Visual semantics and ontology of eventive verbs

Visual Semantics and Ontology of Eventive Verbs

Minhua Ma and Paul Mc Kevitt

School of Computing & Intelligent Systems, Faculty of Engineering
University of Ulster, Derry/Londonderry, BT48 7JL
{m.ma, p.mckevitt}@ulster.ac.uk

Abstract. Various English verb classifications have been analyzed in terms of their syntactic and semantic properties, and conceptual components, such as syntactic valency, lexical semantics, and semantic/syntactic correlations. Here the visual semantics of verbs, particularly their *visual roles*, somatotopic effectors, and level-of-detail, is studied. We introduce the notion of *visual valency* and use it as a primary criterion to recategorize eventive verbs for language visualization (animation) in our intelligent multimodal storytelling system, CONFUCIUS. The visual valency approach is a framework for modelling deeper semantics of verbs. In our ontological system we consider both language and visual modalities since CONFUCIUS is a multimodal system.

1 Introduction

A taxonomic classification of the verb lexicon began with syntax studies such as Syntactic Valency Theory and subcategorization expressed through grammatical codes in the Longman Dictionary of Contemporary English [13]. The classification ground has recently shifted to semantics: lexical semantics [6], conceptual components [9], semantic/syntactic correlations [12], and intrinsic causation-change structures [1]. Here we introduce visual criteria to identify verb classes with visual/semantic correlations.

First, in section 2 the intelligent multimodal storytelling system CONFUCIUS is introduced and its architecture is described. Next, in section 3 we review previous work on ontological categorization of English verbs. Then we introduce the notion of *visual valency* and expound CONFUCIUS' verb taxonomy, which is based on several criteria for visual semantics: number and roles of visual valency, somatotopic effectors, and level-of-detail, in section 4. Finally, section 5 summarizes the work with a discussion of possible future work on evaluation of the classification through language animation, and draws comparisons to related research.

2 Background: CONFUCIUS

We are developing an intelligent multimedia storytelling interpretation and presentation system called CONFUCIUS. It automatically generates 3D animation and speech from natural language input as shown in Figure 1. A prefabricated objects

knowledge base on the left hand side includes the graphics library such as characters, props, and animations for basic activities, which is used in *animation generation*. The input stories are parsed by the *surface transformer, media allocatior* and *Natural Language Processing* (NLP) modules. The natural language processing component uses the Connexor Functional Dependency Grammar parser [10], WordNet [6] and LCS (Lexical Conceptual Structure) database [4]. The current prototype visualizes single sentences which contain action verbs with visual valency of up to three, e.g. *John gave Nancy a book*, *John left the restaurant*.

The outputs of *animation generation*, *Text to Speech* (TTS) and *sound effects* combine at *synchronizing & fusion*, generating a 3D world in VRML. CONFUCIUS employs temporal media such as 3D animation and speech to present stories. Establishing correspondence between language and animation, i.e. language visualization, is the focus of this research. This requires adequate representation and reasoning about the dynamic aspects of the story world, especially about eventive verbs. During the development of animation generation from natural language input in CONFUCIUS, we find that the task of visualizing natural language can shed light on taxonomic classification of the verb lexicon.

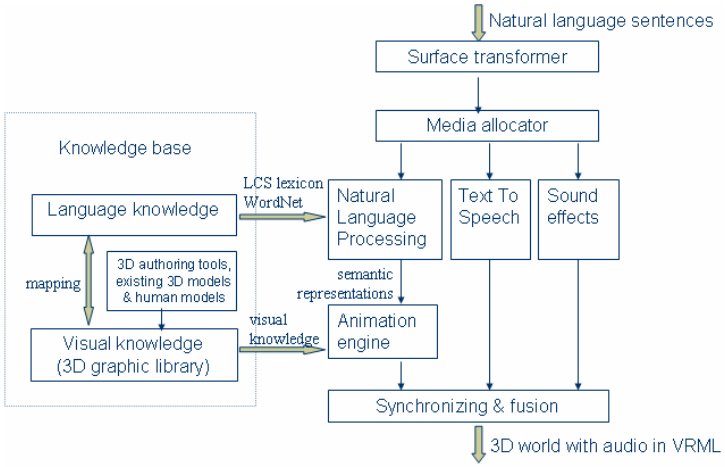


Fig. 1. Architecture of CONFUCIUS

3 Ontological Categories of Verbs

3.1 Syntactic Perspective: Valency and Aspectual Classes

In 1980s, the Longman Dictionary of Contemporary English (LDOCE) was the most comprehensive computational lexicon with a description of grammatical properties of words. It had a very detailed word-class categorization scheme, particularly for verbs. In addition to part-of-speech information LDOCE specifies a subcategorization description in terms of types and numbers of complements for each entry. In LDOCE grammar codes separate verbs into the categories: D (ditransitive), I (intransitive), L

(linking verb with complement), T1 (transitive verb with an NP object), T3 (transitive verb with an infinitival clause as object), etc. These grammar codes implicitly express verb subcategorization information including specifications on the syntactic realization of verb complements and argument functional roles.

The notion of valency is borrowed from chemistry to describe a verb's property of requiring certain arguments in a sentence. Valency fillers can be both obligatory (*complements*) and optional (*adjuncts*): the former are central participants in the process denoted by the verb, the latter express the associated temporal, locational, and other circumstances. Verbs can be divided into classes based on their valency.

There are different opinions on the type of a verb's valency fillers. Leech [11] raises the idea of *semantic valency* to operate on a level different from surface syntax. Semantic valency further developed to the theory of thematic roles in terms of which semantic role each complement in a verb's argument structure plays, ranging from Fillmore's [7] case grammar to Jackendoff's [9] Lexical Conceptual Structure (LCS). The term *thematic role* covers a layer in linguistic analysis, which has been known by many other names: theta-role, case role, deep grammatical function, transitivity role, and valency role. The idea is to extend syntactic analysis beyond surface case (nominative, accusative) and surface function (subject, object) into the semantic domain in order to capture the roles of participants. The classic roles are *agent*, *patient* (*theme*), *instrument*, and a set of locational and temporal roles like *source*, *goal* and *place*.

Having a set of thematic roles for each verb type, Dixon [3] classifies verbs into 50 verb types, each of which has one to five thematic roles that are distinct to that verb type. Systemic Functional Grammar [8] works with 14 thematic roles divided over 5 *process types* (verb types). Some linguists work out a minimal thematic role system of three highly abstract roles (for valency-governed arguments) on the grounds that the valency of verbs never exceeds 3. Dowty [5] assumes that there are only two *thematic proto-roles* for verbal predicates: the *proto-agent* and *proto-patient*. Proto-roles are conceived of as *cluster-concepts* which are determined for each choice of predicate with respect to a given set of semantic properties. Proto-agent involves properties of volition, sentience/perception, causes event, and movement; proto-patient involves change of state, incremental theme, and causally affected by event.

The ontological categories proposed by Vendler [14] are dependent on aspectual classes. Vendler's verb classes (activities, statives, achievements, and accomplishments) emerge from an attempt to characterize a number of patterns in aspectual data. Formal ontologies such as DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), SUMO (Suggested Upper Merged Ontology) and CYC all assume the traditional aspectual (temporal) classification for their events (processes).

3.2 Semantic Perspective: WordNet and Dimension of Causation

The verb hierarchical tree in WordNet [6] represents another taxonomic approach based on pure lexical semantics. It reveals the semantic organization of the lexicon in terms of lexical and semantic relations. Table 1 lists the lexicographer files of verbs in WordNet 2.0, which shows the top nodes of the verb trees.

Asher and Lascarides [1] put forward another lexical classification based on the dimension of causal structure. They assume that both causation and change can be

specified along the following four dimensions so as to yield a thematic hierarchy such as the one described in the lattice structure in Figure 2.

Table 1. WordNet verb files

Lexicographer file	Contents
verb.body	grooming, dressing, bodily care
verb.change	size, temperature change, intensifying
verb.cognition	thinking, judging, analyzing, doubting
verb.communication	telling, asking, ordering, singing
verb.competition	fighting, athletic activities
verb.consumption	eating and drinking
verb.contact	touching, hitting, tying, digging
verb.creation	sewing, baking, painting, performing
verb.emotion	feeling
verb.motion	walking, flying, swimming
verb.perception	seeing, hearing, feeling
verb.possession	buying, selling, owning
verb.social	political/social activities & events
verb.stative	being, having, spatial relations
verb.weather	raining, snowing, thawing, thundering

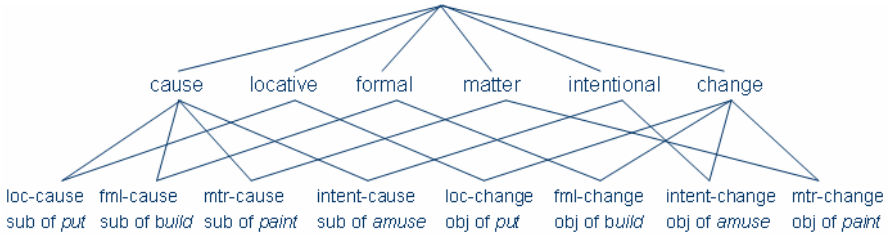


Fig. 2. Dimension of causation-change

- *locative*: specifying the causation of motion, e.g. put
- *formal*: specifying the creation and destruction of objects, e.g. build
- *matter*: specifying the causation of changes in shape, size, matter and colour of an object, e.g. paint
- *intentional*: specifying causation and change of the propositional attitudes of individuals, e.g. amuse, persuade

3.3 Semantic-Syntactic Correlations: Levin's Verb Classes

Besides purely syntactic and purely semantic methodologies, parallel syntactic-semantic patterns in the English verb lexicon have been explored as well since it is discovered that words with similar meaning, i.e. whose LCSs [9] are identical in terms of specific meaning components, show some tendency toward displaying the same syntactic behavior. Levin's [12] verb classes represent the most comprehensive description in this area. She examines a large number of verbs, classifies them accord-

ing to their semantic/syntactic correlations, and shows how syntactic patterns systematically accompany the semantic classification.

4 Visual Semantics and Verb Classes

In order to identify the full set of meaning components that figure in the visual representation of verb meaning, the investigation of semantically relevant visual properties and ensuing clustering of verbs into classes needs to be carried out over a large number of verbs. Here we identify three visual factors concerning verb categorization: (1) *visual valency*, (2) somatotopic effectors involved in action execution (visualization) and perception, and (3) level-of-detail of visual infomation. Eventive verbs are categorized according to involved somatotopic effectors, visual semantic roles (e.g. obligatory argument number and classes, humanoid vs. non-humanoid roles), and the level-of-detail they indicate.

Verbs belonging to the same class in our classification are visual "synomyms", i.e. they should be substitutable in the same set of animation keyframes, through not necessarily in exactly the same visualization. Visualization of action verbs could be an effective evaluation of the taxonomy.

4.1 Visual Valency

Visual valency refers to the capacity of a verb to take a specific number and type of *visual arguments* in language visualization (3D animation). We call a valency filler a *visual role*. We distinguish two types of visual roles: human (biped articulated animate entity) and object (inanimate entity), since they require different process in animation generation. Visual valency sometimes overlaps with syntactic and semantic valency. The difference shown in 1-3 is the number of obligatory roles. It is obvious that visual modalities require more obligatory roles than surface grammar or semantics. What is optional in syntax and semantics is obligatory for visual valency.

1) *Neo* pushed *the button*.
   syntactic valency 2, subject and object
   semantic valency 2, agent and theme
   visual valency 2, human and object
2) *Michelle* cut *the cloth (with scissors)*.
   syntactic valency 2, subject, object, optional PP adjunct
   semantic valency 2, agent, theme, optional instrument
   visual valency 3, 1 human and 2 objects, all obligatory
3) *Neo* is reading.
   syntactic valency 1, subject
   semantic valency 1, agent (and optional source)
   visual valency 2, 1 human and 1 object, all obligatory

Therefore, three visual valency verbs subsume both syntactic trivalency verbs such as *give* and syntactic bivalency verbs such as put (with goal), cut (with instrument), butter (with theme, in *butter toast*) and, an intransitive verb may turn up three visual

valency, e.g. dig in *he is digging in his garden* involves one human role and two object roles (the instrument and the place).

We classify visual roles into atomic entities and non-atomic entities based on their decomposablility, and further subclassify non-atomic roles into human roles and object roles.

4.2 Somatotopic Factors in Visualization

The second visual factor we consider in our verb taxonomy is somatotopic effectors. Psychology experiments prove that the execution, perception and visualization of action verbs produced by different somatotopic effectors activate distinct parts of the cortex. Moveover, actions that share an effector are in general similar to each other in dimensions other than the identity of the effector. Recent studies [2] investigate how action verbs are processed by language users in visualization and perception, and prove that processing visual and linguistic inputs (i.e. action verbs) associated with particular body parts results in the activation of areas of the cortex involved in performing actions associated with those same effectors.

On these theoretical grounds, we take effectors into account. However, we only distinguish facial expression (including lip movement) and body posture (arm/leg/torso) in our ontological system (Figure 3). Further divisions like distinction between upper/lower arm, hands, and even fingers are possible, but we do not make our taxonomy too fine-grained and reflect every fine visual distinction. Here is an example of using somatatopic effectors to classify action verbs *run*, *bow, kick*, *wave*, *sing*, *put*:

- Facial expression – *sing, laugh*
- Body posture
  - Leg – *run, kick*
  - Arm – *wave, put*
  - Torso – *bow*

4.3 CONFUCIUS' Verb Taxonomy

The verb categories of CONFUCIUS shown in Figure 3 represent a very minimal and shallow classification based on visual semantics. Here we focus on action verbs. Action verbs are a major part of events involving humanoid performers (agent/experiencer) in animation. They can be classified into five categories: (1) one visual valency verbs with a human role, concerning movement or partial movement of the human role, (2) two visual valency verbs (at least one human role), (3) visual valency ≥ 3 (at least one human role), (4) verbs without distinct visualization when out of context such as trying and helping verbs, (5) high level behaviours or routine events, most of which are political and social activities/events consisting of a sequence of basic actions.

We further categorize the class of one visual valency verbs (2.2.1.1) into 'body posture or movement' (2.2.1.1.1) and 'facial expressions and lip movement' (2.2.1.1.2) according to somatotopic effectors. The animation of class 2.2.1.1.1 usually involves biped kinematics, e.g. walk, jump, swim, and class 2.2.1.1.2 subsumes communication verbs and emotion verbs, and involves multimodal presentation.

These verbs require both visual presentation such as lip movement (e.g. *speak*, *sing*), facial expressions (e.g. *laugh*, *weep*) and audio presentation such as speech or other communicable sounds.

1. On atomic entities
  1.1. Movement/rotation: change physical location (position or orientation), e.g. bounce, turn
  1.2. Change intrinsic attributes such as shape, size, color, texture, and even visibility, e.g. bend, taper, (dis)appear
  1.3. Visually unobserved change: temperature change, intensifying
2. On non-atomic entities
  2.1. No human role involved
    2.1.1. Two or more individual objects fuse together, e.g. melt (in)
    2.1.2. One objects divides into two or more individual parts
      e.g. break (into pieces), (a piece of paper is) torn (up)
    2.1.3. Change sub-components (their position, size, color, shape etc), e.g. blossom
    2.1.4. Environment events (weather verbs), e.g. snow, rain, thunder, getting dark
  2.2. Human role involved
    2.2.1. Action verbs
      2.2.1.1. One visual valency (the role is a human, (partial) movement)
        2.2.1.1.1. Biped kinematics, e.g. go, walk, jump, swim, climb
          2.2.1.1.1.1. Arm actions, e.g. wave, scratch
          2.2.1.1.1.2. Leg actions, e.g. go, walk, jump
          2.2.1.1.1.3. Torso actions, e.g. bow
          2.2.1.1.1.4. Combined actions
        2.2.1.1.2. Facial expressions and lip movement, e.g. laugh, fear, say, sing, order
      2.2.1.2. Two visual valency (at least one role is human)
        2.2.1.2.1. One human and one object (vt or vi+instrument/source/goal), e.g. trolley (lexicalized instrument)
          2.2.1.2.1.1. Arm actions, e.g. throw, push, open, eat
          2.2.1.2.1.2. Leg actions, e.g. kick
          2.2.1.2.1.3. Torso actions
          2.2.1.2.1.4. Combined actions, e.g. escape (with source), glide (with location)
        2.2.1.2.2. Two humans, e.g. fight, chase, guide
      2.2.1.3. Visual valency ≥ 3 (at least one role is human)
        2.2.1.3.1. Two humans and one object (inc. ditransitive verbs), e.g. give, buy, sell, show
        2.2.1.3.2. One human and 2+ objects (vt + object + implicit instrument/goal/ theme), e.g. cut, write, butter, pocket, dig, cook
      2.2.1.4. Verbs without distinct visualization when out of context
        2.2.1.4.1. trying verbs: try, attempt, succeed, manage
        2.2.1.4.2. helping verbs: help, assist
        2.2.1.4.3. letting verbs: allow, let, permit
        2.2.1.4.4. create/destroy verbs: build, create, assemble, construct, break, destroy
        2.2.1.4.5. verbs whose visualization depends on their objects, e.g. play (harmonica/football), make (the bed/troubles/a phone call), fix (a drink/a lock)
      2.2.1.5. High level behaviours (routine events), political and social activities/events, e.g. interview, eat out (go to restaurant), call (make a telephone call), go shopping
    2.2.2. Non-action verbs
      2.2.2.1. stative verbs (change of state), e.g. die, sleep, wake, become, stand, sit
      2.2.2.2. emotion verbs, e.g. like, disgust, feel
      2.2.2.3. possession verbs, e.g. have, belong
      2.2.2.4. cognition, e.g. decide, believe, doubt, think, remember
      2.2.2.5. perception, e.g. watch, hear, see, feel

Fig. 3. Ontology of events on visual semantics

There are two subcategories under the two visual valency verbs (2.2.1.2) based on which type of roles they require. Class 2.2.1.2.1 requires one human role and one ob-

ject role. Most transitive verbs (e.g. *throw*, *eat*) and intransitive verbs with an implicit instrument or locational adjunct (e.g. *sit* on a chair, *trolley*) belong to this class. Verbs in class 2.2.1.2.2, such as *fight* and *chase*, have two human roles.

Class 2.2.1.3 includes verbs with three (or more than three) visual roles, at least one of which is a human role. The subclass 2.2.1.3.1 has two human roles and one (or more) object role. It subsumes ditransitive verbs like *give* and transitive verbs with an implicit instrument/goal/theme (e.g. *kill*, *bat*). The subclass 2.2.1.3.2 has one human role and two (or more) object roles. It usually includes transitive verbs with an inanimate object and an implicit instrument/goal/theme, e.g. *cut*, *write*, *butter*, *pocket*. The visual valency of verbs conflating with the instrument/goal/theme of the actions, such as *cut, write, butter, pocket, dig, trolley*, have one more valency than their syntactic valency. For instance, the transitive verb *write* (in *writing a letter*) is a two syntactic valency verb, but its visualization involves three roles, *writer*, *letter*, and an implicit instrument *pen*, therefore it is a three visual valency verb.

There is a correlation between the visual criteria and lexical semantics of verbs. For instance, consider the intransitive verb *bounce* in the following sentences. It is a one visual valency verb in both 4 and 5 since the PPs following it are optional. The visual role in 4 is an *object*, whereas in 5 it is a *human* role. This difference coincides with their word sense difference (in WordNet).

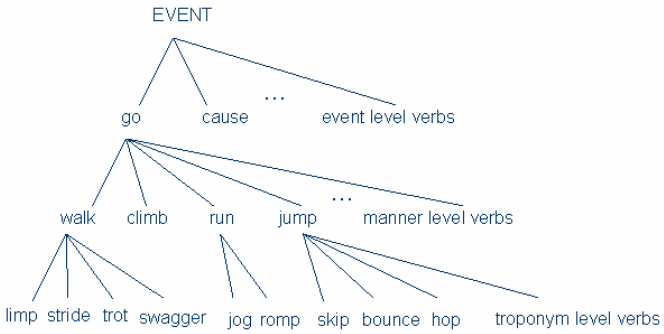
4) The ball *bounced* over the fence.
   WordNet sense: 01837803. Hypernyms: jump, leap, bound, spring
   CONFUCIUS verb class 1.1
5) The child *bounced* into the room.
   WordNet sense: 01838289. Hypernyms: travel, go, move
   CONFUCIUS verb class 2.2.1.1.1

4.4 Level-Of-Detail (LOD) -- Basic-Level Verbs and Their Troponyms

The classes from 2.2.1.1.1.1 through 2.2.1.1.1.4 are the most fine-grained categories in Figure 3. They can be further classified based on *Level-of-Detail* (LOD). The term LOD has been widely used in relation to research on levels of detail in 3D geometric models. It means that one may switch between animation levels of varying computation complexity according to some set of predefined rules (e.g. viewer perception).

Let's have a look at the *verbs of motion* in Levin's [12] classes. They subsume two subclasses: *verbs of inherently directed motion* (e.g. *arrive*, *come*, *go*) and *verbs of manner of motion* (e.g. *walk*, *jump*, *run*, *trot*). We find that there are actually three subclasses in *verbs of motion*, representing three LODs of visual information as shown in the tree in Figure 4. We call the high level *event level*, the middle level *manner level*, and the low level *troponym level*. The event level includes basic event predicates such as *go* (or *move*), which are *basic-level verbs* for atomic objects. The manner-of-motion level stores the visual information of the manner according to the verb's visual role (either a human or a non-atomic object) in the animation library. Verbs on this level are basic-level verbs for human and non-atomic objects. The troponym level verbs can never be basic-level verbs because they always elaborate the manner of a base verb. Visualization of the troponym level is achieved by modifying

animation information (speed, the agent's state, duration of the activity, iteration) of manner level verbs.



Levels	Verbs	Basic-level verb for ...
event level	go, move	atomic object
manner level	walk, jump	human/non-atomic object
troponym level	limp, stride	human

Fig. 4. Hierarchical tree of verbs of motion

In the following examples, 6a is a LCS-like representation of *John went to the station*. The predicate *go* is on the event level. The means of going, e.g. by car or on foot, is not specified. Since the first argument of *go* is a HUMAN, we cannot just move John from one spot to another without any limb movement, the predicate *go* is not enough for visualization a human role. We need a lexical rule to change the high-level verb to a basic-level verb, i.e. change *go* to *walk*, when its visual role is human (6b), because walking is the default manner of movement for human beings. In 7a the predicate *run* is enough for visualizing the action since it is a basic-level verb for human.

6) John *went* to the station.
   a) [EVENT go ([HUMAN john],[PATH to [OBJ station]])]
   b) [EVENT walk ([HUMAN john],[PATH to [OBJ station]])]
7) John *ran* to the station.
   a) [EVENT run ([HUMAN john],[PATH to [OBJ station]])]

This approach is involved with the visualization processes. The manner-of-motion verbs are stored as key frames of involved joint rotations of human bodies in the animation library, without any displacement of the whole body. Therefore *run* is just *running in place*. The first phase of visualization is finding the action in animation files and instantiating it on the first argument (i.e. the human role) in the LCS-like representation. This phase corresponds to the manner level (run) in the above tree. The next phase is to add position movement of the whole body according to the second argument (PATH). It makes the agent move forward and hence generates a *real* run. This phase corresponds to the event level (go) in the tree.

The structure in Figure 4 is applicable to most troponyms, *cook* and *fry/broil/braise/micro-wave/grill*, for example, express different manners and instruments of cooking.

5 Conclusion

In many ways the work presented in this paper is related to that of Levin [12]. However, our point of departure and the underlying methodology are different. We categorize verbs from the visual semantic perspective since language visualization in CONFUCIUS provides independent criteria for identifying classes of verbs sharing certain aspects of meaning, i.e. semantic/visual correlations. A visual semantic analysis of eventive verbs has revealed some striking influences in a taxonomic verb tree. Various criteria ranging from visual valency, somatotopic effector, to LOD are proposed for classifying verbs from the language visualization perspective. Future research should address evaluation issues using automatic animation generation and psychological experiments.

References

1. Asher, N., Lascarides, A.: Lexical Disambiguation in a Discourse Context. *Journal of Semantics*, 12(1): 69-108, (1995)
2. Bergen, B., Narayan, S., Feldman, J.: Embodied verbal semantics: evidence from an image-verb matching task. *Proceedings of CogSci*, Boston ParkPlaza Hotel, Boston (2003)
3. Dixon, R.M.W.: *A new approach to English Grammar on semantic principles*. Oxford: OUP (1991)
4. Dorr, B. J., Jones, D.: Acquisition of semantic lexicons: using word sense disambiguation to improve precision. Evelyne Viegas (Ed.), *Breadth and Depth of Semantic Lexicons*, Norwell, MA: Kluwer Academic Publishers (1999) 79-98
5. Dowty, D. R.: Thematic proto-roles and argument selection. *Language*, 67(3): 547-619, (1991)
6. Fellbaum, C.: A semantic network of English verbs. *WordNet: An Electronic Lexical Database*, C. Fellbaum (Ed.), Cambridge, MA: MIT Press (1998) 69-104
7. Fillmore, C. J.: The case for case. *Universals in Linguistic Theory*, E. Bach and R. Harms (Eds.), New York: Holt, Rinehart and Winston (1968) 10-88
8. Halliday, M.A.K.: *An Introduction to Functional Grammar*. London: Edward Arnold (1985)
9. Jackendoff, R.: *Semantic Structures*. Current studies in linguistics series. Cambridge, MA: MIT Press (1990)
10. Järvinen, T., Tapanainen, P.: A Dependency Parser for English. Technical Report, No.TR-1, Department of General Linguistics, University of Helsinki (1997)
11. Leech, G.: *Semantics.* Cambridge University Press (1981)
12. Levin, B.: *English verb classes and alternations: a preliminary investigation*. Chicago: The University of Chicago Press (1993)
13. Procter, P.: *Longman dictionary of contemporary English*. London: Longman (1987)
14. Vendler, Z.: *Linguistics and Philosophy*. Ithaca, NY: Cornell University Press (1967)