

REVIEWS

Hans C. Boas (ed.). *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Mouton de Gruyter, Berlin, 2009. x + 350 pages. ISBN 978-3-11-021296-9. List price € 9995, US\$ 140.00.

FrameNet, extensively described in *IJL* 16(3) in 2003, is proving its significance by being emulated in other languages. This book, a compendium of papers, describes several of the first such excursions, along with several theoretical discussions that underlie this adoption in languages other than English. The emphasis is on the book's subtitle, in that the various papers provide a rich description of methods and applications that can inform the use of frame semantics in lexicography. The main title can be a little confusing, since the FrameNets are not really multilingual, but simply refers to implementation of the FrameNet approach in other languages, and the lexicography is not really computational, although the implementations are all computer-based. Finally, the lexicography does not correspond to current lexicographic efforts in constructing dictionaries, but what perhaps may emerge in the future if a frame semantic perspective is fully adopted by lexicographers. I will elaborate on these points below.

In *The Oxford Guide to Practical Lexicography*, Atkins and Rundell (2008) describe the incorporation of linguistic theory in lexicography, particularly frame semantics (pp. 144–9). They express the belief that a frame semantics perspective aids in the valency descriptions of lexical items. While they show a sample entry in a dictionary that identifies very general placeholders for surrounding elements (e.g., *ask (somebody) for something*), these descriptions do not really incorporate the frame element characterizations that are central to FrameNet. The present book does not provide any detailed methods by which frames can be further implemented in the creation of ordinary dictionaries. Thus, the book is best viewed as a set of further theoretical developments in frame semantics, rather than a guide to be used in current lexicography.

This book contains 11 papers, consisting of an overview and four parts: (1) principles of constructing multilingual FrameNets, (2) FrameNets for typologically diverse languages, (3) methods for automatically creating new FrameNets, and (4) integrating semantic information from other resources. The overview and Part 1 present the case for a multilingual perspective of FrameNet. Parts 2 and 3 focus on FrameNets for specific languages (Spanish, Japanese, Hebrew, German, and French). Part 4 provides perspectives of other researchers on the question of a universal representation of meaning (i.e., the notion of an interlingua). In describing the individual papers, I will highlight methodological contributions. Collectively, these papers provide many interesting perspectives which can benefit future

lexicographers, developers of FrameNets for other languages, and computational linguists researching frame semantics, semantic components, or semantic roles.

One overall comment about the papers concerns their currency. Except for two reprinted papers, the papers are an outgrowth of presentations at workshops in 2005 and 2006 (Boas (2009)), with some updating that reflects more current work up to 2009. Many papers contain links to web sites that are no longer active or projects that have not changed since the efforts described in the papers. This becomes problematic when proposed continuations of projects and further explorations of ideas have not occurred. In addition, it is difficult to find information about current developments in multilingual FrameNets. Searching the Internet for FrameNet does not yield results which identify FrameNet projects being developed for other languages. That is, it is necessary to search for a specific language (e.g., Swedish FrameNet), even for the projects described in this book (i.e., Spanish FrameNet or Hebrew FrameNet), which contain inactive links. Communications with FrameNet staff (Fillmore (2009); Baker (2009)) indicate that FrameNet projects are under way for at least ten languages beyond those covered in the book.

Another general comment concerning the papers describing FrameNet projects is that many of the investigators spent time working with the FrameNet researchers at International Computer Science Institute in Berkeley. Since they were organizationally apart from the FrameNet team, they were able to bring their individual perspectives in performing their own projects. Thus, the findings and lessons in the various papers are likely to have wide applicability, not only for those developing language-specific FrameNets, but also those performing research in frame semantics.

The overview and Part 1 of the book establish the assertion that FrameNet constitutes a suitable basis for a multilingual database. The overview (*Introduction: Recent trends in multilingual computation lexicography*, Hans Boas, pp. 1–34) provides a short history of multilingual databases, the basic structure of FrameNet, and a short summary of each of the papers. The first two papers in Part 1 (*A bilingual lexical database for Frame Semantics*, Thierry Fontenelle, pp. 37–57 and *Semantic frames as interlingual representations for multilingual lexical databases*, Hans Boas, pp. 59–100) are reprints of IJL papers, respectively, 13(4): 232–248 (2000) and 15(4): 445–478 (2005). Fontenelle describes procedures for using a bilingual dictionary for exploiting data to flesh out frames and frame elements, applying Mel'čuk lexical functions. Boas describes the essential feature of FrameNet that allows for the construction of multilingual FrameNets, namely, the fact that a FrameNet database consists of two components, one providing the machinery for FrameNet (which can be reused in creating a FrameNet for another language) and one containing the lexical items specific to a language. He points out that an essential task in constructing a FrameNet for a language other than English lies mostly in identifying

fine-grained syntactic realizations of frame elements in the other language (while generally being able to reuse the core frame elements). The importance of this point is eventually given support in the following papers implementing FrameNet in the various languages included in this book. The final paper of this part (*The Kicktionary – A multilingual lexical resource of football language*, Thomas Schmidt, pp. 101–32) implements a domain-specific (i.e., football or soccer) FrameNet. This implementation is the one truly multilingual FrameNet (containing English, French, and German lexical units) in the book. He notes that the domain is independent of language, but that language has effects on the frame element specifics. He identifies two issues with using FrameNet in multiple languages: the problem when lexical units have no translation equivalents and the need for other types of semantic relations than those encompassed in FrameNet (such as the kinds of relations included in WordNet, including synonymy, hypernymy, and meronymy), particularly for nouns.

Part 2 includes three papers describing implementation of FrameNets for typologically diverse languages: Spanish, Japanese, and Hebrew. The first paper (*Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon*, Carlos Subirats, pp. 135–62), although nominally not typologically different from English, exemplifies many of the core issues in extending FrameNet to other languages. First, Subirats demonstrates how he was able to use the FrameNet workflow model of tagging sentences with their frame elements for target lexical units; this paper describes the acquisition of the necessary corpus and the applicability of the FrameNet machinery for this new language. Second, this implementation underscores that many language-specific differences will emerge, primarily in nuances in the applicable frame elements for a given frame and in language divergences where different aspects of frame elements may be realized in different syntactic constructions. This FrameNet implementation is also intended to tag full texts (i.e., all the frames triggered by lexical units, rather than focusing on a single target in a sentence); thus, the paper describes issues arising when particular syntactic constituents of the sentence are shared by more than one frame.

The second paper in Part 2 (*Frame-based contrastive lexical semantics in Japanese FrameNet: The case of risk and kakeru*, Kyoko Hirose Ohara, pp. 163–82) focuses on issues pertaining to a single frame. Ohara's paper reinforces the point raised by Subirats, that lexical units in different languages may have different central notions and hence not be equivalent. He calls into question the notion that FrameNet can be used as a true interlingual mechanism, since it will fail at the level of the individual lexical units. The third paper in Part 2 (*Typological considerations in constructing a Hebrew FrameNet*, Miriam Petrucci, pp. 183–205) presents the types of considerations that are relevant in the beginning stages of an effort to implement FrameNet in another language. She identifies the challenges faced for a language in which few linguistic resources have been developed, particularly the need to develop an appropriate

infrastructure for a Hebrew FrameNet. But again, she underscores the basic issues of underspecification, polysemy, frame differences, and frame element membership.

Part 3 includes two papers describing methods for automatically creating new FrameNets, in German and French, highlighting some differences from the basic FrameNet workflow model. The first paper (*Using FrameNet for the semantic analysis of German: Annotation, representation, and automation*, Aljoscha Burchardt et al., pp. 209–44) presents an extensive array of tools used in developing a new FrameNet. Many of these tools are publicly available and can be used, if not specifically, at least for providing an in-depth set of examples for use in implementing a FrameNet. This paper covers the creation of a corpus-driven resource, with the possibility of creating proto-frames on the fly, and addresses questions about dealing with such items as support verb constructions (which differ from English), idioms, and metaphors. The paper also describes techniques for creating lexicons from the corpora and how to deal with the core issues of missing frames and different syntactic realizations. The paper also describes a tool for visualizing cross-language frames and their valences. The second paper in this part (*Cross-lingual labeling of semantic predicates and roles: A low-resource method based on bilingual L(atent) S(ematic) A(nalysis)*, Guillaume Pitel, pp. 244–84) focuses on early stages in implementing a French FrameNet and presents a set of novel techniques for bootstrapping a FrameNet from various language resources. The FrameNet lexicon for French is constructed from existing French-English word-by-word translation resources, with the English FrameNet entries translated into French to give an initial identification of the applicable frames for the French FrameNet; this facilitates a subsequent stage of manual pruning. The next stage involves the use of a bilingual aligned corpus and the latent semantic analysis technique with a metric by which words with semantic relations are considered close to each other. While this approach is not deemed conclusive, it provides a more rapid approach to building a FrameNet for another language. At the same time, it also identifies the problems of null alignments and non-frame-conserving translations.

Part 4 includes two papers that consider the use of other resources for multilingual lexicography. The first paper (*Interlingual annotation of multilingual text corpora and FrameNet*, David Farwell et al., pp. 287–317) describes a project for interlingual annotation of six bilingual parallel corpora. The main objective of this project was the development of an interlingua. The paper lays out steps used to provide several layers of annotation (syntax; ontology, roles, and semantic representation; and the final layer providing the interlingual representation of meaning). Tools and interannotation (evaluation) metrics are described for the first annotation layers, but the final annotation remained

some
ing th
H (199
perspe
the En
Assess
Grid
describ
and th
(KIF
They a
tions.
alterna
efforts
It is
these d
explora
Most
than re
provide
within
age is
the sup
with it.
I find
FrameN
expans
and fra
to man

Referen
Atkins, B
Oxford
Baker, C
Boas, H.
Fillmore,

Kenneth
CL Rese
ken 3 (199
doi:10.1017
Advance

sources the basic
frame element

initially creating
differences from
FrameNet for
automation,
of tools used
available and
set of exam-
the creation of
frames on the
support verb
The paper
and how to
realizations.
frames and their
(semantic pred-
(Semantic)
an implement-
for bootstrap-
Net lexicon for
and translation
French to give an
FrameNet; this
stage involves the
analysis technique
transferred close to
provides a more
At the same
and non-frame-

resources for multi-
multilingual
describes a
The main
The paper lays
ontology, roles,
the interlingual
(evaluation) metrics
remained

somewhat speculative in this paper. No further updates were found in search-
ing the web. The second paper (*Universals and idiosyncrasies in multilingual
WordNets*, Piek Vossen and Christiane Fellbaum, pp. 319–45) presents the
perspectives gained from the implementation of a multilingual WordNet, in
the EuroWordNet project and in subsequent efforts in the Global WordNet
Association. Vossen and Fellbaum describe a project for a Global WordNet
Grid, that is, an ontology as the basis for a universal concept index. They
describe the components of this ontology, making use of specific formalisms
and ontological efforts (i.e., the use of the Knowledge Interchange Format
(KIF) expressions and the Suggested Upper Merged Ontology (SUMO)).
They also identify the usual types of problems with crosslingual representa-
tions: mismatches across languages, lexicalization differences, and syntactic
alternations. Although global WordNet efforts are continuing, no further
efforts in the development of the grid have occurred.

In summary, while the papers in this book are slightly dated, particularly
those in the final part, they provide a compendium of methods that can be
explored in the development of a FrameNet for any non-English language.
Most importantly, the collection provides an overall perspective on this task
that cannot be gained by web searches. Almost as important, the collection
provides perspectives that can be used in research on frame semantics, even
within a single language. A prominent criticism of FrameNet is that its cover-
age is unknown. The papers in this book help considerably in providing
the scope of this problem and in identifying steps that can be used in dealing
with it.

I found that in reviewing this book, as one engaged in using and investigating
FrameNet, I was led to many online investigations and searches which greatly
expanded my understanding of many of the issues surrounding FrameNet
and frame semantics. I expect that other readers will similarly be stimulated
to many paths of investigation.

References

- Atkins, B. T. S. and Rundell, M. 2008. *The Oxford Guide to Practical Lexicography*.
Oxford: Oxford University Press.
Baker, C. 2009. Personal communication.
Boas, H. 2009. Personal communication.
Fillmore, C. 2009. Personal communication.

Kenneth C. Litkowski
CL Research, Damascus, Maryland USA
ken@clres.com
doi:10.1093/ijl/lec034
Advance access publication 9 December 2009