

Toward an Open Multilingual Shared Semantic Lexicon The Shared Wordnet

Francis Bond

Linguistics and Multilingual Studies
Nanyang Technological University

NICT, Japan

`bond@ieee.org`

2012-06-13

GWC-2012

Overview

- Research on Computational Linguistics and Nanyang

Bond Lab

- Toward an Open Multi-lingual WordNet
- What is a free license and why does it matter?
- A survey of released wordnets and their licenses
- The effects of license choice
- Discussion

Bond Lab Members

- Petter Haugereid (MT and HPSG)
- Matthieu Morey (MRS comparison and Graph Matching) Erasmus Post Doc visitor
- Fan Zhenzhen (Chinese HPSG)
- Tan Liling (Multilingual Corpus Construction)
- Lea Frermann (Cross-lingual Disambiguation) Erasmus MA visitor
- Dominikus Wetzel (rephrasing for SMT: negation) Erasmus MA visitor
- Suerya, Nurriil, Helmy (Wordnet Bahasa)
- Huini, ShuWen, Jeanette, Yui Ting (Chinese/Japanese Wordnets)

The Core Problem of MT (& NLU)

(1) 頭 を 掻いた
atama wo kaita
head ACC scratched
“I scratched **my** head.”

- The Japanese text doesn't say
 1. That 掻く should be *scratch*, not *shovel*, *row*, ...
 2. Who scratched
 3. That 頭 should be *head*, not *boss*, *top*, ...
 4. That *head* needs a possessive pronoun
 5. Whose head it is
- A native speaker of Japanese would know (2,5), could deduce (1,3)
- A native speaker of English knows (4)
- ? How do we teach a computer?

Languages Mark Different Things

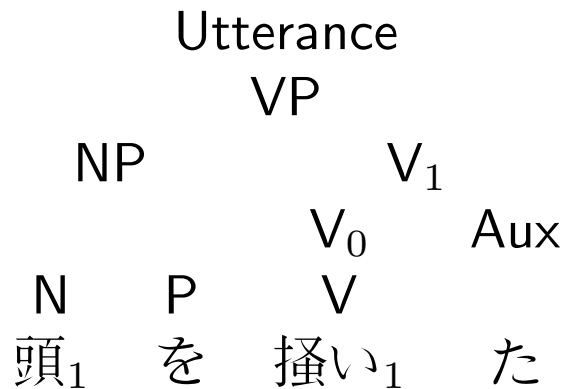
- Most languages care about **possession**
 - English: pronouns
my head
 - Japanese: politeness, evidentiality
your honorable head vs my head
I itch vs you seem to itch
 - Russian: reflexives
I scratch self head
 - Swedish: definiteness
I scratch the head (head-et)
- Shared level somewhere beyond syntax
This is the level that interests me: **semantics**

Exploit Cross-lingual Differences

- Build on a core of hand written monolingual and cross-lingual resources
- Learn as much as we can from other languages
 - Structural Semantics
 - Lexical Semantics

Current Work: Rich Annotation

- (2) 頭 を 掻いた
atama wo kaita
head ACC scratched
“I scratched my head.”



Syntax & Lexical Semantics

atama ₁	is-a	bodypart
kaku ₁	is-a	itch
kaku	ARG1	zero-pronoun
kaku	ARG2	atama
kaku	TENSE	past

Structural Semantics

What does this give us?

- As a scientist — semantics motivates syntax
- As an engineer
 - We want semantic output and we want higher accuracy
 - Syntax helps with WSD (Stevenson)
 - Semantics helps with parse ranking (Fujita et al, Egirre et al)
 - We want to look at **joint models**

Current research: Annotation

- Empirical NLP needs data
- There is no corpus marked up with what I want (Ontonotes is close)
 - Structural Semantics
Predicate-Arguments, Deep Case, Quantification
 - Lexical Semantics
Context dependent sense information from an ontology
- With this data we could start to answer some of the questions
 - Still need more discourse/context

-
- So we add the information
 - HPSG Tree banking
 - WordNet Sense annotation
 - Open data

Bootstrapping Annotation Bilingually

- Languages are ambiguous in different ways

(3) ϕ 頭を掻いた

I scratched my head

- We can use the two sentences to disambiguate each other:
 - Zero-pronoun is 1st person ($I \leftrightarrow \phi$)
 - $kaku_1 \leftrightarrow scratch_3$
 - $atama \leftrightarrow head$ could be boss or body part or top, not grammatical head
- This also works for some syntactic disambiguation

Open Research Questions

- How much can be disambiguated in this way?
depends on language pair, translation quality
Ja↔En, Tanaka Corpus reduce ambiguity to $\approx 30\%$

- What is the best level for alignment?
 - deeper processing aligns better
 - less robust, so less cover

- Currently experimenting with Japanese and English
 - HPSG grammars for parsing DELPH-IN
 - WordNet for the sense inventory Global WordNet Association
 - MT system for deep alignment DELPH-IN
 - n-grams for shallow alignment MOSES
 - MRS n-grams for hybrid alignment DELPH-IN/MOSES

Concrete Projects

- Wordnets (Global Wordnet association)
 - **Japanese WordNet** (with Toyohashi, NICT)
 - **Wordnet Bahasa** (with MMU, BPPT)
 - Chinese WordNet (with South-Eastern)
 - English Wordnet (with Princeton)
- Grammars (HPSG: DELPH-IN)
 - Japanese grammar: NTU + Darmstadt Poly + Washington
 - Chinese: DFKI + Shanghai + NTU
 - Eng: Stanford

-
- NTU Multilingual Corpus
 - JEC sentence aligned word-net tagged
 - JECIVK sentence aligned (JECI will be tagged)
 - Cross-lingual disambiguation (sense, structure)
 - MA: Tan Liling
 - Japanese-English MT
 - Korean-Japanese-English MT (with KHU in Korea)

Semantic Transfer-based MT

- HPSG grammars for parsing and generation
- stochastic models at every step
(parsing, transfer, generation, end-to-end)
- core of hand-written rules
- open rules (some quite complex) learned from corpora
- limited coverage, high quality (beating MOSES 58 to 42))
- Now J-E, soon to add K-J-E, like to add C-E, M-E

Jaen Latest results

	BLEU	METEOR	HUMAN
JaEn First	16.77	28.02	58
MOSES	30.19	31.98	42

BLEU Comparison of Jaen loaded with the Combined rules, and MOSES (1194 items)

- S: 偏見は持つべきではない。
- R: We shouldn't have any prejudice.
- M: You should have a bias.
- J: I shouldn't have prejudice.

Cross Lingual Disambiguation

Reduced ambiguity to 30% (for sentences that parse and partially translate: 71%).

	English		Japanese	
	Prec	F	Prec	F
Included	0.820	0.897	0.804	0.887
First Rank	0.659	0.791	0.676	0.803
MRR	0.713	0.829	0.725	0.837

Accuracy and F-scores for disambiguation performance of our system. Recall was 99% in every case. 'Included': inclusion of the gold parse in the reduced set of parses or not. 'First Rank': ranking of the preferred parse as top in the reduced list. 'MRR': mean reciprocal rank of the gold parse.

Bahasa WordNet (i)

- Standard Extend Model (add words to English synsets)
MIT license (or CC BY)?
- 14,000 Senses (automatically)
 - learned from FEM, KAMI
 - cross lingual Eng-Fre/Chn-Mly (Fre, Chn WN)
 - with Hypernym (Goi-Taikei for KAMI, Fre for FEM)
- 5,000 hand checked
- Single representation for Malay and Indonesian
 - zsm \subset msa
 - ind \subset msa

Bahasa WordNet (ii)

- Merge with (hitherto unreleased)
 - Malay Wordnet; Indonesian Wordnet

Wordnet	Lang	Synsets	Words	Senses
Indonesian Wordnet	ind	27,506	30,358	57,560
Malay Wordnet	zsm	23,953	23,833	48,996
Wordnet Bahasa	ind	19,316	19,522	48,111
	zsm	19,347	19,572	48,181
Combined	ind	48,689	58,541	133,005
	zsm	38,736	45,664	114,025

The combined wordnet has 8,200 definitions;
85,315 of the senses are shared (ind and zsm).

NTU Multilingual Corpus

- Data Sources
 - 2,500 sentences of tourist data (STB) CEIKJV
 - Cathedral and the Bazaar (CEJ)
 - Sherlock Holmes (CEJ)
 - Mainichi News (CEJ)
 - 2,000 sentences of medical tourist data (STB) ACEIJV
- Sentence segmented; POS tagged
- Wordnet annotation for CEJI (not medical yet)
 - Some annotation done as part of semantics class

WordNets

Motivation

- Would like to compare meaning across languages quantitatively if possible
- Experiments tend to end up measuring resource maturity e.g., semantic correlation of countability, classifiers, valence
- Want a collection of resources of comparative completeness
 - Build our own (Japanese, Malay)
 - Contribute to others
- Resources need to be **accessible** and **usable**
 - legally OK to use
 - of sufficient quality, size and with a documented interface

WordNet

- Princeton WordNet (PWN) is an open-source electronic lexical database of English, developed at Princeton University
<http://wordnet.princeton.edu/>
- Made up of four linked semantic nets, for each of nouns, verbs, adjectives and adverbs
- Wordnets exist for many, many languages
- None are as mature as PWN

Composition

- Lexical items are categorised into ~115,000 glossed **synsets** (= synonym sets)
 1. enrichment -- (act of making fuller or more meaningful or rewarding)
 2. enrichment -- (a gift that significantly increases the recipient's wealth)

- Lexical relations at either the synset level or sense (= combination of lexical item and synset) level

- Strongly lexicalist (originally):
 - synsets only where words exist
 - but many multiword expressions ($\approx 50\%$)

Psycholinguistic Foundations

- Strong foundation on hypo/hypernymy (lexical inheritance) based on
 - response times to sentences such as:
 - a canary {can sing/fly,has skin}*
 - a bird {can sing/fly,has skin}*
 - an animal {can sing/fly,has skin}*
 - analysis of anaphora:
 - I gave Kim a novel but the {book,?product,...} bored her*
 - Kim got a new car. It has shiny {wheels,?wheel nuts,...}*
 - selectional restrictions

Word Meaning as a Graph

- You need a very big graph to capture all meanings

An Accessible Multilingual Wordnet

- To help us in disambiguation when making the Japanese and Bahasa wordnets we needed to link various wordnets
- There were many small idiosyncrasies
- To make it easier for others we have released our combined database + scripts
only for those resources whose license allows it
- Hope to be superseded by a more flexible framework (ILI)
 - That allows new (especially) non-English synsets
 - That allows variants

Current State

Wordnet	Lang	Synsets	Words	Senses	Core	Licence
Arabic WordNet	arb	10,165	14,595	21,751	48%	CC BY SA 3.0
Princeton WordNet	eng	117,659	148,730	206,978	100%	wordnet
Persian Wordnet	fas	17,759	17,560	30,461	41%	Free to use
FinnWordNet	fin	116,763	129,839	189,227	100%	CC BY 3.0
WOLF	fre	32,466	37,996	46,188	48%	CeCILL-C
Hebrew Wordnet	heb	5,448	5,325	6,872	27%	GPL
Japanese Wordnet	jpn	57,178	91,959	158,062	95%	wordnet
Wordnet Bahasa	ind	19,260	19,659	48,317	98%	MIT
Wordnet Bahasa	zsm	19,267	19,638	48,321	98%	MIT
OpenWN-PT	por	34,087	35,811	51,471	77%	CC by SA 3.0
Thai Wordnet	tha	73,350	82,504	95,517	81%	wordnet

- <http://casta-net.jp/~kuribayashi/multi/>
- **Coming Soon:** Italian; Spanish, Catalan, Galician, Basque Slovenian; Danish, Norwegian

What is lacking?

- Proper handling of orthographic variants
 - Japanese: 桧, 檜, ひのき, ヒノキ, 火の木 *hinoki*
 - Hebrew, Arabic: with and without diacritics
 - English: color, colour; data base, data-base, database
- Richer morphological information (not just v,a,n,r)
- Substructure for MWEs
- Sense specific frequencies

Multi-lingual issues

- Representation of lexical gaps
 - some concepts are only lexicalized for some languages
- Relations may vary for different languages
- Names for synsets

Next steps

- Standard data formats
 - Sharing tools and corpora
 - Sharing best practice
- Opening up data
- Cooperative Interlingual Index

Issues with linking Wordnets

- English specific synsets
 - identification
 - improvement
- Non-English synsets: the return of the Interlingual-Index
 - *last year*
 - *tsumetai vs samui*
- Language Independent WSD + language dependent WSD

Conclusions

- Please let us know if we have missed your wordnet :-)
- Please release your data in a standard format
 - SQL
 - LMF
 - Tab separated values

What is a free license?

The Open Source Definition

1. Free Redistribution
2. Source Code
3. Derived Works

-
4. Integrity of The Author's Source Code
 5. No Discrimination Against Persons or Groups
 6. No Discrimination Against Fields of Endeavor
 7. Distribution of License
 8. License Must Not Be Specific to a Product
 9. License Must Not Restrict Other Software
 10. License Must Be Technology-Neutral

Why Open?

- Research should be open in principle:

... the principle of openness in research - the principle of freedom of access by all interested persons to the underlying data, to the processes, and to the final results of research - is one of overriding importance.

- Making resources Open Source removes difficulties in distributing work or in continuing work at another institution.
- CL needs serious resources
 - They cannot be built and maintained by a single group
 - Open Source is the only practical way of achieving flexible multi-group collaboration

-
- CL needs standards and historically the successful ones have been created bottom-up.
 - Seeing one's work used by other groups is very rewarding.
 - People are generally enthusiastic about contributing to widely used work.
 - Researchers are evaluated by the impact that their work has. Open Source work generally has more impact.
 - Licensing attempts are extremely time-consuming.
 - In my personal opinion, **Restricting access to information about human language, the great achievement of our species, is immoral.**

Wordnets in the world

Free Wordnets

Name	Language	# Synsets	Release	License
Princeton WN* [€]	English	155,000	1991	WordNet
	Fellbaum (1998): 6,821 citations			
FinnWordNet	Finnish	117,700	2010	WordNet
	Lindén and Carlson. (2010): 0 citations			
Russian WN	Russian	117,000	2004	Wordnet
	Balkova et al. (2008): 15 citations			
Thai Wordnet	Thai	73,593	2007	WordNet
	Thoongsup et al. (2009): 4 citations			
DanNet*	Danish	65,000	2008	WordNet
	Pedersen et al. (2009): 8 citations			
Japanese WN*	Japanese	57,000	2009	WordNet
	Isahara et al. (2008): 24 citations			
Catalan WN*	Catalan	42,000	1999	GPL
	Benítez et al. (1998): 17 citations			

⊙ A subset released under a less restrictive license; [€] A version from EuroWordNet is also available from ELRA

Name	Language	# Synsets	Release	License
MultiWN* [€]	Italian Pianta et al. (2002): 143 citations	38,877	2008	CC BY
LSG	Irish Gaelic Scannell (2003): 11 citations	32,742	2003	GNU FDL
Hindi WN	Hindi Jha et al. (2001): 10 citations	28,687	?	GNU FDL
WOLF	French Sagot and Fišer (2008): 22 citations	22,000	2009	Cecill-C [†]
Bahasa*	Malay/Indonesian Nurriil <i>et al.</i> 2011: —	20,000	2011	CC BY
Spanish WN* [⊙] [€]	Spanish Farreres et al. (1998): 65 citations	15,556	2006	LGPL
Catalan WN* [⊙] [€]	Catalan Benítez et al. (1998): 17 citations	15,556	2006	LGPL

[⊙] A subset released under a less restrictive license; [€] A version from EuroWordNet is also available from ELRA

Name	Language	# Synsets	Release	License
Arabic WN*	Arabic	11,269	2008	CC BY SA
	Black et al. (2006): 28 citations			
Hebrew WN*	Hebrew	5000	2006	GPL
	Ordan and Wintner (2007): 0 citations			

* Results from our survey; Citation counts from Google Scholar (accessed on 2011-09-23)

† A variant of the LGPL

⊙ A subset released under a less restrictive license

€ A version from EuroWordNet is also available from ELRA

Release is the first release under this license

Free for Research Wordnets

Name	Language	# Synsets	Release	License
Chinese WN*	Chinese	115,424	2008	research/com. Xu et al. (2008): 0 citations
KorLex*	Korean	90,000	2007	research/com. subset of Yoon et al. (2009), no reference for nouns only
Spanish WN* [€]	Spanish	62,000	1999	research/com. Farreres et al. (1998): 65 citations
Cornetto* [€]	Dutch	70,371	2009	research/com. Vossen et al. (2008): 19 citations
GermaNet* [€]	German	69,594	2011	research/com. Kunze and Lemnitzer (2002): 52 citations
MWN*	Macedonian	33,276	2010	CC BY NC Saveski and Trajkovski (2010): 0 citations

⊙ A subset released under a less restrictive license; € A version from EuroWordNet is also available from ELRA

Name	Language	# Synsets	Release	License
Ro-WordNet*	Romanian	30,000	soon	no-derive.
	Tufiş et al. (2008): 9 citations			
Czech WN * [€]	Czech	29,000	1999	research/com.
	Pala and Smrř (2004): 34 citations			
SloWnet*	Slovene	20,000	2010	CC BY NC SA
	Fiřer and Sagot (2008): 13 citations			

⊙ A subset released under a less restrictive license; [€] A version from EuroWordNet is also available from ELRA

Available for Research Wordnets

Name	Language	# Synsets	Release	License
KorLex*	Korean	130,878	2007	research/com.
	Yoon et al. (2009): 5 citations			
Estonian* [€]	Estonian	47,000	—	ELRA
	Kerner et al. (2010): 0 citation			
BasqWN	Basque	30,281	?	ELRA
	Pociello et al. (2011): 0 citation			
BulNet* [⊙]	Bulgarian	23,715	2004	ELRA
	Koeva (2008): 3 citations			

[⊙] A subset released under a less restrictive license; [€] A version from EuroWordNet is also available from ELRA

Euro Wordnet

Name	Language	# Synsets	Release	License
EuroWordNet				
English Addition	English	16,361	1999	ELRA
Dutch	Dutch	44015	1999	ELRA
Spanish	Spanish	23370	1999	ELRA
Italian	Italian	48529	1999	ELRA
German	German	15,132	1999	ELRA
French	French	22,745	1999	ELRA
Czech	Czech	22,745	1999	ELRA
Estonian	Estonian	9,317	1999	ELRA
ItalWordNet	Italian	49,360	1999	ELRA
Vossen (1998): 728 citations				

- These are sold by ELRA regardless of the availability of newer free versions.

⊙ A subset released under a less restrictive license; € A version from EuroWordNet is also available from ELRA

Comments

- Most researchers very forward about sharing their data
 - Will share on request
 - Can often download (even for ELRA licensed Wordnets)!
 - Data occasionally redistributed under different license
- Some reasons given for opening data
 - Desire to make language resources publicly available
 - To encourage use and avoid license issuing overhead
- Some reasons given for not opening data
 - Financial — get money to support further construction
 - Quality Control — want to stop low quality forks

Effects of different licenses

Size	Date	Open	Free	Non free
Large	2009	Danish/Thai 8/4		Korean 5
Large	2008	Japanese 24	Dutch 19	
Small	2008	French 22	Slovenian 13	Bulgarian 3

Uptake of a resource partially depends on how **usable** (legally accesible) the resource is.

Exceptions

- Chinese wordnet (Xu et al., 2008) has no citations, despite it being large and **free for research**. We suspect that this is because you must email and ask for permission to use it, which is a substantial barrier to use.
- The Italian wordnet (Pianta et al., 2002) has a very high number of citations. In this case it was developed as part of a multilingual wordnet with several other languages, thus giving it a large citation group.
- Finally, the Thai wordnet (Thoongsup et al., 2009), has relatively few citations, in this case it is also a part of a

large project (the Asian Wordnet: Sornlertlamvanich et al. (2008)) which gets more citations (10).

Reasons for not releasing

- Quality Control
- Non-free resources used in construction
- Sustainability (covering construction costs)

Best Practice

- Open, standard license
- Direct download
- Standard, documented format

Thanks

We would like to thank:

- The compilers of the Global WordNet Association's *Wordnets in the World* page
- Everyone who responded to the survey
- The Creative Commons Catalyst Grant: *Assessing the effect of license choice on the use of lexical resources*
- All the Wordnet developers



References

Balkova, V., Sukhonogov, A., and Yablonsky, S. (2008). Russian wordnet: From UML-notation to internet/intranet database implementation. In Sojka, P., Pala, K., Smrž, P., Fellbaum, C., and Vossen, P., editors, *Proceedings of the Second International WordNet Conference — GWC 2004*, pages 31–38, Brno.

Benítez, L., Cervell, S., Escudero, G., López, M., Rigau, G.,

and Taulé, M. (1998). Methods and tools for building the Catalan wordnet. In *ELRA Workshop on Language Resources for European Minority Languages*, Granada, Spain.

Black, W., Elkateb, S., Rodriguez, H., Alkhalifa, M., Vossen, P., Pease, A., Bertran, M., and Fellbaum, C. (2006). The Arabic wordnet project. In *Proceedings of LREC 2006*.

Farreres, X., Rigau, G., and Rodríguez, H. (1998). Using wordnet for building wordnets. In *COLING-ACL Workshop on Usage of WordNet in Natural Language Processing Systems*, Montreal, Canada.

Fellbaum, C., editor (1998). *WordNet: An Electronic Lexical Database*. MIT Press.

Fišer, D. and Sagot, B. (2008). Combining multiple resources to build reliable wordnets. *Text, Speech and Dialogue*, LNCS 2546:61–68.

Isahara, H., Bond, F., Uchimoto, K., Utiyama, M., and Kanzaki, K. (2008). Development of the Japanese WordNet. In *Sixth International conference on Language Resources and Evaluation (LREC 2008)*, Marrakech.

Jha, S., Narayan, D., Pande, P., and Bhattacharyya, P. (2001).

A wordnet for Hindi. In *International Workshop on Lexical Resources in Natural Language Processing*, Hyderabad.

Kerner, K., Orav, H., and Parm, S. (2010). Growth and revision of Estonian wordnet. In Bhattacharyya, P., Fellbaum, C., and Vossen, P., editors, *5th Global Wordnet Conference: GWC-2010*, pages 198–202, Mumbai.

Koeva, S. (2008). Derivational and morphosemantic relations in Bulgarian wordnet. In *Intelligent Information Systems, XVI*, pages 359–389, Warsaw. Academic Publishing House.

Kunze, C. and Lemnitzer, L. (2002). Germanet —

representation, visualization, application. In *LREC*, pages 1485–1491.

Lindén, K. and Carlson., L. (2010). Finnwordnet — wordnet påfinska via översättning. *LexicoNordica — Nordic Journal of Lexicography*, 17:119–140. In Swedish with an English abstract.

Mohamed Noor, N., Sapuan, S., and Bond, F. (2011). Creating the open Wordnet Bahasa. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation (PACLIC 25)*, pages 258–267, Singapore.

Ordan, N. and Wintner, S. (2007). Hebrew wordnet: a test case of aligning lexical databases across languages. *International Journal of Translation*, 19(1):39–58.

Pala, K. and Smrž, P. (2004). Building Czech wordnet. *Romanian Journal of Information Science*, 7:79–88.

Pedersen, B., Nimb, S., Asmussen, J., Sørensen, N., Trap-Jensen, L., and Lorentzen, H. (2009). DanNet — the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation*.

Pianta, E., Bentivogli, L., and Girardi, C. (2002). Multiwordnet:

Developing an aligned multilingual database. In *In Proceedings of the First International Conference on Global WordNet*, pages 293–302, Mysore, India.

Pociello, E., Agirre, E., and Aldezabal, I. (2011). Methodology and construction of the Basque wordnet. *Language Resources and Evaluation*, 45(2):121–142.

Sagot, B. and Fišer, D. (2008). Building a free French wordnet from multilingual resources. In (ELRA), E. L. R. A., editor, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.

Saveski, M. and Trajkovski, I. (2010). Automatic construction of wordnets by using machine translation and language modeling. In Tomaz Erjavec, J. Z. G. L., editor, *In Proceedings of Seventh Language Technologies Conference, 13th International Multiconference Information Society*, volume C.

Scannell, K. P. (2003). Automatic thesaurus generation for minority languages: an Irish example,. In *TALN Workshop on Traitement Automatique des Langues Minoritaires et des Petites Langues*, volume 2, pages 203–212, Batz-sur-Mer.

Sornlertlamvanich, V., Charoenporn, T., Robkop, K., and

Isahara, H. (2008). KUI: Self-organizing multi-lingual wordnet construction tool. In Tanács, A., Csendes, D., Vincze, V., Fellbaum, C., and Vossen, P., editors, *4th Global Wordnet Conference: GWC-2008*, pages 417–427, Szeged, Hungary.

Thoongsup, S., Charoenporn, T., Robkop, K., Sinthurahat, T., Mokarat, C., Sornlertlamvanich, V., and Isahara, H. (2009). Thai wordnet construction. In *Proceedings of The 7th Workshop on Asian Language Resources (ALR7), Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics (ACL) and the 4th International Joint Conference on Natural Language*

Processing (IJCNLP), Suntec, Singapore.

Tufiş, D., Ion, R., Bozianu, L., Ceaşu, A., and Ştefănescu, D. (2008). Romanian wordnet: Current state, new applications and prospects. In *Proceedings of the 4th Global WordNet Association Conference*, pages 441–452, Szeged.

Vossen, P., editor (1998). *Euro WordNet*. Kluwer.

Vossen, P., Maks, I., Segers, R., and Van der Vliet, H. (2008). Integrating lexical units, synsets and ontology in the Cornetto database. In *LREC 2008*, Marrakech, Morocco. European Language Resources Association (ELRA).

Xu, R., Gao, Z., Qu, Y., and Huang, Z. (2008). An integrated approach for automatic construction of bilingual Chinese-English WordNet. In *3rd Asian Semantic Web Conference (ASWC 2008)*, pages 302–341.

Yoon, A., Hwang, S., Lee, E., and Kwon, H.-C. (2009). Construction of Korean wordnet KorLex 1.5. *Journal of KIISE: Software and Applications*, 36(1):92–108.