

First release of the Multilingual Central Repository of MEANING*

Luís Villarejo, Jordi Atserias, Gerard Escudero, German Rigau

TALP Research Center

Jordi Girona Salgado, 1-3.

08034 Barcelona

{luisv,batalla,escudero}@talp.upc.es

IXA Group

Euskal Herriko Unibertsitatea

Donostia.

{rigau}@si.ehu.es

Resumen: Este artículo contiene una breve descripción de los principales componentes software del Multilingual Central Repository de MEANING y su contenido inicial.

Palabras clave: Wordnet, EuroWordnet, MultiWordnet, Adquisición Multilingüe

Abstract: This paper provides a brief description of the main software components of the Multilingual Central Repository of MEANING and their initial content.

Keywords: Wordnet, EuroWordnet, MultiWordnet

1 Introduction

The MEANING project (Rigau et al., 2002)¹ plans to perform three consecutive cycles of large-scale WSD and acquisition processes in five European languages.

The knowledge acquired for each language will be consistently upload and integrated into the respective local wordnets, and then ported and distributed across the rest of wordnets, balancing resources and technological advances across languages.

The Multilingual Central Repository (MCR) will grant the consistency and integrity of all the semantic knowledge produced by MEANING, acting as a multilingual interface for integrating and distributing all the knowledge acquired in MEANING.

1.1 MCR

The MCR follows the model proposed by the EuroWordNet (Vossen, 1998) project. EuroWordNet is a multilingual lexical database with wordnets for several European languages, which are structured as the Princeton WordNet (Fellbaum, 1998).

The first version of the MCR includes:

- ILI
 - WordNet 1.6
 - EuroWordNet Base Concepts
 - EuroWordNet Top Ontology
 - MultiWordNet Domains

- Local wordnets

– English WordNet 1.5, 1.6, 1.7.1

– Basque, Catalan, Italian and Spanish wordnets

- Large collections of semantic preferences

– Acquired from SemCor

– Acquired from BNC

- Instances

– Named Entities

The Multilingual Central Repository Database plans to represent most of the WordNet and EuroWordNet data and properties, including *language independent information* (ILIs, Top Ontology, domains, etc.) as well as *language dependent data* (synsets, variants, etc.). The first release of the MCR database contains four European wordnets and three different English versions and their corresponding links to the ILI²

In order to upload all this information, local wordnets based on WordNet1.5 has been mapped to Wordnet 1.6. As well as the original set of Base Concepts and the Top Concept Ontology. Moreover the EuroWordNet Top Concept Ontology has been expanded top-down through the WordNet1.6 structure.

1.2 Web Interface to the MCR

The MCR web interface is based on the Web EuroWordNet Interface (WEI)³ (Benítez et al., 1998). The interface provides consulting and editing facilities of the data content of the MCR. The basic aim of this tool is to

* This research has been partially funded by the European Commission (MEANING IST-2001-34460)

¹<http://www.lsi.upc.es/~nlp/meaning/meaning.html>

²<http://www.lsi.upc.es/~nlp/tools/mapping.html>

³<http://nipadio.lsi.upc.es/wei.html>