

SEMANTIC ANALYSIS WITH A LATTICE-BASED FRAMENET

Yelda Fırat¹, Özlem Uçar², Yılmaz Kılıçaslan²

¹Çanakkale Onsekiz Mart University, Vocational School of Technical Sciences, Computer Programming, Çanakkale, Turkey

²Trakya University, Faculty of Engineering, Department of Computer Engineering, Edirne, Turkey

Abstract

From a general perspective, models of reality defined using a formal language are known as formal ontologies. These general-purpose models can also be used in specific fields of application. As they are very big by nature, ontologies are prone to errors and omissions if they are formed manually. Furthermore, it might be difficult to adapt an ontology formed in this way to a particular field of application. This paper presents a formal ontology developed and integrated into the FrameNet technology via concept lattices which in turn provide us with a semantic analysis of sentences.

Key words: Formal Ontology, Formal Concept Analysis, Semantic Structures, FrameNet

1. INTRODUCTION

Formal ontology is a reality model which is defined using a formal language. These general-purpose models are used in natural language processing and computational linguistics applications.

The analysis of language, especially semantic analysis, requires a reality model, ontology. Computational linguistics or natural language processing applications make it essential for this ontology to be formal. Because ontologies are big by nature, they are prone to errors and omissions when formed manually. Besides, adaptations will be difficult for an ontology formed this way because of its application requirements.

Background information will be presented about ontologies, thematic roles and concept lattices in the following section. Section 3 offers an analysis of thematic roles using concept lattices with a FrameNet content. The paper culminates with a summary in the concluding section.

2. ONTOLOGIES, THEMATIC ROLES AND CONCEPT LATTICES

Ontologies are an important source of information for determining the conceptual contents of sentences. A crucial component of the conceptual content of a sentence is its predicate-argument structure or, equivalently its thematic-role structure. We maintain that the interaction between ontologies and thematic roles has not been given due attention in the relevant literature.

Gruber (1993), who initiated modern studies on Thematic relations, refers to ‘a specification of a conceptualization’ as an ‘ontology. In his terms, efficient communication requires ontological agreement.

The role of ontologies in natural language analysis and processing have received increasing attention in recent years. In addition to digital dictionaries, ontologies are used as a reality model in semantic work. Kılıçaslan and Güner (2011), where a reality model is developed using WordNet and Wikipedia resources, sets an example for these kind of ontologies.

FormalConcept Analysis is a mathematical theory of conceptual hierarchies. A concept which is given in a formal context consists of features and objects having these features. In such a hierarchy all concepts are organized within a sub-concept /super-concept relationship. Such concept clusters form a concept lattice. Concept lattices, as seen in examples below, can be shown using diagrams.

2.1. Types of ontologies

There are various kinds of ontologies which are used in particular applications. We use *domain-specific* and *thematic ontologies* in this study.

2.1.1. Domain-specific ontologies

For every ontology which is developed using Formal Concept Analysis, objects and features basic constituents. In other words, objects and features are elements shared by all kinds of ontologies.

For example, in Figure 1, objects identified in the concept lattice are John (j), Fido (f), and Tweety (t), and features are animate, smart, two-legged and furry.

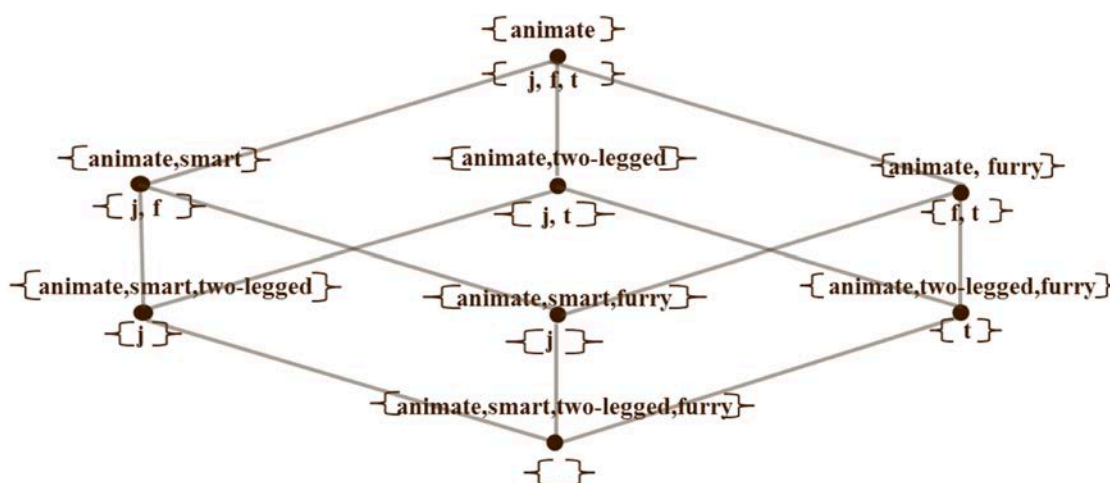


Figure 1. A domain-specific ontology

2.1.2. Thematic ontologies

Thematic roles are the grammatically relevant roles played by the participants of a situation. Jackendoff (1972, 1990) identifies two semantic tiers where thematic roles are organized: thematic tier and action tier.

The thematic tier accommodates the locative roles: Location, Source, Goal, and Path. It is by means of these roles that the location or motion of an object is specified with respect to a locative reference.

The action tier has to do with the roles of arguments that generally answer the question “Who did what to whom?” The two crucial roles of this tier are Agent and Patient.

Below are brief definitions of the thematic roles on the two tiers along with suggestive examples (where the phrase bearing the role defined is underlined):

Theme – an object in motion or in a place

- (1) The cat is in the room.

Agent – a person or a thing which starts an action, the doer of an action

- (2) John broke the glass.

Patient – a person or a thing that is affected by an action

- (3) John broke the glass.

Source – the position from which an object moves

- (4) John went from Istanbul to Ankara.

Goal – the position toward which an object moves

(4) John went from Istanbul to Ankara.

Location – a place where an object is

(5) The cat is in the room.

Path – the trajectory along which an object moves

(6) John walked from a living room to a kitchen through a hallway in a house.

We reformulate the semantic tiers identified by Jackendoff in terms of Formal Concept Analysis. Let us first consider the thematic tier. Take the sentence below:

(7) A child walked from a living room to a kitchen through a hallway in a house.

In the situation the sentence describes, The house, the living room, the kitchen and the hallway played respectively the roles Location, Source, Goal and Path in this situation. But, the living room, the hallway and the kitchen need also be considered as bearing the thematic role Location, as each was a place where the child was at a different time interval of the situation. Moreover, the hallway functioned as the goal to which the child moved at the beginning of the situation and as the source from which the child came out at the end. All these facts can be captured by means of the following concept lattice, where the house, the living room, the kitchen and the hallway are referred to as h, r, k, and w, respectively:

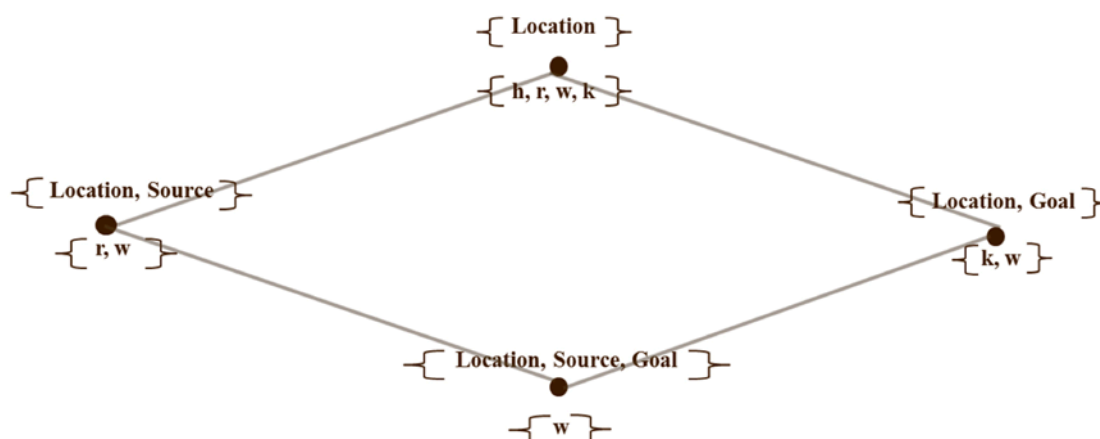


Figure 2. A sample lattice for thematic tier

In some cases, a phrase may take multiple roles, one role from the thematic tier and another from the action tier. The following sentence exemplifies such a case:

(8) A girl hits a boy with a ball.

In this example, the ball plays the theme role. While the girl and the boy play the roles of source and goal, they also assume the roles of agent and patient, respectively. The following figure shows the organization of the locative roles along with the connection points between the two tiers, where g and b are the respective symbols for 'the girl' and 'the boy':

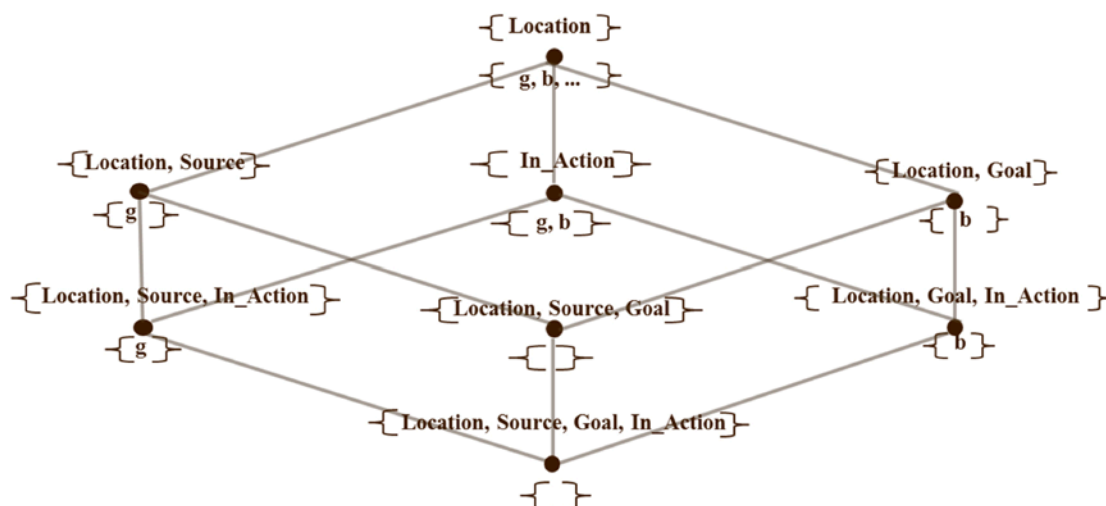


Figure 3. The thematic tier and its interaction points with the action tier

2.2. Modelling thematic roles with concept lattices

As illustrated in the examples above, each node of a concept lattice is a pair of sets, one including a set of attributes and the other a set of objects. A defining condition on these pairs of sets is that all and only the attributes in the set of attributes must be borne by all the associated objects. We will make use of such associations to relate thematic roles to their bearers: the former will be our attributes whereas the latter are the objects. Notice that concept lattices allow us to encode multiple thematic roles on a single object or vice versa.

Moreover, the theoretical framework established in this way enables us to explain certain syntactic phenomena. Consider, for example, the case of so-called double-object constructions, exemplified in (9).

- (9) a. John sent Mary the parrot.
b. John sent the parrot to Mary.

These sentences describe the same event. Yet, the perspectives these sentences present to the event are different. In (9a), Mary is necessarily conceived of as an individual affected by the action. For instance, the following sentence sounds odd, as it is difficult to conceive London as being affected by the action in question:

- (10) ?? John sent London the parrot.

Speaking in terms of thematic roles, this amounts to saying that Mary is both the goal (toward which the parrot was sent) and also the patient (affected by the sending of the parrot). On the other hand, the syntax of (9b) does not impose that kind of constraint on the object of the *to*-preposition. Hence, the following sentence is entirely felicitous:

- (11) John sent the parrot to London.

This difference in perspective can be simply reflected by placing Mary in (9a) in a set of objects bearing the roles of both goal and patient, i.e. at a node representing the intersection of the relevant nodes of the thematic and action tiers. Mary in (9b), on the other hand, is simply placed in a purely thematic node.

3. FRAMENET TECHNOLOGY IN NATURAL LANGUAGE PROCESSING APPLICATIONS

FrameNet is a lexicographic dictionary which was created on semantic frames. FrameNet, which is more comprehensive than traditional dictionaries which make the words' semantic and morphosyntactic (syntactic) information in the clauses interactive, was first created for English by Charles Fillmore and his group of Berkeley University (Basili, Cao, Croce, Coppola & Moschitti 2009).

The FrameNet dictionary has two important components: frames and lexical units that bring these frames into mind. The lexical units in a clause can be composed of verbs, nouns, adjectives, and adverbs, and the elements dependent on the lexical units are called frame elements. Word frames are formed by marking frame elements which are dependent on lexical units in a semantic and morphosyntactic way together with their thematic roles (Gildea & Jurafsky 2002, Tonelli & Pianta 2009).

The meaning of a frame is characterized with semantic and morphosyntactic properties of interrelated semantic frames. Each of these semantic frames is an element with their conceptual roles and they can be called semantic elements as well. Frame elements can be modelled with concept lattices mentioned above.

It can also be stated that FrameNet is an ontology (Scheffczyk, Pease & Ellsworth 2006).

Dissecting the given clauses and marking each segment's morphosyntactic and semantic properties, creating FrameNet database, and stating the semantic roles of the clause segments can be listed within the scope of FrameNet.

In English FrameNet it is possible to find the correct frames of the clauses by enlarging the coverage area of the frames with the relations among the frames. The types of relationships used are the followings: inheritance, sub-frame, using. Below are short descriptions of these relationships

Inheritance: A frame possesses all the properties of another frame. This relationship is a IS-A relationship (Narayana, Fillmore, Baker & Petruck 2002).

Sub-Frame: It is a case in which a frame is a part of another frame. The sub-frame relationship where each part describes itself as a different frame is used to connect existing situations (events) and transitions with each other. (Ruppenhofer, Sunde & Pinkal 2010).

Using: The using relationship can be described as a frame using some lexical units belonging to another frame when needed.

Figure 4 shows an example achieved with the body_movement frame.

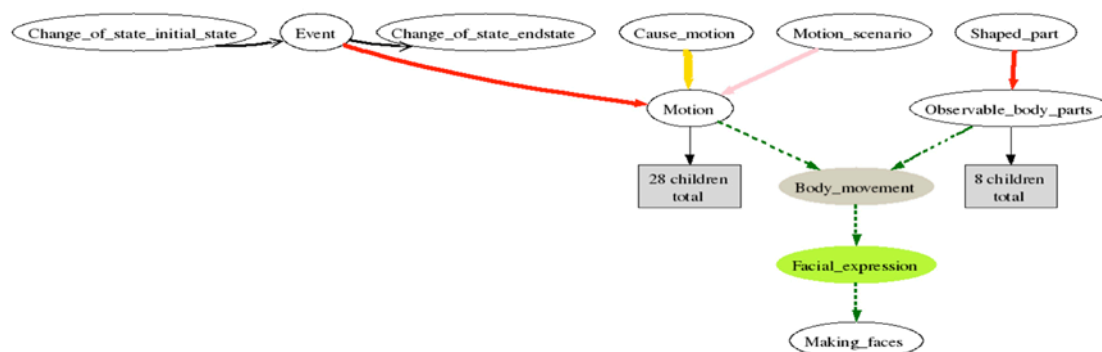


Figure 4. The relationship diagram of the body_movement frame

Below are the meanings of the relationship illustrations:

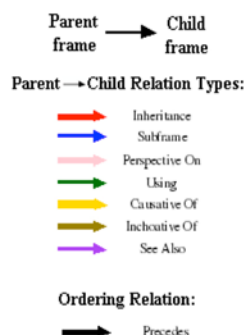


Figure 5. The meanings of the relationship illustrations

In a FrameNet context, the traditionally known thematic roles of Agent and Patient and the semantic roles of frame elements can be easily reached. Let us consider two clauses taken from the English FrameNet again in terms of semantic component structure.

(12) [Chuck_{BUYER}] **bought** [a car_{GOODS}] [from Jerry_{SELLER}] [for \$1000_{PAYMENT}]

(13) [Jerry_{SELLER}] **sold** [a car_{GOODS}] [to Chuck_{BUYER}] [for \$1000_{PAYMENT}]

These clauses describe the same event. In other words, a person transfers the money to another person for shopping a set of things. However, their perspectives prove different from each other. The verbs *sell* and *buy* which give rise to this difference are treated as lexical units within the COMMERCE frame in the English FrameNet. Depending on these actions, differences between the meanings of the frame elements in these two clauses arise. The correct form of the semantic component structures of clauses are achieved through the expression of these differences with the thematic roles assigned to the the frame elements (Fillmore, Johnson & Petruck 2003).

In short, a thematic role can be described as the semantic relationship between an argument and the verb of a clause.

4. CONCLUSION

The semantic analysis of the predicate-argument structure of a clause be done using ontologies organized as concept lattices. With the integration of these lattices into a FrameNet, the argument structure of a clause can be clearly shown.

Concept lattices can serve as formal ontologies for both domain-specific and general-purpose applications. In this paper, we have discussed how to organize thematic roles as concept lattices and how to integrate the latter to a FrameNet. Admittedly, the discussion and presentation is far from being sufficiently detailed. The paper should be considered as an overview of an ongoing work.

REFERENCES

- Basili, R, Cao, D D, Croce, D, Coppola, B & Moschitti, A 2009, 'Cross-Language Frame Semantics Transfer in Bilingual Corpora', *In Computational Linguistics and Intelligent Text Processing*, pp. 332-345.
- Fillmore, C J, Johnson, C R, & Petruck, M R L 2003, 'Background To Framenet', *Intertational Journal of Lexicography*, vol. 16, no. 3, pp. 235-250.

- Gildea, D & Jurafsky, D 2002, 'Automatic Labeling of Semantic Roles', In: *Computational Linguistics*, vol. 28, no. 3, pp. 245-288.
- Gruber, T R 1993, 'A Translation Approach to Portable Ontology Specifications', *Knowledge Acquisition*, 5(2), pp. 199-220.
- Jackendoff, R 1972, *Semantic Interpretation in Generative Grammar*, MIT Press, Cambridge.
- Jackendoff, R 1990, *Semantic Structures*, The MIT Press.
- Kılıçaslan, Y & Güner, ES 2011, 'Filtering Machine Translation Results with Automatically Constructed Concept Lattices', *Proceedings of 8th International Conference on Concept Lattices and Their Applications (CLA 2011)*, Nancy, France, pp. 59-73.
- Narayana, S, Fillmore, C J, Baker, C F, & Petruck, M R L 2002, 'Framenet Meets the Semantic Web: A DAML + OIL Frame Representation', In *AAAI Workshop Proceedings*.
- Ruppenhofer, J, Sunde, J, & Pinkal, M 2010, 'Generating Framenets of various granularities: The Framenet Transformer', *The seventh International Conference on Computational Linguistics and Evaluation (LREC)*, Valletta, Malta, May 19-21.
- Scheffczyk, J, Pease, A, & Ellsworth, M 2006, 'Linking FrameNet to the Suggested Upper Merged Ontology', *Formal Ontology in Information Systems-Proceedings of the Fourth International Conference*, vol. 150, pp. 289-300.
- Tonelli, S, & Pianta, E 2009, 'Three issues in cross-language frame information transfer', *International Conference RANLP*, Borovets, Bulgaria, pp. 441-448.