Creating Models for Specialized Definitions

Selja Seppälä
University at Buffalo

November 20, 2013
Programa de Posgrado en Lingüística
Facultad de Letras
Universidad de Costa Rica
Outline

• Background
• Introducing the problem
• Solutions to the problem
• Proposal: modeling definition contents with BFO
• Methodology
• Conclusion
BACKGROUND
Object of study (1)

Dictionary of Banking

This updated edition of the «UBS dictionary of banking» contains definitions of over 2,400 terms, incorporating the latest information available from our Research department. This represents an increase – compared to the highly successful 2000 edition – of some 280 terms, the majority drawn from the area of wealth management. In future, the dictionary will be updated and expanded on an ongoing basis, so you will continue to benefit free of charge from UBS’s vast expertise in financial theory and practice.

Our aim in preparing this dictionary is to share the vocabulary of banking with our clients. We are convinced that this will make a major contribution to our mutual understanding, by helping to ensure that we share a common language.

The "UBS dictionary of banking" is only provided electronically.

AAA / AAA / AAA / AAA
Highest debt grade in the rating system used by Standard & Poor's for securities in the international financial markets.

ABS / ABS / ABS / ABS
Abbr. for asset-backed security.

ADR / ADR / ADR
American Depositary Receipt. Negotiable registered certificate issued on the US market and evidencing title to non-US equity paper. ADRs are registered with the Securities and Exchange Commission (SEC) and quoted in USD. Holders of ADRs essentially enjoy the same ownership and membership rights as shareholders.
Object of study (2)
Definition of ‘definition’

• A short sentence
• Conveys information about the meaning of domain-specific terms
• Composed of ‘features’ (pieces of information) found in experts’ texts
• Has one genus, and one or more specifiers (differentiae)
• Has a logical form
  – Necessary and jointly sufficient features
  – Typical features
Examples of definitions

ANATOMICAL ENTITY ONTOLOGY

**ligament**

A strap of predominantly extracellular matrix connecting two bones.

OIL SPILL CLEANUP

**net boom**

A boom that is made of netting to facilitate the retention of viscous oils.
Definition writing rules, issues, and needs

• Subject to a few formal rules (punctuation, etc.)
• Only vague principles for selecting the right contents
• Time-consuming, costly, and prone to all kinds of inconsistencies

➔ Create computer software to help authors
  – Comply to the publisher’s formal rules (Seppälä 2006)
  – Include relevant defining contents
Definition writing rules, issues, and needs

• Subject to a few formal rules (punctuation, etc.)
• Only vague principles for selecting the right contents
• Time-consuming, costly, and prone to all kinds of inconsistencies

➔ Create computer software to help authors
  – Comply to the publisher’s formal rules (Seppälä 2006)
  – Include relevant defining contents
INTRODUCING THE PROBLEM
The definition writing activity

• Constituting a corpus of experts’ texts
• Extracting terms and any related information
• Identifying potentially defining information
• Constructing a definition with the relevant pieces of information
5.3.2 Net booms

Netting can be used as a boom to collect viscous oils at sea. Vessels deploy the netting, manoeuvring it to surround and concentrate the oil for recovery. In inshore waters, it may be possible to use nets as a moored boom. Net booming is based on the principle that air and water will pass through the net but not viscous oil. Therefore, there is little or no downward current to carry oil underneath the net and the net is less prone to being blown flat. Because of the lower resistance of nets to water movement, it should also be possible in theory to deploy net booms in faster currents than is possible with conventional booms. Net booms have not yet been fully tested in an actual oil spill but results from field trials have been encouraging.

A net boom consists of a long strip of netting, supported at frequent and regular intervals by poles with floats and weights attached to keep the netting upright (Figure 20).

(Source: http://www.dft.gov.uk/mca/chapter_5.pdf)
5.3.2 Net booms

Netting can be used as a boom to collect viscous oils at sea. Vessels deploy the netting, manoeuvring it to surround and concentrate the oil for recovery. In inshore waters, it may be possible to use nets as a moored boom. Net booming is based on the principle that **air and water will pass through the net but not viscous oil**. Therefore, there is little or no downward current to carry oil underneath the net and the net is less prone to being blown flat. Because of the lower resistance of nets to water movement, it should also be possible in theory to deploy net booms in faster currents than is possible with conventional booms. Net booms have not yet been fully tested in an actual oil spill but results from field trials have been encouraging.

A net boom consists of a long strip of netting, supported at frequent and regular intervals by poles with floats and weights attached to keep the netting upright (Figure 20).

(Source: http://www.dft.gov.uk/mca/chapter_5.pdf)
5.3.2 Net booms

Netting can be used as a boom to collect viscous oils at sea. Vessels deploy the netting, manoeuvring it to surround and concentrate the oil for recovery. In inshore waters, it may be possible to use nets as a moored boom. Net booming is based on the principle that air and water will pass through the net but not viscous oil. Therefore, there is little or no downward current to carry oil underneath the net and the net is less prone to being blown flat. Because of the lower resistance of nets to water movement, it should also be possible in theory to deploy net booms in faster currents than is possible with conventional booms. Net booms have not yet been fully tested in an actual oil spill but results from field trials have been encouraging.

A net boom consists of a long strip of netting, supported at frequent and regular intervals by poles with floats and weights attached to keep the netting upright (Figure 20).

(Source: http://www.dft.gov.uk/mca/chapter_5.pdf)
Example: selecting relevant information for defining the term

5.3.2 Net booms

Netting can be used as a boom to collect viscous oils at sea. Vessels deploy the netting, manoeuvring it to surround and concentrate the oil for recovery. In inshore waters, it may be possible to use nets as a moored boom. Net booming is based on the principle that air and water will pass through the net but not viscous oil. Therefore, there is little or no downward current to carry oil underneath the net and the net is less prone to being blown flat. Because of the lower resistance of nets to water movement, it should also be possible in theory to deploy net booms in faster currents than is possible with conventional booms. Net booms have not yet been fully tested in an actual oil spill but results from field trials have been encouraging.

A net boom consists of a long strip of netting, supported at frequent and regular intervals by poles with floats and weights attached to keep the netting upright (Figure 20).

(Source: http://www.dft.gov.uk/mca/chapter_5.pdf)
Example: composing the definition

5.3.2 Net booms

Netting can be used as a boom to collect viscous oils at sea. Vessels deploy the netting, manoeuvring it to surround and concentrate the oil for recovery. In inshore waters, it may be possible to use nets as a moored boom. Net booming is based on the principle that air and water will pass through the net but not viscous oil. Therefore, there is little or no downward current to carry oil underneath the net and the net is less prone to being blown flat. Because of the low resistance of nets to water movement, it should also be possible in theory to deploy net booms in faster currents than is possible with conventional booms. Net booms have not yet been fully tested in an actual oil spill but results from field trials have been encouraging.

A net boom consists of a long strip of netting, supported at frequent and regular intervals by poles with floats and weights attached to keep the netting upright (Figure 20).

(Source: http://www.dft.gov.uk/mca/chapter_5.pdf)
Observations

Not all pieces of information found in experts’ texts can be:

• considered as potentially defining
• selected as relevant in a definition
Questions raised by definition writing

1. How to distinguish between defining and non-defining features?
2. Among the potentially defining features, how to select the relevant ones?
3. If several genus candidates, which one to select?
Summary of the questions
Summary of the questions

- Floating Net Booms are used for a wide variety of applications across various industries.

(http://www.elastec.com/customcontainment/netboom/)
Because of the lower resistance of nets to water movement, it should also be possible in theory to deploy net booms in faster currents …

Net booms have not yet been fully tested in an actual oil spill but results from field trials have been encouraging.
Summary of the questions

- a boom to collect viscous oils at sea
- air and water will pass through the net but not viscous oil
- consists of a long strip of netting,
- supported at frequent and regular intervals by poles with floats and weights attached to keep the netting upright
Summary of the questions

- Netting
- a boom
Summary of the questions

A boom that is made of netting to facilitate the retention of viscous oils.
A content selection problem

- How to select the relevant kind of definition content?
  ➔ Content selection principles

- How to create rules to help authors select definition contents?
  ➔ Content modeling: represent the contents in a way that
    - Is understandable for definition authors
    - Can be implemented in a computer system
SOLUTIONS TO THE PROBLEM
In practice

• Manuals: general principles on non-circularity and non-tautological definitions
• International standards (ISO704, 1987):
  – Intrinsic: form, dimension, matter, color
  – Extrinsic: provenance, function, location, discoverer, inventor, position
• In practice: experts’ help, knowledge of the domain, background knowledge, experience, intuition
  ➔ Insufficient
In theory: ‘quantitative’ approaches

• Necessary and sufficient conditions
• Prototypical or typical features
  ➔ Account for the relationship between the intension and the extension (logical property)
  ➔ Non-operational rules
In theory: ‘qualitative’ approaches

• Relational approaches (Sager, 1990)
  – Conceptual and other approaches of ontological inspiration
  – Lexico-semantic approaches
  – Cognitive approaches
  – Pragmatic approach

➔ Focus on a limited number of selection principles

➔ Methodological solutions yielding models that are not generalizable
But solution to content representation

Represent contents with models based on categories and relations

**ligament** *(Anatomical Entity Ontology)*

A strap of predominantly extracellular matrix connecting two bones.
PROPOSAL: MODELING DEFINITION CONTENTS WITH BFO
A guiding question

• Definition contents may be influenced by different factors (conceptual, linguistic, cognitive...)

➔ Is there one that can be used as a basis to predict the contents of definitions independently of their domain and language?
Content selection principles

• Extensional dimension
  – Ontological factors
    • Category of the referent (OBJECT, PROCESS, QUALITY, etc.)
    • Type of things or particular (particle accelerator, the LHC)

• Contextual dimension
  – Conceptual system of the domain
  – Background knowledge of the user

• Communicative dimension
  – User needs
  – Communication situation (linguistic expression: quadruped vs. four legged)
Content selection principles

- Extensional dimension
  - Ontological factors
  - Category of the referent (OBJECT, PROCESS, QUALITY, etc.)
  - Type of things or particular (particle accelerator, the LHC)

- Contextual dimension
  - Conceptual system of the domain
  - Background knowledge of the user

- Communicative dimension
  - User needs
  - Communication situation (linguistic expression: quadruped vs. four legged)
Hypothesis

• Content selection partly depends on the type of entity that is the referent of the definition

• Questions:
  – What kinds of entities are there?
  – What are their characteristics?

➔ Philosophers must know!
➔ Use a philosophically motivated ontology
Proposal

• Create definition models based on the BFO categories and their characteristics
• See to what extent these models predict the contents of definitions

➔ ‘Ontological analysis framework’
Why BFO?

• Definitions express relations between things in the world
  ➔ Realist ontology
    (vs. conceptual representation)

• Domain- and language-independent models

• Use of metalanguage (categories and relations)
  ➔ Upper-level ontology
    (vs. applied, domain-specific ontology)
Example: modeling definitions of OBJECTS

**OBJECT**

<table>
<thead>
<tr>
<th>has_part OBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>participates_in PROCESS</td>
</tr>
</tbody>
</table>

**INDEPENDENT CONTINUANT**

<table>
<thead>
<tr>
<th>bearer_of QUALITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>bearer_of REALIZABLE ENTITY</td>
</tr>
<tr>
<td>located_at TEMPORAL REGION</td>
</tr>
<tr>
<td>located_in SITE</td>
</tr>
</tbody>
</table>

**ligament (AEO_0000090)**
(Anatomical Entity Ontology)

<table>
<thead>
<tr>
<th>GEN</th>
<th>is_a OBJECT</th>
<th>A strap</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE</td>
<td>has_part OBJECT</td>
<td>of predominantly extracellular matrix</td>
</tr>
<tr>
<td>SPE</td>
<td>bearer_of REALIZABLE ENTITY</td>
<td>connecting two bones.</td>
</tr>
</tbody>
</table>

**net boom (oil spill cleanup)**

<table>
<thead>
<tr>
<th>GEN</th>
<th>is_a OBJECT</th>
<th>A boom</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE</td>
<td>has_part OBJECT</td>
<td>that is made of netting</td>
</tr>
<tr>
<td>SPE</td>
<td>bearer_of REALIZABLE ENTITY</td>
<td>to facilitate the retention of viscous oils.</td>
</tr>
</tbody>
</table>
Example: modeling definitions of
OBJECTS

<table>
<thead>
<tr>
<th>OBJECT</th>
<th>distilled mustard</th>
<th>sulfur mustard</th>
</tr>
</thead>
<tbody>
<tr>
<td>REF</td>
<td>is_a OBJECT</td>
<td></td>
</tr>
<tr>
<td>GEN</td>
<td>bearer_of</td>
<td>REALIZABLE ENT</td>
</tr>
<tr>
<td>SPE</td>
<td>bearer_of QUALITY</td>
<td>which is an amber brown liquid with an odor similar to that of burning garlic.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INDEPENDENT CONTINUANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>bearer_of QUALITY</td>
</tr>
<tr>
<td>bearer_of REALIZABLE ENT</td>
</tr>
<tr>
<td>located_at TEMPORAL REGION</td>
</tr>
<tr>
<td>located_in SITE</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>pump system (oil spill cleanup)</th>
</tr>
</thead>
<tbody>
<tr>
<td>REF</td>
</tr>
<tr>
<td>GEN</td>
</tr>
<tr>
<td>SPE</td>
</tr>
</tbody>
</table>
Creating definition models (1)

Analyzing BFO categories as relational configurations (RCs):

‘ENTITY+relation+RELATUM’

Ex: OBJECT has_part OBJECT
    OBJECT participates_in PROCESS
Output example: BFO 1.0

CONTINUANT
- is_a ENTITY

INDEPENDENT CONTINUANT
- is_a CONTINUANT
- bearer_of QUALITY
- bearer_of REALIZABLE ENTITY
- located_at TEMPORAL REGION
- located_in SITE

OBJECT AGGREGATE
- is_a INDEPENDENT CONTINUANT
- has_part OBJECT

OBJECT
- is_a INDEPENDENT CONTINUANT
- has_part OBJECT
- participates_in PROCESS

FIAT PART OF OBJECT
- is_a INDEPENDENT CONTINUANT
- part_of OBJECT
Creating definition models (2)

Creating relational models with proper and inherited RCs

**OBJECT**

**INDEPENDENT CONTINUANT**
- is_a SNAP CONTINUANT
- bearer_of QUALITY
- bearer_of REALIZABLE ENTITY
- located_at TEMPORAL REGION
- located_in SITE
- participates_in PROCESSUAL ENTITY
- **OBJECT**
  - is_a INDEPENDENT CONTINUANT
  - has_part OBJECT
  - participates_in PROCESS

- Relations inherited from the entity type INDEPENDENT CONTINUANT
- Relations characterizing the entity type OBJECT
Testing the models

Applying the models to multilingual & multi-domain textual definition corpora

• Segmenting: GEN + SPE
• Annotating: with relational models
• Statistical analyses to see
  – which relations are more relevant for defining
  – if new ones can be added to the models

➤ Relevant models from generic ones
Annotation example (1)

<FICHE langue="en">
  <NI>EN_MA_17</NI>
  <CM>
    <DOMAINE>nettoyage des déversements d'hydrocarbures</DOMAINE>
    <SS-DOM1>lutte en mer</SS-DOM1>
    <SS-DOM2>confinement</SS-DOM2>
  </CM>
  <VE>net boom</VE>
  <DF>
    <GEN REFrelGEN="is_a OBJECT" relationVE="GENRE_PROCH">A boom</GEN>
    <SPE REFrelSPE="has_part OBJECT">that is made of netting</SPE>
    <SPE REFrelSPE="bearer_of FUNCTION">to facilitate the retention of viscous oils.</SPE>
  </DF>
</FICHE>
Annotation example (2)

... <VE>net boom</VE>
<DF>
  <GEN REFrelGEN="is_a OBJECT" relationVE="GENRE_PROCH">A boom </GEN>
  <SPE REFrelSPE="has_part OBJECT">that is made of netting </SPE>
  <SPE REFrelSPE="bearer_of FUNCTION">to facilitate the retention of viscous oils. </SPE>
</DF>
...
Corpus analysis

3 (FR_AB_11) : protéine
(biosynthèse des protéines / cellule eucaryote / protéine)

<table>
<thead>
<tr>
<th>GEN</th>
<th>is_a OBJECT</th>
<th><em>Molécule</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE</td>
<td>has_part OBJECT</td>
<td>constituée d’<em>acides aminés</em> unis entre eux par des <em>liaisons peptidiques</em>, qui adopte une configuration spatiale particulière selon sa fonction.</td>
</tr>
</tbody>
</table>

4 (FR_AB_16) : acide nucléique
(biosynthèse des protéines / cellule eucaryote / acide nucléique)

<table>
<thead>
<tr>
<th>GEN</th>
<th>is_a OBJECT</th>
<th><em>Molécule</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE</td>
<td>has_part OBJECT</td>
<td>constituée d’un enchaînement de <em>nucléotides</em></td>
</tr>
<tr>
<td>SPE</td>
<td>located_in SITE (categ)</td>
<td>disposé le long d’un <em>brin</em> ou deux.</td>
</tr>
</tbody>
</table>

5 (FR_AB_17) : brin
(biosynthèse des protéines / cellule eucaryote / acide nucléique)

<table>
<thead>
<tr>
<th>GEN</th>
<th>is_a OBJECT</th>
<th>Chaîne</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE</td>
<td>bearer_of QUALITY</td>
<td>linéaire</td>
</tr>
<tr>
<td>SPE</td>
<td>has_part OBJECT</td>
<td>constituée de <em>nucléotides</em>, formant les <em>acides nucléiques</em>.</td>
</tr>
<tr>
<td>SPE</td>
<td>autre</td>
<td></td>
</tr>
</tbody>
</table>

6 (FR_AB_21) : séquence de nucléotides / séquence nucléotidique
(biosynthèse des protéines / cellule eucaryote / acide nucléique)

<table>
<thead>
<tr>
<th>GEN</th>
<th>is_a ABSTRACT ENTITY</th>
<th>Ordre</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPE</td>
<td>depends_on OBJECT (categ)</td>
<td>dans lequel les différents <em>nucléotides</em> se succèdent le long d’un <em>brin</em>.</td>
</tr>
</tbody>
</table>
### Analysis grid for OBJECT

<table>
<thead>
<tr>
<th>OBJECT</th>
<th>normal</th>
<th>categ.</th>
<th>other</th>
<th>total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>has_part OBJECT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>participates_in PROCESS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bearer_of QUALITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bearer_of REALIZABLE ENTITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>located_at TEMPORAL REGION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>located_in SITE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Analyzing the results for OBJECT

<table>
<thead>
<tr>
<th>OBJECT</th>
<th>normal</th>
<th>categ.</th>
<th>other</th>
<th>total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>has_part OBJECT</td>
<td>15</td>
<td>1</td>
<td></td>
<td>16</td>
<td>26</td>
</tr>
<tr>
<td>participates_in PROCESS</td>
<td>3</td>
<td>6</td>
<td></td>
<td>9</td>
<td>26</td>
</tr>
<tr>
<td>bearer_of QUALITY</td>
<td>19</td>
<td>3</td>
<td></td>
<td>22</td>
<td>60</td>
</tr>
<tr>
<td>bearer_of REALIZABLE ENTITY</td>
<td>31</td>
<td>1</td>
<td></td>
<td>32</td>
<td>60</td>
</tr>
<tr>
<td>located_at TEMPORAL REGION</td>
<td>0</td>
<td>0</td>
<td></td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>located_in SITE</td>
<td>2</td>
<td>2</td>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>other</td>
<td></td>
<td></td>
<td>14</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td>70</td>
<td>13</td>
<td>14</td>
<td>97</td>
<td>100</td>
</tr>
</tbody>
</table>

Selja Seppälä | November 20, 2013
Pilot corpus analysis

• 240 terminographic definitions in French
• 15 distinct domains
• Randomly extracted from students’ diploma works
  – Only works with the minimum grade to get the diploma
  – Multilingual dictionaries
  – Systematic coverage of the domains
• 16 entity types (out of 37 models)
• 73 relation types (462 occurrences)
Results of the corpus analysis

• ≈ 75 % of the defining features in the corpus correspond to relational configurations from the models (≈ 25 % marked ‘other’)
  ➔ Generic models that are predictive of definition contents

• Within each model represented in the corpus, some relational configurations seem to be more relevant than others
  ➔ Generic models that are relevant

➔ Preliminary results tend to confirm the working hypotheses
CONCLUSION
Conclusion

• Pilot study suggests that BFO-based models are predictive of definitions’ contents
  ➔ Can be used to model definition contents

• When not predictive, BFO categories and relations (its controlled vocabulary) can be used as a fixed metalanguage

• Methodology may reveal ‘typical features’ for defining each category
Advantages

• (Semi-)formal structuring of textual definitions
  ➔ Homogeneity
  ➔ Internal structure linked to the ontology

• No complex BFO labels displayed (only in metadata)
  ➔ Increased readability & user-friendliness

• Allows more complex querying via the definition (through metadata AND text)

• Domain-adaptable through corpus analysis

• Methodology applicable to more specific categories from ontologies extending BFO
Future work

• Conduct large-scale corpus studies
  ➔ Definition annotation manual
• Study the relation between logical definitions in ontologies and their textual counterparts
• Infer correspondence rules
• Possible outcome: program to automatically generate textual definitions from logical ones, and vice versa
Thank you!