Reviewer’s report

Title: Creating a medical dictionary using word alignment: The influence of sources and resources

Version: 1 Date: 13 August 2007

Reviewer: Wendy W Chapman

Reviewer’s report:

General

This was an interesting manuscript with findings that could be useful to others trying to attempt the same task of aligning terminologies from different languages. The study was carried out very methodically, and the authors make interesting and seemingly valid conclusions. I have no suggestions or critiques for the study design. But I have several for the manuscript itself.

-------------------------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

There were three main problems with the manuscript. First, I was never completely sure I understood the overall task of this specific study. I wasn’t sure exactly what sources were being aligned with what other sources (e.g., MeSH English terms with ICD-10 Swedish terms, etc.). Second, there was too much information in this paper—the paper is very long, and I had a hard time keeping track of what was going on and what the results implied. Third, the study was carried out very methodically, and the paper was also presented that way. However, it is a complex study, and I think the paper could be revised to describe the study and its results more clearly. Below, I offer suggestions for the major critiques I mentioned above.

(1) Not clear what the overall task was

It would be helpful in the beginning of the Methods section to describe the overall study task in more detail. What was the input, what is the general word alignment algorithm, and what is the output? Is the output (ideally) filtered by a person or automatically added to the vocabulary? Did you map all English rubrics in all vocabularies you describe to all Swedish rubrics in all of them? Do all of them have Swedish rubrics too? What about those rubrics that are already mapped (e.g., UMLS C000581 may be the concept “pain” and may have a Swedish synonym) (you may answer some of these questions already, and I just got confused along the way). Please also show an example of the input and output (the examples in the appendix were helpful, but there should be an example in the text).

(2) Too much information
Here are some suggestions for shortening and focusing the manuscript:

a) Tables - highlight important points, show summary statistics – the tables are important, but a reader who looks only at the abstract, tables, and conclusion would not know what to make of the tables. Can you highlight important points you make in the text in the table itself (e.g., squares, bold, etc)? Also, average statistics over the different test set partitions would help the reader know the overall performance.

b) You can get away with deleting some of the details in the Alignment tools section.

c) The Terminology partitions section is too long. The discussion was interesting, but I think the point of the section is secondary to the main objective of the paper. A few paragraphs summarizing what you decided and why should be sufficient. I would suggest an online supplement describing the full partition statistics study for those who want the full details.

d) I suggest you reorganize the Methods and Discussion section based on questions rather than batches and partitions (described in (3) below). Such a reorganization should shorten the paper too.

e) The top section on page 19 (number of differences when a sequence of words is treated as a single name…”) could be deleted.

f) I suggest you re-read with the aim of shortening the paper and only keep essential sections and essential sentences

(3) Organization

a) On page 5 you separated the resources into two groups: static and dynamic. You did not explicitly point out what the relationship was between the static and dynamic categories and the static, statistic, and training categories. You also describe word alignment approaches as statistical and linguistic. Categorizing the same things in multiple ways may be too confusing for the reader. Also, I like the word “statistical” better than “statistic” which looked very similar to the word “static”—I had to reread them several times in the beginning to figure out that they were different words.

b) The Batch descriptions in the Methods section were long and were confusing to me. It may work better to organize this whole section by questions and map the batches to the questions they answer. You seemed to have various questions that motivated the batching, such as “how well can resources from a single partition work on the same partition?” and “How well to resources from one partition generalize to other partitions?” I believe structuring this section by the questions then *briefly* describing the batches you use to address the questions will shorten the paper and make it more clear in relation to your conclusions. Similarly, you could structure the discussion section by question rather than partitions and batches. The partitions and batches are only useful to the reader in how they answer the questions you are asking.

c) Although you mostly used consistent terminology throughout the paper, I got
confused and didn’t always understand what each term meant. It would be helpful to define the terms better and maybe even have a table with their definitions to refer to. Here are the terms that confused me:

• rubrics (you did define that, but on page 3, you used it before you defined it, and I didn’t know what it meant)
• sources
• resources
• partitions (in spite of the table showing the partitions, the term was confusing to me, because the word connotes one piece of a whole—not different vocabularies. You might consider changing the term partition to something like “vocabulary” to be more explicit. Even though some of the partitions are comprised of different parts of the same vocabularies, I still think the explicit label “vocabulary” would be better)
• runs
• batch (I was confused by the difference between a partition and a batch)

---------------------------------------------------------------------------------------------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

Abstract: There was a lot of jargon in the abstract, which made it difficult to understand without reading the entire paper. The conclusion was nice and to the point.

Abstract – there were quite a few typos in the abstract, including: “word alignment is depending on” # “word alignment is dependent on”; “This paper report on” # “This paper reports on”; “pairs were manual verified” # “pairs were manually verified”.

First sentence of the Methods paragraph in the Abstract was very confusing. Suggested change: “We automatically aligned terminology systems using static sources, like dictionaries, …”

The introduction was awkward. A few suggestions:
- change “with the recall of 0.77” to “with recall of 0.77” and change precision to be parallel.
- Rewrite last part of that paragraph to: “precision of 0.76, but we hypothesized we could exploit the highly repetitive structure in the terms to achieve even better results.”
- Put comma after phrase in first sentence in Objective: “In automatic word alignment, parallel…”
- Change “how the use of resources” to “how different resources”
- Change “This is done by altering which resources…” to “We accomplish this by altering the resources and medical terminology systems we use.”
Background – par. 1 - change “simply” to “simple”, change “there is” to “there are”, and delete “as translations contain addition of information as well as deletions.”

Bottom of page 5 – delete “It is worth pointing out that”

Top of page 6 – change to “should be applied over all the training material, because inconsistent…”

Middle of page 7 – change “candidate pairs for more” to “candidate pairs of more”

You sometimes use the passive voice, which makes the sentences harder to understand and allows you to leave out details like who did something. Change the passive voice throughout the paper to active voice. A few examples of important places to change is the Manual inspection section on page 7, the experiment setup on page 10, and the dictionary creation section on page 13—who performed the manual inspection, who evaluated the generated list, and who categorized the term-pair candidates?

The use of the terms “complements” on page 8 was confusing to me—I had to think through what the complement was: the rest of the rubrics or the different language.

You use the F-measure to measure reliability—are you familiar with the Hripcsak JAMIA paper that shows that the F-measure is equivalent to kappa when the (unknown) number of TN is large? It would be nice to reference this paper.

Abstract: There was a lot of jargon in the abstract, which made it difficult to understand without reading the entire paper. The conclusion was nice and to the point.

Abstract – there were quite a few typos in the abstract, including: “word alignment is depending on” # “word alignment is dependent on”; “This paper report on” # “This paper reports on”; “pairs were manual verified” # “pairs were manually verified”.

First sentence of the Methods paragraph in the Abstract was very confusing. Suggested change: “We automatically aligned terminology systems using static sources, like dictionaries, …”

The introduction was awkward. A few suggestions:

- change “with the recall of 0.77” to “with recall of 0.77” and change precision to be parallel.
- Rewrite last part of that paragraph to: “precision of 0.76, but we hypothesized we could exploit the highly repetitive structure in the terms to achieve even better results.”
- Put comma after phrase in first sentence in Objective: “In automatic word alignment, parallel…”
- Change “how the use of resources” to “how different resources”
- Change “This is done by altering which resources…” to “We accomplish this by altering the resources and medical terminology systems we use.”

Background – par. 1 - change “simply” to “simple”, change “there is” to “there are”, and delete “as translations contain addition of information as well as deletions.”

Bottom of page 5 – delete “It is worth pointing out that”
Top of page 6 – change to “should be applied over all the training material, because inconsistent…”
Middle of page 7 – change “candidate pairs for more” to “candidate pairs of more”

You sometimes use the passive voice, which makes the sentences harder to understand and allows you to leave out details like who did something. Change the passive voice throughout the paper to active voice. A few examples of important places to change is the Manual inspection section on page 7, the experiment setup on page 10, and the dictionary creation section on page 13—who performed the manual inspection, who evaluated the generated list, and who categorized the term-pair candidates?

The use of the terms “complements” on page 8 was confusing to me—I had to think through what the complement was: the rest of the rubrics or the different language.

You use the F-measure to measure reliability—are you familiar with the Hripcsak JAMIA paper that shows that the F-measure is equivalent to kappa when the (unknown) number of TN is large? It would be nice to reference this paper.

Discretionary Revisions (which the author can choose to ignore)

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Needs some language corrections before being published

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:
I declare that I have no competing interests