Michael Carl

METIS-II: Low-Resource MT for German to English

1 Abstract

METIS-II was a EU-FET MT project running from October 2004 to September 2007, which aimed at translating free text input without resorting to parallel corpora. The idea was to use 'basic' linguistic tools and representations and to link them with patterns and statistics from the monolingual target-language corpus. The METIS-II project has four partners, translating from their 'home' languages Greek, Dutch, German, and Spanish into English.

The paper outlines the basic ideas of the project, their implementation, the resources used, and the results obtained. It emphazises on the German implementation.

2 Introduction

Starting in October 2004, METIS-II was the continuation of METIS-I (IST-2001-32775) Dologlou et al. (2003). Like METIS-I, METIS-II aims at translating free text input by taking advantage of a combination of statistical, pattern-matching and rule-based methods. The METIS-II project has four partners, each translating from their 'home' languages Greek, Dutch, German, and Spanish into English.

The following goals and premises were defined for the project:

1. use 'basic' NLP tools and resources,
2. use bilingual hand-made dictionaries,
3. use a monolingual target-language corpus,
4. use translation units within the sentence boundary,
5. allow different tag sets for SL and TL possible,

Crucially, parallel corpora are not required, and their usage was excluded within METIS-II. The rationale behind this was to develop prototypes of MT systems which would be suitable to translate 'small languages', i.e. language pairs for which parallel texts are difficult to come by. A basic set of NLP tools is nonetheless required for these languages, albeit very basic. The availability of the monolingual target language corpus, from which statistical language models are computed, makes METIS-II a data-driven MT system. These facts set METIS-II apart from mainstream SMT/EBMT systems.

With these goals and requirements, a number of implementations are possible. The METIS-II partners decided therefore to test and compare various implementations of the ideas, which will be outlined in this paper.

Hence, METIS-II consists of a number of modules which can be investigated horizontally, from source language to target language, or vertically, dividing the task into source-language analysis, lexical transfer, target language word-order generation and word-token generation. While the development of the four horizontal translation directions are to a large extent free-standing and independent efforts of the respective METIS-II partners, the consortium has also developed an exchange and interface format to communicate intermediate (i.e. vertical) processing results between the different parallel modules METIS-II (2006, 2007).

In this paper we aim at presenting METIS-II from a vertical and from a horizontal perspective. We discuss each of the parallel processing steps for all language modules involved, thereby showing their common and diverging characteristics.

The project has used a broad set of tools for source text analysis that were available or else easily obtainable by the partners. The Spanish analysis module experiments on using as few linguistic resources as possible - essentially only a lemmatizer and PoS tagger. The Dutch module adds a shallow parser to detect phrases and clauses while the German module includes also "topological" information. The Greek module seeks a more complete syntactic analysis of input.

The Spanish module uses only a bilingual dictionary that had been extracted from a printed Spanish-English dictionary. The Dutch-English dictionary was also compiled from external sources and the Greek-English dictionary was compiled from preexisting machine-readable dictionaries and augmented manually by the most frequent entries from the Hellenic National Corpus. The German-English dictionary is the largest of all the reported sizes and has been collected from unnamed sources over a long period of time. It covers words and both continuous and discontinuous phrases. Unlike other dictionaries, the German dictionary is preprocessed before use essentially through morphological analysis and generation of variants.

Section 6 describes the main resources used for generation and section 7 explains the way(s) how translations are generated in METIS-II. METIS-II follows a "generation-heavy" approach Habash (2004), where most of the hard translation issues are addressed during the generation phase.

The basic resource for generation are target language models, which are extracted from a huge target language corpus (the BNC) and which assist in selecting — and in some cases also in generating the word order of — the best translations. In this respect, the METIS-II core approach resembles Whitelock's (1991; 1992) 'shake-and-bake' method where the "target texts are constructed from a bag of TL basic expressions, whose elements are derived from the analysis of the source text and a set of equivalences of basic expressions"(Whitelock, 1991, p:1). However, while Whitelock uses logical and semantic constraints for 'baking' a target text from the basic expressions, METIS-II relies on statistical and pattern-based language models extracted from the target corpus to consolidate and verify target sentences.

Section 6 shows how the target-language corpus was preprocessed and how language models were conceptualized and extracted from the corpus. These models are built in idiosyncratic ways, with significant differences across language pairs. The Spanish module uses sequences of lemma/tag to validate insertions, deletions and permutations of words, the Greek and Dutch modules consolidate TL word order based on patterns and templates and the German module uses statistical n-grams.

Section 7 deals with the actual translation, the "decoding" of the source language. The overall translation method in METIS-II is creating a set of possible translation solutions and then using statistical methods to find the most probable translations. The language models play a crucial role in the selection process. Section 9 provides a detailed comparison of the differences and similarities across these modules.

Section 8 presents an evaluation of the translation systems using two test sets, the test suites used during development and a EUROPARL fragment, using BLEU, NIST and TER. Results for each language pair, using a well consolidated system such as Systran, are used as topline reference measure to gauge METIS-II results.

3 Background of METIS-II Implementations

In this section we briefly describe the basic ideas behind the implementations of the four translation directions. A linguistically minimal approach is favoured by the Spanish module, while the other modules employ a shallow parser to detect phrases and clauses. The Dutch and Greek modules assume some kind of structural isomorphism of phrases and clauses between the source and the target language, while the German module employs flat re-ordering rules.

3.1 Spanish to English

The approach followed by the Spanish-to-English METIS-II system strives to use as little linguistic resources as possible. The motivation in this case is not the lack of resources for processing Spanish but the desire to experiment in the leanest possible conditions, so that our findings can be applied to other, possibly smaller languages with fewer resources available. Consistently with this purpose, the preprocessing of the Spanish input requires only a tool able to lemmatize and assign morphological tags to each word of the sentence. The Spanish sentence is thus tokenized, tagged and lemmatized, but it is not chunked or analyzed in terms of constituency.

3.2 Dutch to English

For the Dutch-to-English translation pair was chosen an approach that requires a number of tools in order to perform a shallow source language analysis: a tagger, a lemmatizer, and a shallow parser (including a clause detector). We required the target-language corpus to be preprocessed with the same means, so equivalent tools for the target language are needed off line(Vandeghinste (2008)).

3.3 Greek to English

What is crucial within the Greek-to-English METIS-II approach is the notion of pattern, that is, phrasal segments that serve as the basis for modelling both the source (SL) and the target (TL) languages. The patterns roughly correspond to phrasal constituents of a varying size and type, ranging from clauses to sub-clausal level patterns (chunks and contained tokens). This approach, because it reflects the recursive character of natural language is expected to assist more effectively the translation process. Besides, even within the Statistical Machine Translation paradigm that strictly aimed to avoid using phrasal segments, the potential beneficial role of phrase-based models has now been recognized (Carpuat and Wu (2007)).

3.4 German to English

The German METIS-II architecture uses rule-based techniques to generate a graph of partial translation hypotheses and employs statistical techniques to rank the best translation(s) in their context. Word tokens are generated for the n-best translations.

The core idea is similar to Brown and Frederking (1995) who use a statistical English Language Model to combine partial translations produced by three symbolic MT systems. In contrast to their approach, we build the search graph with flat re-ordering rules.

The re-ordering rules generate an acyclic AND/OR graph which allows for compact representation of many different translations. A beam search algorithm tries to find most likely paths in the AND/OR graph. A similar idea for generation was suggested by Langkilde and Knight (1998) who use 2-gram language models to find the best path in a word lattice. Unlike a usual statistical decoder (Germann et al. (2001); Koehn (2004)), our ranker, hence, does not modify the graph and it does not generate additional paths which are not already contained in the graph.

4 Morphological processing

Each of the source languages modules in METIS-II has their individual preprocessing and SL analysis tools which are described in this section. In line with the requirements and philosophy of the project, all language modules use a lemmatizer and PoS tagger to process the source language input. In addition Dutch and Greek use a shallow parser to detect phrases and clauses and German recognizes topological fields. Besides the source language analysis, we have also implemented a reversible lemmatizer for the target language (English) which was used throughout for generation in METIS-II.

lemma	#	PoS	chunks	clauses
{lu=das,	wnrr=1,	c=w,sc=art,	phr=np;subjF,	cls=hs;vf}
,{lu=haus,	wnrr=2,	c=noun,	phr=np;subj,	cls=hs;vf}
,{lu=werden,	wnra=3,	c=verb,vt=fiv,	phr=vg_fiv,	cls=hs;lk}
,{lu=von,	wnrr=4,	c=w,sc=p,	phr=np;nosubjF,	cls=hs;mf}
,{lu=Hans,	wnrr=5,	c=noun,	phr=np;nosubj,	cls=hs;mf}
,{lu=kaufen,	wnra=6,	c=verb,vt=ptc2,	phr=vg_ptc,	cls=hs;rk}

Table 1: Analysis for the German sentence “Das Haus wurde von Hans gekauft” (The house was purchased from Hans).

The German source-language analysis produces a flat sequence of feature bundles which contain chunking and topological information of the sentence Müller (2004). An example of the German analysis is given in table 1.

Among other things, the analysis comprises of a unique word number, the lemma and part-of-speech of the word, as well as morphological and syntactic information. It also contains chunking and topological information. The parser produces a linguistically motivated, flat macro structure of German sentences, as coded by the `cls` feature.

Within the METIS-II project, we have implemented a reversible lemmatizer for English (Carl et al. (2005)) which reads CLAWS5-tagged words and generates a lemma together with two additional features indicating the orthographic properties (O) and the index of the inflection rule (IR). The IR-index serves to memorize the inflection rule which was applied to generate the lemma. Lemmatization rules are used to strip off or modify regular inflection suffixes from word tokens. Table 2 plots two lemmatization examples. A lemmatization lexicon is used for the irregular cases.

TAG	token	⇔	lemma	TAG_O_IR	IR	suffix mapping		
VVG	sniffing	⇔	sniff	VVG_l_1	1	`ffing`	↔	`ff`
VVG	DRESSING	⇔	dress	VVG_c_3	3	`ssing`	↔	`ss`

Table 2: Left: input and output of lemmatization and token-generation, Right: corresponding bi-directional inflection rule which can be used for lemmatization and for token generation.

The lemmatizer uses a single table of 128 lemmatization rules (two of which are shown on the right side in table 2). Each rule specifies the removal or replacement of an ending, conditionally on the TAG of the word and its suffix. Lemmatization and token generation is 100% reversible: a token set {token,TAG} is equivalent to a lemma set {lemma,TAG,O,IR} and both sets can be transformed into each other without loss of information, by reversing the lemmatization rule.

However, during token generation, we usually want to produce word forms from incomplete lemma sets {lemma,TAG}, where the inflection rule *IR* is not known. To generate an educated guess which IR would produce the desired word form, we have counted for each lemma suffix the inflection rules which generated the lemma. A word form would then be generated from a lemma by looking at the ending of the lemma and by applying the most likely reversed inflection rule. With slightly more than 20,000 lemma suffixes the reversible lemmatizer achieves a precision of more that 99.5%. In order to achieve this precision we had to add a few additional tags to the original CLAWS5 tagset, and then re-tagged the BNC[1] with the enhanced tagset. Table 2 plots two lemmatization examples.

5 Bilingual Dictionary

Apart from the resources required for the monolingual source language analysis, there are two other types of resources that were used in METIS-II: a bilingual transfer dictionary and the monolingual target-language corpus. For Spanish, Dutch and Greek the dictionary was compiled from external resources and adapted to the needs of METIS-II.

The German-English METIS-II dictionary contains more than 629,000 entries collected over the past 20 years. In its editable form, dictionary entries are represented as full forms and both language sides are independent. That is, a single word can translate into a single word, a phrase or a discontinuous phrase as in table 3. The German verb *einsperren* for instance, translates into a discontinuous English verb *lock ⟨so.⟩ away*. Entries are coded as flat trees: while the word(s) of the entries represent the leaves of the tree, the features *DE* and *EN* in table 3 are their 'mother nodes', which provide information about the type of the entry.

German	*DE*	English	*EN*
einsperren	verb	lock ⟨so.⟩ away	verb
Anweisung ausführen	verb	execute statement	verb
von ⟨etw.⟩ Kenntnis nehmen	verb	take note of	verb

Table 3: Examples from the German-to-English dictionary

The dictionary undergoes a number of preprocessing steps before the entries can be mapped on a German lemmatized and analysed sentence. 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]Section 6 gives more information on this corpus.

5.1 Morphological analysis and lemmatization of the 'leaves'

With the lacking context of words in a dictionary, the morphological analyser MPROMaas (1996) provides the following ambiguous readings for the word *ausführen*.

lemma	PoS	agreement	morph. structure
ausführen	noun	sg, acc;dat;nom, neut	aus_$führen
ausführen	verb, fin	plu, 1;3, pres	aus_$führen
ausführen	verb, inf		aus_$führen
ausfahren	verb, fin	plu, 1;3, past, subj	aus_$fahren

The symbol '_$' marks the detachable prefix *aus*, and thus illustrates the structure of the word. These readings are then disambiguated and filtered based on the type of the entry.

5.2 Checking internal consistency of the entries

By means of a set of patterns we control whether the analyses of the words (i.e. the leaves of the entry, as in the table above) are consistent with its type. A dictionary entry is consistent if at least one of its readings can be consolidated by a pattern associated to its type; otherwise the entry will be marked obsolete. This process also disambiguates readings and filter those readings that are intended by its type (e.g keeping only the *verb,inf* reading of *ausführen*). The process makes sure that the representations of the entries are consistent with the analysed words of an input text.

5.3 Variant generation

Variants are generated to extend the coverage of the dictionary for nominal and verbal expressions. A variant is an additional translation relation that covers a different realization of a dictionary entry. The verb *ausführen*, for instance, matches a main-clause verb in a non-compositional tense while the variation *führen . . .aus* matches in a subordinate clause. For nominal expressions morpho-syntactic variation for compounding, as e.g.: *Abfertigung des Gepäcks* → *Gepäckabfertigung*, but also coordination, and synonyms are generated (Carl and Rascu (2006)).

6 Target Language Modelling

We have experimented with various ways to use the implicit knowledge encoded in the monolingual target language corpus, and generated different language models. All language models are based on the BNC[2]. The BNC is a tagged collection of texts making use of the CLAWS5 tagset which comprises roughly 70 different tags. As pointed out in

[2]The British National Corpus (BNC) consists of more than 100 million words in more than 6 million sentences `http://www.natcorp.ox.ac.uk/`

section 4, to ensure reversibility of the lemmatized forms we had to add a few tags to the tagset and re-tag the BNC accordingly. The re-tagged BNC was then lemmatized before building the language models. For target language modelling there were, thus, three types of information available: (i) the original word form, (ii) the lemma and (iii) the PoS tag of the words.

In the German-to-English module, we have generated statistical *n*-gram language models. The language models (LMs) were generated using the CMU language modelling toolkit[3] or SRILM toolkit. The functions provided with these toolkits were adapted and integrated into a beam search algorithm as described in section 7. We have experimented with the following parameters:

- number of sentences arbitrarily extracted from the BNC:
  - 100K, 1M, 2M and 5M
- different kinds of statistical language models:
  - token-based LM: using the surface word forms
  - lemma-based LM: using the lemmatized word forms
  - tag-based LM: using the CLAWS5 tags
  - lemma-tag co-occurrence statistics
- 3 and 4-gram for token and lemma LMs and 4 to 7-gram CLAWS5-tag LMs

7 Translating with METIS-II

In line with the different philosophies and the variety of resources, decoding works differently for each of the language pairs. This section illustrates how translations are actually produced for German ↔ English.

In the German-to-English approach, rule-based devices generate an acyclic AND/OR graph, which allows for compact representation of many different translations. A statistical beam-search tries to find the best translation in that graph. Starting from a SL sentence, the graph is constructed in three rule-based steps. The graph is then traversed and translations are ranked. Finally word tokens are generated for the *n*-best translations. The architecture consists of the following five steps:

7.1 German SL Analysis

The *Analyser* lemmatizes and morphologically analyses the SL sentence. It produces a (flat) grammatical analysis of the sentence, detecting phrases and clauses and potential subject candidates as described in section 4, table 1.

[3] This toolkit can be downloaded from `http://www.speech.cs.cmu.edu/SLM_info.html`

7.2 Dictionary Lookup

The analysed SL sentence is then matched on the transfer dictionary. The procedure retrieves ambiguous and/or overlapping entries and stores them in the graph. Matching proceeds on morphemes and lemmatized forms and suited to retrieve discontinuous entries, cf. section 5.

Due to the complexity of discontinuous matches, we only allow discontinuous matches for verbal and nominal entries. In Carl and Rascu (2006) we have described various strategies to reject matched entries if they do not obey a predefined set of criteria.

For verbal entries, various permutations of the words are possible, according to whether the entry occurs in a subordinate clause or in a main clause. We use the field and chunk annotation in the German analysis to validate and filter or reject the matched entries. These criteria are further developed in Anastasiou and Culo (2007) making use of the German topological fields.

To account for a maximum number of different contexts, the dictionary generates all translation hypotheses which are then filtered and graded by the *Ranker* in the context of the generated sentence.

7.3 Word-Order Generation

This step inserts, deletes, moves, and permutes items or chunks in the AND/OR graph according to the TL syntax by means of a rule-based device. The rules take into account phrase and clause segmentation of the SL language sentence as well as word grouping resulting from the dictionary lookup. The modifications in the graph are such that each path containes exactly once the translation(s) of all the words of the source language sentence.

As in the so-called "generation-heavy" translation (Habash (2004)), the rules produce numerous partial translation hypotheses. For our German-to-English module we have currently ca. 50 rules, which are described in more detail in Carl (2007). This "symbolic overgeneration" is then constrained by a statistical ranker making use of several statistical feature functions.

7.4 Ranking and Translation Selection

In this step, the AND/OR graph is traversed to find the most likely translations as a path through the graph. Ranking is a beam search algorithm which estimates each node in the path with a set of feature functions (Och and Ney (2002)) and keeps those target sentence $\hat{e}$ with the highest probability according to equation (1).

$$\hat{e} = \texttt{argmax} \sum_{n} \sum_{m} w_m h_m(\cdot) \tag{1}$$

In equation (1), h_m is a feature function and w_m is a weighing coefficient, while n is the number of non-overlapping translation units matching the SL sentence (including those inserted or deleted in the generation module). Given the rich annotation of our data, there are numerous possibilities for the selection of feature functions, some of which are described in section 6. In the METIS-II evaluations reported in sections 8 we compare different ways to compute translation units and their mapping into the target language.

7.4.1 Token Generation

This step (cf. section 4) generates surface word-forms from the lemmas and PoS tags.

8 Evaluation of METIS-II

The evaluation of METIS-II was performed on two test sets, one consisting of data that had been used throughout the project for development purposes and one consisting of unseen data gathered from a previously existing bilingual corpus (Vandeghinste et al. (2008)). To measure results we used BLEU (Papineni et al. (2002)), NIST (Doddington (2002)) and TER (Snover et al. (2006)). The first two metrics measure edit distance using n-grams, while TER (Translation Error Rate) measures the amount of editing that a human would have to perform to get the translation right.

Each language group constructed a development set consisting of 200 sentences, with material evenly distributed among four different categories: 56 sentences illustrating grammatical phenomena (defined by each site), 48 sentences from newspapers; 48 sentences from encyclopedia articles, or similar sources of non-specialized texts, which provides a homogeneous evaluation framework. We compared Metis translations of this set with Systran translations. Systran is a syntactic transfer, rule-based MT system that has been under development since 1968, with a huge amount of funding from companies and institutions and large development teams. It uses large repositories of rule sets, large dictionaries, full parsers, elaborated algorithmic principles, etc. METIS-II, on the other hand, has been built in 3 years within 4 university groups, as an exploratory effort to build a hybrid MT system with no parallel corpus. Its architecture and components have been subject to much experimentation during the process. It is therefore reassuring that its results, though clearly worse than those obtained with Systran, stand up to the comparison.

In table 4 we plot the results of the German-to-English METIS-II system in two different experimental settings.

In the first experiment (METIS-II_1), we used a basic set of generation rules (cf. section 7). In the second experiment (METIS-II_2), we further developed and refined some generation rules for handling adverbs and negation particles, such as ‘never’, ‘usually’,

	Development set			Europarl test set	
	METIS-II$_1$	METIS-II$_2$	Systran	METIS-II$_2$	Systran
BLEU	0.186	0.223	0.313	0.282	0.396
NIST	5.48	5.32	6.36	6.68	8.05
TER	—	—	—	55.97	42.93

Table 4: DE-EN results for METIS-II and Systran on the Development and the Europarl test set.

extraposition of prenominal adjectives (e.g., "der vom Baum gefallene Apfel" would become "The apple fallen from the tree"), and "um ... zu" constructions. In the ranker we used lemma language models with 3 and 4-grams and tag language models with 4, 5, 6, and 7-grams. We varied weights between 0.01 and 10 for each of the feature functions and kept the combination which provided the best results. This setting was also used to evaluate the Europarl test set. The public version of Systran (Babelfish), however, performs even better than our best setting.

	Europarl	development	difference
NL-EN	0.1925	0.2369	0.0444
DE-EN	0.2816	0.2231	-0.0585
EL-EN	0.1861	0.3661	0.1800
ES-EN	0.2784	0.2941	0.0157

Table 5: Cross-language results on the development and Europarl test set (BLEU).

Table 5 shows that ES-EN is the system that has the most stable performance across test sets, while EL-EN shows the greatest variation. The most surprising result is DE-EN's, which performs better on the Europarl corpus than on the development set. A partial explanation may be that DE-EN has used Europarl type of text to tune lexical weights. Also, the DE-EN development set was chosen to contain hard translation problems so that also Systran performs more poorly on it than on the Europarl test set.

9 Comparison of decoders

This section resumes and compares the characteristics of the METIS-II decoders by looking at how hypotheses about TL word order are generated and how the most likely translation is selected.

9.1 Greedy vs. exhaustive translation modelling

Spanish, Dutch and Greek follow an incremental, non-monotonic approach to 'shake-and-bake', where the target sentence is piece by piece constructed from portions of the 'bag of TL expressions' (Whitelock (1991)) and each portion is in itself locally validated through the target language model. In contrast, the German decoder first produces all possible translation hypotheses in a compact graph representation and then uses language models and a beam searcher to select the best translation as a path through the graph.

9.2 Algorithmic vs. rule-heuristic word re-ordering

The Dutch and German modules employ rules to generate hypotheses of possible TL word-order — particularly for long distance movements. Spanish and Greek chose an algorithmic way to permute TL expressions. The latter approach has potentials of making the systems more language independent, while it is hard to correctly produce discontinuous translation in an algorithmic manner, which seems to be particularly important for Dutch and German.

9.3 Isomorphism vs. local changes

Dutch and Greek assume structure-isomorphism of phrases and clauses in the source and target language, while Spanish and German rely on local re-arrangements of the TL expressions. The former method requires a synchronization of the source- and target language resources, while for the latter, in principle, SL and TL resources may be processed and prepared independently.

9.4 SL vs. TL information for word order hypotheses

Permutation and re-arrangement of TL expressions for the German module is based exclusively on SL information from which these expressions were derived, while for Spanish TL word order hypotheses are based only on the TL information of the expressions. Due to the isomorphism assumption, the Dutch and Greek modules hypothesize TL word order based to some extend on the correlation of SL and TL information.

9.5 Top-down vs. bottom-up vs. flat re-ordering

The Greek module generates translations top-down by applying first larger, more abstract clause pattern models and then establishing the correct word order within each chunk. The Dutch module proceeds bottom-up incrementally consolidating word order from lower level phrases to higher level phrases. Spanish and German use flat re-ordering rules.

10 Conclusions

The paper reports on the underlying ideas, implementation and results of the EU-FET MT project METIS-II running from October 2004 to September 2007. METIS-II aimed at translating free text input using basic linguistic resources and a monolingual target language corpus.

With only a limited amount of work (about 12 man years) we have developed four language pairs, Dutch, German, Greek and Spanish into English. While results of METIS-II are not as good as a well-established MT system such as Systran, which we have chosen as topline reference, they can be considered of an acceptable quality. The paper shows that METIS-II provides a solid framework that can be easily adapted to new language pairs, that can be tuned to particular domains, and that can be upgraded with additional resources as they become available.

The paper describes the language processing tools and bilingual dictionaries of METIS-II which rely on shallow linguistic representations. Within METIS-II we have developed and explored various innovative language models and the paper points out how the models are exploited during translation. While we also give a comparative evaluation of the modules, we feel it is too early to draw ultimate conclusions on the best parameter settings.

We view METIS-II in the bigger context of self-learning systems that learn to translate from textual resources. Instead of learning relations between surface word forms, we maintain that the learned parameters must include linguistic properties of words and sentences for the system to tackle the hard problems of machine translation. Appropriate adaptive and dynamic representation of these parameters together with suitable reasoning mechanisms will ultimately help overcome the shortcomings of today's SMT systems. METIS-II has explored some of the possible avenues, and pointed to further directions that can be followed.

References

Anastasiou, D. and Culo, O. (2007). Using Topological Information for detecting idiomatic verb phrases in German. In *Proceedings of the Conference on Practical Applications in Language and Computers (PALC)*, pages 49–58, Lodz, Poland.

Brown, R. and Frederking, R. (1995). Applying statistical English language modelling to symbolic machine translation. In *Proceedings of the 6th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI)*, pages 221–239, Leuven, Belgium.

Carl, M. (2007). METIS-II: The German to English MT System. In *Proceedings of the 11th Machine Translation Summit*, Copenhagen, Denmark.

Carl, M. and Rascu, E. (2006). A dictionary lookup strategy for translating discontinuous phrases. In *Proceedings of the European Association for Machine Translation*, pages 49–58, Oslo, Norway.

Carl, M., Schmidt, P., and Schütz, J. (2005). Reversibl1e Template-based Shake & Bake Generation. In *Proceedings of the Example-Based Machine Translation Workshop held in conjunction with the 10th Machine Translation Summit*, pages 17–26, Phuket, Thailand.

Carpuat, M. and Wu, D. (2007). How Phrase Sense Disambiguation outperforms Word Sense Disambiguation for Statistical Machine Translation. In *Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI-07)*, pages 43–52, Skövde, Sweden.

Doddington, G. (2002). Automatic Evaluation of Machine Translation Quality using N-gram Co-occurrence Statistics. In *Proceedings of the second Human Language Technologies Conference (HLT-02)*, pages 128–132, San Diego.

Dologlou, I., Markantonatou, S., Tambouratzis, G., Yannoutsou, O., Fourla, A., and Ioannou, N. (2003). Using monolingual corpora for statistical machine translation. In *Proceedings of EAMT/CLAW 2003*, pages 61–68, Dublin, Ireland.

Germann, U., Jahr, M., Knight, K., Marcu, D., and Yamada, K. (2001). Fast Decoding and Optimal Decoding for Machine Translation. In *Proceedings of the 39th ACL and 10th Conference of the European Chapter*, pages 228–235, Toulouse, France.

Habash, N. (2004). The use of a structural n-gram language model in generation-heavy hybrid machine translation. In *Proceeding 3rd International Conference on Natural Language Generation (INLG '04), volume 3123 of LNAI, Springer*, pages 61–69.

Koehn, P. (2004). Pharaoh: a Beam Search Decoder for Phrase-Based Statistical Machine Translation Models. In *Proceedings of AMTA, the Association for Machine Translation in the Americas*, pages 115–124, Washington, DC, USA.

Langkilde, I. and Knight, K. (1998). The Practical Value of n-grams in generation. In *In Proceedings of the 9th International Natural Language Workshop (INLG '98)*, Niagara-on-the-Lake, Ontario.

Maas, H.-D. (1996). MPRO - Ein System zur Analyse und Synthese deutscher Wörter. In Hausser, R., editor, *Linguistische Verifikation, Sprache und Information*. Max Niemeyer Verlag, Tübingen.

METIS-II (2006). Validation/Evaluation framework. Public Report, D5.1, European Commission, FP6-IST-003768, Brussels. `http://www.ilsp.gr/metis2/files/Metis2_D5.1.pdf` [25.Aug.2008].

METIS-II (2007). Validation & Fine-Tuning Results for the first Prototype. Public Report, D5.2, European Commission, FP6-IST-003768, Brussels. `http://www.ilsp.gr/metis2/files/Metis2_D5.2.pdf`[25.Aug.2008].

Müller, F. H. (2004). *Stylebook for the Tübingen Partially Parsed Corpus of Written German (TÜPP-D/Z)*. `http://www.sfb441.uni-tuebingen.de/a1/pub.html`[25.Aug.2008].

Och, F. J. and Ney, H. (2002). Discriminative Training and Maximum Entropy Models for Statistical Machine Translation. In *Proceedings of the 40th annual ACL Conference*, pages 295–302, Philadelphia, PA.

Papineni, K., Roukos, S., Ward, T., and Zhu, W. (2002). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th ACL*, pages 311–318.

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., and Makhoul, J. (2006). A study of translation edit rate with targeted human annotation. In *Proceedings of Association for Machine Translation in the Americas (AMTA 2006)*, pages 223–231.

Vandeghinste, V. (2008). A Hybrid Modular Machine Translation System. Phd thesis, Netherlands Graduate School of Linguistics.

Vandeghinste, V., Dirix, P., Schuurman, I., Markantonatou, S., Sofianopoulos, S., Vassiliou, M., Yannoutsou, O., Badia, T., Melero, M., Boleda, G., Carl, M., and Schmidt, P. (2008). Evaluation of a Machine Translation System for Low Resource Languages: METIS-II. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC)*, page 96, Marrakech, Morocco.

Whitelock, P. (1991). *Shake-and-Bake Translation*. Unpublished Draft.

Whitelock, P. (1992). Shake-and-Bake Translation. In *Proceedings of the COLING92*.