Improving MT Quality: Towards a Hybrid MT Architecture in the linguatec 'Personal Translator'

Gregor THURMAIR
linguatec
Gottfried-Keller-Str. 12
81245 Munich, Germany
g.thurmair@linguatec.de

Abstract

This paper reports on measures to improve the quality of MT systems, by using a hybrid system architecture which adds corpus-based and statistical components to an existing rule-based system backbone. The focus is on improving the accuracy of the dictionary resources.

1 Baseline

Although there have been significant improvements in the last period, the quality of machine translation is still an issue: The acceptance of the technology would improve significantly if the quality of translation would not prevent users from larger scale usage.

Research has not really focused on this issue; instead there were many attempts to start anew, hoping that a change in technology would lead to improved system quality; however, up to now, this has not proven to be the case.

This also holds for the latest developments in statistical machine translation (SMT); it is a development which tries to apply learning techniques on existing bilingual corpus material, and uses already existing translation material to compute translation equivalents for phrases, and complete sentences, on this basis (Knight and Koehn, 2003).

The question how the results compare to existing approaches has recently been studied in a bit more detail (Thurmair, 2005, with a comparison of German-to-English MT), with two main results:

- The overall quality of SMT is outperformed by existing rule-based MT systems.
- The overall quality of both approaches is not yet sufficient, as between 20 and 30% of the evaluated sentences were ranked as being unacceptable.

A closer look at the results shows that the main sources of errors in SMT (about 60%) are related to phenomena like Satzklammer and split verb constructions, non-standard constituent ordering, gapping etc., all of which could be rather easily described in a rule-based framework, while the main sources of errors in rule-based systems (again about 60%) consist of lexical issues, and wrong selection of lexical material, which in turn a corpus-based approach can easily avoid[1]. Examples are

a. Wrong VP treatment and constituent order in SMT:

(1) *Beim Anlegen einer Tabelle wird ein Schlüssel nach programmtechnischen Gesichtspunkten vergeben*[2].

when you create a table ***is*** *a key after a a technical* ***assign*** .

(2) *Die Parameter der Datenbanksystemprüfung können Sie in der R/3-Tabelle DBCHECKORA konfigurieren*[3].

the parameters of the database system check in R /3 configure table DBCHECKORA .

b. Wrong lexical selection in rule-based MT:

(1) *Der zweite Teilbaum beschreibt die Struktur des zu lesenden Datenbestandes*[4].

The second partial tree describes the structure of the ***data stock*** *to be read.*

(2) *Für die Verknüpfung mit Organisationsobjekten müssen Sie kein HR im Einsatz haben*[5].

You don't have to have any HR for the ***bonding*** *with organization objects in use*

As a consequences, a combination of linguistic and statistical techniques into a hybrid system

---

[1] The rest consists mainly in grammatical mistakes; here the problem is less that grammatical structures are not covered but rather that the wrong structures are selected in a given situation.

[2] Reference: *when you create a table, you assign a key according to technical criteria* .

[3] Reference: *configure the database system check parameters in the R/3 table DBCHECKORA*

[4] Reference: *the second sub - tree describes the structure of the data that will be read*

[5] Reference: *you do not have to have HR for the relationships with organizational objects* .

seems to be the most promising approach, and there are two main lines of action in this respect:

- Select a statistical / corpus-based system architecture, and improve it by making the input data more intelligent (using stemmers (Nießen and Ney, 2000), syntax trees, and syntactic relations (Och et al., 2003).
- Select a rule-based architecture and enrich it by corpus analysis results.

Experiences in trying the second alternative are presented in the following sections, starting from the weakest point in the evaluation, namely dictionary and translation selection.

2 Missing Dictionary Entries

Lexical mistakes, in general, result from two sources: No translations exist in the dictionary, and too many translations exist and a wrong one is selected.

2.1 Dictionary Gaps

The most straightforward case is dictionary gaps. In most existing MT dictionaries, surprising entries can be found: Experiments have shown (Dillinger, 2001) that MT dictionaries contain a significant amount of entries which have been spent coding effort on, but are nearly never used. In turn, surprising gaps can be detected in such dictionaries as well.

This situation seems to be due to a coding situation by which dictionary coders selected entries from paper dictionaries, and added them to the MT dictionary. But the situation is fundamentally different today: There is plenty of corpus material available, and coding can be done much more efficiently: Coding one term with frequency 1000 has the same overall effect as coding 1000 terms of frequency 1. Also, the coverage of a dictionary can be evaluated, in comparison to a corpus.

Therefore, corpus-based technologies of monolingual and bilingual term extraction are used to close dictionary gaps, based on frequency information[6]. Such tools combine linguistic and statistical information, and return term candidates with frequency information[7].

In the context of the 'Personal Translator', the missing entries with a frequency over 5000 were identified and added to the system dictionaries.

[6] The linguatec corpus for German and English , collected for the work presented here, consists of 700-800 million word forms each.

[7] cf. Piperidis et al., 1997; Thurmair, 2003

2.2 Special Language Terminology

There will still be a huge amount of terms not represented in MT dictionaries, mainly terminology for special domains. Most MT systems offer terminology import and coding tools, to allow customers to enter such terminology, which often are in the order of magnitude of several ten thousand terms.

Corpus-based techniques here are to be preferred to conventional dictionary entering:

- Experiments in the automotive sector show that even well-accepted special domain dictionaries, in a significant amount of cases, propose translations which sound plausible but are not at all used in the target language; this can easily be verified by searching for them in the internet.

  For de *Schlagfrequenz,* the following alternatives are found in different automotive terminological dictionaries:

Translation	Source Quality	Fre-quency
blow sequence	high	34
beat frequency	mid	7830
impact frequency	mid	0
hammer-blow sequence	mid	3
blow frequency	low	118
lateral runout frequency	high	0
stroke frequency	low	1160

Table 1: Translation proposals for *Schlagfrequenz*

- Often it is required to meet special user terminology requirements: E.g. if users allow for crosslingual searches on their web sites, terms must be translated in a user specific way; standard dictionary translations which do not reflect the terminology of this particular user lead to poor search results in cross-lingual retrieval.

Again, corpus-based work is required, even in cases where bilingual corpora are not available.

2.3 Proper Name Treatment

A third large source of unknown words are proper names. Although they form a considerable amount of the vocabulary, they are not considered to be of too much linguistic interest[8], and only

[8] In the 1800-page English Grammar of (Huddleston and Pullum, 2002, there are hardly 2 pages on proper names.

recent research (Babych and Hartley, 2003, Jimenez, 2001) into proper names shows the potential for quality improvement.

Proper names by definition cannot be stored in dictionaries, as there is a too large and ever growing amount of them. However, end users often are puzzled by wrong analyses of proper names in texts.

1. If proper names are not treated at all, what often happens are errors in parsing, as with other missing lexical elements in the input. It is not just parse failures but also wrong parses due to the fact that the system tries to cope with the lexical gap somehow and ends up with a wrong parse.

2. Such problems can be avoided if proper names are marked to be 'don't-translate' words, as is possible in some systems[9]. Then the proper names undergo some default system treatment (usually: noun with some default values for gender and number). However, this can be incorrect as proper names have syntactic properties: They inflect (like in Russian or German), they differ in number (plurale tantum like *the Hebrides, les Pyrénnées),* they take special prepositions, etc.; so more information is needed than just the default.

3. Therefore, a full named entity recognition component is required to improve the analysis, by providing information about constituency *(He robbed [the Bank of Scotland]* vs. *He robbed [the Bank] [of Scotland]}* and semantic type of proper names. At this stage, it turned out that the standard Named Entity categories must be refined, e.g. in cases of place names which need subtypes, as these subtypes have different linguistic properties (e.g.: country, city: *He lives in France / Paris;* lake, mountain: *He is on Lake Hudson / the Everest).*

Named Entity recognition often uses statistical or shallow parsing technology, and there are two options of integration into an MT system: running as some pre-processor, or being integrated into the full syntactic analysis. Full integration tends to be less robust (in case of parsing errors) but is easier able to cope with homographs (de *Peter Maurer war Maurer* -> en *Peter Maurer was a bricklayer)*[10] or gender issues (Frank et al., 2004). And there is another feature of Named Entity recognisers, which is coreference analysis, which influences conventional MT system structure: Coreference is a feature which is text based, and not sentence-based as most MT systems are.

In the following example, while the first occurrence of *Schneider* is recognised by contextual analysis, sentence-based MT systems fail to identify it in the third sentence, and therefore incorrectly translate the name there:

Das FDP-Mitglied Dr. Schneider lebt in München. Dort ist es heiß. Schneider ist der erste ausländische Politiker.

The FDP member Dr. Schneider lives in Munich. It is hot there. Tailor is the first foreign politician. (instead of: *Schneider is the first foreign politician).*

4. A special challenge consists in the translation of proper names. While it is a common mistake of MT systems to translate proper nouns (en *Mrs. Rice* -> de **Fr. Reis,* de *Hr. Fischer* -> en **Mr. Fisherman),* it is only true for person names that they must not be translated[11]. Dates usually must be translated to accommodate to the respective language's conventions. Places behave differently: some are translated (en *Ivory Coast* -> fr *Côte d'Ivoire* -> de *Elfenbeinküste),* others are not (e.g. *Santiago de Compostela).* Often such place names are put into the dictionary.

The target language proper names can also have different linguistic properties, which is relevant for generation: The *Désert du Thar* is masculine in French but *Thar Wüste* is feminine in German, and so is *Rhône* where even the lemma is identical in both languages. *Balkan* is singular in English but plural in Russian *(Балканы)* . For product names, the gender seems to be dependent on the 'base type': cars like *Renault* default to be masculine in German (derived from *der Wagen)* but feminine in French (derived from *la voiture);* determiner placement is language specific as well: fr *L'Italie* -> de __ *Italien* but fr *La Suisse* -> de *die Schweiz.*

While some of these cases can be handled by default assumptions[12], others are idiosyncratic and require a special resource to describe them.

5. The result of integrating a named entity component into an MT system (the linguatec Personal Translator) was an increase in translation

[9] Babych and Hartley, 2003, tested a named entity recogniser, and marked all entities as do-not-translate words. Systems with large dictionaries even sometimes follow the approach that all unknown words must be proper nouns; however they do not analyse their semantic type.

[10] There are also homographs of different types of proper nouns, e.g. *BMW* (company) vs. *BMW* (product) : *He drove a red BMW* vs. *He met a BMW spokesman.*

[11] albeit transliterated, which opens another problem when translating between Cyrillic or arabic and western scripts. Cf. Virga and Khudanpur, 2003

[12] or by corpus work, cf. Jiménez (2001)

quality for sentences containing proper names by about 30% on average[13]. The main improvements were:

- no erroneous translations of person
- names, esp. in coreference positions
- better contextual adaptations (correct preposition and determiner selection; and correct pronominalisation)
- better parses in a few cases (e.g. segmentation of dates)

Of course the overall quality gain for a given corpus depends on the number of sentences containing proper names, and will be higher in news text translation than e.g. in computer manuals.

3 Selecting the right one from many translation options

While the problem of missing dictionary entries seems to be reducible to a tolerable size (with the special challenge of proper names), the opposite problem is much more difficult to solve, which consists in an improper selection of a target term from a number of candidates. This problem aggravates with growing numbers of dictionary entries and increasing system intelligence. And this is what articles like "Have fun with MT" refer to: *Wortebene* is *word level* and not *word plane,* and *Stromunternehmen* is not a *river expedition* but an *electric power producer.*

The challenge consists on the selection of the proper translation in a given context[14].

3.1 Current Disambiguation Means

1. Global settings by users. Most systems provide options for subject area settings, for customer settings (to cover customer-specific terminology), for locales (to select for *truck*$_{US}$ or *lorry*$_{UK}$*),* for conservative vs. progressive spelling (to select for German *Gemse* vs. *Gämse),* and several other options.

These settings require user interaction, and a level of user skills which often is not available: Translations of search engine results do not ask users for subject area settings (although it could help a lot).

2. Linguistic context description. Such descriptions are coded in the dictionaries as transfer tests; they describe linguistic contexts which trigger special transfer selections:

See (gender = <feminine>) -> sea
See (gender = <masculine>) -> lake
ausführen (dir. object = <person> ->take out
ausführen (dir. object=<program>) -> *execute*

Such tests can be described as configurations of feature settings of underspecified tree structures[15]. Translation candidates are compared, in a specific order, to the input trees, and if their test configuration matches the input tree then this translation is selected.

Such a technique has two problems to solve:

- In case of parse failures, the structures with which the transfer candidates are compared are erroneous, so the comparison may fail, and a poorer translation is selected
- There are many cases of underspecification, i.e. the information which would trigger a transfer selection is not present: In cases where de *Bank* (plural *Bänke)* -> en *bench / benches* de *Bank* (plural *Banken)* -> en *bank / banks* but the sentence contains only a singular *(er steht vor der Bank),* then the system cannot apply the test, and randomly has to pick a translation, which can be wrong.

3.2 Automatic Subject Area Selection

To overcome the problem that not even the options which can be provided by the system (especially subject area selection) are used, a topic identification component has been added to the MT system, to guess to what subject area a text would be assigned[16].

1. There are two main lines of technology to build topic identification, or text classification, systems (cf. Jackson & Moulinier, 2002): Selecting classification features (usually words) from an example corpus by machine learning techniques, or using manually selected key words describing the respective topic. While the former crucially depends on the similarity of test and runtime text material, and therefore is less robust, the later depends on a careful selection of key words and tends to have a too small keyword basis.

In contexts where an MT system must translate internet material, the selection of a corpus which would be sufficiently similar to the texts to be translated at runtime is a very challenging task.

[13] a total of 1500 sentences was evaluated in three language directions, 15% of which contained proper names.

[14] Dictionaries for human lookup show an even wider range of translation possibilities than MT dictionaries, which requires an even more elaborate disambiguation mechanism.

[15] For examples cf. (Thurmair, 1990). An attempt to define a kind-of-standard representation for this has been made in OL1F, cf. (McCormick, 2001)

[16] A similar approach for disambiguation was followed in (Samiotou and Kranias, 2004)

2. In an MT environment, the most plausible option seems to use the system dictionary as a resource. However, although dictionaries are sensitive for subject area selection, they follow a different approach:

- They use subject area tags only in case disambiguation is needed; and for 1:1 translations a subject area assignment is not really necessary as the respective translation is selected anyway. For a classification, however, this is a drawback.
- Also, there are subject areas containing only very few terms (only the ones which need to be disambiguated), which is not suitable for good classification either.

So, MT dictionaries can be a good starting point, but more intelligence must be spent.

3. Therefore, a different approach was taken: A large text corpus was searched, starting with some seed terms (like "*sports football hockey racing"),* and the system returned the best correlated terms (both single and multiwords) to the seed words. From those the experts selected the ones which they believed to describe the topic best, and repeated this procedure. For each of the about 40 topics, between 200 and 700 terms per language were collected to describe it.

The classification is implemented in such a way that it gives the best (or the several best) subject areas if they match a given threshold, and gives no indication if it is not sure, and leave it to the users to decide.

4. The evaluation of the component was satisfactory: For a test corpus of several hundred documents, the correct subject area was returned in over 80% of the cases, and no false positives were returned. In a system which still allows users to select subject areas, and only provides a fallback in case they don't, this is quite acceptable.

However, correct subject area recognition is just a prerequisite for proper selection of translation alternatives by the MT system. It depends on the organisation of the dictionaries what use of this information the system can make, and how sensitive it is to subject area coding[17].

5. During the evaluation, it also turned out that a subject area code rather means that a given translation alternative is rather unlikely <u>outside</u> of a certain subject area, but it does not mean that <u>within</u> a subject area this translation is always correct. Many general vocabulary terms occur in specific domains both with their special and their general meaning, like in the automotive domain:

en *project* -> de *Restaurierungsobjekt vs. Projekt*
de *Übersetzung* -> en *translation vs. gear ratio*

As a result, a subject area test, even if the subject area is recognised correctly, is not the most reliable information for transfer selection, Additional means need to be used.

3.3 Neural Transfer

1. When observing human behaviour in transfer selection, it can be seen that people often refer to the context, in particular the conceptual context, to explain that „even in automotive domain, '*Übersetzung'* in the context of '*documentation', 'language'* and other such terms can only be '*translation*', not '*gear ratio'". The question is if such human behaviour can be modelled in an MT system to improve transfer selection using conceptual context.

The task is similar to word sense disambiguation, but applied not to abstract word senses (as in WordNet) but to concrete word senses as represented in different translations. It requires the identification of conceptual contexts which indicate a certain word sense, and consequently a certain translation of a term.

2. As a consequence, all dictionary entries with more than one translation were evaluated, and 'clear' cases like

en *teacher*$_{masculine}$ -> de *Lehrer*
en *teacher*$_{feminine}$ -> de *Lehrerin*

were eliminated. From the remaining set, several hundred candidates were selected for further analysis. Each of them was looked up in a standard dictionary to make sure that the most important readings of the term were represented.[18]

3. For each term, a corpus lookup was done, using the linguatec corpus, resulting in a couple of thousand contexts per term. Each of these contexts was assigned a reading of the word in question, to enable the formation of clusters of concepts for each reading. These clusters were then statistically analysed to identify the most distinctive terms for a given reading, and represented as a neural network. This is why we call this kind of transfer 'neural transfer'.

Examples of the effect are shown in the following texts, for different translations of *fan* and

[17] There are related problems to this, e.g. what to do if a text contains a term with just one translation which belongs to a subject area that has not been selected.

[18] Terms which happen to have the same translation for all readings (like en *cell* -> de *Zelle)* will be added in following versions.

of *coach* into German *Fan* vs. *Ventilator* and *Trainer* vs. *Bus,* respectively:

(1) en *The fans make noise. The whole club was already drunk when they came to the stadium to support their soccer heroes, although their coaches had to leave.*

de *Die Fans machen Lärm. Der ganze Klub war schon betrunken, als sie zum Stadium kamen, um ihre Fußballhelden zu unterstützen, obwohl ihre Trainer abfahren mussten.*

(2) en *The fans make noise. Their rotor does not distribute the air evenly, and the electric motor is not in full operation. All the coaches full of tourists were disappointed.*

de *Die Ventilatoren machen Lärm. Ihr Rotor verteilt die Luft nicht gleichmäßig, und der elektrische Motor ist nicht in vollem Betrieb. All die Busse voll von Touristen waren enttäuscht.*

The first sentence is translated differently in the two contexts, although both times identical in the source language.

4. The next task was the integration of the neural networks into the MT system. There are two challenges here:

- Like in proper name recognition, neural transfer needs more context than just a sentence; systems with a only sentence-based architecture create artificial limitations. So more context must be looked at than just one sentence.
- The neural transfer must be integrated into the transfer selection architecture of the MT systems. The existing heuristics, based on linguistic tests, are quite valid in many cases, and can be applied more easily than the neural transfers. But there are cases where such tests require extensions; and it turned out to be a special task to integrate the neural transfers into the test sequence of a word with multiple transfers.

5. Although only the first fraction of the dictionary has been treated this way, the evaluation is very positive.

To demonstrate the effect of this transfer, the following short story shows a state-of-the-art MT translation (from German):

All sheets have already decreased in our green plant. There a house with a narrow course and two small rooms stands. This was referred recently of a recent pair as its summer house. The pieces of furniture came on two vices which were hard loaded. The pair had a menu for dinner with five walks yesterday. First cooked lenses were served up. As dessert they had an almond court. The only one what disturbed the young people was the loud aviation and sums of the brakes, mosquitoes and hornets. A brake flew even into the face of the young woman, she then flew into her hair and she stopped on the pony next to the eyebrow there. Since its man has an artistic vein and very much fate, he took an admission fast. The photograph is supposed to appear on the next Sunday in the local leaf. The woman looked good, with her short rock, the high sales and the pink-coloured sweater with the irresistible cutting. The pair Sundays always comes into the church to the fair.

It seems to be difficult to build a coherent and meaningful semantic representation of this text without a reference to its source. This is due to the fact that many words which have different senses in the German source text are translated incorrectly. Applying neural transfer technology, the linguatec Personal Translator produces the following result (the translation of the German words which caused the mis-translations in the standard translation above are underlined):

All leaves have already fallen off in our park. A house with a narrow corridor and two little rooms stands there. This was recently obtained from a young couple as her vacation home. The furniture came on two trucks which was loaded heavily. The couple had a menu for dinner with five courses yesterday.
First cooked lentils were served up. As a dessert they had an almond dish. The only thing which disturbed young people was loud flying and sums of the horseflies, gnats and hornets. A horsefly flew even into the face of the young woman, she then flew into her hair and she stopped on the bangs next to the eyebrow there. Since her husband has an artistic bent and much dexterity, he made a photo fast. The photo shall be published in the local newspaper next Sunday. The woman looked good with her short skirt, the high heels and the pink-coloured jumper with the irresistible neckline. The couple on Sunday always comes into the church for the mass.

A preliminary evaluation was done as follows: In the German-to-English system, 30 concepts were randomly selected for the tests, and texts containing these concepts were downloaded from the internet, without reading disambiguation. The texts contain 165 occurrences of the test concepts. Of those, 162 (98%) were correctly translated, using neural transfer. Without neural transfer, just 92 (56%) were correct, so there is an improvement in quality of more than 40%.

Of course the real quality gain depends on the frequency of such concepts in the complete corpus.

It should be noted that correct translations on one place sometimes create follow-up improvements: Once de *Hörer* (en *listener* vs. *receiver)* is correctly disambiguated, the verb de *auflegen* (en *establish* vs. *put_down)* profits from this fact, and the translation is *put down the receiver* (instead of *establish the listener).*

6. This technology can be further improved not just by broadening its coverage. Several observations can be drawn from the existing work:

- There is a dependency between automatic subject area recognition and neural transfer as both techniques operate on the same data, and reinforce each other's calculations.
- Not all candidates are equally suitable for this kind of transfer; the ones for which the contexts can be clearly distinguished work best.
- Manual postediting / cleanup of the results of statistical clustering improves the translation results.
- There is room for optimisation by combining linguistic and neural transfers, e.g. neural transfer can be more precise for the readings of *"run"* if the transitive vs. the intransitive reading can first be distinguished by the syntactic analysis.

4 Conclusion

These examples show that the quality of MT systems is not yet at its limits; it also shows that it will develop in an evolutionary process rather than in a completely new technology.

The most promising approach seems to consist in hybrid system architectures, like the one chosen for the current version of 'Personal Translator', enriching rule-based approaches (which model the language competence) by corpus-based and statistical techniques (modelling the language performance aspects), as presented above.

5 Acknowledgements

The work presented here was mainly carried out by Vera Aleksić (who also composed the story), Alois Baumer and Thilo Will.

References

B. Babych, A. Hartley. 2003. *Improving Machine Translation Quality with Automatic Named Entity Recognition.* Proc. EACL-EAMT, Budapest.

M. Dillinger. 2001. Dictionary Development Workflow for MT: Design and Management. Proc. MT Summit, Santiago

A. Frank, Chr. Hoffmann, M. Strobel. 2004. *Gender Issues in Machine Translation.* Univ. Bremen

P. Fung. 1995. *A pattern matching method for finding noun and proper noun translations from noisy parallel corpora:* Proc. ACL Cambridge

R. Huddleston, G. Pullum. 2002. *The Cambridge Grammar of the English Language.* Cambridge Univ. Press

P. Jackson, I. Moulinier. 2002. *Natural Language processing for Online Applications.* Amsterdam (J. Benjamins).

M. Jiménez. 2001. *Generation of Named Entities.* Proc. MT Summit, Santiago

K. Knight, Ph. Koehn. 2003: *Introduction to Statistical Machine Translation.* Tutorial MT Summit 2003, New Orleans

S. McCormick. 2001. The structure and content of the body of an OL1F v.2 File. www.olif.net

A. Meyers, M. Kosaka, R. Grishman. 2000: *Chart-Based Transfer Rule Application in Machine Translation,* Proc. COLING, Saarbrücken

S. Nießen, H. Ney. 2000. *Improving SMT Quality with Morpho-syntactic Analysis.* Proc. COLING 2000, Saarbrücken

F. Och, D. Gildea, S. Khudanpur, et al. 2003. *Syntax for Statistical Machine Translation.* John Hopkins Summer Workshop. www.clsp.jhu.edu/ws03/groups/translate

St. Piperidis, S. Boutsis, J. Demiros. 1997. *Automatic Translation Lexicon Generation from Multilingual Texts.* Proc. AAAI 1997.

St. Richardson, W. Dolan, A. Menezes, J. Pinkham. 2001. Achieving Commercial-quality Translation with Example-based Methods. Proc. MT Summit VIII, Santiago

A. Samiotou, L. Kranias. 2004.: *Exploiting Parallel Texts for Populating TM & MT Databases.* Proc. Workshop 'The Amazing Utility of Parallel and Comparable Corpora', at LREC Lisbon

Gr. Thurmair. 1990. *Complex lexical transfer in METAL.* Proc. TMI 3, Austin, Tx.

Gr. Thurmair. 2003. *Making Term Extraction Tools Usable.* Proc EAMT-CLAW Dublin.

Gr. Thurmair. 2005. *Hybrid architectures for Machine Translation Systems.* LREC Journal 1, 2005 (to appear)

P. Virga, S. Khudanpur. 2003. *Transliteration of Proper Names in Cross-Language Applications.* Proc. SIGIR Toronto