Pedrazzini [1999] gives a more technical presentation of this procedure. An example of a realistic application is described by Zappatore & ten Hacken [2000].

In the project on which the results presented here are based, the aim is to produce large-scale WM dictionaries for Italian and English. As a starting point, WM rule databases for these languages were available. Shared lexicographic guidelines valid for both languages are used to support consistent lexicographic decisions. This means that in determining, for instance, whether a particular word is a compound, the basic criteria do not refer to language-specific properties.

3 Word Formation in Word Manager

In the organization of the WM lexicon, word formation can be seen as the central structuring process. This entails a number of obvious differences between the treatment of word formation in traditional lexicography and in WM. In traditional lexicography word formation rules are only implicit. Even if the structure of a word is indicated, e.g. [im [possible]], the underlying rules are not formulated. In WM, *impossible* is classified as an instance of (1) prefixation to an adjective (2) involving the prefix *in-* (3) which is changed in this case by a regular phonological process to *im-*. Word formation rules in WM are available both procedurally and declaratively. The latter means that *impossible* is automatically part of classes defined by (1-3) and their superclasses. These classes can be retrieved as sets of all their entries existing in the database. The former means that the word formation rules can be used as instructions for the productive formation of new lexemes. This organization of the lexicon provides powerful instruments both for its use in practice and for its validation, cf. [Ten Hacken 1998; 2002].

As a consequence of the separation of morphology from syntax and semantics, word formation in WM is modelled as a relationship between lexemes rather than individual readings of lexemes. This approach requires special care in determining which forms and meanings should be grouped together as a lexeme. The problem has two sides. First the optimal balance in determining the forms and senses taken together as a lexeme must be found, second the distinction between word formation and etymology has to be optimized.

This means that historical considerations may play a role in the description as far as they influence the future development of the vocabulary.

As to the first side of the problem, the differences in the inflectional paradigm require the distinction of two lexemes *walk* and *bank*, one for the verb and one for the noun, but the transitive and intransitive uses of *walk* and the two homonymous senses of the noun *bank* are collapsed. In the case of the verb *lie*, the homonymy coincides with a difference in past forms, so that two verbal lexemes are distinguished.

As to the other side of the problem, the relationship between etymology and word formation is an intricate one, as argued also by Dworkin [1985]. In some cases they may interact to give conflicting indications as to the structure of a word. Borrowed items such as English *kindergarten* and *rendezvous* are complex entries in the source languages they were borrowed from. *Kindergarten* is a compound in German, *rendez-vous* a nominalized phrase in French. Both words constitute parts of larger word families in their source language. In English, however, they are simple entries, because none of their components is a lexeme in the target language.

It has sometimes been stipulated, e.g. [Van Marle 1985], that the status of a word as borrowed is of central influence in determining whether a word formation process is productive or not. In the view defended here, however, the fact that *kindergarten* is analysed as a simple entry in English is only due to the impossibility of its emergence from a word formation process available in English. In the absence of the base words *kind* and *garten* in English, compounding cannot produce *kindergarten*.

An example illustrating the difference between the two points of view is the analysis of English *conjectural*. This word is a borrowing from Latin, where it appears as *coniecturalis*. There is also a word *conjecture*, borrowed from Latin as well. The formation of adjectives from nouns by the suffix *-al* in English is widely attested, but there are very few examples of native words undergoing this process, e.g. *coastal*. In van Marle's analysis, the classification of the process as productive in English would be doubtful and depend crucially on such examples of *coastal*. In our view, the several hundred of instances provide ample evidence that the process exists as part of the English word formation system. Independently of examples such as *coastal*, the relationship between *conjectural* and *conjecture* should be recognized as a word formation relationship. After a certain number of cases in which pairs of words have been borrowed, the regularity is reconstructed as a rule of word formation in English. The emergence of the process can be termed a reconstruction because the process in English is based on the results of a similar process in Latin.For the items borrowed before the reconstruction of the process, a reanalysis takes place, resulting in the classification as a complex entry rather than a simple entry. Similar processes can be found in neo-classical word formation, cf. [Petropoulou & ten Hacken 2002].

A somewhat different situation is found in the emergence of suffixes such as *-gate* from *Watergate*, cf. [Joseph 1998]. Although in new formations such as *Iran-Contragate* and *Monicagate* the last element is definitely related to the last syllable of *Watergate*, there is no reconstruction of a process in borrowing and presumably no reanalysis. *Watergate* first underwent a metonymical meaning extension from the name of a building to an event

happening there. Then one element of the extended meaning was associated with the last part of the form and used in other contexts. The element *water* in *Watergate*, however, is not a designation of the kind of affair *Watergate* refers to, so that the latter remains a simple entry. Items such as *Monicagate* are entered as complex lexemes, but they cannot be related to *Watergate* by a word formation rule. The relationship between *Watergate* and *Monicagate* is purely etymological.

4 Heuristic Principles

In the project "Word Formation as a Structuring Device of the English and Italian Lexicons: A Large-Scale Exploration" small teams are working on the two languages, specifying lexemes as entries by classifying them in terms of the WM rule databases for English and Italian. The first decision in the specification of an entry is whether the entry is simple or complex. A simple entry, i.e. a lexeme not resulting from a word formation rule, is assigned to an inflection class (IRule). The IRule generates the inflectional paradigm. A complex entry is assigned to a word formation rule (WFRule). The WFRule models the underlying word formation process and assigns the resulting entry to an IRule for generating the inflectional paradigm. The specification process is supported by a menu which prompts the lexicographer to select a WFRule, find the appropriate base lexeme(s) in the database, and select any affixes involved in the process. On the basis of this information, the system generates the resulting lexeme.

Explicit guidelines are necessary to ensure consistency in the way decisions are taken within each team and over time. The establishment of common guidelines for both languages makes it possible to use the resulting databases in comparative studies as well. In the case of the interaction of word formation and etymology, the main decision for which support is required is whether a particular entry is simple or complex. It has been our goal to describe the word formation relationships as they are synchronically experienced. This is not to say that diachronic considerations do not play a role, but they are only considered to the extent they influence the development of the vocabulary in future. In order to achieve this, the following heuristic principles were adopted:

- *Describe the object language*: Arguments for establishing a word formation relationship between two lexemes A and B can only be grounded in the object language.

- *Ignore the morphology of the source language*: If A and B are lexemes in the source language and B is derived from A, this can never by itself be a reason to derive B from A in the object language.

- *Maximize word formation in the object language*: If there is a choice between treating A and B as independently borrowed or as related by a word formation rule, the latter analysis is preferred.

The general operation of these principles can be illustrated on the basis of the examples of *conjectural* and *kindergarten*. In the case of *conjectural*, the English word pair is *conjecture* (A) and *conjectural* (B), corresponding to Latin *coniectura* (A) and *coniecturalis* (B). The object language in this case is English, so that the Latin derivational relationship cannot influence our analysis. As in this case there is a process which derives adjectives from nouns by suffixation with *-al*, the last principle tells us to derive *conjectural* from *conjecture* in this

way. In the case of *kindergarten*, German has the combination of *Kind* and *Garten* in the role of A, and *Kindergarten* as B. In English, however, there is no lexeme for A corresponding to *kindergarten* as B. The absence of a process in English means that there is no choice of the type referred to in the last principle, so that *kindergarten* in English is a simple entry.

5 Exemplification

Let us now look at some more challenging cases, illustrating the effects of the principles chosen in practice. As a first example we take the verb *inspect* and the nouns *inspector* and *inspection*. Before their first appearance in English, all three are attested (with minor formal divergence) in French, and their origin is Latin. According to the OED, the first to be borrowed was *inspection*, attested from the 14th century, the other two followed in the early 17th century. From a synchronic perspective, however, *inspector* and *inspection* are recognized as deverbal nouns involving the suffixes *-or* and *-ion*. These suffixes belong to the English word formation system. If the verb *inspect*, the deverbal noun-forming suffix *-ion*, and the noun *inspection* exist, it is no longer relevant whether the word formation process took place in English or in French (or Latin) or in which order the words were borrowed.

The verb *inspect* belongs to a class whose analysis has been a matter of some controversy in the literature. Thus, Aronoff [1976:11-14] proposes that verbs such as *refer, remit, resume, receive,* and *reduce* consist of a prefix and a base. His main argument is the regular allomorphy found in word formation processes. All of *remit, demit, commit, transmit, submit, admit,* and *permit* form nominalizations in *-mission* and adjectives in *-missive*. This is not a property of the phonological string, as shown by the behaviour of *vomit*, but it can be associated with the formative *-mit*. Similar alternations occur for *-sume* and *-sumpt-*, *-ceive* and *-cept-*, and *-duce* and *-duct-*. If *admit* is analysed as complex, there seems to be no reason not to analyse *inspect* as complex too, relating it to *respect* in the same way as *admit* is related to *transmit*. Such an analysis is also adopted by Lieber [1990].

The opposite view is represented by Marchand [1969:5f.]. Most recently, also Bauer [2001:108f.] rejects the analysis of items such as *transmit* into two morphological entities. He refers to a study by Marslen-Wilson et al. [1994] based on psycholinguistic experiments with lexical priming. For words such as *govern, governor,* and *government*, they investigated to what extent exposure to one of them cognitively activates the others. They found that there are two directions of activation, both from *govern* to *governor* and *government*, and from *governor* or *government* to *govern*. As for the relationship between *governor* and *government*, the findings are more equivocal. Marslen-Wilson et al. claim there is no priming effect, but Bauer refers to other studies in which a priming effect is found in such cases. There is no discussion whatsoever about the results for pairs such as *conceive* and *receive*. No priming is ever attested. These results can be transposed immediately to our example with *inspect*. They indicate that *inspection*, and *inspector* should be related to *inspect*, but that for *respect* no relationship to *inspect* is warranted.

On the basis of our principles, we cannot construct an argument for a morphological relationship between *inspect* and *respect*. If *inspect* is taken as B, there is no lexeme **spect*

in English to take the role of A. The fact that in Latin there is a verb *spectare* underlying *inspectare* is not relevant for the English analysis. A direct relationship between *inspect* as A and *respect* as B or the reverse is not possible because there is no English word formation process substituting *re-* for *in-* or the reverse. The fact that priming experiments support this lack of relationship suggest that the heuristic principles adopted lead to the desired conclusion. If no psycholinguistic relationship in the speaker's mind can be found, the relationship is unlikely to affect the future development of the vocabulary.

In the discussion of the relationship between *inspect* and *inspection*, the problem of the phonological change of the final stem consonant was disregarded. As WM is used for written forms and the orthographic representation hides the change, the modification can in fact be neglected. This raises the question of how to treatcases in which the phonological modification is also written. An example is the pair *pretend* and *pretension*. The question in this case is whether there is an English morphological process changing *d* to *s* in this context. Such a change is not a regular component of English phonology. In the particular morphological context of *-ion* suffixation, however, it is well attested. The suffixation process resulted from reconstruction on the basis of a large number of borrowings from Latin. The Latin suffix *-ion* (realized as *-io* in the nominative) regularly attached to what Aronoff [1994] calls the *t*-stem of Latin verbs. As Aronoff's term indicates, this stem usually ends in *-t* and this consonant is affected by a regular phonological process if followed by *-i*. Many verbs borrowed from Latin were borrowed in their stronger form, i.e. an intensified version, morphologically based on the *t*-stem as well. For instance, Latin *spectare* is the intensified form of *specere*. For the reconstruction of *-ion* suffixation, the basis was a large number of pairs of verbs, mostly ending in *-t* because they were intensive forms, and nouns, mostly ending in *-tion*, because *-ion* attached to the *t*-stem. Therefore the consonant alternation was naturally included as part of the process adding *-ion*.

In order to represent this aspect of word formation, the WM formalism has a mechanism of SRules ("spelling rules"). They can be stated on a general basis or individually. In the former case they are part of the rule database and apply automatically from the point of view of the lexicographer. An example is the change of *in-* to *im-* in the formation of *impossible*. Individual SRules are specified by the lexicographer for exceptional cases. As a rule, the influence of *-ion* on the final stem consonant can be ignored in writing, so that no general SRule is required. For pairs such as *pretend* and *pretension*, an individual SRule is used.

The fact that the SRule mechanism is available in the lexicographic specification phase immediately raises the question of how its use should be constrained. While it is obviously well applied in the consonant alternation in the pair *pretend* – *pretension*, there are no constraints in the formalism barring its application to the alternations found in *father* and *paternal*. Putting aside here the exact analysis of the suffix involved, let us concentrate on the relationship between the forms *father* and *pater*. They are representations of an Indoeuropean stem Beekes [1995:158] gives as *$ph_2t\check{Z}r$ (here h_2 is a laryngeal and $\check{Z}$ is a long vowel). The form *father* is the result of a number of changes in Germanic languages not shared by Latin. An SRule changing *father* into *pater* in a suffixation process would actually reverse these processes. This cannot be the intention of a morphological description. Fortunately, the principle of maximizing the morphology of the object language in no way

forces us to write such an SRule. The relationship between *father* and *paternal* is semantic and etymological, but there is no morphological process in English relating them. Whereas *father* is of Germanic origin, *paternal* is a 17th century borrowing from Late Latin. Therefore they are encoded in WM as unrelated simple entries.

In Italian, the corresponding case of *padre* and *paternale* shows one of the central differences between English and Italian in this context. Italian has Latin both in the role of ancestor language, reflected in *padre,* and as a source of borrowing of learned words, as in *paternale.* The gradually increasing distance between classical Latin and what is now Italian does not allow the recognition of rigid borderlines between synchronic and diachronic phonological relationships. The phonologically conditioned alternation between *pater* and *patr* is found in Latin, with the latter occurring in the accusative *patrem.* The only historical phonological change affecting the source of *padre* but not *paternale* is the change from *-t-* to *-d-.* As shown by Dante's use of *patre,* this change took place in the history of Italian, not in the transfer from Latin to Italian. Moreover, the relationship between Italian *padre* and *paternal* is experienced in a way the one between *father* and *paternal* in English is not. Therefore, the adjective *paternale* is encoded as related to *padre* in terms of word formation, with an SRule covering the difference in form.

6 Conclusion

Of the different processes involved in the extension of the lexicon, word formation processes are the only ones which can be formalized so as to anticipate the future development of the vocabulary. Encoding the existing lexicon in such a way that word formation rules are taken as the organizing principle offers a good basis for a well-founded description of these processes. In doing so, word formation should be distinguished from etymology. This is not always straightforward, because borrowing as a competing process interferes with word formation.

The approach adopted here consists in the establishment of a number of heuristic principles, accompanied by exemplification. The principles exclude relationships based merely on morphological processes in the source language, but do not exclude the reconstruction of such processes on the basis of a sufficiently large number of pairs of words borrowed. The borrowings giving rise to the reconstruction are subsequently reanalysed. A similar process is found in neo-classical word formation.

Exemplification and discussion of problem cases within the project resulted in intersubjective consistency in the specification of word formation relationships in the large Word Manager lexicon databases for English and Italian (over 40,000 entries in each language at the time of writing). The main difference between the two languages is the different role of Latin with respect to their development. The learned borrowings in Italian interact with the lexemes which came into the language through its descent from Latin, so that more processes can be reconstructed. This reconstruction is warranted because the processes play a role in the further development of the vocabulary.

Acknowledgement

References

[Aronoff 1976] Aronoff, Mark H. 1976. *Word Formation in Generative Grammar*, Cambridge, Mass., MIT Press.

[Aronoff 1994] Aronoff, Mark H. 1994. *Morphology by Itself: Stems and Inflectional Classes*, Cambridge, Mass., MIT Press.

[Bauer 2001] Bauer, Laurie 2001. *Morphological Productivity*, Cambridge, Cambridge University Press.

[Beekes 1995] Beekes, Robert S.P. 1995. *Comparative Indo-European Linguistics: An Introduction*, Amsterdam, Benjamins.

[Crystal 1995] Crystal, David 1995. *The Cambridge Encyclopedia of the English Language*, Cambridge, Cambridge University Press.

[Dworkin 1985] Dworkin, Steven N. 1985. *Etymology and Derivational Morphology: The Genesis of Old Spanish Denominal Adjectives in -ido*, TÝbingen, Niemeyer.

[Ten Hacken & Domenig 1996] ten Hacken, Pius & M. Domenig, 1996. Reusable Dictionaries for NLP: The Word Manager Approach, in: *Lexicology*, 2, De Gruyter, Berlin.

[Ten Hacken 1998] ten Hacken, Pius 1998. 'Word Formation in Electronic Dictionaries', *Dictionaries* 19, DSNA, Cleveland.

[Ten Hacken 1999] ten Hacken, Pius 1999. 'Two Perspectives on the Reusability of Lexical Resources', *McGill Working Papers in Linguistics* 14, McGill University, MontrŽal.

[Ten Hacken 2002] ten Hacken, Pius 2002. 'Word Formation and the Validation of Lexical Resources', in: *Proceedings of LREC 2002* - Language Resources & Evaluation Conference, Las Palmas, 27 May - 2 June 2002.

[Joseph 1998] Joseph, Brian D. 1998. 'Diachronic Morphology', in: Spencer, Andrew & Zwicky, Arnold M. (eds.) *The Handbook of Morphology*, Oxford, Blackwell.

[Lieber 1990] Lieber, Rochelle 1990. *On the Organization of the Lexicon*, New York, Garland.

[Marchand 1969] Marchand, Hans 1969. *The Categories and Types of Present-Day English Word-formation: A Synchronic-Diachronic Approach*, 2nd edition, MÝnchen, Beck.

[Van Marle 1985] van Marle, Jaap 1985. *On the paradigmatic dimension of morphological creativity*, Dordrecht, Foris.

[Marslen-Wilson et al. 1994] Marslen-Wilson, William; Komisarjevsky Tyler, Lorraine; Waksler, Rachel & Older, Lianne 1994. 'Morphology and Meaning in the English Mental Lexicon', *Psychological Review* 101, American Psychological Association, Washington, DC.

[Matthews 1974] Matthews, Peter H. 1974. *Morphology: An Introduction to the Theory of Word Structure*, Cambridge, Cambridge University Press.

[Pedrazzini & ten Hacken 1998] Pedrazzini, Sandro & ten Hacken, Pius 1998. 'Centralized Lexeme Management and Distributed Dictionary Use in Word Managerⁿ', in: Schršder, Bernhard; Lenders, Winfried; Hess, Wolfgang & Portele, Thomas (eds.) *Computers, Linguistics and Phonetics between Language and Speech, Proceedings of the 4th Conference on NLP, Konvens'98, Bonn, Germany*, Frankfurt am Main, Lang.

[Pedrazzini 1999] Pedrazzini, Sandro 1999. 'The Finite State Automata's Design Patterns', in: Champarnaud, Jean-Marc; Maurel, Denis & Ziadi, Djelloud (eds.) *Automata Implementation, Third International Workshop on Implementing Automata, WIA'98, Rouen, France*, Berlin, Springer.

[Petropoulou & ten Hacken 2002] Petropoulou, Evanthia & ten Hacken, Pius 2002. 'Neoclassical word formation in an electronic dictionary', in: *Proceedings of Euralex 2002*, K¿benhavn, 13-17 August 2002.

[Zappatore & ten Hacken 2000] Zappatore, Daniela & ten Hacken, Pius 2000. 'Word Manager and Banking Terminology: Industrial Application of a General System', in: Heid, Ulrich; Evert, Stefan; Lehmann, Egbert & Rohrer, Christian (eds.) *Proceedings of the Ninth Euralex International Congress, Euralex 2000*, Stuttgart, Institut fŸr Maschinelle †bersetzung, UniversitŠt Stuttgart.