Spoken Language Understanding; a survey

Renato De Mori

LUNA IST contract no 33549

IEEE ASRU Kyoto Dec 11th, 2007
Summary

• THE SIGN TO MEANING PROCESS

• WORDS TO CONCEPTS (SEMANTIC CONSTITUENTS) TRANSLATION

• SEMANTIC GRAMMARS

• SEMANTIC COMPOSITION AND INFERENCE

• CONFIDENCE, CORPORA ANNOTATION AND LEARNING
THE SIGN TO MEANING PROCESS
**Introduction**

Epistemology, the science of knowledge, considers a datum as basic unit.

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

Computer epistemology deals with observable facts and their representation in a computer.

Natural language interpretation by computers performs a conceptualization of the world using computational processes for composing a meaning representation structure from available signs and their features.
Some problems and challenges in SLU

• meaning representation,

• definition and representation of signs,

• conception of relations between signs and meaning and between instances of meaning,

• processes for sign extraction, generation of hypotheses about units of meaning and constituent composition into semantic structures,

• robustness and evaluation of confidence for semantic hypotheses,

• automatic learning of relations from annotated corpora,

• collection and semantic annotation of corpora.
SLU and NLU share the goal and some types of signs of obtaining a conceptual representation of natural language sentences.

Specific to SLU is the fact that

• signs to be used for interpretation are coded into signals with other information such as speaker identity.

• spoken sentences often do not follow the grammar of a language; they exhibit self corrections, hesitations, repetitions and other peculiar phenomena.

• SLU systems contain an ASR component and must be robust to noise due to the spontaneous nature of spoken language, errors introduced by ASR and its difficulty in detecting sentence boundaries.
Meaning representation

Semantic theories have inspired the conception of *Meaning Representation Languages* (MRL).

MRLs have a syntax and a semantic (Woods, 1975) and should, among other things:

represent **intension** and **extension**, with defining and asserting properties, use **quantifiers** as higher operators, lambda abstraction And make it possible to perform **inference**

**Frame** languages define computational structures (Kifer et al., JACM, 1995) and can be seen as **cognitive structuring devices** (Fillmore, 1968, 1985) in a semantic construction theory.
Frames as computational structures (intension)

A frame scheme with defining properties represents types of conceptual structures (intension) as well as instances of them (extension). Relations with signs can be established by attached procedures (S. Young et al., 1989).

{address
  loc   TOWN
        .......attached procedures
  area  DEPARTMENT OR PROVINCE OR STATE
        .......attached procedures
  country  NATION
            .......attached procedures
  street  NUMBER AND NAME
            .......attached procedures
  zip    ORDINAL NUMBER
            .......attached procedures
}

IEEE ASRU
A convenient way for asserting properties, and reasoning about semantic knowledge is to represent it as a set of logic formulas.

\[(\exists x) \left\{ \text{instance of (x, address)} \land \text{loc(x, Avignon)} \land \text{area(x, Vaucluse)} \land \text{country(x, France)} \land \text{street(x, 1 avenue Pascal)} \land \text{zip(x, 84000)} \right\}\]

A frame instance (extension) can be obtained from predicates that are related and composed into a computational structure.

Frame schemata can be derived from knowledge obtained by applying semantic theories.

Interesting theories can be found, for example in (Jackendoff, 1990, 2002) or in (Brackman 1978, reviewed by Woods 1985)
Schemata contain collections of properties and values expressing relations. A property or a role are represented by a slot filled by a value.

\{a0001
  instance_of
  loc
  area
  country
  street
  zip
  address
  Avignon
  Vaucluse
  France
  1, avenue Pascal
  84000
\}
Semantic networks
Entity relations plus structural descriptions represented by logic formulas are proposed in KL-ONE (Brachman, 1978).
Process overview

An integrated solution: the blackboard architecture (Erman et al., ACM Comp. Surveys 1980)

- Learning
  - Long Term Memory: AM, LM, interpretation, KSs
  - Speech to conceptual structures and MRL
  - Short Term Memory
    - Signs
    - Words
    - Concept tags
      - Concept structures
        - MRL description
  - Dialogue
Interpretation problem decomposition

Problem reduction representation is context-sensitive

Interpretation is a composite decision process. Many decompositions are possible involving a variety of methods and KSs, suggesting to consider a modular approach to process design.

Robustness is obtained by evaluation and possible integration of different KSs and methods used for the same sub-task.
Levels of processes and application complexity

Translation from words to basic conceptual constituents

Semantic composition on basic constituents

Context-sensitive validation

Combination of level processes may depend on the application
Hypothesize a lattice of concept tags for semantic constituents and compose them into structures. Detection vs. translation.
WORDS TO CONCEPTS (SEMANTIC CONSTITUENTS) TRANSLATION
Generation of semantic constituent hypotheses

Bien alors donc c'est d'accord j'en je voudrais réserver...

null
{}  
response
{oui}
command-tache
{reservation}

...du quatre au sept avril dans cet hotel à le cap sud

temps-date
{04/04}
temps-date
{07/04}
objetBB
{hotel}
nom-hotel
{cap sud}

spoken sentence  
concepts  
attributes  
spoken sentence  
concepts  
attributes
ASR algorithms compute probabilities of word hypotheses using finite state language models.

It is important to perform interpretation from a lattice of scored words and to take, possibly redundant, word contexts into account (Drenth and Ruber, 1997, Nasr et al., 1999). Other interesting contributions are in (Prieto et al., 1993, Kawahara et al., 1999).

Finite state approximations of context-free or context-sensitive grammars (Pereira, 1990, reviewed in Erdogan, 2005), Finite state parser (TAG) with application semantics (Rambow et al. 2002).
This architecture is used also for separating in domain from out domain message segments (Damnati, 2007) and for spoken opinion analysis (Camelin et al., 2006). The whole ASR knowledge models in this way a relation between signal features and meaning.
noise tolerant models
Hypothesis generation from lattices

An initial ASR activity generates a word graph (WG) of scored word hypotheses with a generic LM.

The network is composed with WG resulting in the assignment of semantic tags to paths in WG

\[ SEMG = WG \circ \left( \sum_{c=0}^{C} Y CLM_c \right) \]

\[ SWG = OUTPROJ(SEMG) \]

(Special issue Speech Communication, 3 2006, Béchet et al., Furui)
Word graph WG

circa 20 euros du Trocadéro
autour/NEAR
/RANGE

du/€

deu/€

vingt/€

un/€

€/€

Trocadéro/€

Trocadéro/<PLACE>

euros/<PRICE>

Place IN([Thing LOC(type: square, value: Trocadéro)])
In (Papineni et al., 1998) statistical translation models are used to translate a source sentence $S$ into a target, artificial language $T$ by maximizing the following probability:

$$
Pr(T|S) = \frac{Pr(S|T)P(T)}{Pr(S)}
$$

The central task in training is to determine correlations between groups of words in one language and groups of words in the other. The source channel fails in capturing such correlations, so a direct model has been built to directly compute the posterior probability $P(T|S)$.

Interesting solutions also in (Macherey et al., 2001, Sudoh and Tsukada, 2005 for attribute/value pairs, LUNA)
Possibility of having features from long-term dependences

Results for LUNA from Riccardi, Raymond, Ney, Hann

\[ p(y \mid x) = \frac{1}{Z(x)} \exp \left( \sum_{c \in C} \sum_{k} \lambda_k f_k (y_{i-1}, y_i, x, i) \right) \]

\[ Z(x) = \sum_{y} \exp \left( \sum_{c \in C} \sum_{k} \lambda_k f_k (y_{i-1}, y_i, x, i) \right) \]

\[ f_k (y_{i-1}, y_i, x, i) = \begin{cases} 
 1 & \text{if } y_i = \text{ARRIVECITY} \\
 0 & \text{otherwise} 
\end{cases} \]

and \( x_i \ldots x_{i-1} \) contain{arrive\to}
Method comparison and combination

- Results on the French MEDIA corpus, LUNA project, NLU RWTH Aachen results
- Approaches:
  - Linear chain CRF
  - FST
  - SVM
  - Log-linear on positional level
  - MT
  - SVM with tree kernel

Comparison

<table>
<thead>
<tr>
<th>model</th>
<th>attribute CER [%]</th>
<th>attribute VALUE CER [%]</th>
<th>attribute VALUE SER [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRF</td>
<td>11.8</td>
<td>16.2</td>
<td>23.0</td>
</tr>
<tr>
<td>log-linear</td>
<td>14.9</td>
<td>19.3</td>
<td>26.4</td>
</tr>
<tr>
<td>FST</td>
<td>17.9</td>
<td>21.9</td>
<td>28.1</td>
</tr>
<tr>
<td>SVM</td>
<td>18.5</td>
<td>22.2</td>
<td>28.5</td>
</tr>
<tr>
<td>MT</td>
<td>19.2</td>
<td>23.3</td>
<td>27.6</td>
</tr>
</tbody>
</table>

Incremental oracle performance

<table>
<thead>
<tr>
<th>model</th>
<th>attribute CER [%]</th>
<th>attribute VALUE SER [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRF</td>
<td>11.8</td>
<td>17.7</td>
</tr>
<tr>
<td>+log-linear</td>
<td>9.8</td>
<td>15.6</td>
</tr>
<tr>
<td>+FST</td>
<td>8.3</td>
<td>13.8</td>
</tr>
<tr>
<td>+SVM</td>
<td>7.6</td>
<td>12.9</td>
</tr>
<tr>
<td>+MT</td>
<td>7.0</td>
<td>12.3</td>
</tr>
</tbody>
</table>
Sequential approach with 1-best ASR

Comparison of interpretation results obtained in the MEDIA corpus 1 best ASR output

Concept error rate (CER)

Conditional Random Fields 25.2 %
Finite State Transducers 29.5 %
Support Vector Machines 29.6 %

CER close to 20 when N-best concepts (N<10) are obtained with FSMs. Possibility of further improvement by combination with CRFs and using dialog constraints
History

Systems developed in the seventies reviewed in (Klatt, 1977) and the eighties, early nineties (EVAR, SUNDIAL) mostly performed syntactic analysis on the best sequence of words hypothesized by an ASR system and used non probabilistic rules, semantic networks, pragmatic and semantic grammars for mapping syntactic structures into semantic ones expressed in logic form.

In the nineties, the need emerged for testing SLU processes on large corpora that could also be used for automatically estimating some model parameters. Probabilistic finite-state interpretation models and grammars were also introduced for dealing with ambiguities introduced by model imprecision.
The probability $P(CW)$ is computed using Markov models as

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

Semantic Classification trees

(Kuhn and De Mori, 1995)
SEMANTIC GRAMMARS
Interpretation of written text can be seen as a process that uses procedures for translating a sequence of words in natural language into a set of semantic hypotheses (just constituents or structures) described by a semantic language.

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

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

Interesting discussion in (Jackendoff, 1990) Each major syntactic constituent of a sentence maps into a conceptual constituent, but the inverse is not true.
Adding semantic building structures to cfg

Categorial grammars (Lambek, 1958)
Montague Grammars (Montague, 1974)
Augmented Transition Network Grammars (Woods 1970)
Semantic grammars for SLU (Woods, 1976)

Tree Adjoining grammars (TAG) integrate syntax and logic form (LF) semantics. Links can be established between the two representations and operations carried out synchronously (Shabes and Joshi, 1990).
A robust fallback module has been incorporated in successive versions (Delphi Bates et al., 1994).

The system developed at SRI consists of two semantic modules yoked together: a unification-grammar-based module called "Gemini", and the "Template Matcher" which acts as a fallback if Gemini can't produce an acceptable database query (Appelt, 1996).

When a sentence parser fails, constraints on the parser are relaxed to permit the recovery of parsable phrases and clauses (TINA Seneff, 90). Fragments are then fused together.

The linguistic analyzer **TINA**, (MIT, Seneff, 1989), has a grammar written as a set of probabilistic context free rewrite rules with constraints.

The grammar is converted automatically at run-time to a network form in which each node represents a syntactic or semantic category.

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).

**History grammars** (Black et al., 1993)

**Robust partial parser**
Pragmatic grammars

\[
S \rightarrow \text{NP} \text{ VP} \\
\text{NP} \rightarrow \text{PERSON NP} \\
\text{NP} \rightarrow \text{FLIGHT NP} \\
\text{PERSON NP} \rightarrow \text{name} \\
\text{PERSON NP} \rightarrow \text{pronoun} \\
\text{PERSON NP} \rightarrow \text{determiner} \text{ worker} \\
\text{FLIGHT NP} \rightarrow \text{FLIGHT} \\
\text{FLIGHT NP} \rightarrow \text{FLIGHT LOC PHRASE} \\
\text{FLIGHT} \rightarrow \text{flight} \text{ flight num} \\
\text{FLIGHT} \rightarrow \text{determiner flight} \\
\text{LOC PHRASE} \rightarrow \text{SOURCE LOC} \\
\text{LOC PHRASE} \rightarrow \text{DEST LOC} \\
\text{VP} \rightarrow \text{VP1} \\
\text{VP} \rightarrow \text{VP1 TIME PP} \\
\text{VP1} \rightarrow \text{RESERVE VP} \\
\text{VP1} \rightarrow \text{DEPART VP} \\
\text{VP1} \rightarrow \text{ARRIVE VP}
\]
Parsing with ATIS stochastic semantic grammars

```
Please show me the flights to Boston on Monday.
```

Non-terminal nodes:
- show
- flight
- Dest.
- Date

Terminal nodes:
- Show indicator
- Flight Indicator
- Dest. Indicator
- City Name
- Date Indicator
- Day
- Please show me
- the flights
- to
- Boston
- on
- Monday
The Hidden Understanding Model (HUM) system, developed at BBN, is based Hidden Markov Models (Miller et al., 1994).

In the HUM system, after a parse tree is obtained, bigram probabilities of a partial path towards the root, given another partial path are used. Interpretation is guided by a strategy represented by a stochastic decision tree. The semantic language model employs tree structured meaning representations: concepts are represented as nodes in a tree, with sub-concepts represented as child nodes.

\[
Pr(M|W) = Pr(W|M)Pr(M)/Pr(W)
\]

M: meaning
**Hidden vector state 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. (He and Young, 2006)

An example of **state vector** representing a path for a composition to the start symbol S is:

\[
\begin{bmatrix}
\text{CITY} \\
\text{FROM\_LOCATION\_} \\
\text{FLIGHTS} \\
\text{S}
\end{bmatrix}
\]
Microsoft stochastic grammar

Semantic structures are defined by schemata. Each schema is an object (Y.Y. Wang, A. Acero, 2003).

Object structures are defined by an XML schema. Given a semantic schema, a semantic CFG is derived using templates. Details of the schemata are learned automatically.

An entity is the basic component of a schema which defines relations among entities. An entity consists of a head, optional modifiers and optional properties defined recursively so that they finally incorporate a different sequence of schema slots. Each slot is bracketed by an optional pre-amble and post-amble which are originally place holders.
Semantic parsing is discussed in (Tait, 1983).

A semantic first parser is described in (Lytinen, 1992).

A race-based parser is described in (McRoy and Hirst, 1990).

The Delphi system (Bobrow et al., 1990), contains a number of levels, namely, syntactic (using Definite Clause Grammar, DCG), general semantics, domain semantics and action.

Rules transform syntactic into semantic representations

Recent works introduce actions in parsers for generating predicate/argument hypotheses. Strategies for parsing actions are obtained by automatic learning from annotated corpora (FrameNet, VerbNet ....)
Recently, classifiers were proposed for detecting concepts and roles. Such detection process was integrated with a stochastic parser (e.g. Charniak 2001).

A solution using this parser and tree-kernel based classifiers for predicate argument detection in SLU is proposed in (Moschitti et al. ASRU 2007).

Other relevant contributions on stochastic semantic parsing can be found in (Goddeau and Zue. 1992, Goodman. 1996, Chelba and Jelinek, 2000, Roark, 2002, Collins, 2003)

Lattice-based parsers are reviewed in (Hall, 2005)
Semantic building actions in parsing

Use tree kernel methods for learning argument matching (Moschitti, Raymond, Riccardi, ASRU 2007)
Important questions

There is no evidence yet that there is an approach that is superior to all others.

Where are the signs? Are they only words?

Many system architectures are ASR + NLU

How effective is the use of syntactic structures with spoken language and ASR?

How important are inference and composition? Relevant NLU literature exists on these topics.

To what extent can they be used?
SEMANTIC COMPOSITION AND INFERENCE
Semantic composition and dependencies

*a hotel in Toulouse with a swimming pool* *this hotel must be close to the Capitole*

WP2

- a hotel
- in Toulouse
- swimming pool
- this hotel
- close to
- the Capitole

WP3

Semantic composition

ID=1, frame: reservation
frame-elements:

- lodging=hotel
- location=Toulouse
- facility=swimming pool

ID=2, frame: reservation
frame-elements:

- lodging=hotel
- location=close-to-Capitole

Coreference

INF_status="new" related="no"/> INF_status="given" antecedent="ID1" ambiguity="unambiguous" />

Dialog act

da-tag-1="statement"
From constituents to structures

Lattice of interpretations
(to $C_{WP4}$)

Decision Module

Confidence Evaluation

Search for interpretation hypotheses

Lattice of concept hypotheses

*with context information*

(from $C_{WP2}$)

Interpretation strategy

Confidence knowledge

Semantic composition knowledge
Frame representation can be derived from semantic networks and logic. They are computational structures (Kifer et al., JACM, 1995) and also cognitive structuring devices (Fillmore, 1985) in a semantic construction theory.

In (Jackendoff 1990), major conceptual categories also called semantic parts of speech can be elaborated into a function and arguments. Functions can be represented by action frames and arguments by roles.

In KL-ONE (Brachman, 1978) each concept is characterized as a configuration of parts (roles) in specified relationships. Structured taxonomy with inheritance and action parts attached to concept nodes. Constraints on parts are represented by structural descriptions.
A convenient way for asserting properties, and reasoning about semantic knowledge is to represent it as a set of logic formulas. A frame instance (extension) is obtained from predicates that are related and composed into a computational structure. Basic composition units are semantic constituents. They are hypothesized by a sequence labelling process using knowledge acquired by machine learning for which two main approaches have been followed.

Use of k-order generative probabilistic models of paired input sequences and label sequences, for instance hidden Markov models (HMMs) or multilevel Markov models. Generative models are trained to maximize the joint probability of the training data, which is not as closely tied to the accuracy metrics of interest.
Another approach views the sequence labelling problem as a sequence of classification problems, one for each of the labels in the sequence. The classification result at each position may depend on the whole input and on the previous $k$ classifications.

The sequential classification approach can handle many correlated features, as demonstrated in work on maximum-entropy, and a variety of other linear classifiers, including winnow, AdaBoost, and support-vector machines. Furthermore, they are trained to minimize some function related to labeling error, leading to smaller error in practice if enough training data are available.

Conditional random fields (CRFs) bring together the best of generative and classification models. They can accommodate many statistically correlated features of the inputs, and they are trained discriminatively.
Conditional random fields (CRFs) bring together the best of generative and classification models. Like classification models, they can accommodate many statistically correlated features of the inputs, and they are trained discriminatively. But like generative models, they can trade off decisions at different sequence positions to obtain a globally optimal labeling.

If using different models the oracle error rate is reduced, it is worth investigating suitable combinations of methods and models for hypothesizing constituents (a sort of shallow parsing) and for composing them. Different combinations for different composition levels may lead to better results than just using a single approach.

Furthermore, useful confidence indicators can be obtained with multiple views;
Interpretation is problem solving performed by a *composite decision process* which replaces the set of attached procedures.

Problem reduction representation is context-sensitive. Many decompositions are possible involving a variety of methods and KSs, suggesting to consider a *modular approach* to process design.

Possible role instances are hypothesized from constituents and words. Composition is driven by the *support of relations* between supports of constituents (e.g. MEDIA specifiers hypothesized with CRFs)

*Robustness* is obtained by evaluation and possible integration of different KSs and methods used for the same sub-task.
Frame structures and slot chains

Instances of semantic structures are represented by slot chains (Koler, Pfeiffer, 1998)

\[ F_j[r_{jk}(G_x[r_{vk}(v_{xkh})])] \]

\[ \sigma(F_j, v_{xkh}) = \{ (F_j, v_{xkh})/ r_{jk}(F_j, G_x) \land \sigma(G_x, v_{xkh}) \} \]
Composition

\[ \Gamma_j : \text{REQUEST.}\left[\text{agent(speaker), recipient (system), theme (KNOW [theme ITEM [theme (LODGING [])])}\right] \]

\[ G_x : \text{LODGING [ldg_structure (HOTEL[]), ldg_room (ROOM[]), ldg_lux (good)]} \]

Obtained by inference after constituent detection

\[ \text{Speaker(user) } \land \text{ chambre-standing[bon] } \Rightarrow \]

\[ \text{LODGING [ldg_structure (HOTEL[]), ldg_room (ROOM[]), ldg_lux (good)]} \]
Support for Composition

REQUEST.[agent(speaker), recipient (system), theme (KNOW
[theme ITEM [theme (LODGING [ldg_structure (HOTEL[]),
ldg_room (ROOM[]), ldg_lux (good)])])]

Composition is performed if there is a support in the data for their relation

\[ \sup \{ R \left[ \sup \left( \Gamma_j \right), \sup \left( G_x \right) \right] \} \]

Relation support have general word patterns (e.g. specification, inclusion…) which are often independent from the application domain
**Soft constraints**

In (Koller and Pfeffer, 1998) is noticed that one of the limits of the expressive power of frames is the inability to represent and reason about **uncertain and noisy** information.

In **probabilistic frame-based systems**, a frame slot $S$ of a frame $F$ is associated a facet $Q$ with value $Z$: $Q(F,S,V)$. A **probability model** is part of a facet as it represents a **restriction** on the values $V$.

It is possible to have a probability model for a slot value which depends on a slot chain, or, in general, on other values (**Probabilistic version of structural descriptions**)
Frame instance probability

It is shown (Koller, 1998) that it possible to construct a Bayesian network (BN) from a list of dependencies (F1.A <- F2.B) if the resulting dependency graph is acyclic. A Conditional Probability Table (CPT) is associated with each dependency).

The probability of a frame instance can be computed as follows:

\[
P\left\{\Gamma_{i,j}, C_{i,j}, W_{i,j} \mid Y_{i,j}\right\} = P\left[\Gamma_{i,j}\right] \frac{P\left[W_k \mid C_k, R(\gamma_{i,j,k})\right]}{P(W_k)} \frac{\prod_{k=1}^{K} P\left[W_k \mid C_k, R(\gamma_{i,j,k})\right]}{P(W_k)} \frac{P\left[W_{i,j} \mid Y_{i,j}\right]}{P(W_k)}
\]

Frame_instance, concepts_for_slots supports relation_to_slot
Dependency graph with cycles

If the dependence graph has cycles, then possible worlds can be considered. A general method for computing probabilities of possible worlds based on Markov logic networks (MLN) is proposed in (Richardson, 2006).
Probabilistic models of relational data

Probability of relational data can be estimated in various ways, depending on the data available and on the complexity of the domain.

For simple domains it is possible to use a naïve Bayes approach. Otherwise, it is possible to use the disjunctive interaction model (Pearl, 1988), or relational Markov networks (RMN) (Taskar, 2002).

Methods for probabilistic logic learning are reviewed in (De Raedt, 2003).
Frame-based resources


Verb lexicon **PropBank** (Palmer, 2003).

**VerbNet** (Kipper et al., 2000) is a manually developed hierarchical verb lexicon based on the verb classification of Levin (1993). For each of 191 verb classes, including around 3000 verbs in total, VerbNet specifies the syntactic frames along with the semantic role assigned to each slot of a frame.
Semantic knowledge representation

Semantic descriptions may have connectives, co referential (descriptions attached to a slot are attached to another and vice-versa), declarative conditions. Attached procedures may perform different types of actions.

Verbs are fundamental components of natural language sentences. Roles can cases. Roles can also be properties of structured entities or arguments for functions. Descriptions based on predicate/argument structures can be derived.

Temporal representations can be made in higher order logic with lambda abstraction (Crouch and Pulman, 1993). With procedural attachment, complex knowledge can be represented as in schemata (S. Narayanan, 1999)
Partial parsing, also called \textit{chunking}, is proposed for mapping the verb arguments onto subcategorization frames that can be extracted automatically, for example, from \textit{WordNet} (Miller, 1995).

\textit{MindNet} (Richardson et al., 1998) produces a hierarchical structure of semantic relations (\textit{semrels}) from a sentence using a words in a machine readable dictionary.
DIALOGUE ACTS AND TASK REPRESENTATION
Short-term memory structures

- Dialogue state representation
  - Dialogue turn representation
  - Understanding actions
  - Negotiation model
    - Understanding actions
      - Task representation
        - Task model
          - Constituents
Speech acts

Negotiation dialogues are characterized by a hierarchy of illocutory (speech) acts (Chang, 2004).

They are discourse actions identified by verbs, other lexical units or implied by other concepts expressed in a sentence.

These speech acts (SA) determine the sentence type. Various attempts have been made to identify SAs which are domain independent.

A possible taxonomy of them is formulated in the Dialogue Act Markup in Several Layers (DAMSL).
Speech acts

In (Cohen and Perrault, 1979), a notation of formulating dialogue acts as plan operators is proposed.

A negotiation dialogue follows a partially ordered plan represented by a Hierarchy of Tasks (HT) (Sacerdoti, ijcai75).

Each task is characterized by a SA whose effect is the instantiation, modification or finalization of conceptual structures required for performing transactions.

HT is a generative structure of possible sequences of SAs characterizing the sentences of a dialogue with which a system and a user negotiate for defining a possible transaction.
Speech acts

The main purpose of a service is to satisfy a user goal.

If a service can satisfy many goals, it has to hypothesize/identify actual user goals and, for each goal consider a mean to achieve it.

Such a mean can be a plan whose actions are executed following a policy and have the objective of gathering all the necessary details for specifying an instance of a goal which corresponds to a user intention

In the considered applications the goals are performing transactions and the dialogue involves negotiations represented by non-linear, partially ordered hierarchies of tasks whose possible sequences can be generated by rules
Negotiation dialogues

N_Dialogue := Open - Negotiation - Commit - Close

Negociation := Formulation (Formulation | Repair)*

Formulation := (Assert | Request | Propose | Maybe)
(Assert | Request | Propose | Maybe)*

Request := (Know | Reserve | Confirm) (Know | Reserve | Confirm)*

Repair := (Repeat + Hold + Correct)* (Repeat + Hold Correct + Reject + PartialReject)

Commit := Accept
Dialogue turn representation

words c'est bien ça

constituent command-dial[confirmation-demande]:

Freme instance CONFIRM. [theme (ITEM)])
Task representation

SESSION [ theme (TRANSACTION [], INFORMATION[])]

TRANSACTION [theme (RESERVATION[]), status (completeincomplete), proposed (Y,N) ]

INFORMATION [theme (enum(LODGING []), enum (RESTAURANT []))]

RESERVATION [ customer (PERSON []), theme (LODGING [], RESTAURANT []), time (PERIOD []) ]

LODGING [ loc [LOCATION], type [HOTEL], element [enum(ROOM)], facilities° [enum (FACILITY)], luxury° (value)]
REQUEST [agent (user), theme (ITEM[])] \land

\neg \ EXIST_INSTANCE_OF (SESSION) ->

instantiate SESSION [theme INFORMATION[theme (ITEM.theme)]]

where ITEM.theme is the value of the theme of ITEM
MODULAR SYSTEMS
**Combinations of approaches NLU**

Rule-based approaches to interpretation suffer from their brittleness and the significant cost of authoring and maintaining complex rule sets.

Data-driven approaches are robust. However, the reliance on domain-specific data is also one of the significant bottlenecks of data-driven approaches.

Combining different approaches makes it possible to get the best out of them. Simple grammars are used for detecting possible clauses, then **classification-based parsing** completes the analysis with inference (Kasper and Hovy, 1990).

**Shallow semantic parsing** was proposed by (Gildea and Jurafsky, 2002, Hacioglu and Ward, 2003, Pradhan et al. 2004)
In (Wang et al., 2002), stochastic semantic grammars are combined with classifiers for recognizing concepts. Their combination with ROVER (the hypothesis which gets the majority of votes wins). SVM alone resulted to be the best even if ROVER is applied. Important improvement was found by replacing certain words with their semantic categories found by the parser. Concepts detected in this way are used to filter the rules of the semantic grammar applied to find slot fillers.
A parser based on tagging actions producing non-overlapping shallow tree structures is proposed in (Hacioglu, K. (2004), at lexical, syntactic and semantic levels to represent the language.

The goal is to improve the portability of semantic processing to other applications, domains and languages.

The new structure is complex enough to capture crucial (non-exclusive) semantic knowledge for intended applications and simple enough to allow flat, easier and fast annotation.
The use of just a grammar is not sufficient, (Bangalore et al.,) because recognition needs to be more robust to extragrammaticality and language variation in user’s utterances and the interpretation needs to be more robust to speech recognition errors. For this reason, a class-based trigram LM is built with in-domain data.

In order to improve recognition rates, sentences are generated with the grammar to provide data for training the classifiers.

In (Shapire et al. 2005), authors explore the use of human-crafted knowledge to compensate for the lack of data in building robust classifiers.
In (Sarikaya et al, 2004), a system is proposed which generates an N-best (N=34) list of word hypotheses with a dialogue state dependent trigram LM and rescores them with two semantic models.

1 An Embedded context-free semantic Grammar (EG) is defined for each of 17 concepts and performs concept spotting by searching for phrase patterns corresponding to concepts.

2 A second LM, called Maximum Entropy (ME) LM (MELM), computes probabilities of a word, given the history, using a ME model.
SPEECH ACTS
Sentence boundary detection

Using prosody (Shriberg et al., 2000)

Approaches to boundary detection have used finite-state sequence modeling approaches, including Hidden Markov Models (HMM) and Conditional Random Fields (CRF) (Roark et al. 2006)

Sentences are often short, providing relatively impoverished state sequence information.

A Maximum Entropy (MaxEnt) model that did not use state sequence information, was able to outperform an HMM by including additional rich information.

Features from (Charniak, 2000) parser were used.
Sentence classification

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.

• Connectionist models (Gorin et al. 1995)
• SVD (Chu-Carroll and Carpenter, 1999)
• Latent Semantic Analysis (LSA) (Bellegarda 2002)
• SVM, cosine similarity metric (used in IR) and Beta-classifier (IBM, 2005, 2006)
• Cluster of sentences is proposed in (He and Young, 2006)
Béchet et al. ICASSP 2007

\[ \Gamma_k \text{ is a composition} \]

\[ P(S_k|\Gamma_k S_{k-1}) \]

\[
P(S|Y) = \sum_{\Gamma} P(S\Gamma|Y) = \sum_{\Gamma} P(S_k\Gamma_k|H_kY)P(H_k|Y)
\]

\[ P(S_k\Gamma_k|H_kY) \approx \max_{W_k, C_k} P(\Gamma_k|C_k)P(C_k|W_k)P(W_k|Y_k) \times \]

\[ P(S_k|\Gamma_k S_{k-1}) \]

FT/LIA System 3000

IEEE ASRU

Kyoto Dec 11th 2007
CONFIDENCE AND LEARNING
unsupervised semantic role labelling

Interpretation modules have parameters estimated by automatic learning (Chronus, Chanel, HUM and successor systems).

Semantic annotation is time consuming. The process should be semi-automatic starting with bootstrapping (e.g., Hindle and Rooth, 1993; Yarowsky, 1995; Jones et al., 1999).

Initially make only the role assignments that are unambiguous according to a verb lexicon ((Kate and Mooney, 2007).

A probability model is created based on the currently annotated semantic roles.

When unlabeled test examples are also available during training, a transductive framework for learning can further improve the performance on the test examples.
Active Learning

Hakkani-Tür, Riccardi Gorin, 2002)
Certainty-Based Active Learning for SLU
Confidence

Evaluate confidence of components and compositions

\[ P(\Gamma|\Phi_{\text{conf}}) \]

\( \Phi_{\text{conf}} \) represents the confidence indicators or a function of them.

Notice that it is difficult to compare competing interpretation hypotheses based on the probability \( P(\Gamma|Y) \) where \( Y \) is a time sequence of acoustic features, because different semantic constituents may have been hypothesized on different time segments of stream \( Y \).
Confidence measures

Two basic steps:

1) generate as many features as possible based on the speech recognition and/or natural language understanding process and

2) Estimate correctness probabilities with these features, using a combination model.
Define confidence-related situations

Consensus among classifiers and SFST is used to produce confidence indicators in a sequential interpretation strategy (Raymond et al. 2005, 2007). Classifiers used are SCT, SVM, adaboost. Committee-Based Active Learning uses multiple classifiers to select samples (Seung et al. 1992)
Committee-Based Active Learning

Call classification (Tur, Schapire, and Hakkani-Tür, 2003)
Unsupervised Learning

(Tur and Hakkani-Tür, Riccardi and Hakkani-Tür, 2003)
Co-Training

Assume there are multiple views for classification

1. Train multiple models using each view

2. Classify unlabeled data

3. Enlarge training set of the other using each classifier’s predictions

4. Goto Step 1
Combining Active and Unsupervised Learning

Train a classifier using initial training data

While (labelers/data available) do

Select $k$ samples for labeling using active learning

Label and add these selected ones to the training data and retrain

Exploit the unselected data using unsupervised learning

Update the pool.
Adaptive Learning in Practice

(Riccardi et al, 2005)
Solutions for applications

The simple use of semantic constituents is sufficient for applications such as call routing, utterance classification with a mapping to disjoint categories and perhaps to speech-to-speech translation and speech information retrieval.

Semantic composition is useful for applications like spoken opinion analysis, call routing with utterance characterization (finer-grain comprehension), question/answering, inquiry qualification.

A broad context is taken into account for context-sensitive validation in complex spoken dialog applications and inquiry qualification considering an utterance as a set of sub-utterances and the interpretation of one sub-utterance being context-sensitive to the others.
Conclusions

A modular SLU architecture can exploit the benefits of combined use of CRFs, classifiers and stochastic FSMs, which are approximations of more complex grammars.

Grammars should perhaps be used in conjunction with processes having inference capabilities.

Recent results and applications of probabilistic logic appear interesting, but its effective use for SLU still has to be demonstrated.

Annotating corpora for these tasks is time consuming suggesting that it is suitable to use a combination of knowledge acquired by a machine learning procedures and human knowledge.
Conclusions

Robustness, incremental learning, portability are important and open issues.

SLU is not only used in human-machine dialogs. Other applications are for opinion analysis, indexing, summarization, retrieval.

When SLU is used in dialog, interpretation strategies should provide hypotheses with confidence indicators, taking into account dialogue context, communication principles, types of actions and goals, types of sources.
THANK YOU