

Extraction of Selectional Preferences for French using a Mapping from EuroWordNet to the Suggested Upper Merged Ontology

Dennis Spohr

Institut für Linguistik/Romanistik
Universität Stuttgart
Stuttgart, Germany
dennis.spohr@ling.uni-stuttgart.de

Abstract. This paper presents an approach to extracting selectional preferences of French verbal predicates with respect to the ontological types of their arguments. Selectional preference is calculated on the basis of Resnik’s measure of *selectional association* between a predicate and the class of its argument [1]. However, instead of using WordNet synsets to express sortal restrictions (cf. [2]), we employ conceptual classes of the Suggested Upper Merged Ontology (SUMO; [3]) that have been automatically mapped to synsets of the French EuroWordNet [4] in a generic way that is in principle applicable to all WordNets which are linked to the Inter-Lingual-Index.

1 Introduction

Lexical-semantic NLP and with it semantic lexicons have become increasingly important over the last decades, and the contribution of (Euro-)WordNet [5,4] and FrameNet [6] within this field is of course so fundamental and well-known that it need not be discussed here. However, recent years have further seen a strong tendency towards interfacing such resources with knowledge bases or taxonomies of general knowledge, both commonly referred to as *ontologies*. Well-known examples of such efforts are e.g. [7], who linked EuroWordNet’s Inter-Lingual-Index to a number of *base concepts* and a *top ontology* as integral part of the EuroWordNet project, [8] who mapped Princeton WordNet to the Suggested Upper Merged Ontology (SUMO), and [9] who linked FrameNet and SUMO. Moreover, the recent Global WordNet Grid¹ is pursuing such efforts on a considerable scale to create mappings from SUMO to all existing WordNets.

One of the main reasons why such approaches are so important is that while resources like (Euro-)WordNet and FrameNet attempt to model lexical-semantic knowledge, ontologies try to mediate common knowledge or knowledge of the world. Therefore, linking these two types of resources may be able to bridge the gap between language-dependent lexical knowledge and language-independent facts or statements about the world.

Such statements appear to have a more *universal* character, and this is what makes combinations of ontological and lexical-semantic resources interesting for the formulation of selectional restrictions or preferences. We believe that a statement like “*X prefers*

¹ http://www.globalwordnet.org/gwa/gwa_grid.htm

subjects of type Human or CognitiveAgent” is – from a meta-linguistic perspective – more informative than saying “*X prefers the subjects {human_1, individual_1, mortal_1, person_1, someone_1, soul_1} or ...*”. In this paper, we present a general methodology for mapping EuroWordNets to the SUMO ontology by using both an existing mapping from Princeton WordNet 1.6 to SUMO [8] and the linking of the EuroWordNets to the Inter-Lingual-Index [7]. We apply our methodology to the French EuroWordNet and extract *sortal* selectional preferences that are calculated on the basis of an established measure of *selectional association* between a predicate and the classes of its argument [1]. Section 2 of this paper introduces some background on WN and SUMO, and Resnik’s approach to selectional preference extraction. In Section 3, we will present our methodology for mapping EWN to SUMO, and Section 4 gives details on how we extract selectional preferences based on this mapping. After an evaluation of the mapping methodology and selectional preference extraction, we conclude in Section 6 and briefly discuss ways to apply and further extend our approach.

2 Background

2.1 WordNet and the Suggested Upper Merged Ontology

In this section, we will briefly discuss work that has been done on linking WordNet to the Suggested Upper Merged Ontology. [8] have created such a mapping for version 1.6 of WordNet, and have in subsequent years released new mappings for each new version of WordNet, with the latest release in summer 2007 for WN3.0. In creating their linking, [8] have decided to use the following three mapping relations: *synonymy* (equivalence, indicated by ‘=’; cf. Section 3.2), *hypernymy* (subclass-superclass relation, indicated by ‘+’), and *instantiation* (indicated by ‘@’). In contrast to the approach presented in their paper, we try to create mappings automatically, though relying heavily on their manual preparatory work.

As was mentioned in the introduction, the Global WordNet Grid initiative, which was launched in early 2006, is trying to provide WordNet-SUMO mappings for all existing WordNets. The current state, as of late 2007, comprises mappings for 5,000 English base concepts, as well as for the Spanish and Catalan WordNets.

2.2 Selectional Preference Extraction

The measure we use for the calculation of selectional preference is that of [1]², who uses the notion of *relative entropy* known from information theory [11]. The strength of selectional preference $S_R(p)$ of a predicate p with respect to a grammatical relation R is defined as follows.

SELECTIONAL PREFERENCE STRENGTH:

$$S_R(p) = \sum_c Pr(c|p) \log \frac{Pr(c|p)}{Pr(c)}$$

² See [10] for a recent survey of several other approaches to selectional preference acquisition.

The better $Pr(c)$ approximates $Pr(c|p)$, the closer $\log \frac{Pr(c|p)}{Pr(c)}$ is to 0, i.e. the less influence p has on its argument, and therefore the less strong is its selectional preference. The selectional preference strength is on the one hand an indicator as to “how much information [...] predicate p provides about the conceptual class of its argument” ([1]: p. 53). On the other hand, it is used for *normalising* the selectional preference values of a predicate, in order to be able to compare the values of different predicates: a predicate that is generally weak in showing preferences will thus receive a higher value if it really shows a preference for a particular conceptual argument class. Selectional preference for a particular conceptual class is calculated in the form of *selectional association* $A_R(p, c)$ between p and the class c of its argument. Its definition is given below.

SELECTIONAL ASSOCIATION:

$$A_R(p, c) = \frac{1}{S_R(p)} Pr(c|p) \log \frac{Pr(c|p)}{Pr(c)}$$

As Resnik points out, the fact that text corpora have usually not been annotated with explicit and unambiguous conceptual classes requires some sort of distribution of frequencies among the possible conceptual class of a word. The following formula calculates the frequency of predicate p and class c .

FREQUENCY PROPAGATION:

$$freq_R(p, c) \approx \sum_{w \in c} \frac{count_R(p, w)}{classes(w)}$$

This means that the actual frequency count of a word w , which stands in relation R (e.g. verb-object) to p , is distributed equally among the classes c which w is a member of. In a hierarchical resource such as SUMO, this also has the effect of *propagating* the *freq* value up the hierarchy: if w is a member of class c , it is, of course, also a member of the superclasses of c , and thus $freq(p, c)$ is also added to all the superclasses of c .

3 Mapping EuroWordNet to SUMO

In this section, we will present how the French EuroWordNet has been mapped onto conceptual classes of the Suggested Upper Merged Ontology. The general methodology of creating the mapping to the French EuroWordNet is described in the following subsection. Although a quite recent mapping to version 3.0 of WordNet exists, we decided to use the very first mapping – namely that of WordNet version 1.6 – as a starting point. The reasons for doing so mainly concern the sensemaps between different versions of WordNet, and are explained in detail in Section 3.2 below.

3.1 General Methodology

As was just mentioned, we use the mapping of SUMO to version 1.6 of WordNet in order to link the French EuroWordNet to SUMO. The French EWN itself – as is the case with

all EuroWordNets – is linked to the Inter-Lingual-Index, a set of concepts that is intended to be largely language-independent (cf. [7]). A crucial prerequisite for our approach to function is that the identifiers of entities in the Inter-Lingual-Index correspond to synset identifiers in version 1.5 of WordNet. For example, entity 00058624-n of the Inter-Lingual-Index, which is glossed by “*the launching of a rocket under its own power*”, corresponds to synset {*décollage_1, lancement_d’une_fusée_1*} in the French EWN and to {*blastoff_1, rocket_firing_1, rocket_launching_1, shoot_1*} in WN1.5. Starting from these observations, i.e. the mapping of SUMO to WN1.6 and the linking of the French EWN to the Inter-Lingual-Index (\approx WN1.5), the only remaining task that is left is to move from WN1.5 to WN1.6. In order to do this, we can avail ourselves of the sensemap files that came with the 1.6 release of WordNet, which indicate the changes from WN1.5 to WN1.6. Ignoring particular mapping issues for the moment (see Section 3.2 below), the resulting EuroWordNet entries look like the one shown in Figure 1. The structure is based on the format suggested by the Global WordNet Grid. The whole mapping process is summarised in Figure 2.

```

<SYNSET>
<POS>n</POS>
<SYNONYM>
  <LITERAL>organisme
    <SENSE>1</SENSE>
  </LITERAL>
  <LITERAL>forme de vie
    <SENSE>1</SENSE>
  </LITERAL>
  <LITERAL>\^etre
    <SENSE>2</SENSE>
  </LITERAL>
  <LITERAL>vie
    <SENSE>11</SENSE>
  </LITERAL>
</SYNONYM>
<ILI>00002728-n</ILI>
<HYPERONYM>00002403-n</HYPERONYM>
<SUMO>Organism
  <TYPE>=</TYPE>
</SUMO>
<DEF>any living entity</DEF>
</SYNSET>

```

Fig. 1. Sample EuroWordNet entry of synset 00002728-n ({*organisme_1, forme_de_vie_1, être_2, vie_11*}) after the mapping

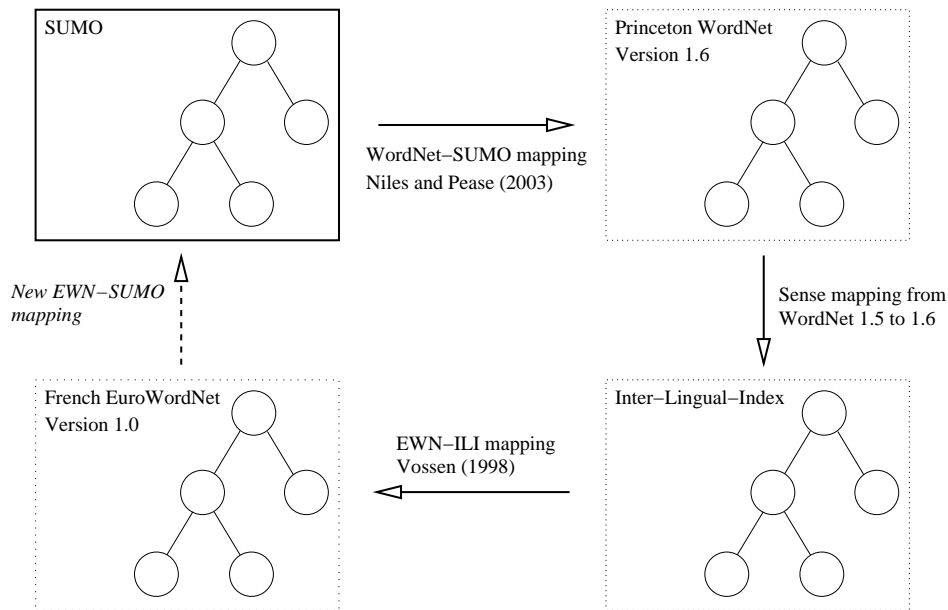


Fig. 2. Process of mapping the French EuroWordNet to SUMO (clockwise from top left)

3.2 WordNet Sensemaps

Whenever updates of WordNet are released, the updated version comes with files that, among others, indicate changes in the structure of the synsets. For example, synset 00058624-n from above has been split in the step from WN1.5 to WN1.6: {shoot_1} is now a member of synset 00078261-n, {blastoff_1} of synset 00065319-n, and {rocket_firing_1, rocket_launching_1} of synset 00065148-n. Therefore, version 1.6 contains new synsets that did not exist in WN1.5, and further cases in which a synset is reorganised thus that some of its items belong to different synsets in the updated version. The primary problem for the task of mapping such instances comes from the fact that individual members of a synset do not have unique identifiers themselves, but only the synset as a whole³. Therefore, when a synset has been split, it is not possible to automatically determine the correct position at which the synset has to be split in a different language, or even whether it has to be split at all. Moreover, each update comes with a large number of such changes, and therefore using the most recent mapping between SUMO and WN3.0, which is without a doubt desirable, would multiply the inaccuracies in the mapping right from the start. Just imagine a case where a synset has been split e.g. from WN1.5 to WN1.6, and the new synset is then split again when going to WN1.7, and so on.

³ This is, of course, not a problem of the WordNet approach, but rather of the fact that there is no one-to-one mapping between languages.

The decision that was made for cases like these is to assign to the original synset two (or more if necessary) SUMO classes: first the one that has been mapped to this synset, and second the ones to which the new (or relevant existing) synsets have been mapped in WN1.6. The justification of this decision is based on the assumption that on a level as abstract as that of SUMO conceptual classes, a “slight” reorganisation of the synsets and some of their items should not lead to significant conceptual clashes, as this would imply that grave errors had been made when putting the respective senses into one synset in the first place. In Figure 3 below, which depicts the entry of synset 00058624-n after the mapping, we see that the two SUMO classes that have been assigned to this synset do at least remotely fit the senses: more specific than *Impelling* and *Motion*, and equivalent to *Shooting*. Of course, a qualitative evaluation is needed to determine the degree of inaccuracy that is introduced. However, such an evaluation would rely heavily on manual inspection and could therefore not be carried up to this moment.

```

<SYNSET>
...
<ILI>00058624-n</ILI>
<HYPERONYM>00058381-n</HYPERONYM>
<SUMO>Impelling
  <TYPE>+</TYPE>
</SUMO>
<SUMO>Motion
  <TYPE>+</TYPE>
</SUMO>
<SUMO>Shooting
  <TYPE>=</TYPE>
</SUMO>
</SYNSET>

```

Fig. 3. Part of the EuroWordNet entry of synset 00058624-n (*{décollage_1, lancement_d'une_fusée_1}*) after the mapping

4 Extraction of Selectional Preferences

4.1 Corpus extraction

The (potential) nominal arguments of the verbal predicates have been extracted from a portion of more than 350 million tokens from the French *Agence France-Presse* corpus licensed by the Linguistic Data Consortium⁴. The corpus has been part-of-speech tagged using the French *TreeTagger* parameter files [12] and has been stored in the widely-used *Corpus Workbench* format [13]. Figure 4 below shows the CQP query that extracted potential direct objects of *'manger'*.

⁴ <http://www ldc.upenn.edu/>

```

1 [pos="DET:(ART|POS)"]? [pos="AD(V|J)"]{0,3} [pos="N(A|O)M"]
2 [pos="DET:(ART|POS)"]? [pos="AD(J|V)"]{0,3}
3 ([pos="PRP.*"] [pos="DET:(ART|POS)"]? [pos="AD(V|J)"])?
4 [pos="N(A|O)M"] [pos="AD(V|J)"]{0,3}
5 [pos="VER.*"]{0,2} [pos="ADV"]{0,2} [lemma="avoir|faire"]?
6 [lemma="manger" & pos!="VER:ppre"]
7 [pos="ADV" & lemma!="que"]{0,2}
8 [pos="DET:(ART|POS)"]? [pos="AD(V|J)" & lemma!="que"]{0,3}
9 [pos="NUM"]?
10 [pos="N(A|O)M" & lemma!="(lundi|mardi|mercredi|jeudi|
11 vendredi|samedi|dimanche|janvier|f\`evrier|mars|avril|mai|
12 juin|juillet|ao\`ut|septembre|octobre|novembre|d\`ecembre)"];

```

Fig. 4. CQP query extracting direct objects of 'manger'

We have decided to use a quite rigid syntactic structure, and therefore the query contains both the potential subject and direct object although only one of them is focussed on at a time. Lines 1-4 in Figure 4 represent the subject position – with the head of the subject noun phrase at the end of in line 1 –, and the direct object is described in lines 10-12. The verbal predicate, in this case 'manger', is shown in line 6. The results of this query, when grouped e.g. by object, look like the following (see Table 1).

Table 1. Results of the query in Figure 4 after grouping by object

Word	Frequency	Word	Frequency
pain ('bread')	16	revenu ('revenue')	2
enfant ('child')	8	chose ('thing')	2
plat ('dish')	4	nourriture ('nutrition')	2
glace ('ice')	4	partie ('part')	2
poisson ('fish')	3	méchoui (≈ "Arabian dish")	1
chapeau ('hat')	3	victuaille ('comestible')	1
cœur ('heart')	3	pélican ('pelican')	1
steak ('steak')	2	vipère ('viper')	1
poussin ('poult')	2	raisin ('grape')	1
abat ('innards')	2	cervelle ('brains')	1
singe ('monkey')	2	sandwich ('sandwich')	1
soupe ('soup')	2	hamburger ('hamburger')	1
feuille ('leaf')	2	christmas ('christmas')	1

4.2 Storage and Retrieval

Before we calculated the selectional preferences, we converted the file containing the SUMO-EuroWordNet mappings to OWL (Web Ontology Language; cf. [14]). We have

further created a “class only” OWL version of SUMO based on the XML version of SUMO that is distributed with the KSMSA ontology browser⁵ (version 1.0.9.1.1). The reasons for not using the OWL version available from the SUMO project site⁶ are (i) that it is difficult to process by the “standard” ontology editing tool Protégé [15], which is mainly due to the fact that SUMO was originally written in the far more expressive *Suggested Upper Ontology Knowledge Interchange Format* (SUO-KIF⁷) and contains, e.g., entities which are one-place predicates and two-place predicates at the same time and therefore occur in both class and property hierarchies, and (ii) that processing a class hierarchy for frequency propagation is far more straightforward and intuitive than processing a mixed hierarchy (see below). Therefore, if a synset had been mapped onto a SUMO concept in an instance relation (cf. ‘@’ in Section 2.1 above), it was still created as an OWL class with a subclass relation to the SUMO concept. We believe that the cognitive differences between the instantiation and hypernymy relations (cf. [8]) can be neglected for this purpose. The two files (SUMO and the EWN-SUMO mapping) were then stored as an RDFS database in the Sesame Framework [16]. The main reason for doing all this is that we thus have the benefit of using OWL’s – and of course RDF’s – built-in subsumption and inheritance mechanism, which is very advantageous since the frequencies for the calculation of selectional preferences have to be propagated along the hierarchy (cf. Section 2.2). A further benefit is that we are thus able to use the Protégé OWL API⁸ and the Sesame API⁹ in order to perform the propagation of frequencies.

4.3 Calculation of Selectional Preferences

In order to calculate the selectional association between the verbal predicate and its arguments, it is necessary to first calculate *prior* values, i.e. to propagate the frequencies of all arguments irrespective of the verbal predicate up the SUMO hierarchy. For each word in the list (cf. Table 1), the synsets it belongs to are looked up in the database. If it belongs to more than one synset, which is typically the case, then its frequency is divided by the number of readings (cf. Section 2.2 above). After that, for each of these synsets, first its equivalent SUMO classes are extracted, and then the frequency is propagated up the hierarchy along the *direct* superclass relationship. In case of multiple inheritance, i.e. one class having more than one direct superclass, the frequency is divided by the number of direct superclasses, similar to what has already been explained for the different readings of a word. The result is a structure in which every SUMO class has an associated *prior* value. As was already mentioned in Section 2.2, the same is done in order to determine the *posterior* values for the words occurring as arguments of a given verbal predicate, and these are then compared with the *prior* values. Section 5.2 below shows and discusses results for two French verbal predicates.

⁵ <http://virtual.cvut.cz/ksmsaWeb/browser/title/>

⁶ <http://www.ontologyportal.org/>

⁷ <http://suo.ieee.org/SUO/KIF/suo-kif.html>

⁸ <http://protege.stanford.edu/plugins/owl/api/>

⁹ <http://www.openrdf.org/doc/sesame/users/ch07.html>

5 Evaluation

5.1 Evaluation of SUMO Mapping

Table 2 below shows the results of the mapping procedure. Lines 1-3 in the table display the total number of synsets in the French EuroWordNet, as well as the numbers of those which have or have not received a SUMO mapping. Of those 22,351 synsets which have been assigned a SUMO class (cf. lines 4-6), 98.54% have been assigned exactly one class, whereas 0.96% have been mapped to two and 0.50% to three or more SUMO classes¹⁰. In line 8 we see that almost 55% of the synsets that have been assigned SUMO classes occurred in multiple sensemaps, but were all mapped onto synsets belonging to the same SUMO class, while only 1.46% were mapped onto two or more SUMO classes (cf. line 9). This means that only 1.46% are in principle able to cause “conceptual clashes” when retaining the strategy presented in Section 3.2. Table 3 displays the 20 most frequent SUMO classes that have been mapped to synsets in the French EuroWordNet.

Table 2. Number of SUMO mappings according to different types

	<i>Type</i>	<i>Frequency</i>	
		<i>abs</i>	<i>rel</i>
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%

5.2 Evaluation of Preference Extraction

In the following, we will discuss the results for the selectional preference extraction of direct objects of *'lire'* (*'read'*) and *'manger'* (*'eat'*). These words were chosen because

¹⁰ One synset even received 16 SUMO classes. This was due to the fact that the English synset contained the highly polysemous *'cut'*, which was split into 26 new synsets in the step from WN1.5 to WN1.6. The fact that this number is reduced to 16 mappings shows that many of them are still covered by the same conceptual class in SUMO.

Table 3. Distribution of the top 20 assigned SUMO classes

<i>Type</i>	<i>Frequency</i> ¹¹	
	<i>abs</i>	<i>rel</i>
SubjectiveAssessmentAttribute	1,293	5.78%
Device	1,088	4.87%
Artifact	689	3.08%
Motion	583	2.61%
OccupationalRole	555	2.48%
Communication	478	2.14%
Human	460	2.06%
Food	441	1.97%
SocialRole	404	1.81%
Process	379	1.70%
IntentionalProcess	361	1.62%
IntentionalPsychologicalProcess	276	1.23%
Text	247	1.11%
City	246	1.10%
StationaryArtifact	243	1.09%
NormativeAttribute	238	1.06%
EmotionalState	227	1.02%
Clothing	223	1.00%
DiseaseOrSyndrome	220	0.98%
FloweringPlant	205	0.92%

we believe them to show strong selectional preferences as far as their direct objects are concerned. Thus they may serve as proof-of-concept cases for our approach.

The *selectional preference strength* $S_{obj}(lire)$, i.e. the preference strength of 'lire' wrt. to the direct object relation (cf. Section 2.2) is 1.37296, whereas $S_{obj}(manger)$ is 3.46397. This means that 'manger' generally shows a stronger preference wrt. to its direct object than 'lire'. The effect of this is that if 'lire' shows a preference for a particular SUMO class, this preference will weigh more than a preference of 'manger', since it is generally weaker wrt. preferential behaviour. This is due to the fact that the value of the selectional association between a predicate p and the class c of its argument (cf. Section 2.2 above) is normalised by the selectional preference strength of p .

Table 4 shows that 'lire' shows a strong preference for objects of type "Text", and further preferences for "ContentBearingPhysical" and "LinguisticExpression". After these three items, the figures indicate a bigger gap to the next entity. As far as 'manger' is concerned, it shows a very strong preference for direct objects of type "Food", with the second best class ("SelfConnectedObject") reaching just over half of the score for "Food". Looking at these results, it is fair to say that they do match our intuitions.

¹¹ The frequency indicates the number of synsets which have been mapped directly onto the respective SUMO class, so no accumulation of frequency counts along the SUMO hierarchy was made, since that would, of course, leave the top 20 slots in the table to the top 20 nodes in the hierarchy. A synset such as 00058624-n (cf. examples above), which has been mapped onto three different SUMO classes, counts for each of these classes.

Table 4. Selectional preferences of 'lire' ('read') and 'manger' ('eat') wrt. direct objects

<i>SUMO concept c</i>	$A_{obj}(lire, c)$	<i>SUMO concept c</i>	$A_{obj}(manger, c)$
Text	0.3868	Food	0.2179
ContentBearingPhysical	0.2548	SelfConnectedObject	0.1253
LinguisticExpression	0.2431	NonFullyFormed	0.0779
Disseminating	0.1259	Object	0.0745
Communication	0.1083	DevelopmentalAttribute	0.0662
Stating	0.0840	Meat	0.0599
Word	0.0611	Animal	0.0453
Noun	0.0601	Organism	0.0403
Artifact	0.0582	OrganicObject	0.0402
ContentBearingProcess	0.0541	Vertebrate	0.0382
OccupationalRole	0.0498	FruitOrVegetable	0.0368
LinguisticCommunication	0.0409	WarmBloodedVertebrate	0.0295
CorpuscularObject	0.0377	Arachnid	0.0258
name	0.0368	Monkey	0.0239
Book	0.0311	BodySubstance	0.0222
Proposition	0.0295	Mammal	0.0216
SelfConnectedObject	0.0285	AnatomicalStructure	0.0201
Physical	0.0225	Fish	0.0190
FamilyGroup	0.0218	CorpuscularObject	0.0187
destination	0.0156	PlantAnatomicalStructure	0.0184

6 Conclusion

We have presented a generic method for mapping EuroWordNets to the Suggested Upper Merged Ontology and have shown its application to the French EuroWordNet. The mapping procedure builds on existing work on SUMO and version 1.6 of Princeton WordNet [8], EuroWordNet's Inter-Lingual-Index [7] and WordNet's sensemap files. The resulting mapping was used in the calculation of selectional preferences of French verbal predicates with respect to nominal arguments. Preference extraction within an experimental setup shows promising results for the French verbs 'manger' ('eat') and 'lire' ('read').

In the future, we intend to carry out a qualitative evaluation on a larger scale, both for the mapping procedure (cf. Section 3.2) and the extraction of selectional preferences. The ultimate goal is to use the extracted selectional preferences for word sense disambiguation of verbal predicates as well as their arguments, and we will work on this in the near future. Moreover, we plan to consider extracting pairs of subjects and objects in order to calculate preferences of a direct object given the subject and vice versa. Finally, it would be interesting to see how the mapping methodology performs when applied to EuroWordNets other than French, provided that they are linked to the Inter-Lingual-Index as well. We do, however, expect our methodology to be generic enough to be applied to other languages without any major issues.

Acknowledgements

The research described in this work has been carried out as part of the project '*Polysemy in a Conceptual System*' (project B5 of SFB 732) and was funded by grants from the German Research Foundation. I should like to thank Adam Pease, Achim Stein, Piek Vossen, Sabine Schulte im Walde and Christian Hying for their valuable comments and suggestions at the outset of this work, as well as the two anonymous reviewers for helping to improve the structure and content of the paper.

References

1. Resnik, P.: Selectional preference and sense disambiguation. In: Proceedings of the ACL SIGLEX Workshop on Tagging Text with Lexical Semantics: Why, What, and How?, Washington, DC (1997) 52–57
2. Li, H., Abe, N.: Generalizing Case Frames using a Thesaurus and the MDL Principle. *Computational Linguistics* **24**(2) (1998) 217–244
3. Niles, I., Pease, A.: Towards a Standard Upper Ontology. In Welty, C., Smith, B., eds.: Proceedings of the 2nd International Conference on Formal Ontology in Information Systems (FOIS-2001), Ogunquit, ME (2001)
4. Vossen, P., ed.: EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer Academic Publishers (1998)
5. Fellbaum, C., ed.: WordNet: An Electronic Lexical Database. MIT Press (1998)
6. Baker, C.F., Fillmore, C.J., Lowe, J.B.: The Berkeley FrameNet project. In: Proceedings of the ACL/COLING, Montreal (1998)
7. Vossen, P., Bloksma, L., Rodriguez, H., Climent, S., Calzolari, N., Roventini, A., Bertagna, F., Alonge, A., Peters, W.: The EuroWordNet Base Concepts and Top Ontology. (1998)
8. Niles, I., Pease, A.: Linking Lexicons and Ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In: Proceedings of the 2003 International Conference on Information and Knowledge Engineering (IKE '03), Las Vegas, NV (2003)
9. Scheffczyk, J., Pease, A., Ellsworth, M.: Linking FrameNet to the SUMO Ontology. In: Proceedings of the 4th International Conference on Formal Ontology in Information Systems (FOIS-2006), Baltimore, MD (2006)
10. Schulte im Walde, S.: The Induction of Verb Frames and Verb Classes from Corpora. In Lüdeling, A., Kytö, M., eds.: *Corpus Linguistics. An International Handbook. Handbooks of Linguistics and Communication Science*. Mouton de Gruyter, Berlin (To appear)
11. Kullback, S., Leibler, R.A.: On information and sufficiency. *Annals of Mathematical Statistics* **22** (1951) 79–86
12. Stein, A., Schmid, H.: Etiquetage morphologique de textes français avec un arbre de décisions. *Traitement automatique des langues* **36**(1-2) (1995) 23–35
13. Christ, O.: A modular and flexible architecture for an integrated corpus query system. In: Proceedings of the 3rd International Conference on Computational Lexicography (COMPLEX '94), Budapest (1994)
14. Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D.L., Patel-Schneider, P.F., Stein, L.A.: OWL Web Ontology Language Reference. Technical report (2004)
15. Knublauch, H., Musen, M.A., Rector, A.L.: Editing description logic ontologies with the Protégé OWL plugin. In: Proceedings of DL 2004, Whistler, BC (2004)
16. Broekstra, J., Kampman, A., van Harmelen, F.: Sesame: A generic architecture for storing and querying RDF and RDF Schema. In: Proceedings of the 1st International Semantic Web Conference (ISWC '02), Sardinia (2002)