View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Archivio istituzionale della ricerca - Università degli Studi di Venezia Ca' Foscari

Inducing frames in the Italian Lexicon

Rodolfo Delmonte
Department of Language Sciences
Università Ca' Foscari - Ca' Bembo
30123 - VENEZIA (It)
E-mail: delmont@unive.it
Website: project.cgm.unive.it

Abstract

In this chapter we will present a computational approach to the problem of lexical access based on FrameNet and similar resources. The starting point will be FrameNet (hence FN) in its latest release 1.3, which is rightly regarded as the most reliable source of lexical knowledge for semantic and syntactic processing. FN is a lexical knowledge repository containing knowledge representation of the world allowing inference to be fired at different levels, close to a full-fledged ontology much in the same vein as WordNet.
In commenting FN peculiar features we will compare it to a number of other similar lexical resources: LCS (Lexical Conceptual Structures), WordNet, Levin's WordList, PropBank and VerbNet. Each of these publicly available lexica contains some feature that make it comparable to FN, even though they all lack in its fine-grained overall data description. In particular, even though they are all handcrafted, they all allow some type of inference to be fired directly; some contain such an induction mechanism at the level of lexical structure, some other as additional superstructure.
In particular we will focus on Dorr's Lexicon of Conceptual Structures which constitutes a valid companion to FN and other lexical resources. We will introduce the underlying theoretical framework in some detail seen that this is also what we have been using in our work on the lexicon of Italian. One feature in particular will be made object of further scrutiny: the notion of Causality and the way it is represented at a lexical level. One section of the chapter will be devoted to explore this notion in detail and the way in which it is tackled in FN and LCS.
Eventually, we will describe our effort of building one such resource in late '80s for Italian. In that effort we tried to reconcile knowledge of the world and linguistic semantic and conceptual information needed to produce highly structured lexical entries as the ones provided by FN and LCS. Besides presenting the algorithm and its linguistic components, we will highlight those features of our approach that make it different and, in some cases, computationally more perspicuous than the others. As will be shown, the premises we adopted can lead in some cases to a more complete and richer representation.

1. Introduction

In this paper we will present a computational approach to the problem of lexical access based on FrameNet(hence FN), which may be regarded the lexical repository closest to a description of the world as it is. In this sense, FN is a database containing knowledge representation of the world which allows inference to be fired at different levels, close to a full-fledged ontology much in the same vein as WordNet.
However, from a computational point of view, FN is considered a too sparse source of lexical data to serve as a useful tool in open domain, unrestricted text processing (see Green et al., 2004). The current version of FN is useful for a certain number of different text processing tasks, like using Frames and Frame Elements to produce more perspicuous semantic roles associated to the output of a dependency parser. Eventually there is always the possibility to extend it manually – if needed - to other lexical units

which share/evoke the same concept/frame. This is certainly beneficial in case small texts in highly restricted domains are the target of the analysis.
In commenting FN peculiar features we will compare it to a number of other similar lexical resources: LCS (Lexical Conceptual Structures), WordNet, Levin's WordList, PropBank and VerbNet. Each of these publicly available lexica contains some feature that make it comparable to FN, even though they all lack its fine-grained overall data description. In particular, even though they are all handcrafted, they all allow some type of inference to be fired directly; some contain such an induction mechanism at the level of lexical structure, some other as an additional superstructure.
In our perspective, however, lexical knowledge associated to a given predicate or lexical entry, is gathered from sentence level or propositional level syntactic/semantic structure and is constituted of the following linguistic items:
1. Entities, Properties and Events
2. Events or Situations
3. Participants (arguments)
4. Semantic Roles (the situation)
5. Perspective or Point of View
6. Temporal Extension of the Event

As a consequence of that, there is a number of issues that are strictly related to lexical acquisition that need explanation:

7. Meaning of lexical entry is related to the actual world or not
8. Events carry consequences on the state of affairs described
9. Relations of events to spatiotemporal locations of arguments may change or not
10. Complexity of lexical meaning contained in the lexical entry

Differently from the theory put forward by Fillmore in FN, we assume that lexical knowledge is not attached to single lexical units, which "evoke Frames". Fillmore's underlying theory is Construction Grammar (hence CxG) (see Croft, Kay et al., Fillmore et al., Hudson). CxG assumes that linguistic knowledge is all lexically fixed by grammatical constructions which are syntactic templates where form and content are coupled: no derivations and no semantic de-compositional processes are needed. Phonology, prosody, semantics, pragmatics and indeed syntax are part of grammatical constructions which are organized in a hierarchical set of models. However, if this theory were really computable and mentally or cognitively plausible, one would have to assume an inventory of some million different such grammatical constructions covering all possible semantically and pragmatically relevant combinations of atomic morphemic units with their corresponding sounds. CxG uses notions such as Default Inheritance and Complete Inheritance to induce generality in the network constituted by grammatical constructions: however, the model of CxG proposed by FN and more accepted seems the Usage-based model. This model is based on inductive learning, i.e. linguistic knowledge is acquired bottom-up from real sentences. Generalization on similar structures is obtained through use of recurrent constructions. In such a framework there is no Universal Grammar, no Typological Universals seen that mapping from form to content is strictly language and construction dependent. Eventually, no formal distinction between lexical (semantic) and syntactic structures.
Rather, we assume that computability is a cognitively sound gauge of the psycholinguistic plausibility of any theory: human languages are characterized by "Recursivity", and make use of semantic types and compositionality of meaning allows all possible combinations to be computed smoothly. Words need to be computed w.r.t. their linguistic context which alone can guarantee the semantic representation to match with predicate-argument structure (hence PAS) in the lexicon. From PAS Frames can be evoked and Frame Elements be associated to arguments of the governing predicate.
Principles underlying CxG are listed in Holmes and Hudson (2001:2) and reported here below, where the authors match and combine principles formulated in Kay & Fillmore (the CG manifesto) and in Hudson's Word Grammar:

• The goal of linguistic theory is "to account for the entirety of each language", including "non-core" patterns as well as the central core.

• No distinction is assumed (or found) between 'rules' and 'lexical items', so a linguistic theory must include "an explicit system of representation, capable of encoding economically and without loss of generalization all the constructions (or patterns) of the language, from the most idiomatic to the most general".
• The list of constructions is the database of "an explicit, non-derivational (constraint based) grammar", so the grammar is generative (explicit) but not derivational (transformational).
• Syntactic patterns are intimately bound to semantic ones so that "syntactic and semantic information is represented within a single feature structure"; each grammatical construction is "a conventional association of linguistic form and content".
• Complex patterns in sentence structure are generated by the interaction of a multiplicity of individually much simpler patterns. In CG the simpler patterns are called 'constructions', so the grammar must be able to integrate "both constructions and the words, phrases and sentences of the language which they license - which we call 'constructs' ..." (It is true that the terms "construction" and "construct" have not been generally used in WG, but they apply perfectly to the very simple basic patterns of WG and the more complex patterns that they license. In both theories the term 'inherit' is used for the relation of a construct to its licensing constructions.)
• Semantic structures must show the fine grain of lexical semantics as well as the broader structures due to syntax; for example, the analysis of GIVE must include "a set with four members, each ... representing a minimal predication, consisting of a frame plus its participants or arguments...". The semantic structure must accommodate pragmatic information such as illocutionary force (e.g. request for information) and presupposition (e.g. that the scene described is "incongruous" as in the famous What's X doing Y? construction).
In our perspective, syntactic structure is independently built and has independent properties and principles that help in the semantic interpretation process. Even though we are also lexicalist – as LFG is one such theory – we believe in the usefulness of lexical and syntactic rules and in their independence of semantic rules. Some of the rules depend on lexical representations, some others are part of any semantic theory and are applied at propositional level through some Logical Form representation.
This will be discussed in the following sections: Section 2 presents details of Frame theory and its main features. In the same section we also present similar resources available which describe other important elements of lexical semantic representation. Section 3 is devoted to discussion of the underlying linguistic framework motivating our approach. Section 4 of this paper is dedicated to a preliminary proposal for Lexicon induction from semantic primitives and grammatical categorization. In this proposal we will point to shortcomings of FN and address them by referring to the Italian lexicon we built in the past.
We will assume that in order to construct a lexicon containing the information listed above, we should address the following linguistic categories:

- grammatical categories – derived from a categorization of reality into entities – nouns –, events – verbs and nominals -, attributes – adjectives, adverbials, and nouns –
- semantic categories, like negation, quantifiers;
- discourse level categories, like deictics, definiteness, conjunctions for coordination and subordination at propositional level;
- syntactic categories - encoding the arity of predicate-argument structures as they are interpreted in situations;
- aspectual categories – encoding the internal temporal structure of events (as expressed both by verbal and deverbal nominals);
- semantic conceptual categories – classifying types of events in relation to the (un)reality they encode;
- selectional restrictions – encoding the typicality of event participants in inherent semantic features as they are represented in an ontology or connected encyclopedic database of entities and their semantic interrelationships;
- grammatical constraints – encoding socalled syntactic and anaphoric binding constraints on arguments of predicate and dependent predicates only for propositional arguments, though.

This set of primitives has been used for lexical induction in a program that automatically produces full-fledged lexical representations in Italian. In Delmonte(1989) - but see also Carrier et al. 1993 - we presented a mapping algorithm that starting from a fixed number of classes and a syntactic encoding of argument structure could derive automatically via a certain number of linguistic rules, both grammatical function, semantic roles and conceptual representations. In conclusion, starting from FrameNet and its companion lexical resources available online, we want to test hypotheses on the nature of lexical knowledge acquisition and its function in language understanding and generation, by simulating its creation and usage at a computational level.

2. Frames and Syntactic-Semantic Representations

FrameNet is designed as an ontology of frames, i.e. representations of prototypical situations. Each frame provides a set of predicates (nouns, verbs or adjectives) by which it can be evoked and a set of semantic roles – or frame elements - which correspond to categories of entities or concepts that occur in the situation. Like other projects, FrameNet has predominantly concentrated on building a large manually annotated corpus. The corpus, a subset of the British National Corpus, currently contains about 135.000 instances of 795 frames. However, for role semantics to become relevant for language technology, robust and accurate methods for automatic semantic role assignment are needed. In recent years, a number of studies has investigated this task on the FrameNet corpus (Gildea and Jurafsky, 2002; Fleischman et al., 2003; Chen and Rambow, 2003). For each frame, the following information is given:
1. the name of frame,
2. a description (usually including example sentences),
3. a list of the lexical units (LUs) in the frame (if any), and
4. the names and descriptions of each frame element (FE), usually including example sentences.
5. Example sentences which are either taken from the Brown Corpus or the Penn Treebank and come with POS, Syntactic constituency labels and Grammatical Functions.

In turn Frame Elements

- may belong to the Core or the Periphery;
- they may be Extrathematic, Core-Unexpressed;

Of all the FEs or Semantic roles associated to example sentences, limited though to verb LUs, we have the following distribution in terms of Core, Periphery and Extrathematic:

- 61390 roles are core
- 6560 roles are extrathematic
- 16100 roles are peripheric
- finally, 1091 verbal entries are empty (there are no examples associated)

Semantic Roles which appear in all three types are the following ones,

ADDRESSEE, CARRIER, CAUSE, CONTAINER, DIRECTION, DISTANCE, EVIDENCE,, GOAL, INITIAL_STATE, INSTRUMENT, LOCATION, MEANS, MEDIUM, OUTCOME, PATH, PATH_SHAPE, PURPOSE, ROLE, SOURCE, STATE, VEHICLE

Semantic Roles which appear in only two of such types – either core/extrathematic, or core/periphery, or periphery/extrathematic – are the following ones,

Area, Category, Characterization, Charges, Circumstances, Components, Configuration, Containing_Event, Cotheme, Court, Degree, Depictive, Difference, Dimension, Duration, Event, Experience, External_Cause, Feature, Final_Category, Final_Value, Firearm, Form, Goods, Group, Handle, Initial_Category, Intermediary, Internal_Cause, Judge, Legal_Basis, Lessor, Locus, Manipulator, Manner, Material, Message, Mode_Of_Transportation, Money, Occasion, Parameter, Perceiver_Passive, Place, Point_Of_Contact, Purpose_Of_Theme, Rate, Reason, Resource, Result, Resulting_Action, Seller, Sleep_State,Speaker, Standard, Support, Time, Topic

Finally Semantic Roles which may only appear in one such syntactic type are the following ones where we divide Core/C from Extrathematic/E and Periphery/P:

C
Arguer, Authority, Body_Part, Cognizer, Entity, Impactor, Interlocutor, Item, Part, Participant, Partner, Party, Performer, Phenomenon, Side, Theme
E
Beneficiary, Concessive, Duration_Of_End/Final_State, Iteration, Ground, Recipient, Subregion
P
Criteria, Value

FN is accompanied by examples which are represented as constituent structures. Constituency labels used in FN are the following:

2nd, 3rd, a, ajp, apos, avp, cni, dni, inc, ini, n, np, poss, pp, pping, ppinterrog, quo, sabs, sbrst, sfin, sforto, sing, sinterrog, srel, sto, sub, swhether, unknown, vpbrst, vpfin, vping, vpto

where in some cases we see the use of a lexical category (a, n, poss). Grammatical Functions labels are only 4,

appositive, DEP, EXT, OBJ

and they may have Indefinite Null Instantiation (INI), Definite Null Instantion (DNI) and Constructional Null Instantiation (CNI). Constructional Null Instantiation, CNI is used to mark omitted subjects in passives, imperatives, gerunds, infinitives and objects of some imperatives. DNI is used to mark arguments omitted when they are presupposed as known. INI is used with existential cases with intransitivized objects etc. Inchoative uses of a LU is distinguished from the causative use. But both passive, middle and ergative uses are not present.
In addition to that, FN has built a network of hierarchical Frame-to-Frame relations with super frames called “scenarios” – currently 23 have been created - to allow for inferencing to take place. They are listed here below:

Inherits From:, Is Inherited By:, Subframe of:, Has Subframes:, Precedes:, Is Preceded by:, Uses:, Is Used By:, Perspective on:, Is perspectivized in:, Is Causative of:, See Also:

However not all the inferential functions are filled as they should be in a real ontology. The highest or top level of the hierarchy is constituted by Scenarios, which may be dubbed as complex Frames and connect together a certain number of Frames. Semantic inheritance relations are encoded through the two basic relations, “Inherits From” and “Is Inherited By”. The currently encoded Scenarios are the following one,

Attempting_scenario, Causation_scenario, Change_of_phase_scenario, Change_of_state_scenario, Commerce_scenario, Crime_scenario, Employee's_scenario, Employer's_scenario, Employment_scenario, Getting_scenario, Giving_scenario, Knot_creation_scenario, Lose_possession_scenario, Measure_scenario, Motion_scenario, Obligation_scenario, Receiving_scenario, Requirement_scenario, Resolve_attempt_scenario, Safety_scenario, Searching_scenario, Shooting_scenario, Transfer_scenario

There are 107 Frames directly connected to the Scenarios. Producing a handcrafted highly reliable deep lexical representation as FN is very labor-intensive and time-consuming. The number of Frames of the current release has increased from version 1.2 by a 62% - from 488 to 795 – thus indicating that the possibility to cover all facets of reality by producing frames and linking them to the evoking predicates and the example sentences is an effort which gives slow but steadily increasing results. However this is still to be regarded as a small inventory of all possible frames and the FN team itself does not expect to have a full inventory of frames until a substantial proportion of the general-purpose vocabulary of English has been analyzed.

FN constitutes the last and more perspicuous level of representation of knowledge of the world and the least generalizing one, on top of strictly domain-related pieces of reality expressed in natural language. The frames are organized into hierarchies where the more specific frames inherit some properties from the more general ones. In general, some semantic areas are covered only by a general frame, some others by a combination of specific frames and some by a combination of specific and general frames.

As said above, the current version of FN is useful for a certain number of different text processing tasks, like using Frames and Frame Elements to produce more perspicuous semantic roles associated to the output of a dependency parser. Eventually there is always the possibility to extend it manually – if needed - to other lexical units which share/evoke the same concept/frame. However this task, when based completely on FN and its set of examples, is poorly executed due to the small number of lexical units listed for each Frame (see Pado & Boleda, 2004). The same researchers remark the lack of uniformity in the argument structures associated to each Frame, which adds up to the poor predictability that can be associated to the evoking predicates of a given frame. Other problems related to FN are due to the lack of a one-to-one mapping between evoking LUs, their syntactic realization in example sentences and semantic roles (Frame Elements). In fact, on the one side FN proposes a general scheme for frame-evoking LUs by associating Frame Elements to portions of text – actual words - in example sentences; on the other side it tries to reconcile generality with specificity by increasing the number of example sentences. This should cover the problem of lack of selectional restrictions associated to Frame Elements. Example sentences are however only ancillary or subsidiary pieces of information in the overall theoretical framework: their syntactic structure, grammatical relations and grammatical categories play no restrictive role on the way in which Frame Elements – for instance – have been encoded in the corresponding frame evoked by the governing predicate.

In this respect, one of the basic problems presented by the use of examples for the purpose of machine or automatic learning of semantic roles, is the presence of a great number of Null elements as indicated by the grammatical representation, which makes the mapping from Frame Elements to actual sentences and viceversa, poorly executable and consequently poorly predictive/predictable. Null elements are the result both of syntactic Alternations, in the sense of Levin, and syntactic transformation or derivation. Most common structures are,

- Passivized structures where the Agent has been deleted;
- Intransitivized structures where the Object has been left unexpressed and the subject is an agent;
- Middle structures where the Object has been left unexpressed and the sentence is non factual;
- Inchoativized structures are represented in the FEs;
- Ergativized structures where the Object has been raised to Subject and the deep subject is left unexpressed.

As Green et al. remark in their paper, using FrameNet for open texts processing reveals its limits as far as coverage is concerned: low recall affects both the number of semantic frames, but also the frame-evoking capacity of each list of verb units, which is very low. No verbs are listed for over 30% of all frames, while another 10% or so list only 1 or 2 verbs. This must be regarded as a secondary or side-effect of the overall theoretical framework, which takes Lexical Units and not the Predicate Argument Structure or eventually the sentence in which LUs are contained as the primary source of lexical knowledge representation.

2.1 Lexical Resources and Ontologies

To cope with the problem of sparcity, a number of other lexical resources have been proposed in the past and others are being proposed now. We start by commenting the paper by Green R. et al.(2005) where the authors comment in their Conclusions that "... sets of verbs evoking a common semantic frame can be induced from existing lexical tools." However as the authors have to admit, the lexical resources available on English suffer from the problem of sparseness which affects both Recall and Precision in the last resort.

Many recent annotation efforts for English have focused on pieces of the larger problem of semantic annotation, rather than producing a single unified representation like the Prague Dependency Tectogramatical Representation (Hajicova & Kucerova, 2002).

Other important resources are the following ones, which will briefly comment below:

- PropBank (Palmer et al, 2005) annotates predicate argument structure anchored by verbs.
- NomBank (Meyers, et. al., 2004a) annotates predicate argument structure anchored by nouns.
- TimeBank (Pustejovsky et al, 2003) annotates the temporal features of propositions and the temporal relations between propositions.
- The Penn Discourse Treebank (Miltsakaki et al 2004a/b) treats discourse connectives as predicates and the sentences being joined as arguments.

2.1.1 PropBank: The Penn Proposition Bank focuses on the argument structure of verbs, and provides a corpus annotated with semantic roles, including participants traditionally viewed as arguments and adjuncts. An important goal was providing consistent semantic role labels across different syntactic realizations of the same verb, as in *the window* in *[ARG0 John] broke [ARG1 the window]* and *[ARG1 The window] broke*. Arg0 and Arg1 are used rather than the more traditional Agent and Patient to keep the annotation as theory-neutral as possible, and to facilitate mapping to richer representations. Finally, the corpus contains 5 different types of Arg. Coarse-grained sense tags, based on groupings of WordNet senses, are being added, as well as links from the argument labels in the Frames Files to FrameNet frame elements.

2.1.2 NomBank: The NYU NomBank project can be considered part of the larger PropBank effort and is designed to provide argument structure for instances of about 5000 common nouns in the Penn Treebank II corpus (Meyers, et. al., 2004a). PropBank argument types and related verb Frames Files are used to provide a commonality of annotation.

2.1.3 TimeBank: The Brandeis TimeBank corpus focuses on the annotation of all major aspects in natural language text associated with temporal and event information (Day, et al, 2003, Pustejovsky, et al, 2004). Specifically, this involves three areas of the annotation: temporal expressions, event-denoting expressions, and the links that express either an anchoring of an event to a time or an ordering of one event relative to another. Identifying events and their temporal anchorings is a critical aspect of reasoning, and without a robust ability to identify and extract events and their temporal anchoring from a text, the real aboutness of a text can be missed. The core of TimeBank is a set of 200 news reports documents, consisting of WSJ, DUC, and ACE articles, each annotated to TimeML 1.2 specification. It is currently being extended to AQUAINT articles. The corpus is available from the timeml.org website.

2.1.4 WordNet is a handcrafted lexical database that is based on the hyperonymy-hyponymy relation – but comprises also other semantic relations like antonymy, meronymy, a causal relation, etc. – to shape the structure of an English lexicon and build a lexical semantic network where each entry is organized as a synset, i.e. a list of close synonyms. Each entry is also accompanied by a gloss which however, contrary to what happens in LDOCE, uses a non-restricted vocabulary. The use of a restricted vocabulary would enable its computability in terms of word sense disambiguation. The same problem affects another important lexical resource, FrameNet, which is enriched with glosses which however don't use a restricted vocabulary.

2.1.5 Levin's Verb Classes. According to the hypothesis stated in (Levin, 1993), syntactic features of verbs are semantically determined and thus syntactic behavior of verbs can be used for their semantic classification. Levin describes syntactic behavior of verbs with respect to possible syntactic alternations and semantic classes are built from verbs that undergo a certain number of alternations. An alternation means a change in the surface realization of the argument structure of a verb. Levin’s list contains 258 semantic verb classes.

2.1.6 The LCS database (Dorr, 2001) was designed as a semantic representation of predicates and propositions. It describes the semantics of verbs as a combination of semantic structure and semantic content – semantic structure is characteristic for all verbs from one semantic group whereas particular verbs can differ in their semantic content. The lexical item is an oriented rooted graph that bears information on its subject, its objects (arguments) and its 'modificators' and on their obligatoriness / optionality. In addition, their thematic roles are stated as well as restrictions on conceptual categories (also called conceptual POS, as e.g. 'thing', 'event', 'state', 'place', 'purpose', 'manner', 'time').
LCS uses the following 12 THETA-ROLES labels for Logical Arguments:
AG, EXP, TH, SRC/SRC(), GOAL/GOAL(), INFO, PERC/PERC(), PRED/PRED(), LOC/LOC(), POSS, TIME/TIME(), PROP
and the following 7 role labels for Logical Modifiers:
MOD-POSS, BEN, INSTR, PURP(), MOD-LOC, MANNER(), MOD-PROP, PARTICLE
Apart from Purpose and Manner all other roles modifers are attested both with and without preposition. Particle indicates other particle handled via collocations. The presence of () indicates that the roles can be preceded by preposition. Together they amount to 19 different labels. However when composed together with preposition and obligatoriness or optionality marked by comma or underscore, we reach the number 250 different Thematic Grids, which have been mapped to the theta roles of PropBank.

2.2 Lexicalized Causal Meaning: LCS and FrameNet

As noted above, one of the problems that faces FN lexical representations is the lack of a direct mapping into Levin’s classes. Syntactic lexical alternations are very useful to detect Frame Elements which are not primitives but are related to passivized, intransitivized, or inchoativized structures derived from an underlying basic lexical entry. Causal meaning is however encoded in FN. Clear cases of this situation include causative-inchoative pairs which is encoded in the hierarchical link “Is Causative of”. So although causative-inchoative pairs are not in the same frame, the FrameNet database provides an explicit link between the paired frames via a frame-to-frame relation Causative of.
Generally speaking, causes need causative verbs to be expressed or else discourse markers. From a conceptual point of view, all predicates belonging to the following aspectual classes may contain an abstract CAUSE operator:

- accomplishments
- non reversible accomplishments
- gradual accomplishments
- achievements
- punctual achievements

Predicates belonging to these classes may also be regarded as containing two additional features:

- intentionality
- animacy

However, it may be proven that Causative Relations may be non-intentional and have a non-animate causer as it would happen with all natural events. FN does not provide aspectual information for single LUs but only what can be referred to Frames in its entirity, i.e. relations intervening between the arguments (optional and obligatory ones) of the governing predicates. In this sense, Causality in FE is only a property of the overall Frame. However, from a computational point of view the lexical semantic notion of causality is very important to help define Causality Relations in sentences and texts.
In this perspective, we follow Ray Jackendoff by postulating the existence of a certain number of primitive lexical operators among which is the CAUSE operator. For instance, the sentence,

- John built a house

may be conceptually decomposed and represented into the following representation,

CAUSE(John-SUBJ,
BE-IN-THE-WORLD/EXISTENCE(house-OBJ),
$Evs_{i,j}$ (Ev_i …Ev_j))

Which can be paraphrased as follows,
John caused the house to be in existence/world in the time span intervening between Ev_i and Ev_j, where the final event represent the Accomplished Result State. More on these topics below.
Predicates belonging to STATE or ACTIVITY cannot be used to express Causes and consequently cannot be decomposed into the lexicalized conceptual causal operator. On the contrary they may be regarded the preferred target of RESULT clauses. In the LCS lexicon, on the total number of lexical items present, which amount to 9000 entries, over 5000 have a CAUSE operator incorporated in their lexical meaning representation.
FrameNet uses the following FRAMEs related to Causation to classify verb predicates:

- *Causation*, which is used for Verbs, Nouns and one Adjective, 'causative';
- then another 26 Frames which are used to define Verbs and Nouns:

Cause_begin_motion, Cause_change, Cause_change_of_consistency, Cause_change_of_phase, Cause_change_of_position_on_a_scale, Cause_confinement, Cause_expansion, Cause_fluidic_motion, Cause_harm, Cause_impact, Cause_motion, Cause_of_shine, Cause_temperature_change, Cause_to_amalgamate, Cause_to_be_dry, Cause_to_be_sharp, Cause_to_be_wet, Cause_to_continue, Cause_to_end, Cause_to_experience, Cause_to_fragment, Cause_to_make_noise, Cause_to_make_progress, Cause_to_move_in_place, Cause_to_resume, Cause_to_start, Cause_to_wake

All predicates thus described – evoked by each frame - will have a Causer or an Agent as Subject. The number of predicates thus classified in FrameNet is however very small, less than 548 over 10195 examples, 349 of which are constituted by verb predicates.
LCS entries contain cross reference to Levin verb classes, to WordNet sense, to PropBank argument list which have been mapped to a more explicit label set of Semantic Roles which can be regarded more linguistically motivated than the ones contained in FrameNet which are more pragmatically motivated. However, for our purposes, LCS notation is more perspicuous because of the presence of the CAUSE operator, and is more general: on a total number of 9810 lexical entries – and 4868 different linguistic forms -, 5000 contain the CAUSE operator. On the contrary, in FrameNet on a total number of 10000 lexical entries, only 333 are related to a CAUSE Frame; if we search for the word "cause" in the definitions contained in all the Frames the number increases to 789 but is still too small compared to LCS.
Here below is one example of LCS entries:

```
(:DEF_WORD "prevent"
 :CLASS "059"
 :WN_SENSE (("1.5" 01387332)
        ("1.6" 01669882))
 :PROPBANK ("arg0 arg1 arg2(from)")
 :THETA_ROLES ((1 "_ag_th_prop(from)"))
 :LCS (cause (* thing 1)
     (go circ (* thing 2)
    ((from 3) circ (thing 2) (at circ (thing 2) (* nil 27))))
     (prevent+ingly 26))
 :VAR_SPEC ((27 (aspect prog))) )
```

To recover the same information from FrameNet, either the definition or the Frame Elements had to be inspected. In particular, Prevent is part of the Frame PREVENT which has Preventing_Cause as Frame Element. "Purge" is part of the Frame Removing and in order to know that it is a causative verb one needs to read the definition which paraphrases the content of the Frame Elements by saying that The Agent is a person (or other force) that causes the Theme to move.

3. From FrameNet to a Linguistically-based Fully Subcategorized lexicon

As said above, we intend to test hypotheses on the nature of lexical knowledge and lexicon acquisition. We assume that the representation of lexical knowledge in a system for text understanding is equivalent to that present in the mind of an individual, i.e. it is psycholinguistically relatable to the one possessed by a speaker of the language. As said above, differently from what is assumed by FN theory, where Frames are evoked by LUs, we believe that Frame or frame-like representations and more generally any lexical representation at all, can only be derived from sentence or propositional level interpretation. It is thus a type of knowledege which is not merely word-related but profoundly structurally-related both syntactically and semantically.

There are two basic components or subtheories onto which our lexical representations are built: one is the syntactic-semantic component, the other is the semantic-conceptual component. We will address them separately in the subsections below.

3.1 Linguistic Theories and Computational Issues

The questions we pose ourselves in implementing the programs needed to simulate lexical knowledge and acquisition are approximately the following ones: what is the set of primitives required to build lexical descriptions which we call forms - in accordance with LFG - which make available to native speakers both syntactic and semantic information required to parse and recognize sentences of their mother tongue in the acquisition phase. Also, since we believe that low level inferences are made unconsciously by speakers, what information are required in order to automatically build a conceptual representation to be used by the inferencing mechanism?

In this sense, lexical forms are tightly connected to the nature of the human and computational parser and the inference mechanism which uses them. However, we believe the format of lexical forms should be universal and internally motivated, rather than externally conditioned.

Chomsky's projection principle:

> "Representations at each syntactic level (LF, and D- and S-structure) are projected from the lexicon, in that they observe the subcategorisation properties of lexical items." (1981: 29, see also Chomsky 1986).

The initial hypotheses verify whether Grammatical Functions rather than Semantic Roles can be taken as primitives, or if the opposite applies (see Wilkins, 1988). The two hypotheses are derived from Chomsky's and Bresnan's theories: both take for granted the existence of syntactic categories and an X-bar system to build main constituents: however, in Chomsky's system thematic roles are mapped directly on to s-structures which are specified in terms of syntactic constituents, whereas in Bresnan's model thematic (or rather semantic) roles are only subsidiary on f-structures which act as interface onto c(onstituent)-structures.

According to both theories, words, when they are first heard by the child, are categorized into lexical categories and concatenated into major constituents to form the first structural representation, onto which lexical forms are projected in order to proceed to the semantic interpretation. Whereas Bresnan introduces grammatical functions as a restriction on c-structure wellformedness and as an interface to lexical projections made up both of grammatical functions and semantic roles, Chomsky uses an enriched notion of c-structure, i.e. S-structure. This is meant to capture the same generalizations of Bresnan's two levels: annotated c-structure and f-structure, through the principle-and-parameter approach. In addition, Chomsky introduces the level of D-structure, which is "...directly associated with the lexicon. It is a 'pure' representation of theta-structure, expressing theta relations through the medium of the X-bar-theoretic conditions in accordance with the projection principle"(1988:2).

Furthermore, a lexical entry must specify "enough of its properties to determine its sound, meaning, and syntactic roles... it should not contain redundant information, for example, about the quality of the vowel, properties of action verbs generally, or the fact that together with its complement it forms a VP."(ibid.,2). In Chomsky's terms then, the lexicon is separated from D-structure, and both are related to S-structure by means of iterated application of Affect(Move)-Alpha. This segmented view or modular view of the grammar is completely opposite to the unitary view proposed in Bresnan, in which the lexicon is the repository of all information relating c-structure to its interpretation. The lexical form of an item includes its phonetic, syntactic and semantic properties all in one single format, the only exception being constituted by derived forms (the Affect-Alpha application) which are built once and for all by Lexical Redundancy Rules and as such are listed in the lexicon in the same form as the underived or original one, thus receiving the same interpretation in terms of predicate argument structure and semantic roles assignment.

In Movement theories then, in order to address a level of representation like D-structure or LF-structure one must go through the interaction with S-structure, or phrase structure representation; on the contrary, in LFG, c-structure is separated from f-structure and both are separated from lexical structure and may be addressed separately. Also, Thematic Structure is separated from syntactic functions and both are separated from discourse functions. Syntactic functions are associated with semantic roles and discourse functions according to principles of Lexical Mapping(see Bresnan & Kanerva, 1989:23-28). Subject and object correspond virtually to any semantic role and may even be nonthematic; oblique arguments, on the contrary, have fixed semantic roles, as Bresnan & Kanerva note(ibid.,25), implying by this that there must be some additional information which is contextual in kind: in particular, this may be marked by the preposition used, obviously when no ambiguity may arise - which as we shall see in more detail below is far from unfrequent. The only fixed fact is the possibility to classify certain semantic roles relatively to the coupling with syntactic function: for instance, Agent may never be [+objective], like Locative. However, the main difference between the two theories lies in the fact that in Chomsky's theory theta-roles are syntactically relevant elements; on the contrary, in Bresnan's framework, semantic roles are just semantic objects related to Predicate Argument Structure, or simply Argument Structure.

We shall follow the second position. In our system, Semantic Roles are partly derived from a mapping with Grammatical Function, as in Bresnan's system, and partly reflect the wealth of information contained in semantic and conceptual classes which are associated to each verbal predicate. These classes contain information about the meaning of the verb, meaning which can be represented partly in terms of Semantic Roles and partly in terms of Conceptual Representations, as we shall see in more detail below. Even though we think it sensible to assume that it is impossible to transpose all world or encyclopaedic knowledge into classes or roles, we believe it a scientifically interesting hypothesis trying to define these levels of representation within a global system of text understanding in order to see what kind of information must be made available to the semantic interpreter. In this sense, we think FN can be deemed the most important attemp in that direction.

3.2 From Linguistic Theory to Conceptual Representation

In accordance with the principle of meaning decomposition, we assume that concepts denoted by lexical items are made up of primitive concepts which can be expressed by the use of a very limited number of templates. The granularity of the description depends strictly on the (sub)domain and the aim of the task at hand. For instance, abstract concepts like “responsible” or “responsibility” when dealt with in a legal subdomain require a specification of preconditions which is different from what is expected in a generic domain.

A method for the decomposition of lexical information should represent a principled way to organize a taxonomy of the concepts in a language, categorized by sets of features, which however are tightly interleaved with argument structure and the syntactic nature of each argument.

Basic constituents for our conceptual representations are spatial primitives on the basis of analogical relations existing in the spatiotemporal realm which is at the heart of the meaning of all verbs and deverbal nouns and adjectives. According to Jackendoff and Gruber, human beings seem to base their descriptions of any kind of experience on some crucial concepts drawn directly from what might be called the spatial semantic field. Similarly, temporal sequence is both perceived and expressed on the model of

spatial sequence. Events and states are located in time - on a timeline - just as things and entities are located in space. The same prepositions are used both for spatial and temporal expressions.
In theory, it should be possible to describe the basic conceptual components of meaning of any verb given a finite number of spatial primitives and of modalities attached to them. Modalities describe a bit/portion of meaning of a lexical item when decomposed into conceptual primitive functions, and adds to them a certain modality. This is not to be confused with negation, which is itself an operator preceding and having scope over conceptual functions, as implied by the meaning of lexical items – more on this below.
When analysing utterances, we are interested in interpreting their meaning: this implies three basic things,
A. knowing if the meaning has to be related to the actual world or not, and if so, where should it pick up its reference from;
B. knowing what consequences the action or state of affairs described by the predicate-argument structure of the predicate(s) appearing in the utterance brought about, i.e. what can be inferenced. This is partly contained in the conceptual decomposition; and partly extractable from aspectual classes, tense, and the presence of time adverbials, as we already saw;
C. knowing the internal structure of temporal representation which varies basically according to lexical representation, but as we saw, can undergo dramatic structural modification in case of the intervention of surface level constituents of meaning, since we know that temporal interpretation must be built compositionally.
As for the conceptual level of representation, we take Jackendoff's theoretical framework as the basis of our work. However in our framework, conceptual representations are not primitives, they are derivative on low level representations and depend heavily on the semantic/pragmatic class of the verb. As we saw, also theta or semantic roles are derivative in our system.
To discuss this point briefly, let us consider a few predicates. Copulative verbs like BE and HAVE in our lexicon are assigned different lexical forms in case they may appear in different structural contexts: BE may be inserted or may appear with an open complement, a generic XCOMP, or it may have an infinitive as SUBJect/prop. On the contrary, the verb HAVE only appears with NCOMP/prop.
The result of this choice is that there is no way to map the semantic and conceptual interpretation directly onto a specific lexical form, these being dependent on the kind of lexical complement analysed in a particular sentence. For instance, the existential-locative meaning of the verb BE would have to be mapped once a PCOMP with locative meaning is analysed: however, this meaning does not depend on the presence of a particular preposition alone. It is both geared on selectional restrictions on the head noun and the preposition. In "John is in a hurry" we just want to say that John is an EXPeriencer and that there is a STATE in which he is in, and this state is a PCOMP whose head is "hurry". This would correspond to the Italian counterpart which uses HAVE rather than BE, "Gino ha fretta", and an NCOMP rather than a PCOMP.
Thus, there is no way to map the "possessive meaning" of HAVE onto a specific lexical form, since this meaning depends strictly speaking on the kind of nominal heads being analysed in the SUBJect and the NCOMP. Thus "John has a fever" means again that John is EXPeriencing a particular STATE; on the contrary, "John has a car" will have to be computed with John as OWNER and the NCOMP as indicating POSSession.
At the same time, we want to be able to map different conceptual representation onto separate syntactic lexical forms in case it is needed: take for instance, the Italian "chiedere" for "ask" which has only one underlying lexical form made up of the same combination of functional labels but with different semantic roles:

i. pred(chiedere [LET(<address>(GO(REP(FROM<informtn><sourceA>))])
ii. pred(chiedere [CAUSE(<agent>(REP(GO-POSSinf(FROM<theme>(GO-circ(FROM <ownerA>))])

where we see in i. the conceptual structure for ask/inquire for information and in ii. the structure for ask as demand or request someone to be allowed or given. These representations derive from the corresponding syntactic representations where we see that no difference exists at the level of constituency and functional labels:

i. pred(chiedere,tr,achiev,inform,[np/subj1/addres/[+hum],np/obj1/informtn/[+abst],obj2/source/a/[]])
ii. pred(chiedere,tr,achiev,ask_poss[np/subj1/agent/[+hum],np/obj1/theme_affect/[-ani],obj2/owner/ a/[]])

Thus semantic role and ontological categories are determined by the semantics of the head, its aspectual class and selectional restrictions which in turn depend on syntactic and semantic classes of the predicate.
Conceptual representations are of paramount importance when we consider the case of predicates containing negation or some other modality, like "difficult" in a predicate like MANAGE, or else "guilt" in MURDER. Predicates belonging to the class of "performatives" may be decomposed into primitives which involve the assignment of properties as the main meaning of the verb. To this aim, consider a predicate like MARRY, which could be assigned a conceptual representation as the following one,

pred(marry, [CAUSE(<actor>(GO[to_exist(Property(x))] (PATH-TO-IDENT <actee> (GO [to_exist (Property(y))])))) {[(Gender=mas, Property=husband ; Gender=fem, Property=wife), Property(x) ≠ Property(y)])

In other words, to marry someone corresponds to cause a new property to be existent in the world and to associate this property to the actor or actee of the event according to their gender. Consequently, a predicate like DIVORCE would inherit from process verbs like END its basic meaning, i.e. to divorce a wife or husband means to put an end to a property, basically:

pred(divorce, [CAUSE(<actor>(GO[to_end(Property(x))] (PATH-TO-IDENT<actee> (GO[to_end (Property(y))])))) {[(Gender=mas, Property=husband ; Gender=fem, Property=wife), Property(x) ≠ Property(y)])

Consider also predicates like PREVENT, HINDER, PROHIBIT, FORBID, BAN, DESTROY, END, LACK, INTERRUPT, BEGIN which contain information related to negation and permission rather than simply causation, a semantic component which is important in the semantic interpretation of the sentence. We associate the following descriptions to some of these predicates,

pred(distruggere[CAUSE(<agent>(GO[to-nonexist]<theme_aff>)))])/ destroy
pred(incominciare[GO-CIRC(PATH TO-CIRC(CAUSE(<agent>(GO-circ[to-exist]<propA>))]) / begin
pred(interrompere[CAUSE-abort(<agent>(STAY-circ(AT<theme>)))]) / interrupt
pred(impedire[CAUSE(<agent>NOT(GOexten[to-exist]<theme>(GO-circ(TO<goalA>)))))]) / prevent
pred(ostacolare[CAUSE(<agent>NOT(LET(GO-circ<theme>)))])/hinder
pred(mancare[NOT(BE(<theme_bound>(PLACE-ATposit<ncomp>)))])/ lack

3.3 Conceptual Representations and World Knowledge

Another fundamental property of our conceptual representations is the event structure which is automatically generated from aspectual information. There are basically three types of event structure:

- **states**, represented as $\{(e_n,t_n)\}$
- **activities,** represented as $\{(e_1,t_n))\}$
- **achievements,** represented as $\{(e_1,t_1))\}$
- **gradual accomplishments,** represented as $\{(e_1,t_n),(e_2,t_2)\}$
- **accomplishments,** represented as $\{(e_1,t_1), e_1,t_n,(e_2,t_2)\}$

Each representation contains two indices, one E for subevent time, and another T for the actual temporal interval in which the current event takes place. The subscript may indicate: beginning subevent, 1 and ending subevent, 2 into which the overall event described by the lexical entry is subdivided. The presence of a N subscript indicates openness or lack of temporal closure of the event: this applies to all types of actions apart from Achievements which are punctual and have no temporal extension whatsoever.
The full inventory of conceptual primitives produced for our lexicon is included here below and is represented in the table 1. where we also associate one conceptual representation and some example verbs:

exten, subjct, hyper, manip, factive, evaluat, proprty, ment_act, process, measu_maj, percpt, perfect(end/exist), perform, inform(at/to), possess(to), possess(from), inform(at/from), measu_min, posit(at/to), react, posit/origin(from), exten_neg, let, coerc, ask_poss, at_posit, not_exten, not_let, dir, touch, divide, hold, hole(into), dir_difclt, unite, go_against, ingest, perform_to, rep_contr, not_react, dir_tow, over, go_through, color, quantity, follow.

TABLE 1. SEMANTIC CLASSES AND CONCEPTUAL REPRESENTATIONS

0 = exten	(STAY/GO/IDENTexten(AT/TO	lavorare
1 = subjct	(STAY!GO/IDENTin_subj_mind(REP	credere
2 = hyper	(ORIENThyper(TOWARD	fingere
3 = manip	(STAY/GO/IDENTcirc(AT/TO	minacciare
4 = factive	(IDENTfactv(REP(TR	sapere
5 = evaluat	(ORIENTeval(TOWARD	piacere
6 = proprty	(STAY/GO/IDENTpropr(AT/TOWARD	divenire,assomigliare
7 = ment_act	(STAY/GO/IDENTcirc(AT/TO	dimenticare,comprendere
8 = process	(STAY/GO/IDENTcirc(REP(AT/FROM	continuare
	(GOcirc(TO[exist]/[end]	finire, interrompere
9 = measu_maj	(STAY/GO/IDENTmeas(AT/TO-[major]	crescere
10 = percpt	(STAY/GO/IDENT(REP(AT/TO	mostrare,esibire
11 = perfect	(GO(TO[end] - (GO(TO[exist]	finire, creare
12 = perform	(GOcirc[perf]	battezzare
13 = percpt	(REP(STAY/GO/IDENT(AT/TO	ascoltare, udire
14 = inform	(STAY/GO/IDENT(REP(AT/TO	capire, comunicare
15 = possess	(STAY/GOposs(AT/TO	dare, vendere, possedere
16 = possess	(GOposs(FROM	comprare, ricevere
17 = inform	(STAY/GO/IDENT(REP(AT/FROM	domandare, rispondere, sapere
18 = measu_min	(STAY/GO/IDENTmeas(AT/TO-[minor]	diminuire
19 = posit	(STAY/GOposit(AT/TO	andare, stare
20 = react	(STAY/GOreact(AT/TO	resistere, contraddire
21 = posit	(GOposit(FROM	venire, arrivare
22 = exten_neg	(GO(TO[nonexist]	distruggere, uccidere
23 = let	(LET(GOcirc	permettere, aiutare
24 = coerc	(GOcoerc(TO[exist]	costringere
25 = ask_poss	(REP(GOinf(FROM	chiedere
26 = at_posit	(STAY(AT	essere - locativo
27 = not_exten	(NOT(GOexten(TO[exist]	mancare, escludere
28 = not_let	(NOT(LET(GOcirc	ostacolare, vietare
29 = dir	(GO(AFTER	seguire, corteggiare
30 = touch	(STAY/GOtouch(AT/TO	toccare,accarezzare
31 = divide	(GOsegmnt(TO	separare
32 = hold	(STAY/GOhold(AT/TO	tenere,acchiappare
33 = hole	(GOhole(INTO	perforare, bucare
34 = dir_dfct	(GO(AFTER[difclt]	inseguire
35 = unit	(STAY/GOunite(AT/TO	unire
36 = go_against	(GO(AGAINST	attaccare assalire

37 = ingest	(GOingest(INTO	ingerire, ingoiare
38 = perform_to	(GOperf(TO	incaricare
39 = rep_contr	(REP(STAY/GO/IDENTcontr(AT/TO	controllare
40 = not_react	(NOT(GO/STAYreact(TO -	cedere,subire
41 = dir_tow	(GO(TOWARD	dirigere, deviare
42 = go_over	(GO(OVER	sorpassare, vincere
43 = go_through	(GOexten(FROM(TO	andare, tradurre
44 = colour	(GOexten(FROM(TO	dipingere
45 = quantity	(GOexten(FROM(TO	andare, tradurre
46 = follow	(GOexten(FROM(TO	andare, tradurre

4. The Computational Lexicon of Italian

Work on a Lexicon of Italian – the LIFVE - started in the '80s thanks to a contract with DIGITAL Eq. to produce a TTS for unrestricted text. At first we actually produced a fully subcategorized lexicon of Italian with reference to most frequent 5000 lemmata as registered in current FLI. Followed by the first Treebank worldwide on the basis of 40,000 word corpus, BUT fully manual. Subsequently, from the beginning of the '90s we worked on a more extended version of the LIFVE. We classified a list of 18,000 Italian verbs entries (approximately 10,000 different lemmata, and 5,000 different roots) for computational purposes, producing a highly fine-grained categorization of syntactic and semantic features of arguments of the predicate. At a national level, an initiative called The Ilex project was started by three research centres ITC-IRST, University of Tourin, University of Venice and the goals were formulated in a very ambitious way, by setting ourselves the following incremental sizes of lemmata: 1000 lemmata end of 1996; 10000 lemmata end of 1997 ; 30000 lemmata end of 1998.
Lexical Information for ILEX was constituted by POS and syntactic-semantic class information; morphological and phonological information; semantic information. However, work stopped after the first year, 1995. Each research unit, we too, continued working on the peculiarities of Italian verb classes.

4.1 LIFVE - a lexical transducer

With LIFVE we wanted to produce a lexical encoder, a fully automatic annotation tool, which allowed the linguist annotator to easily encode all syntactic and semantic aspects of lexical knowledge. In particular, we organized the algorithm into a sequence of operations with the following properties:
- Hierachy of syntactic-semantic choices
- Highly efficient and rigorous encoding
- All information in one single record
- Linguistic rules for the multilevel transducer

Suppose now that we want to characterize a computational lexicon: this should be made in reference to what a parser needs when analysing a text. For instance, in order to produce predictions as to what the internal constituents of a given predicate might be - verb guidance as is commonly called, and consequently the structure of the VP governed by a given verb, we want to have access to the verb predicate argument structures which might correspond to the following information:
1. predicate; 2. syntactic class; 3. aspectual class; 4. semantic class; 5. the list of arguments contains for each argument:
  a. syntactic constituency; b. grammatical function; c. semantic role;
  d. selectional restrictions or a control equation for open functions;
  e. no information at all for propositional arguments.

The final result for the verbs "give" and "give_up" is the following record of information:
i. give, ditrans, achiev, transf_poss, [np/subj/owner/[human], np/obj/poss/[object], pp/obj2/recip/[human]].
ii. give_up, trans, achiev, abort, [np/subj/agnt/[human], np/obj/theme/[activ, object]].

As discussed above, and also proposed by FN, a number of derived structures can reasonably be expected to descend from such the lexical representation of verbs belonging to the class of trans(itive) or ditrans(itive). However, none of these structure will be represented explicitely in our lexicon seen that they can all be derived automatically from lexical rules, as indicated in LFG theoretical framework. Among these structures we note the following:
- passive constructions for transitive verbs;
iii. The present was given to Mary (general passive structure)
iv. Mary was given a present (special passive rule for English)
- dative shift
v. John gave Mary a present (specialized for Germanic languages)
- generation of a quantified oblique argument in case of agentless passive;
vi. The book was sold to help the students
- ergative or inchoative constructions again from transitive verbs;
vii. The book sold (*to help the students)
Il palazzo si è mosso / the palace self has moved
- reflexive constructions from transitive verbs;
viii. Gino si è lavato (la faccia) / John washed (his face)
- object intransitivized constructions from transitive verbs that allow it and are therefore marked as such in terms of object argument NP optionality;
ix. John is reading
- direct object NP deletion or indirect object2 PP deletion from psychic verbs, adequately marked in the lexicon, inducing thus the interpretation of generic assertions;
x. La musica rende _ felici / Music makes (everybody) happy
- postverbal subjects for focusing purposes with unaccusatives, freely in Italian;
xi. E' arrivato Gino / *has arrived John
- postverbal subjects for topicalizing purposes, freely in Italian and English;
xi. … disse Gino / … said John
- preverbal sentential argument dislocation with impersonal predicates;
xii. That John is happy results from what I heard yesterday
- controlling secondary predication with transitive verbs is expressed in the lexicon in Italian where it is limited to a restricted class of predicates - "dipingere" for resultatives, "considerare, ritenere" with individual level predication and "rendere" for stage level predicates; in English it is constrained by aspectual class of the predicate - only change of state verbs allow freely resultatives, i.e. accomplishments and achievements (see Rapoport);
xiii. John hammered the metal flat (*happy, *tall)
John painted his house red (*high)
- middle constructions are not freely generable from the lexicon. In particular, aspectual classes are relevant - stative verbs cannot form middles; only certain adverbs are allowed; only affected themes may be used, however (see Hoekstra & Roberts); the same applies to impersonal constructions;
xiv. *This lesson knows easily
Bureaucrats bribe easily (*evidently)
This truck loads easily
Alongside the specification of optionality for certain arguments - the feature "lexs" - our computational lexicon lists a number of interesting transitive predicates. Here below we list some examples with special argument specifications for double object constructions, adverbial objects, and secondary predication: all these cases require the parser to search in the argument list for a given functional or semantic type. In the case of adverbial object NP no passive is allowed, nor can it constitute a legal argument for wh- movement; for secondary predication we shall have an open function which predicates on the object NP, as happend in verbs like "considerare"/consider.

4.2 Transitivity

Theoretically speaking, we can assume that transitivity is the main syntactic class we want to characterize in our lexicon, the remaining classes being only derivative on it. In fact, from our work on the main general lexicon of Italian, we found that transitive verbs constitute the great majority of verbs, and their number is by far superior to intransitive verbs. Over a number of 13,000 verb entries we see that intransitive classes are only approximately 3,000, the remaining belonging to one of the transitive classes, as follows:

Tr = transitive [6700]; tr_cop = transitive+predicative argument [112]; tr_perc = transitive_perceptive [24]; ditr(+preps) = ditransitive [386]; psych1= psychic 1 [59]; psych2 = psychic 2 [251]; psych3 = psychic 3 [19]; inac = unaccusative [935]; inerg = unergative [1612]; rifl = reflexive [890]; rifl_rec = reflexive reciprocal [203]; rifl_in = reflexive inherent [304]; erg_rifl = ergative reflexive [1742]; imp = impersonal [30]; imp_atm = impersonal atmospheric [32]; cop = copulative [8]; mod = modal [5]; C_mov = motion verb + another class [255]; C_prop = propositional verb + another class [210];

Most of the classes have labels which are self-explanatory. Some of the classes have specific values like the one of psych(ic) verbs, which we divided into three subclasses both for syntactic and semantic reasons (but see Delmonte, 1989):
- psych1 verbs assign the Experiencer role to the subject;
- psych2 verbs assign the Experiencer role to the direct object;
- psych3 verbs are intransitive and assign the Experiencer role to an oblique argument;

Another class that needs clarification is the one of,
erg(ative)-refl(exive) - reflexive verbs which have as subject, the object of the transitive or psychic corresponding verb.
Lately, we did additional lexical research on the properties of specific lexical classes: in particular, we focussed our attention on verbs governing sentential complements, including both tensed and untensed propositions. The final classification ended up with over 100 different syntactic-semantic classes. Differences regarded mainly the following linguistic items:
a. optionality of the complementizer;
b. type of the complementizer;
c. obligatoriness of subjunctive mood;
d. obligatoriness of negation;
e. propositional type;
f. optionality of the sentential complement;
g. idiomatic lexical forms.
From basic syntactic classes, the 19 listed above, we posited the need to introduce more fine-grained labels where we specialized the following ones:
h. ditr(ansitive)1 - <SUBJ, OBJ, COMP>
i. ditr(ansitive)2 - {<SUBJ, OBJ2, COMP>; <SUBJ, OBL, COMP>}
We then added a specific class for negative transitive verbs:
l. ntr_propint(negative transitive + question + subjunctive mood)
Examples of idiomatic lexical forms are the following ones:
o. avere, aux_prop_a-[ebbi, ebbe, ebbero]
p. dare, ditr2_prop-[a_bere]
q. far_sì, tr_prop
r. valere_la_pena, inac_propx
where prop stands for proposition and propx for infinitive lacking a complementizer.

Going back to the general remarks on transitivity, if we look at nontransitive predicates, we might regard them as conflated forms of either transitive or copulative corresponding predicates. In line with what Hale and Keyser(1991) discuss in their paper, we could also assume that a lot of verbs are

conflated forms in which a light verb (fare, stare, avere, mettere etc.) and either an adjective or a noun are made into a verb. We can have verbs like “figliare” (avere un figlio) for animals; verbs which allow cognates object as in “vivere, correre” from “fare una vita” or still “vivere una vita” with the repetition of the internal object; or verbs like “pranzare, cenare, etc.”, or “pescare, bagnarsi, tuffarsi, etc.” presumably made up of a light verb “fare” and the internal noun denoting the kind of activity or event; weather verbs “piovere, nevicare, lampeggiare, tuonare, etc.”
Besides, we know that derivational rules build verbs from adjectives, as discussed for Italian in Scalise(1984) with parasynthetic derivation - “annerire, arrossire, appiattire, etc.” with the obvious meaning “cause to become ADJ” where ADJ is the underlying adjective base form. The same thing could be said for an extended number of inherent reflexive verbs like “arrabbiarsi, ammalarsi, indurirsi, etc.” which could be treated as a form of a copulative verb, using “diventare” as a light verb and the corresponding adjective as a predication of the subject.
Finally, we have psychic verbs which are clearly decomposable into a light verb and a noun, as in “addolorare, interessare, divertire, meravigliare, stupire, piacere etc”; this can be explained by the fact that in the argument structure of the underlying noun there is always the presence of a beneficiary, or an emotional causer and an experiencer. Thus we can say:
i. il dolore/l’interesse/lo stupore/il piacere di Gianni per Maria/*di Maria
which is equitable to,
ii. Maria addolora/interessa/stupisce/piace a/ Gianni

4.3 Lexicon and Semantics

The relations which we assume as relevant from the point of view of semantic interpretation and which can reasonably be encoded in a computational lexicon are the following:
- eventually, the possibility to interpret the structure of the event denoted by the sentence in which the predicate is analysed as a whole, depends compositionally on the basic lexical aspectual class, tense specifications, definiteness and quantification of the object NP, the presence of temporal adjuncts.
- relations intervening between aspectual class and the presence of a certain tense morpheme;
- relations intervening between aspectual class and the presence of temporal adjuncts;
- relations intervening between aspectual class and object NP quantificational properties;
Semantic roles are derived from argument structure and semantic field information as well as from aspectual class; the following is a complete list of roles derivable in our lexical representations:

Argument semantic roles associated to NP and to non NP
Agent, Head (only for adjectives), Theme_affect, Appost, Poss, Theme_unaffect, Theme_emot, 8-Causer_emot, Actor, Exper, Theme_bound (subj of predication), Trans_obj, Istigat, Address, Source_info, Goal. Specfn (only for nouns), Receiv, Owner, Perciv, Possesn, Locat, Causer (non-human), Informtn, Patnt, Nattrib, Recipnt, Ex_owner, Comit, Instr, Source, Med_exch, Ratio, Benef, Temp, Direc, Malef, Mattr, Locat/manner, Locat_exten, Asp_progr, Manner, Adjnct, Prop, Propint, Propf, Quest, Excl

5. The classification algorithm

As described in detail below, we start with a description of lexical entries which takes into account their categorial status and reflects their syntactic behaviour in terms of major constituents, according to the predictions made by an X-bar system (see Jackendoff, 1977). The result is a set of fully specified Subcategorization Frames - in the sense that the categorial status of the subject or external argument is also included - with selectional restrictions expressed in terms of inherent semantic features.
However this is not sufficient to account for Functional Structures, which in our model are derivative on syntactic constituency; also, since functional control had to be specified at a lexical level (we chose to include this explicitely for each controlled argument rather than deriving it from a default rule and

specifying control for exceptions, because this introduced a lack of uniformity in lexical formats) - this is specified by an additional feature on the controlled argument.
As we said, in LFG grammatical functions are specified in lexical forms to produce new lexical entries by means of Lexical Redundancy Rules. Some such rules, for instance Passivization, are activated by the presence of two arguments,with function SUBJECT the first and OBJECT the second. Given a parser which implements general conditions of functional completeness and functional coherence, each lexical form for a verb thus defines a set of grammatical contexts in which the verb can be lexically inserted.
Finally, since syntactic constituency does not make available information on the semantic and logic nature of arguments of predicates, we specify syntactic classes like Unaccusative, Unergative, Copulative and so on, to tell predicative from non predicative arguments apart.
What is more important, the level of predicate argument structure containing information relative to Theta-roles - which is derivative on lexical forms containing grammatical functions - requires information as to the nature of semantic classes, like Aspectual classes. Theta-roles assignment precedes the creation of Conceptual Representations used to work out inferences.
In sum, there is a set of computer programs building on a primitive description of lexical items based on categorial status, main constituents subcategorized, inherent features, syntactic and semantic classes. It produces a number of representations divided up into four levels: a syntactic level, or Level 1, a functional level Level 2, a thematic level Level 3 and a conceptual level Level 4. Each level builds on top of the lower one using restrictions based on the information present at that level and class information available. For instance, syntactic classes are only available at Level 2 in order to assign grammatical functions; semantic classes are only available at Level 3 and 4 in order to build semantic roles and conceptual representations. As to conceptual representations, they are clearly an expansion of semantic roles, which are abstractions and reductions on the finer grained semantic structures of predicates, and can likewise be produced with some additional information constituted by semantic classes.
The classification regards an enlarged version of the frequency lexicon of written Italian which we call LIFVE, consisting in approximately 2000 verbs, 4000 nouns and 1000 adjectives which however amount finally to 12,000 entries when different contexts of use are taken into account. In addition, we have a list of about 500 function words(see Delmonte, 1989). The output of LIFVE is used by our system for text understanding GETA_RUN and for the generation of conceptual representations which will be commented further on in this paper (but see also Delmonte, 1995).
The basic classification (see also Carrier et al. 1993) is intended to provide information to be used by more than one level of representation. Level 1 is a complete syntactic representation with a rich number of features partly used by Level 2, partly by Level 3, partly by lexical redundancy rules, and partly by Level 4. Some of these features are associated with the predicate and some with its arguments: the former are features such as [+RAIS], [+INAC], [+PERC], [+FACT]; the latter are inherent semantic features and include the following, [+ANIM, -ANIM, +HUM, -HUM, +ABST, -ABST]. Only three features can have agentive and causative meaning, [+ANIM, +HUM, -HUM], where +Anim, stands for animal kind, +Hum for human kind, and -Hum for natural powers. These three features individuate then possible causers, the first two intentional, and the latter non-intentional. Level 1 lexical classification is obtained by an interactive program written in C language, which presents a series of windows with multiple choice menus. The classifier - a linguist - is thus required to provide what we define as full subcategorization frames which basically expand the classical notion of syntactic subcategorization and extend it to include explicit reference to the subject argument, a distinction between argumental XP arguments and predicative XP arguments, plus a number of syntactic and semantic features which are meant to allow the translation of each level of representation. They are also used by the parser to impose restrictions on the class of semantically compatible and appropriate modifiers and adjuncts.
Level 2 is an annotation in lexical-functional grammar (LFG) terms of complete grammatical functions onto major syntactic constituents: this level is coupled with a number of lexical redundancy rules, which are expressed as rules in the parser, in order to build derived structures such as passivized and pronominalized structures, inchoativized and intransitivized structures and so on.

Level 3 is a translation into the semantic roles augmented by a system of aspectual features. Aspectual classes are accounted by features such as STATES, ACTIVITIES, ACHIEVEMENTS, ACCOMPLISHMENTS, and others. They are derived from the more primitive features: ±imperative, ±temporal dependent, ±ingressive, and the interaction with syntactic classes. In turn, both syntactic classes and aspectual classes contribute to the labelling of arguments in terms of Theta-roles as appearing in Level 3 representations. In terms of thematic relations, the subject of activity verbs cannot be a Theme and can be either an Agent or an Experiencer. In the class of activity predicates, achievements can be distinguished from accomplishment in that the latter class requires an agentive subject whereas the former does not. It is part of the nature of accomplishments that they involve a result or an end, and these can be expressed by the direct object or by another argument strictly subcategorized by the verb. Thus, an activity verb like "run" becomes an accomplishment if there is a specific distance to overcome or a specific goal to attain. This implies that verb aspectual classes are closely connected with the semantics of the arguments of a predicate. As to syntactic classes, PSYCHic verbs receive a labelling of their arguments in terms of Theta-roles that assigns THEME to the Subject argument and EXPERIENCER to the Object one. Level 4 is a classification in conceptual representation derived strictly from Level 3 plus information on semantic categories. Level 1 lexical classification is listed in the following Table which shows the basic items included in our system of representation:

5.1 Rules for the transduction from Grammatical Functions into Semantic roles

Here below we indicate some of the rules for decoding from functional labels into theta-roles. The translation from functional specifications into semantic roles works according to the following scheme:
- all SUBJECT arguments are translated into AGENT role
- all OBJECT arguments are translated into THEME role

This rule constitutes the default value assigned to Subject and Object arguments. This default rule however is applied after all exceptions have been computed. The exceptions make use of a number of features available in Level 1 representation and can be divided up into syntactic and semantic exceptions to the default rule.

1. Exceptions for Subject decoding:

a. Syntactic
- the verb is classified as Psychic1 ----> CAUSER_EMOTIONAL
- the verb is classified as Psychic2 ----> EXPERIENCER
- the verb is classified as Inherent Reflexive ----> THEME_AFFECTED
- the verb is classified as Perception verb ----> PERCEIVER
- the verb is classified as Raising verb ----> THEME_BOUND

etc.

b. Semantic
- the features associated to the Subject are NOT [+HUMAN] ---> CAUSER
- the verb is classified as Unergative and is STATIVE --> THEME_UNAFFECTED
- the verb is classified either as Inherent Reflexive or as Psychic and the feature associated to the Subject is [+HUMAN] ---> THEME_EMOTIONAL
- the verb is classified as Copulative and stative ----> THEME_BOUND
- the verb is classified as stative and [+human] ----> EXPERIENCER

etc.

2. Exceptions for Object decoding

a. Syntactic
- the verb is classified as agentive ----> THEME_AFFECTED
- the verb is classified as Inherent Reflexive or Psychic1 ----> EXPERIENCER
- the verb is classified as Psychic2 ----> THEME_EMOTIONAL

- the verb is classified as Raising ----> THEME_BOUND
- the noun is classified as pure common Noun ----> AGENT

b. Semantic

- the verb is classified as "accomplishment, pol=1" -----> THEME_EFFECTED
- the verb is classified as "informational" ----> THEME_UNAFFECTED
- the verb is classified as stative ----> THEME_UNAFFECTED
- the verb is classified as "manipulative" and the selection is [+human] ----> GOAL

etc.

The remaining decoding instructions are quite simple and are derived straightforwardly from functional labels. Some difficulties derive from ambiguous prepositions for Oblique argument decoding, and need manual intervention.

Among the conceptual/semantic classes there two classes which are underspecified: EXTENSIONAL and PERLOCUTORY. An extensional predicate may have to be further specified in case, for instance, the domain requires a notion of ingesting a solid vs drinking a liquid. The perlocutory or performative class includes all predicates which ensue in an event which has a social outcome, which require a certain social status in order to be adequately performed, which have a legal domain as inherent and so on.

The need to increase the number of theta/semantic roles is determined by the requirements of both conceptual representations and syntactic and semantic parsing. The basic idea is that whenever an internal argument/object is actually to be computed as an Affected Theme there must be an Agent as Subject. Real Affected Themes are those which undergo a change of state as a result of the action denoted by the predicate. There is a great number of rules of grammar which are sensitive to the partition of OBJects into Affected Themes versus non-Affected Themes. In particular, all predicates of change of position, of possession, of information do not activate Affected Themes - thus no Resultative may be selected as a semantically adequate secondary predication of the OBJect. The same applies, obviously for psychic predicates. Also consider the partition of Agents vs. Recipients where the former give rise to CAUSE as a conceptual template, while the latter requires LET. Finally all alternates like go/come, sell/buy, and others discusses by B.Levin(1986) including the anti-causative alternation. For this reason we have increased the number of theta-roles which map onto SUBJect, OBJect and other grammatical function on the basis of conceptual fields. Alternates are accounted for by the presence of the semantic attribute of polarity: kill, has polarity=0, while create has pol=1; buy has polarity=0, while sell has polarity=1 because the seller possesses the object at time t1 whereas the buyer does not possess it, and so on and so forth.

5.2 Some Rules from the Transducer Algorithm

```
abbassare        1 1 1 9 - 1 0 % 3 4 - 1 0 % 2 4      /lower/
abbassare,tr,activ,measu,[np/subj1/agent/[+hum,-hum],np/obj1/theme_affect/[-ani,-hum]]
                     if (c == '0') {
                     if (act) {strcat(output, "STAYexten(AT"); }
                     else if (td || ingr) {strcat(output, "GOexten(TO"); }
                     else strcat(output, "IDENTexten(AT");
abbassare[CAUSE(<agent>(STAYmeas(PLACE-AT<theme_affect>)){(e1(t1,t2,t3))}]
credere          1 1 2 1 - 1 0 % 3 - 5 0 % 1              /believe/
credere,tr,statv,subj,[np/subj1/exper/[+hum],scomp/propints[subj=subjx]]
                     else if (c=='1') {
                     if (act) {
                     strcat(output, "(STAY-IN SUBJ-MIND(REP"); }
                     else if (td || ingr) {
                     strcat(output, "(GO-IN SUBJ-MIND(REP"); }
                     else strcat(output,"(IDENT-IN SUBJ-MIND(REP");
```

credere[BE(<exper>(IDENT-IN SUBJ-MIND(REP<prop_nonpred>)){(en,tn)}]

6. What are FRAMES and Conceptual Representations good for??

In Delmonte(1990) we considered CRs as the link from semantics as specified in a Discourse Model to world knowledge by means of a mapping algorithm with inference rules that allowed reasoning to be performed. As shown above, each conceptual representation carries an event structure which is used for temporal reasoning

arrivare[BE(<th_unaff>(STAY$_{posit}$(AT<locat>))){(e_n,t_n),(e_1,t_1)}]/arrive

Temporal Reasoning works on interval semantics and the inferential procedure has a set of basic rules which decompose CRs on more elementary predicates like BE, as follows:

[STAY ([X],[AT Y]) from t1 to t2] => [BE ([X],[AT Y]) at t3] cond = t1<t3<t2

where the restriction on spatial location of the argument of the main predicate states that in order to Arrived at a given location Y at a given time t1, the participant entity X needs to Be there at time t3 which precedes t1.

6.1 Mapping into the KB from the DM by means of CRs

Knowledge representation is fed dynamically from the list of facts produced by our system for text understanding stored in a Discourse Model, where anaphoric processes are taken care of. Facts listed in the DM only carry information related to Semantic Roles associated to a given entity in a given situation and governed by a given verb predicate. In addition to that, properties may be associated to the same entity in the same situation, where also a polarity may appear – index 1, argument slot 4 in fact. Verbal predicates are associated to a specific semantic index, K, to which the overall situation is finally linked by the predicate **linkverb**.

```
generate :-
fact(_,isa,[arg:K,arg:ev],1),
fact(K,Pred,[Agent:X,Locat:Y],1),
associate_primary_function_and_roles(K, Pred, Agent:X, Locat:Y).
```

Primary function and roles is activated by unification with lexical semantic information associated to main predicate – GO.

```
associate_primary_function_and_roles(K, go, Agent:X, Locat:Y):-
   role_saturation(go, Agent, Locat, SuperFunct, PrimFunct, ExtraRoles),
   db_mapping(SuperFunct, PrimFunct, [X, ExtraRoles,Y], K).
```

Role saturation is the actual transfer of conceptual information and is where the mapping from semantic roles to conceptual information takes place:

```
role_saturation(go, th_aff, locat, SuperFunct, PrimFunct, [ExtraRoles])
:-
go[GOexten( <th_aff>(FROM<source>(TO<locat>))){(en,tn)}],
SuperFunct=CAUSE, PrimFunct=GO, ExtraRoles=source,!.

db_mapping(MainPrimitive,PrimaryFunction,[Agent|ArgumentRoles],SemanticIndex)
:-
```

R::MainPrimitive and theme:Agent and primitiveIntroduced:Primitive, introduce_specific_primitive(PrimaryFunction, [Agent |ArgumentRoles], Primitive), linkverb(SemanticIndex,R),!.

Finally the specific primitive is associated to the conceptual meaning representation which contains reference to the “spatial” semantic field:

introduce_specific_primitive(GO, [Agent,From,To], Primitive):-
Primitive=(S::primGO and theme:Agent and from:From and to:To and semanticfieldgo:spatial),!.

linkverb(A,B) :- A::anything and prim:B.

7. CONCLUSIONS

We have shown how inferences can be drawn both for event structures, event participants and temporal structures by linking PAS derived from syntactic and semantic processing of text and enriched with semantic role labels. The inventory of semantic role labels we use is much smaller than the one proposed by FN with its FEs. Besides, conceptual representations have been automatically induced from a small number of syntactic, aspectual and semantic templates.

References

Baker, Collin F., Charles J. Fillmore, and John B. Lowe (1998): The Berkeley FrameNet project. In Proceedings of the COLING-ACL, Montreal, Canada.
Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Blackwells.
Carrier J., Randall J.(1993), Lexical Mapping, in Reuland & Abraham(eds) (1993), Knowledge and Language, Vol II, Kluwer Ac. Dordrecht, 119-142.
Chomsky, Noam. 1970. "Remarks on nominalizations." in Jacobs and Rosenbaum (1970).
Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.
Chomsky, Noam. 1986. *Knowledge of language*. New York: Praeger.
Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet project. In Proceedings of COLING-ACL-98, Montreal, Canada. John Chen and Owen Rambow. 2003. Use of deep linguistic features for the recognition and labeling of semantic arguments. In Proceedings of EMNLP-03, Sapporo, Japan.
Croft, W. (1998). The structure of events and the structure of language. In Tomasello, M.(ed.), The New Psychology. London: Lawrence Erlbaum. 67-92.
Delmonte R., 2007(to be published). *Computational Linguistic Text Processing*, Book II – Logical Form, Semantic Interpretation, Discourse Structure, Reasoning and Question Answering, Nova Science Publishers, New York.
Delmonte, R., 2004. Text Understanding with GETARUNS for Q/A and Summarization, *Proc. ACL 2004* - 2nd Workshop on Text Meaning & Interpretation, Barcelona, Columbia University (2004) 97-104.
Delmonte R.(1989), From Subcategorization Frames to Thematic Roles, in R.Delmonte(ed.)(1989), *ibid.*, Chapt.IV, pp.167-212.
Delmonte R.(1990), Semantic Parsing with an LFG-based Lexicon and Conceptual Representations, *Computers & the Humanities*, 5-6, pp.461-488.
Delmonte R.(1995), Lexical Representations: Syntax-Semantics interface and World Knowledge, in *Rivista dell'AI*IA (Associazione Italiana di Intelligenza Artificiale)*, Roma, pp.11-16.
Delmonte R.(1996), Lexical Representations, Event Structure and Quantification, *Quaderni Patavini di Linguistica* 15, 39-93.
Delmonte R., G.Ferrari, A.Goy, L.Lesmo, B.Magnini, E.Pianta, O.Stock, C.Strapparava(1996), ILEX - Un dizionario computazionale dell'italiano, *Atti Convegno Nazionale AI*IA Cibernetica e Machine Learning*, Napoli, pp.27-30.
Delmonte R.(1997), Rappresentazioni lessicali e linguistica computazionale, *Atti SLI, Lessico e Grammatica - Teorie Linguistiche e applicazioni lessicografiche*, Roma, Bulzoni, pp.431-462.
Bianchi D., Delmonte R.(1997), Rappresentazioni concettuali nella comprensione di storie, *Atti Convegno Nazionale AI*IA Apprendimento Automatico e Linguaggio Naturale*, Torino, pp.95-98.
Delmonte R.(ed.)(1989), *Lessico, Stutture e Interpretazione*, UniPress, Padova.

Delmonte R.(1992), *Linguistic and Inferential Processing in Text Analysis by Computer*, UP, Padova.
Dorr Bonnie J. and Mari Broman Olsen, 1997. Deriving Verbal and Compositional Lexical Aspect for NLP Applications, *Proc. 35th ACL*, 151-158.
Ellsworth, Michael, Katrin Erk, Paul Kingsbury, and Sebastian Pado (2004): PropBank, SALSA, and FrameNet: How design determines product, Proceedings of the LREC 2004 Workshop on Building Lexical Resources from Semantically Annotated Corpora, Lisbon.
Fellbaum Christiane (Ed.) 1998. Introduction. In C. Fellbaum (Ed.), *WordNet: An Electronic Lexical Database*. MIT Press.
Fillmore, Charles J. and B.T.S. Atkins (1998): FrameNet and lexicographic relevance, Proceedings of the First International Conference on Language Resources and Evaluation, Granada, Spain.
Fillmore, Charles J. and Collin F. Baker (2001): Frame Semantics for Text Understanding. In Proceedings of WordNet and Other Lexical Resources Workshop, NAACL, Pittsburgh, June, 2001.
Fleischman, Michael, Namhee Kwon, and Ed Hovy. 2003. Maximumentropy models for FrameNet classification. In Proceedings of EMNLP-03, Sapporo, Japan. Daniel Gildea and Dan Jurafsky. 2002. Automatic labeling of semantic roles. Computational Linguistics, 28(3):245–288. Jeffrey Gruber. 1965. Studies in lexical relations. MIT Working Papers in Linguistics, Cambridge, MA.
Gildea Daniel and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3): 245–288.
Goldberg, A. (1995). Constructions. A Construction Grammar Approach to Argument Structure. Chicago: University of Chicago Press.
Goldberg, A. (1998). Patterns of experience in patterns of language. In Tomasello, M.(ed.), The New Psychology. London: Lawrence Erlbaum. 203-219.
Green, Rebecca and Bonnie J. Dorr, and Philip Resnik. 2004. Inducing frame semantic verb classes from WordNet and LDOCE. *42nd Annual Meeting of the Association of Computational Linguistics*, Proc.ACL-SIGLEX, Workshop on Deep Lexical Acquisition, 57-66.
Green R., Pearl L, B.J.Dorr, 2001. Mapping WordNet Senses to a Lexical Database of Verbs, submitted to 39th ACL, Toulouse France.
Green Rebecca and Bonnie J. Dorr. 2004. Inducing a semantic frame lexicon from WordNet data. Workshop on Text Meaning and Interpretation, *42nd Annual Meeting of the Association of Computational Linguistics*.
Eva Hajicova and Iyona Kucerova. 2002. Argument/Valency Structure in PropBank, LCS Database and Prague Dependency Treebank: A Comparative Pilot Study. In *Proceedings of the Third International Conference on Language Resources and Evaluation*, pages 846–851.
Heringer, H. J. (1993). Dependency syntax - basic ideas and the classical model. In Jacobs, J., Stechow, A. v., Sternefeld, W., & Venneman, T.(eds.), Syntax - An International Handbook of Contemporary Research, Volume 1. Berlin: Walter de Gruyter. 298-316.
Holmes, J. (2002). Lexical Properties of English Verbs. PhD. UCL, London.
Holmes, Jasper and Richard Hudson, Constructions in Word Grammar, manuscript, version of October 2001.
Hudson, R. (1990). English Word Grammar. Oxford: Blackwell.
Jackendoff R.(1972), Semantic Interpretation in Generative Grammar, The MIT Press, Cambridge, MA.
Jackendoff R.(1976), Towards an Explanatory Semantic Representation, Linguistic Inquiry 7, 89-150
Jackendoff R.(1980), Belief Contexts Revisited, Linguistic Inquiry 11(2).
Jackendoff R.(1983), Semantics and Cognition, MIT Press, Cambridge Mass.
Jackendoff R.(1985a) Believing and Intending: Two Sides of the Same Coin, Linguistic Inquiry 16, 445-460.
Jackendoff R.(1985b), Multiple Subcategorization and theTheta-Criterion: the Case of Climb, Natural Language and Linguistic Theory 3, 271-296.
Jackendoff R.(1987a), Consciousness and the Computational Mind, The MIT Press, Cambridge Mass.
Jackendoff R.(1987b), The Status of Thematic Relations in Linguistic Theory, Linguistic Inquiry 18, 369-411.
Jackendoff, Ray S. 1992. "Parts and boundaries." in Levin and Pinker (1992). 9-46.
Jackendoff R.(1993a), The Combinatorial Structure of Thought: The Family of Causative Concepts, in Reuland & Abraham(eds), op.cit., 31-50.
Jackendoff R.(1993b), On the role of Conceptual Structure in Argument Selection: A Reply to Emonds, Natural Language and Linguistic Theory 11, 2, 279-312.
Kay, P. and Fillmore, C. (1999). Grammatical constructions and linguistic generalizations: The what's X doing Y? Construction. Language 75. 1-33.
Levin Beth, 1993. *English Verb Classes And Alternations: A Preliminary Investigation*, The University of Chicago Press, University of Chicago.

Levin, Beth and Steven Pinker. 1992. *Lexical and conceptual semantics*. Oxford: Blackwell.
Pado, Sebastian , Gemma Boleda, 2004. The Influence of Argument Structure on Semantic Role Assignment, ACL2004, EMNLP, pp.103-110.