A Dictionary Lookup Strategy for Translating Discontinuous Phrases

Michael Carl and Ecaterina Rascu

Institut für Angewandte Informationsforschung, Martin-Luther-Str. 14, D-66111 Saarbrücken, Germany

{carl | kati}@iai.uni-sb.de

Abstract

Translation of discontinuous phrases is a major challenge in Machine Translation. Within METIS-II we developed a dictionary lookup strategy by mapping the items of a dictionary entry on non-adjacent words in an input text. Mapping is controlled through so-called contextual rejection, i.e. inappropriate mappings are discarded if they fail to satisfy a predefined set of constraints. We present various dictionary preprocessing steps to transform the entries into a suitable and more effective format for lookup. Then we describe the dictionary matching of discontinuous phrases. We illustrate this process on German verbs with detachable prefixes and support verb constructions.

1 Introduction

This paper describes a dictionary lookup strategy for the German to English component of the METIS-II MT system. METIS-II[1] (Dologlou et al., 2003) is a research prototype of a hybrid statistical machine translation system, financed under the EU STREP programme. Hand-crafted bilingual dictionaries are used to transfer lexical tokens (words, multi-word units and phrases) into the target language. A large target language corpus is consulted to adjust the transferred token according to the target language syntax and/or to rank the generated translation candidates. Within METIS-II, a number of modules have been proposed and developed for re-ordering and ranking the translation candidates (Badia, Boleda, Melero, & Oliver, 2005; Carl, Rascu, & Schmidt, 2005; Dirix, Schuurman, & Vandeghinste, 2005; Markantonatou, Sofianopoulos, Spilioti, & Tambouratzis, 2005). The general layout of the architecture is shown in figure 1.

Translation of discontinuous phrases is a major challenge when translating from German into another language. In German there are not only detachable (verbal) prefixes that produce 'non-monotonic' translations (Turcato & Popowich, 2003) but also a large number of light and semantically weak verbs that participate in support verb constructions (SVCs). In such constructions, (e.g. *in Gefahr bringen = in danger bring = endanger*) the verb is semantically empty while the noun carries the information.

Homonymy and changing word order may further complicate the translation process. Detached prefixes and prepositions have the same form — but can be distinguished on the basis of their different syntactic position within the sentence. For instance the verb *bringen* can participate in a support verb construction together with the noun *Gefahr* or it can occur as a 'full' verb on its own with a different meaning (i.e. *bring*) even if the listed noun (i.e. *Gefahr*) co-occurs in the same sentence.

In addition, German has productive inflection, derivation and composition patterns that make the enumeration of all word forms or even of all different lemmas in a dictionary impossible.

There are thus at least two problems when matching a German sentence on a dictionary: i) find the 'right' translation entry

[1] http://www.ilsp.gr/metis2/

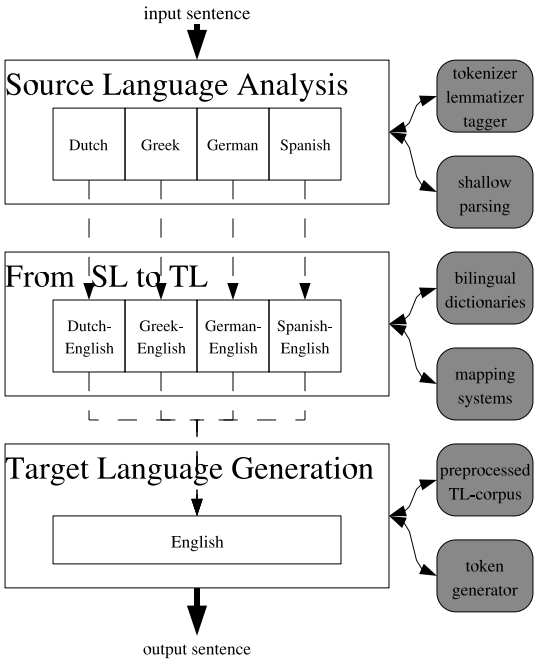

Figure 1: General System Flow and Resources Used in METIS-II

even if the words are differently inflected and and have different derivations and ii) find the right collocation of words even if they are non-adjacent and/or homonym.

The paper discusses strategies to overcome these problems by i) mapping the morphological structure of the words rather than their lemmas or surface forms and ii) by mapping syntactic variants to account for different realizations of a dictionary entry in a sentence.

While these mechanisms increase recall and coverage of the dictionary, there is also a need to control and reject matched dictionary entries if they do not satisfy the required word order and derivational or inflectional properties.

The paper proposes solutions to these mechanisms. We first present the German-English dictionary, outline its format and potentials. Then we describe a number of processing steps that transform the text version of the dictionary into a database by extracting features and representing them suitably for the matching process. We discuss the representation in the TL and find that the dictionary in itself is a segmentation device that produces structures which are neither isomorphic with the SL chunks nor do they represent a linguistically motivated TL structure. Section 4 describes dictionary lookup and the transfer of features into the target language. Section 5 describes contextual rejection for support verb constructions whereas section 6 presents a small experiment carried out to evaluate the lookup strategy.

2 The German English Dictionary

Our German-English dictionary contains more than 460.000 entries collected over the past 20 years. Dictionary entries are represented in the form of feature bundles surrounded by curly brackets as shown in examples (1) to (3). The German (*de*) or English (*en*) side of an entry represents either a single word or a continuous or discontinuous phrase in which words are separated by an underscore. Additional language specific classification information is coded in the attributes *mde* and *men* respectively.

(1) {*de=einsperren,mde={c=verb}, en=lock_⟨so.⟩_away,men={c=verb}*}.
(2) {*de=Anweisung_ausführen,mde={c=verb}, en=execute_statement, men={c=verb}*}.
(3) {*de=von_⟨etw.⟩_Kenntnis_nehmen,mde={c=verb},en=take_note_of,men={c=verb}*}.

Both language sides are independent. That is, a single word can translate into a single word, a phrase or a discontinuous phrase. The German verb *einsperren* for instance, translates into a discontinuous English verb *lock_⟨so.⟩_away*. Note that the entries are actually flat trees: while the words represent the leaves of the tree, the features *mde* and *men* are their mother nodes.

Discontinuous entries are coded as coherent strings in the dictionary but they need not appear as continuous word sequences in a sentence. In fact, they do not even need to be marked as discontinuous in the dictionary. For instance, while *einsperren* is coded in the dictionary as a continuous sequence of morphemes, it can actually appear as *sperren ... ein* in a German main clause, where *sperren* is the verb and *ein* its detached prefix. Meta-information en-

closed in pointed brackets serves to mark some properties of a compulsory sequence in the sentence. In example (3) the meta-information ⟨*etw.*⟩ denotes a dative NP to occur between the preposition *von* and the noun *Kenntnis*. Note also that dictionary entries do not require the same number of 'arguments' in their left and right sides. It is thus a matter of TL generation where the object denoted by ⟨*etw.*⟩ is realized.

The components of the entry *Anweisung ausführen* in (2) can be distributed in several ways in a German sentence. In a subordinate clause, the finite verb *ausführen* follows the object *Anweisung* and thus forms a SOV structure. Other constituents, as for instance adverbial modifiers or relative clauses may be inserted between the object and the verb as shown in (4). In the German main clause the verb precedes the object while detachable verb prefixes follow it. This requires a complete re-organisation of the entry since the order of the words in a sentence will be reversed as shown in (5): the object follows the verb *führen* while the prefix *aus* is the last element in the finite sentence.

(4) **Anweisung** *mehrfach hintereinander* **ausführen**

(5) **führen** *eine* **Anweisung** *mehrfach* **aus**

Thus, discontinuous phrases can occur in different permutations in a sentence. Nevertheless, only canonical forms should be included in a dictionary and special dictionary lookup and matching strategies are needed to account for such peculiarities. The following sections discuss the coding and preparation of the dictionary as well as the way entries are mapped on an input text within the present architecture.

3 Dictionary Preprocessing for German

The dictionary undergoes a number of preprocessing steps before the SL items in the dictionary can be mapped on a text. The source and the target language sides of the dictionary pass through a multi-layered fully automatic compilation step. For the SL side this involves:

1. Morphological analysis and lemmatisation of the 'leaves'
2. Checking internal consistency of the entries
3. Variant generation
4. Indexation of matching and consolidation features

The four attributes of the dictionary entries mentioned in section 2 (*de, mde, en, men*) are separated and further processed. For the German side (*de*), these steps are explained in the following sections.

3.1 Morphological Analysis of German Dictionary Entries

The *de* slots of the dictionary are morphological analyzed and lemmatised. The morphological analyzer MPRO provides the following output for the word *ausführen*[2]:

{*lu=ausführen,c=noun,ehead={nb=sg, case=acc;dat;nom,g=n},ls=aus_$führen*};
{*lu=ausfahren,c=verb,vtyp=fiv,nb=plu,per=1;3, tns=past,mode=subj,ls=aus_$fahren*};
{*lu=ausführen,c=verb,vtyp=fiv,nb=plu,per=1;3, tns=pres,ls=aus_$führen*};
{*lu=ausführen,c=verb,vtyp=inf,ls=aus_$führen*}.

A non-detachable prefix is marked with "$" detachable prefixes are marked with "_$".

The word *ausführen* is thus ambiguous with respect to its lemma, part of speech, and other information.

3.2 Consistency Checking

The mother node of an entry encodes the type of the entry. For instance, *mde={c=verb}* encodes a verbal entry and *mde={c=p}* encodes prepositions. Each type is associated with a set of morpho-syntactic patterns. For instance, an entry of type *mde={c=verb}* must end with an infinite verb while a preposition ends with

[2]The features are: lu – lemma, ehead – nominal inflection information (number, case, gender) ls – morphological structure.

a preposition. A dictionary entry is consistent iff a pattern associated to its type can be consolidated in the morphological analysis of its leaves.

As shown above, MPRO analyses the word *ausführen* as a noun, or a finite or non-finite verb. The finite verb can be derived from the verb *fahren* (*drive*) or *führen* (*lead*). Since the meta-information specifies that the dictionary entry is a verb (i.e. *mde*={*c*=*verb*}), we can eliminate all but one reading of the word.

We thus control whether MPRO analyses of the words (i.e. the leaves of the entry) are consistent with the type codings in the meta-information, detect errors and inconsistencies as well as classify and quantify the quality of the entries. In addition, we can disambiguate readings and filter those readings that are intended by the type (i.e. delete 3 readings of *ausführen*). On the other hand, by reanalyzing the dictionary entries each time a database is compiled, we are sure that the representations of the entries are consistent with the analysed words of an input text.

Entries that do not comply with the consistency criteria are marked and are manually corrected by a lexicographer.

3.3 Variant Generation

To increase the coverage of the dictionary, a number of variants are generated in a next step. A variant is essentially an additional translation relation that covers a different realization of a dictionary entry. For instance, we generate two variants for the verb *ausführen*. One variant is suited to match a main-clause verb in a non-compound tense while the other matches sub-ordinate clause patterns as explained in examples (4) and (5).

Generating lexical variants when compiling the dictionary will increase the size of the database. It has the advantage that the variation process can be better controlled. Moreover, complex variation phenomena can be better accounted for in a preprocessing step than on the fly variation since time consuming operations will not be carried out during runtime. While the number of indexes in the database increases only slightly, searching the augmented database is logarithmic. Producing variants on the fly at runtime directly affects computation time in a linear manner. (see (Carl, Rascu, Haller, & Langlais, 2004) for more details).

Currently the system deals with variation for nominal and verbal expressions. Essentially two types of variation are considered for nominal expressions: morpho-syntactic variation including compounding, coordination etc. and synonymy. Examples of morpho-syntactic variation are given in (6) to (9) and have been discussed in detail in (Carl, Haller, Horschmann, Maas, & Schütz, 2002; Carl et al., 2004):

(6) *Abfertigung des Gepäcks*
→ *Gepäckabfertigung*
(7) *Gepäckabfertigung*
→ *Abfertigung des Gepäcks*
(8) *Anzahl Mitarbeiter*
→ *Mitarbeiteranzahl*
(9) *Schreib- und Übersetzungsbüros*
→ *Schreibbüro und Übersetzungsbüro*

Variation patterns for verbal expressions are generated in a similar way. For instance, in (10) the main clause variant of the verb *ausführen* is generated by allocating separate nodes for the prefix (*aus*) and the verb (*führen*) . The verb *ausführen* has thus two instantiations in the compiled dictionary: a main clause and a subordinate clause variant.

(10) *ausführen* → *führen ... aus*
(11) *von_⟨etw.⟩_Kenntnis_nehmen*
→ *nehmen_von_⟨etw.⟩_Kenntnis*
(12) *von_⟨etw.⟩_Kenntnis_nehmen*
→ *davon_Kenntnis_nehmen*
(13) *von_⟨etw.⟩_Kenntnis_nehmen*
→ *nehmen_Kenntnis_davon*

Examples (11) to (13) show variants of a SVC. In the main clause variant of the support verb construction in (11), the finite verb precedes its nominal argument. The variant in (12) specifies that the PP *von_⟨etw.⟩* is replaced by the adverbial *davon*[3]. The main clause variant for the transformation illustrated in (12) is given in (13).

[3] *davon* is an anaphor and refers to a clause or phrase which is not part of the SVC.

Besides the variation phenomena described in this section some of the discontinuous structures may be further modified. Possible modifications are confirmed or eliminated through the so-called context rejection rules presented in section 4.3.

4 Dictionary lookup

There are in principle two different methods to map non-continuous phrases on a (necessarily continuous) dictionary entry: i) regroup those words in a sentence that belong to the discontinuous phrase into one unit so that they map the entry in the dictionary or ii) map the items of a dictionary entry on the non-adjacent words in the sentence and then discard wrong mappings.

The first method is only manageable if the required knowledge resources are available to re-group the distributed words into a coherent string or structure. That is, to join *aus* and *führen*, as in example (5) into one structure, one would have to know with high certainty that they belong together and form one meaning entity.

This approach was favored for instance within CAT2, a sideline of EUROTRA (e.g. (Haller, 1993; Streiter, 1996)) where even support verbs, prepositions and articles are 'dissolved' into features and appropriate translations are re-generated in the target language. The approach resembles interlingua MT which analyses the source language string into a structure that describes its content specification. The target language sentence is generated starting from these content values (Kuhn, 1994). A major problem in this approach is that the dictionary entries and input sentences have to be consistent not only in terms of their morphological representations but also with respect to their syntactic structure and content specifications, which is even more difficult to ensure and control.

We follow the latter approach, where scattered words of (discontinuous) phrases are mapped onto dictionary entries. The method is better suited for 'shallow' approaches since multiple overlapping entries can be found and decisions on the best match can be postponed to later stages of processing. Dictionary lookup is probably best understood as an instance of abductive reasoning[4]: dictionary entries are considered facts; matching a sentence on the dictionary is a process of proving or disapproving the presence of these facts in the sentence. From the perspective of the sentence it is investigated which translation relations fit best the whole of the sentence. If no exact matching entries are found, those translation relations are kept and processed further that provide the best explanation for the observations in the sentence.

A major challenge for dictionary lookup is how to deal with incompleteness. Even the most complete dictionary is likely to contain translation relations only for a subset of words in a language. Particularly for German, due to inflection, derivation and compounding, one cannot even expect all lemmas to be enumerated in the dictionary.

During dictionary compilation, all those features that are important to prove the presence of the lexical fact in the input sentence are made explicit and available (cf. section 3). In addition, a set of rules are used at runtime to consolidate or disapprove entries by examining the context of the matched items.

Dictionary lookup works in three successive steps which are explained in the following subsections:

1. morpheme retrieval
2. lexical delta
3. contextual rejection

4.1 Morpheme Retrieval

In this step, all entries are retrieved that map any continuous or discontinuous sequence of morphemes in the input sentence. For instance, assume entry (10) is in a dictionary together with its main clause variant. Any occurrence of *führen* and *aus* in one sentence, irrespectively of their distance, trigger retrieval of this entry. However, for a sentence containing the morphemes *aus* and

[4]Abduction is often defined as inference to the best explanation, see discussion on `http://www.cs.bris.ac.uk/~flach/ECAI96/ECAI96report.html`

führen in another order than in any of the variants, this match would be rejected.

(Langlais & Gotti, 2006) restrict the number of matched entries by allowing the parts of a discontinuous phrase to occur only within a window of six words. While such a strategy retrieves entries even with intervening modifiers, a window technique, whatsoever its size, is not suited for many of the German discontinuous phrases. The distribution of the words are rather guided by syntactic constraints and the properties of the phrase itself, as explained in section 5.

In contrast to most statistical techniques, we match morphemes instead of surface forms or lemmas. This may lead to a large number of retrieved entries which all share the same base morphemes such as participles: *ausgeführt*, adjectives: *ausführende*, adverbs *ausführbar* nouns: *Ausführen*, *Ausführer*, *Ausgeführte*, *Ausführung* etc.

It is not in every case desirable that all these entries are retrieved from the dictionary (see some more possible matches in table 1). We therefore compute a lexical delta between the matched word in the sentence and the dictionary entry and keep only the closest matches.

4.2 Lexical Delta

There are several reasons for not having all possible derivations for all base morphemes in a dictionary. First, for German this is almost impossible since derivation is highly productive (cf. Table 1). Second, and probably more important, in many cases their translations are identical. Thus, possible English translations of the verb *ausführen* are *to execute* and *to export* etc. The German participle *ausgeführt* would be translated as *executed* and *exported* respectively while nominal derivations will be *execution* and *exportation* and so on. Depending on the abilities to generate inflected and derived forms in the target language, it is thus conceivable to cluster several word forms in one dictionary entry and transfer inflection and derivation information though a second chanel.

If we observe the morphological analysis of the various derivations and word forms generated from *ausführen*, it resembles a factoring (see table 1). Even though all words are derived from the same morphemes *aus* and *führen* some are closer derivations than others. We compute the difference for each pair of words that is derived from the same morpheme and only retain translation relations from close neighbours. The underlying assumption is that similar derivations and PoS tags will also have similar translations into the TL.

We consider a dictionary entry consolidated, iff its lemma, PoS and gender match the analysed word in a sentence. For instance, the two word forms *ausführen* and *ausführe* differ only in their inflectional information (i.e. number and person, see table 1). The form *ausführe* found in a sentence would thus fully consolidate a dictionary entry *ausführen*. Additional inflection information (number, person, degree of adjectives, etc.) will be transferred into the target language where a token generator produces the correct word form (Carl, Schmidt, & Schütz, 2005).

Starting from the verbal form in the first line of table 1, increasingly different and more remote word forms which are also likely to have different translations are listed on each line. Depending on which word forms are contained in the dictionary and which word forms are found in the text, only translations of the most similar derivations are retained.

4.3 Contextual Rejection

Contextual rejection is a mechanism which discards retrieved entries if the context of the matched word does nor satisfy predefined constraints. It relies on a set of KURD rules which have access to the mother node of the entry, the dictionary entry including the meta-information, and the information of the analysed sentence. The context of a matched entry are words that fill the gap between two (or more) morphemes/words of a discontinuous entry and those to its left or right. For instance, a matching nominal multi-word entry is rejected if the components of the entry are

Word	Lemma	PoS	Derivation	Degree	Inflection
ausführen	ausführen	verb	—	—	per=1;3,tns=pres
ausführe	ausführen	verb	—	—	per=1,nb=sg,tns=pres
ausgeführt	ausführen	verb	ptc2	—	
ausgeführtes	ausgeführt	adj	ptc2	base	nb=sg,case=n;a
ausgeführteren	ausgeführt	adj	ptc2	comp	
Ausführender	ausführend	adj	ptc1	base	nb=sg;plu
ausführendem	ausführend	adj	ptc1	base	nb=sg,case=dat
ausführend	ausführend	adv	ptc1	—	
ausführbar	ausführbar	adv	∼bar	base	
ausführbarer	ausführbar	adv;adj	∼bar	comp	
Ausführer	ausführer	noun	∼er	—	nb=sg
Ausführung	ausführung	noun	∼ung	—	nb=sg
Ausführbarkeit	ausführbarkeit	noun	∼bar∼heit	—	nb=sg

Table 1: Analysed words and their lemmas which share the same base morphemes *aus_$führen*

not all within the same nominal chunk.

(14) {*de=Abbau_der_Ozonschicht, mde={c=noun},en=ozone_depletion, men={c=noun}*}.

(15) **Abbau der** *arktischen* **Ozonschicht**

Assume the dictionary entry (14). The head of the term *Ozonschicht* can be modified in the matched sentence, for instance by adjectives as in example (15). While we would like to validate the entry despite the intervening adjective *arctic*, we want to reject the entry if the words co-occur 'by accident' in the same sentence and are actually unrelated. This would be the case if the words occurred in different noun phrases.

Similarly, a detached prefix must occur at the end of the main-clause in which the finite verb is realized. A matched verbal entry would be rejected if this is not the case. In the next section we develop more examples for this mechanism.

5 Contextual Rejection for Support Verb Constructions

Contextual rejection is a very powerful mechanism for dealing with support verb constructions and with idioms which allow for a range of combinations.

A SVC consists of a nominal part, usually a noun in the accusative or a prepositional phrase, and a semantically weak support verb. Some SVCs are lexicalised and consequently relatively fixed in their internal structure, whereas others are much more flexible. Flexibility is dealt with at various levels. Fixed predicative nominal constituents are usually marked as such during morphological analysis. Thus, the prepositional phrases in the SVCs *in Anspruch nehmen* and *zum Ausdruck bringen* are analysed as single nodes with a complex lexical structure, i.e. {*c=w,sc=pred,ls=in_Anspruch*} and {*c=w,sc=pred,ls=zum_Ausdruck*}.

Other SVCs allow for more variation and a need emerges to control and consolidate the retrieval of the entries by looking at the context of the matched words. Contextual rejection works on the basis of a set of constraints, which specify which combinations involving SVCs are allowed or disallowed. One such constraint specifies that all elements in a SVC must appear within one clause, either main or subordinate. Therefore, the matched entry *zur Kenntnis nehmen* is rejected for input (16). Rejected matches are underlined, successful ones are in bold face.

(16) *Zu diesem Schluss kommen zwei Berichte, von denen der Bundesrat gestern Kenntnis genommen hat.*

In case meta-information, as provided in pointed brackets in example (2), is not available, default constraints are applied. More specific constraints are used for SVCs in which the prepositional phrases are used as predicative nouns and the contracted

determinate article is encoded together with the preposition (e.g. *zum* or *zur*), the noun may only be modified through adjective phrases. Consequently the SVC in (17) is validated whereas the ones in (18a) and (19a) are correctly rejected.

(17) *Die Ereignisse wurden* **zur** *allgemeinen* **Kenntnis genommen**.

(18 a) *Der Stadtrat hat von diesem Bericht zu wenig Kenntnis genommen.*

(19 a) *Der Stadtrat hat von der Zustimmung zur Einheitsgemeinde Kenntnis genommen.*

(18b) *Der Stadtrat hat* **von** *diesem Bericht zu wenig* **Kenntnis genommen**.

(19b) *Der Stadtrat hat* **von** *der Zustimmung zur Einheitsgemeinde* **Kenntnis genommen**.

For entries like (3) containing meta-information, constraints verify if the specifications are met for the respective slot or not. In this case the dative NPs occurring in (18b) and (19b) after the preposition *von* are confirmed and the proposed match involving SVC (3) are validated.

Rejection of matched entries through contextual constraints is thus used to validate the 'surviving' matches.

6 Evaluation of the Dictionary Lookup Strategy

The dictionary lookup strategy outlined in the previous sections was tested on two small test texts *T1* and *T2*. *T1* contains 101 sentences with a total of 111 annotated examples of verbs with detachable prefixes. *T2* contains 103 sentences, each with one annotated support verb construction. The examples in the test texts were taken from the *Deutscher Wortschatz* corpus[5] (Quasthoff, 1998). All annotated verbs and SVCs are to be found in our German-English dictionary.

We computed precision as the ratio of the *correct* recovered items over all recovered items (*correct* / (*correct* + *noise*)), and recall as the ratio of the *correct* recovered items over all annotated items (*correct* /(*correct* + *misses*)). The f-score is (*2 * precision * recall*) / (*precision* + *recall*). An item is considered correctly retrieved if it was correctly matched on the corresponding dictionary entry and was subsequently successfully validated through contextual rejection rules. Table 2 summarises the results of the experiment.

	T1	*T2*
f-score	0.93	0.98
precision	1.00	1.00
recall	0.86	0.96
correct	96	99
noise	0	0
misses	15	4

Table 2: Evaluation

High precision could be achieved for looking up and validating both verbs with detachable prefixes and support verb constructions. However, while ensuring high precision, too restrictive contextual rejection rules lead to decrease in recall. The items that were correctly looked up in the dictionary but wrongly rejected by the contextual rejection rules were counted as *misses*. There were 4 such cases for *T1* and 1 for *T2*. In case of *T1*, failure in matching the appropriate dictionary entries is the main cause for the *misses* whereas in *T2*, the *misses* are mainly due to uncovered variants of the respective dictionary entries (cf. section 3.3). One such example shown in (20) is the unimplemented variant detected for dictionary entry (3).

(20) *von_⟨etw.⟩_Kenntnis_nehmen*
→ *Kenntnis_davon_nehmen*

7 Conclusion and Future Work

The paper presents a generic dictionary lookup strategy which is able to map discontinuous dictionary entries on shallow processed input text. A number of control mechanisms are discussed. These mechanisms are triggered through the information

[5] www.wortschatz.uni-leipzig.de

provided by the entry and control the appropriateness of the matched entry. In the first implementation of the system we achieve figures for f-score between 0.93 and 0.97.

References

Badia, T., Boleda, G., Melero, M., & Oliver, A. (2005). An n-gram approach to exploiting monolingual corpus for MT. In *Proceedings of second ebmt workshop.* Phuket, Thailand.

Carl, M., Haller, J., Horschmann, C., Maas, D., & Schütz, J. (2002). The TETRIS Terminology Tool. *TAL, Structuration de terminologie*, *43*(1).

Carl, M., Rascu, E., Haller, J., & Langlais, P. (2004). Abducing Term Variant Translations in Aligned Texts. *Terminology*, *10*(1), 103–133.

Carl, M., Rascu, E., & Schmidt, P. (2005). Using template grammars for shake & bake paraphrasing. In *Proceedings of eamt 2005.*

Carl, M., Schmidt, P., & Schütz, J. (2005). Reversible Template-based Shake & Bake Generation. In *Proceedings of second ebmt workshop.* Phuket, Thailand.

Dirix, P., Schuurman, I., & Vandeghinste, V. (2005). METIS-II: Example-based machine translation using monolingual corpora - System description . In *Proceedings of second ebmt workshop.* Phuket, Thailand.

Dologlou, Y., Markantonatou, S., Tambouratzis, G., Yannoutsou, O., Fourla, A., & Ioannou, N. (2003). Using Monolingual Corpora for Statistical Machine Translation: The METIS System. In *Proceedings of the eamtclaw 03: Controlled language translation.*

Haller, J. (1993). CAT2, Vom Forschungssystem zum präindustriellen Prototyp . In H. P. Pütz & J. Haller (Eds.), *Sprachtechnologie: Methoden, werkzeuge, perspektiven* (p. 282-303). Hildesheim: Georg Olms AG. (URL: http://www.iai.uni-sb.de/en/cat-docs.html)

Kuhn, J. (1994). *Die Behandlung von Funktionsverbgefügen in einem HPSG-basierten Übersetzungsansatz* (Verbmobil Bericht 66). Universität Stuttgart.

Langlais, P., & Gotti, E. (2006). EBMT by Tree-Phrasing. *Machine Translation, To appear*(1-2).

Markantonatou, S., Sofianopoulos, S., Spilioti, V., & Tambouratzis, Y. (2005). Monolingual Corpus-based MT using Chunks. In *Proceedings of second ebmt workshop.* Phuket, Thailand.

Quasthoff, U. (1998). Projekt Deutscher Wortschatz. In G. Heyer & C. Wolff (Eds.), *Linguistik und neue medien.* DUV.

Streiter, O. (1996). *Linguistic modeling for multilingual machine translation.* Aachen: Shaker Verlag.

Turcato, D., & Popowich, F. (2003). What is Example-Based Machine Translation. In *Recent advances in example-based machine translation.*