CREATING A BILINGUAL PSYCHOLOGY LEXICON FOR CROSS LINGUAL QUESTION ANSWERING, A FOLLOW UP STUDY

Andrea Andrenucci
Department of Computer and System Sciences, Stockholm University
andrea@dsv.su.se

Keywords: Internet services, Natural Language Interfaces, Data Mining, Cross Lingual Question Answering

Abstract: This paper discusses a follow-up study aimed at investigating the extraction of word relations from a medical parallel corpus in the field of Psychology. Word relations are extracted in order to create a bilingual lexicon for cross lingual question answering between Swedish and English on a medical portal. Six different variants of the corpus were utilized: word inflections with and without POS tagging, syntactically parsed word inflections, lemmas with and without POS tagging, syntactically parsed lemmas. The purpose of the study was to analyze the quality of the word relations obtained from the different versions of the corpus and to understand which version of the corpus was more suitable for extracting a bilingual lexicon in the field of psychology. The word alignments were evaluated with the help of reference data (gold standard) and with measures such as precision and recall.

1. INTRODUCTION

Users of medical portals in general, regardless of their background, value the possibility of formulating their information needs in their own native language. In Question Answering this is possible with the help of Machine Translation (MT), which converts user questions into the language of the texts from where the answers are extracted. This paradigm is called Cross Language Question Answering, CLQA (Aunino, Kuuskoski and Makkonen, 2004).

The Web4health medical portal (http://web4health.info) supports cross language question answering (CLQA). User questions are translated into English with the help of Systran's MT system (http://www.systransoft.com) and are then used to retrieve answers from the knowledge base of the portal. One problem with the existing implementation is that Systran implements medical lexicons which are not tailored to the specific domain of the portal, i.e. psychology and psychotherapy. The aim of this research is to produce a bilingual lexicon for Swedish and English that overcomes this gap. For this purpose we have investigated the possibility of automatically extracting word relations from a parallel corpus (Swedish and English), which consists of Web4health's knowledge base. The corpus was extracted in two versions, one version consisting of words in their inflected forms and another version consisting of word lemmas. For both versions we also provided three variants: 1) a variant annotated with part of speech (POS) tagging, 2) a variant with POS-tagging and syntactically parsed 3) a "as-is" variant (i.e. without POS tagging nor parsing). The purpose of the study was to analyze the quality of the word relations obtained from the different versions of the corpus and the quality of the relations with different word frequencies. This was done in order to understand which version of the corpus was more suitable for extracting a bilingual lexicon.

The texts were aligned at the paragraph, sentence and word level with the Uplug toolkit (see section 3), a collection of tools for processing parallel corpora, developed by Jörg Tiedemann (2003a). Uplug utilizes both statistical and linguistic information in the alignment process. The alignments were evaluated at the word level with the help of reference data (gold standard), which were constructed before the word alignment process. The gold standard was created with a frequency based sampling approach (see section 4.2).

The paper is structured as follows: section two describes related research in the field of cross lingual question answering within medicine. Section three summarizes the knowledge base and the Uplug toolkit. Section four and five describe the pilot study and its quantitative results. The paper is concluded with a discussion of the results (section six) and the paper conclusions (section seven).

2. RELATED RESEARCH

Several projects have focused on developing lexical resources for the medical domain. Marko et al. (2006) created multilingual medical lexicons mapping monolingual lexicons (in French, English and German) to one another. No parallel corpora were utilized. The researchers mapped terms in each monolingual lexicon to interlingua representation (i.e. a morphosemantic representation) of the terms. This methodology is similar to pivot alignment (Borin 99) and differs from the approach of this research that utilizes parallel corpora without mapping terms to intermediate representations.

Nyström et al. (2006) developed a medical English-Swedish dictionary utilizing word alignment of several international medical terminology resources such as MeSH (Medical Subject Headings) and ICF (International Classification of Functioning, Disability And Health). There are two major differences between Nyströms work and our research. Nyström produced a dictionary with medical terms for medical practitioners; our purpose is to produce a lay dictionary (consumer vocabulary) of terms that are understandable for common people.

Nyström also utilized metrics of precision and recall but only considered entirely correct alignments. This approach works well when it comes to evaluating single word alignments (SWUs), but is too coarse for the evaluation of Multi Word Units (MWUs), which often imply partially correct results (Tiedemann, 2003b, p. 26). Since we are even interested in semi automatically extracting MWUs we even considered partially correct links (links that have at least one correct word on source and target side).

Baud et al. (1998) also built a bilingual medical lexicon aligning the French and English International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10). Unlike this research, no syntactic parsing was performed and no recall results were provided.

3. THE KNOWLEDGE BASE AND THE UPLUG TOOLKIT

The Web4health medical portal (http://web4health.info) is well established among the medical portals on the Web. Psychiatrists and psychotherapists from five different European countries (Italy, Sweden, Holland, Greece and Germany) use the portal to jointly develop a set of semantically classified Web pages that answer questions in matters of psychological and psychotherapeutic advice. Users consult the knowledge base submitting questions in natural language, which are then matched against pre-stored FAQ-files (Frequently Asked Questions).

The Uplug toolkit (Tiedemann, 2003a) is a collection of tools for processing parallel corpora. Its main functionality consists of sentence and word alignments of bilingual texts. The main idea behind Uplug's alignment process is to utilize both linguistic and statistical information in order to extract word relations. Each individual piece of information is called a clue, $C_i(s,t)$, and is defined as a probability that indicates an association between two sets of words s and t in parallel texts. Formally it is defined as a weighted association A between s and t, where w_i is used to weight and normalize the score of A_i:

$$C_i(s,t) = P(a_i) = w_i A_i(s,t) \qquad (1)$$

All clues are then combined in an overall measure, which is defined as the disjunction of all indications:

$$C_{all}(s,t) = P(a_{all}) = P(a_1 \cup a_2 \cup ... \cup a_n) \qquad (2)$$

Clues are not mutually exclusive. The addition rule for probabilities generates the following formula for a disjunction of two clues:

$$P(a_1 \cup a_2) = P(a_1) + P(a_2) - P(a_1 \cap a_2) \qquad (3)$$

Two main types of clues are considered: basic (static) clues, whose value is constant for a pair of lexical items and dynamic clues, whose values are learned dynamically during the alignment process. Basic clues include co-occurrence coefficients (the Dice coefficient, Tiedemann 1999), string similarity coefficients (the longest common subsequence ratio, Melamed 1995) and GIZA++ clues (Och and Ney, 2003), based on IBM models (Brown et al., 1993) and Hidden Markov Model. Dynamic clues include patterns of POS labels, phrase types and word positions. The system aligns first sentences and words with the basic clues and then utilizes the aligned links as training data in order to learn new dynamic clues and improve the quality of the alignments. For instance, examining POS tags in source and target language, it is possible to estimate the probabilities of translation relations between words that belong to certain word classes.

A huge advantage of the Uplug tool is that it supports the dynamic construction of alignments with multi word units (MWUs), i.e. noun phrases, idiomatic expressions and other phrasal constructions that should not be split up in the alignment process (Tiedemann, 2003b, p. 18).

4. IMPLEMENTATION OF THE STUDY

This section outlines how the study was conducted. It describes how the corpora were annotated and prepared for the alignment process (section 4.1), plus how the results were extracted and evaluated (section 4.2).

4.1 The Corpus Selection and Annotation

The parallel corpus of this study consists of the knowledge base of the Web4health portal, i.e. FAQs, or question/answer pairs, in the source language (Swedish) and the target language (English). The Swedish corpus consists of circa 135301 tokens and the English counterpart of circa 143118 tokens.

Prior to utilizing the randomly chosen texts, we scanned and proofread the material and, when necessary, corrected it to ensure its completeness and correctness. This was a difficult and time consuming task, since the documents in the repository are often translated freely and the structure of the texts tend also to differ, with sentences or phrases that are available in one language only.

Prior to starting the alignment process, some preliminary work was needed in order to prepare the corpora. Since the FAQ documents are annotated with HTML tags, the texts had first to be cleaned up by the existing tags and then converted into plain text. The Uplug toolkit was then used for encoding the texts with ISO88591 for Latin1 (which includes Swedish and English) and annotating them with XML Corpus Encoding Standard (XCES) (Ide and Priest-Dorman, 2000). Sentence splitting and tokenization were included in this step.

A version of the bilingual corpus was lemmatized with the CST Lemmatizer (Jongejan and Haltrup, 2005) which is a trainable, rule-based tool that works with languages that utilize inflectional suffixes, such as Swedish and English.

The Trigrams'n Tags tagger (Brants, 2000) was utilized to annotate the POS-tagged versions of the corpora. TnT was chosen since it is the tagger that has the highest overall accuracy among data-driven taggers and succeeded best in the annotation of both known and unknown words in Swedish (Megyesi, 2000).

The tagger was trained on Swedish (Megyesi 2002) using the StockholmUmeå Corpus (SUC, 1997), and utilized for the labels the PAROLE annotation scheme (Ejerhed and Ridings, 1995), a tagset that include part-of-speech and morphological features such as gender and number of the words. The Penn Treebank corpus and its tagset (Marcus, Santorini och Marcinkiewicz, 1993), which also encodes morphological information such as number, were utilized for the English language.

The English part of the bitext was parsed with GROK (Baldrige J.), an open source library for Natural Language Processing. The Swedish part was parsed with a context free grammar parser for Swedish (Megyesi 02).

4.2 Evaluation method and the Gold Standards

After aligning the different versions of the corpus at the sentence level, capital letters were converted to non-capital letters in order to improve precision of the word-level alignment. Once the word alignment was finished, a table, with word-pair frequencies sorted in descending order, was constructed for each corpus version in order to see which alignments occurred more often. These frequency tables were later utilized for analyzing the evaluation results (see section 5 and 6). Two main evaluation techniques are utilized when it comes to evaluating word alignment (Ahrenberg et al., 2000): automatic evaluation with a reference alignment (Gold Standard) or manual evaluation by experts. Automatic evaluation was preferred since reference alignments can be re-utilized and it is possible to control the process of selecting the reference data, focusing for instance on certain word types or words from certain frequency ranges (Merkel, 1999).

Our reference data (or gold standard) were aligned manually according to detailed guidelines (Merkel, 1999). They were compiled by randomly selecting word samples from the parallel corpus. The word samples were limited to content units (phrases and content words, i.e. words with a full meaning of their own). We applied a frequency balanced approach, i.e. we grouped entries according to the following frequency ranges: 40 entries with frequency above 10, 40 entries with frequency 7-9, 40 with frequency 5-6, 40 with frequency 3-4 and 40 with frequency 1-2.

The GS included links of type “regular” (standard), “fuzzy” (somehow semantically overlapping but with different POS or different degrees of specification) and “null” (omissions). Complex MWU links were also included.

As stated of Ahrenberg et al. (2000), word alignment can be viewed as a retrieval problem. For this reason, when evaluating the quality of the alignments, it is appropriate to apply measures from the field of information retrieval such as precision and recall. By precision it is meant the ratio of correctly aligned items in proportion to the number of aligned items and by recall the ratio of correctly aligned items in proportion to the total number of correct items (reference data). However a problem of these measures is that they do not handle *partially correct links*, i.e. links that have at least one correct word on source and target side, since links are either considered as entirely right or entirely wrong. This approach works well when it comes to evaluating single word alignments, but is too coarse for the evaluation of MWUs, which often imply partially correct results (Tiedemann, 2003b, p. 26).

In order to overcome this deficiency we chose to apply refined metrics of precision and recall (Tiedemann, 2003b, p. 68) that measure the *degree of correctness* of the proposed links. They calculate a partiality value Q that is proportional to the number of words that are in common between the proposed alignments and the reference data:

$$Q_x^{precision} = \frac{\left|aligned_{src}^x \cap correct_{src}^x\right| + \left|aligned_{trg}^x \cap correct_{trg}^x\right|}{\left|aligned_{src}^x\right| + \left|aligned_{trg}^x\right|}$$

$$Q_x^{recall} = \frac{\left|aligned_{src}^x \cap correct_{src}^x\right| + \left|aligned_{trg}^x \cap correct_{trg}^x\right|}{\left|correct_{src}^x\right| + \left|correct_{trg}^x\right|}$$

$aligned_{src}^x$ is the set of source language words and $aligned_{trg}^x$ the set of target language words in link proposals for a reference link x in the GS. $correct_{src}^x$ and $correct_{trg}^x$ define the sets of source and target words of reference link x. Precision (P) and recall (R) are then defined with the help of Q:

$$R = \sum_{x=1}^{X} \frac{Q_x^{recall}}{|correct|} \qquad P = \sum_{x=1}^{X} \frac{Q_x^{precision}}{|aligned|}$$

$|aligned|$ is the total number of correct, incorrect and partially correct links in relation to the GS and $|correct|$ represents the size of the GS.

These metrics handle also partially correct links in a more fine grained way, unlike other coarser approaches (e.g. the PLUG metrics, Ahrenberg et al., 2000) that penalize partially correct links with a constant value without considering the degree of correctness of the links.

5. QUANTITATIVE RESULTS

Our study produced the quantitative results that are shown in table 1 below. In the next section we discuss the results with the help of the data from the frequency tables and elicit the differences between each version of the corpora.

The results in table 1 present some interesting differences between the corpora. For what concern word inflections, the POS tagged corpus and the shallow parsed corpus had better precision/recall for low frequency MWUs and SWUs (frequency rate 1-2, 3-4). With higher frequency rates “untagged” word inflections achieved slightly better results than POS tagged and parsed alignments.

Alignments based on lemmas had in general worse statistical results than word inflections. This is a surprising effect that partly contradicts the results of our pilot study that bootstrapped this research (Andrenucci 2007). However the pilot study utilized a corpus that was only a small subset of the whole parallel corpus: the Swedish sample corpus consisted of 12800 tokens and the English counterpart of circa 13000.

Low frequency (1-2) lemmas achieved worse precision results than POS tagged and syntactically parsed lemmas.

With higher frequencies (5-6, 7-9, >10) the “as-is” lemmas gathered better precision results for both SWUs and MWUs. Syntactically parsed lemmas had lower precision results than POS tagged lemmas in all frequency rates.

Table 1: Precision and Recall results

Frequency rate	**Word inflections Precision**	**Word inflections Recall**	**POS Words Precision**	**POS Words Recall**	**Parsed Words Precision**	**Parsed Words Recall**
MWUs 1-2	52,55	64,39	63,38	84,75	65,62	84,94
MWUs 3-4	73,71	88,81	74,10	90,17	76,27	90,5
MWUs 5-6	71,39	86,34	70,60	86,27	69,52	86,25
MWUs 7-9	67,50	85,59	64,31	83,59	63,9	83,46
MWUs >10	67,35	85,19	63,90	83,06	61,24	82,83
SWUs 1-2	73,15	87,37	76,30	90,40	77,69	90,40
SWUs 3-4	82,01	89,91	88,02	99,07	88,39	99,23
SWUs 5-6	86,14	96,49	81,75	93,50	81,91	93,57
SWUs 7-9	86,90	96,30	84,12	95,12	84,17	95,07
SWUs >10	87,66	98,08	85,83	97,36	79,70	94,57
Frequency Rate	**Lemmas Precision**	**Lemmas Recall**	**POS Lemmas Precision**	**POS Lemmas Recall**	**Parsed Lemmas Precision**	**Parsed Lemmas Recall**
MWUs 1-2	47,51	68,02	58,12	74,02	56,52	72,79
MWUs 3-4	42,01	65,68	43,23	66,88	41,52	63,60
MWUs 5-6	53,29	66,93	40,35	56,83	39,36	52,09
MWUs 7-9	56,30	78,01	53,03	74,61	51,56	73,71
MWUs >10	54,72	72,30	52,36	71,28	51,99	70,31
SWUs 1-2	50,20	59,85	52,73	62,88	51,21	62,12
SWUs 3-4	60,52	74,44	55,31	72,22	54,64	72,22
SWUs 5-6	59,33	72,97	56,54	71,17	52,55	70,97
SWUs 7-9	58,65	72,81	57,54	70.18	53,93	67,54
SWUs >10	55,67	72,44	53,86	67,95	50,76	6,51

6. DISCUSSION AND ANALYSIS

As a complement to the statistical data presented in table 1 we analyzed the frequency tables extracted from the alignments and compared the results, trying to elicit the similarities and the differences among the different versions of the corpus. We discuss our analysis with the help of some examples of the word relations that are presented in tables 2 and 3.

Lemmas with POS Tags VS Lemmas without POS/Parse Tags

The statistical results for Lemmas with and without POS/Parsing and the examination of the frequency tables clarified some points of difference. POS-tagged lemmas were more precise when aligning compound words (which were included among the MWUs) with low frequency rate (1 or 2).

Low frequency POS tagged and syntactically parsed MWUs had fewer additions, i.e. words that are occurring in the alignments but that are not present in the reference links (e.g. “nedsatt minneskapacitet - memory deficit” VS “nedsatt minneskapacitet - memory deficit failure” see table 2), and less incorrect links.

The POS tagging and syntactic parsing proved to be useful in aligning words consisting of dissimilar strings and with low co-occurrence frequency, but sharing the same POS (e.g. two nouns: “matstrupe - oesophagus” 1-2 VS an adjective and a noun: “överkänslig - oesophagus”).

Alignments consisting of SWUs achieved fewer addictions in the corpus without POS nor parsing, particularly with higher frequency rate (“fördom - prejudice” >10, “alkoholist - alcoholic” >10, “toalett - toilet” 9-7, “tanke- thought” 9-7, see table 2). POS tagged lemmas produced slightly better results than syntactically parsed lemmas in all frequency rates.

Table 2: Alignment examples of lemmas with and without POS/Parsing

Lemmas no POS	Lemmas with POS	Parsed Lemmas
Frequency >10		
alkoholist – alcoholic	alkoholist - alcoholic drinker abuse	alkoholist - alcoholic drinker abuse
Fördom – prejudice	Fördom närstående - prejudice relative	Fördom närstående - prejudice relative
Frequency 7-9		
toalett - toilet	toalett – fear toilet	toalett - fear toilet
tanke – thought	tanke – effect thought use	tanke – effect thought use
Frequency 1-2		
överkänslig - oesophagus	matstrupe – oesophagus	matstrupe – oesophagus
nedsatt minneskapacitet - memory deficit failure	nedsatt minneskapacitet - memory deficit	nedsatt minneskapacitet – memory deficit
problem humör svängning - problem mood swing	humörsvängning - mood swing	humörsvängning - mood swing

Inflected words with POS, syntactic parsing VS Inflected words

As shown in table 1, alignments of inflected words with POS and syntactic parsing obtained, in comparison to inflected words without POS, better precision, recall results as well as a higher number of correct links in lower frequency rates (1-2 and 3-4) for both MWUs and SWUs. The morphological information helped to disambiguate the gender of Swedish adjectives in noun phrases, including them in the alignment when they agreed with the head noun and their inclusion was necessary to build a conceptual unit ("dåligt uppförande - misbehaviour" VS "uppförande - misbehaviour" 1-2, where "dåligt" means "bad" and "uppförande" means "behaviour", see table 3). It also helped to link nouns with the same number ("barndomsupplevelser - childhood experiences" VS "barndomsupplevelser - childhood"). For what concerns single word units the morphological information was helpful for aligning words sharing the same definiteness ("förmågan - the ability" VS "förmågan - ability") or POS (e.g. two adjectives: "felaktiga - inappropriate" instead of a noun and an adjective "antaganden - inappropriate").

The POS based and parsed word relations had also better alignments among phrasal verbs that consisted of a verb and a particle in Swedish and a verb in English ("tänka ut - decide" VS "tänka - decide"; "klara av -handle" VS "klara - handle"). They even provided better alignments of verbs in passive forms ("omvandlas – be converted" VS "omvandlas intas med i det här - converted enters through into").

Table 3: Alignment examples of inflected words with and without POS/Parse

Word Inflections	Word Inflections with POS	Parsed Word Inflections
Frequency 1-2		
uppförande – missbehaviour	dåligt uppförande - missbehaviour	dåligt uppförande - missbehaviour
svårighetsgrad - reflect severity	svårighetsgrad – severity	svårighetsgrad – severity
Frequency 3-4		
förmågan - ability	förmågan – the ability	förmågan - the ability
magkatarr - organs gastritis	magkatarr - gastritis	magkatarr – gastritis
Frequency 5-6		
hetsätare - compulsive eater	hetsätare bli frisk - eater compulsive eater	hetsätare bli frisk - eater compulsive eater
Frequency 7-9		
alkoholproblem - alcohol problems	människor alkoholproblem motstånd - people alcohol problems attempts	Människor alkoholproblem motstånd - people alcohol problems attempts
Frequency >10		
tvångstankar - obsessive thoughts	tvångstankar ögonsmärtor - obsessive thoughts	tvångstankar ögonsmärtor - obsessive thoughts
vanföreställningar –delusions	brukarna vanföreställningarna -delusions	brukarna vanföreställningarna – delusions

It is interesting to point out that, for what concerns "as-is" word inflections, it was more difficult for low frequency words with poor string similarity coefficients to get precise alignments. The system tended to attach extra words to equalize the different string lengths. The information contained in POS and syntactic analysis overcame this problem though (ex: svårighetsgrad – reflect severity VS svårighetsgrad - severity).

The statistical figures of "as-is" word inflections are higher than POS—tagged and shallow parsed word inflections for frequency rates 5-6, 7-9 and >10 (see table 1). With higher frequency rates the higher co-occurrence coefficient values (the Dice coefficient, see chapter 3) compensated the lack of morphological information in "as-is" word inflections (e.g. tvångstankar - obsessive thoughts VS tvångstankar ögonsmärtor - obsessive thoughts, see table 3) and the role played by POS and shallow parsing was also less important in the tagged corpora. The broader the amount of dynamic clues the lesser was the quality of the alignments (within the same frequency rate) for high frequency rates.

Inflected words with POS and syntactic parsing tags VS Lemmas with POS and syntactic parsing

As we stated earlier, alignments based on lemmas had in general worse statistical results than word inflections. The truncation caused by the lemmatization process influenced the alignment at the paragraph and sentence level, worsening in some cases even the quality of sentence alignments (since the system utilize a length-bases sentence alignment algorithm, Church 93). Lemmatization influenced also the alignment at the word level. The removal of number, definiteness and gender information in Swedish nouns and adjectives, obtained through lemmatization, determined a coarser POS tagging, affecting the dynamic clues and worsening in particular the quality of the produced alignments. For instance removing the gender suffix in adjectives made it more difficult to individuate the nouns the adjectives referred to, causing less precise alignments in comparison to inflected forms (e.g. "uppförande - misbehaviour" instead of "dåligt uppförande - misbehaviour", where "dåligt" means bad and "uppförande" means behaviour).

Errors deriving form the lemmatization caused erroneous POS tagging (for instance removing the suffix "t" from Swedish adverb "vanligt", as if they were adjectives referring to "ett-"words, made the tagger mark those words as adjectives instead of adverbs). This also affected dynamic clues such as word positions and POS-patterns, and thus influenced the quality of the alignments.

Alignments of proper nouns such as medicine names with similar names in both languages (Concerta - Concerta, Trifluoperazin - Trifluoperazine), were generally correct both for lemmas and word inflections in all frequency rates, since those nouns were not truncated by lemmatization and their string similarity/length coefficients were very high .

7. CONCLUSIONS

This paper has examined the extraction of word relations from a medical parallel corpus in order to create a bilingual lexicon for cross lingual question answering between Swedish and English. Six different variants of the sample corpus were created: word inflections, word inflections with POS tagging, word inflections with syntactic parsing, lemmas, lemmas with POS tagging, lemmas with syntactic parsing. The results of this follow-up study partly confirm the results of our pilot study. POS tagging and shallow parsing enhances the quality of the alignments of both SWUs and MWUs, particularly for units with low frequency (1-2, 3-4). The role of morphosyntactic information was particularly important when aligning dissimilar strings sharing the same POS or phrase (both SWUs and MWUs). The information about gender, number and definiteness contained in the suffixes of word inflections was particularly crucial for the quality of alignment of low frequency MWUs. Considering that the medical domain is characterized by multi-word terms that are either unknown to generic lexicons or have meanings specific to this domain (Rinaldi et al, 2004), it is advisable to utilize corpora with syntactically parsed inflections as source for the extraction of bilingual lexicons for extracting low-frequency MWUs. If the target consists of words with higher frequency rates it is advisable to utilize "as is" word inflections.

Unlike our pilot study the results of the lemmatized alignments were much lower than word inflections. It is thus not advisable to lemmatize the corpora prior to extracting the corpora, since the lemmatization process creates problems at the sentence, paragraph and word level, influencing negatively the quality of the extracted alignments (at least with the tools that we utilized). Lemmatization should be applied after extracting the word alignments, in order to group together words sharing the same base form in the source language or target language and facilitate the extraction of synonym lists in both languages.

REFERENCES

Ahrenberg, L., Merkel, M., Sågvall Hein, A., Tiedemann, J., 2000. Evaluation of Word Alignment Systems. In *LREC'00, 2nd International Conference on Linguistic Resources and Evaluation*.

Andrenucci A., 2007. Creating a Bilingual Psychology Lexicon for Cross Lingual Question Answering - A Pilot Study. In *ICEIS'2007*. INSTICC Press.

Aunino L, Kuuskoski R., and Makkonen J., 2004. Cross-Language Question Answering at the University of Helsinky. In *CLEF' 04, Cross Language Evaluation Forum*.

Baldridge J. GROK An open source library for natural language processing http://grok.sourceforge.net

Baud R., Lovis C., Rassinoux AM., Michel PA, Scherrer JR., 1998.Automatic Extraction of Linguistic Knowledge from an International Classification. *Studies in health technology and informatics*, 52.

Borin L. Pivot Alignment. In *NoDaLiDa'99*.

Brants, T., 2000. TnT – A statistical Part-of-Speech Tagger. In *ANLP-2000, 6th Conference on Applied Natural Language Processing*.

Brown, P., Della Pietra S., Della Pietra V., and Mercer R., 1993. The mathematics of statistical machine translation Parameter estimation. *Computational Linguistics, 19*, 263-311

Church, K., 1993. Char Alignment, a program for aligning texts at the character level. In *ACL'93*.

Ejerhed, E. and Ridings, D., 1995. Parole and SUC, http://spraakbanken.gu.se/parole/sgml2suc.html

Germann, U., 2003. Greedy decoding for statistical machine translation in almost linear time. In *HLT-NAACL'03*.

Ide, N. and Priest-Dorman, G., 2000. *Corpus encoding standard – document CES 1*. Technical report, Vassar College, LORIA/ CNRS. Vandoeuvre-les-Nancy, France.

Jongejan, B. and Haltrup, D., 2005. *The CST Lemmatiser*, Copenhagen University, Denmark.

Lindberg, D., Humphreys, B. and McCray, A., 1993. The Unified Medical Language System. *Methods of Information in Medicine, 32*, 281-291.

Marcus, M., Santorini, B., and Marcinkiewicz, M., 1994. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19*.

Marko K, Baud R, Zweigenbaum P, Merkel M, Toporowska-Gronostaj M, Kokkinakis D, Schulz S: Cross-Lingual Alignment of Medical Lexicons. In *LREC 2006*.

Megyesi, B., 2000. Comparing Data-Driven learning algorithms for PoS tagging of Swedish. In *NoDaLiDa2001, 13th Nordic Conference on Computational Linguistics*.

Megyesi, B., 2002. *DataDriven Syntactic Analysis Methods and Applications for Swedish*. PhD Thesis, Kungliga Tekniska Högskolan. Sweden.

Megyesi, B., 2002. Shallow parsing with POS taggers and linguistic features. *Journal of machine leaning research, special issues on shallow parsing.*

Melamed, D., 1995. Automatic evaluation of uniform filter cascades for inducing N-best translation lexicons. In *3rd Workshop on Very Large Corpora*.

Merkel, M., 1999. *Annotation Style guide for the PLUG link annotator*. Technical Report, Linköping, University, Linkökping.

Nyström, M. Merkel M., Ahrenberg L., and Zweigenbaum P., 2006. Creating a medical English-Swedish dictionary using interactive word alignment. BMC Medical Informatics and Decision Making, 6(35).

Och FJ, Ney H., 2005. A Systematic Comparison of Various Statistical Alignment Models. *Computational Linguistics 29(1)*, 19-51.

Rinaldi, F., Dowdall, J., Schneider, G., and Persidis, A., 2004. Answering Questions in the Genomics Domain. In *ACL'04, Workshop on Question Answering in Restricted Domains*. ACL press.

SUC, 1997. *SUC 1.0 Stockholm Umeå Corpus, Version 1.0*. Umeå University and Stockholm University, Sweden.

Tiedemann, J., 1999. Word alignment – step by step. In *NODALIDA'99, the 12th Nordic Conference on Computational Linguistics*.

Tiedemann, J., 2003a. Combining Clues for Word Alignment. In *EACL'03, 10th Conference of the European Chapter of the ACL*. ACL press.

Tiedemann, J., 2003b. *Recycling translations. Extraction of lexical data from parallel corpora and their application in natural language processing*. PhD thesis, Uppsala University, Sweden.