



Terminologies and Ontologies
July 4, 2011

Improving the mapping between MedDRA and SNOMED CT

Fleur Mougin

LESIM, INSERM U897, ISPED
University Bordeaux Segalen, France



Natalia Grabar
STL, CNRS UMR 8163,
University Lille 3, France



Marie Dupuch
CRC, University Pierre et Marie Curie,
INSERM U872, Paris, France



Introduction

- Spontaneous reporting of ADRs depends on healthcare professionals → small proportion of the existing ADRs
- Solutions: exploring clinical reports
- Requirements:
 - To link clinical reports with pharmacovigilance databases
 - To map SNOMED CT with MedDRA
- Existing mapping in the UMLS: 42%
 - Objective: improving this mapping through an automatic lexical-based approach

Resources

- MedDRA: 86,842 terms structured into 5 hierarchical levels
- SNOMED CT: 291,205 current concepts (750,880 synonyms) compositional and follow the post-coordination approach
- UMLS[®] (2010AA)
 - Metathesaurus[®]
 - Over 150 source vocabularies (incl. MedDRA and SNOMED CT)
 - > 2 million concepts (clusters of synonymous terms)
 - Semantic Network
 - 133 semantic types (ST) organized in a tree structure
 - Aggregated into 15 coarser semantic groups (SGs)
 - Each Metathesaurus concept has a unique identifier and is assigned at least one ST

Methods

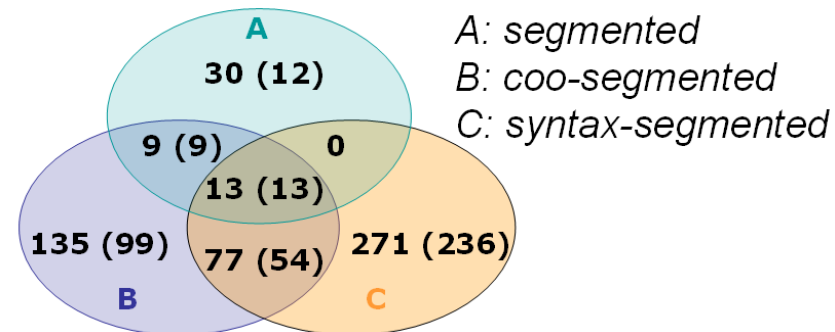
- Preparing and mapping the terms
 - MedDRA terms = from UMLS concepts without SNOMED CT term
 - Lexical approach applied to terms
 - Segmentation into words
 - Normalization (punctuation, derived forms, synonyms, ...)
 - Direct mapping + Mapping after a decomposition on stopwords + Mapping after a decomposition on stopwords with a special processing of the coordination
- Filtering mappings according to their SGs
 - Possible for 1-1 mappings (a MedDRA term for a SNOMED CT concept)
 - Elimination if the SG of the MedDRA term \neq SG of the SNOMED CT concept
- Evaluating the mappings
 - Quantitatively:
 - Number of 1-1 mappings and full mappings (all MedDRA components could all be mapped to one or more SNOMED CT concepts)
 - Comparison of full mappings obtained by the three segmentation sets
 - Qualitatively: assessment of the quality of mapping as “correct”, “incorrect”, or “hierarchically-related”

Results

- Mapping: 30,023 MedDRA terms (23,102 UMLS concepts)

	Direct	Segmented	Coo-segmented	Syntax-segmented
# of MDR components		28,227	30,116	21,056
# of full mappings		52	234	361
# of 1-1 mappings	308	10	211	137

- Direct approach: 199 correct mappings (64.6%), 45 incorrect (14.6%), and 64 hierarchically-related (20.8%)
- Comparing the segmentation approaches



Discussion

- Findings
 - New and correct mappings: more complete mapping between MedDRA and SNOMED CT
 - Compositionality of the MedDRA terms
- Limitations
 - Use of NLP tools may cause wrong segmentations
→ incorrect mappings
 - Synonymous pairs may provide a correct link in some but not in all the contexts
- Benefits
 - Exploitation of the SGs: useful to eliminate wrong mappings (1/4)
 - Identification of inconsistencies in the UMLS

Thanks for your attention

fleur.mougin@isped.u-bordeaux2.fr

marie.dupuch@crc.jussieu.fr

natalia.grabar@univ-lille3.fr