Thai FrameNet Construction and Tools

Dhanon Leenoi, Sawittree Jumpathong, Peerachet Porkaew and Thepchai Supnithi
Human Language Technology Laboratory
National Electronics and Computer Technology Center
Pathumthani, Thailand
{dhanon.leenoi; sawittree.jumpathong; peerachet.porkaew; thepchai.supnithi}@nectec.or.th

Abstract

This paper presents the construction of Thai FrameNet through a combination approach: the expand and the merge. The alignment of the original FrameNet is made through the former whilst the latter manifests the indigenous Thai concepts. After a combination approach is implemented, frames are organised in the Thai FrameNet database and linked by seven frame-to-frame relations. To envisage frame relations, a visualised tool was developed. Substantially, this tool assists the frame developers to invent the genuine Thai frames and to place on Thai FrameNet structure.

Keywords

Thai FrameNet; Frame Semantics; Annotation Tool; Visualised Tool

1 Introduction

The Berkeley FrameNet (Ruppenfer et al, 2010; Baker, 1998), a lexical resource based on the brainchild of Charles J. Fillmore, Frame Semantics (Fillmore, 1982; Fillmore, 1985; Geeraerts, 2010), provides a number of semantically and syntactically annotated sentences from which reliable evidence can be revealed on the combinatorial possibilities of each word in each of its senses, through manual annotation of example sentences. A 'frame', in Frame Semantics, corresponds to scenario involving the interaction of its participants, so-called Frame Elements (FEs), which are fine-grained semantic role labels.

FrameNet is one of the important resources for Natural Language Processing. Successfully, it has been applied to question answering systems (Gildea and Jurafsky, 2002; Agrawal and Mukherjee, 2009; Jia and Tai, 2008), and to the research on word sense disambiguation (Carroll and McCarthy, 2000; Carroll et al, 2001; Ye and Baldwin, 2006), machine translation (Boas, 2002; Boas, 2002) and information retrieval (Mohit and Narayanan, 2003). Additionally, FrameNet information, the database reflecting the facts of the valency description as evidenced in corpus, is exactly what lexicographers need to be aware of when writing the dictionary entry (Sue and Micheal, 2008; Sue, 2008).

Given the afore-mentioned benefits, similar analyses of Japanese (Ohara et al, 2003; Ohara et al, 2004; Saito et al, 2008) and Chinese (Chen and Fung, 2004; Chen and Fung, 2010), Spanish (Subirats and Petruck, 2003; Subirats, 2009) and Italian (Lenci, Johnson and Lapesa, 2010), German (Burchardt et al, 2009) and French (Mouton, de Chalendar, Richert, 2010), Bulgarian (Koeva, 2010) and Hebrew (Petruck, 2009) are in progress, closely associated with

the Berkeley FrameNet. Nothing, nevertheless, has been done with FrameNet for the Thai language.

This research aims at manual constructing the very first Thai FrameNet (henceforth, TFN) related to the original FrameNet through a combination approach: the expand and the merge (Vossen, 1999). The former is applied for the original FrameNet alignment whilst the latter vividly reflects Thai conceptual scenarios. Subsequently, Thai frames are collected and connected by frame-to-frame relations.

2 Thai FrameNet Construction

Meticulously, the TFN associated with the archetypal FrameNet which composes of 1,010 frames is handcrafted by expand approach. Additionally, the genuine Thai frames are created from the ground up through merge approach.

2.1 Expand Approach

This approach, the Berkeley FrameNet 1.3 frames are translated (using bilingual dictionaries) into equivalent frames in the Thai language. Since the alignment of the original creation, expand approach has been applied as follows:

1. Translating Frame Names and Definitions: The original frame names together with their defintions are manually translated from English to Thai. For example, INVADING 'The Invader enters a Land in an aggressive attempt to cripple or dominate its people and its government' are translated to ' การบุกรุก' /ka:nbùkrúk/ ' ผู้บุกรุกเข้าบุกรุกเข้าไปยังพื้นที่ ด้วยความพยายามอย่างก้าวร้าวเพื่อทำลาย หรือควบคุมประชาชนและฝ่ายปกครองของพื้นที่นั้น' /phû: bùk rúk kʰ âu pai พื้นที่นั้น' /phû: bùk rúk kʰ âu pai yaŋ pʰ ɯ́:n tʰ î: dûai kʰ a:m paʔ ya: ya:m yà:ŋ kâ:u rá:u pʰ ɯ̂:a tʰ am la:i rɯ̌: kʰ û:ap kʰ u:m praʔ tɕʰ a: tɕʰ ɔ :n lɛ̀ʔ pʰ à:i pok kʰ rɔ :ŋ kʰ ɔ̌:ŋ pʰ ɯ́:n tʰ î: nán/.

2. Translating Lexical Units: Semi-automatically, Lexical Units (hereafter, LUs), words evoking frame, are translated to Thai with four translation resources: a) LEXiTRON dictionary, b) Nontri dictionary, c) HOPE dictionary and d) Thai thesaurus, namely Khlang Kam. As a result, only correct senses are retained. For example, when three LUs, 'invade' 'invasion' and 'overrun', are translated, twelve Thai words are retrieved, i.e., การบุกรุก /ka:n bùk rúk/, การรุกราน /ka:n rúk ra:n/, การโจมตี /ka:n tɕo:m ti:/, บุกรุก /bùk rúk/, /bùk rúk/, รุกราน /rúk ra:n/, บุก /bùk/, โจมตี /tɕo:m ti:/, รุกล้ำ /rúk lám/, ย่ำยี /yâm yi:/, ล่วงล้ำ /lû:aŋ lám/, lám/, แพร่หลาย /pʰ rɛ̂: lă:i/ and ส่วนเกิน /sù:an kɤ̀ :n/. However, แพร่หลาย /pʰ rɛ̂: lă:i/ and ส่วนเกิน /sù:an ส่วนเกิน /sù:an kɤ̀ :n/ which are incorrect senses are deleted from the list.

3. Extracting corpus sentences: Subsequently, sentences containing LUs are extracted from BEST corpus (Kosawat et al, 2009). For example, ' กองทัพนโปเลียนบุกออสเตรีย' /kɔ :ŋ /kɔ :ŋ tʰ áp náʔ po: li:an bùk ʔɔ́stri:a/ which means 'Napoleon's troops invaded Austria' is extracted.

4. Annotating linguistic information: Later, the target sentences are manually annotated by three linguistic information: a) semantic role (FEs), b) phrase type and c) grammatical function as shown in Table 1.

กองทัพนโปเลียน	บุก	ออสเตรีย
'Napoleon's troops	invaded	Austria'
INVADER		LAND
NP		NP
SUBJECT		COMPLEMENT

Table 1. Linguistic Information Annotation

2.2 Merge Approach

This approach, the TFN is done through local resources. Frames and their relations are first developed separately, after which the equivalence relations to the Berkeley FrameNet 1.3 are generated. Manifesting the indigenous Thai concepts, merge approach is implemented as follows:

1. Defining Frame: Frames are defined through the Thai wisdom particularly the articles on Thai culture and custom, festival and folklore. For example, LOY KRATHONG ' ลอยกระทง' /lɔ :i kra͡ tʰ oŋ /, 'River Goddess worship ceremony annually held in Thailand on a full moon night in November', is invented.

2. Finding Lexical Units: From the autochthonous Thai documents, the lists of LUs are made by linguists. For example, ' กระทง' /kraʔ tʰ oŋ / ' วันเพ็ญเดือนสิบสอง' /wan pʰ en pʰ en duɯ:an sìp sɔ̌:ŋ / ' ขึ้นสิบห้าค่ำเดือนสิบสอง' /kʰ ûn sìp hâ: kʰ âm duɯ:an sìp sɔ̌:ŋ / ' นางนพ ' นางนพมาศ' /na:ŋ nòp pʰ aʔ mâ:t/ ' เผาเทียนเล่นไฟ' /pʰă:u tʰ i:an lên fai/ ' จองเปรียง' /t͡ɕ ɔ :ŋ /t͡ɕ ɔ :ŋ pri:aŋ / ' พลุตะไลไฟพะเนียง' /pʰ lúʔ tàʔ lai fai pʰ áʔ ni:aŋ / and so on are listed. listed.

3. Extracting corpus sentences: From a large number of sentences in BEST corpus, the example sentences holding the target LUs are excerpted. For example, ' เขาเดินไปลอยกระทง กระทงที่ริมแม่น้ำเจ้าพระยา' /kʰ ău dɤ :n pai lɔ :i kra͡ tʰ oŋ tʰ î: rim mɛ̂: ná:m t͡ɕ â:u pʰ aʔ ya:/ 'He 'He floats a raft at Chaopraya River's bank' is extracted.

4. Annotating linguistic information: Afterwards, linguisitc information is manually annotated to the extracted sentences.

2.3 Caveat

Knowing how to analyse data and abandon a failed idea is an important kind of decision linguists have to make. For TFN developers, four cruxes of the matter should be concerned.

1. Defining or Translating Frame:
    a. Prototypical meaning: To define frame name, an umbrella term should be regarded. For example, ' การบุกรุก' /ka:nbùkrúk/ and ' ลอยกระทง' /lɔ :i kra͡

tʰ oŋ / are suitable cover terms for 'INVADING' and 'LOY KRATHONG' respectively because these two typical words represent such frames meaningfully.

2. Finding or Translating Lexical Units:
   a. Polysemy: Thai has many words which are polysemous, a word with multiple meanings. As mentioned earlier, frame-evoking words are listed. Only correct meanings must be retained whilst incorrect ones must be deleted. For example ' แพร่หลาย' /pʰ rɛ̂: lă:i/ and ' ส่วนเกิน' /sù:an kɤ :n/ which which are incorrect meanings are deleted from FE's list of 'INVADING'.

3. Extracting corpus sentences:
   a. Sentence Boundary: Unlike English or Japanese, Thai has obscure sentence boundaries (Aroonmanakun, 2007; Supnithi, 2010). When sentences are extracted from BEST corpus, the manual determination is required. For example, ' คืนวันเพ็ญเดือนสิบสองปีนี้ เขาขอหนีความวุ่นวายจากฝูงชน เขาเดินไปลอยกระทงที่ริมแม่น้ำ เขาเดินไปลอยกระทงที่ริมแม่น้ำเจ้าพระยามาครับ หนีคนแน่นๆ และไม่ต้องรีบลอย รีบดู รีบเดินเหมือนทุกปี' /kʰ ɯ :n wan เหมือนทุกปี' /kʰ ɯ :n wan pʰ en dɯ :an sìp sɔ̌:ŋ pi: ní: kʰă:u kʰɔ̂: nǐ: kʰwa:m wûn wa:i wa:i tɕ à:k fū :ŋ tɕʰ ɔ :n kʰă:u dɤ :n pai lɔ :i kra tʰ oŋ tʰ î: rim mɛ̂: ná:m tɕ â:u pʰ aʔ ya: ma: kʰ ráp nǐ : kʰ ɔ :n nɛ̂n nɛ̂n lɛ̀ mâi tôŋ rî:p lɔ :i rî:p du: rî:p dɤ :n mɯ̌:an tʰ úk pi:/ 'Full moon night in November this year, he prefer eluding from crowd, he floats a raft at Chaopraya River's bank. Escape from throng, there is no hurriedly floating, seeing and walking like every year.' is extracted. However, only ' เขาเดินไปลอยกระทงที่ริมแม่น้ำเจ้าพระยา' /kʰă:u dɤ :n pai lɔ :i dɤ :n pai lɔ :i kra tʰ oŋ tʰ î: rim mɛ̂: ná:m tɕ â:u pʰ aʔ ya:/ 'He floats a raft at Chaopraya River's bank' is selected.

4. Annotating linguistic information:
   a. Serial Verb: In Thai, two or more verbs or verb phrases with shared nominal arguments are put in juxtaposition without any linker. For example, ' เขาเดินไปลอยกระทง' /kʰă:u dɤ :n pai lɔ :i kra tʰ oŋ / 'He walk go float a float a raft', three verbs, ' เดินไปลอย' /dɤ :n pai lɔ :i/ 'walk go float' are juxtaposed. In this case, ' ลอย' /lɔ :i/ 'float' is considered as the target main main verb.
   b. Phrasal Verb: Inseparably, many verbs in Thai are followed by preposition. For example, ' กลมกลืนกับ' /klo:m klɯ :n kàp/ 'be in harmony with', ' เรียกร้องให้' ' เรียกร้องให้' /rî :ak rɔ̀:ŋ hâi/ 'request for', ' จัดการกับ' /tɕ àt ka:n kàp/ 'deal with'. with'. In this case, a combination of verb and preposition is considered as a unit.
   c. Compound Noun: Compounding is a process of word formation based on the combination of two or more words which appear as independent words in language. For example, ' เขาไปอาบอบนวด' /kʰă:u pai ʔà:p ʔòp nû:at/ 'He goes to goes to a massage parlour', the compond noun ' อาบอบนวด' /ʔà:p ʔòp nû:at/ 'massage parlour' is made up from three verbs, ' อาบ' /ʔà:p/ 'bath', ' อบ' /ʔòp/ 'vapour bath' and ' นวด' /nû:at/ 'massage'. This sentence may be

misinterpreted as serial verb, 'He goes to bath, to vapour bath and to massage'.

After a combination approach is implemented, frames are collected in the TFN database and connected by frame-to-frame relations.

3 Thai FrameNet Ontology and Relations

3.1 Thai FrameNet Ontology

Thai frames are organised in the TFN database. Each frame composes of five components as shown in Figure 1 and 2:

1. Frame Name: Defining conceptually scenario, frame is named.
2. Frame Definition: Clearly, to explain the meaning of frame, definition is used.
3. Frame Element: Two kinds of FE are determined: a) core element is one that instantiates conceptually necessary component of a frame whilst b) non-core element is optional component.
4. Lexical Unit: Frame-evoking words are listed.
5. Frame Example Semtence: Example sentences are annotated by three annotation layers: a) semantic role (FEs), b) phrase type and c) grammatical function.

3.2 Frame Relations

Linking frames in specific ways, seven types (with ten subtypes) of frame relations are defined as shown in Table 2.

1. Inheritance: Experienced through ontologies, the Inheritance relation corresponds to 'IS-A' relations.
2. Perspective_on: The Perspective_on relation indicates the presence of at least two different point-of-views taken on the neutral frame.
3. SubFrames: A single frame may represent a whole sequence of events, each of which represented through a single frame. The SubFrame relation is utilised to model complex scenes.
4. Precedes: Temporal precedence is encoded between the different subframes of a complex frame.
5. Using: Exclusively, the Using relation is used for cases in which a part of the scene evoked by the child refers to the parent frame.
6. Causative_of and Inchoative_of: To mark the lexical aspect of verbs, Causative_of and Inchoative_of are used.
7. See_also: To help human readers, the See_also relation represents groups of frames which are similar and should be carefully differentiated, compared and contrasted.

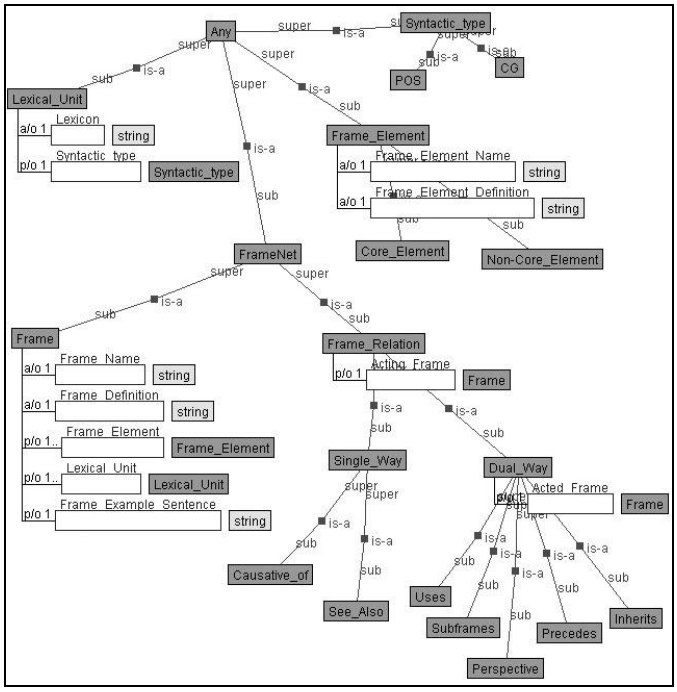

Figure 1. Thai FrameNet Ontology.

Frame Types	Frame Subtypes
Inheritances	Is inherited by
	Inherits from
Perspective_on	Perspective on
	Is Perspectivized in
SubFrames	Has SubFrame
	SubFrame of
Precedes	Precedes
	Is preceded by
Using	Uses
	Is used by
Causative_of and Inchoative_of	-
See_also	-

Table 2 Frame relations

Figure 2. Thai FrameNet Organisation.

4 Thai FrameNet Tools

4.1 Thai FrameNet Annotation Tool

The Thai FrameNet Annotation Tool is developed for linguistic annotation on the sentences which are extracted corresponding to the lexical units (LUs) in each of its frames. The frame developers who are linguists can annotate a) semantic role (FEs), b) phrase type and c) grammatical function to the target sentences. Subsequently, the result is collected into the database in XML format. The system architecture of the annotation tool consisting of three modules is shown in Figure 3.

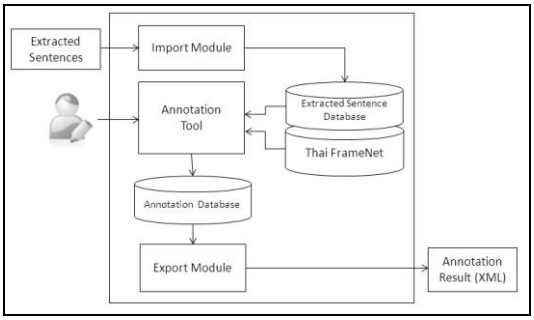

Figure 3.The annotation system overview.

1. Import Module: The Import Module allow users to import the extracted sentences to the database. Users have to specify the corresponding frame and its lexical unit for each sentence before import to the database.

2. Annotation Module: The Annotation Module allows users to annotate a) frame element (FEs), b) phrase type and c) grammatical function of word or phrase of a sentence. Only frame elements corresponding to frame of the sentence are shown for annotation. Figure 4 shows the user interface for the Annotation Module. The annotation results are stored in the Annotation Database.

3. Export Module: The Export Module allow user to export annotation results in XML format. An example of an annotation result is shown in Figure 5.

Figure 4. TFN annotation tool.

```
<extracted_sentence id=1 frame=การค้าขาย word=ขาย>
<sentence>เขา ขาย กล้วย</sentence>
<annotation_result>
      <chunk>
         <word>เขา</word>
         <semantic_role>Seller [Slr]</semantic_role>
         <phrase_type>NP</phrase_type>
         <grammartical_function>SUBT</grammartical_function>
      </chunk>
      <chunk>
         <word>กล้วย</word>
         <semantic_role>Goods [Gds]</semantic_role>
         <phrase_type>NP</phrase_type>
         <grammartical_function>OBJT</grammartical_function>
      </chunk>
</annotation_result>
</extracted_sentence>
```

Figure 5. Annotation result in XML Format.

4.2 Thai FrameNet Visualised Tool

Vividly, Thai frame-to-frame relations and details will be envisaged through a graphical visualised tool which composes of three components: (1) user connection (2) system structure and (3) Thai FrameNet database as shown in Figure 6.

Increasing efficiency, JavaScript library, namely theJIT, has been applied for graphic production. The visualised system structure consists of five modules as shown in Figure 7.

1. Frame List Module: With the height of 50 frames at maximum, frame names will be shown.

2. Frame Relation Module: In this module, Thai frame-to-frame relations conjointly whose frequencies will be displayed.

3. Visualised Frame Module: Visibly clear, the target frame will always be shown on the centre. The graphical display of frame can be moved along the user dictation.

4. Frame Connection Module: Obviously, frames related to the target will be listed; moreover, they can automatically be changed by user command on visualised frame module.

5. Frame Element Module: The list of FEs which is associated to the target frame will be shown in this module.

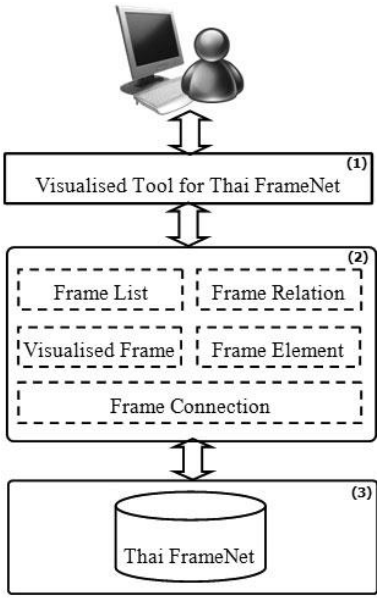

Figure 6. The architecture of TFN visualised tool.

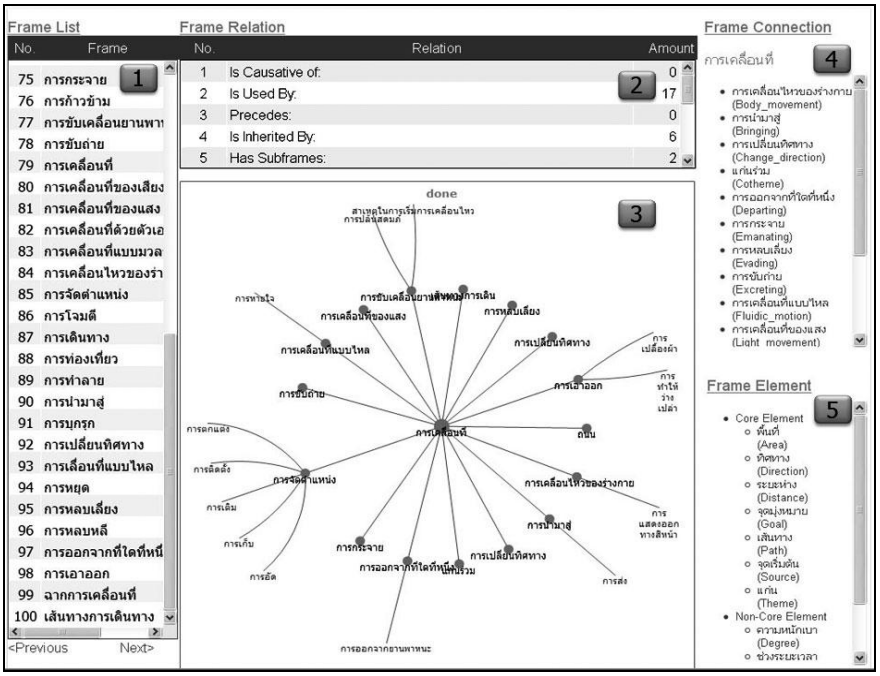

Figure 7. The visualised system structure.

4.3 Result from users

Fast access, work faster. Previously, if the frame developers who are the linguists need to investigate frame information and its related frames, the access time was approximately one minute for all related frame needs to be opened one by one. Afterwards the TFN visualised tool having been exploited, the linguists can directly access to the target and associated frames on user demand. The access time has been decreased to merely five seconds. Furthermore, all related frame shown, it is easier for frame developers to invent the genuine Thai frame and to locate in TFN data structure.

5 Conclusion and Future Works

Through a combination approach, this research is the first TFN construction. The TFN closely associated to the original FrameNet is made in way of expand approach. Equally important, merge approach reflects the genuine Thai concepts. All frames are linked by seven frame-to-frame relations.

In future, the TFN should be enlarged in particular reference to Thai wisdom. The TFN users can suggest new Thai frames or add more linguistic information through the TFN suggestion system. In addition, the TFN verification system should be developed to check the suggested frames. We believe the findings could shed the great light on Thai wisdom.

6 Acknowledgements

We owe a lot to Charles J. Fillmore, Collin F. Baker and all FrameNet team at ICSI, Berkeley, without whom no discovery on the Thai FrameNet has been made.

7 References

Agrawal, A. and Mukherjee, A. Question Answering using FrameNet. Available from:http://www.cse.iitk.ac.in/users/ashagr/webpage/courses/cs365_project_report.pdf [2009, April 19]

Aroonmanakun, W. Thoughts on Word and Sentence Segmentation in Thai. In Proceedings of the SNLP2007. Pattaya, Thailand, pp 85-90.

Baker, C. ,Fillmore, C. and Lowe, J. The Berkeley FrameNet project. In Proceedings of the COLING-ACL, Montreal, Canada, 1998.

Boas, H. Bilingual FrameNet Dictionaries for Machine Translation. In González Rodríguez, M., and C. Paz Suárez Araujo (eds.), Proceedings of the Third International Conference on Language Resources and Evaluation. Las Palmas, Spain. Vol. IV: 2002, pp 1364 – 1371.

Boas, H. Using Multi-lingual FrameNet Databases for Machine Translation. In German Research Center for Artificial Intelligence, Saarland University, Saarbrücken, Germany, June 2002.

Burchardt, A. et al. Using FrameNet for the semantic analysis of German: Annotation, representation, and automation. In Hans C. Boas (Ed), Multilingual FrameNets in Computational Lexicography: Methods and Applications. Mouton de Gruyter, Berlin, 2009.

Carroll, J. and McCarthy, D. and J. Priss. Disambiguating Noun and Verb Senses Using Automatically Acquired Selectional Preferences. In Proceeding of the SENSEVAL-2 Workshop at ACL/EACL'01. 2001, Toulose, France.

Carroll, J. and McCarthy, D. Word Sense Disambuguation using Automatically Acquired Verbal Preferences. In Computers and the Humanities, Senseval Special Issue, 1-2 (34), 2000, pp 109-114.

Chen, B. And Fung, P. Automatic Construction of an English-Chinese Bilingual FrameNet. In Proceedings of Human Language Technology conference 2004 (HLT/NAACL-04). Boston, Mass: May 2004, pp 29-32.

Chen, B. And Fung, P. BiFrameNet: Bilingual Frame Semantics Resource Construction by Cross-lingual Induction. Available from: http://acl.ldc.upenn.edu/C/C04/C04-1134.pdf [2010, May 8]

Fillmore, C. Frame semantics. In Linguistics in the Morning Calm, Seoul, Hanshin Publishing Co. 1982, 111-137.

Fillmore, C. Frames and the semantics of understanding. In Quaderni di Semantica, Vol. 6.2, 1985, pp 222-254.

Geeraerts, D. Theories of Lexical Semantics. Oxford University Press, 2010.

Gildea, D. and Jurafsky, D. Automatic Labeling of Semantic Roles. In Computational Linguistics. Vol 28.3, 2002, pp 245-288.

Jia, J. and Tai, Y. The Design and Implementation of Question Analysis for Q&A System Based on Chinese FrameNet ;New Technology of Library and Information Service; 2008-06

Koeva, S. Lexicon and Grammar in Bulgarian FrameNet. In Proceedings of the seventh international conference on Language Resources and Evaluation (LREC 2010), Malta: 2010, pp 63-67.

Kosawat, K. et al., BEST 2009: Thai Word Segmentation Software Contest, Open Conference on the SNLP2009. Bangkok, Thailand, pp. 83-89.

Lenci, A., Johnson, M. and Lapesa, G. Building an Italian FrameNet through Semi-automatic Corpus Analysis. In Proceedings of the seventh international conference on Language Resources and Evaluation (LREC 2010), pp-pp. Malta: European Language Resources Association (ELRA).

Mohit, B. and Narayanan, S. Semantic extraction with wide-coverage lexical resources. In Proceedings of the Human Language Technology Conference (HLT-NAACL) 2003, Edmonton, Canada.

Mouton, C., de Chalendar, G. and Richert, B. FrameNet translation using bilingual dictionaries with evaluation on the English-French pair. In Proceedings of the seventh international conference on Language Resources and Evaluation (LREC 2010), Malta: 2010, pp 88-91.

Ohara, K. et al. The Japanese FrameNet Project: A Preliminary Report. In Proceedings of Pacific Association for Computational Linguistics (PACLING'03), Halifax, Canada. August, 2003, pp 249-254.

Ohara, K. et al. The Japanese FrameNet Project: An introduction. LREC 2004. The Fourth international conference on Language Resources and Evaluation.In Proceedings of the Satellite Workshop "Building Lexical Resources from Semantically Annotated Corpora, pp 9-11. Lisbon, Portugal. May, 2004.

Petruck, M. Typological considerations in constructing a Hebrew FrameNet. In Hans C. Boas (Ed), Multilingual FrameNets in Computational Lexicography: Methods and Applications. Mouton de Gruyter, Berlin, 2009.

Ruppenfer, J. et al. FrameNet II Extended Theory and Practice. Available from: http://framenet.icsi.berkeley.edu/[2010, June 1]

Saito, H. et al. "The Japanese FrameNet Software Tools." LREC2008. In Proceedings of the 6th International Conference on Language Resources and Evaluation.

Subirats, C. and Petruck, M. Surprise: Spanish FrameNet! In E. Hajicova, A. Kotesovcova & Jiri Mirovsky (eds.), Proceedings of CIL 17. CD-ROM, 2003, Prague: Matfyzpress.

Subirats, C. Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon.In Hans C. Boas (Ed), Multilingual FrameNets in Computational Lexicography: Methods and Applications. Mouton de Gruyter, Berlin, 2009.

Sue, A. and Micheal, R. The Oxford Guide to Practical Lexicography. Oxford University Press, 2008.

Sue, A. Then and Now: Competence and Performance in 35 Years of Lexicography. In Fontenelle, Thierry (Ed), Practical Lexicography: A Reader. Oxford University Press, 2008.

Supnithi, T. A Supervised Learning based Chunking in Thai using Categorial Grammar. Open Conference on 8th Asian language workshop in the COLING 2010, Beijing, China, 2010.

Vossen, P. EuroWordNet General Document Version 3 Final.University of Amsterdam.EuroWordNet LE2-4003, LE4-8328. 1999.

Ye, P. and Baldwin, T. Verb Sense Disambiguation Using Selectional Preferences Extracted With a State-of-the-art Semantic Role Labeler. In Proceeding of the 2006 Australasian Language Technology Workshop (ALTW2006). Sydney, Australia.