LexiPass methodology: a conceptual path from *frames* to *senses* and back

Alessandro Oltramari

Laboratory for Applied Ontology (ISTC-CNR);
Faculty of Cognitive Science, University of Trento,
Via Solteri 38, 38100 Trento (Italy) - c/o Polo Tecnologico
oltramari@loa-cnr.it

Abstract

In this paper we claim that an integration of FrameNet and WordNet will improve interoperability, user-friendliness and usability of both lexical resources. If the former provides a sophisticated representational structure compared to a narrow lexical coverage, the latter - on the other side - supplies a dense network of word senses and semantic relations although not supporting advanced accessibility (i.e., via frames). According to the integration perspective we present in the paper, we introduce *LexiPass* methodology, which combines Burchardt's tool 'WordNet Detour of FrameNet' with basic statistical analysis, enabling *frame-guided search and extraction* of domain synsets from WordNet.

1. Introduction

Princeton WordNet[1] (Fellbaum 1998) covers several *domains*, namely groups of homogeneous terms referring to the same topic (art, geography, aeronautics, sport, politics, biology, medicine, etc.). In recent years there have been interesting and fruitful attempts to annotate WordNet with domain/topical information in order to improve the overall accessibility to the dense semantic database[2].

A semantic lexicon can be also analyzed from a different perspective, focusing on *frames* (to be conceived as orthogonal to domains):

> Frames are data-structure for representing a stereotyped situation, like being in a certain kind of living room, or going to a child's birthday party. Attached to each frame are several kinds of information. Some of this information is about how to use the frame. Some is about what one can expect to happen next. Some is about what to do if these expectations are not confirmed. (Minsky 1997)

FrameNet[3] aims at providing a lexical account of this kind of 'schematic representations of situations[4]'(Ruppenhofer, Ellsworth et al. 2005). Let's see a sketchy example.

If you point to the **discussion** frame, namely an abstraction of a state of affairs where discussants talk about something in a given place at a given time, you will find several instances in FrameNet (generically called 'lexical units') of different roles (or *frame elements*): i.e., the nouns 'student' and 'advisor' instantiate the ***interlocutor*** role in the frame **discussion**. In principle, the same *lexical unit* may belong to heterogeneous and distinct frames, thus instantiating different roles: the noun 'student', for example, also instantiates the ***person*** role in the **people** frame and so on and so forth.

This paper focuses on the links between WordNet and FrameNet, fostering an integration perspective of the two lexical resources.

[1] http://wordnet.princeton.edu/
[2] i.e. http://multiwordnet.itc.it/english/home
[3] http://framenet.icsi.berkeley.edu
[4] Involving various participants with a specific role.

2. Integrating FrameNet and WordNet

Frames can be considered as 'conceptual pathways' for accessing to a lexical database and extract context-specific information: in principle, if one needs to build a specific lexicon of domain terms (including verbs, nouns, adverbs, adjectives) or populate a domain ontology with lexical entries, she could exploit the domain-related frames in FrameNet in order to retrieve appropriate lexical units. Nevertheless, this picture does not fit to the real potentialities of FrameNet, since this resource contains only unstructured lexical units (8900), compared to a relatively massive network of frames (625); on the contrary, Princeton WordNet incorporates more than 150000 lexical units (organized in a semantic network of synsets) but lacks of a frame-like organization.

This short outline points out that

1) FrameNet is strongly bound to its limited coverage and thus partially inadequate for broad lexical information retrieval;
2) WordNet does not support advanced accessibility (i.e. via frames), although providing a thick *web* of word senses and semantic relations for several domains.

An integration of FrameNet and WordNet then becomes an important requirement for improving interoperability, accessibility and usability of both these lexical resources: Aljoscha Burchardt's tool, 'WordNet Detour of FrameNet'[5] does supply a first step toward such an integration (Burchardt, Erk et al. 2005). Using a specific algorithm, this tool can associate WordNet synsets with FrameNet frames, ranking the results by assigning weights to the discovered connections.

According to this direction of research, we propose a methodology that exploits Burchardt's tool and basic statistical analysis in order to foster a *frame-guided search and extraction* of domain synsets from Princeton WordNet. In the next paragraph we will introduce this methodology referring to an on-going work on the domain of cognition.

[5] http://www.coli.uni-saarland.de/~albu/cgi-bin/FN-Detour.cgi

3. *LexiPass* methodology

If a user wants to extract from WordNet a set of terms referring to the mental realm[6], first of all she needs to focus on the branch whose top node is 'psychological_feature'[7]: in the correlated downward taxonomy she will actually find lexical units referring to cognitive, perceptual and affective entities (mental states, mental representations, emotional features, etc.) organized in suitable synsets[8]. Nevertheless, this simple procedure needs some *add-ins* in order to favour the pulling out of a suitable group of lexical elements linked to the mental domain. Since the rationale behind WordNet semantic network is not strictly inspired by domain-arrangement, we can easily predict that further terms will be definitely found in different locations inside the database (an interwoven network which does not simply rely on topical information). How to find them? How to build a sort of *lexical compass* that can be of assistance in plotting the intended *course* through WordNet senses?

According to our present study, the following methodological steps (to be jointly indicated as *LexiPass* methodology) help to individuate domain-related synsets in the lexical database:

1. use Burchardt's tool 'WordNet Detour of FrameNet' to label every hyponym of a given synset with suitable frames;
2. perform a statistical analysis of the frame distribution[9] within the hyponyms of a given synset. Distinguish within:
   a. Focal Frames
      Frames that occur more than the value of the sum between the mean and standard deviation of the frame distribution;
   b. Peripheral Frames
      Frames that occur more than the value of the mean and less than the sum between the mean and the standard deviation of distribution;
   c. Bare Frames
      Frames that occur less than the mean of the distribution;
3. search in FrameNet[10] every lexical unit associated with Focal Frames (and possibly with Peripheral Frames) which **do not correspond** to the lexical units composing the hyponyms of the considered synset, suitably expanding them in order to isolate correspondent synsets[11];
4. explore and evaluate hyponyms/hyperonyms of the discovered synsets and include domain-related ones into the domain partition of WordNet[12]

We are actually exploiting this methodology to build the lexical layer of a knowledge base for the affective and cognitive domain, ACKIRA[13], whose upper layer is a domain ontology, COMET[14] (Oltramari and Ferrario 2004), linked to DOLCE[15] (Masolo, Gangemi et al. 2002). The idea of ACKIRA emerges from the need of a conceptual clarification from the standpoint of formal ontology of the entities and concepts that play a role in agent technologies for information systems, like mental attitudes, intentional objects, cognitive processes, emotional states, etc. Moreover, as reflected by its acronym, DOLCE has a clear *cognitive orientation*, in the sense that it aims at capturing the ontological categories underlying natural language and human commonsense. Hence, we do not intend DOLCE and ACKIRA's categories to account for the intimate nature of the world, but we rather see them as cognitive artefacts ultimately depending on human perception, cultural imprints and social conventions.

4. Some experimental observations

In a test case consisting of ~430 hyponyms of 'psychological_feature', labeled according to point 1., we found five Focal Frames, namely **awareness**, **sensation**, **feeling**, **desiring** and **state** (represented by the highest five rectangles in fig.1) and four Peripheral Frames, that are **age**, **goal**, **emotion_directed**, **process** (according to point 2.). Let's focus for example on **desiring** frame, from the first group[16]. As the FrameNet documentation explains[17], the generic situation represented by **desiring** deals with "an experiencer that desires that an event occur": this frame clearly concerns the mental domain. Moving to point 3. we find that within the lexical units associated to this frame, there is 'aspiration#1'. Browsing WordNet database we interestingly discover that aspiration#1 ('a will to succeed') does not belong to the 'psychological_feature'-rooted branch, but on the contrary it has 'abstraction#6' as superordinate synset ('a general concept formed by extracting common features from specific examples'). Exploring this branch we are able to locate other relevant synsets for the mental domain, as the following flat list points out:

- **Trait#1**: 'a distinguishing feature of your personal nature'
- **Drive#5**: 'the trait of being highly motivated'
- **Enterprise#3**: 'readiness to embark on bold new ventures'
- **Status seeking#1**: 'a drive to acquire power'

[6] Nouns like 'belief', 'desire', 'emotion', 'feeling'; verbs like 'fear', 'love'; adjectives like 'joyful', 'sad' etc.
[7] In the first releases of WordNet this synset used to be a *unique beginner*, namely a top node of the network.
[8] For the sake of simplicity, in this paper we are using 'synset' in a metonymic sense too, which means not only for indicating sets of synonymous terms but also each single term belonging to a given synset.
[9] Measure how many times every single frame occurs in a selected partition of WordNet (see fig.1).
[10] http://sato.fm.senshu-u.ac.jp/fn23/notes/fullMenuFrame.html
[11] Bare Frames, the least frequent ones, have to be excluded as "noise".
[12] Of course, at a certain level of generality, hyperonyms are domain-independent.
[13] Affective and Cognitive Knowledge Base for Interaction Resources in Agent Technlogies.
[14] Computational Ontology of Mental Entities.
[15] Descriptive Ontology for Linguistic and Cognitive Engineering.
[16] And most relevant, for definition.
[17] http://sato.fm.senshu-u.ac.jp/fn23/helpF/Desiring.html

- **Aggressiveness#1**: 'the quality of being bold and enterprising'
- **Emotionlessness#2**: 'absence of emotion'
- **Folly#1**: the trait of acting stupidly or rashly

From an ontological viewpoint these results show up that the extreme richness of WordNet should be complemented by a deep ontological clarification in order to be usefully and comprehensively exploited. Thus, considering the above-listed terms, **qualities** of persons (indicated by folly#1, aggressiveness#1,...) must be distinguished from generic **states** (**emotionlessness#2**) persons may participate to. Without entering in details here, we can simply say that according to DOLCE's distinctions[18] qualities and states definitely have different characterizations: states do occur at a certain space-time location, being kinds of events, while qualities do not happen (like the state of fear or joy) but are attached – let's say - to people participating in these states.
The test-case presented in this paragraph depicts a situation where *LexiPass* methodology supports the user in finding lexical entries combining WordNet and FrameNet features.

5. Conclusions

So far, preliminary results are encouraging and first experimental data show that the integration between WordNet and FrameNet can open interesting directions of research for developing a new generation of enhanced computation lexicons, extremely useful also as infrastructure in ontology-driven technologies for Semantic Web[19].
Future work will concern a wide-ranging exploitation of the *LexiPass* methodology in the development of ACKIRA.

6. Acknowledgements

I would like to thank Christiane Fellbaum, George Miller, Ben Haskell, Randee Tengi, Susanne Wolf and Helen Langone for the fruitful discussions and comments which inspired this work. This research has been substantially conceived and performed during my visiting period at the Princeton's Cognitive Science Laboratory (March-May 2005).
This work was supported by the FP6 IST project METOKIS (Methodology and Tools Infrastructure for the Creation of Knowledge Units) and PAT (Provincia Autonoma di Trento) project MOSTRO (Modeling Security and Trust Relationships within Organizations).

7. References

Burchardt, A., K. Erk, et al. (2005). A WordNet Detour to FrameNet. Sprachtechnologie, mobile Kommunikation und linguistische Resourcen. H.-C. S. Bernhard Fisseni, Bernhard Schröder and Petra Wagner. Frankfurt am Main, Peter Lang. **8:** 408-421.

Fellbaum, C., Ed. (1998). WordNet - An Electronic Lexical Database, MIT Press.

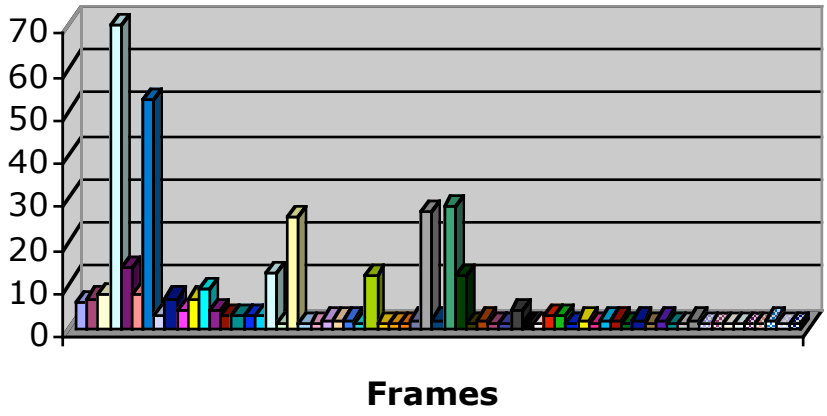
Masolo, C., A. Gangemi, et al. (2002). WonderWeb Deliverable D17: The WonderWeb Library of Foundational Ontologies.

Minsky, M. (1997). A Framework for Representing Knowledge. Mind Design. P. Winston, MIT Press**:** 111-142.

Oltramari, A. and R. Ferrario (2004). Towards a Computational Ontology of Mind. *Formal Ontology in Information Systems (FOIS 2004)*, Torino, IOS Press.

Ruppenhofer, J., M. Ellsworth, et al. (2005). FrameNet: Theory and Practice.

[18] But a similar distinction is also provided by SUMO (Suggested Upper Merged Ontology, http://www.ontologyportal.org/) and OpenCyc (http://www.opencyc.org/).
[19] http://www.w3.org/2001/sw/



FRAME	INSTANCES
Capability	6
Leadership	7
Observable_body_parts	8
Awareness	70
Age	14
Memory	8
Sensation	53
Documents	3
Reason	7
Expertise	4
Perception_experience	7
Communicate_categorization	9
Linguistic_meaning	3
Appearance	3
Project	3
Goal	13
Intentionally_affect	1
Feeling	26
Experiencer_subject	1
Emotion_directed	12
Destroying	1
Expectation	1
Reliance_on_expectation	1
Health_response	2
Desiring	27
State	28
Process	12
Aggregate	1
Certainty	2
Posture	1
Inclination	1
Judgement	4
Subjects_stimulus	1
Biological_urge	3
Temperature	1
Attention	2
Perception	1
Subjective_influence	2
Cogitation	2
Scrutiny	1
Request	2
Deciding	1
Locale	2
Craft	1
People	1
Change_of_leadership	1
Gizmo	2
Hear	2
SUM	**424**
MEAN	**5,047619**
MEDIAN	**2**
STANDARD DEVIATION	**10,48338**
VARIANCE	**109,9013**
AVERAGE ABSOLUTE DEVIATION	**5,504535**

Fig. 1: Frame Distribution in WordNet 'Psychological feature' hierarchy (graph & chart)[20]

[20] Due to space limits, we report here only a partial overview of frame distribution.