Evaluating bilingual medical terminologies with word alignment methods

Mikael Nyström[a], Magnus Merkel[b], Håkan Petersson[a], Hans Åhlfeldt[a]

a Department of Biomedical Engineering, Linköping University, Sweden

b Department of Computer and Information Science, Linköping University, Sweden

LiU expanding reality

Introduction

Translations of medical terminology systems are necessary if the systems are used in countries with different languages.

Manual translations are expensive, but one solution to cut costs are to reuse already translated medical terminology systems for translation of new systems. The systems already existing in two languages are word aligned to generate lexical resources for reuse in new translations. The translation of new systems can then be semi-automatic instead of manual.

Terms in medical terminology systems have a more repetitive structure than natural language texts and the quality of the generated resources are therefore expected to be higher for medical terminology systems.

In an earlier study we used word alignment to generate a medical English-Swedish dictionary from a union of medical terminology systems[1]. In this study we explore the influence the different terminology systems and resource types have on the results.

Objective

The method used for the automatic word alignment of the terminology systems uses resources generated from the terminology systems.

In this study the aligned partitions and the partitions used to generate the resources of the terminology systems are varied. The variations give the possibility of evaluating similarities and differences within and between the terminology systems.

The similarities are mainly based on whether the contents in the rubrics consist of unique words or not, if the structures of the rubrics are repetitive and if the structure in the translated rubrics remains unchanged compared to the original rubric.

Background - Word alignment

Word alignment is the task of finding correspondences on the word or phrase level between a source text and its translation. Languages are not structured in identical ways across language borders and translations contain additions as well as omitted information. It is therefore no straightforward mapping techniques between two languages.

The standard approaches to word alignment in parallel corpora are statistical approaches and linguistic approaches. The statistical approach use probabilistic translation models. The linguistic approach often use rules for segmentation into lexical units, bilingual dictionaries and rules on word order and positions, as well as rules on corresponding part-of-speech labels.

The primary usefulness of word alignment is that it will recover and present candidate word and term pairs from a potentially previously unknown domain through a text corpus.

Methods - Alignment tools

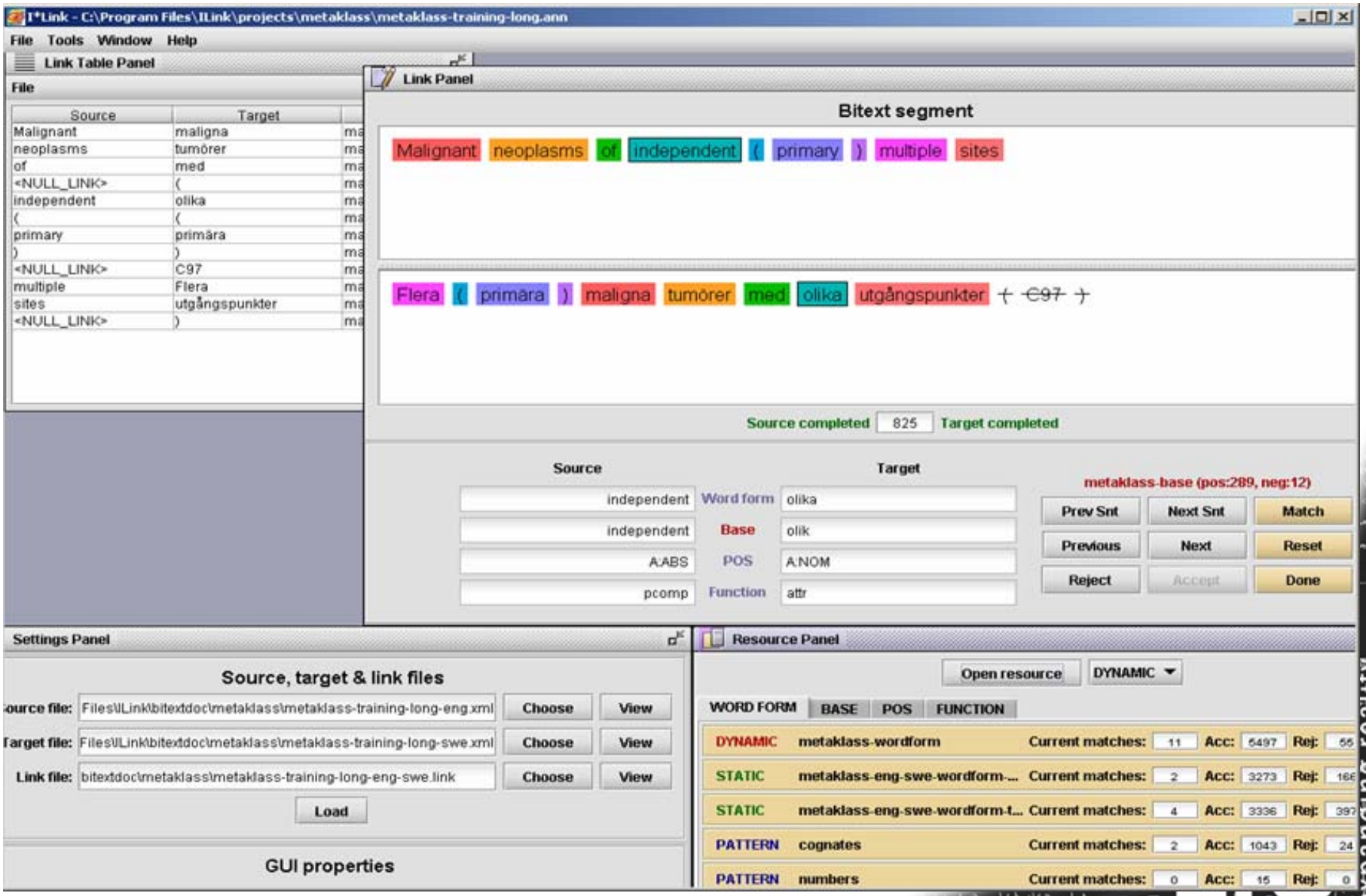
The ITools suite is used for the word alignment [2-4]. First morpho-syntactic tagging and statistical processing are performed. Then the user creates word alignment training data in an interactive environment. After that automatic alignment is done in a primarily linguistically oriented manner where different types of resources are combined in a voting procedure to select the best possible alignments.

The static resources used are dictionaries with a total of 22,500 general English-Swedish entries and pattern resources for common and unwanted parts-of-speech correspondences and standard punctuation.

The dynamic resources used are generated during the training stage. These resources are supposed to reflect the specific characteristics of the used material. The dynamic resources are built on four levels simultaneously, namely word form, lemma form, parts-of-speech and syntactic function levels.

The “statistic resources” used are probabilistic dictionaries produced using statistical analysis on the parallel texts with the GIZA++ tool kit.

Methods - Interactive word alignment tool



Methods - Preparations

The preferred English and officially translated Swedish rubrics were extracted from the used medical terminology systems. An exploration was then carried out to find partitions of the systems with assumed similar quality for the word alignment.

Considering the exploration the systems were divided into 8 partitions and ordered from partition 1 with the highest assumed quality to partition 8 with the lowest. Partition 1 was selected to be included in the static resources as a dictionary.

Statistical resources for the automatic word alignment were generated from each of the partitions 2–8.

From each of the partitions 2–8, 5 % were randomly sampled to a training set and 5 % to a test set. These sets were interactively aligned following a word alignment style guide. The training sets were used to build training resources and the test sets as gold standards for calculating the automatic word alignment's recall, precision and F-score.

Materials - Terminology partitions

Parti-tion	Content	Rubrics	English rubric average number (standard deviation) of words	Swedish rubric average number (standard deviation) of words
1	MeSH, one word in English or Swedish rubric	13,514	1.5 (0.7)	1.0 (0.1)
2	MeSH, more than one word in English and Swedish rubric	5,568	2.6 (0.8)	2.3 (0.7)
3	ICF, whole	1,496	4.7 (2.5)	4.2 (2.8)
4	KSH97-P, whole	968	4.0 (2.5)	3.5 (2.4)
5	ICD-10, except chapter 2 level 4	10,791	5.2 (3.0)	5.2 (3.4)
6	NCSP, except chapter N	4,137	5.8 (2.7)	5.0 (2.5)
7	ICD-10, chapter 2 level 4	713	3.6 (2.2)	6.3 (2.7)
8	NCSP, chapter N	1,388	9.4 (2.6)	7.7 (2.3)

Methods - Automatic word alignment

The test sets were automatically word aligned in 4 batches and the partitions used for building the resources and to automatically word align were altered. The same partitions were used to generate the statistic and dynamic resources. The main assumptions to examine in the batches are

1. The performance of the automatic word alignment when a partition is aligned with resources generated from the same partition. The performance reflects the similarities and differences inside each partition.

2. The performance of the automatic word alignment when generalizing resources from one partition to other partitions. The performance reflects the similarities and differences between a partition and the union of all partitions, which can be seen as an approximation of the medical language in healthcare records.

3. If the performance of the automatic word alignment improves monotonously when resources from more and more partitions are used or if some partitions worsen the performance.

4. The performance of the automatic word alignment for the different partitions when resources from all partitions are used. The performance reflects the similarities and differences between the union of all partitions and a partition.

Results

Batch	Run	Reso-urce parti-tion	Align parti-tion	Recall	Preci-sion	F-score
1	1	2	2	0.72	0.70	0.71
	2	3	3	0.80	0.80	0.80
	3	4	4	0.71	0.68	0.69
	4	5	5	0.80	0.78	0.79
	5	6	6	0.83	0.78	0.80
	6	7	7	0.63	0.58	0.60
	7	8	8	0.84	0.85	0.84
2	1	2	2–8	0.65	0.65	0.65
	2	3	2–8	0.62	0.63	0.62
	3	4	2–8	0.64	0.64	0.64
	4	5	2–8	0.75	0.73	0.74
	5	6	2–8	0.67	0.67	0.67
	6	7	2–8	0.64	0.65	0.64
	7	8	2–8	0.64	0.65	0.64

Batch	Run	Reso-urce parti-tion	Align parti-tion	Recall	Preci-sion	F-score
3	1	2	2–8	0.65	0.65	0.65
	2	2–3	2–8	0.67	0.66	0.66
	3	2–4	2–8	0.69	0.67	0.68
	4	2–5	2–8	0.77	0.74	0.75
	5	2–6	2–8	0.80	0.76	0.78
	6	2–7	2–8	0.80	0.77	0.78
	7	2–8	2–8	0.81	0.77	0.79
4	1	2–8	2	0.73	0.71	0.72
	2	2–8	3	0.81	0.81	0.81
	3	2–8	4	0.81	0.76	0.78
	4	2–8	5	0.81	0.78	0.79
	5	2–8	6	0.84	0.79	0.81
	6	2–8	7	0.74	0.65	0.69
	7	2–8	8	0.85	0.85	0.85

Discussion and Conclusion

Larger resource partitions generate resources that only to some extent have higher quality, so only adding larger resource partitions can be an expensive way to improve the results.

Partition 5, which contains the major parts of ICD-10, seems to cover the medical domain best of the included terminology systems. It can also be used to generate the best general resources of the included partitions. The large size of partition 5 is probably the main reason.

Systematic differences in structure between rubrics and their translations make the rubrics harder to align. The automatic word alignment can be improved by filtering out these rubrics.

Of the investigated partitions, two of them are harder to align than the others, namely partition 2 and 7.

References, acknowledgments and contact

[1] Nyström M, Merkel M, Ahrenberg L, Zweigenbaum P, Petersson H, Åhlfeldt H. Creating a medical English-Swedish dictionary using interactive word alignment. BMC Med Inform Decis Mak. 2006 October 12;6(35).
[2] Foo J, Merkel M. Building standardized term bases through automated term extraction and advanced editing tools. To be published in the Proceedings of the International Conference on Terminology; 2006 November 16-17; Antwerp; 2006.
[3] Ahrenberg L, Merkel M, Petterstedt M. Interactive word alignment for language engineering. In: Copestake A, Hajic J, editors. Tenth conference on European chapter of the Association for Computational Linguistics; 2003 April 12-17; Budapest; 2003. p. 49-52.
[4] Merkel M, Petterstedt M, Ahrenberg L. Interactive Word Alignment for Corpus Linguistics. In: Archer D, Rayson P, Wilson A, McEnery T, editors. Proceedings from Corpus Linguistics; 2003 March 29-31; Lancaster; 2003. p. 533-42.

This work was performed in the framework of the EU-funded Network of Excellence entitled Semantic Interoperability and Data Mining in Medicine (SemanticMining).

Mikael Nyström, Department of Biomedical Engineering, Linköping University, SE-581 85 Linköping, Sweden. mikny@imt.liu.se.