SPOKEN LANGUAGE UNDERSTANDING: A SURVEY

Renato De Mori

LIA – BP 1228 – 84911Avignon CEDEX 9 (France)
renato.demori@univ-avignon.fr

ABSTRACT

A survey of research on spoken language understanding is presented. It covers aspects of knowledge representation, automatic interpretation strategies, semantic grammars, conceptual language models, semantic event detection, shallow semantic parsing, semantic classification, semantic confidence, active learning

Index Terms— Spoken language understanding, conceptual language models, spoken conceptual constituent detection, stochastic semantic grammars, semantic confidence measures, active learning.

1. INTRODUCTION

Epistemology, the science of knowledge, considers a datum as basic unit. A datum can be an object, an action or an event in the world and can have time and space coordinates, multiple aspects and qualities that make it different from others. A datum can be represented by an image or it can be abstract and be represented by a concept. A concept can be empirical, structural, or an a-priori one. There may be relations among data.

Computer epistemology deals with observable facts and their representation in a computer. Knowledge about the structure of a domain represents a datum by an object and groups objects into *classes* by their properties.

Natural language refers to data in the world and their relations. Sentences of a natural language are sequences of words. Words of a sentence have associated one or more data conceptualizations also called *meanings* which can be selected and composed to form the meaning of the sentence.

Semantics deals with the organization of meanings and the relations between signs or symbols and what they denote or mean.

Computer semantics performs a conceptualization of the world using well defined elements of programming languages. Programming languages have their own syntax and semantic. The former defines legal programming statements, the latter specifies the operations a machine performs when an instruction is executed. Specifications are defined in terms of the procedures the machine has to carry out. Semantic analysis of a computer program is essential for understanding the behavior of a program and its coherence with the design concepts and goals. Formal logics can be used to describe computer semantics.

Computer programs conceived for interpreting natural language differ from the human process they model. They can be considered as approximate models for developing useful applications, interesting research experiments and demonstrations. Semantic representations in computers usually treat data as *objects* respecting logical *adequacy* in order to formally represent any particular interpretation of a sentence. Even if utterances, in general, convey meanings which may not have relations which can be expressed in formal logics ([45], p. 287), formal logics have been considered adequate for representing natural language semantics in many application domains. Logics used for representing natural language semantic should be able to deal with *intension* (the essence of a concept) and *extension* (the set of all objects which are instances of a given concept).

Computer systems interpret natural language for performing actions such as a data base access and display of the results and may require the use of knowledge which is not coded into the sentence but can be inferred from the system knowledge stored in long or short term memories. It is argued in [135] that a specification for natural language semantics requires more than the transformation of a sentence into a representation. In fact, computer representations should permit, among other things, legitimate conclusions to be drawn from data [72].

Interpretation may require the execution of procedures that specify the truth conditions of declarative statements as well as the intended meaning of questions and commands [135]. Procedures are executed by an *interpretation strategy*.

Spoken Language Understanding (SLU) is the interpretation of signs conveyed by a speech signal. This is a difficult task because meaning is mixed with other information like speaker identity and environment. Natural language sentences are often difficult to parse and spoken messages are often ungrammatical. The knowledge used is often imperfect and the transcription of user utterances in terms of word hypotheses is performed by an Automatic Speech Recognition (ASR) system which makes errors. Strategies of the first SLU systems performed transformations from signals to words, then from words to meaning. Some strategies were successively propose d to transform signals into basic semantic constituents to be further composed into semantic structures.

This paper reviews the history of SLU research with particular attention to the evolution of interpretation paradigms, influenced by experimental results obtained with evaluation corpora. This review integrates and complements reviews in [22,73].

2. COMPUTER REPRESENTATIONS OF MEANING

Computer representation of meaning is described by a Meaning Representation Language (MRL) which has its own syntax and a semantic. MRL should follow a representation model coherent with a theory of epistemology, taking into account, *intension* and

extension, relations, reasoning, composition of semantic constituents into structures, procedures for relating them with signs. The semantic knowledge of an application is a *knowledge base (KB)*. A convenient way for reasoning about semantic knowledge is to represent it as a set of logic formulas. Formulas contain variables which are bound by constants and may be typed. An object is built by binding all the variables of a formula or by composing existing objects.

Semantic compositions and decisions about composition actions are the result of an inference process. Basic *inference problem* is to determine whether $KB \models F$ which means that KB *entails* a formula F, meaning that F is true in all possible variable assignments (worlds) for which KB is true.

In [135], the possibility of representing semantic relations with links between classes and objects is discussed. The formulas in a KB describe concepts and their relations which can be represented in a network called *semantic network*. A semantic network is made of nodes corresponding to entities and links corresponding to relations. This model combines the ability to store factual knowledge and to model associative connections between entities [135]. Examples of relations are composition functions [44].

The structure of semantic networks can be defined by a graph grammar. Computer programming classes and objects called *frames* can be defined to represent entities and relations in semantic networks. In a frame based MRL, grammar of frames is a model for representing semantic entities and their properties. Such a grammar should generate frames describing general concepts and their specific instances. Part of a frame is a data structure which represents a concept by associating to the concept name a set of roles which are represented by *slots*. Finding values for roles corresponds to fill the frame slots. A *slot filler* can be the instance of another frame. This is represented by a pointer from the filler to the other frame. The semantic of an MRL can be described by procedures for generating instances of entities and relations. This characterizes procedural semantics. Procedures for slot filling as well as for frame evocation use methods.

Different frames may share slots with similarity links. There may be *necessary* and *optional* slots. *Fillers* can be obtained by *attachment* of procedures or detectors (of e.g. noun groups), *inheritance*, default.

Procedures can also be attached to slots with the condition in which they have to be executed. Examples of conditions are *when-needed*, *when-filled*. Slots may contain expectations or replacements (to be considered if slots cannot be filled).

Descriptions are attached to slots to specify constraints. Given a slot-filler for a slot, the attached description can be inferred. Descriptions can be instantiations of a concept carrier and can inherit its properties. Descriptions may have connectives, coreferential (descriptions attached to a slot are attached to another and vice-versa), declarative conditions.

Verbs are fundamental components of natural language sentences. They represent actions for which different entities play different roles. Actions reveal how sentence phrases and clauses are semantically related to verbs by expressing cases for verbs. A *case* is the name of a particular *role* that a noun phrase or other component takes in the state or activity expressed by the *verb* in a sentence. There is a case structure for each main verb. Attempts were made for mapping specific *surface cases* into a deep semantic representation expressing a sort of semantic invariant. Many deep semantic representations are based on *deep case n-ary relations* between concepts as proposed by Fillmore [31]. *Deep case* systems have very few cases each one representing a basic semantic constraint.

In [81], schemas containing roles and other information are proposed as active structures to model events and capture sequentiality.

A popular example of MRL is the Web Ontology Language (OWL) [87] which integrates some of the most important requirements for computer semantic representation..

A heterarchical architecture based on a KB made of situation-action (production) rules is described in [29].

3. SYNTACTIC AND SEMANTIC ANALYSIS FOR INTERPRETATION

An initial, considerable effort in SLU research was made with an ARPA project started in 1971. The project is reviewed in [55] and included approaches mostly based on Artificial Intelligence (AI) for combining syntactic analysis and semantic representation in logic form. Systems of this project generate a sequence of word hypotheses with an ASR system and perform interpretation with the same approaches used for written text. It was assumed, as stated for example in [128], that a semantic analyzer has to work with a syntactic analyzer and produce data acceptable to a logical deductive system. This is motivated by arguments, for example in [44], that each major syntactic constituent of a sentence maps into a conceptual constituent, but the inverse is not true. For example, adapting the notation in [44], a sentence requiring a restaurant near the Montparnasse metro station in Paris can be represented by the following bracketed conceptual structure expression:

Γ:[Action REQUEST ([Thing RESTAURANT], [Path NEAR ([Place IN ([Thing MONTPARNASSE])])]]

The formalism is based on a set of categories. Each category, e.g. Place can be elaborated as a Place-function, e.g. IN and an argument.

The expression Γ can be obtained from a syntactic structure like this:

Ψ:[S[VP [V give, PR me] NP [ART a, N restaurant] PP[PREP near, NP [N Montparnasse, N station]]]]

Assuming that natural languages are susceptible to the same kind of semantic analysis as programming languages, in [78], it is suggested that each syntactic rule of a natural language generative grammar is associated with a semantic building procedure that turns the sentence into a logic formula.

An association of semantic building formulas with syntactic analysis is proposed in categorical grammars conceived for obtaining a surface semantic representation [62].

Semantic knowledge is associated, in this case with lexical entries and logic formulas are composed by actions performed during parsing. The use of a lexicon with Montague grammars is discussed in detail in [26].

Organization of lexical knowledge for sentence interpretation has been recently the object of investigation. VerbNet [54], is a manually developed hierarchical verb lexicon. For each verb class, VerbNet specifies the syntactic frames along with the semantic role assigned to each slot of a frame. Modelling joint information

about the argument structure of a verb is proposed in [123]. In the WordNet Project [75], a word is represented by a set of synonymous senses belonging to an alphabet of *synsets*. It can be used for word sense disambiguation.

Suitable procedures can be attached to frames to generate logical sentences from slots filled are filled. Details on the use of syntax and semantics for natural language understanding can be found in [2].

Slot filling procedures can be executed under the control of a parser or, in general, by precondition-action rules. As natural language is context sensitive, procedural networks for parsing under the control of Augmented Transition Network Grammars (ATNG) were proposed. ATNGs [134] are augmentations of Transition Network Grammars (TNGs). TNGs are made of states and arcs. The input string is analyzed during parsing from left to right, one word at a time. The input word and the active state determine the arc followed by the parser. Arcs have types, namely CAT (to read an input symbol), PUSH (to transfer the control to a sub-network) and POP (to transfer the control from a sub-network to the network that executed the PUSH to it).

In ATNGs condition testing and register setting actions are associated to certain arcs. Actions set the content of registers with linguistic feature values and can also be used for building parse trees. It is also possible to introduce actions of the type BUILD associated to an arc to compose a parse tree or to generate semantic interpretations. Different ATNGs can be used in cascade for parsing and interpretation. An arc type TRANSMIT transfers syntactic structures from the syntactic to the semantic ATNG.

If a portion of a parse tree can be mapped into a semantic symbol of an MRL, then this symbol could be used as a nonterminal in a grammar which integrates syntactic and semantic knowledge. In [135], syntactic, semantic and pragmatic knowledge are integrated into *procedural semantic* grammar networks in which symbols for sub networks can correspond to syntactic or semantic entities.

In [139], TNGs are proposed as procedural attachment to frame slots. A chart parser can be activated for each TNG under the conrol of the interpretation strategy. In [133], a search algorithm was implemented in which the TNG was employed during ASR decoding.

In [127] a best first parser is used. Its results trigger activations in a partitioned semantic network with which inferences and predictions are performed by spreading node activation through links. Tree Adjoining grammars (TAG) also integrate syntax and logic form (LF) semantics [114].

Classification based parsing may use Functional Unification grammars (FUG), Systemic Grammars (SG), or Head Driven Phrase Structure Grammars (HDPSG) which are declarative representations of grammars with logical constraints stated in terms of *features* and *category structure*. Semantics may also drive the parser, causing it to make *attachments* in the parse tree. Semantics can resolve ambiguities and translate English words into semantic symbols using a *discriminant net* for disambiguation.. A interesting example of interleaving syntax and semantics in a parser is proposed in [25].

Semantic parsing is discussed in [144]. A semantic first parser is described in [143].

Simple grammars are used for detecting possible clauses, then classification-based parsing completes the analysis with inference [51].

Early experiment is SLU made it clear the necessity of analyzing portions of a sentence when the complete sentence could not be analyzed. Problems of this type may be due to the fact that spoken language very often does not follow a formal grammar, hesitations and repetitions are frequent and available parsers do not ensure full coverage of possible sentence even in the case of written text.

As grammar coverage was limited for input speech, in [136] ATNGs were proposed to interpret parts of a sentence using a middle out analysis of the input words. A scope specification is associated with grammar actions. Parsing can proceed to the left or to the right of the input word. Scope specification indicates a set of states the parser has to have passed through before the action can be safely performed. If this is not the case, the action is delayed. In [112], it is proposed to relax parser constraints when a sentence parser fails. This will permit the recovery of phrases and clauses that can be parsed. Fragments obtained in this way are then fused together.

More complex systems using fallback were proposed. Noticeable examples are The Delphi system [10] and the Gemini system [46]. They are described in some detail in ([22], ch. 14).

4. FINITE STATE PROBABILISTIC MODELS FOR INTERPRETATION

Even if there are relations between semantic and syntactic knowledge, integrating these two types of knowledge into a single grammar formalism may not be the best solution. Many problems of automatic interpretation in SLU systems arise from the fact that many sentences are ungrammatical, the ASR components make errors in hypothesizing words and grammars have limited coverage. These considerations suggest that it is worth considering specific models for each conceptual constituent.

In addition to partial parsing [51] and back-off, in the Air Travel Information System (ATIS) project, it was found useful to conceive devices for representing knowledge whose imprecision is characterized by probability distributions. It was also found useful to obtain model parameters by automatic learning using manually annotated corpora. This works as far as manual annotation is easy, reliable and ensures a high coverage.

Stochastic finite-state approximations of natural language knowledge are practically useful for this purpose. Finite-state approximations of context-free grammars are proposed in [89]. Approximations of TAG grammars are described in [97]. A review of these approximations is provided in [28].

Let assume that a concept C is expressed by a user in a sentence W which is recognized by an ASR system, based on acoustic features Y. This can be represented as follows: $Y \rightarrow_e W \rightarrow_e C$. Symbol $\rightarrow_e$ indicates an evidential relation meaning that if Y is observed then there is evidence of W and, because of this, there is evidence of C.

There are exceptions to this chain of rules, because a different concept C' can be expressed by W and Y can generate other hypotheses W' which express other concepts. Furthermore, C can be expressed by other sentences W_j which can be hypothesized from Y. The presence of C in a spoken message described by Y can only be asserted with probability:

$$P(C|Y) \approx \frac{1}{P(Y)}\left[\sum_j P(Y|W_j)P(CW_j)\right]$$

Let assume now that C is a sequence of hypotheses about semantic constituents, the following decision strategy can be used to find the most likely sequence C' as follows:

$$C' = \arg\max_C P(C/Y) = \arg\max_C P(Y/W)P(CW)$$

Word hypotheses are generated by an ASR system using a *probabilistic language model (LM)*.

A solution based on the above introduced concepts is implemented in the system called Chronus [90]. The core of this system is a stochastic model whose parameters are learned from a corpus in which semantic constituent are associated to sentence chunks. The *conceptual decoder* at the core of Chronus is based on a view of utterances as generated by an HMM-like process whose hidden states correspond to meaning units called *concepts*. Thus, understanding is a decoding of these concepts hidden in an utterance. In the Chronous system, the probability P(CW) is computed as follows.

$$P(CW)=P(W|C)P(C)$$

P(C) is obtained with *concept bigram probabilities*

Examples of learning algorithms for finite state transducer can be found in [91].

In [27], it is proposed to extract concept hypotheses from a word lattice. Each concept hypothesis is extracted with a probabilistic *conceptual semantic context-free grammar*.

The CHANEL system [60], performs the following computation:

$$P(CW)=P(C|W)P(W)$$

CHANEL learns semantic interpretation rules by means of a forest of specialized decision trees called *Semantic Classification Trees* (SCTs). The required annotation only consists in listing the concepts present in a sentence.

There is an SCT for every elementary concept. An SCT is a binary tree with a question associated to each node. Questions span an entire sentence. They are generated and selected automatically. Probability P(C|W) is obtained from the counts of times the leaf corresponding to the pattern that matched with W is reached. Notice that W can be an entire sentence. Different conceptual constituent hypotheses can be generated by different sentence patterns that share some components. A frame-based semantic representation is generated by rules. In the ATIS domain a frame instance is expressed by a single spoken message.

Specific *conceptual language models* can be used in ASR decoding [95, 145] to obtain semantic constituent hypotheses, possibly a lattice of them, directly from the signal rather than from word hypotheses. A compound LM can be obtained by integrating a generic n-gram LM with specific LMs, one for each semantic constituent. Specific LMs can be accepted by stochastic finite-state machines (FSM). Variable N-gram Stochastic Automata (VNSA) and their use for hypothesizing semantic constituents are proposed in [101].

In [83], weighted finite state machine (WFSM) are proposed whose edges are labelled with words. A path in the WFSM represents a phrase. Word n-grams and WFSMs can be combined and regarded as a hidden Markov model (HMM).The model construction starts with sentence parsing. The first step of the construction consists in partially parsing a training corpus in order to recognize sequences of words as phrases. The training corpus is first annotated with part of speech (POS) tags. At the end of this process, each word of the corpus is associated to its most probable part of speech. The annotated corpus is then partially parsed using a greedy finite state partial parser. The parser gathers together adjacent words composing a phrase of a given type (noun phrase, verb phrase..). Different grammars are used to recognize phrases of different nature and length. The second step is the construction of phrase classes. It consist in grouping together into classes phrases of the same category. The third step consists in merging together classes having a close internal distribution.

Finite state models can be made more robust by modifying the original topology to take into account possible insertions, deletions and substitutions. Insertion of words not essential for characterizing a semantic constituent can be modeled by groups of syllables [21].

Recent advances in research on stochastic FSM made it possible to generate a probabilistic lattice of conceptual constituent hypotheses from a probabilistic lattice of word hypotheses.

In [99]; a stochastic finite-state conceptual language model CLM_j is conceived for every semantic constituent. An initial ASR activity uses a generic LM, indicated as GENLM, for generating a graph WG of word hypotheses. An automaton AWG is derived for this graph. A sequence W of word hypotheses is scored by its likelihood.

A knowledge source, is built by connecting all the CLM_j in parallel. Such a knowledge source is composed with WG leading to an automaton SEMG:

$$SEMG = WG \circ \left(\bigcup_{c=0}^{C} CLM_c \right)$$

operator $\circ$ indicates composition. CLM_0 is a generic model. Arcs of SEMG are labelled by pairs of symbols. The first symbol of the pair is a word *w* with associated its likelihood. The second symbol of the pair can be the empty symbol, the beginning of a semantic tag or the end of a semantic tag. A semantic tag represents any semantic constituent or structure for which a relation with a word pattern has been identified.

The *support* of a *concept* c_j, $\sup(c_j)$, is the union of all the paths going from the beginning to the end of WG. Supports for different concepts can overlap.

In order to obtain the concept tags representing hypotheses that are more likely to be expressed by the analyzed utterance, SEMG is projected on its outputs leading to a weighted Finite State Machine (FSM) containing only indicators of beginning and end words of semantic tags. The resulting FSM is then made deterministic and minimized leading to an FSM SWG given by:

$$SWG=OUTPROJ(SEMG)$$

where OUTPROJ represents the operation of projection on the outputs followed by determinization and minimization.

A network of conceptual LMs has been used directly in the ASR decoding process [21]. The whole ASR knowledge models in this way a relation between signal features and meaning.

Conceptual hypotheses in the lattice obtained by this projection can be further processed for performing semantic composition and

inference. In [53], an automaton extracts key phrases from continuous speech and converts them to commands for a multi-modal interaction with a virtual fitting room. Finite state LM for interpretation are discussed in [92], Interesting results can also be found in [138]. Integration of semantic predictors in statistical LMs is proposed in [16].

LMs based on Latent Semantic Analysis (LSA) capture some semantic relationship between words. LSA maps the words and histories into a semantic space using Singular Value Decomposition (SVD) technique [6]. Word similarities are measured with distance metrics such as the inner product between vectors. A similar technique was proposed for hypothesizing semantic components in a sentence [15].

A solution with which relevant improvements were observed in large corpora experiments is proposed in [130]. Super abstract role values (superarv) are introduced to encode multiple knowledge sources in a uniform representation that is much more fine-grained that parts of speech (POS).

In [121], a hierarchy of LMs is proposed for interpretation. The introduction of three new ways to use semantic information in LMs is presented in [28].

The introduction of three new ways to use semantic information in LMs is presented in [28].

Finite state models are used to obtain a concept LM score which is interpolated with the n-gram LM score. In a second approach, semantic parse information is combined with n-gram information using a two-level statistical model. In the third approach, features are used for computing the joint probability of a sentence and its parse with a single maximum entropy (ME) model.

5. STOCHASTIC GRAMMARS FOR INTERPRETATION

The rules of a grammar assert the truth of a non terminal symbol given the truth of other terminal and non terminal symbols. The assertion of the presence of a semantic constituent or compound also depends on the assertion of syntactic structures and words. It is thus possible, in principle, to introduce nonterminal symbols which represent semantic entities into a natural language grammar and hypothesize their presence in a sentence with a parsing strategy. Grammars of this type should be context-sensitive and parsing strategies should provide inference capabilities. Nevertheless, context-free grammars or grammars capable of representing certain degrees of context-sensitivity may be adequate for a grop of applications. Furthermore, development of new types of grammars and parser capable of taking into account imprecision made it attractive to integrate syntactic and semantic knowledge into stochastic semantic grammars. Grammars may capture relations between words and semantic constituents as well as knowledge for composing constituents into structures. These grammars can be augmented to contain structure building knowledge and perform logic operations.

Stochastic context-free grammars (SCFG) can generate sentences of any length. Parsing these sentences is an activity that involves the application of a finite number of rules. Sequences of their application can be modelled by a finite state structure and the history of the rules applied before a given rule can be summarized by finite feature sets. Sequences of rule applications and their probabilities are considered in history grammars [8] making them a more accurate probabilistic LM.

For SLU, the linguistic analyzer TINA was proposed. It is written as a set of probabilistic context free rewrite rules with constraints, which is converted automatically at run-time to a network form in which each node represents a syntactic or semantic category [111]. The probabilities associated with rules are calculated from training data, and serve to constrain search during recognition (without them, all possible parses would have to be considered).

A robust matcher was obtained by modifying the grammar to allow partial parses [112]. In robust mode, the parser proceeds left-to-right as usual, but an exhaustive set of possible parses is generated starting at each word of the utterance.

The Hidden Understanding Model (HUM) is inspired by (but not formally equivalent to) Hidden Markov Models [76]. In the HUM system, after a parse tree is obtained, bigram probabilities of partial path towards the root, given another partial path are used. Interpretation is guided by instructions represented by a stochastic decision tree.

Let M be the meaning of an utterance, represented by one or more semantic structures, and let W be the sequence of words that convey this meaning. Hypotheses are scored by the following probability:

$$Pr(M|W) = Pr(W|M)Pr(M)|Pr(W)$$

For given W, the M that maximizes Pr(M|W) can be found by maximizing Pr(W|M)Pr(M), since Pr(W) is fixed. Pr(M) can be estimated from a *semantic language model* that specifies how meaning expressions are generated stochastically; Pr(W|M) can be estimated from a *lexical realization model* that specifies how words are generated, given a meaning. The semantic language model employs *tree structured meaning representations*: concepts are represented as nodes in a tree, with sub-concepts represented as child nodes. Interpretation is guided by a strategy represented by a stochastic decision tree. Each terminal node is the parent of a word or of a sequence of words. Note that unlike Chronus, HUM allows arbitrary nesting of concepts.

Chart parsers were used to analyze portions of sentences in a middle-out strategy and to produce a forest of sub-trees when the parser could not process an entire sentence. The problem of computing the probability of a partial parse when a stochastic CFG is used was investigated in [19] and it was shown that only upper-bounds for parse probabilities can be obtained.

Other examples on the use of semantic grammars can be found in [80]. Parsing word graphs is proposed in [122].

Most grammars have hand-crafted rules which might then be augmented with corpus statistics. Parsing with these grammars suffers from limited coverage.

At Cambridge University [42], an approach based on SCFGs was proposed which does not require fully annotated data for training. The proposed solution considers a hidden vector state (HVS) model. Each vector state is viewed as a hidden variable and represents the state of a push-down automaton. Such a vector is the result of pushing non-terminal symbols starting from the root symbol and ending with the pre-terminal symbol. Non-terminal symbols correspond to semantic compositions like FLIGHTS while pre-terminal symbols correspond to semantic constituents like CITY.

In [129] it is observed that the remarkable robustness exhibited by the auditory system may be attributed to the use of a detection based mechanism. A new formulation is proposed that performs

concept hypothesization in conjunction with ASR decoding under the control of a SCFG.

Combination of semantic and syntactic structures in LM is proposed in [11]. Lexicalized stochastic grammars and head-driven statistical parsers are presented in [14,18]. Partial parses are proposed in [13,106] to enhance robustness. They use a top-down strategy, conditioning word prediction on previously hypothesized structures. Several learning systems have been developed for semantic parsing. These systems use supervised learning methods which only utilize annotated sentences.

In [52], a semi-supervised learning system for semantic parsing using a support vector machine (SVM), is described. Given positive and negative training examples in some vector space, an SVM finds the maximum-margin hyperplane which separates them. When new unlabeled test examples are also available during training, a transductive framework for learning uses them for adapting the SVM classifier.

In [86], statistical translation models are used to translate a source sentence S into a target MRL. Interesting solutions for semantic interpretation using a machine translation approach can be found in [69]. Sudoh and Tsukada, [117] propose a statistical NLU model that can be trained using loose correspondence between pairs of a word sequence and a set of concepts associated at the sentence level. Concepts are represented as attribute/value pairs.

6. MODULAR SEMANTIC INTERPRETATION

Semantic interpretation involves operations of different types performing, among other things, a sort of syntactic analysis, generation of MRL descriptions and inference. Approaches purely based on grammars show limitations is assuring adequate coverage and ability to deal with ungrammatical sentences, hesitations and corrections, imprecision of the ASR component.

In order to increase interpretation accuracy, it appears useful to perform different operations with suitable modules, each using specific methods, models and strategies.

Following ideas about local parsing [1], interesting results were found on the generation of semantic constituents using finite-state models and different types of specific classifiers. Depending on the domain complexity, constituent hypotheses can be composed into semantic structures with semantic grammars, logical inference, and situation-action rules.

Semantic constituent hypotheses are generated with *shallow semantic parsing* using classifiers trained with recent machine learning algorithms. The contribution of different interpretation features is scored with exponential models.

Shallow semantic parsing with the goal of creating a domain independent meaning representation based on a predicate/argument structure was first explored in detail by Gildea and Jurafsky, [33], Pradhan, [93].

Most of the approaches to shallow parsing use features and perform classification and can be divided into two broad classes: Constituent-by-Constituent (C-by-C) or Word-by-Word (W-by-W) classifiers [37].

In C-by-C classification [93], the syntactic tree representation of a sentence is linearized into a sequence of non-terminals syntactic constituents. Then, each constituent is classified into one of several arguments or semantic roles using features derived from its respective context. In the W-by-W method, features are obtained with a bottom-up process for each word after chunking a sentence into phrases.

In [119] a method of unsupervised semantic role labelling is proposed for large corpora. The approach starts with "bootstrapping" by making role assignments that are unambiguous according to a verb lexicon. Then, iteratively, a probability model is created based on the currently annotated semantic roles. This model is used to assign roles having sufficient evidence which are added to the annotated set. The procedure is repeated and probability thresholds are adapted until all predicate arguments have been assigned roles. Class back-off probabilities are used when detailed probabilities cannot be reliably estimated. Interpretation can benefit from useful collections of linguistic information.

A lexicon can be used for semantic role labelling which lists the possible roles for each syntactic argument of each predicate. A predicate lexicon is available for FrameNet [3], and a verb lexicon is available for PropBank [68].

VerbNet [54] specifies, for each verb class, the corresponding syntactic frames along with the semantic role assigned to each slot of a frame.

Various feature-based methods have been proposed for identifying and classifying predicates and arguments and for extracting relations using kernel methods and maximum entropy models [49,118].

In [140], a combination is proposed of partial parsing, also called *chunking*, with the mapping of the verb arguments onto subcategorization frames that can be extracted automatically, for example, from WordNet [75].

MindNet [103] produces a hierarchical structure of semantic relations (*semrels*) from a sentence using words in a machine readable dictionary. These structures are inverted and linked with every word appearing in them, thus allowing performing matching and computing similarities by spreading activation.

Results in [94] with SVM classifiers have shown that there is a significant drop in performance when training and testing on different corpora.

Committee-Based Active Learning uses multiple classifiers to select samples [113]. The concurrent use of SCT, *boosingt* [109] and SVM classifiers is proposed in [100] to increase classification robustness.

A cascade of classifiers for a two step interpretation strategy is proposed in [66]

In [109], the possibility is considered of using human-crafted knowledge to compensate for the lack of data in building robust classifiers. The AdaBoost algorithm proposed for this task combines many simple and moderately accurate categorization rules that are trained sequentially into a single, highly accurate model. AdaBoost is entirely data-driven and requires an adequate amount of data for training. A new modification of boosting is proposed that combines and balances human expertise with available training data. The basic idea of the approach is to modify the loss function used by boosting to balance two terms, one measuring fit to the training data, and the other measuring fit to a human-built model.

For the interpretation of written text, assigning arguments to predicates has been considered as a tagging problem for which various supervised machine learning techniques have been proposed [9,33,37,93]. Some of the features are the predicate, the syntactic category of a phrase and its position with respect to the predicate, the head-world, named entities, other features of the parse tree.

In [93], the parsing problem is formulated as a multi-class classification problem and uses an SVM classifier whose scores are converted to probabilities using a sigmoid function. For each sentence being parsed, an argument lattice is generated. A Viterbi search is performed through the lattice combining the probabilities computed from the SVM output with the LM probabilities, to find the maximum likelihood path.
The issue of combining model-driven grammar-based and data-driven approaches has been considered in [131].
At ATT [4], a *mixture language model* for a multimodal application is described with a component trained with in-domain data and another obtained with data generated by a grammar. Understanding is the recognition of the sequence of predicate|argument tags that maximizes P(T|W) where T is the tag sequence and W the sentence. An approximation is made by considering bigrams and trigrams of tags
At IBM [107], a system is proposed which generates an N-best list of word hypotheses with a dialogue state dependent trigram LM and rescores them with two semantic models. An Embedded context-free semantic Grammar (EG) is defined for each concept and performs concept spotting by searching for phrase patterns corresponding to concepts. As a result, semantic hypotheses are generated by filling a number of slots in a frame representation. Decision among these hypotheses is made based on maximum word coverage.
Trigram probabilities are used for scoring hypotheses with the EG model. Concept tags are placed at the beginning and end of the corresponding phrases in a sequence of word hypotheses. The resulting score of a hypothesis is P(W,C).
A second LM, called Maximum Entropy (ME) LM (MELM), computes probabilities of a word, given the history, using an ME model.
The use of classifier in spoken opinion analysis is described in [5].

7. DIALOG ACT AND SENTENCE CLASSIFICATION

A *speech act* is a dialogue fact expressing an action. Speech acts and other dialog facts to be used in reasoning activities have to be hypothesized from discourse analysis. Different classifiers for speech acts, goal and roles are proposed in [30]. Dialogue acts (DA) are meaningful discourse units, such as statements and questions. Dialogue acts and other dialogue events, such as subjectivity expressions, are related to discourse segments which may contain many sentences. For this reason, in order to make statistical models for DA hypothesization it is useful to introduce features of various types, such as lexical, segment, numerical.
Various techniques have been proposed for DA modeling and detection. Among them, it is worth mentioning semantic classification trees [71], Decision trees [115], hidden Markov models (HMM) [115], fuzzy fragment-class Markov models [137], neural networks [105,115], maximum entropy models [115].
In [105], dialog acts are hypothesized by a search process based on the Viterbi algorithm. There is an HMM source for every dialog act DA which generates sequences of words W. The emission probability is given by:

$$\Pr(W \mid DA) = \frac{\Pr(DA \mid W)\Pr(W)}{\Pr(DA)}$$

and the probability Pr(DA|W) is obtained by a neural network fed by words and prosodic features and trained using the Kullback-Leibler divergence as error measure.
In [120], Pr(DA|W) is obtained from a finite-state model automatically trained using SCTs.
Words and dialogue facts can be related to query communication goals with belief networks [74].
Graphical models are proposed in [47]. The focus is on dynamic Bayesian networks. For joint segmentation and classification of DAs, a technique based on a Hidden-Event Language Model (HELM) is described in [142].
A more accurate event detection is obtained if sentence boundaries are identified in spoken messages containing more than one sentence. Approaches to this task have used Hidden Markov Models (HMM) [110] and Conditional Random Fields (CRF) [67].
Call routing is an important and practical example of spoken message categorization. In applications of this type, the dialog act expressed by one or more sentences is classified to generate a *semantic primitive action* belonging to a well defined set.
A solution to spoken message categorization is proposed in [35]. Knowledge is represented by a network used for mapping words or phrases into actions. The network computes a score for every action hypotheses when fed with words or phrases. Phrases are obtained with grammar fragments. A single-layer association network is considered whose parameters are estimated with a training corpus. An overview of early versions of the *How may I help you* application can be found in [36]. The application has evolved with the introduction of new classification and learning methods.
More recent solutions for document type (and sentence type) hypothesization were proposed using Latent Semantic Analysis (LSA) [7,15].
In [61], discriminative training is proposed for natural language call routers. In [34] a method is proposed for estimating the LM probabilities with a criterion that optimizes end-to-end performance of a natural language call routing system. In [92], the problem of categorical classification of actions from speech input is investigated. A dialogue model is introduced with state-dependent LMs.

8. PROBABILISTIC LOGIC AND INFERENCE FOR SLU

In practical applications, SLU is part of a dialogue system whose objective is the execution of actions to satisfy a user goal. Actions can be executed only if some pre-conditions are asserted true and their results are represented by post-conditions. Preconditions for actions depend on instances of semantic structures composed by previous dialogue actions.
Control strategies for interpretation determine how semantic structures are built, how expectations are defined and how knowledge structures are matched with input data in the presence of constraints and imprecision.
A control strategy can be called *constructive* if it gradually builds data structures using a basic queue called *agenda* where pointers to partial interpretations, called *theories*, are stored in an order dependent on the scores assigned to the theories.
Early approaches to SLU used semantic representations in terms of partitioned semantic networks [127].
In the last two decades, most SLU applications did not require to perform complex semantic compositions and were mostly based

on semantic grammars, shallow semantic parsers and sentence categorization.

Recently, in [21] a logic based solutions was proposed for making inferences about user intensions in telephone applications. A dialogue manager (DM) of a vocal service has a state model. A set of states is active at turn k of a dialogue. The system interprets a dialogue turn message in two phases.

In the first one, a word-to-constituent transducer translates a word lattice into a constituent lattice. In the second phase, a set of precondition-action rules are also encoded as a transducer that transform concept hypotheses into state transitions. Different states can be reached with different probabilities. The N-best states are then processed by DM to determine the next dialogue action.

Instances of constituents can be structured into probabilistic frames. In probabilistic frame-based systems [57], it is possible to have a probability model for a slot value which depends on a slot chain. It is also possible to inherit probability models from classes to subclasses, to use probability models in multiple instances and to have probability distributions representing structural uncertainty about a set of entities. It is shown that it possible to construct a Bayesian network (BN) for a specific instance-based query and then perform standard BN inference.

A general method based on Petri nets for probabilistic inference on frames is proposed in [82].

Methods for probabilistic logic learning are reviewed in [23].

If different logical worlds have to be considered, then possible world probabilities have to be estimated. The computation of probabilities of possible worlds is discussed in [84],[88] (p. 459). A general method for computing probabilities of possible worlds based on Markov logic networks (MLN) is proposed in [104].

9. SEMANTIC CONFIDENCE

The posterior probability $P(\Gamma|Y)$, where Y is a time sequence of acoustic features is not the best reliability indicator for a hypothesis [116]. In fact, acoustic, lexical, language and semantic models introduce various degrees of imprecision. Furthermore, suitable confidence indices should also take into account information that is not coded in Y, such as the coherence of the available hypotheses with the entire dialogue history, including system prompts and repairs.

It is important to design algorithms for computing the probability $P(\Gamma|\Phi_{conf})$ that an interpretation Γ is correct given Φ_{conf} which represents a set of confidence indicators or functions of them.

In [12], important issues related to confidence metric for ASR and SLU are discussed. They refer to the identification of errors and confidence features, feature combination and use, evaluation.

Confidence measures for ASR are reviewed [41]. Confidence measures for in SLU are proposed in [85]. Confidence measures for ASR and SLU are reviewed in [32, 48].

The majority of the approaches share two basic steps:

- generate as many features as possible based on the speech recognition and|or natural language understanding process,
- estimate correctness probabilities with these features

Typically, confidence measures depend on the particular application and its domain.

Using the posterior probabilities, obtained with acoustic and language models, of words supporting the interpretation [64], the probability that a conceptual structure can be evaluated [58].

Lin and Wang [65] propose a concept-based probabilistic verification model, which also exploits concept N-grams.

In order to achieve more accurate scoring depending on the context, in [93] it is proposed to create confidence models for semantic frames using previous system prompts in addition to the features obtained from the speech recognition results.

Among the methods for fusing confidence scores it is worth mentioning Fisher linear discriminant analysis [50], decision trees, neural networks and SVM [141].

In [100] an interpretation strategy implemented by a decision tree is proposed. At a node of the tree, a decision unit DU_j is applied. Node units make decisions based on the values of specific confidence indicators, including the consensus among observations taken from different view points.

Following ideas proposed for committee based active learning [113], some sesemantic confidence indicators are based on the agreement of semantic interpretations obtained by different methods using FSMs, and classifiers of the type SCT, SVM and *boostexter*.

In [108], both word, and concept level confidence annotations are considered. Two methods are proposed that use two sets of statistical features to model the presence of semantic information in a sentence. The first relies on a semantic tree where node and extension scores are used. Scores are based on the assumption that sentences that are grammatically correct and likely to be free of recognition errors tend to be easier to parse and should receive high confidence. The second technique is based on joint maximum entropy modelling words of a sentence and the semantic parse tree. Different maximum entropy techniques are used to combine semantic and lexical information features depending on the type of parsing performed. Lattice based posterior probabilities are combined with semantic features in a probabilistic framework for each word or concept and dialog state information.

Word lattices can be further processed and formatted into a Confusion-Network (CN) structure. In [70], an algorithm for the generation of confusion networks (CN) has been proposed. An alternative CN generation algorithm has been proposed in [40].

Speech recognition systems encounter more difficulties when trying to recognize short words as compared to longer words. In a word lattice, the ASR system tends to generate a large number of hypotheses of the same word with different lengths, start frames and acoustic scores. It is frequent that word hypotheses having significantly different time lengths are grouped in the same class, with short words becoming possible alternatives to much longer words. Following these observations some modifications suitable for confidence evaluation in SLU were proposed in [77].

Error correction is proposed in [98]. In [56], probabilities $P^{Rel}(c_i, c_j)$ of the relations between instantiations of concepts in the same spoken sentence are defined and related to the mutual information of constituent hypotheses in a sentence.

Confidence scoring, has been applied to detect errors in intention recognition results and has proved useful for dialogue management [24,58]. If the detection is successful, the system can safely avoid unnecessary confirmations for reliable slots and put high priority in asking questions about unreliable or unfilled ones. In [43] it is proposed to incorporate discourse features into the confidence scoring of intention recognition results. A number of discourse-related features (called discourse features) are introduced that characterize the contextual adequacy of slot values

In [96], multiple candidate hypotheses from different sources (e.g. deep syntactic parsing and shallow topic classification) are evaluated and assigned overall confidence scores using features at multiple levels (e.g. acoustic, semantic and context-based).
A *discourse coherence* measure [63], based on topic consistency across consecutive utterances is obtained with interutterance distance based on the topic consistency between two utterances The confidence measures are incorporated into the utterance verification framework by combining them in the computation of an overall posterior probability.

10. RECENT RESULTS IN ADAPTIVE LEARNING FOR SLU

Knowledge for SLU is imprecise and incomplete. Once an application is deployed, many errors can be ascribed to SLU knowledge imprecision.
It is useful to adapt systems to fast variations in feature statistics and learn new events with minimum supervision. Instead of assuming a fixed and given training data as in the passive learning used in the approaches reviewed so far, in adaptive learning samples are dynamically determined with automatic methods.
Methods for adaptive learning are active learning, unsupervised learning and their combination.
Part of errors due to SLU knowledge imprecision can be detected by introducing suitable confidence indicators. The corresponding messages can then be used as samples for updating SLU knowledge. Such an activity is known as active learning.
Approaches to active learning rely on two basic method types: certainty-based and committee-based methods. An initial system is developed with *certainty-based methods* using a small set of annotated examples [17]. Such a system is used for interpreting unannotated examples. Confidence indicators are obtained for these examples and the examples with the lowest certainties are proposed to human labelers for annotation. C*ommittee-based methods* consider a set of classifiers trained with a small set of annotated examples [20]. A new set of unannotated instances is presented to the classifiers. The samples for which different classifiers provide the most different interpretations are selected for human inspection and annotation.
Applications of certainty-based learning to sentence classification are proposed in [38].
With committee-based learning better results were obtained for sentence classification using SVM and Boostexter classifiers [124].
A committee-based method, which is applicable to multiview problems (i.e., problems with several sets of uncorrelated attributes that can be used for learning is *co-testing* [79]. In co-testing, the committee of classifiers is trained using different views of the data.
In [39], a method is proposed in which a bootstrapped model is built with selected samples of relevant text obtained by transcriptions from conversational systems and data retrieved from web sites. The boostrap model is updated by an iterative process which combines unsupervised and active learning. Unsupervised learning involves decoding followed by model building. This is implemented by co-training with the assumption that there are multiple views for classification. Multiple models are trained using the views. Unlabelled data are classified with all the models. The training set of a classifier is then composed using other classifier's predictions. A confidence score is computed for active learning and used to select utterances for manual annotation.
In [102], an active learning method is proposed based on *selective sampling* and *error rate prediction* as function of the training examples.
Interpretation model adaptation is proposed in [125].
A multitask learning method is presented in [126] for natural language intent classification. The already labelled data are reused across applications while training so that collaboration among methods improves learning results.
LM adaptation to the prediction of concepts is proposed in [59].
Discriminative training of acoustic and language models using the Maximum Mutual Information (MMI) or Minimum Classification Error (MCE) criteria have been used for language model adaptation in spoken dialogue systems. The learning objective in SLU systems is to minimize concept error rate which does not reduce to minimizing word error rate which has been the objective of previous MMI and MCE applications [132]..

11. FUTURE PERSPECTIVES

History of SLU has shown an evolution from the use of high precision, non-probabilistic, low coverage semantic human-crafted KSs to the introduction of modular, complex, probabilistic KSs some of them obtained with automatic learning using manually annotated corpora. Research on new KS paradigms, probabilistic frames and inference for SLU will provide more effective KSs.
In the future, it will be interesting to consider a more careful evaluation of cost and performance of manual annotation vs. manual composition of KSs.
Modular architectures should implement cooperation between KSs making an effective use of human linguistic knowledge, machine learning algorithms, linguistic resources, available data.
Optimal or sub-optimal, parallel or sequential decision strategies should use system capabilities to assess the confidence of the interpretations they produce.
In spite of the imprecision of the modules used in the SLU chain, it is possible to develop useful applications in limited domain. Thanks to effective confidence indicators, SLU results can be evaluated. Specific dialog actions can be performed when confidence is not high. By switching to a human operator when verification is not satisfactory, it is possible to achieve, in some cases, good automation rates.
Better KSs and strategies will make it possible to consider domains more complex that the ones of today applications.

12. ACKNOWLEDGEMENTS

This work was supported by the European Union (EU), Project LUNA, IST contract no 33549.

13.REFERENCES

1. P. Abney. (1991) Parsing by chunks. In R. C. Berwick, S. P. Abney, and C. Tenny, eds, *Principle-Based Parsing: Computation and Psycholinguistics,* pp. 257-278. Kluwer, Dordrecht.
2. J. Allen (1987). *Natural language Understanding*, The Benjamin/Cummings Publishing Company, Menlo Park CA.
3. C. Baker, C. Fillmore, and J. Lowe. 1998. The Berkeley Framenet project. COLING-ACL- 1998.
4. S. Bangalore and M. Johnston (2004) Balancing data-driven and rule-based approaches in the context of a multimodal conversational system. HLT-NAACL, pp. 33-40

5. F. Béchet , G. Damnati, F., N. Camelin, R. De Mori (2006) Spoken opinion extraction for detecting variations in user satisfaction IEEE/ACL Workshop on Spoken Language technology, Aruba.
6. J. R. Bellegarda, (2000) Large vocabulary speech recognition with multi-span statistical language models", *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 1, pp. 76-84, Jan. 2000.
7. J.R. Bellegarda and K.E.A. Silverman (2001) Data-driven semantic inference for unconstrained desktop command and control.Eurospeech, Aalborg, Denmark, pp.455-459
8. E. Black, F. Jelinek, J. D. Lafferty, D. M. Magerman, R. L. Mercer, S. Roukos (1993) Towards History-Based Grammars: Using Richer Models for Probabilistic Parsing. ACL, p. 31-37
9. D. Blaheta and E. Charniak. 2000. Assigning function tags to parsed text. *NAACL*, pp. 234–240.
10. R. Bobrow; R. Ingria and D. Stallard (1990). Syntactic and semantic knowledge in the DELPHI unification grammar. Proc. *Speech and Natural language Workshop* : 230-236, Hidden Valley, PA, Morgan Kaufmann Inc., Palo Alto, CA.
11. R. Bod (2000) Combining Semantic and Syntactic Structure for Language Modeling. ICSLP, Beijing, China
12. L. Chase (1997) Error-responsive feedback mechanisms for speech recognition. Ph.D. thesis, Carnegie Mellon Univ., Pittsburgh, PA,USA
13. E. Charniak, (2001), Immediate-head parsing for language models. ACL, pp. 116–123.
14. C. Chelba and F. Jelinek (2000) Structured language modeling. *Computer Speech and Language*, 12-4:283-332
15. J. Chu-Carroll and B. Carpenter (1999) Vector-based natural language call routing, *Computational. Linguistics* 25, (3), pp. 361–388.
16. N. Coccaro and D. Jurafsky (1998) Towards better integration of semantic predictors in statistical language modelling. ICSLP, pp. 2403-2406.
17. D. Cohn, L. Atlas, and R. Ladner, (1994) Improving generalization with active learning, *Machine Learning*, vol. 15, pp. 201–221,
18. Collins, M. (1999) Head-Driven Statistical Models for Natural Language Parsing. PhD thesis, University of Pennsylvania, Philadelphia, PA, USA.
19. Corazza, R. De Mori, R. Gretter, . and G. Satta (1994). Optimal Probabilistic Evaluation Functions for Search Controlled by Stochastic Context-free Grammars. *IEEE Trans. on Pattern Analysis and Machine Intelligence*,16 (10) : 1018-1027.
20. I. Dagan and S. P. Engelson,(1995) Committee-based sampling for training probabilistic classifiers. 12th Int. Conf. Machine Learning, pp. 150–157
21. G. Damnati, F. Bechet, R. de Mori, (2007) Spoken language understanding strategies on the France Telecom 3000 voice agency corpus, IEEE ICASSP, Honolulu, Hawaii
22. R. De Mori, (1998) *Spoken dialogues with computers*. Academic Press.
23. L. De Raedt and K. Kersting (2003) Probabilistic Logic Learning. ACM SIGKDD exploration newsletter, (5):1.
24. K. Dohsaka, N. Yasuda and F. Aikawa,, (2003) Efficient spoken dialogue control depending on the speech recognition rate and system database. Eurospeech, Geneva, Switzerland, pp. 657–660.
25. J. Dowding, J. Gawron, et al., 1993). Gemini: A Natural Language System for Spoken-Language Understanding. *Spoken Language Systems Technology Workshop,*MIT, Cambridge, Mass., pp. 20-23.
26. D. Dowty (1979). *Word meaning and Montague grammar*. Reidel, Dordrecht, the Netherlands.
27. E.W. Drenth and B. Ruber Context-dependent probability adaptation in speech understanding. *Computer Speech and Language*, 11(3):225-252
28. H Erdogan, R. Sarikaya, S.F. Chen, Y. Gao and M. Picheny (2005) Using Semantic Analysis to improve Speech Understanding Performance. *Computer Speech and Language*, 19(3):321-344.
29. L. D. Erman, F. Hayes-Roth, V. R. Lesser et R. D. Reddy.(1980) The Hearsay-II Speech Understanding System : Integrating Knowledge to Resolve Uncertainty. *ACM Computing Surveys*, 12(2):213-253.
30. J. Eun, M. Jeong, G. Geunbae Lee (2005) A Multiple Classifier-based Concept-Spotting Approach for Robust Spoken Language Understanding. Eurospeech, Lisbon, Portugal, pp. 3441-3444
31. C. J. Fillmore (1968). The case for case. in E. Bach and R. Harms eds. *Universals in linguistic theory*, Holt, Rinehart and Winston, New York, 1968.
32. S. Furui, (2003) Robust Methods in Automatic Speech Recognition and Understanding Proc. Eurospeech, Geneva, Switzerland, pp. 1993,1998.
33. D. Gildea and D. Jurafsky. (2002). Automatic labeling of semantic roles. *Computational Linguistics*, 28(3):245–288.
34. V. Goel, H. K. J. Kuo, S. Deligne, C. Wu (2005) Language model estimation for optimizing end-to-end performance of a natural language call routing system Proc. IEEE ICASSP, Philadelphia, PA, USA, pp. I 565-568
35. L. Gorin (1995) On automated language acquisition. *Journal of Acoustical Society of America*, 97(6):3441-2461
36. L. Gorin, G. Riccardi, and J. H. Wright, (1997) How may I help you? *Speech Communication*, vol. 23, no. 1–2, pp. 113–127.
37. K. Hacioglu and W. Ward. (2003) Target word detection and semantic role chunking using support vector machines. HLT-HLT-NAACL, Edmonton, Alberta, Canada
38. D. Hakkani-tur, g. Riccardi and a. Gorin (2002) Active learning for automatic speech recognition. IEEE ICASSP, Orlando, FLA, USA
39. D. Hakkani-Tur G. Tur, M. Rahim G. Riccardi (2004) Unsupervised and active learning in automatic speech recognition for call classification IEEE ICASSP, Montreal, Que, Canada, pp.I-429-432.
40. D. Hakkani-Tur, F. Bechet, G. Riccardi and Gokhan Tur (2006) Beyond ASR 1-Best: Using Word Confusion Networks for Spoken Language. Understanding, *Computer Speech and Language* 20(4):495-514.
41. T. Hazen, T. Burianek, J. Polifroni, and S. Seneff (2000) Recognition confidence scoring for use in speech understanding systems, *Automatic Speech Recognition Workshop*, Paris, France, pp. 213–220.
42. Y. He and S. Young (2006) Spoken language understanding using the Hidden Vector State Model. *Speech Communication* 48, 262–275
43. R. Higashinaka, N. Miyazaki, M. Nakano and K. Aikawa, (2004). Evaluating discourse understanding in spoken dialogue systems. *ACM Trans. Speech and Language Processing* 1 (1), 1–20.
44. R. Jackendoff (1990). *Semantic Structures*, The MIT Press, Cambridge Mass.
45. R. Jackendoff (2002). *Foundations of language*, Oxford University Press, Oxford UK.
46. E., Jackson, D. Appelt, J. Bear, R. Moore and A. Podlozny (1991). A template matcher for robust natural language interpretation. *Speech and Natural language Workshop*, 190-194, Morgan Kaufmann, Los Altos, CA., USA
47. G. Ji and J. Bilmes (2005) Dialog act tagging using graphical models. IEEE ICASSP, pp.I 33-36.
48. H. Jiang (2005) Confidence measures for speech recognition: a survey. *Speech Communication*, 45(4):455-470.
49. N. Kambhatla (2004) Combining Lexical, Syntactic, and Semantic Features with Maximum Entropy Models for Extracting Relations, ACL, pp. 177-180 (poster),, Barcelona, Spain
50. S.O. Kamppari, T.J. Hazen (2000). Word and phone level acoustic confidence scoring. IEEE ICASSP, Istambuk, Turkey, pp. 1799–1802.
51. R.T. Kasper and E.H. Hovy (1990). Performing integrated syntactic and semantic parsing using classification. *Speech and Natural language Workshop*, 54-59, Hidden Valley, PA, Morgan Kaufmann, Los Altos, CA.,USA

52. R. J. Kate and R. J. Mooney (2007) Semi-Supervised Learning for Semantic Parsing using Support Vector Machines. NAACL HLT 2007, pp. 81–84,
53. T., Kawahara,K. Tanaka and S. Doshita (1999) Virtual fitting room with spoken dialogue interface. *ESCA Workshop on Interactive Dialog in Multi-Modal Systems*, Kloster Irsee Germany, pp.5-8.
54. K. Kipper, H. T. Dang, and M. Palmer. 2000. Class based construction of a verb lexicon. AAAI 2000
55. D.H. Klatt (1977), "Review of the ARPA speech understanding project". *Journal of the Acoustical Society of America*, 62(6), pp. 2405-2420.
56. C. Kobus, G. Damnati, L. Delphin-Poulat, R. De Mori (2006) Exploiting semantic relations for a Spoken Language Understanding application, ICSLP, Pittsburgh, Pennsylvania, PA, USA. pp. 1029-1032
57. D. Koller and A. Pfeffer (1998) Probabilistic frame-based systems. AAAI98, pp. 580–587, Madison, Wisc., USA.
58. K. Komatani, T. Kawahara (2000) Flexible mixed-initiative dialogue management using concept-level confidence measures of speech recognizer output. COLING, Vol. 1, pp. 467–473.
59. R. Kneser, J. Peters (1997) Semantic Clustering for Adaptive Language Modeling. ICASSP'97, Munich, Germany, p. 779
60. R. Kuhn and R. De Mori (1995). The Application of Semantic Classification Trees to Natural Language Understanding. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17 : 449-460.
61. H.K.J. Kuo and C.H. Lee (2003) Discriminative training of natural language call routers. *IEEE Trans. on Speech and Audio Processing*, SAP-11(1):24-35
62. J. Lambek (1958). The mathematics of sentence structure. *American mathematical monthly*, 65 : 154-170.
63. I. Lane and T. Kawahara (2005) Utterance Verification Incorporating In-domain Confidence and Discourse Coherence Measures. Eurospeech, Lisbon Portugal, pp.421-424
64. R. Lieb, T. Fabian, G. Ruske and M. Thomae (2004) Estimation of Semantic Confidences on Lattice Hierarchies. ICSLP, Jeju Island, Korea,
65. Y.C. Lin and H.M. Wang (2001) Probabilistic concept verification for language understanding in spoken dialogue systems. Eurospeech, Aalborg, Denmark, pp. 1049–1052.
66. Wei-Lin Wu, Ru-Zhan Lu, Hui Liu, Feng Gao (2006) A Spoken Language Understanding Approach Using Successive Learner, ICSLP, Pittsburg, PA, USA, pp.1906-1909
67. Y. Liu, A. Stolcke, E. Shriberg, and M. Harper (2005) Using conditional random fields for sentence boundary detection in speech. *ACL*, pp. 451–458.
68. M. Palmer, D. Gildea, and P. Kingsbury.(2003) The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics*.
69. K. Macherey, F.J. Och and H. Ney (2001) Natural language understanding using statistical machine translation. Eurospeech , Aalborg, Denmark, pp. 2205-2208
70. L. Mangu, E. Brill, and A. Stolcke (2000) Finding Consensus in Speech Recognition: Word Error Minimization and Other Applications of Confusion Networks. *Computer Speech and Language* 14(4):373-400.
71. M. Mast et al, (1996) Dialog act classification with the help of prosody. ICSLP, , Philadelphia, PA, USA.
72. J. McCarty and P. J. Hayes (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, Ed. by B. Meltzer and D. Michie, Edinburgh University Press.
73. M. McTear (2006) Spoken language understanding for conversational dialog systems, IEEE/ACL Workshop on Spoken Language Technology Aruba.
74. H.M. Meng, W. Lam and C. Wai (1999). To believe is to understand. Eurospeech, Budapest, Hungary.
75. G.A. Miller (1995) WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39-41
76. S. Miller, R. Bobrow et al (1994). Statistical Language Processing Using Hidden Understanding Models. *Spoken Language Technology Workshop,* 48-52, Plainsboro, New Jersey, Los Altos, CA., USA
77. B. Minescu, G Damnati, F. Béchet, R. De Mori (2007) Conditional use of Word Lattices, Confusion Networks and 1-best string hypotheses in a Sequential Interpretation Strategy. Interspeech, Antwerpen, Belgium
78. R. Montague (1974). *Formal Philosophy*. Yale University press, New Haven,Conn., USA
79. A. Muslea, (2000) Active Learning with MultipleViews, Ph.D. dissertation, Univ. Southern California, Los Angeles, CA, USA
80. A. Nagai, Y. Ishikawa and K. Nakajima (1994) A semantic interpretation based on detecting concepts for spontaneous speech understanding. ICSLP, Yokohama, Japan, pp. 95-98.
81. S. Narayanan. (1999). Moving right along: A computational model of metaphoric reasoning about events. AAAI Menlo Park, CA, USA.
82. S. Narayanan. (1999) Reasoning about actions in narrative understanding. IJCAI. Morgan Kaufmann Press.
83. A. Nasr, Y. Estéve, F. Béchet, T. Spriet, R. De Mori (1999) A language model combining n-grams and stochastic finite state automata. Eurospeech, Budapest, Hungary, pp :2175-2178
84. N. Nilsson (1986) Probabilistic logic, *Artificial Intelligence* 28: 71-87,
85. C. Pao, P. Schmid, and J. Glass, (1998) Confidence scoring for speech understanding systems, ICSLP, Sydney, NSW, Australia.
86. K.A. Papieni, S. Roukos and R.T. Ward R.T. (1998) Maximum likelihood and discriminative training of direct translation models. IEEE ICASSP, Seattle WA
87. P. F. Patel-Schneider, P. Hayes and I. Horrocks (2003) OWL Web Ontology Language Semantics and Abstract Syntax, W3C working Draft.
88. J. Pearl (1988) Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA, USA.
89. F. Pereira (1990) Finite-state approximations of grammars Speech and Natural language workshop, Hidden Valley, PA, pp. 12-19
90. R. Pieraccini, E. Levin and C.H.Lee (1991). Stochastic Representation of Conceptual Structure in the ATIS Task. Proceedings of the, 1991 *Speech and Natural Language Workshop*, 121-124, Los Altos, CA.
91. R. Pieraccini, E. Levln, E. Vidal. (1993) Learning How To Understand Language. Eurospeech, pp. 1407-1412. Berlin, Germany
92. A. Potamianos, S. Narayanan and G. Riccardi,(2005) Adaptive Categorical Understanding for Spoken Dialogue Systems *IEEE Trans; on Speech and Audio Processing*,SAP-13 (2) : 321-329
93. S. S. Pradhan, H. Ward, K. Hacioglu, J.H. Martin and D. Jurafsky (2004) Shallow Semantic Parsing using Support Vector Machines. HLT-NAACL, Boston, Mass., USA, pp.233-240.
94. S. S. Pradhan, H. Ward, and J.H. Martin (2007) Towards Robust Semantic Role Labeling. NAACL HLT, Rochester NY, USA pp. 556–563,
95. N. Prieto, E. Sanchis and L. Palmero (1994) Continuous speech understanding based on automatic learning of acoustic and semantic models. ICSLP, Yokohama, Japan
96. M. Purver, F. Ratiu, L. Cavedon (2006) Robust Interpretation in Dialogue by Combining Confidence Scores with Contextual Features. ICSLP, Pittsburgh, PA, USA, pp. 1-4
97. O. Rambow, S. Bangalore, T. Butt, A. Nasr, R. Sproat (2002) Creating a Finite State Parser with Application Semantics. COLING, Taipei
98. C. Raymond, F. Béchet , N. Camelin, R. De Mori, G. Damnati (2005) Semantic interpretation with error correction IEEE ICASSP, Philadelphia, PA, USA
99. C. Raymond, F. Béchet R. de Mori and G. Damnati, (2006) On the use of finite state transducers for semantic interpretation, *Speech Communication*, 48(3): 288-304.

100. C. Raymond, F. Béchet, N. Camelin, R. De Mori and G. Damnati (2007) *Sequential decision strategies for machine interpretation of speech*, *IEEE Trans. on Speech and Audio Processing*, 15(1):162-171.
101. G. Riccardi, R. Pieraccini and E. Bocchieri (1996) Stochastic automata for language modeling. Computer Speech and Language, 10(4):265-293.
102. G. Riccardi and D. Hakkani-Tur (2005) Active Learning: Theory and Applications to Automatic Speech Recognition *IEEE Trans. on Speech and Audio Processing*,SAP-13 (4) : 534-545
103. S.D. Richardson, W. B. Dolan, and L. Vanderwende. (1998). MindNet: Acquiring and Structuring Semantic Information from Text. ACL-COLING, Montreal, Canada, pp. 1098-1102.
104. M. Richardson and P. Domingos (2006) Markov Logic Networks, *Machine Learning*, 62:107-136.
105. K. Ries (1999) HMM and neural network based speech act detection. IEEE ICASSP Phoenix, AZ, USA.
106. B. Roark (2001) Probabilistic top-down parsing and language modeling. *Computational Linguistics* 27 (2), 2–28.
107. R. Sarikaya, Y. Gao and M. Picheny (2004) A Comparison of Rule--Based and Statistical Methods for Semantic Language Modelling and Confidence Measurement. HLT-NAACL, Boston, Mass, USA, pp. 65-68
108. R. Sarikaya, Y. Gao, M. Picheny and H. Erdogan. (2005) Semantic Confidence Measurement for Spoken Dialog Systems *IEEE Trans. on Speech and Audio Processing*, 13 (4) : 534-545.
109. R. E. Schapire, M. Rochery, M. Rahim, and N. Gupta (2005) BoostingWith Prior Knowledge for Call Classification *IEEE Trans. on Speech and Audio Processing*,SAP-13 (2) : 174-182
110. E. Shriberg, A. Stolcke, D. Hakkani-Tur, and G. Tur, (2000) Prosody-based automatic segmentation of speech into sentences and topics, *Speech Communication*, pp. 127–154, 2000.
111. S. Seneff S. (1989). TINA: A Probabilistic Syntactic Parser for Speech Understanding Systems. IEEE ICASSP, 2 : 711-714. Glasgow, UK.
112. S. Seneff (1992). A Relaxation Method for Understanding Spontaneous Speech Utterances. Proceedings of the, 1992 *Speech and Natural Language Workshop*, Los Altos, CA.
113. H. S. Seung, M. Opper, H. Sompolinsky (1992) Query by Committee. COLT 1992 287-294
114. Y. Shabes Y. and A. K. Joshi (1990). Two recent developments in tree adjoining grammars: Semantic and efficient processing.. *Speech and Natural language Workshop*, 48-53, Los Altos, CA.
115. A. Stolcke *et al.*, Dialog act modelling for conversational speech. AAAI Spring Symp. on Appl. Machine Learning to Discourse Processing, 1998, pp. 98–105.
116. K. Sudoh and M. Nakano (2005) Post-dialogue confidence scoring for unsupervised statistical language model training. *Speech Communication*, 45(4):387-400).
117. K. Sudoh and H. Tsukada (2005) Tightly Integrated Spoken Language Understanding using Word-to-Concept Translation. Eurospeech, Lisbon Portugal, pp.429-432.
118. J. Suzuki, H. Isozaki and E. Maeda (2004) Convolution Kernels with Feature Selection, for Natural Language Processing Tasks. ACL, pp. 120-127, Barcelona, Spain
119. R. S. Swier, and S. Stevenson (2004) Unsupervised Semantic Role Labelling. EMNLP ,Barcelona, Spain, pp. 95—102
120. K. Tanigaki and Y. Sagisaka (1999) Robust speech understanding based on word graph interface. *ESCA Workshop on Interactive Dialog in Multi-Modal Systems*, Kloster Irsee Germany, pp. 45-48, 1999.
121. M. Thomae, T. Fabian, R. Lieb and G. Ruske (2005) Hierarchical Language Models for One-Stage Speech Interpretation. Eurospeech, Lisbon, Portugal, pp. 3425-3428
122. M. Tomita (1986) An Efficient Word Lattice Parsing Algorithm for Continuous Speech Recognition, IEEE ICASSP, Tokyo, Japan, p. 330
123. K. Toutanova, A. Haghighi, C. D. Manning (2005) Joint learning improves semantic role labeling. ACL, Ann Arbor, Michigan, USA. pp.: 589 - 596
124. G. Tur, R.E. Shapire and D. Hakkani-Tur (2003) Active learning for spoken language understanding. IEEE ICASSP, Hong Kong, China, pp. I-275,279
125. G. Tur (2005) Model adaptation for spoken language understanding. IEEE ICASSP, Philadelphia, PA., USA, pp.I 41-44
126. G. Tur (2006) Multitask learning for spoken language understanding. IEEE ICASSP, Toulouse (France), pp.585-588
127. Walker, D. (1975) The SRI speech understanding system. *IEEE Trans. On Acoustics, Speech, And Signal Processing*, ASSP-23, NO- *5,* pp. 397-416
128. D L. Waltz (1981) Toward a Detailed Model of Processing for Language Describing the Physical World. IJCAI, Vancouver (BC), Canada, pp. 1-6
129. K. Wang (2004) A detection based approach to robust speech understanding IEEE ICASSP, Montreal Canada, May, pp.I-413-416
130. W.Wang, Y. Liu and M.P. Harper, (2002) Rescoring effectiveness of language models using different levels of knowledge and their interaction. IEEE ICASSP, Orlando, FLA, USA, pp785-789
131. Y.Y. Wang and A. Acero (2003) Combination of CFG and N-Gram Modeling in Semantic Grammar Learning Eurospeech, Geneva, Switzerland
132. Y.Y. Wang and A. Acero (2006) Discriminative Models for Spoken Language Understanding. ICSLP, Pittsburg, PA., USA, pp. 2426-2429
133. W. Ward and S. Issar(1994). Integrating Semantic Constraints into the Sphinx-II Recognition Search. IEEE ICASSP,pp. 17-19., Adelaide, Australia.
134. W.A. Woods, (1970) Transition Network Grammars for Natural Language Analysis, *Communications of the ACM*, Vol. 13:10.
135. W.A. Woods (1975). What's in a link? in D.G. Bobrow and A. Collins Eds*, Representation and understanding* , Academic Press, New York.
136. W.A. Woods, et al (1976) Speech Understanding Systems. Bolt, Beranek and Newman Inc., Cambridge, MA., USA, Final Report, Vol. IV, V
137. C.-H.Wu, G.-L. Yan, and C.-L. Lin, (2002) Speech act modelling in a spoken dialog system using a fuzzy fragment-class Markov model. *Speech Communication*, 38 (1-2), pp. 183– 199, 2002.
138. Wutiwiwatchai and S. Furui (2006) A multi-stage approach for Thai spoken language understanding. *Speech Communication* 48 305–320
139. S. R. Young, A.G. Hauptmann, W. H. Ward, E.T. Smith and P. Werner (1989). High level knowledge sources in usable speech recognition systems. *Communications of the ACM*, 32 (2) : 183-194.
140. K. Zechner(1998) Automatic construction of frame representations for spontaneous speech in unrestricted domain. ACL-COLING, Montreal, Canada, pp. 1448-1452.
141. R. Zhang and A. I. Rudnicky (2002) Improve Latent Semantic Analysis based Language Model by Integrating Multiple Level Knowledge. ICSLP, Denver, CO, USA.
142. M. Zimmermann and et al., (2005) Toward joint segmentation and classification of dialog acts in multi-party meetings. *2nd MLMI*, Edinburgh, UK.
143. Lytinen S. (1992). Semantic-first natural language processing. Proc. *National Conference on Artificial Intelligence* , 111-116 San Jose CA.
144. Tait J..I. (1983). Semantic Parsing and syntactic constraints. In K. Sparck Jones and Y. A. Wilks Eds. *Automatic natural language parsing.* Ellis Horwood/Wiley, Chichester.
145. M. Mohri (1997) Finite-state transducers in language and speech processing Computational Linguistics, 23(2):270-310