

국가 IT 온톨로지 구축을 위한 상위 온톨로지 설계에 관한 기초연구

류병래
충남대학교 언어학과
ryu@cnu.ac.kr
<http://ling.cnu.ac.kr/>

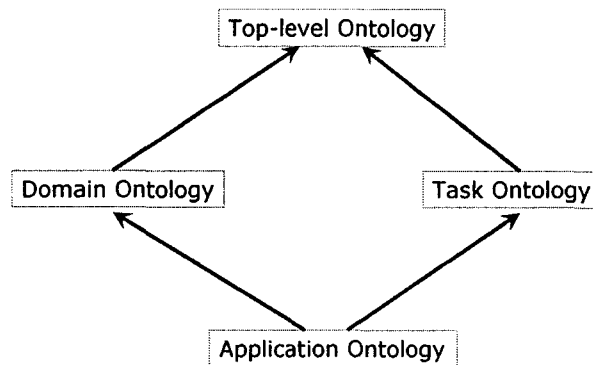
초록

본 연구는 『국가 IT 온톨로지 인프라 기술개발』 연구에서¹⁾ 진행하고 있는 정보기술 (Information Technology, 이하 IT) 온톨로지 구축을 위한 상위 온톨로지 (Upper Ontology)의 설계에 관한 기초연구이다. 본 연구의 목적은 구축하려는 IT 온톨로지의 상위층위 클래스를 어떻게 구성해야 하는가에 대한 답을 구하는 것이다.

온톨로지(ontology)란 그리스어 'ontos' (being)과 'logos' (words)에 그 어원을 두고 있다. 이는 원래 철학 (philosophy), 더 구체적으로는 형이상학(metaphysics)의 한 분야로서, 존재 (being, existence)의 본질에 관한 학문이다. 존재의 본질을 밝히기 위해서는 세계를 구성하고 있는 각 개체의 정의와 범주화 그리고 분류, 그들 사이에 성립하는 관계 등에 대한 이해가 명확해져야 한다. 최근 온톨로지는 철학 뿐만 아니라 언어학, 문헌정보학, 인공지능, 자연언어처리, 최근에는 시맨틱웹에 이르기까지 다양한 분야에서 활발한 논의가 이루어지고 있는 주제이다.

언어학에서는 인지의미론이나 어휘의미론, 범주론, 형식의미론 등의 분야에서 온톨로지에 대한 논의가 이루어지고 있으나, 개념적 공감대가 아직은 명확하게 서 있지 않은 상태라고 평가된다. 그러나 일반적으로 '공유된 개념화에 대한 형식적이고도 명시적인 기술 (A formal, explicit specification of a shared conceptualization)로 이해가 되고 (Gruber (1993)), 그 구체적인 모습은 어느 특정 영역을 모형화하는 데에 사용될 수 있는 공유 어휘 (A shared vocabulary, which can be used to model a domain)로 나타난다. 이 공유 어휘는 개체들의 유형과 개념, 속성, 관계 등을 가리킨다.

1) KAIST, POSTECH, 숭실대, 영남대, 경북대, 부산대, 울산대, 충남대, 한국외대 등 국내 대학교와 ETRI 등 연구소, 그리고 기업체인 (주) Saltlux와 (주) Korea WiseNut 등 국내 기관이 영국의 University College of Dublin과 독일의 University of Karlsruhe와 공동으로 국제공동연구 컨소시엄을 구성하여 2006년부터 2010년까지 5년간 수행하는 연구이다.



[그림 1] 특정목적에 의존적인 정도에 따른 온톨로지의 분류 (Guarino (1994:9))

Guarino (1994:9)에 따르면, 상위 온톨로지 (Top-level Ontology, 혹은 Upper Ontology)는 매우 일반적인 개념을 기술하게 되고, 이미 구축되어 있는 온톨로지의 최상위 개념이 연결될 수 있는 일반개념을 제공해야 한다. 상위 온톨로지의 구축에서 가장 문제가 되고 있는 점은 이미 여러 개의 상위 온톨로지들이 존재하고 있으니 각 상위 온톨로지에서도 가장 일반적인 개념들을 분류하는 기준이 상이하다는 것이다.

영역 온톨로지 (Domain Ontology)는 이미 용어가 암시하고 있는 바와 같이 의학, 약학, 공학, 법학, 자동차 등 특정 영역에 한정된 특화된 온톨로지를 말한다. 영역 온톨로지는 영역 내의 개념에 대한 어휘들과 그들 사이의 관계, 그리고 그 영역에서 일어나는 행위와 그 영역에 고유한 기본 원리나 이론 등에 대한 어휘를 포함하게 된다. 영역 온톨로지에서의 개념들은 일반적으로 상위 온톨로지에서도 정의된 개념들을 특화하는 방식을 취하게 된다.

업무 온톨로지 (Task Ontology)는 진단이나 일정짜기, 판매 등 일반적인 업무나 행위에 대한 어휘를 기술한다. 여기에서도 상위 온톨로지에서도 정의된 용어들을 특화하게 되는데, 동일한 영역이거나 동일한 영역이 아니더라도 일반적으로 업무수행과 관련된 문제를 해결하는 용어들을 포함한다.

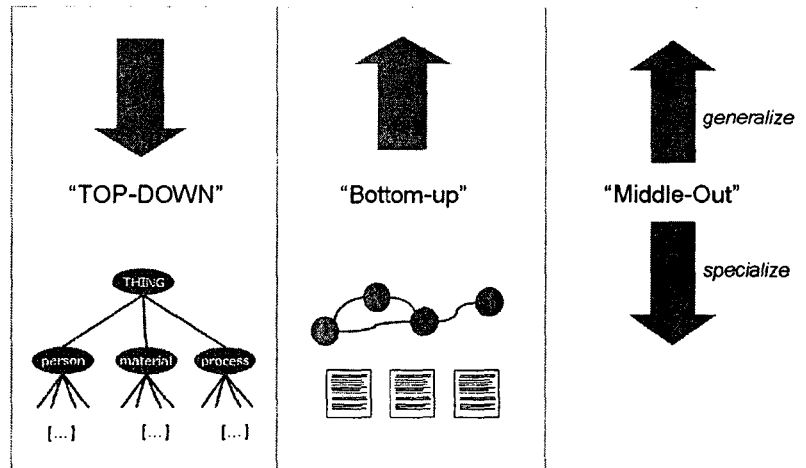
응용 온톨로지 (Application Ontology)는 특정 응용분야에서 요구되는 지식을 모형화하는 데 필요한 모든 정의들을 기술한다. 응용 온톨로지는 흔히 영역 온톨로지와 업무 온톨로지의 어휘들을 해당 응용분야에 맞게 특화하게 된다.

요약하면 상위 온톨로지와 영역 온톨로지는 상호 대비되는 온톨로지이며, 업무 온톨로지는 행위 (doing)에 대한 온톨로지라는 측면에서 사건 (event)나 객체 (thing)에 대한 온톨로지인 개체 온톨로지 (Object Ontology)와 대비된다고 볼 수 있다. 다른 시각에서 온톨로지는 중량 온톨로지 (Heavy-weight ontology)와 경량 온톨로지 (Light-weight ontology)로 구분하기도 한다. 이 구분에 따르면, 중량 온톨로지는 특히 논리적인 방법으로 개념의 정의와 관계들을 완전하게 기술한 온톨로지를 말하고, 이와는 반대로 경량 온톨로지는 가장 전형적이라 할 수 있는 상하관계 (is-a relation)만을 포함하는 부분적으로만 기술된 온톨로지를 말한다.

상위 온톨로지는 모든 실세계에 대한 인간의 지식을 포괄해야 하기 때문에, 엄격하게 이야기해서 세계에 대한 지식이 확립되어야 하는 점을 전제로 하고 있다. 이러한 이유 때문에 상위 온톨로지의 설계는 매우 어려운 일이라고 할 수 있다. 그러나 일반적으로 상위 온톨로지를 설계하는 원리로 다음과 같은 사항들이 제안되어 있다 (Gómez-Pérez et al. (2004) 참고).

- A small number of concepts can cover the world.
- Concepts can be used in lower ontologies.
- Concept should be general and abstract.

온톨로지의 구축방법론은 일반적으로 [그림 2]에 도표로 제시한 바와 같이 세 가지 방법론이 알려져 있다. 본 연구에서는 IT 온톨로지 구축을 위한 방법론으로서 상위 온톨로지 (Upper Ontology)와 영역온톨로지를 결합하는 새로운 방법론에 대한 가능성을 점검해 보고자 한다.

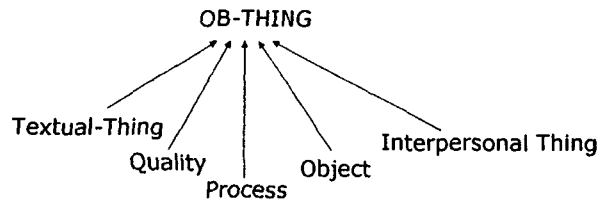


[그림 2] 온톨로지 구축방법론

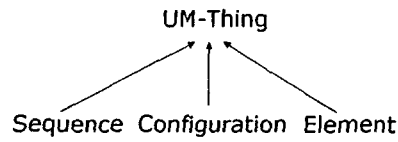
시도해 보고자 하는 새로운 방법론은 이미 존재하고 있는 상위 온톨로지를 수정하여 도입하고, 또 이미 존재하고 있는 영역 온톨로지를 결합하여 기본적으로 양쪽에서 사이를 좁혀 가는 방법으로서, 개념적으로 위의 세 번째 방법론과 정반대의 절차를 밟는 방법론이다.

우선 상위 온톨로지를 접근법에 따라 분류하고 그들의 장단점을 살펴 본 다음에, IT 온톨로지의 구축에 가장 적합한 구현환경을 갖춘 하나의 온톨로지를 정하도록 한다. 그런 다음 IT 각 분야의 영역 온톨로지라고 할 수 있는 온톨로지를 선별하여 두 온톨로지를 결합하는 방법을 모색해 보기로 한다.

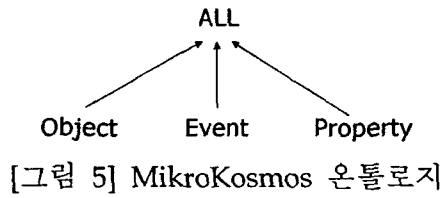
상위 온톨로지를 본 연구에서는 크게 세 가지로 구분하고자 한다. 첫째, 형식적 접근법에 의한 온톨로지, 둘째, 언어학적 접근법에 의한 온톨로지, 세 번째 경험적 접근법에 의한 온톨로지가 그것이다. 형식적 온톨로지에 대한 예로 본 연구에서는 Sowa's top-level ontology, DOLCE 온톨로지를 고찰하기로 하고, 언어학적 온톨로지의 예로서 Penman Upper Model, WordNet을 고찰한다. 또 경험적 접근법에 의한 온톨로지로서 SUMO, Cyc, EDR 온톨로지를 살펴 본다. 각 온톨로지의 개략적 상위 구성은 다음과 같다.



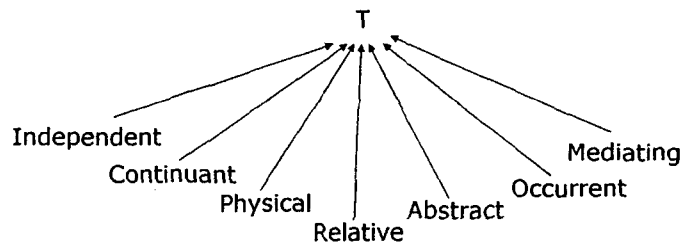
[그림 3] SENSUS 온톨로지



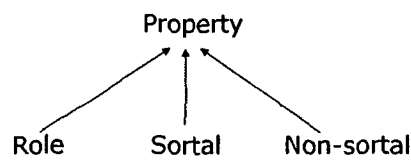
[그림 4] GUM 온톨로지



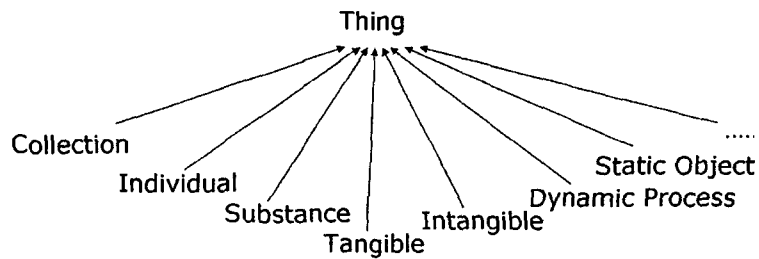
[그림 5] MikroKosmos 온톨로지



[그림 6] SOWA 온톨로지



[그림 7] Guarino & Welty (2000) 온톨로지



[그림 7] Cyc 온톨로지

온톨로지의 구축에서 있어서 가장 중요한 부분은 온톨로지의 설계와 검증이다. 왜냐하면 정확하게 구축되어 있지 않거나 상호모순이 되는 정보를 담고 있는 온톨로지는 실제 응용기술에 적용될 수도 없을 뿐더러 적용된다 하더라도 틀린 정보를 제공하게 될 것이기 때문이다. 온톨로지 기술개발에서 개발과 검증이 중요한 이유를 요약하면 다음과 같다. 다음에 제시할 항목들은 대부분 이미 대용량의 온톨로지를 개발한 경험이 있는 외국의 학자들이 정리해 놓은 것들의 일부이나, 이는 『국가 IT 온톨로지 인프라 기술개발』에서도 필연적으로 해결하거나 미리 방지해야 할 문제로 보인다.

■관계를 기술하는 과정에서 "instance_of"와 "subclass_of"를 구분하는 문제: 예를 들어, "철수"는 "포유동물"의 "instance_of"이고 "인간"은 "포유동물"과 "subclass_of" 관계를 갖는다.

■"part_of"와 "subclass_of"를 구분하는 문제: 예를 들어, "바퀴"와 "자동차"는 "part_of" 관계에 있지 "subclass_of" 관계에 있지는 않다. 이러한 오류는 기경험자의 보고에 의하면 OWL에서 매우 빈번하게 발생한다고 알려져 있다. 왜냐하면, OWL은 "rdf:type"이라는 관계가 있어 편리하게 쓰이지만, 그에 해당하는 "part_of"는 가지고 있지 않기 때문이다.

■"사건"을 "관계"로 모형화하는 문제: 이론언어학 문헌 (Davidson (1967), Parsons (1991) 등 참고)에는 "사건" (event)을 1차술어논리를 이용하여 기술되어 편리하게 사용되나 이를 OWL과 같은 언어에서는 "사건"의 진술을 "관계"에 대한 논항으로 허용하지 않기 때문에 추론이 불가능하다. 이를 극복하는 것도 중요한 연구과제 중의 하나이다.

■온톨로지 구축 기준의 확립문제 ("Ontological promiscuity"): 새로운 용어 (terms)를 온톨로지의 구축과정에서 도입할 때에는 이미 존재하는 다음 용어와 충분히 그리고 명확하게 구분이 되어야 한다. 그 기준을 제시하지 못한 채 구축을 한다면 온톨로지는 구조화되지 않은 자료더미에 불과할 것이다. 이런 의미에서 온톨로지 구축의 기준을 확립하고 설계하는 것은 중요하고, 이 과정에서 언어학자나 인지과학자들의 역할이 중요할 것으로 보인다.

■언어와 개념을 혼돈하는 문제: 인간들은 하나의 개념을 표현하는 데 둘 이상의 낱말, 즉 언어 (language)을 사용하기도 한다. 그러나 온톨로지 구축에서 동일한 개념을 지칭할 때 서로 다른 형식용어를 사용하지 않도록 설계를 해야 한다.

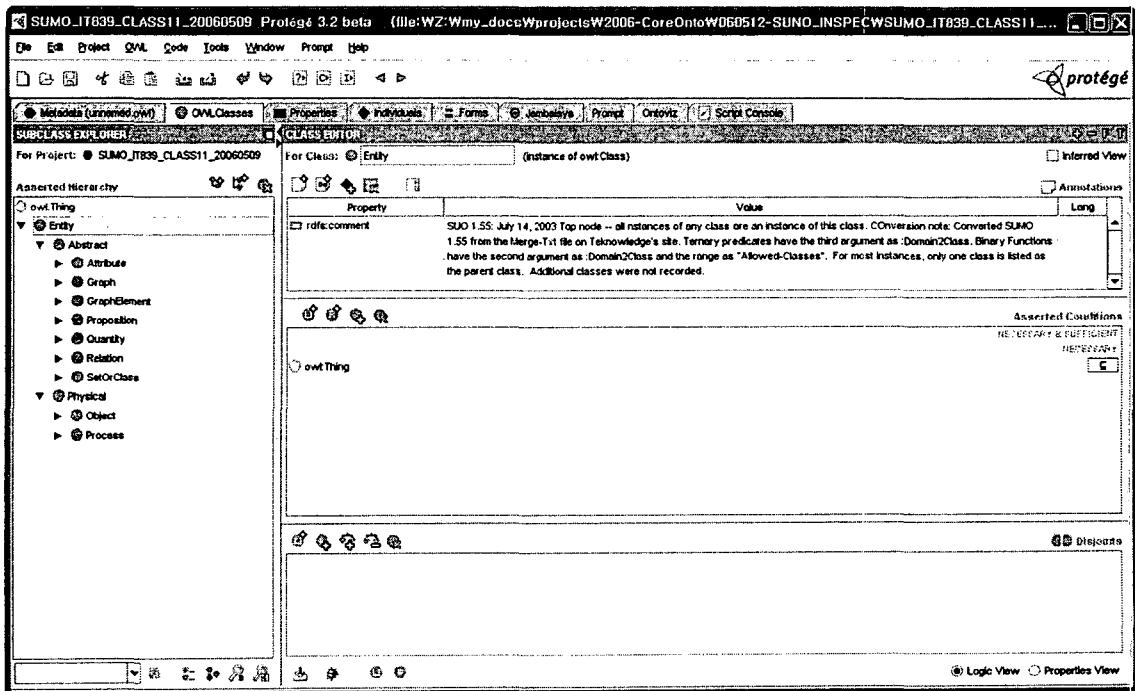
■역할(role)을 "class"로 모형화하는 문제: 개체들은 시간에 따라 부류(class)는 동일하게 유지하지만, 역할 (role)은 시간의 추이에 따라 변할 수 있다. 예를 들어 철수가 선생님이라면 (instance Cholswu Teacher)로 형식언어로 표현될 수 있으나, 철수가 은퇴한 경우는 철수는 선생님의 인스턴스가 더 이상 아니므로 이 경우 철수를 선생님이 지칭하는 것은 오류이다. 중간층위 온톨로지 구축에서 WordNet synsets을 그대로 온톨로지로 사상해 오는 경우, 이런 문제가 발생할 수 있다고 알려져 있고 우리나라에서 온톨로지를 독자적으로 개발할 경우 기존의 CoreNet 등 자원을 검증작업이 없이 그대로 사상해 올 경우 잠재적으로 커다란 문제가 될 것으로 보인다.

■재사용 문제: 온톨로지 구축의 좋은 방법은 검증된 기존의 자원을 최대한 활용하는 것으로 알려져 있다. 그렇지 않을 경우 시간을 낭비할 수 있고 표준화되지 않은 용어나 표상방법을 사용할 여지가 커 결국은 검증된 기존의 자원을 적절하게 사용하여 구축한 것보다 질이 떨어질 가능성이 크다고 지적되어 왔다.

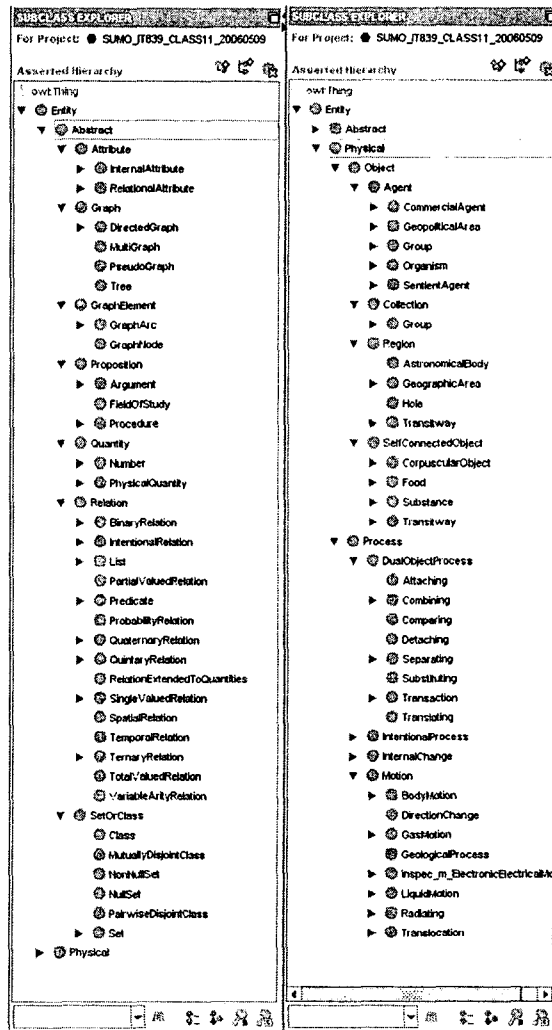
■기타:온톨로지 구축 상의 제반 문제는 현재로서는 위에 제시한 몇 가지의 단순

하지만 부정적인 파급효과가 큰 문제에 국한되어 있으나, 추후에는 언어학적인 쟁점사항들이 크게 부각될 것으로 예측할 수 있다. 예를 들어, 양상을 표현하는 문제("can", "may", "should") 그리고 인식론적 문제를 표현하는 문제("obligation", "permission") 등도 해결해야 할 문제로 부각될 것으로 보인다.

이러한 알려진 난점들을 극복하면서 어떻게 IT 온톨로지를 구축하느냐 하는 문제는 계속된 논의에서 해결해야 하겠지만, 우선 본 연구에서는 상위 온톨로지로서 SUMO (Suggested Upper Merged Ontology)와 INSPEC 온톨로지를 결합시키는 시도를 온톨로지 구축/검증/통합 도구인 Protégé를 사용하여 해 보기로 한다. [그림 8]과 [그림 9]는 실제 구축과정을 보여주는 그림이다.



[그림 8] Protégé로 실행시킨 SUMO와 INSPEC을 결합시킨 OWL 파일



[그림 9] SUMO의 최상위 클래스 Abstract와 Physical

참고문헌

신효필 (2004), 지식기반(Knowledge Base)으로서의 온톨로지(Ontology)와 시맨틱웹 (Semantic Web). 한국정보처리학회 학회지 11:64-75.

양정진 (2003), 시맨틱 웹에서의 온톨로지 공학. 한국정보과학회 학회지 21:28-35.

한국과학기술원 전문용어언어공학연구센터 (2005), CoreNet - 한국어 어휘의미망. 대전: KAIST Press.

최호섭, 옥철영 (2004), 정보검색 시스템과 온톨로지. 한국정보과학회 학회지 22:62-71.

Baker, Collin F., Charles J. Fillmore, and John B. Lowe (1998), The Berkeley FrameNet Project. Paper presented at The COLING-ACL, Montreal, Canada.

Fellbaum, Christiane (1998), Semantic Network of English Verbs. In *WordNet: An*

- Electronic Lexical Database*, ed. Christiane Fellbaum, 69-104. Cambridge, Massachusetts: The MIT Press.
- Gómez-Pérez, Asunción, Mariano Fernández-López & Oscar Corcho (2004), *Ontological Engineering*, Berlin, Springer Verlag.
- Gruber, Thomas R. (1993a), A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition Journal* 5:199-220.
- Gruber, Thomas R. (1993b), Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *Paper presented at The International Workshop on Formal Ontology*, Padova, Italy.
- Guarino, Nicola (1998), Formal Ontology in Information Systems, In: Guarino, N. (ed.) *The 1st International Conference on Formal Ontology in Information Systems (FOIS '98)*, Trento, Italy, IOS Press, Amsterdam, 3-15.
- Guarino, Nicola and Chris Welty (2000), A Formal Ontology of Properties. In, Dieng, R., and Corby, O., eds, *Proceedings of EKAW-2000: The 12th International Conference on Knowledge Engineering and Knowledge Management*. Spring-Verlag LNCS Vol. 1937:97-112. October, 2000.
- Hamp, Birgit, and Helmut Feldweg (1997), GermaNet - a Lexical-Semantic Net for German. Paper presented at The ACL/EACL-97 workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP applications, Madrid.
- Noy, Natalya F., and Deborah L. McGuinness (2001), *Ontology Development 101: A Guide to Creating Your First Ontology*, ed. Technical Report: Stanford Knowledge Systems Laboratory.
- Ruppenhofer, Josef, Michael Ellsworth, Miriam R. L. Petruck, and Christopher R. Johnson (2005), *FrameNet: Theory and Practice*.
- Talmy, Leonard (1985), Lexicalization patterns: Semantic Structure in Lexical Forms. In *Grammatical Categories and the Lexicon. Language Typology and Syntactic Description*, ed. Timothy Shopen, 57-149: Cambridge Uni. Press.