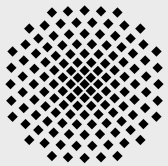
A General Methodology for Mapping EuroWordNets to the Suggested Upper Merged Ontology

Dennis Spohr

dennis.spohr@ling.uni-stuttgart.de

Institut für Linguistik/Romanistik
Universität Stuttgart

6th Language Resources and Evaluation Conference 2008
Marrakech, Morocco



Motivation

Polysemy in a Conceptual System

- Project at University of Stuttgart, funded by DFG
- Creation of a ***lexical semantic resource of polysemous French verbs***
- ***Formal*** lexical semantic descriptions

- Encode ***selectional argument restrictions*** for ***WSD***
- Express restrictions wrt. ***ontological types***, rather than WordNet synsets

Problem

No mapping between a French lexical resource and an ontology

Motivation

Polysemy in a Conceptual System

- Project at University of Stuttgart, funded by DFG
- Creation of a ***lexical semantic resource of polysemous French verbs***
- ***Formal*** lexical semantic descriptions

- Encode ***selectional argument restrictions*** for ***WSD***
- Express restrictions wrt. ***ontological types***, rather than WordNet synsets

Problem

No mapping between a French lexical resource and an ontology

Outline

- Motivation
- **Existing Mappings from WordNets to Ontologies**
- **Creating the Mapping from the French EWN to SUMO**
  - Prerequisites
  - Methodology
  - Results
- **Conclusion**

EuroWordNet and the EWN Top Ontology — Vossen et al., 1998

- Linking of EuroWordNet's ***Inter-Lingual-Index*** to a set of ***Top Concepts*** as part of the EWN project

Princeton WordNet and SUMO — Niles and Pease, 2003

- Mapping between ***Princeton WordNet 1.6*** and the ***Suggested Upper Merged Ontology***

The Global WordNet Grid — Horák et al., 2008

- ***Goal:*** Provide mappings from (all) WordNets to ***SUMO***

EuroWordNet and the EWN Top Ontology — Vossen et al., 1998

- Linking of EuroWordNet's ***Inter-Lingual-Index*** to a set of ***Top Concepts*** as part of the EWN project

Princeton WordNet and SUMO — Niles and Pease, 2003

- Mapping between ***Princeton WordNet 1.6*** and the ***Suggested Upper Merged Ontology***

The Global WordNet Grid — Horák et al., 2008

- ***Goal:*** Provide mappings from (all) WordNets to ***SUMO***

EuroWordNet and the EWN Top Ontology — Vossen et al., 1998

- Linking of EuroWordNet's ***Inter-Lingual-Index*** to a set of ***Top Concepts*** as part of the EWN project

Princeton WordNet and SUMO — Niles and Pease, 2003

- Mapping between ***Princeton WordNet 1.6*** and the ***Suggested Upper Merged Ontology***

The Global WordNet Grid — Horák et al., 2008

- ***Goal:*** Provide mappings from (all) WordNets to ***SUMO***

(Euro)WordNet — Vossen, 1998

- EWN synsets *linked to Inter-Lingual-Index (ILI)*
- ILI represents *version 1.5 of Princeton WordNet*
- New releases of Princeton WN are related via sensemaps

SUMO/WordNet mapping — Niles and Pease, 2003

- Links to synsets of WN1.6 - WN3.1

Idea

- Go from French EWN to ILI ($\approx$ WN1.5)
- Use *sensemap files* to go from WN1.5 to WN1.6
- Use *SUMO/WN1.6 mapping*

(Euro)WordNet — Vossen, 1998

- EWN synsets *linked to Inter-Lingual-Index (ILI)*
- ILI represents *version 1.5 of Princeton WordNet*
- New releases of Princeton WN are related via sensemaps

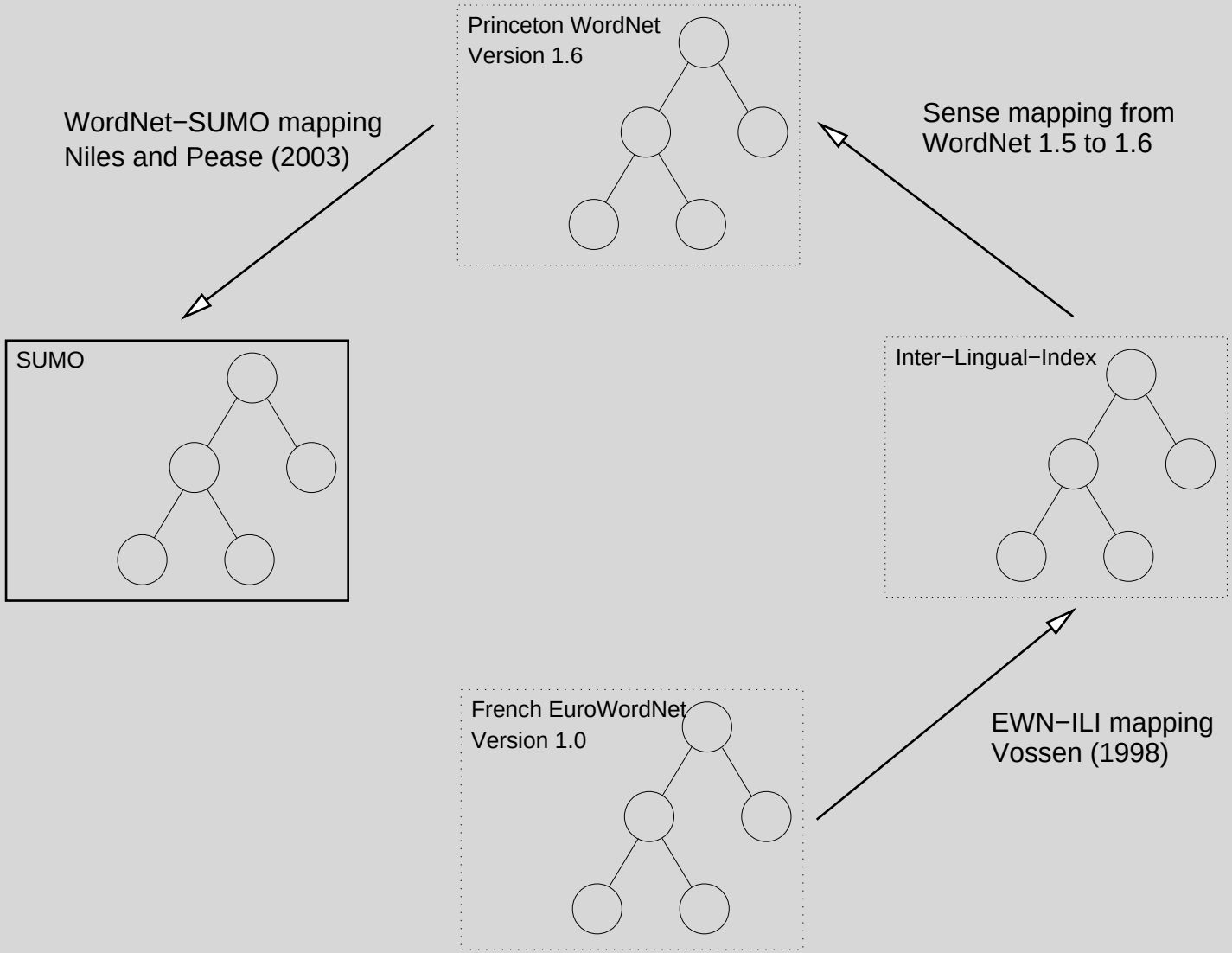
SUMO/WordNet mapping — Niles and Pease, 2003

- Links to synsets of WN1.6 - WN3.1

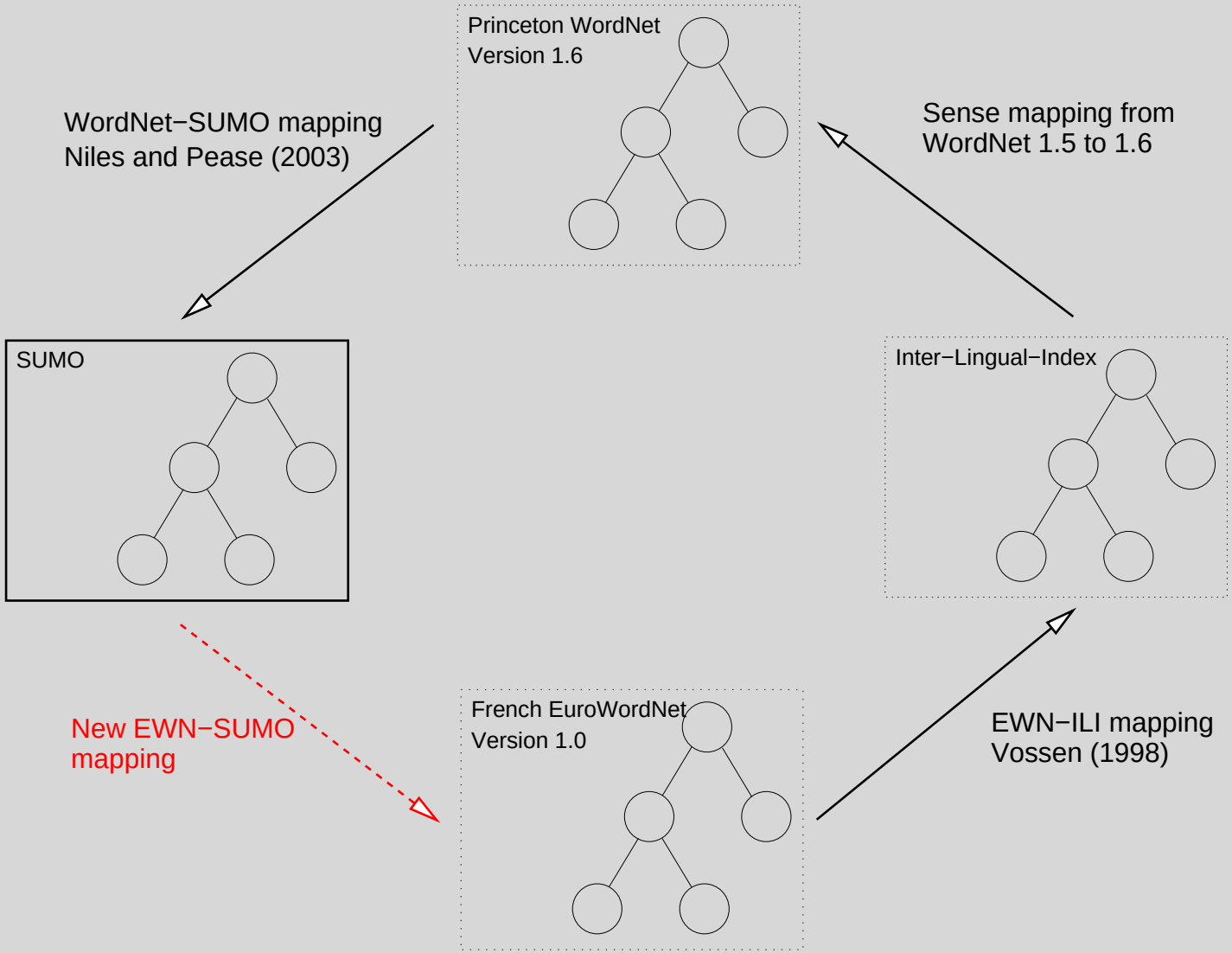
Idea

- Go from French EWN to ILI ($\approx$ WN1.5)
- Use *sensemap files* to go from WN1.5 to WN1.6
- Use *SUMO/WN1.6 mapping*

Mapping from SUMO to EuroWordNet



Mapping from SUMO to EuroWordNet



French EuroWordNet — Vossen (ed.), 1998

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI> 00002403-n </ILI>
</SYNSET>
```

Sensemap — Fellbaum (ed.), 1998

```
entity%1:03:00::; 00002403  entity%1:03:00::; 00001740
```

WordNet-SUMO mapping — Niles and Pease, 2003

```
 00001740  03 n 02 entity 0 ... | anything having
existence (living or nonliving)  &%Physical=
```

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI>00002403-n</ILI> <SUMO>Physical<TYPE>=</TYPE></SUMO>
</SYNSET>
```

French EuroWordNet

VOSSEN (ED.), 1998

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI> 00002403-n </ILI>
</SYNSET>
```

Sensemap

FELLBAUM (ED.), 1998

```
entity%1:03:00::; 00002403  entity%1:03:00::; 00001740
```

WordNet-SUMO mapping

NILES AND PEASE, 2003

```
 00001740  03 n 02 entity 0 ... | anything having
existence (living or nonliving)  &%Physical=
```

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI>00002403-n</ILI> <SUMO>Physical<TYPE>=</TYPE></SUMO>
</SYNSET>
```

French EuroWordNet

VOSSEN (ED.), 1998

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI> 00002403-n </ILI>
</SYNSET>
```

Sensemap

FELLBAUM (ED.), 1998

```
entity%1:03:00::; 00002403  entity%1:03:00::; 00001740
```

WordNet-SUMO mapping

NILES AND PEASE, 2003

```
00001740  03 n 02 entity 0 ... | anything having
existence (living or nonliving)  &%Physical=
```

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI>00002403-n</ILI> <SUMO>Physical<TYPE>=</TYPE></SUMO>
</SYNSET>
```

French EuroWordNet

VOSSEN (ED.), 1998

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI> 00002403-n </ILI>
</SYNSET>
```

Sensemap

FELLBAUM (ED.), 1998

```
entity%1:03:00::; 00002403  entity%1:03:00::; 00001740
```

WordNet-SUMO mapping

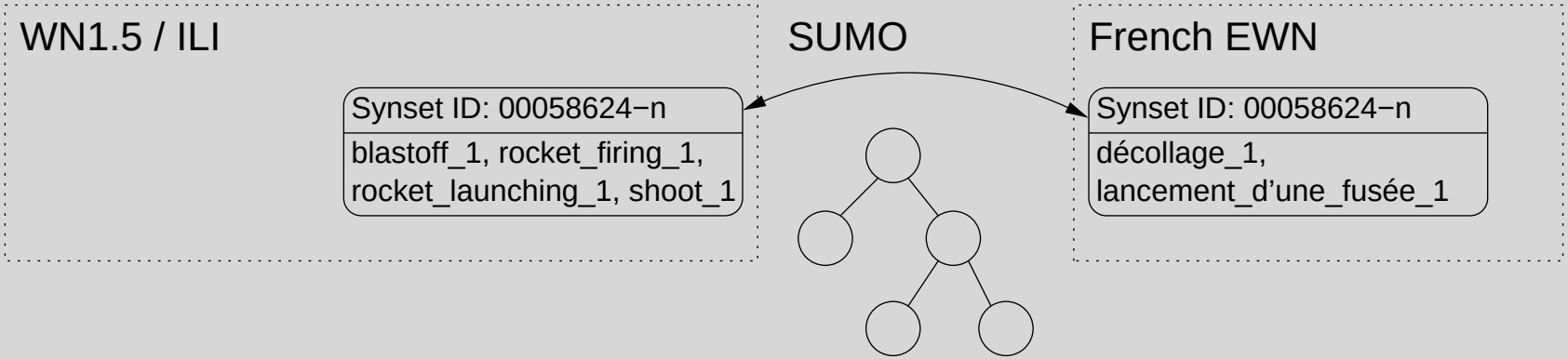
NILES AND PEASE, 2003

```
00001740  03 n 02 entity 0 ... | anything having
existence (living or nonliving)  &%Physical=
```

```
<SYNSET> ...
 <LITERAL>entité<SENSE>1</SENSE></LITERAL>
 <ILI>00002403-n</ILI> <SUMO>Physical<TYPE>=</TYPE></SUMO>
</SYNSET>
```

Some issues

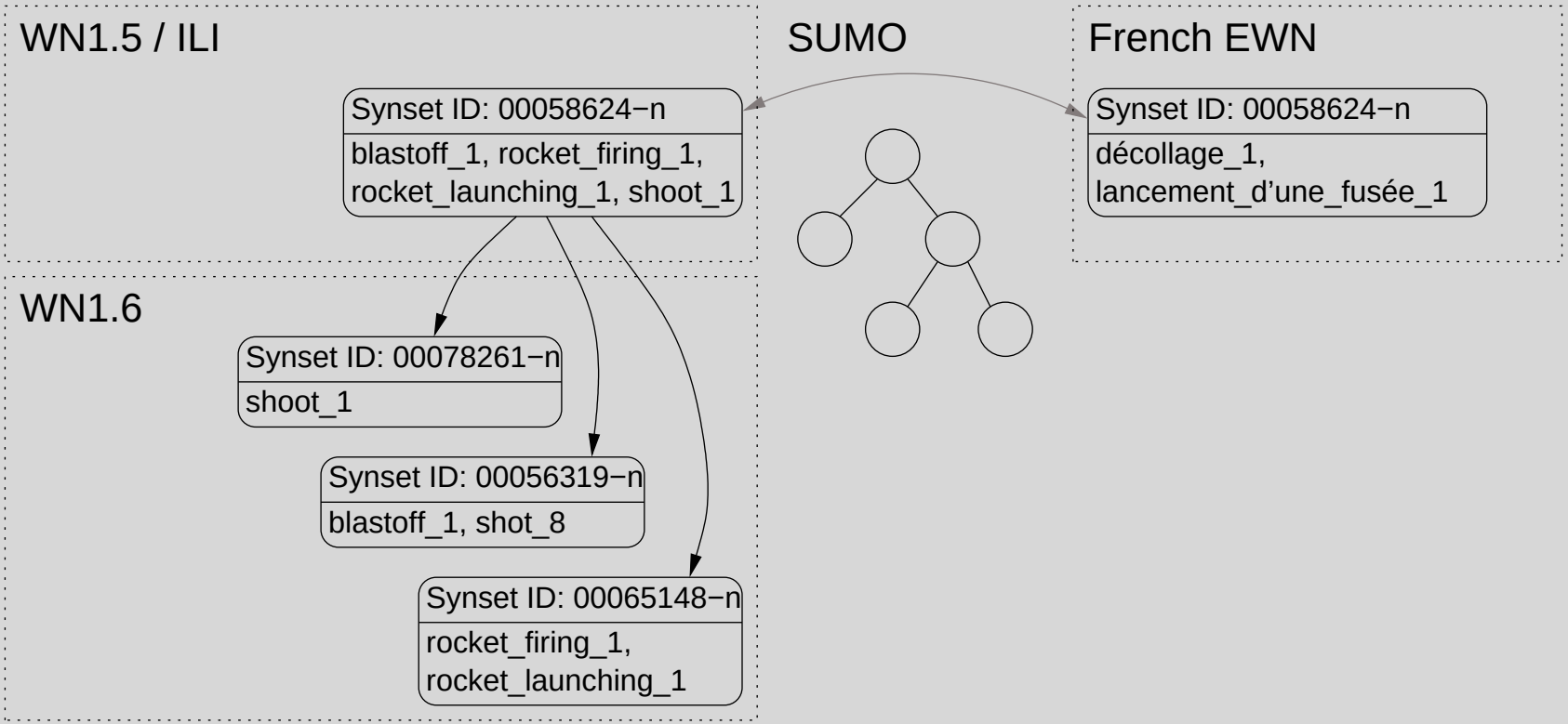
Synset splitting from WN1.5 to 1.6



- Where to split the French synset?
- Which SUMO classes should the French synset map to?

Some issues

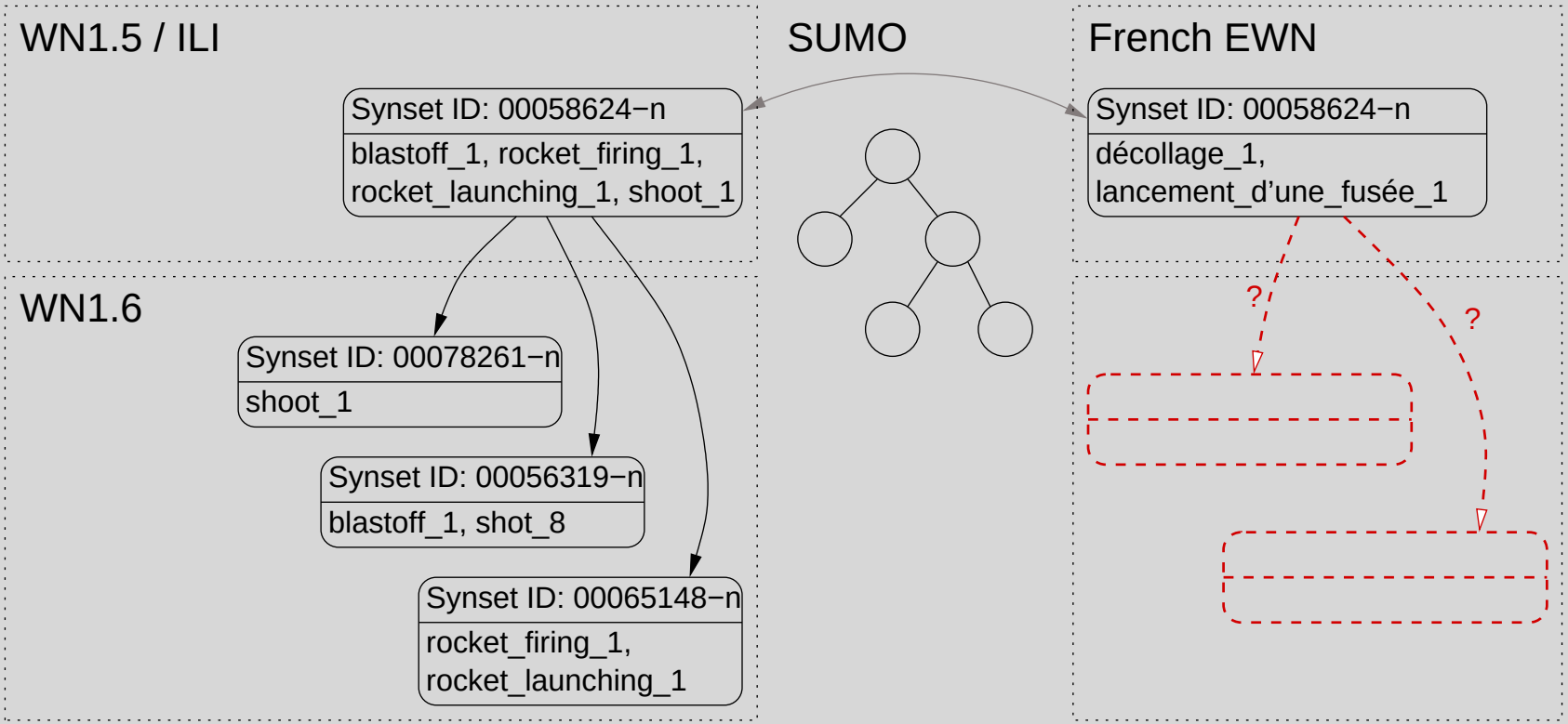
Synset splitting from WN1.5 to 1.6



- Where to split the French synset?
- Which SUMO classes should the French synset map to?

Some issues

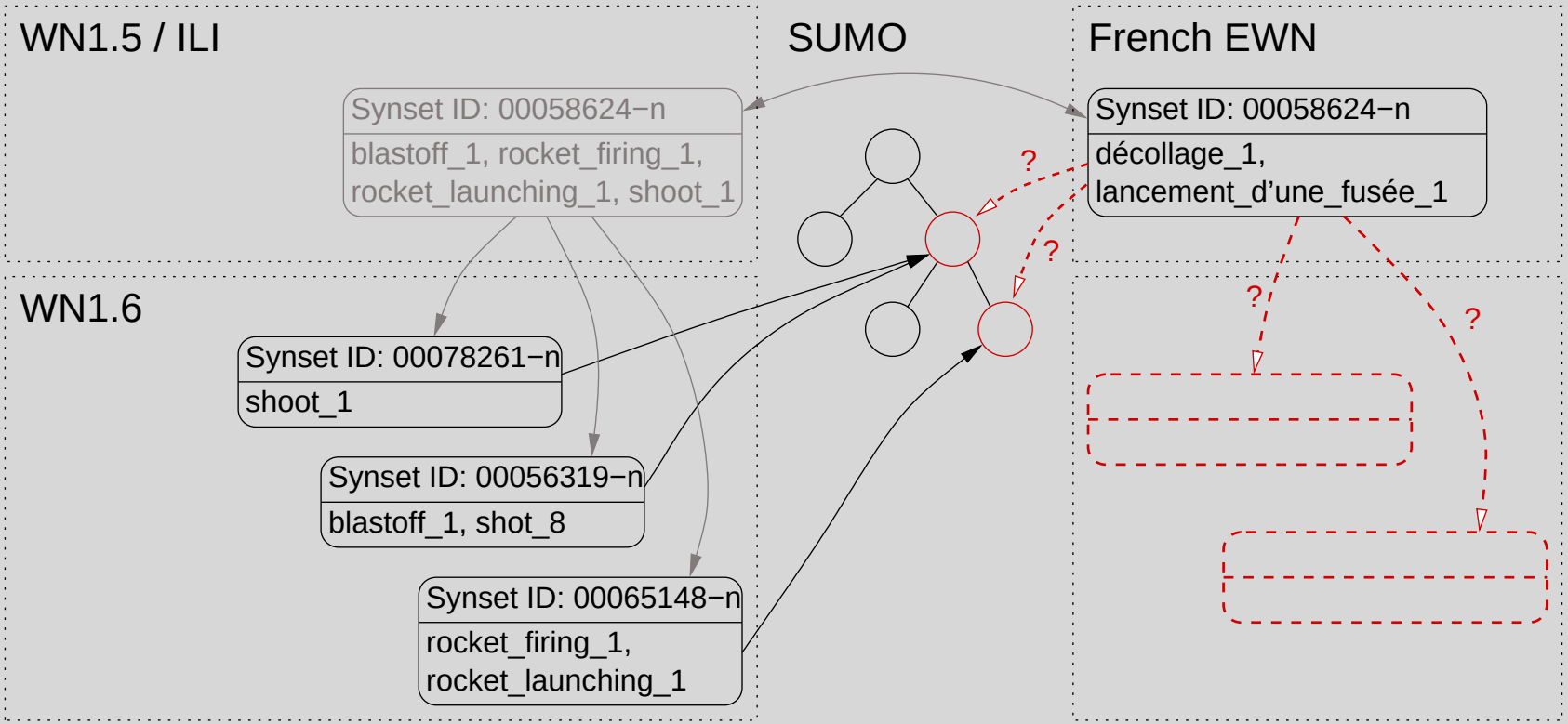
Synset splitting from WN1.5 to 1.6



- Where to split the French synset?
- Which SUMO classes should the French synset map to?

Some issues

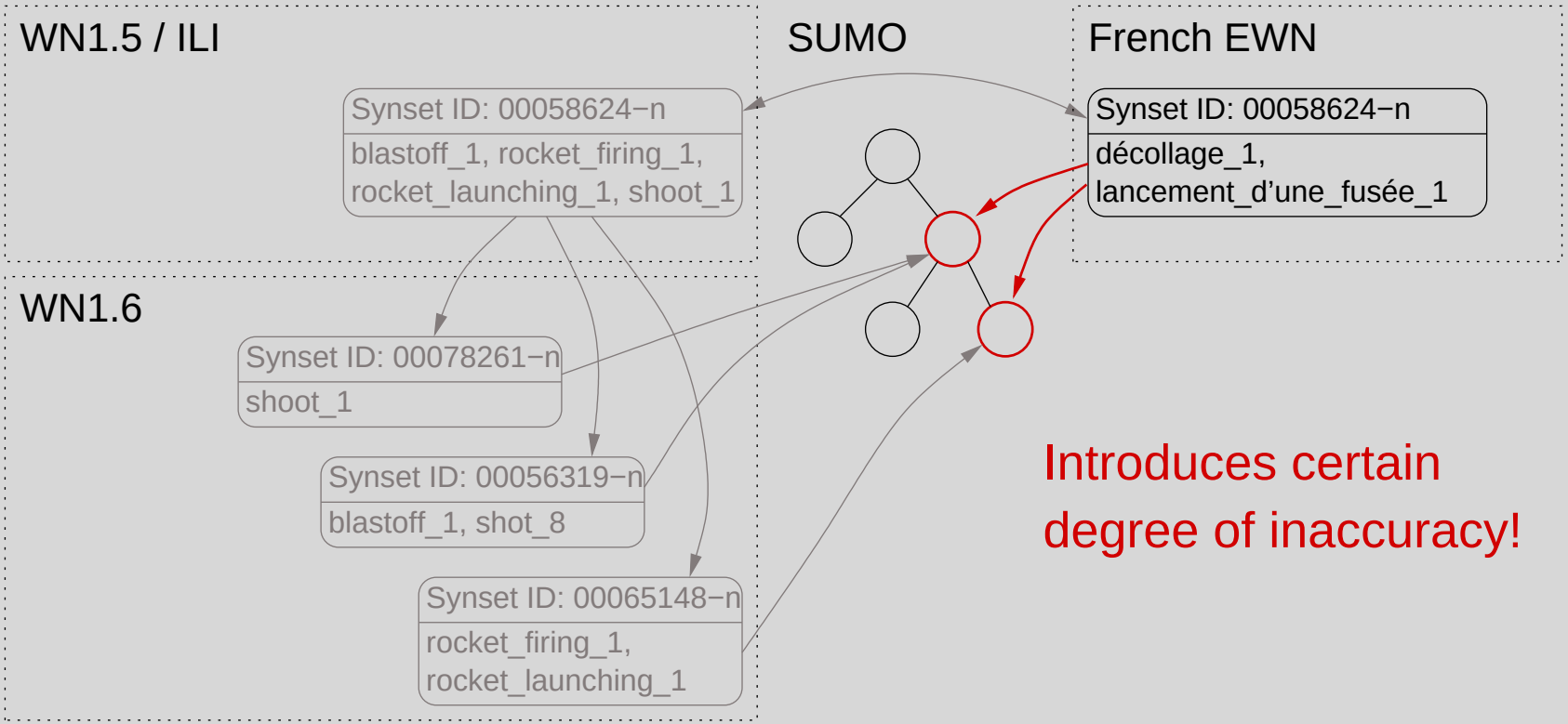
Synset splitting from WN1.5 to 1.6



- Where to split the French synset?
- Which SUMO classes should the French synset map to?

Some issues

Synset splitting from WN1.5 to 1.6



- Where to split the French synset?
- Which SUMO classes should the French synset map to?

Overall mapping results

	Type	*Frequency*	
		abs	*rel*
1	Synsets in French EWN	22,745	100.00%
2	. . . with SUMO mapping	22,351	98.27%
3	. . . without SUMO mapping	394	1.73%
	Of those with SUMO mapping		
4	. . . with one mapping	22,026	98.54%
5	. . . with two mappings	214	0.96%
6	. . . with three or more mappings	111	0.50%
7	. . . with only one sensemap	9,739	43.57%
8	. . . with more than one sensemap but only one SUMO class	12,287	54.97%
9	. . . with more than one sensemap and more than one SUMO class	325	1.46%

Analysis

Of the 394 without SUMO class . . .

. . . ***323*** are from the ***new technology domain***, e.g. *adresse d'inter-réseau* (*'network address'*), *applet* or *cache mémoire* (*'cache memory'*)

. . . ***23*** are ***collocational*** or ***idiomatic***, e.g. *tenir compte de* (*'to account for'*), *vendre la mèche* (*'to reveal a secret'*; lit. *'to sell the fuse'*) or *saigner quelqu'un à blanc* (*'to exploit someone'*; lit. *'to bleed someone to white'*)

. . . ***8*** from the ***food domain***, e.g. *petit four* (a specific kind of pastry) or *sauce au chocolat fondu* (a specific kind of chocolate sauce)

Analysis

Of the 394 without SUMO class . . .

. . . ***323*** are from the ***new technology domain***, e.g. *adresse d'inter-réseau* (*'network address'*), *applet* or *cache mémoire* (*'cache memory'*)

. . . ***23*** are ***collocational*** or ***idiomatic***, e.g. *tenir compte de* (*'to account for'*), *vendre la mèche* (*'to reveal a secret'*; lit. *'to sell the fuse'*) or *saigner quelqu'un à blanc* (*'to exploit someone'*; lit. *'to bleed someone to white'*)

. . . ***8*** from the ***food domain***, e.g. *petit four* (a specific kind of pastry) or *sauce au chocolat fondu* (a specific kind of chocolate sauce)

Analysis

Of the 394 without SUMO class . . .

. . . ***323*** are from the ***new technology domain***, e.g. *adresse d'inter-réseau* (*'network address'*), *applet* or *cache mémoire* (*'cache memory'*)

. . . ***23*** are ***collocational*** or ***idiomatic***, e.g. *tenir compte de* (*'to account for'*), *vendre la mèche* (*'to reveal a secret'*; lit. *'to sell the fuse'*) or *saigner quelqu'un à blanc* (*'to exploit someone'*; lit. *'to bleed someone to white'*)

. . . ***8*** from the ***food domain***, e.g. *petit four* (a specific kind of pastry) or *sauce au chocolat fondu* (a specific kind of chocolate sauce)

- Motivation
- Existing Mappings from WordNets to Ontologies
- Creating the Mapping from the French EWN to SUMO
  - Prerequisites
  - Methodology
  - Results
- **Conclusion**

Conclusion

- Mapping from French EuroWordNet to SUMO
- Fairly straightforward methodology using existing mappings of Princeton WordNet

---

- Mapping from French EWN to SUMO did not exist before
- Methodology is applicable to all EuroWordNets
- Resulting resource can be used e.g. for calculation of selectional preferences

Outlook

Updating to the latest SUMO-WN mapping

- We didn't use the latest mapping since ***degree of inaccuracy increases*** when going from 1.6 to 3.0
- Basically three options:
  1. Accept the resulting inaccuracy and evaluate the manual effort needed to resolve it
  2. Apply methodology only to those synsets that have not been split, at the cost of a heterogeneous resource
  3. Manually create a direct mapping from the French EuroWordNet to the latest version of Princeton WordNet

References

C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of ACL/COLING*, San Francisco, CA.

C. Fellbaum (ed.). 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

A. Gangemi, R. Navigli, and P. Velardi. 2003. The OntoWordNet Project: Extension and Axiomatization of Conceptual Relations in WordNet. *LNCS 2888*, Springer.

A. Horák, K. Pala, and Adam Rambousek. 2008. The GlobalWordNet Grid Software Design. In *Proceedings of the 4th Global WordNet Conference*. Szeged, Hungary.

I. Niles and A. Pease. 2003. Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of IKE '03*, Las Vegas, NV.

D. Spohr. 2008. Extraction of Selectional Preferences for French using a Mapping from EuroWordNet to the Suggested Upper Merged Ontology. In *Proceedings of the 4th Global WordNet Conference*, Szeged, Hungary.

P. Vossen (ed.). 1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers.