Semantic Lexical Resources Applied to Content-based Querying – the OntoQuery Project

Bolette S. Pedersen & Patrizia Paggio

Center for Sprogteknologi
Njalsgade 80, 2300 S
{bolette patrizia}@cst.dk

Abstract

This paper deals with the exploitation of the lexical and conceptual knowledge coded in the SIMPLE-DK lexicon in the methodology for content-based querying developed by the OntoQuery project. SIMPLE-DK has proven a rich and flexible lexical resource, which the project has taken advantage of in several ways. Firstly, the paper explains how the ontology provided by SIMPLE is used by the current project prototype to derive conceptual descriptors on which to base the matching of documents to user queries. Furthermore, it discusses how selectional restrictions and qualia roles, both coded in SIMPLE, can be used to construct an ontological grammar to build more complex descriptors.

1. Introduction

This paper deals with the exploitation of the lexical and conceptual knowledge coded in the SIMPLE-DK lexicon for context-based querying in the research project OntoQuery (Andreasen *et al.* 2000).

OntoQuery is developing a methodology for content-based querying and retrieval based on the extraction and evaluation of conceptual content from noun phrases (NPs) in texts and queries. Both extraction and evaluation presuppose the availability of an ontology, and the SIMPLE-DK lexicon has been chosen as the lexical backbone of the project since it provides it with a well-defined and carefully established top ontology. Nutrition has been selected as a first test domain. Therefore, an ontology for this domain has been established and merged with the SIMPLE top ontology. The hyponymy and synonymy relations coded in the resulting ontology are applied in the current OntoQuery prototype to derive semantic descriptors representing the content of NPs, and the ontology is also used by the matching algorithm. Furthermore, a more complex model is being developed in which selectional restrictions and qualia roles from SIMPLE-DK, are used to construct an ontological grammar which allows richer descriptors to be generated in which complex concepts can be built by combining simple domain concepts with semantic roles.

The approach advocated in OntoQuery is in line with previous experiments concerned with the expansion of concepts based on the use of wordnets as done for English in Voorhees (1993), Voorhees (1994), Smeaton & Quigley (1996) and in a cross-lingual environment in Gonzales *et al.* (1998). However, OntoQuery is not just replicating the results achieved in those projects. SIMPLE-DK is in fact based on quite a different approach to lexical semantics and provides therefore different concept expansion possibilities. Moreover, the use of an ontological grammar drawing on information coded in SIMPLE-DK, constitutes an important enhancement, and a step towards a more complex semantic analysis of texts.

After a short presentation of the methodology for content-based querying developed in OntoQuery in Section 2, we describe the Danish SIMPLE lexicon in Section 3: we give examples of qualia structure instantiation and selectional restrictions for different word senses and explain how the ontology created for OntoQuery has been constructed on the basis of SIMPLE. In Section 4 we give an account of the way in which the ontology is used by the matching algorithm implemented in the current OntoQuery prototype. Finally in Section 5, we address the issue of how selectional restrictions and qualia roles can provide the basis for the formulation of an ontological grammar. Section 6 contains a brief conclusion.

2. Content-based querying and *ontotypes*

The aim of the linguistic and semantic analysis carried out in OntoQuery is to identify the concepts corresponding to the NPs that occur in texts and queries. The choice of focussing on NPs, is motivated by the fact that these have a clear conceptual content that can be captured in an ontology-based framework. On the contrary, verb phrases are not taken into account since they behave much less clearly from the point of view of ontological structuring. For the same reason, our analysis is further restricted to those parts of the noun phrase that contribute to its conceptual content, i.e. the head noun and the modification provided by adjectives, genitives and prepositional phrases, thus excluding determiners and for the moment also relative clauses.

The ontological descriptor corresponding to a linguistic expression is an algebraic term, or *ontotype*, associated with a node in a lattice of concepts (Nilsson 2001). For example, the ontotype for *fedtdepoter hos børn* (fat deposits in children) is the complex concept resulting from the combination (via the meet operator x) of the atomic concept *depot* (deposit) with the concepts *fedt* (fat) and *barn* (child) by means of the relations CON (contains) and LOC (located-in).

(1) (*depot* x (CON: *fedt*) x (LOC: *barn*))

The relation-concept pairs (CON: *fedt*) and (LOC: *barn*) – also referred to as semantic roles – are valid restrictions of the concept *depot*, and the resulting complex concept can thus be regarded as a subtype of this. Valid combinations of concepts and semantic roles are expressed in an ontological grammar the specific formulation and implementation of which is still under discussion in the project (see Nilsson 2001b and Paggio 2001 for a discussion). Experiments are being carried out with a number of different formalisms to find an adequate framework. One such framework is typed feature structures as implemented for example in the LKB system (Copestake 1999), where semantic restrictions can be expressed as semantic feature types to constrain the application of syntax rules. In the example under consideration, the concept *depot* inherits the following restrictions from the more general *container* concept:

(2)
$$\begin{bmatrix} \textit{container} & \\ \text{SELECTS} & \left\langle \begin{array}{l} \begin{bmatrix}\text{CON} & \textit{substance}\end{bmatrix} \\ \begin{bmatrix}\text{LOC} & \textit{entity}\end{bmatrix} \end{array} \right\rangle \end{bmatrix}$$

The feature SELECTS indicates a list of semantic roles that can be associated with a given concept, here a contains and a located-in role.

In order to map NPs onto ontological descriptors, various NLP techniques are applied. Queries and texts are tokenised and part-of-speech tagged; then all NPs are recognised and words are lemmatised and replaced by the corresponding concepts in the ontology (Pedersen *et al.* 2001). An NP parser must then apply syntax rules and semantic constraints to produce semantic descriptors similar to that in (1). Currently however, the linguistic and semantic analysis is less fine-grained, and each NP is represented as a set of concepts, with the relations in between them left undefined. Thus, instead of the descriptor in (1), the analysis of the NP *fedtdepoter hos børn* will currently produce the underspecified descriptor shown below:

(3) (*depot, fedt, barn*)

Descriptors extracted from a user query are evaluated by comparing them with descriptors previously extracted from texts, and the results are scored based on how distant concepts are from each other in the domain ontology. A more formal description of this process is given in Section 4. In the following section, we describe the principles behind the SIMPLE-DK lexicon, especially the semantic information relevant to the generation of semantic descriptors.

3. The Danish SIMPLE Lexicon

The lexical and ontological resources applied in OntoQuery originate from the Danish SIMPLE lexicon developed within the EU-project SIMPLE (Semantic Information for Multifunctional Plurilingual Lexica). This project aimed at providing harmonised semantic lexicons for Natural Language Processing for 12 of the European languages (Lenci *et al.* 2000). The project developed an extension of the LE-PAROLE lexicons, which contain 20,000 entries with corresponding morphological and syntactic information for each of the 12 languages that were covered in the PAROLE project, cf. (Ruimy *et al.* 1998).

The language specific encodings in SIMPLE are performed on the basis of a unified, ontology-based semantic model – the so-called SIMPLE model – representing an extended qualia structure based partly on Pustejovsky (1995), partly on experience in previous lexical projects such as Genelex, WordNet (Miller et al. 1990) and EuroWordNet (Vossen (ed.) 1999). A general design model is thus provided allowing for the encoding of a large amount of semantic information such as ontological typing, domain information, qualia structure, argument structure, event structure and selectional restrictions.

Consider for illustration the entry for the concept *kanin* (rabbit) in Figure 1.

Semantic Unit	*kanin_EAN* (rabbit)
Definition:	*en lille hare som har forholdsvis korte ører, holdes som kæledyr* (a small hare with relatively small ears; often kept as a pet)
Corpus example:	*Klitområderne vil få en ny indvandring af kaniner* 'the dune areas will get a new invasion of rabbits'
Ontological type:	Earth animal
Supertype:	Animal
Domain:	General
Formal quale:	Is_a = *hare* (hare)
Agentive quale:	Nil
Telic quale:	Nil
Constitutive quale:	Nil
Complex :	*kanin_SUF* (Systematic Polysemy**)**

Figure 1: The semantic unit for the 'animal' reading of *kanin* (rabbit)

Note that a systematic polysemy relation is established to the substance food reading of rabbit via the relation *complex*. The food reading has a richer qualia structure encompassing also its coming about as well as its function, as seen in Figure 2 – factors which are not encoded for natural kinds.

Semantic Unit	*Kanin_SUF* (rabbit)
Definition:	*kød af kanin* (the meat from a rabbit)
Corpus example:	*Mærkeligt nok er kanin ikke særlig brugt kød herhjemme* 'strangely enough, rabbit is not a very used kind of meat in our country'
Ontological type:	Substance food
Supertype:	Substance
Domain:	Nutrition
Formal quale:	Is_a *kød* (meat)
Agentive quale:	Made_by *tilberede* (prepare)
Telic quale:	Used_for *spise* (eat)
Constitutive quale:	Nil
Complex :	*Kanin_EAN* (Systematic Polysemy**)**

Figure 2: The semantic unit for the 'food' reading of *kanin* (rabbit)

For an example of a lexical entry for deverbal nouns which include selectional restrictions, consider Figure 3 below.

Semantic Unit	*behandling_PUA* (treatment)
Definition:	*Forsøg på at helbrede nogen el. lindre deres smerter* (attempt of curing somebody or relieve their pains)
Corpus example:	*Undersøgelse og behandling af kræftpatienter blev støttet med 30 mill* (examination and treatment of cancer patients were supported by 30 mill.)
Ontological type:	Purpose act
Unification Path:	Relational act/Telic
Domain:	Medicine
Argument Structure	ARG1 ARG2 ARG3 INSTRUMENT
Selectional Restrictions	ARG1= Human ARG2 = Disease ARG2P = Human OR Animal INSTRUMENT = concrete entity
EventType	Process
Formal quale:	Is_a = *handling* (act)
Agentive quale:	Nil
Telic quale:	Purpose = *helbredelse* (curing)
Constitutive quale:	Nil
Systematic Polysemy	Nil
Synonymy	Nil

Figure 3: The semantic unit for *behandling* (treatment)

Several operations have been performed on the Danish SIMPLE lexicon in order to make it into an operational language-specific ontology on which content-based querying in Danish could apply. First of all, the hyponymy relations (*is_a*) encoded in the Danish SIMPLE lexicon have been used to establish a preliminary general language ontology for Danish which again maps onto the SIMPLE top ontology of approx. 140 meta-concepts. The boundary between the top ontology, which is shared by all twelve language-specific SIMPLE lexicons, and the rest of the ontology, is clearly marked by the fact that English words are used as top concept names, whereas Danish words are used for the bottom part.

For illustration, the 'food' reading of *kanin* (rabbit) is linked to the superconcepts in the ontology as shown in Figure 4.

```
Top
    Telic
    Concrete entity
        Food
            Substance Food
                mad (food)
                    kød (meat)
                        kanin (rabbit)
```

Figure 4: The hyperonyms of the food reading of *kanin*

Note that the top concept Food is a so-called unified type and therefore has multiple coordinates. It inherits thus both from the superconcept Concrete entity and from the superconcept Telic (which according to Pustejovsky's qualia roles encompasses the *function* of food, i.e. that it is to be eaten). Since the Danish SIMPLE lexicon contains only 10,000 frequency based semantic senses, additional senses that are central to the structure of the ontology (typically Danish top-ontology or near top-ontology concepts) have been added.

A second operation concerns the addition of a specific ontology within the specialised domain of *nutrition* which has been chosen as the test domain for the system. Nutrition texts from a Danish encyclopaedia are used as a basis, meaning that the ontology is built bottom-up mainly with the perspective of a layman, since this is the perspective adopted in the texts. The nutrition ontology consists of approx. 1,000 concepts which are attached to the SIMPLE top ontology primarily under the top concepts Substance, Substance Food and Disease. The internal structure of the nutrition ontology is highly taxonomical; for illustration, consider in Figure 5 the superconcepts of *b-vitamin*.

```
Top
  Concrete entity
    Substance
      ernæringsstof (nutrition substance)
        mikronæringsstof (micronutrition substance)
          vitamin (vitamin)
            vandopløseligt vitamin (watersoluble v.)
              b-vitamin (vitamin b)
```

Figure 5: The hyperonyms of *b-vitamin*

4. Applying the ontology in searching

As briefly mentioned above, the concepts found in a query are downwards-expanded by the search engine using the *is-a* and *synonymy* relations from the ontology (Andreasen 2000) to find matching NPs in the texts.

If for instance a query *'er der b-vitaminer i kornprodukter ?'* (is there vitamin B in corn products ?) is input to the OntoQuery Prototype, the analysis shown in Figure 6 is produced (for space reasons, only the first two hits out of 23 are shown).

OntoQuery Prototype

Tagger final state: er/V_PRES der/UNIK b-vitaminer/N i/PRÆP kornprodukter/N

NP-recognizer:
[V_PRES er] [UNIK der] [NP [N b-vitaminer]] [PRÆP i] [NP [N kornprodukter]]

Query: er der B-vitaminer i kornprodukter

(B-vitamin), (kornprodukt)

- 0.90 **niacin**: Pellagra forekommer fortrinsvis hos ulandsbefolkninger, der indtager en ensidig kost bestående af majs eller andre kornprodukter, hvori niacin er fast bundet, og hos personer, der indtager en kost med lavt proteinindhold eller er i længerevarende behandling med lægemidler, som hæmmer omdannelsen af tryptofan til nikotinamid.

(nikotinamid),(kost),(tryptofan),(pellagra),(ensidig,kost),(majs),(korn produkt),(person),(lav,proteinindhold), (behandling),(lægemiddel)

- 0.50 **niacin**: niacin, (sammentrækning af eng. nicotinic acid vitamin), fællesbetegnelse for nikotinsyre ($C_6H_5O_2N$) og nikotinamid, der indgår i gruppen af B-vitaminer, tidligere benævnt B_7-vitamin.

(sammentrækning),(vitamin),(nikotinsyre),(nikotinamid),(gruppe),(B -vitamin),(tidlig)

Figure 6: Search results for the query *er der b-vitaminer i kornprodukter?*

Query evaluation is based on two-level order-weighted aggregation, where the two levels are determined by the set of sets structure of the descriptions representing the texts to be retrieved. The similarity of the concepts to be aggregated is computed by a naive approach based on distance in the ontology so that a query concept X matches another concept Y by $1 - d(X,Y) / 1$ (see Andreasen 2001 for more details). Thus a concept X in the query matches X in a text by 1.0, an immediate sub-concept by 0.9, a second-level sub-concept by 0.8, etc. Therefore, in Figure 6 above, *b-vitamin* matches *nikotinamid* by 0.8 since it is a subconcept to *niacin* which is again a subconcept to *b-vitamin*, whereas *kornprodukt* obviously matches *kornprodukt* by 1.0. The degree to which the set of sets ((b-vitamin) (kornprodukt)) representing the content of the query, matches the content of the text retrieved is aggregated by simple average to yield 0.90. This example shows the usefulness of the ontology since a traditional pattern matching procedure would not be able to establish a relationship between *b-vitamin* and *nikotinamid*.

The first preliminary evaluation of the system (see Pedersen & Paggio (forthcoming)) show very similar results as regards the way in which the retrieved texts pattern together. Typically, for each query, very few texts are retrieved with a score between 0.90 and 1.00. In contrast, a relatively large number of the retrieved texts have a score of 0.50 and below. From an evaluation point of view, the retrieved texts with a score from 0.90 to 0.95 are of particular interest: they do not contain exactly the same concepts as in the queries but rather specifications of these, but they are still very good answer candidates to the queries put forward. In other words, in a retrieval system without ontological knowledge, these texts would in most cases have been 'hidden' in the large group of retrieved texts with a score of 0.50 or below, if retrieved at all.

5. Exploiting selectional restrictions and qualia structure

As we saw, the ontotype for an NP is not just a set of concepts, but a complex term consisting of a head concept plus a set of relation-concept pairs according to a conceptual grammar where valid concept combinations are defined. Such a representation constitutes a richer basis on which to define a matching algorithm to be used by the search engine. In this section, we focus on how a conceptual grammar can build on the selectional restrictions and qualia roles coded in SIMPLE.

Let us consider the example *behandling af børn med overvægt* (treatment of children with overweight [problems]). In OntoQuery, the selectional restrictions associated with the concept *behandling* state that the patient (PNT) role must be filled out by a *human* or an *animal*, and the instrumental (BMO: by-means-of) by a *concrete-entity*. The two relations correspond to ARG2P (second participant) and INSTRUMENT in the SIMPLE entry shown in Figure 3. These restrictions license a patient interpretation of the PP *af børn* (of the children) and rule out an instrumental interpretation of *med overvægt* (with overweight [problems]). The point is of course that *med overvægt* cannot play the instrumental role since *overvægt* (overweight) is not a *concrete-entity*. The concept *overvægt* has *disease* and *state* as superconcepts. It combines with the preposition *med* to express a characteristic (CHR). The correct representation is shown in (3), where the final PP is analysed as a characteristic of the concept *barn* (child).[1]

(3) (*behandling* x (PNT :(*barn* x (CHR: *overvægt*))))

[1] The right-low attachment of (CHR: *overvægt*) to the PP *af børn* rather than to the head noun *behandling* cannot be handled by the conceptual grammar. Thus the system would produce two analyses. A structural preference rule could possibly be used to pick the correct one.

Selectional restrictions are relevant for deverbal nouns like *behandling* (treatment). For a concept like *depot* (deposit) there is no corresponding verbal event to derive selectional restrictions from. However, the qualia roles can be resorted to. Thus, the PP *af vitaminer* in (4) below fills out the ‘contains’ relation in the *constitutive* qualia role of *depot*. Likewise in *fedtdepoter*, the same role is filled out by the first noun component of the compound. All the examples in (4) display the same basic semantic structure:

(4) (*container* x (CON: *substance*))

depoter af vitaminer (deposits of vitamins)
vitamindepoter (vitamin deposits)
fedtdepoter (fat deposits)
depoter af K-vitamin (deposits of vitamin K)

Thus, the semantic restriction can be expressed in a more general way by associating it with the superconcept *container*, as was shown in Section (2). In general, the idea is to extract such restrictions from the qualia structures of very specific concepts and wherever possible, associate them with concepts placed higher up in the ontology. Examples of such general combinations of a concept with semantic roles valid in the nutrition domain are:

(5) (*substance* x (LOC: *concrete-entity*)
x (CBY: *event*)
x (SRC: *concrete-entity*)
x (POF: *concrete-entity*))

The semantic relations used are LOC for located-in, CBY for caused-by, SRC for source, and POF for part-of. In contrast, the combination (*substance* x (TMP: *time*)) is not allowed by the ontological grammar (see Nilsson 2001b for a discussion).

Currently, the OntoQuery project is working with domain-specific knowledge, and the prototype can be used to query a collection of texts belonging to the domain covered. However, we are aware of the fact that this is a somewhat artificial restriction and that the model would gain in generality were it able to deal with a broader range of texts. In this connection, the qualia structure is an obvious source to be looking at to disambiguate between word senses. An example is that of *kanin* (rabbit), which has, as we saw earlier, at least two different regular polysemous senses: the ‘animal’ and the ‘food’.[2] Apart from having different hyperonyms, the two senses also differ with respect to their qualia structures.

To start with the sense most relevant to the nutrition domain, the ‘food’ one, its telic role tells us that it is used for eating, and the agentive that it is made by cooking. In (6) below, the agentive role is in fact made explicit by the verb *steg* (to cook) indicating that we are dealing with the food sense of rabbit:

[2] A ‘fur’ sense of rabbit is in fact also coded in SIMPLE. For a discussion of its semantics, see Pedersen & Paggio (forthcoming).

(6) *Luk maven med kødnåle og steg <kaninen> i ovnen ca. 11/2 time 2 timer.*
(Close the stomach with skewers and cook the rabbit in the oven for about 1 1/2 to 2 hours)

Let us now consider the “default”, animal interpretation of *kanin*. The only qualia relevant to this sense is the formal one, in other words it is the *is-a* relation. Corpus examples displaying this sense are shown in (7) and (8):

(7) *På dyrehospitalet i Valby har jeg set en tam <kanin> med brækket ryg blive opereret for en sum, der ...*
(At the animal hospital in Valby I have seen a tame rabbit with a broken back operated on for a sum that ...)

(8) *det er sjovt at sidde i haven og pludselig få besøg af høns og <kaniner> eller en nysgerrig kalkun*
(it is fun to sit in the garden and suddenly be visited by hens and rabbits or a curious turkey)

A strong disambiguating element is the presence of other concepts filling out the same argument position and belonging to the ‘animal’ type, like hens and turkeys in (8). But selectional restrictions also play a role: only an animal (and not a type of food) can be operated on as in (7).

Thus, by enriching the OntoQuery ontology with the information coded in the qualia roles of the lexical items contained in SIMPLE, the system would be able to generate richer semantic descriptors than those currently implemented, and also be in a better position to distinguish between word senses.

6. Conclusions

The application of the Danish SIMPLE lexicon to content-based querying in the OntoQuery project, has confirmed that SIMPLE constitutes a flexible and very rich lexical and ontological source. We have seen that the SIMPLE ontology, after the extension provided by a domain-specific sub-ontology, is used by the OntoQuery search engine to perform query expansion and description matching. Furthermore, SIMPLE provides other types of semantic information – selectional restrictions and qualia structure – which the project is exploiting in the formulation of a conceptual NP grammar with the purpose of generating richer semantic descriptors on which to base the retrieval and ranking of relevant texts.

Acknowledgements

The work presented in this paper builds on work carried out by a large group of researchers. In particular, the Danish SIMPLE lexicon has been developed at CST by the following main staff members: Sanni Nimb, Sussi Olsen, Anna Braasch, Britt Keson, and Bolette S.

Pedersen. The OntoQuery prototype has been developed at the University of Roskilde under the leadership of Troels Andreasen, whereas the nutrition ontology has been developed mainly by Nikolaj Oldager at the Danish Technical University. The language technology components of the OntoQuery prototype have been developed at CST by Dorte Haltrup and the authors with contributions from Hanne Thomsen and Bodil Nistrup Madsen from the Copenhagen Business School. Finally, the formal logic used by the search engine is the work of Jørgen Fischer Nilsson from the Danish Technical University, whereas the ideas concerning a conceptual grammar have been developed by Jørgen Fischer Nilsson in cooperation with Per Anker Jensen, Southern Danish University, and Carl Vikner, Copenhagen Business School.

References

Andreasen, T., J. F. Nilsson, H. Thomsen, 2000. "Ontology-based Querying", in: H.L. Larsen *et al.* (eds.) *Flexible Query Answering Systems, Recent Advances*, Physica-Verlag, Springer, pp. 15-26.

Andreasen, T., 2001. "Query Evaluation Through Concept Descriptions", in: Thomsen (ed.) *OntoQuery Workshop Proceedings, LAMBDA,* Copenhagen Business School.

Copestake, A., 1999. *The (new) {LKB} system – version 5.2*. CSLI, Stanford.

Gonzales J., F. Verdejo, C. Peters, N. Calzolari. 1998. "Applying EuroWordNet to Cross-lingual Text Retrieval", in: *Computers and the Humanities Vol. 32*: 185-207, Kluwer Academic Publishers, The Netherlands.

Lenci, A., N. Bel, F. Busa, N. Calzolari, E. Gola, M. Monachini, A. Ogonoski, I. Peters. W. Peters, N. Ruimy, M. Villegas, A. Zampolli. 2000." SIMPLE – A General Framework for the Development of Multilingual Lexicons", in: T. Fontenelle (ed.) *International Journal of Lexicography Vol 13*. pp. 249-263. Oxford University Press.

Miller, G. R. Bechwithm C. Feldbaum, D. Gross, K.J. MIller " An On-line Lexical Database", in: *Internationl Journal of Lexicography 3(4), p. 235-244*.

Nilsson, J.F.,2001. "A Logico-Algebraic Framework for Ontologies, in: P. Jensen & P. Skadhauge *Ontology-based INterpretations of Noun Phrases, Proceedings from the First International OntoQuery Workshop.* University of Southern Denmark, Kolding.

Nilsson, J.F., 2001b."Generative Ontologies, Ontological Types, and Conceptual Grammar", in H. Thomsen *Proceedings from OntoQuery Workshop on Ontologies and Search, LAMBDA pp. 45-53*, Copenhagen Business School.

Paggio, P. "Parsing in OntoQuery – Experiments with LKB", in: P. Jensen & P. Skadhauge *Ontology-based Interpretations of Noun Phrases, Proceedings from the First International OntoQuery Workshop.* 2001. University of Southern Denmark, Kolding.

Paggio, P; Pedersen, B.S.; and Haltrup, D., 2001. "Applying Language Technology to Content-based Querying" The Ontoquery Project, in Proceedings from Workshop on Artificial Intelligence for Cultural Heritage and Digital Libraries pp. 75-79. Università di Bari, Italy.

Pedersen, B.S., Paggio, P., forthcoming. "A Danish Semantic Lexicon and its Application in Content-based Querying" in: *CST Working Papers.* Center for Sprogteknologi, Copenhagen.

Pustejovsky, J., 1995. *The Generative Lexicon*, Cambridge, MA, The MIT Press.

Ruimy, R., N. O. Corazzari, E. Gola, A. Spanu, N. Calzolari, A. Zampolli, 1998. "The European LE-PAROLE Project: The Italian Syntactic Lexicon", in: *First International Conference on Language Resources & Evaluation pp 241-249*, 1998. Granada, Spain.

Smeaton, A. & A. Quigley, 1996. "Experiments on Using Semantic Distances between Words in IMage Caption Retrieval", in : Proceedings of the 19th International Conference on Research Development in IR.

Voorhees, E. M., 1993. "Using WordNet to disambiguate word senses for text retrieval." In: Korfhage, Robert, Edie Rasmussen and Peter Willett, eds., Proceedings of the 16th Annual ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburgh, pp. 171 - 180.

Voorhees, E. M., 1994. "Query expansion using lexical-semantic relations". In: Croft, W. Bruce and C. J. van Rijsbergen, eds., *Proceedings of the 17th Annual ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 61 - 69.

Vossen, P. (ed.), 1999. *EuroWordNet, A Multilingual Database with Lexical Semantic Networks*, Kluwer Academic Publishers, The Netherlands.