



VU Research Portal

Introduction to the Special Issue on the BalkaNet Project

Vossen, P.J.T.M.

published in

Romanian Journal of Information Science and Technology
2004

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Vossen, P. J. T. M. (2004). Introduction to the Special Issue on the BalkaNet Project. *Romanian Journal of Information Science and Technology*, 7(1-2).

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Introduction to the Special Issue on the BalkaNet Project

The Princeton WordNet, developed by George Miller and his team, was initially intended as an implementation of the mental lexicon. It represents a semantic network of concepts with semantic relations, where each concept is motivated by a set of English synonyms or synsets. EuroWordNet added a completely new perspective to the Princeton WordNet, namely “multilinguality”. The vocabularies of 7 other languages were organized in a similar way and, more importantly, they were related to each other via a so-called Inter-Lingual-Index or ILI.

Obviously, there will not be any person on this world that has a multilingual mental lexicon for 8 complete languages. Still, the multilingual perspective opens up new possibilities for validating psycholinguistic claims about the organization of the mental lexicon. Fascinating new and old questions immediately come to mind. Are concepts language dependent? Does language affect thinking and reasoning? How universal are concepts, how universal are lexicalization patterns? Is the model of Princeton WordNet valid for all languages? Is the structure Princeton WordNet valid for all languages? Are the relations sufficient for all languages?

EuroWordNet made an early attempt to investigate these questions by defining a set of Based Concepts that play a major role in relating many concepts of most languages and their formalization with a shared top-ontology. A few years after EuroWordNet, the BalkaNet project continued and intensified the discussion. Slavic languages are a bit more alien to English than the other West-European languages. Both morphological structures and cultural background are slightly different. What will this mean for the model? What adaptations are needed?

The reader should not expect that the BalkaNet project specifically focused on answering these questions. The primary goal is the development of the valuable lexical resource for the individual languages and the linking to the other languages via the ILI. Nevertheless, the model was applied and validated by the implementation itself. The Base Concepts have been incorporated and extended. Both the ILI and the methodologies for building wordnets have successfully been applied. This has resulted in a unique resource where people can continue to study these issues.

The consortium has taken up the ideas from EuroWordNet but also investigated new methods and ideas. The parallel *1984* corpus has been used for the development of wordnets but more importantly also for the verification of the cross-lingual perspective. Where wordnets are separately linked to the abstract ILI, the *1984* corpus

relates language use via an interlingual novel. The multilingual relations in the ILLI can be tested by the actual translation equivalences and the corpus can be used to derive many more information for the synsets in each language.

In addition to the many tools and sub-resources developed by the separate partners, a major achievement is the maturation of the VisDic tool. The VisDic database is widely used not only in the BalkaNet project but also outside the project. It is currently the standard database for developing wordnets, enabling formal and conceptual consistency checking and supporting various formats.

The more languages are added to the multilingual wordnet database, the more redundancy of information is represented. The BalkaNet project has effectively exploited these possibilities. The redundancies are important steps to investigate universalities. The new languages also exhibit new language features: explicit morphological structures that reveal relations that have so far not been expressed by other wordnets. The richness of the database, even though the wordnets are not yet complete, will be an inspiration for the development of many more wordnets. It will also inspire many scholars to study the structures and investigate the opportunities.

The extension of the EuroWordNet database with 5 new languages also has an important political and socio-cultural impact. Borders are not only removed physically and politically but, perhaps more importantly, by enabling communication. The BalkaNet project has connected the East with the West of Europe in a very special and unique way.

Piek VOSSSEN

Irion Technologies BV,
Bagijnhof 80, P.O. Box 2849,
2601 CV Delft, The Netherlands
E-mail: piek.vossen@irion.nl