

Identification of Idioms by Machine Translation: a Hybrid Research System vs. Three Commercial Systems

Dimitra Anastasiou
Institut für Angewandte Informationsforschung (IAI)
Martin Luther Str. 14, 66111 Saarbrücken, German
dimitra@d-anastasiou.com

Abstract

We compare three commercial Machine Translation (MT) systems, Power Translator Pro¹, SYSTRAN², and T1 Langenscheidt³, with the research hybrid, statistical and rule-based system, METIS-II, with respect to identification of idioms. Firstly, we make a distinction between continuous (adjacent constituents) and discontinuous idioms (non-adjacent constituents). Secondly, we describe our idiom resources within METIS-II, the system's identification process, and we evaluate the results with simple techniques. From the translation outputs of the commercial systems we deduce that they cannot identify discontinuous idioms. We prove that, within METIS-II, the identification of discontinuous idioms is feasible, even with low resources.

1 Introduction

Wehrli (1998) points out in speaking about some current commercial translation systems that „a simple glance at some of the current commercial translation systems shows that none of them can be said to handle multi-word expressions in an appropriate fashion“. Furthermore, the manual of Power Translator Pro explicitly warns the users to avoid inputting sentences containing idioms if they expect high-quality translation.

Section 2 discusses a basic distinction between continuous and discontinuous idioms on the basis of their syntactic realization. In section 3 we introduce the research system, METIS-II, and refer to its idiom identification process. Then, we evaluate this process arriving at the result that METIS-II can translate successfully both continuous and discontinuous idioms. In section 4 we provide some general information about the commercial systems and give their translation outputs of input sentences containing an idiom.

2 Realization of Continuous and Discontinuous Idioms

The distinction between continuous and discontinuous idioms is noteworthy. The idioms are called continuous when their constituents are adjacent to each other (see Example 1). By contrast, when an alien element is inserted between two idiom's constituents, the idioms are called discontinuous (see Example 2). The alien elements in Example 2 are an adverb, *oft* (often), and a prepositional phrase (PP), *wegen Stress* (due to stress). A subordinate clause could be inserted too. The kind of elements which can split the adjacency of the idiom's constituents is outside the scope of this paper. More information can be found in (Anastasiou & Carl, 2008).

As for the frequency of the two realizations of idioms, we measured their distribution in our German corpus of 486 sentences containing an idiom and we found that 73,8% is covered by continuous and 26,2% by discontinuous idioms.

- (1) *Niemand will **auf die Nase fallen**.*
*Nobody wants to **come a cropper**.*
- (2) *Er **fällt oft wegen Stress auf die Nase**.*
*He often **comes a cropper** due to stress.*

¹ http://www.lec.com/listProductFamily.asp?product_family=Power-Translator-Pro

² <http://www.systranet.com/systran/net>

³ http://www.langenscheidt.de/katalog/reihe_langenscheidt_t_volltextuebersetzer_version__625_0.html

3 Hybrid Statistical and Rule-based MT System METIS-II

In the following sections we provide information about our idiom resources, describe the idiom identification and translation process of METIS-II, and evaluate the results.

3.1 METIS-II Resources

METIS-II is considered as a hybrid system since it combines statistical tools and linguistic rules. It has Dutch, German, Greek and Spanish as source languages (SL), and British English as target language (TL). It uses the British National Corpus (BNC) as well as resources, which are language-specific for both SL and TL, such as bilingual dictionaries, tokenizer, part-of-speech (PoS) tagger, chunker, lemmatizer, and manually constructed matching rules. The matching rules „transfer“ the structures from the SL to TL.

We experiment with the German-English language pair and focus on idiom processing. The low resources we use for the idiom processing within METIS-II are the following:

- 1) Bilingual (German-English) idiom dictionary of 871 entries;
- 2) Monolingual (German) corpus of 486 sentences containing an idiom;
- 3) Syntactic rules according to the German topological field model.

Our dictionary and corpus are described in detail in (Anastasiou & Carl, 2008).

3.1.1 Idiom Dictionary

As for the dictionary entries, 72% of the total amount are idiomatic verb phrases (iVPs). The German lemmas are verb-final (see Examples 3, 4). It is noteworthy that the English lemma-translation counterparts omit the infinitive particle *to*. It should also be noted that the language sides are independent, i.e. a German idiomatic multi-word expression (MWE) can have an English one-word counterpart (Example 4), or vice versa.

	Lemmas:	PoS-tags
(3)	<i>auf die Nase fallen</i> <i>come a cropper</i>	iVP iVP
(4)	<i>ins Auge fassen</i> <i>envisage</i>	iVP verb

3.1.2 Idiom Corpus

As far as our German idiom corpus is concerned, it is assembled from four different resources:

- 1) A subset of the German-English Europarl corpus⁴ (Koehn, 2005);
- 2) Manually constructed sentences;
- 3) Sentences extracted from the Web;
- 4) Sentences extracted from the digital dictionary of the German language in the 20th century (DWDS⁵);

All sentences of the corpus contain an idiom. The corpus is manually annotated. There is one pair of angled brackets for continuous idioms (Example 5) and (at least) two pairs of brackets for discontinuous idioms (Example 6). In the first case, the brackets are placed at the beginning and the end of the whole continuous idiom (Example 5), whereas in the latter case, we set them at the beginning and the end of each idiom's constituent (Example 6).

- (5) *Niemand will <MWE id=1> auf die Nase fallen </MWE id=1>.*
Nobody wants to come a cropper.

⁴ We take only the German sentences into account.

⁵ DWDS stands for *das Digitale Wörterbuch der deutschen Sprache des 20. Jh.*: <http://www.dwds.de/>

- (6) *Er* <MWE id=1> **fällt** </MWE id=1> *oft wegen Stress*
 <MWE id=1> **auf die Nase** </MWE id=1>.
He often comes a cropper due to stress.

3.1.3 Syntactic Rules

As for the syntactic rules, they are based on the German topological field model. We now introduce the topological fields, into which the German main clause is divided:

- i) Pre-field (*Pf*);
- ii) Left bracket (*Lb*);
- iii) Middle field (*Mf*);
- iv) Post-field (*Pf*);
- v) Right bracket (*Rb*).

The verb forms occur in the left and right bracket. The finite verb occurs always in the left bracket. When a modal/auxiliary verb is placed in the left bracket, the infinite verb/participle form occurs in the right bracket. The pre-field is occupied mostly by one constituent, whereas many constituents are present in the middle field, even in a relatively free word order. One or more subclauses often populate the right bracket. More information about the German topological field model can be found in Drach (1964), DUDEN (1998), and Dürscheid (2000).

Our syntactic rules are based on this field model. Firstly, we observe sentences containing an idiom on the basis of the field model, i.e. we see where the idiom's verb and constituents can occur, and which alien elements can be inserted between them. The topological field model is actually taken into account for the discontinuous idioms (Examples 7a, 7b, 7c, 7d).

- (7a) *Er* [**fällt**]_{Lb} *oft wegen Stress* [**auf die Nase**]_{Mf}.
 (7b) *Er ist oft* [**auf die Nase**]_{Mf} *wegen Stress* [**gefallen**]_{Rb}.
 (7c) [**Auf die Nase**]_{Pf} *wegen Stress* [**fällt**]_{Lb} *er oft*.
 (7d) [**Auf die Nase**]_{Pf} *ist er wegen Stress oft* [**gefallen**]_{Rb}.

Secondly, we interpret this topological field model into rules, in order to apply to sentences containing an idiom. Within METIS-II we load these rules, the idiom dictionary and corpus, and then the system reads the sentences of the corpus, trying to „match“ the idiom contained in each sentence and translates it with reference to the dictionary. „Match“ in this context means „identify as an idiom“. The system starts by matching the first idiom's participant. Then, it continues matching each of the remaining idiom's participants, either when there are no gaps (continuous idioms) or with gaps (discontinuous idioms). In the latter case, we have to name the topological syntactic field in which the alien element occurs, in order for the system to know where it should ignore matching the alien element(s). There is only one rule needed for continuous idioms (see Rule 9) and four for discontinuous idioms. The rules for discontinuous idioms are as many as four, because the discontinuous patterns of an idiom in a sentence are four, shown in (7a, 7b, 7c, and 7d). We now provide only one rule for discontinuous idioms (Rule 8) which identifies the discontinuous idioms in their most common⁶ syntactic discontinuous pattern (7a):

- (8) *Discontinuous pattern LbMf*
 (a) **match=yes, field=Lb, PoS=verb,**
 (b) [**match=no, last word=no, field =Mf**]*,
 (c) **match=yes, last word =yes, field =Mf,**
 (d) : *identify as discontinuous iVP.*

The rule (8) contains three „constraints“, (a), (b), (c), and one command (d). The constraints (a) and (c) have the attribute-value pair, **match=yes**, to make clear to the system that they are parts of the idiom which should be „matched“. The elements of the middle field which do not belong to the idiom should be ignored (**match=no**). The rule interprets the following discontinuous syntactic structure:

- i) Constraint (a): The idiom's verb (**PoS=verb**) occurs in the left bracket (**field=Lb**);

⁶ The rules based on the remaining three discontinuous patterns (7b, 7c, 7d) are similar.

- ii) Constraint (b): Arbitrarily many words (asterisk*) which do not belong to the idiom can follow the verb in the middle field (*field =Mf*);
- iii) Constraint (c): The idiom's constituent (PP, NP, or combination NP-PP) follows in the middle field (*field =Mf*) and is the last matched word (*last word =yes*).

As for the continuous idioms, their constituents occur *en bloc*, thus the rule (9) is simple:

- (9) *Continuous pattern en bloc =*
- (a) *match=yes, last word=no,*
- (b) *match=yes, last word =yes*
: identify as continuous iVP.

The first constraint (a) shows that all idiom's parts should be matched. The system keeps on matching every following idiom's constituent until it finds the last one (*last word =yes*) which should be matched (*match=yes*) (see Constraint b).

In the following section we explain why these rules are helpful for the translation of idioms.

3.2 METIS-II Translation of Idioms

The translation process of idioms within METIS-II entails the three following steps:

- 1) SL analysis;
- 2) SL-to-TL matching;
- 3) TL generation.

The first step of the translation process is the shallow analysis of SL sequences, sentences, or texts. Tokenization, part-of-speech (PoS) tagging, lemmatization, and chunking are the SL analysis stages. In these stages, idioms are processed in the same way as all lexical units. These stages are performed automatically by means of the appropriate morphological tools of the institute IAI.

As far as the SL-to-TL matching is concerned, two resources are taken into consideration:

- i) The bilingual idiom dictionary;
- ii) The syntactic matching rules.

A basic prerequisite for the matching is that the idiom is stored in the bilingual idiom dictionary. METIS-II then takes into account the syntactic rules which are based on the topological field model (shown in 3.1.3). METIS-II is able to identify the idiom only by means of these rules. As aforementioned, the system reads the corpus or other input sentences, to which it matches the idiom (continuous or discontinuous), which is then "captured" and considered as a unity.

Also, two very important tools for SL-to-TL matching (not restricted to idiom matching) within METIS-II are *Expander* and *Ranker* (see Carl, 2007).

Expander is a rule-based software tool which reverses the allomorphy of the SL and TL at the lexical or structural level. For example, it adjusts the word order according to the grammatical rules of each language, e.g.:

- (10) *Sie haben mir reinen Wein eingeschenkt*
They have told me the plain truth

The input sentences and their translation outputs within METIS-II, after the *Expander's* function, are the following:

- (11) **input:** *Niemand will auf die Nase fallen.*
output: *Nobody wants to come a cropper.*
- (12) **input:** *Er fällt oft wegen Stress auf die Nase.*
output: *He often comes a cropper due to stress.*

Ranker is a statistical tool which functions similarly as a decoder used in statistical MT (SMT). It computes the most likely target sentences in a log-linear fashion (Och & Ney, 2002). In our case, the *Ranker* gives the following translation versions too, as there is another English lemma (*fall flat on one's face*) for the same German lemma (*auf die Nase fallen*):

- (13) *Nobody wants to fall flat on their face.*
- (14) *He often falls flat on his face due to stress.*

As for the TL generation, it is performed by using the BNC as a data set of examples. The BNC must be pre-processed at the same level as the input sentence. It is tokenized, tagged, lemmatized, and chunked (see Dirix et al., 2005). The BNC helps in disambiguating between various translation possibilities and it is used to retrieve the TL word order (Vandeghinste et al., 2005). The token generator has been described in Carl & Schütz (2005). We do not focus much on TL idiom generation, but on the matching of the idiom to the input. Since the idiom is matched, then it is also correctly generated in the TL.

3.3 METIS-II Evaluation

We evaluate the matching/identifying of idioms with simple techniques. When the idiom is identified, this is called *hit* and when it is not, this is called *miss*. When the idiomatic phrase is used in its literal meaning⁷, but the MT system identifies it as an idiom, this is called *noise*.

As for the evaluation techniques, we compute precision (Pr) as the ratio of the correct items, *hits*, over *hits* and *noise* items:

$$\text{Pr} = \frac{\text{hits}}{\text{hits} + \text{noise}}$$

and recall (Re) as the ratio of the correct items over *hits* and *misses* items:

$$\text{Re} = \frac{\text{hits}}{\text{hits} + \text{misses}}$$

The fscore is the result of the following formula:

$$\text{fscore} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

We provide two tables considering the evaluation figures of both continuous idioms (Table 1) and discontinuous idioms (Table 2) of the sentences which are included in the corpus data sets.

	Recall	Precision	fscore
Europarl Corpus	98,3%	96,8%	96,8%
Manually constructed and website examples	99%	96,2%	97,4%
DWDS	98,9%	96,7%	97,4%

Table 1. Evaluation figures for continuous idioms in the corpus data sets

	Recall	Precision	fscore
Europarl Corpus	88,2%	78,9%	83,2%
Manually constructed and website examples	95,7%	84,8%	88,8%
DWDS	92,5%	90,2%	90,6%

Table 2. Evaluation figures for discontinuous idioms in the corpus data sets

From the tables it can be seen that the evaluation figures for continuous idioms of all techniques are higher than these for the discontinuous idioms. This is attributed to the fact that discontinuous idioms are more difficult to identify due to their spread components through the sentence.

⁷ We exclude the cases where a sentence contains an „idiom“ which is literally used, the so-called „idiom counter examples“. We examine these cases separately. This subject is outside the scope of this paper.

4 Identification of Idioms by Three Commercial MT Systems

In this section we provide some information about the companies and the language pairs of three commercial MT systems. Also, we present the same input examples, as within the METIS-II, as well as the commercial systems' translation outputs. We tested only a small sample of 50 sentences, because, even after adding more examples, recall and precision were still lower than 5% and 10% respectively; we also could not advance the translation quality by writing rules.

4.1 Power Translator Pro

It was Globalink Inc., then Lernout and Hauspie Speech Products and now Language Engineering Company (LEC) that publishes Power Translator Pro.

Pro versions translate English texts into French, German, Italian, Portuguese, Spanish, and vice versa. The online translation software uses English as Interlingua and thus gives access to 21 languages and over 300 language pairs⁸.

We tested the version Power Translator Pro 7.0 (Binder, 2000). As far as the treatment of idioms is concerned, the shortcoming of Power Translator Pro is the attribution of the idiomatic phrases. There is not any attribution category for verb phrases, where most idioms belong to. The lemmas of the attribution category *verb* are limited to one word, thus users cannot attribute a multi-word idiomatic verb phrase (iVP) as *verb* and thus the system cannot identify any idiom (continuous or discontinuous) when the verb is inflected.

In Pro 7.0, idiom pairs are included in the main dictionary, as there is not a separate idiom dictionary available. Also, users have to add the idiom pairs to both language sides, which is time-consuming.

Below are the Pro's 7.0 translation outputs of two input sentences (15a, 15b), before and after adding the idiom entry, *auf die Nase fallen* – *come a cropper*, to the dictionary:

(15a) *Niemand will **auf die Nase fallen**.*
*before: Nobody wants **on the nose fall**.*
*after: Nobody wants **come a cropper**.*

(15b) *Er **fällt** oft wegen Stress **auf die Nase**.*
*before/after: It **falls** often because of stress **on the nose**.*

We see that, in both cases, when the idiom is not stored in the dictionary, Pro 7.0 literally translates the input sentences. Referring to (15a), after adding the idiom entry to the dictionary, the system identifies the idiom and translates it correctly. The output of the input sentence (15b), even after adding the idiom entry to the dictionary, is the same as before. Thus, Pro 7.0 cannot handle this case when the PP-component of the idiom, *auf die Nase*, is placed at the end of the sentence.

We will briefly discuss the continuous idiom of (15a). Its identification is actually not such a difficult task, as the idiom is realized in the sentence in the same form as in the dictionary entry, *auf die Nase fallen*. However, the whole sentence is not correctly translated, as the English infinitive construction of the idiom does not contain the particle *to*. If we add the infinitive particle *to* to the English idiom in the dictionary, then the whole sentence would be correctly translated. METIS-II overcomes this „problem“ without storing all English verbs with the particle *to*, which saves time. Also, there is the case when the idiom *auf die Nase fallen* occurs in German subordinate clause (Example 16), where the verb *fallen* is finite, 3rd plural and not infinitive. Then the English translation with the infinitive construction is wrong.

(16) *Er sagt, dass sie immer **auf die Nase fallen**.*
*He says that they always **come a cropper**.*
He says that they always **to come a cropper.*

The example (16) shows that the storing of the German idiomatic multi-word expression (MWE) with the English *to*-idiomatic MWE does not bring satisfying results.

⁸ This software is called “Translate DotNet” and is mainly used to translate documents, e-mails, webpages, blogs, and instant messages: <http://www.lec.com/translation-subscriptions.asp#dotnet>

4.2 SYSTRAN

The software of the company SYSTRAN is based on over four decades of expertise and is used by global corporations, Internet portals, and public agencies, such as the US Intelligence Community and the European Commission.

The MT system SYSTRAN has over 35 available language pairs and 20 vertical domains. It contains many subject-specific dictionaries and preserves the original layout of the user's documents.

We have tested a rather old version, the SYSTRAN Premium 4.0, so we are cautious about current advances of this system. As Power Translator Pro 7.0, it does not contain a separate idiom database. Below are Premium's 4.0 translation outputs of the same input sentences:

- (17a) *Niemand will **auf die Nase fallen**.*
*before: Nobody wants **on the nose fall**.*
*after: Nobody wants **come a cropper**.*

The output of the sentence (17a), after adding the idiom to the dictionary, is the same as the corresponding output of Power Translator Pro 7.0. Also, if we add to the dictionary *to come a cropper*, the translation would be the following: *Nobody wants to come a cropper*. We also input the same German sentence with the finite verb, as shown in (16) above, and there was the same output as Pro 7.0.

Moreover, we tested an input sentence, where the verb is inflected in second position (17b); the idiom is still continuous.

- (17b) *Er **fällt auf die Nase**.*
It comes a cropper.

Although the idiom is identified and correctly translated, there is a fundamental grammatical mistake, i.e. the German relative pronoun *er* (3rd person singular, masculine) is translated as *it*.

It is bizarre that in the case where the inflected verb (other than in 3rd plural) follows the PP, for example in a German subordinate clause, the idiom could not be identified and was literally translated (*auf die Nase fällt – on the nose falls*).

As for the example (17c) containing the discontinuous idiom, SYSTRAN 4.0, as Power Translator Pro 7.0, cannot identify the idiom, when alien elements occur between the idiom's parts.

- (17c) *Er **fällt oft wegen Stress auf die Nase***
*before/after: It often **falls** because of stress **on the nose***

4.3 T1 Langenscheidt

Langenscheidt is a privately-held publishing company in the field of language resource literature. The first Langenscheidt software for PC, T1 Standard, came out on the market in 1984. We tested the version T1 Professional. It translates between English – German, Spanish – German, and French – German. Dictionary entries consist of a lexeme, its translation equivalents, and its PoS with corresponding information. T1 Professional includes a Translation Memory (TM) and thus identifies exact and fuzzy matches, and newly translated sentences by using different colors in the screen output. For every source sentence it can present a choice of up to 3 different target sentences from the TM, if that many are found. T1 Professional contains in its TM two external modules, 5,000 phrases/sentences for business letters and a huge idiom collection of 71,000 pairs, which is derived from *Langenscheidt's Handwörterbuch Englisch*. Some idioms are complete in themselves, but most idioms consist of sentence fragments. Again the same input examples and the system's outputs follow:

- (18a) *Niemand will **auf die Nase fallen***
*before: Nobody wants **onto the nose fall***
*after: Nobody wants **onto the nose fall***

- (18b) *Er **fällt oft wegen Stress auf die Nase***
*before/after: It often **falls** because of stress **onto the nose***

As we more or less expected from the example (18b), the translation of the discontinuous idiom is not feasible; thus, T1 Professional translates it literally. What astonishes us is that the output of (18a), even after adding the idiom to the dictionary, does not change at all. The idiom has to appear with

exactly the same context as it is stored in the idiom collection, in order to be correctly translated. Thus, since translations are always done on complete sentences and T1's idioms occur with specific context in the collection, T1's idiom collection is not meant for automatic translation, but only for manual look-up, as Volk (1998) emphasizes. We also tried to add idiomatic VPs to the main lexicon, whose entries are subject to MT analysis, transfer and generation, but it accepts only one-word verbs.

5 Discussion

Among the small number of examples tested, METIS-II performs better than the commercial MT systems for three reasons:

- 1) METIS-II identifies and translates the continuous idiom correctly, not only in the morpho-syntactic form it is stored in the dictionary, as Pro and SYSTRAN do, but also in other forms, too. This is attained through the *Expander* tool, which takes into account the allomorphy of the languages. T1 cannot identify any continuous idioms, as the idiom module is used only for manual look-up.
- 2) The processing of discontinuous idioms is not feasible at all by any of the three commercial MT systems. METIS-II achieves by means of the syntactic matching rules almost more than 90% recall and 80% precision.
- 3) METIS-II identifies and accordingly translates correctly the sentences containing idioms, continuous or discontinuous, even if their verb is inflected or the idiom's participants undergo syntactic transformations.

6 Summary and Future Work

In this paper, we described the hybrid statistical and rule-based research system METIS-II. We referred to our idiom resources, the system's idiom translation process, and evaluation of results. The evaluation of METIS-II by using simple techniques, gave (almost) always more than 80% recall, precision, and fscore, for both continuous and discontinuous idioms.

We also had a short look at three commercial MT systems: Power Translator Pro, SYSTRAN, and T1 Langenscheidt, and how they translate sentences containing idioms. After adding the idiom entry to the dictionary, two of them gave satisfying translation outputs having identified the idiom. However, this is the case only for continuous idioms. In the case of the discontinuous idioms, all three commercial MT systems were incapable of identifying and accordingly translating the idiom.

In the future, we plan to increase the recall and precision within METIS-II, eliminating *noise* and *misses*. We will add more entries to the idiom dictionary and enrich our corpus with more sentences containing both continuous and mainly discontinuous idioms, in order to set high standards to face the difficult task of automated idiom matching and translation.

7 References

Anastasiou, D.; Carl, M., (2008), "A Lexicon of Shallow-typed German-English MW-Expressions and a German Corpus of MW-Expressions Annotated Sentences", in: *Towards a Shared Task for Multiword Expressions* (MWE 2008), Workshop at the LREC 2008, Marrakech, Morocco, 14-19.

Binder, M., (2000), "Review of L&H Power Translator Pro 7.0 software: Make Yourself Understood - Software Review – Evaluation", in: *Home Office Computing*.

Carl, M.; Schütz, J., (2005), "A Reversible Lemmatizer/Token-generator for English", in: *EBMT Workshop 2005, MT Summit X*, Phuket, Thailand.

Carl, M., (2007), "METIS-II: The German to English MT System", in: *MT Summit XI*, Copenhagen, Denmark, 65-72.

Dirix, P.; Schuurman, I.; Vandeghinste, V., (2005), "METIS: Example-Based Machine Translation Using Monolingual Corpora - System Description", in: *EBMT Workshop 2005, MT Summit X*, Phuket, Thailand, 43-50.

Dürscheid, C., (2000), *Syntax: Grundlagen und Theorien*, Wiesbaden.

Koehn, P., (2005), “Europarl: A Parallel Corpus for Statistical Machine Translation”, in: *MT Summit X*, Phuket, Thailand, 79-86.

Och, F.J.; Ney H., (2002), “Discriminative Training and Maximum Entropy Models for Statistical Machine Translation”, in: *Proceedings of the 40th Annual ACL Conference*, 295-302.

Vandeghinste V.; Dirix P.; Schuurman I., (2005), “Example-based Translation without Parallel Corpora: First experiments on a prototype”, in: *EBMT Workshop 2005, MT Summit X*, Phuket, Thailand, 135-142.

Volk, M., (1998), “The Automatic Translation of Idioms. Machine Translation vs. Translation Memory Systems”, in: Nico Weber (Ed.): *Machine Translation: Theory, Applications, and Evaluation. An assessment of the state of the art*. St. Augustin: Gardez-Verlag.

Wehrli, E., (1998), “Translating Idioms”, in: *17th COLING 1998*, Vol. 2, 1388-1392.