Assisting the translation of the CORE Subset of SNOMED CT into French

Hocine ABOUDOUNE, Tayeb MERABTI, Stéfan J. DARMONI, Michel JOUBERT

LERTIM, Faculty of Medicine, University of Aix-Marseille 2, France
CISMeF, Rouen University Hospital, France
Background

• Problem list is part of the problem-oriented health record
  – Each institution owns its problem-list vocabulary

• UMLS problem list CORE Subset
  – CORE stands for Clinical Observations Recording and Encoding
  – 14,000 terms coming from 7 vocabularies are mapped to 6,800 UMLS concepts

• CORE Subset of SNOMED CT
  – 81% are mapped to SNOMED CT
Objective

• To propose an automated method to assist a translation of the CORE Subset (of SNOMED CT) into French
Material (1/2)

• UMLS Metathesaurus
  – MRCONSO
  – MRREL
    • primary_mapped_to / from
    • mapped_to / from
    • other_mapped_to / from

• CORE Subset of SNOMED CT
  – disorder: 3,794 concepts
  – finding: 752 concepts
  – procedure: 396 concepts
Material (2/2)

• Our mappings operate exclusively on preferred terms (PTs), but we exploit synonyms

• French terminologies
  – SNOMED International: 107,900 PTs
  – ICD10: 9,306 PTs
  – MeSH: 25,186 PTs
  – MedDRA: 18,209 PTs
Method

CORE Subset

MRCONSO

MRREL

Legend

Exact match

Explicit mapping

SNOMED Int

ICD10

MeSH

MedDRA

Michel JOUBERT - LERTIM
## Results (1/2)

Number and percentage of each terminology by axis in the CORE Subset

<table>
<thead>
<tr>
<th>Terminologies</th>
<th>Disorder</th>
<th>Finding</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMED Int.</td>
<td>3,277 / 86%</td>
<td>522 / 69%</td>
<td>262 / 66%</td>
</tr>
<tr>
<td>ICD10</td>
<td>2,733 / 72%</td>
<td>477 / 63%</td>
<td>7 / 2%</td>
</tr>
<tr>
<td>MeSH</td>
<td>2,151 / 57%</td>
<td>364 / 48%</td>
<td>118 / 30%</td>
</tr>
<tr>
<td>MedDRA</td>
<td>2,505 / 66%</td>
<td>495 / 66%</td>
<td>162 / 41%</td>
</tr>
</tbody>
</table>
Results (2/2)

Number of PTs in the union of the four terminologies aligned with PTs of the CORE Subset

<table>
<thead>
<tr>
<th>Axes</th>
<th># of PTS of the 4 Terminologies</th>
<th># of PTS of the CORE Subset</th>
<th>% of PTS of the Core Subset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disorder</td>
<td>3,463</td>
<td>3,794</td>
<td>91%</td>
</tr>
<tr>
<td>Finding</td>
<td>632</td>
<td>752</td>
<td>84%</td>
</tr>
<tr>
<td>Procedure</td>
<td>291</td>
<td>396</td>
<td>73%</td>
</tr>
<tr>
<td>Total</td>
<td>4,386</td>
<td>4,942</td>
<td>89%</td>
</tr>
</tbody>
</table>
Discussion

• Our method produces at least one proposal for the translation of 4,386 of the 4,942 CORE Subset concepts (89%)

• Synonymy is a symmetric relationship between terms
  – if two terms share the same concept identifier, the synonyms of the first are considered as synonyms of the second one, and conversely
  – for a given concept, the method builds a set of terms made of preferred terms and synonyms originating from different terminologies
Conclusion

• The automated method we propose for assisting the translation of the CORE Subset works at a conceptual level
  – is not dependent on French
  – can be reused for another natural language than French
    • on condition that the number of terminologies in this language integrated in the UMLS is sufficient
The end

Thank you for your attention