Proceedings of the 16th EAMT Conference, 28-30 May 2012, Trento, Italy

Creating Term and Lexicon Entries from Phrase Tables

Gregor Thurmair
Linguatec
g.thurmair@linguatec.de

Vera Aleksić
Linguatec
v.aleksic@linguatec.de

Abstract

This document describes a tool which extracts term and lexicon entries from SMT phrase tables, without further reference to monolingual data. It applies filters to such tables, and builds lexicon entries from the 'good' candidates. Error rates of the tool can be as low as 7.3%, accumulated from source, target, and transfer errors.[1]

1 Introduction

It is a common understanding that machine translation systems need to be adapted to the domain and text type they are supposed to translate. For knowledge-driven systems, such adaptation is done by means of lexicon update: The domain terminology is identified, and coded as a special additional lexicon repository, loaded at runtime. In the age of data-driven technology, terminology is extracted from corpus data, and so are translation equivalents for the found terms.

1.1 Task

The task of the P2G (phrasetable2glossary) tool is to create proper bilingual lexicon entries from comparable corpus data; the technique should be usable for special domains, and should create output which can be imported into a backend (rule-based) MT system.

The question what the target of a bilingual extraction component is, is difficult to define. Real term banks, even in the same domain, contain very different material, depending on the subdomain and focus, and the skills of the translators involved. As a result, the term extraction process will always contain a step whereby humans investigate a term list and decide which entry candidates they want to keep for term bank import.

The task of a term extraction tool is to prepare this candidate list. The quality of the extraction tool is determined by the effort it makes to go through this list.

The approach of P2G consists of the following steps:

- Step 1: Extract phrases with a good chance of being translations of each other. This means to apply word and phrase alignment to the input. Tools exist which do this.
- Step 2: Not all phrases are well-formed terms. Therefore, the term candidates are filtered on several levels:

  Frequency filter: only phrases with a frequency and translation probability above a given threshold are considered as candidates.

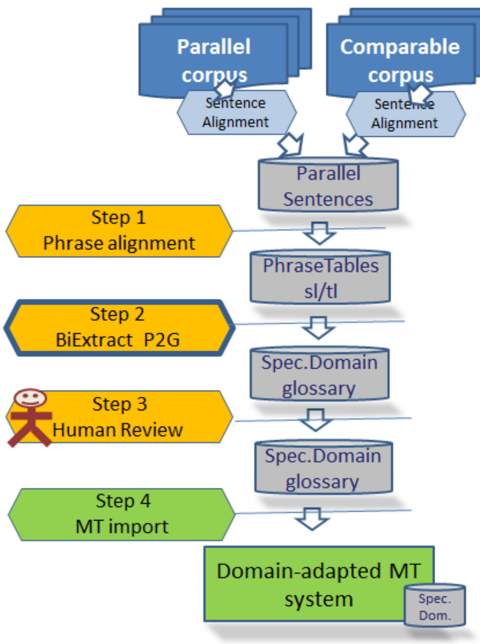

Fig. 1: Lexicon extraction workflow

[1] This work was done in the context of the FP7-ICT projects ACCURAT (248347) (system core) and the PANACEA (248064) (language extensions).

Linguistic filter: have an internal linguistic structure. Only candidates that match this structure are legal term candidates.

Lexicon filter: This is an optional user-defined component which allows eliminating non-terms.

- Step 3: The resulting list will be given to human post-editing, to correct erroneous system decisions. The quality of the tool is a key factor for the efficiency of this step.
- Step 4: The correction result will be imported into a rule-based MT system (e.g. Linguatec's '*Personal Translator'*). Care must be taken that all the annotations required by the backend MT system are available.

The focus of this paper is on step 2, term and lexicon extraction.

1.2 Related Work

There is an abundance of literature on bilingual term extraction. In the present context, we focus on papers which use phrase alignment for term extraction.

Macken et al. 2008 use linguistic pre-processing on the SL and TL side, and try to identify chunks from which they can conclude phrase similarity. They report an error rate between 15% and 33%, for the automotive domain. Our approach has a much smaller error rate, and does not need any corpus pre-processing.

Ideue et al. 2011 first extract term candidates from SL and TL texts, and then try to find matches in bilingual phrase tables, which they score according to different measures. They have a very small evaluation set (only 100 terms); however, the argument would be that

a. if a string is a term then it *must* show up in the aligned phrases somehow,
b. if it shows up in the phrase tables then it *must* be able to be extracted from there, and no reference to any source and target sentences is required
c. as a consequence, no comparison / distance between sentence-based and phrasetable-based terms needs to be computed.

In turn, our approach needs *only* aligned phrases as input, and tries to find the good terms in them.

Wolf et al. 2011 have a similar objective than the present report, namely using phrase tables for RBMT lexicon improvement; they use a full RBMT analysis (and generation) component to identify translation candidates in the phrase tables, by exploring if a phrase table entry matches constraints imposed by the MT tree. They do not report evaluation results for term extraction but only for overall MT quality improvements; however they share a lot of aspects (like the need to create MT-compatible entries) with the present work.

Our approach is more robust, as it does not need a full MT system for term identification, and does not require 'phrase-table-external' term candidates; it applies linguistic patterns which are usable by most RBMT systems, and provides annotations which should enable a straight-forward lexicon import.

All these approaches follow the standard approach towards bilingual term extraction, which is a two-step procedure: First *identification* of term candidates in the source language, and then *mapping* of source to target term candidates. Usually the corpus data need to be preprocessed, from the level of lemmatization / POS tagging (Caseli/Nunez 2006) to the level of logical form creation (Menezes/Richardson 2001); this is always a source of error.

1.3 Approach

The system presented here takes the opposite approach: It does mapping *first* (using state-of-the-art phrase aligners), and *then* it does extraction from the aligned phrases, by applying filters to the phrases. This approach follows the following considerations:

1. If a (monolingual) source language term candidate does not have a correspondence in the target language, it is unlikely that it is really a term. In turn, this means that if something is a term (i.e. a relevant concept) in a bilingual setup, then it *must* show up in the alignment results, and the alignment can be used as a filter for term candidates.
2. The best available alignment tools produce translation tables which contain all possible term mappings (and beyond that many phrases which would not be considered as proper terms). So most of the correct term candidates *will* be represented in such translation tables.
3. As a result, the task consists in identifying 'good' term candidates from phrase table input. This is achieved by applying different *filters* to such input to extract the good terms.

Therefore, the approach reverses the identification and mapping steps, and identifies term candidates only from alignment results. The *only* source of input therefore is a set of aligned phrases, as produced by standard aligners. No monolingual extraction is needed.

2 Mode of Operation

As mentioned earlier, the approach is to apply filters on input records of aligned phrases, whereby formats of different alignment tools are supported as input.
Three filters are applied, as shown in Fig. 2.

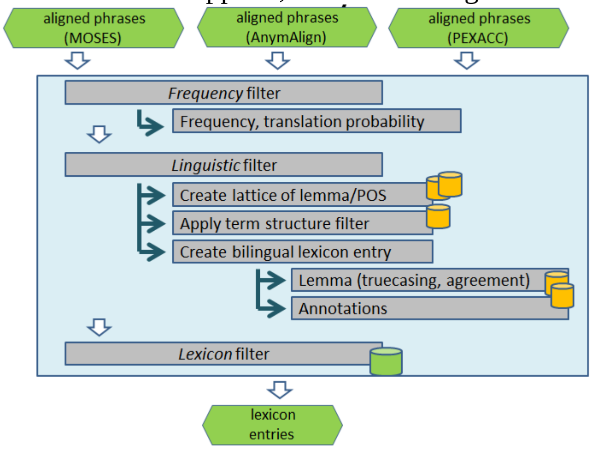

Fig. 2: Operation flow of the P2G system

The filters are:

- A **Frequency filter**: Only phrases with a given frequency and / or translation probability are accepted as term candidates
- A **Linguistic filter**: Only phrases which have certain linguistic properties are acceptable.

  If a candidate passes the linguistic filter, it is brought into the right lexicon form, in terms of lemma creation, assignment of annotations, etc.
- The **Lexicon filter** compares the lexicon entries just produced with a filter resource. This way, candidate entries can be removed which are already known, or are not wanted, or should not be part of the output for some other reason.

Details are given in the following sections.

2.1 Frequency Filter

As the system does not itself create alignments (i.e. translation candidates), it must rely on the efficiency of the statistical alignment tools from which it receives the aligned candidates. The first step is therefore to identify the best translation proposals, in terms of recall (as many terms as possible) and precision (as good translations as possible).

Two factors influence the translation quality of the P2G tool: the selection of the alignment tool, and the selection of the thresholds for frequency and translation probability.

For the alignment tool, it can easily be seen that GIZA++ only is insufficient, as no multi-word entries are found, which form close to 50% of a lexicon / term list, esp. in narrow domains. So the focus was on phrase alignment tools, which also give superior quality in translation (Och/Ney 2004). To create phrase alignment, two alignment methods were tried out[2]:

- Giza++ and **MOSES** (cf. Koehn 2010), creating Phrase Tables. From the *LT_automotive* input data (cf. below), a phrase table with about 7.97 mio entries was built.
- Phrases as produced with **Anymalign (**Lardilleux/Lepage 2009**)**. Anymalign created about 3.14 mio word/phrase pairs from the same input data.

It soon turned out that if **frequency** is not considered, too much noise would be in the output. Therefore, frequency (on source and target side) is used and set to > 1.

For the **translation probability**, tests were done to find the optimal recall / precision combination.

The two alignment systems were compared, using different values for the translation probability. For evaluation, a random set of term candidates manually inspected[3], and the errors in alignment / translation were counted[4]. The results are given in Table1.

Tool	transl. prob.	no entries	errors
MOSES	p > 0.8	12.000	5.54%
MOSES	0.6 < p < 0.8	3.900	5.42%
MOSES	0.4 < p < 0.6	20.000	55.11%
AnymAlign	p > 0.7	12.600	46.91%
AnymAlign	p > 0.8	10.900	47.56%

Table 1: Translation errors for different alignment methods and probabilities

It can be seen that the MOSES Alignment has a much better quality, and is in the reach of being usable; AnymAlign error rates are ten times higher. For AnymAlign, taking a higher threshold (0.8 instead of 0.7) does not improve alignment quality. Overall MOSES input with a

[2] Input from PEXACC (Ion et al. 2011) for comparable corpora is also supported.
[3] Entries starting with the letters C, F, and S.
[4] There are always unclear cases among translations (e.g. transfers usable only in certain cases); they were not counted as errors. Errors are only clearly wrong translations; however a range of subjectivity remains.

threshold of 0.6 for P(f|e) seems to give best results for term extraction, for this size of phrase tables[5], with an overall error rate of about 5.5%: It increases recall without reducing precision.

It should be noted that alignment errors result from external phrase alignment components, and are just 'inherited' by the current extraction system. However, they count in the overall workflow evaluation: Incorrect translation proposals lead to significantly higher human reviewing effort.

2.2 Linguistic Filter

Not all phrase aligned candidates which pass the frequency filter are linguistically meaningful. So only the ones which can be terms, or lexicon entries, are extracted[6]. Most such terms have an internal linguistic structure, described by a part-of-speech tag sequence. So the internal structure of the linguistic filter is:

- Create a word lattice for the input string, providing the different readings for each of the input words
- Match the input lattice to the legal term patterns, on source and target side;
- Create a lexicon entry for candidates with a successful match on both source and target side, with proper lemma and its annotations.

a. Word lattice

First, each candidate input phrase is tokenized and normalized in spelling and casing[7].

Next, each token is lemmatized to find its base form and part-of-speech tag. Lemmatization is basically done by lexicon lookup. Unknown words are handled by a POS-defaulting component; for German unknown words, a decomposer component is called to find a known head word. This procedure is documented in (Thurmair et al. 2012).

As tokens can have multiple readings, the result of this procedure is a word lattice consisting of the respective readings of each of the single words of a candidate. This procedure is language-specific, and is done on both source and target side.

b. Term Pattern matching

From the word lattice, all possible POS sequences are created, and compared to the legal term structure patterns.

The patterns go significantly beyond the 'usual suspects'; they were collected as the result on an inspection of a large terminological database. For German, patterns for the structures are provided[8] as shown in Fig. 3.

Term ::= AdP? NoC (NoC | NP | PP)?
AdP ::= Ad | VbP
NP ::= Dt (AdP)? NoC
PP ::= (Ap Dt? AdP? NoC) | (ApPD AdP? NoC)

Fig. 3: Term structure for German

The maximum length of such patterns is set to 6 members; longer terms are hardly ever found in term banks, and are even rarer in running texts.

The pattern filters are of course language-specific; e.g. in German and Greek, patterns must be foreseen which cover post-head NP's in genitive case, French and Spanish patterns cover both prenominal and postnominal adjectives, etc.

The matching strategy is a simple best-first approach, i.e. it returns the first match. It could be improved by sorting the multiword patterns according to frequency, and/or giving weights to the different POS readings of an input word. However such extensions would only marginally affect the results, and would not avoid the most frequent errors of this filter (cf. the evaluation below, section 3).

The pattern filter is applied to the candidates on both the source and target side, independently of each other, to be able to map a source language single word (e.g. a German compound) to a target language multiword expression. If both side candidates pass the filter, then the sequence of readings corresponding to the matching patterns is given to the entry creation module.

c. Term and Lexicon Entry Creation

All entries which have passed the filter so far must be brought into a proper canonical form. The creation of lexicon entries for source and target consists of two parts:

[5] However, this changes with the size of the phrase table, cf. section 5.5 below.

[6] As a consequence, there are phrases in the phrase table which are perfectly valid translations, however would never be found in a term bank.

[7] Normalization in casing is problematic as it also lowercases proper names. However, *not* doing it would lead to significant errors due to the fact that phrase tables contain many capitalized non-propername words. The output would contain pseudo-doublets from capitalized and non-capitalized term proposals. Example: '*Financial debt' where* lowercased '*financial debt'* can also be found.

[8] Not covered: Proper nouns (*Lufthansa Service Center*), and terms containing conjunctions (*Facts and Figures*), as the backend MT system cannot cope with some of such structures.

- Creating proper **lemmata**. This is required for both term and lexicon use.
- Creating proper **lexicon entries**. This is relevant if the extracted terms are to be integrated into MT systems; such systems usually require certain annotations (at least part of speech information).

Lemma creation implies the creation of a canonical form for the entry. This has two aspects:

- **Truecasing** of all lemma parts: Proper names and German common nouns should be capitalized, the other forms lowercased.
- Production of the **canonical form** of the lemma.

  The *head* (or the term if it is a single word) is lemmatized, and the lemma is given as canonical form. In multiword entries, the head position is given in the pattern.

  The *modifiers* in a multiword entry are treated as follows:
  Head-modifying adjectives must be set into gender-number-agreement with their head (it '*cardiopatia coronarica'*, es '*cuestión política'*)[9]. Therefore the production of the lemma of multiword entries requires knowledge about the gender of the head. To provide this, a special component (gender defaulter) has been added to the system which consults an appropriate resource; depending on the gender of the noun, the adjective is inflected[10].
  The *post-head modifiers* of the multiword stay in their inflected form: de '*Oberfläche mit speziellen Farbpigmenten*', en '*surface with special color pigments*' would leave the PP untouched.

Based on these two principles, the multiword lemma is composed[11].

It should be noted that the step of creating canonical forms can create duplicates (e.g. if a phrase table contains one entry for a singular and another one for a plural noun). Such duplicates must be eliminated before the final list is output.

Lexica go beyond term lists as their entries need **annotations**. The lexicon entries in P2G show the following annotations:
All of them have a lemma, a part of speech, and a reading number, as these elements constitute an entry. In addition, they have annotations which depend on a feature called '*entrytype'*, with values '*singleword'*, '*compound'*, '*multiword'*.

Single word entries are annotated with gender (in German) and inflection; this information is either taken from the lexicon, or defaulted.

Multiword entries and compounds (i.e. the agglutinated German compounds) share the same entry structure; they provide: the head position, the sequence of lemmata, and the sequence of parts of speech of which the multiword consists. These annotations allow for a successful identification of multiword terms in texts.

Of course, the lexicon must contain much more information; however this goes beyond what the term extraction can contribute. In turn, the use which can be made of the provided annotations depends on the single backup MT systems and their import possibilities: Most systems can use (or even require) POS information, but e.g. not all multiword term patterns are supported (e.g. terms containing conjunctions). Tests on transfers, like in (Caseli/Nunez 2006), are not created, however.

The final output of the linguistic filter consists either of complete *lexical* entries (for MT import), or of *term* entries (for human lookup), depending an output format parameter.

2.3 Lexicon Filter

Before human post-editors select the entries which they really want to keep, a possibility has been created to remove unwanted term candidates. Such entries could be:

- Candidates which are already known; they need not be reviewed a second time
- Candidates which do not belong to a specific domain (e.g. automotive); the filter then would be a general-domain lexicon, letting pass only narrow-domain words
- Candidates which contain certain stopwords (like en '*large'*)
- Candidates which are known to be irrelevant.

The system offers the option to apply a filter which blocks this kind of entries. Users would

[9] In German, there are even two options, the weak inflection (<das> '*niedrige Zinsniveau'*) or the strong one (<ein> '*niedriges Zinsniveau'*).Both can be found in dictionaries; the strong inflection is more difficult as it requires knowledge of the head noun gender; unfortunately this is the form expected by the backend MT system.

[10] The system uses a static inflection resource for this.

[11] These heuristics for truecasing and for lemma creation leave room for errors, e.g. in cases where the prenominal adjective is in comparative form (de '*der frühere Präsident*' -> *'*der frühe Präsident*'), or in cases where the head should be in plural (en '*facts & figures*' -> *'*fact & figure*'). However, they show the best performance overall.

provide the filter data themselves; only non-matching entries pass the lexicon filter.

3 Evaluation

3.1 Methodology

As explained above, it is difficult to evaluate a term extraction tool vis-à-vis a gold standard; term extraction always depends on the knowledge and interest of the users. Despite of a research focus on the selection of relevant entries (‘termhoood’, cf. Vu et al. 2008, Wong et al. 2007, Kit 2002), there will always be a step where users review the list of candidates produced by the extraction tool, and select the entries they want to keep.

While there is no clear view which entries *should* be in the term list, on the other side, there is agreement on which candidates should *not* be presented, and be considered as noise: wrong translations, the same entry in singular and plural form, or in capitalized and lowercased spelling, etc. It is *this* type of entry, which a term extract evaluation should focus on: Creation of only ‘good’ term candidates. This is what the following evaluation does.

3.2 Data

Several corpora were used for testing, related to several projects:

- The PANACEA corpora for environment, prepared by DCU: (*DCU_ENV*) and labour legislation (*DCU_LAB*)[12]
- Corpora in the Health and Safety domain, collected by Linguatec (*LT_H&S*) in different languages
- A corpus on automotive texts, collected by Linguatec (*LT_autom.*)
- The ACCURAT corpora for automotive, in two versions, prepared by DFKI: *DFKI_adapt* and *DFKI_lexacc*[13].

The size, languages treated, size of phrase tables created, and number of glossary entries extracted is given in Table 2.

3.3 Evaluation Procedure

From all corpus data sets, term candidates were extracted by the P2G system. From these candidates, term candidates were selected randomly. These candidates were evaluated manually by two evaluators.

[12] cf. Mastropavlos / Papavassiliou. 2011.

[13] cf. ACCURAT Deliverable D4.2: Improved baseline SMT systems adjusted for narrow domain. 2012

Corpus	lang.	No. sentences	Phr.Tab. size
DCU_ENV	en-fr	29 K	0.4 M
DCU_LAB	en-fr	21 K	0.8 M
LT_H&S	en-fr	52 K	2.9 M
LT_H&S	en-es	48 K	2.6 M
LT_H&S	en-it	40 K	2.1 M
LT_H&S	en-pt	14 K	0.6 M
LT_autom.	en-de	155 K	7.97 M
DFKI_adapt	en-de	1483 K	85.0 M
DFKI_lexacc	en-de	1595 K	83.9 M

Table 2: Test corpora (no. sentences, phrasetable size, size of extracted glossaries)

Overall, 99 K bilingual term candidates were extracted of which 17.2 K (17%) were manually evaluated; details are given in Table 2 below.

3.4 Results

First, speed was measured for the corpora. Depending on the frequency filter, the system processes between 45K (no filter) and 170K (0.8 filter) entries per second on a standard PC. This would be fast enough for practical use.

As for quality and errors, two kinds or errors are distinguished in the evaluation:

- *Translation* errors, i.e. the candidates are not translations of each other. These errors are produced by the aligners, as explained above. For the final tests, MOSES was selected as alignment method, with a translation probability threshold set to 0.6 and a frequency threshold set to >1.
- *Lemma and annotation* errors; these errors are created by the P2G tool. They are obviously language-specific; an error analysis is given below.

Table 3 shows the evaluation results. The average error rate of the complete P2G system is 9.26%, varying from 7.3 to 14.4%.

Translation errors: Translation errors vary from 1.5% to 12.7%, with 5.1% on average.

Translation errors seem to correlate with the size of the phrase tables[14]: Larger phrase tables show a lower translation error rate for the extracted terms. This is not particularly surprising, as more data usually lead to better performance.

Translation errors are produced by MOSES alignment, and are not accessible to the P2G tool; however, they increase the total error rate.

[14] DCU_ENV and DCU_LAB need to be considered in more detail.

	PhrTab size	Gloss. size	Transl. error	P2G error	Total error
DCU_ENV	400	2.8	5.2%	1.3%	7.8%
DCU_LAB	800	4.5	4.9%	1.2%	7.3%
LT_H&S fr	2.900	10.7	11.3%	1.3%	13.9%
LT_H&S es	2.600	13.2	10.9%	0.4%	11.6%
LT_H&S it	2.100	9.9	9.8%	2.3%	14.4%
LT_H&S pt	600	4.4	12.7%	0.4%	13.5%
LT_autom.	7.970	15.7	5.7%	2.8%	10.3%
DFKI_adapt	85.000	23.2	1.5%	3.3%	8.0%
DFKI_lexacc	83.900	23.3	1.7%	3.1%	7.9%

Tab. 3: Evaluation results: Phrase Table size (K entries), size of extracted glossaries (K entries), error rates of translation, of P2D, and combined error rates

P2G errors: P2G errors vary from 0.4% to 3.3%, depending on the languages involved[15], with an average error rate of 2.1%. Main of errors are:

- errors in linguistic filtering: either homograph words pass the filter (en '**are permanent'* as '*are'* etc. has also a noun reading; similar it '*sono'* in '**sono piccolo*', etc.). Or patterns pass the filter which are no terms but happen to have the 'right' structure: en '**strategy for example*', it '**formazione a favore*', de '**Flüchtlings-fonds für den Zeitraum'*.
- errors in lemma creation: either errors in casing (en '**fujitsu'*, '**flemish port'*), mostly due to lexicon gaps, or errors in agreement, (de '**freundlicher Wort'*, *fr '*force élevées'*, *es '*animal infectados'*).

Many of these errors can be corrected by improvements of the backend components (dictionary, gender defaulters etc.), which would bring the P2G error rate down by an estimated 1%.

The P2G errors do not depend on the size of the data; they are language-dependent of course: Errors in German result from more complicated gender agreement; in Italian, homograph problems, in English casing problems are the main error source. Variations of error rates within one language in the different test sets do not seem to be significant.

Total errors: As the output of the system is a bilingual lexicon, i.e. description of two source terms plus their translation, the error rates accumulate, so the overall error rate of the tool is two P2G errors plus translation errors; the total error rate is somewhat linear to the translation error rate. In total it is between 7.3% and 14.4%, which means that 8 entries out of 100 need to be corrected by human reviewers. This can be considered a reasonable result of a term extraction component.

3.5 Recall Issues

Another observation is that the number of phrase table entries containing good terms decreases with the size of the phrase table: As Table 2 shows, the extraction factor for smaller tables is about 150 phrases per 'good' term, while for the large tables it is about 3600, producing only 23.000 terms. So, either these tables contain more irrelevant entries, or the translation probability factors need to be adjusted in relation to the size of the phrase table.

A comparison between the terms of *DFKI_lexacc* and *DFKI_adapted* showed that there was a difference of about 15% in the output entries, meaning that there are at least 15% undetected 'good' terms in the data.

As a consequence, the translation probability threshold for the frequency filter should be set depending on the size of the phrase table. To test this, the *DFKI_lexacc* data were split into packages depending on the translation probabilities. In each package, about 1000 entries were manually evaluated. The result is shown in Table 4.

Translation . probability.	no entries found	error rate
p > 0.8	5.900	2.11%
0.6 < p < 0.8	20.500	0.58%
0.4 < p < 0.6	54.900	2.33%
0.2 < p < 0.4	58.100	4.03%
0.0 < p < 0.2	1.001.900	59.69%

Tab. 4: Error rates and probabilities in large phrase tables (DFKI_lexacc)

The results show that the entry sets with a probability > 0.4 have basically the same error rate (the 0.58% may be due to some data idiosyncrasies); entry sets from 0.2 to 0.4 have a slightly increased error rate, and entries < 0.2 cannot be used.

This means that recall can be improved dramatically by lowering the probability threshold, with no or just minimal loss in precision, cf. Table 5.

translation probability	no. entries retrieved	expected translation error rate
P (f|e) > 0.4	67.664	2.25 %
P (f|e) > 0.2	109.418	3.53 %

Tab. 5: Recall improvement for large phrase tables

[15] P2G supports the languages en de fr es it pt

As a result, the P2G term extraction tool can produce a 110 K bilingual glossary from phrase tables where 92 out of 100 entries are correct (7.7% total error rate[16]).

Schalterkontakt	No	switch contact	No
Schalterraum	No	switch room	No
Schaltfinger	No	shift finger	No
Schaltgabel	No	shift fork	No
Schaltgetriebe eines Fahrzeugs	No	gearbox of a vehicle	No
Schalthebel	No	shift lever	No
Schalthebelanordnung	No	shift lever assembly	No
Schalthebel-Positionssensor	No	shift lever position sensor	No
Schalthebelsystem	No	shift lever system	No
Schalthebelvorrichtung	No	shift lever device	No
Schaltintervall	No	shift interval	No
Schaltkabel	No	shift cable	No
Schaltkanal	No	shifting channel	No
Schaltknopf	No	button	No
Schaltkolben	No	shifting piston	No
Schaltkraft	No	switching force	No
Schaltkraftsensor	No	gear shift sensor	No

Fig. 4: Example term output (automotive domain)

References

Caseli, H., Nunes, M., 2006: Automatic induction of bilingual resources for machine translation: the ReTraTos project. Machine Translation 20,4

Daille, B., Morin, E., 2005: French-English Terminology Extraction from Comparable Corpora. Proc. IJCNLP 2005

Fung, P., McKeown, K., 1997: Finding Terminology Translations from Non-parallel Corpora. Proc. 5th Annual Workshop on Very Large Corpora (VCL 97), Hong Kong

Gamallo Otero, P., 2007: Learning Bilingual Lexicons from Comparable English and Spanish Corpora. Proc MT Translation Summit Copenhagen

Gamallo Otero, P., 2008: Evaluating Two Different Methods for the Task of Extracting Bilingual Lexicons from Comparable Corpora. Proc. LREC Workshop on Comparable Corpora, Marrakech

Ideue, M., Yamamoto, K., Utiyama, M., Sumita, E., 2011: A Comparison of Unsupervised Bilingual Term Extraction methods Using Phrase Tables. Proc. MT Summit XIII, Xiamen

Ion, R., Ceauşu, A., Irimia, E., 2011: An Expectation Maximization Algorithm for Textual Unit Alignment. Proc. 4th Workshop on Building and Using Comparable Corpora (BUCC), Portland, USA

Kit, Ch., 2002: Corpus Tools for Retrieving and Deriving Termhood Evidence, Proc. 5th East Asia Forum of Terminology

Koehn, P., 2010: Statistical Machine Translation. Cambridge University Press.

Lardilleux, A., Lepage, Y., 2009: Sampling-based multilingual alignment. Intern. Conf. on Recent Advances in Natural Language Processing (RANLP 2009), Borovets, Bulgaria

Macken, L., Lefever, E., Hoste, V,.2008: Linguistically-based sub-sentential alignment for terminology extraction from a bilingual automotive corpus. Proc. 22nd COLING, Manchester

Mastropavlos, Nikos; Papavassiliou, Vassilis. (2011). Automatic Acquisition of Bilingual Language Resources. Proceedings of the 10th International Conference on Greek Linguistics. Komotini, Greece

Menezes, A., Richardson, St.D., 2001: A best-first alignment algorithm for automatic extraction of transfer mappings from bilingual corpora. Proc. ACL / DMMT

Morin, E., Prochasson, E., 2011: Bilingual Lexicon Extraction from Comparable Corpora Enhanced with Parallel Corpora. Proc. BUCC, Portland, Oregon, USA

Och, F., Ney, H., 2004: The Alignment Template Approach to Statistical Machine Translation. Computational Linguistics 30,4

Rapp, R., 1999: Automatic identification of word translations from unrelated English and German corpora. Proc. 37th ACL, College Park, Maryland

Robitaille, X., Sasaki, X., Tonoike, M., Sato, S., Utsuro, S., 2006: Compiling French-Japanese Terminologies from the Web. Proc. 11th EACL, Trento

Thurmair, Gr., 2003: Making Term Extraction Tools Usable. Proc. CLT, Dublin

Thurmair, Gr., Aleksić, V., Schwarz, Chr., 2012: Large-scale lexical analysis. Proc. LREC Istanbul

Vu, Th, Aw, A.T., Zhang M., 2008: Term Extraction Through Unithood And Termhood Unification Proc. IJCNLP 2008, Hyderabad, India

Weller, M., Gojun, A., Heid, U., Daille, B., Harastani, R., 2011: Simple methods for dealing with term variation and term alignment. Proc TIA 2011: 9th International Conference on Terminology and Artificial Intelligence, Paris, France

Wolf,P., Bernardi, U., Federmann, Chr., Hunsicker, S., 2011: From Statistical Term Extraction to Hybrid Machine Translation. Proc EAMT Leuven

Wong, W., Liu, W., Bennamoun, M., 2007: Determining Termhood for Learning Domain Ontologies using Domain Prevalence and Tendency, Proc. Sixth AusDM 2007), Gold Coast, Australia. CRPIT 70

[16] Two times the average P2G of 2.1% plus the translation error rate of 3.53%