Corpus-based Error Detection in a Multilingual Medical Thesaurus

Roosewelt L. Andrade[a,b], Edson Pacheco[a,b], Pindaro S. Cancian[a,b], Percy Nohama[a,b], Stefan Schulz[b,c]

[a] Paraná University of Technology (UTFPR), Curitiba, Brazil

[b] Pontificial Catholic University of Paraná, Master Program of Health Technology, Curitiba, Brazil

[c] University Medical Center Freiburg, Medical Informatics, Freiburg, Germany

PUCPR

UNIVERSITÄTS FREIBURG KLINIKUM

Thesaurus

- Controlled Vocabulary for document indexing and retrieval

- Assigns semantic descriptors (concepts) to (quasi-)synonymous terms

- Contains additional semantic relations (e.g. hyperonym / hyponym)

- Examples: MeSH, UMLS, WordNet

- Multilingual thesaurus: contains translations (cross-language synonymy links)

Multilingual Thesaurus Management

- International team of curators

- React to new terms and senses

- Decide which terms are synonymous / translations

- Decide which senses of a term have to be accounted for in the domain

- Requires quality assurance measures

Case study: Morphosaurus

- Medical subword thesaurus
- Organizes subwords (meaningful word fragments) in multilingual equivalence classes:

  – #derma   = { derm, cutis, skin, haut, kutis, pele, cutis, piel, … }
  – #inflamm = { inflamm, -itic, -itis, phlog, entzuend, -itis, -itisch, inflam, flog, inflam, flog, ... }

- Maintained at two locations: Freiburg (Germany), Curitiba (Brazil)
- Lexicon curators: frequently changing team of medical students

Morphosaurus Structure

- Thesaurus: ~21.000 equivalence classes (MIDs)

- Lexicon entries:
  - English: ~23.000
  - German: ~24.000
  - Portuguese: ~15.000
  - Spanish :   ~11.000
  - French: ~ 8.000
  - Swedish: ~10.000
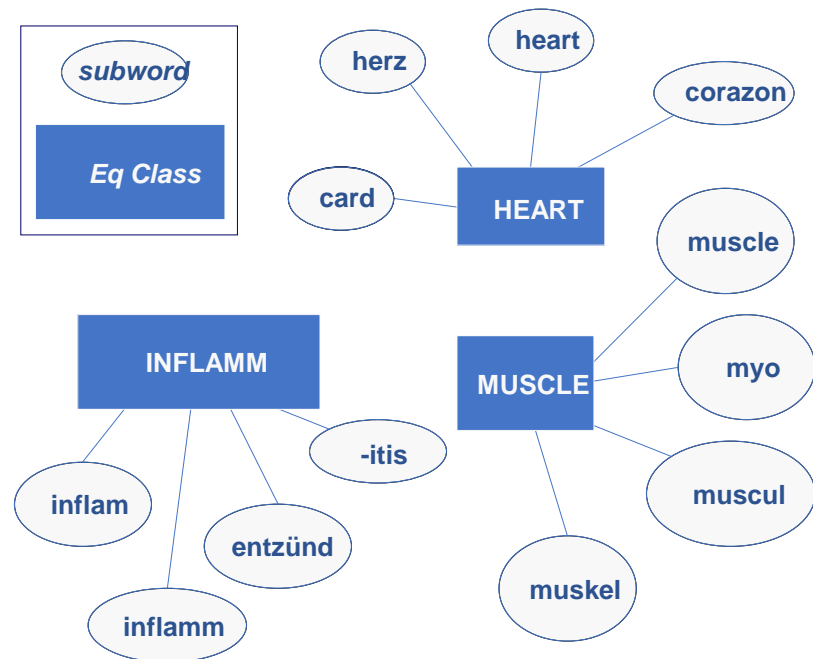
Segmentation:

Myo | **kard** | **itis**

Herz | **muskel** | **entzünd** | ung

Inflamm | ation of the **heart musc**

Indexation:

#muscle #heart #inflamm

#heart #muscle #inflamm

#inflamm #heart #muscle

Morphosemantic Normalization

Original Document	Orthographic Normalization	Morphological Segmentation	Semantic Normalization
High TSH values suggest the diagnosis of primary hypothyroidism while a suppressed TSH level suggests hyperthyroidism.	high tsh values suggest the diagnosis of primary hypothyroidism while a suppressed tsh level suggests hyperthyroidism.	high tsh value s suggest the diagnos is of primar y hypo thyroid ism while a suppress ed tsh level suggest s hyper thyroid ism.	**#up#** tsh **#value#** #suggest# **#diagnost#** **#primar#** **#small#** **#thyre#** **#suppress#** tsh **#nivell#** #suggest# **#up#** **#thyre#** .
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose, ein supprimierter TSH-Spiegel spricht dagegen für eine Schilddrüsenüberfunktion.	erhoehte tsh-werte erlauben die diagnose einer primaeren hypothyreose, ein supprimierter tsh-spiegel spricht dagegen fuer eine schilddruesen-ueberfunktion.	er hoeh te tsh - wert e erlaub en die diagnos e einer primaer en hypo thyre ose, ein supprim iert er tsh - spiegel spricht dagegen fuer eine schilddrues en ueber funktion.	**#up#** tsh - **#value#** #permit# **#diagnost#** **#primar#** **#small#** **#thyre#** , **#suppress#** tsh - {#mirrorr# **#nivell#**} #speak# **#thyre#** **#up#** #function# .
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário, enquanto níveis suprimidos de TSH sugerem hipertireoidismo.	a presenca de valores elevados de tsh sugere o diagnostico de hipotireoidismo primario, enquanto niveis suprimidos de tsh sugerem hipertireoidismo.	a presenc a de valor es elevad os de tsh suger e o diagnost ico de hipo tireoid ismo primari o, enquanto niveis suprimid os de tsh suger em hiper tireoid ismo.	#actual# **#value#** **#up#** tsh #suggest# **#diagnost#** **#small#** **#thyre#** **#primar#** , **#nivell#** **#suppress#** tsh #suggest# **#up#** **#thyre#** .

Additional Challenges for Morphosaurus

- Properly delimit subword entries so that they are correctly extracted from complex words

- Create consensus about the scope of synonymy classes, especially with regard to highly ambiguous lexicon entries

Morphosaurus Quality Assurance

- Content quality: Identify content errors in the thesaurus content
  - Implement a quality process to fix these errors
  - Show positive impact
- Process quality: Detect and prevent user action anomalies
  - actions that consume effort without any positive impact : uncoordinated edit / update / delete "do undo" transactions done by different people)
  - see Paper Bitencourt *et al.*, Session 089

Testbed: Parallel Medical Corpora

- Apply morphosemantic indexing to the Merck manual in English, Spanish, Portuguese, German

- Hypothesis: nearly identical frequency distribution of Morphosaurus identifiers (MIDs) for each language

- Problematic MIDs can be spotted by comparing MID frequencies for each language pair

Testbed: Parallel Medical Corpora

Original Document	Orthographic Normalization	Morphological Segmentation	Semantic Normalization
High TSH values suggest the diagnosis of primary hypothyroidism while a suppressed TSH level suggests hyperthyroidism.	high tsh values suggest the diagnosis of primary hypothyroidism while a suppressed tsh level suggests hyperthyroidism.	high tsh value s suggest the diagnos is of primar y hypo thyroid ism while a suppress ed tsh level suggest s hyper thyroid ism.	**#up#** tsh **#value#** #suggest# **#diagnost#** **#primar#** **#small#** **#thyre#** **#suppress#** tsh **#nivell#** #suggest# **#up#** **#thyre#** .
Erhöhte TSH-Werte erlauben die Diagnose einer primären Hypothyreose, ein supprimierter TSH-Spiegel spricht dagegen für eine Schilddrüsenüberfunktion.	erhoehte tsh-werte erlauben die diagnose einer primaeren hypothyreose, ein supprimierter tsh-spiegel spricht dagegen fuer eine schilddruesen-ueberfunktion.	er hoeh te tsh - wert e erlaub en die diagnos e einer primaer en hypo thyre ose, ein supprim iert er tsh - spiegel spricht dagegen fuer eine schilddrues en ueber funktion.	**#up#** tsh - **#value#** #permit# **#diagnost#** **#primar#** **#small#** **#thyre#** , **#suppress#** tsh - {#mirrorr# **#nivell#**} #speak# **#thyre#** **#up#** #function# .
A presença de valores elevados de TSH sugere o diagnóstico de hipotireoidismo primário, enquanto níveis suprimidos de TSH sugerem hipertireoidismo.	a presenca de valores elevados de tsh sugere o diagnostico de hipotireoidismo primario, enquanto niveis suprimidos de tsh sugerem hipertireoidismo.	a presenc a de valor es elevad os de tsh suger e o diagnost ico de hipo tireoid ismo primari o, enquanto niveis suprimid os de tsh suger em hiper tireoid ismo.	#actual# **#value#** **#up#** tsh #suggest# **#diagnost#** **#small#** **#thyre#** **#primar#** , **#nivell#** **#suppress#** tsh #suggest# **#up#** **#thyre#** .

Scoring of MID Imbalance

f_1, f_2 frequencies of MIDs in language 1 and 2

degree of imbalance

$$S_d = \frac{|f_1 - f_2|}{|f_1 + f_2|}$$

mean relative frequency

$$S_a = \left(\frac{f_1 + f_2}{f_{x1} + f_{x2}}\right)_{\max}$$

$$S = \frac{2S_d + S_a}{3}$$

Score used for MID ranking per language pair

For each l_1, l_2: MIDs ranked by Score

```
| MID                        | MIDCod  | f1    | f2    | Sa     | Sd     | S      |
| peopleriixypa              | 500783  | 6352  | 0     | 0,1466 | 1,0000 | 0,7155 |
| fromiwiixxa                | 060077  | 4676  | 0     | 0,1079 | 1,0000 | 0,7026 |
| icasikprrr                 | 023555  | 0     | 3022  | 0,0697 | 1,0000 | 0,6899 |
| lttroriiyira               | 500805  | 10    | 3331  | 0,0771 | 0,9940 | 0,6884 |
| mostiizrpwa                | 009536  | 2783  | 0     | 0,0642 | 1,0000 | 0,6881 |
| enteikywjw                 | 028616  | 0     | 2069  | 0,0477 | 1,0000 | 0,6826 |
| icakiirwy                  | 200568  | 0     | 1945  | 0,0449 | 1,0000 | 0,6816 |
| sometimerijixja            | 501071  | 1708  | 0     | 0,0394 | 1,0000 | 0,6798 |
| pressureiiipkza            | 000329  | 1833  | 2     | 0,0423 | 0,9978 | 0,6793 |
| ciamikpzzp                 | 023993  | 0     | 1461  | 0,0337 | 1,0000 | 0,6779 |
| likelihoodrijjzwa          | 501196  | 905   | 0     | 0,0209 | 1,0000 | 0,6736 |
| askijrkja                  | 201521  | 889   | 0     | 0,0205 | 1,0000 | 0,6735 |
| ppippwra                   | 303365  | 6     | 1508  | 0,0349 | 0,9921 | 0,6730 |
| overipjqqka                | 031442  | 798   | 0     | 0,0184 | 1,0000 | 0,6728 |
| relievrijkqxa              | 501247  | 618   | 0     | 0,0143 | 1,0000 | 0,6714 |
| mientorijwyp               | 501683  | 0     | 529   | 0,0122 | 1,0000 | 0,6707 |
| physisijqkiia              | 014200  | 501   | 0     | 0,0116 | 1,0000 | 0,6705 |
| logikipxzyw                | 037986  | 0     | 495   | 0,0114 | 1,0000 | 0,6705 |
```

Experimental Modification of Workflow

- Discussion of 100 most highly ranked MID imbalances for each language pair

- Classification of Problems

- Documentation

- Correction in consensus

Problems identified

Reason for MID high score	Portuguese / English	German / English	Spanish / English
Ambiguous lexemes	0.23	0.38	0.14
Missing or dispensable MID	0.49	0.18	0.53
Same Sense in Different MIDs	0.06	0.12	0.19
One MID with Different Senses	0.04	0.05	0.06
No problem	0.11	0.10	0.04
Unclassified problem	0.07	0.17	0.04

Modified Workflow

•Discussion of highly ranked MID imbalances

Summative Evaluation

- Did the targeted error spotting and resolution have an impact on the "in vivo" performance of the thesaurus ?

- Benchmark: OHSUMED document collection (user queries with relevant MEDLINE abstract assigned), queries translated to German, Spanish, Portuguese

- Target: monitoring of Timelines of Eleven Point Average measure (precision values at recall *0, 0.1, 0.2,…0.9, 1.0*)

Summative Evaluation Results
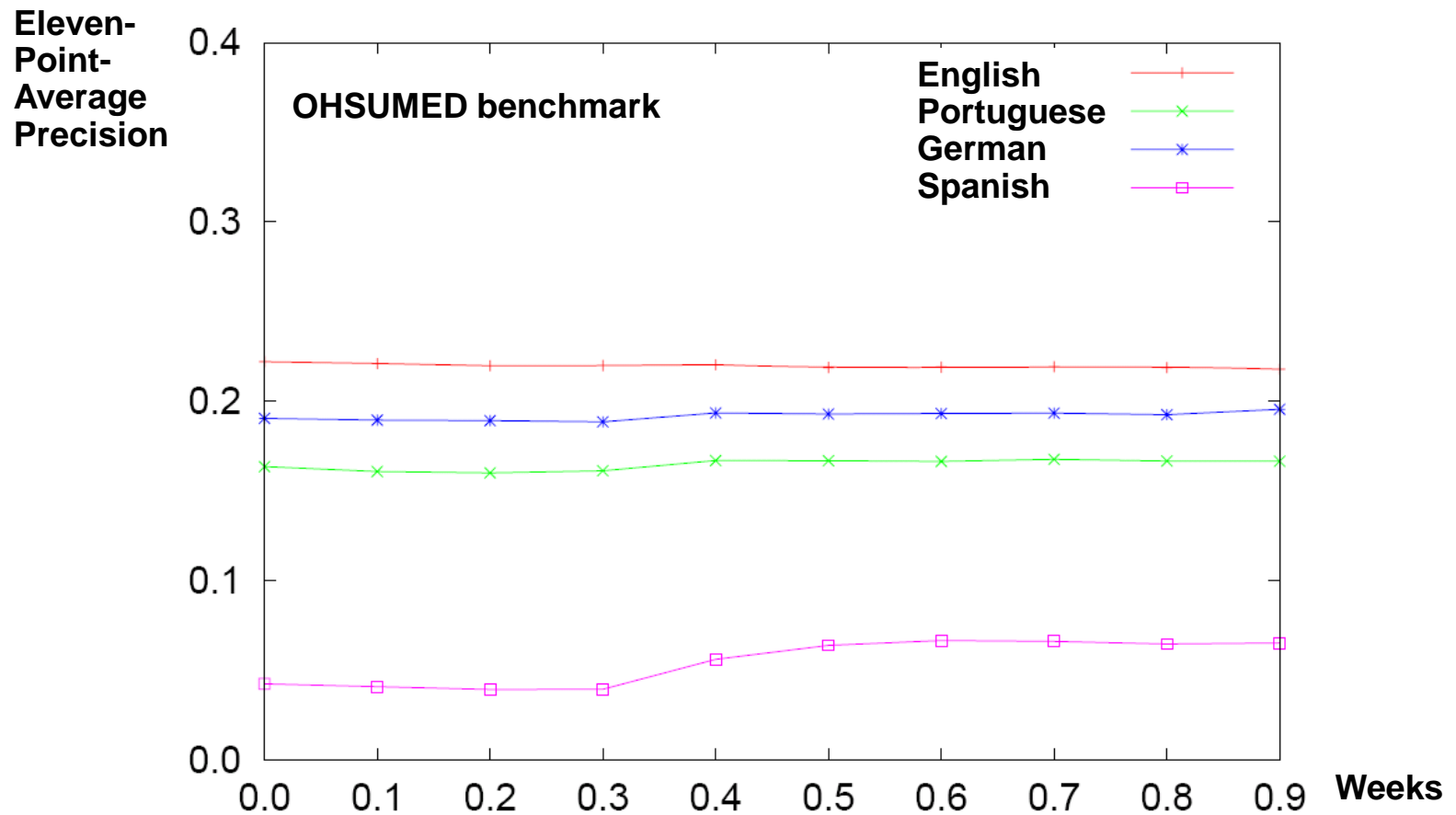
Discussion of Results

- Most problems detected corresponded to real errors (approx. 90%)

- Consensus could be found in most cases

- "In vivo" evaluation showed a clear increase in IR performance for Spanish, at the time the less consolidated language in Morphosaurus

Conclusions / Recommendations

- Error detection by MID imbalance score proved useful

- New workflow productive for error elimination

- Recommendations
  - Record MID score over time
    - avg for each language pair
    - avg for each language
    - for each MID
  - Generate alerts for every MIDs which exhibits an increase in imbalance above a tolerance interval
  - Continue monitoring AvgP11 values
  - Include into Morphosaurus editing environment

Medical Thesaurus Anomaly Detection by User Action Monitoring
Session 089

Jeferson L. Bitencourt [a,b], Pindaro S. Cancian[a,b], Edson Pacheco[a,b], Percy Nohama[a,b], Stefan Schulz[b,c]

[a] Paraná University of Technology (UTFPR), Curitiba, Brazil

[b] Pontificial Catholic University of Paraná, Master Program of Health Technology, Curitiba, Brazil

[c] University Medical Center Freiburg, Medical Informatics, Freiburg, Germany