Spoken Language Understanding; a survey

Renato De Mori

Laboratoire d'informatique
Université d'Avignon

LUNA IST contract no 33549

Information Society Technologies

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

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* }

Frame instances (extension)

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

$$(\exists x)\left\{\begin{array}{l}\text{instance\_of}(x,\text{address}) \wedge \text{loc}(x,\text{Avignon}) \wedge \text{area}(x,\text{Vaucluse}) \wedge \\ \wedge\ \text{country}(x,\text{France}) \wedge \text{street}(x,1\ \text{avenue Pascal}) \wedge \text{zip}(x,84000)\end{array}\right\}$$

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)

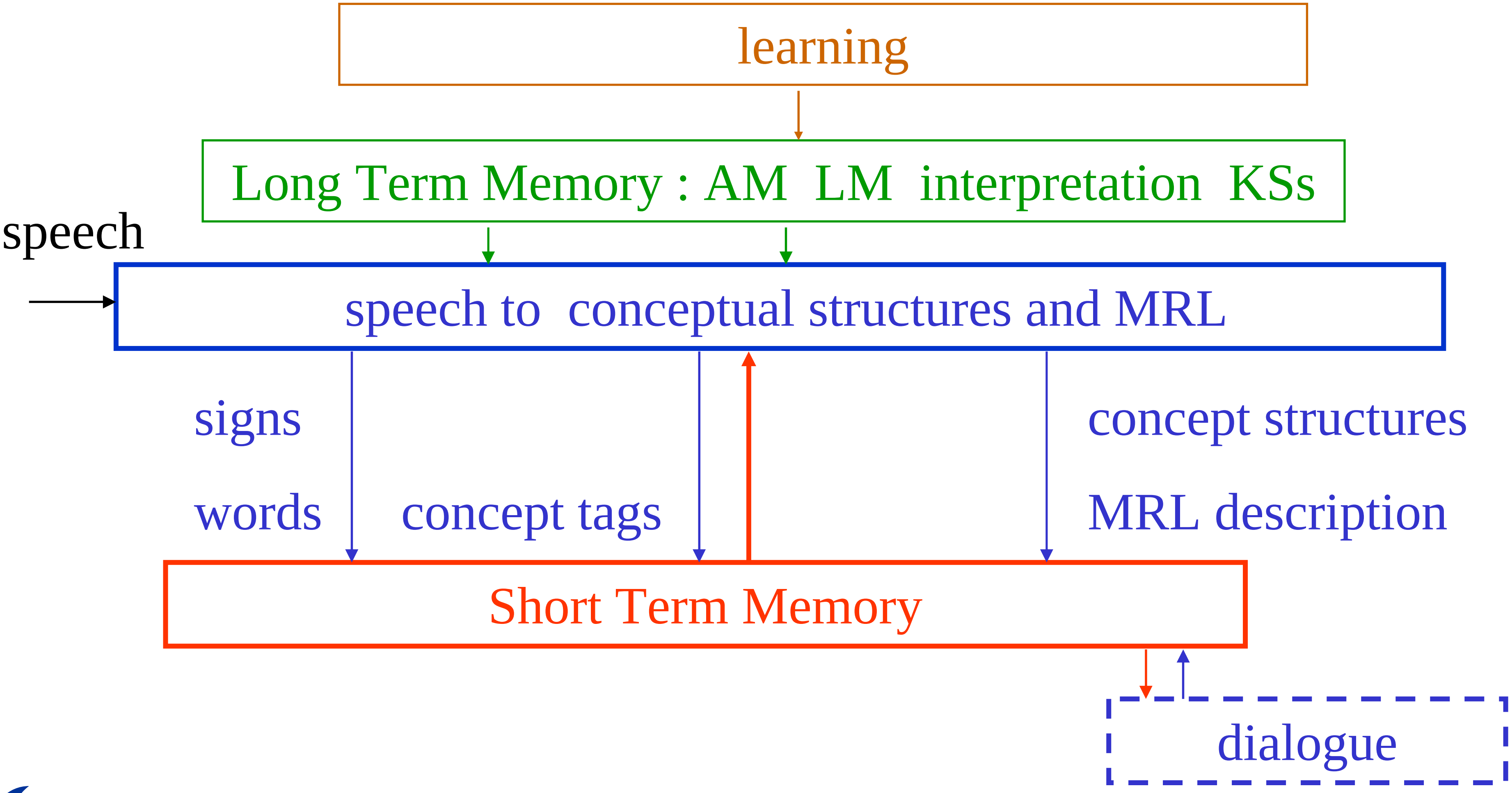
Frame instance

Schemata contain collections of properties and values expressing relations. A property or a role are represented by a slot filled by a value

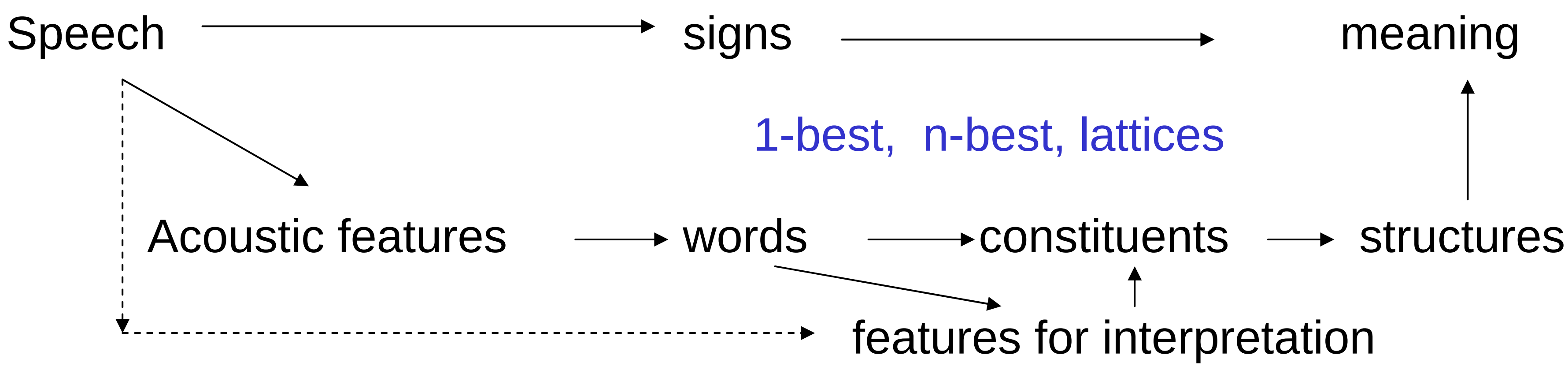
{a0001	
instance_of	address
loc	Avignon
area	Vaucluse
country	France
street	1, avenue Pascal
zip	84000
}	

Process overview

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



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

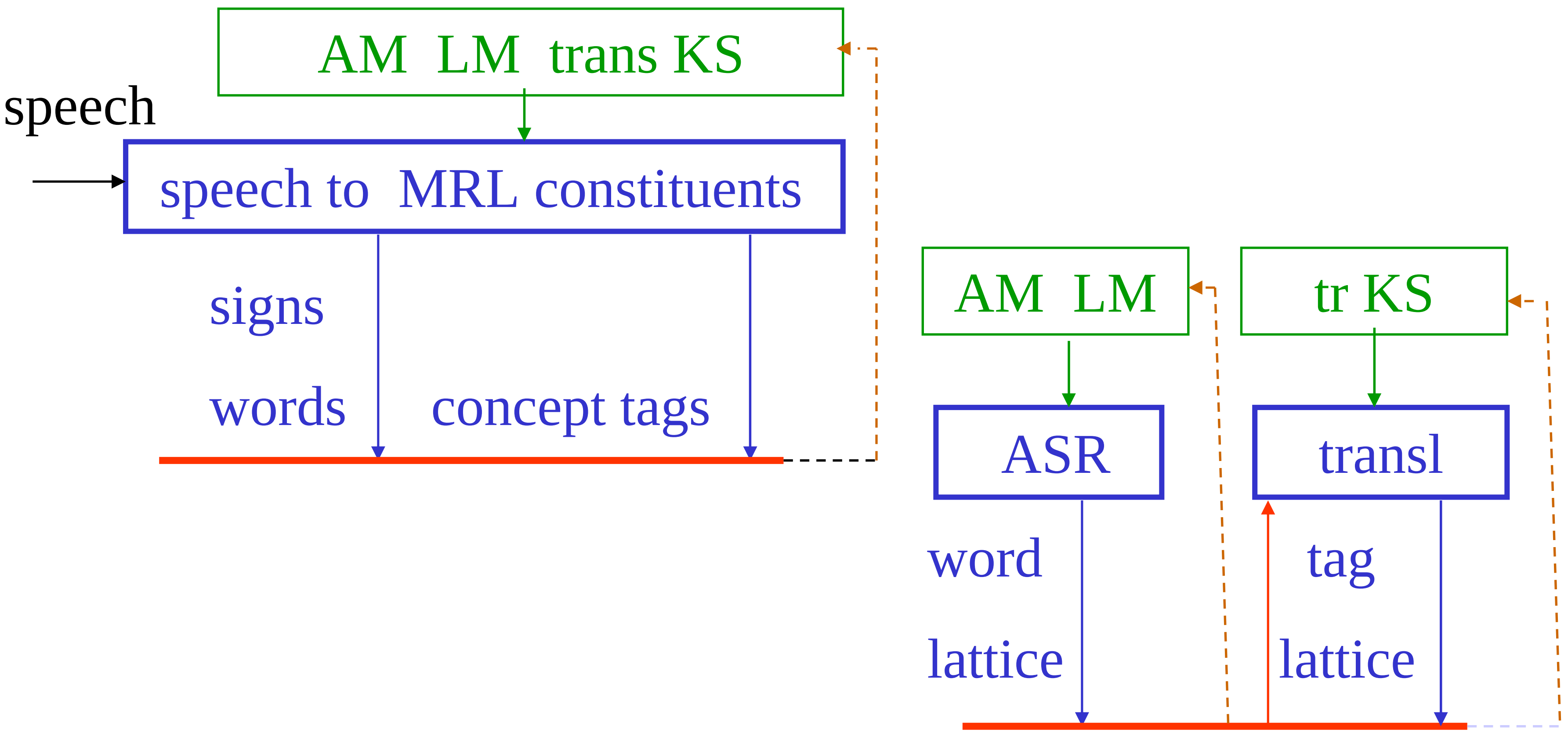
Semantic composition on basic constituents

Context-sensitive validation

Combination of level processes may depend on the application

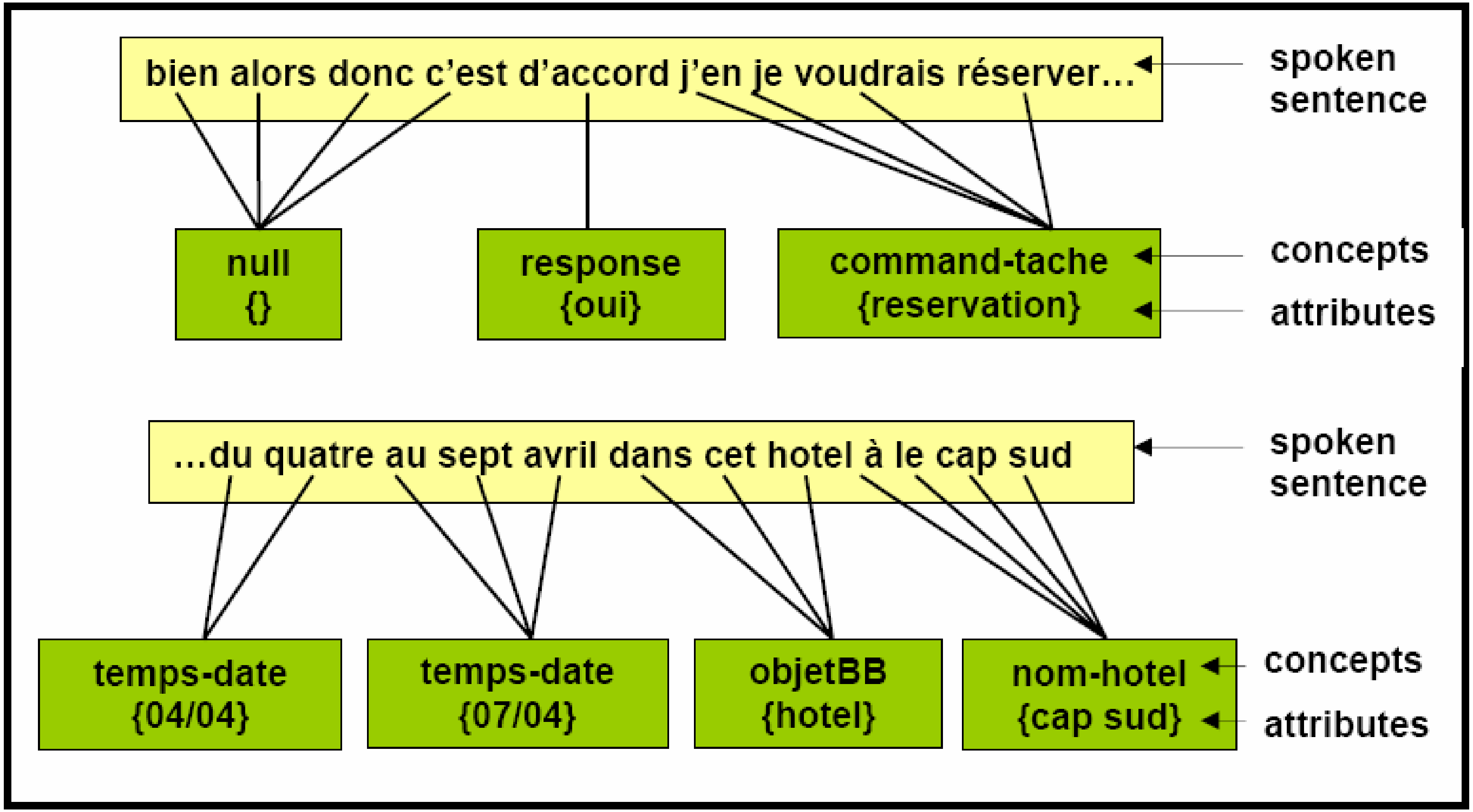
From signs to constituents

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



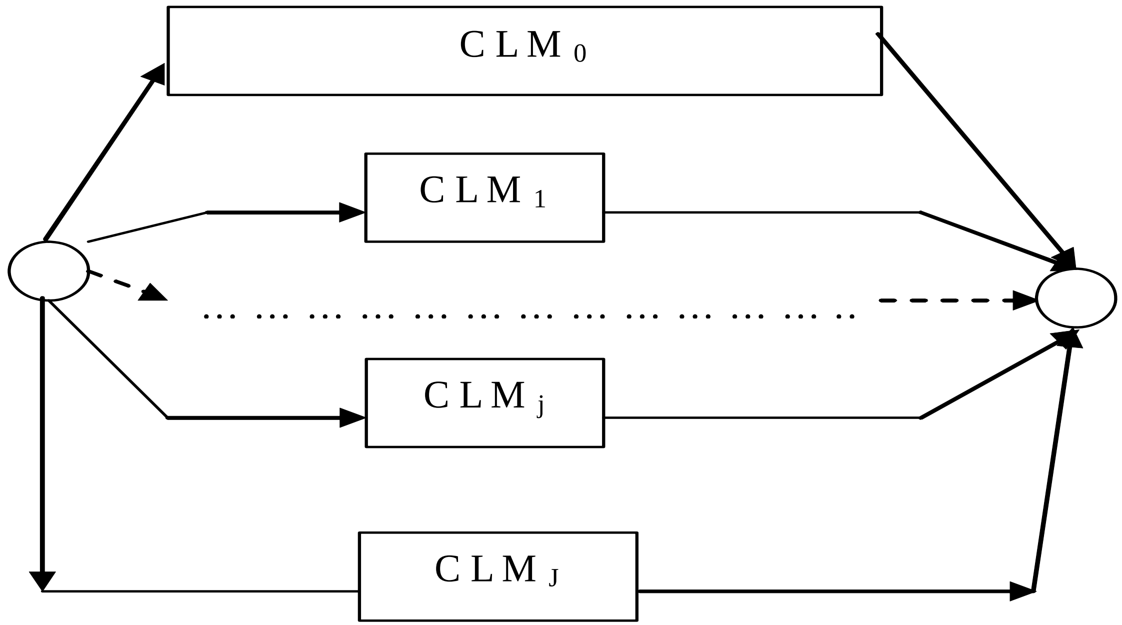
Finite-state conceptual language models

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).

Conceptual Language Models



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.

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

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

SWG=OUTPROJ(SEMG)

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

NL - MRL translation

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 group 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).

Intresting solutions also in (Macherey et al., 2001, Sudoh and Tsukada, 2005 for attribute/value pairs, LUNA)

CRF

Possibility of having features from long-term dependences

Results for LUNA from Riccardi, Raymond, Ney, Hann

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

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

$$f_k(y_{i-1}, y_i, x, i) = \begin{cases} 1 & \text{if } y_i = ARRIVE.CITY \\ & \text{and } x_i \ldots x_{i-1} \text{ contains } \{\text{arrive} \mid \text{to}\} \\ 0 & \text{otherwise} \end{cases}$$

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

Raymond C., Riccardi G. "Generative and Discriminative Algorithms for Spoken Language Understanding", Proc. INTERSPEECH, Antwerp, 2007.

Moschitti A., Riccardi G., Raymond C. "Spoken language understanding with kernels for syntactic/ semantic structures", Proc. IEEE ASRU, Kyoto, 2007.

Comparison

model	attribute CER [%]	attribute SER [%]	attribute/value CER [%]	attribute/value SER [%]
CRF	11.8	17.7	16.2	23.0
log-linear	14.9	22.2	19.3	26.4
FST	17.9	24.7	21.9	28.1
SVM	18.5	24.5	22.2	28.5
MT	19.2	24.6	23.3	27.6

Incremental oracle performance

model	attribute CER [%]	attribute SER [%]
CRF	11.8	17.7
+log-linear	9.8	15.6
+FST	8.3	13.8
+SVM	7.6	12.9
+MT	7.0	12.3

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. Possiiblity of further imprevement by combination with CRFs and using dialog constraints

Demo

LUNAVIZ

History

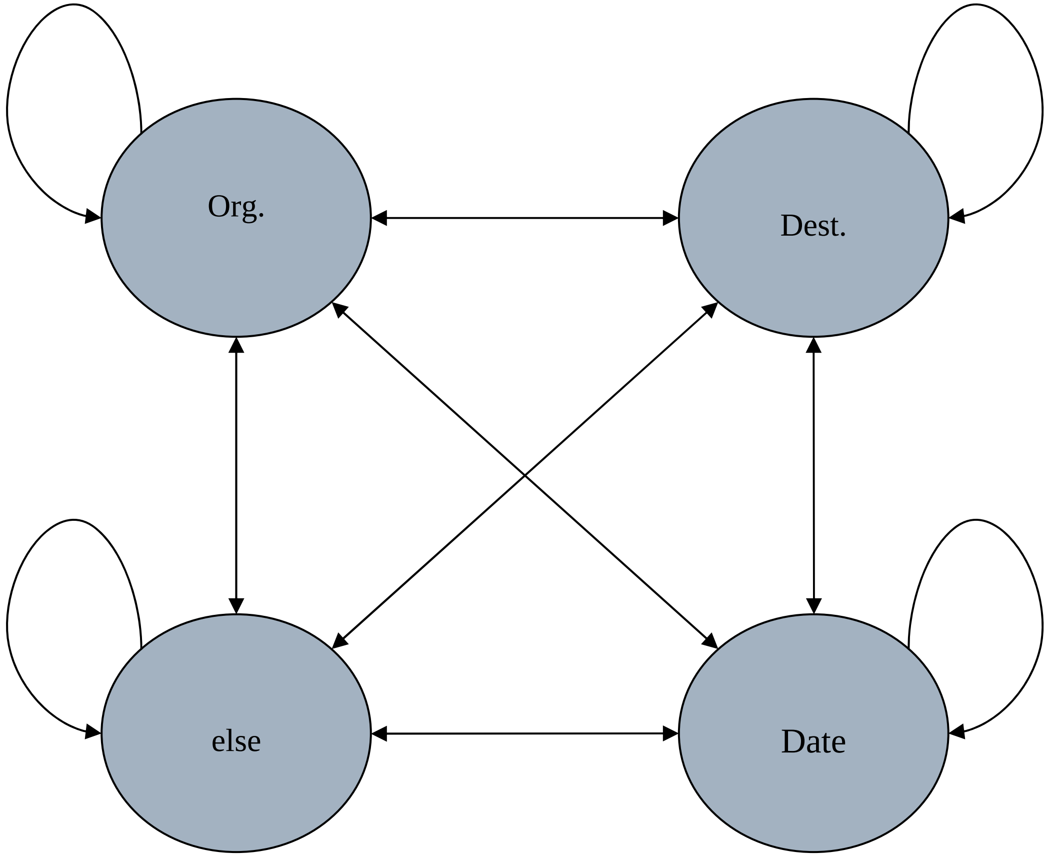
Systems developed in the seventies reviewed in (Klatt, 1977) and the eighties, early nineties (EVAR, SUNDIAL) mostly performed syntactic analysis on the best sequence of words hypothesized by an ASR system and used non probabilistic rules, semantic networks, pragmatic and semantic grammars for mapping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

Probabilistic interpretation in the Chronous system

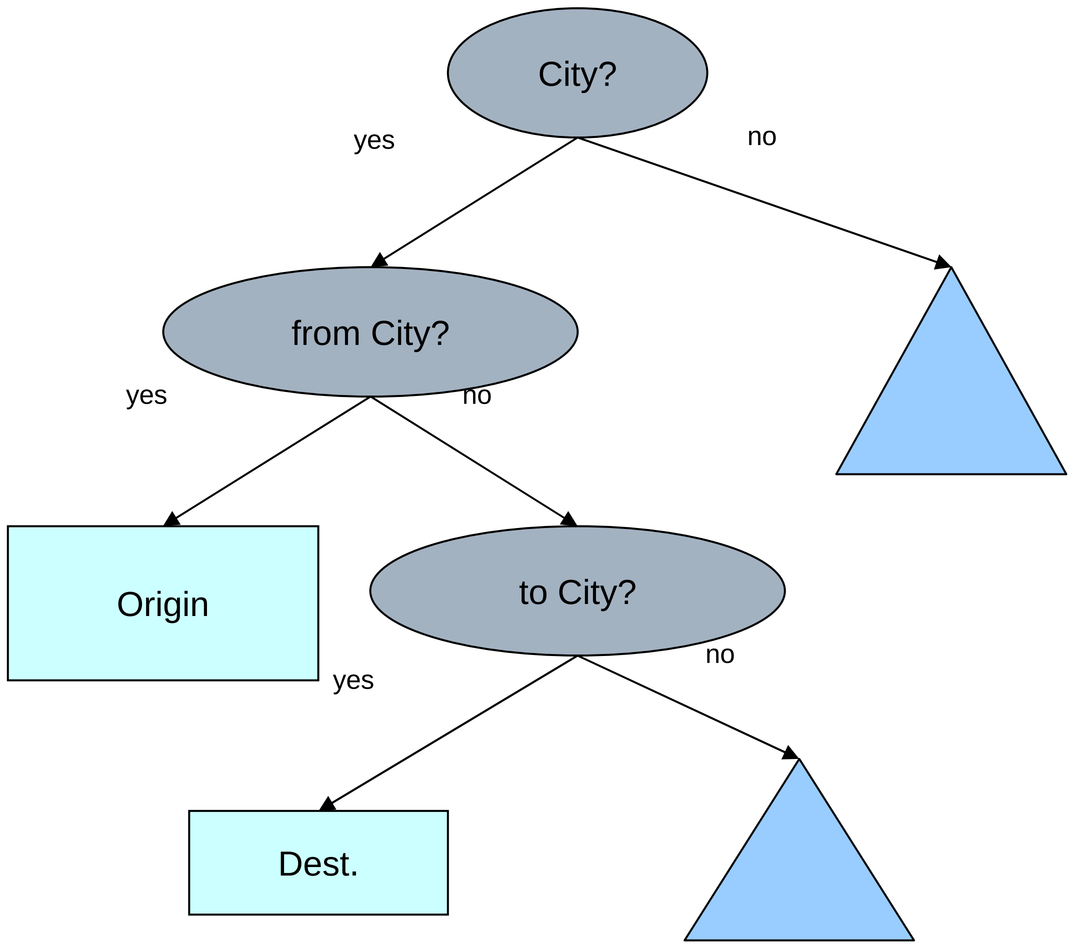
The probability P(CW) is computed using *Markov models* as

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



(Pieraccini et al., 1991, Pieraccini, E. Levin, E. Vidal, 1993).

Semantic Classification trees



(Kuhn and De Mori, 1995)

SEMANTIC GRAMMARS

Interpretation as a translation process

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.

Using grammars for NLU

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).

Robust parsing (early ATIS)

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.

Local parsing (Abney, 1991).

Stochastic semantic context-free grammars

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

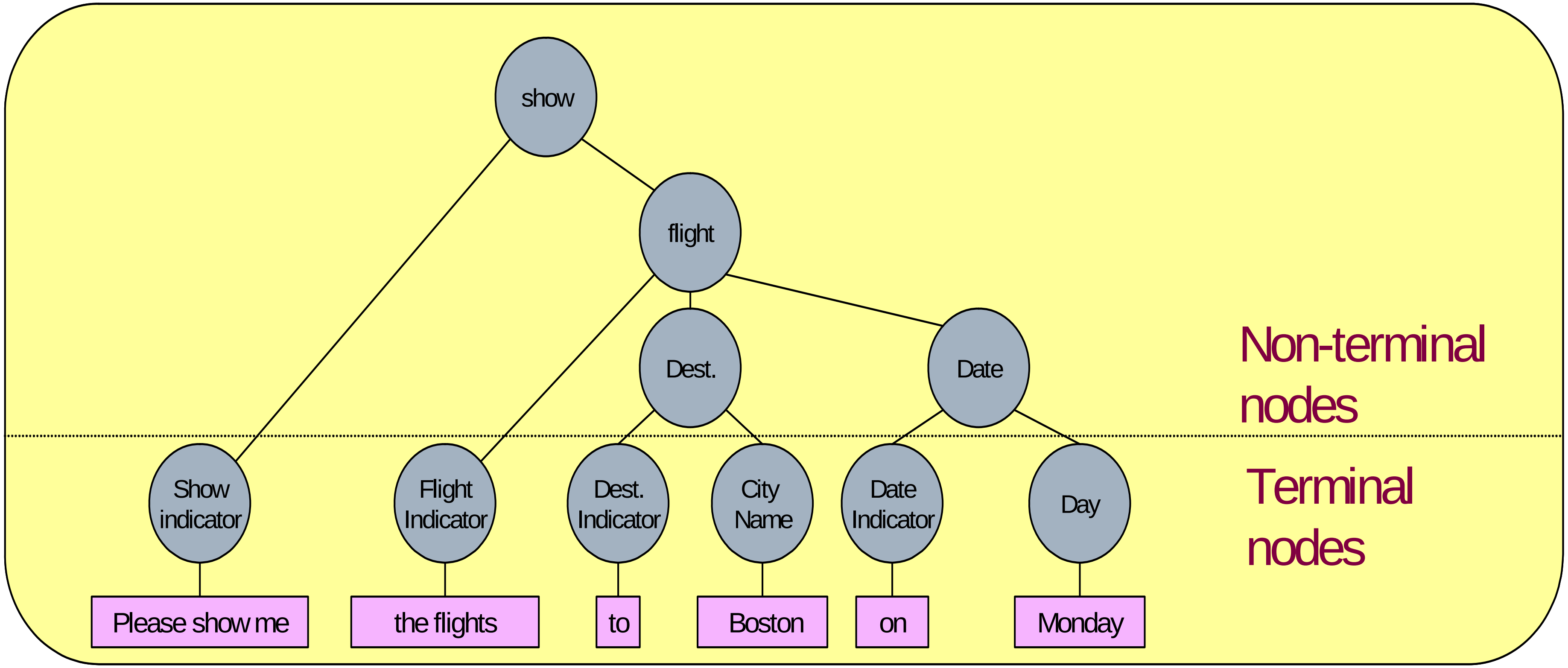
The grammar is converted automatically at run-time to a network form in which each node represents a syntactic or semantic category.

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

Parsing with ATIS stochastic semantic gramamrs



Stochastic semantic context-free grammars

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.

Concurrent or sequential use of syntax and semantic knowledge

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 ….)

Predicate/argument structures and parsers

Recently, classifiers were proposed for detecting concepts and roles.

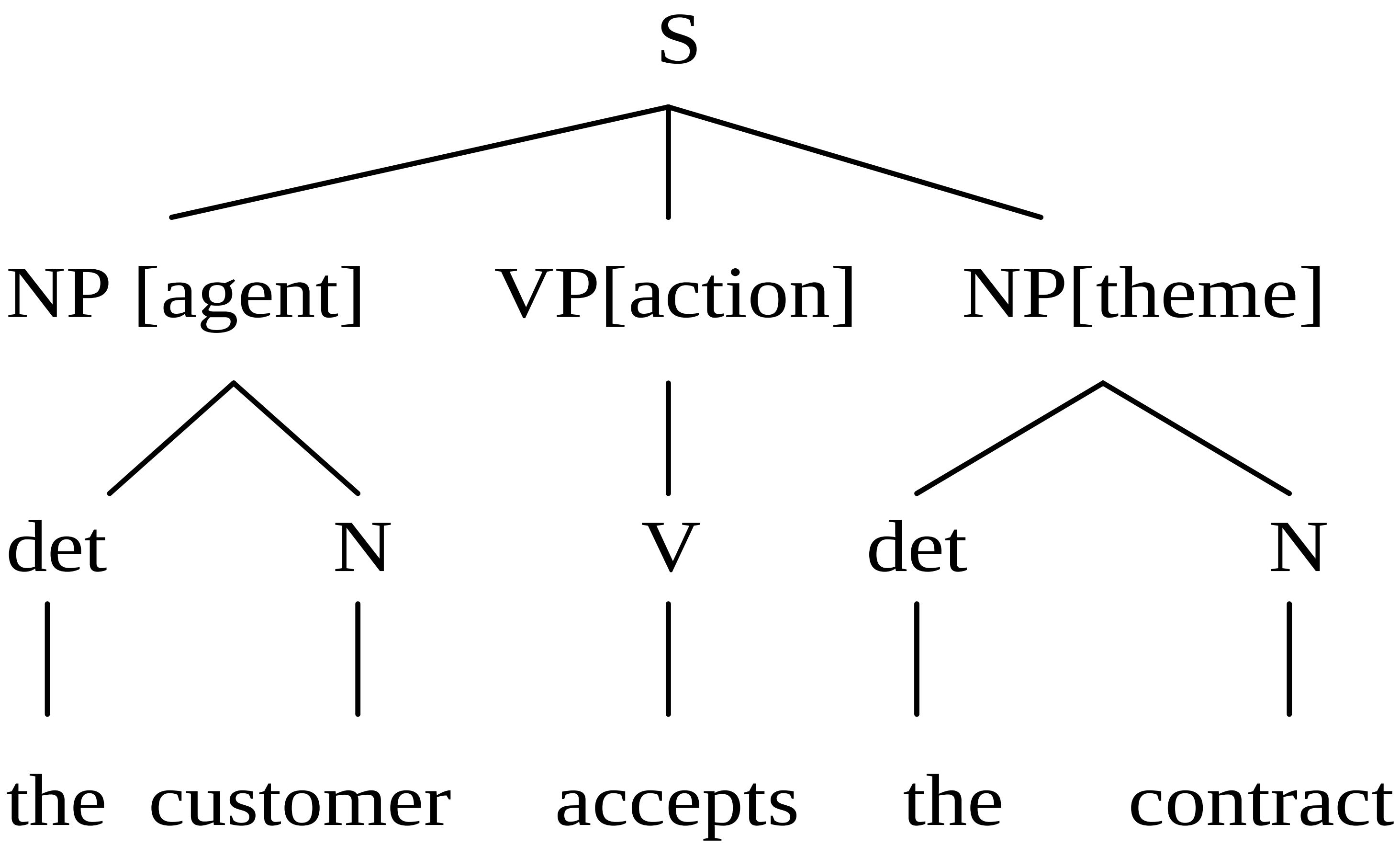
Such detection process was integrated with a stochastic parser (e.g. Charniak 2001).

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

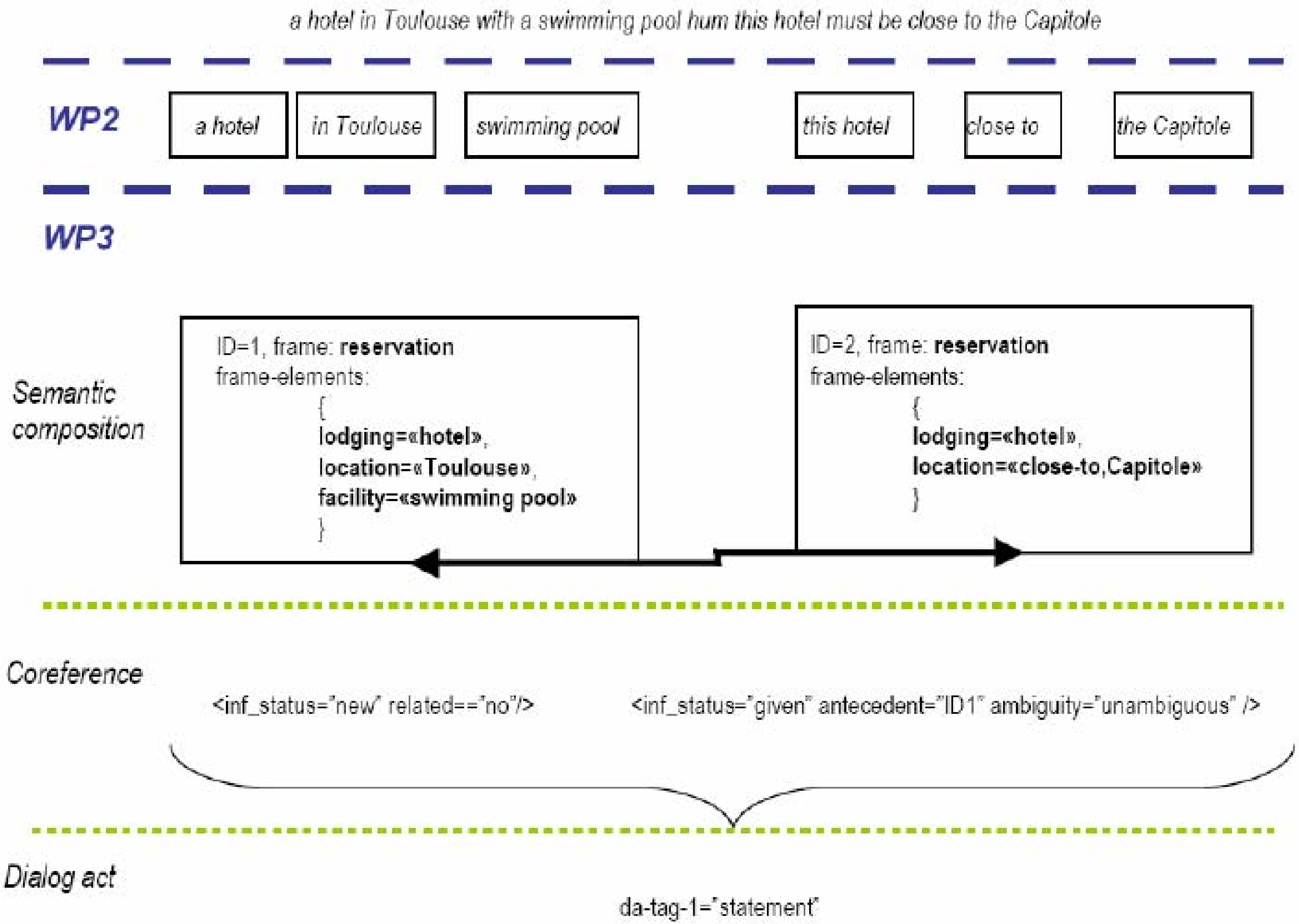
How effective is the use of syntactic structures with spoken language and ASR?

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



Composition features

If composition is performed when semantic constituents have been hypothesized, then it is important to identify words and features that support the fact that a constituent hypothesis is the slot-filler of a frame instance.

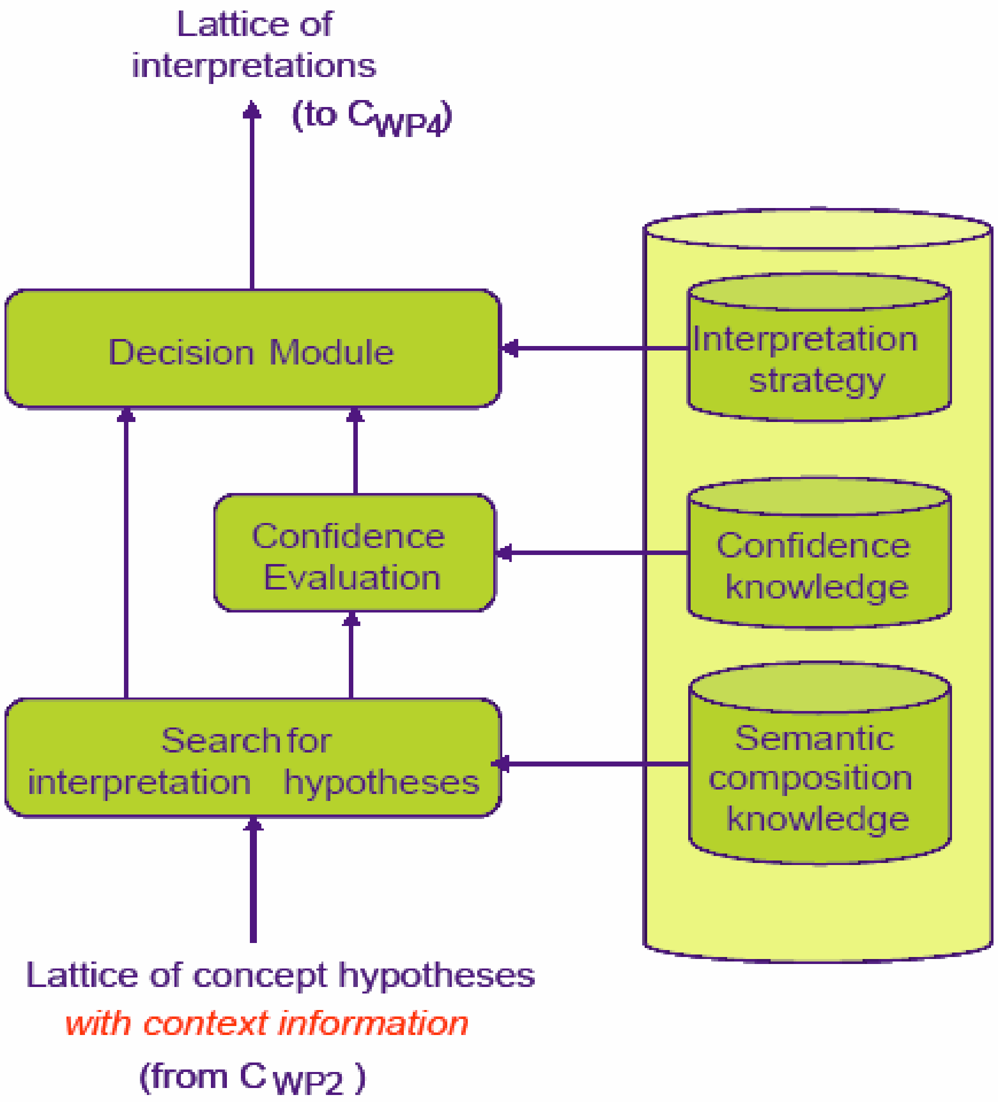
$$W_k \rightarrow R(C_k, \gamma_{i,j,k})$$

Automatic annotation of the MEDIA corpus has been performed using models trained after boostrapping. Annotations were validated using relational LMs

Demo

FRIZ

From constituents to structures



Simple frame probabilistic model

In (Thompson et al., 2003) it is suggested that a frame F is instantiated by a predictor word S and roles R are related to phrases C.

S $\longrightarrow$	F $\longrightarrow$	[R1 ……. Rn]	*roles*
predictor	*frame*	[C1 ……. Cn]	*chunks*

Probability model with Markov assumption

$$P(C, R, F, S) = P(S)P(F|S)P(R|FS)P(C|RFS)$$

$$P(R|FS) \approx \prod P(R_i|R_{i-1}F)$$

$$P(C|RFS) \approx P(R|FS) \approx \prod P(C|R) = \prod P(C_i|R_i)$$

Logic based approaches to interpretation

Logic based approaches to NLU were proposed for representing semantic knowledge and performing inference on it.

In (Norvig, 1987) inferences are considered for asserting implicit meaning of a sentence or implicit connections between sentences.

In (Palmer, 1983), it is suggested to detect relationships between semantic roles by inference.

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. Probability distributions were introduced in slot facets to represent constraints on possible role values. An algorithm was proposed for obtaining a Bayesian Network (BN) from a list of dependences between frame slots.

Probabilistic frame based systems

In probabilistic frame-based systems, (Koller 1998) a frame slot S of a frame F is associated a facet Q with value Z: Q(F,S,Y).

A probability model is part of a facet as it represents a restriction on the values Y.

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.

$Y_k \rightarrow W_k \rightarrow$

Dependency graph with cycles

$Y_k \rightarrow W_k \rightarrow C_k \rightarrow \gamma_{i,j,k}$

$Y_\ell \rightarrow W_\ell \rightarrow C_\ell \rightarrow \gamma_{i,j,\ell}$

$Y_m \rightarrow W_m \rightarrow C_m \rightarrow \gamma_{i,j,m}$

Acoustic_evidence support concept filled-slot

If the dependence graph has cycles, then possible worlds can be considered. The computation of probabilities of possible worlds is discussed in (Nilsson, 1986). 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).

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)

Microsoft SLU

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

Colorado

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.

ATT

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

Goal frames

Predicate/argument sets contribute to form a frame when the resulting structure has a specific meaning.

For some applications, the only useful composition is a frame representing dialog act whose components are semantic constituents.

{ TEST (CONNECTS (SUBJ AC) (PATH (ORIG TORONTO) (DEST DALLAS)))}

Application goals can be represented by frames which constrain the aggregation of predicate/argument pairs to specify system actions.

Dialog act detection can be performed when constituents have been hypothesized

Speech acts

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.

- Semantic classification trees [Mast *et. al* '96], (Wiebe at al., 1997)
- Decision trees [Stolcke *et. al* '98, Ang *et. al* '05],
- HMMs [Stolcke et. al '98],
- Classification trees (Tanigaki and Sagisaka, 1999),
- Neural networks [Stolcke *et. al* '98, Wang *et. al* '99]
- Fuzzy fragment-class Markov models [Wu *et. al* '02]
- Maximum entropy models [Stolcke *et. al* '98, Ang *et. al* '05]
- Bayesian belief networks (Bilmes et al., 2005),
- Bayesian belief model (BBM) (Li and Chou, 2002)

Dialog event tagging

In (Zimmermann et al., 2005) prosodic features (pause durations) are used in addition to word dependent events.

A Hidden-Event Language Model (HELM) is used in a process of simultaneous segmentation and classification.

After each word, the HE-LM predicts either a non-boundary event or the boundary event corresponding to any of the DA types under consideration

Mapping words into actions (Potamianos et al., 1999, Meng et al., 1999).

Latent Semantic Analysis is proposed in (Bellegarda, 2002, Zhang and Rudnicky, 2002)

Sentence boundary detection

Using prosody (Shriberg rt 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)

Use of dialog constraints for composition

User goals can be represented by frames. A plan for achieving each goal can be represented by a sequence of states. If different goals are hypothesized in a dialog control agenda (e.g. Rudnicky, 2001), then the set of the corresponding plans are represented by a finite state machine.
Different states can be reached with different probabilities.
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 phase, a word-to-constituent transducer translates a word lattice into a constituent lattice. In the second phase, a set of precondition-action rules, encoded as a transducer, transforms concept hypotheses into state transitions.

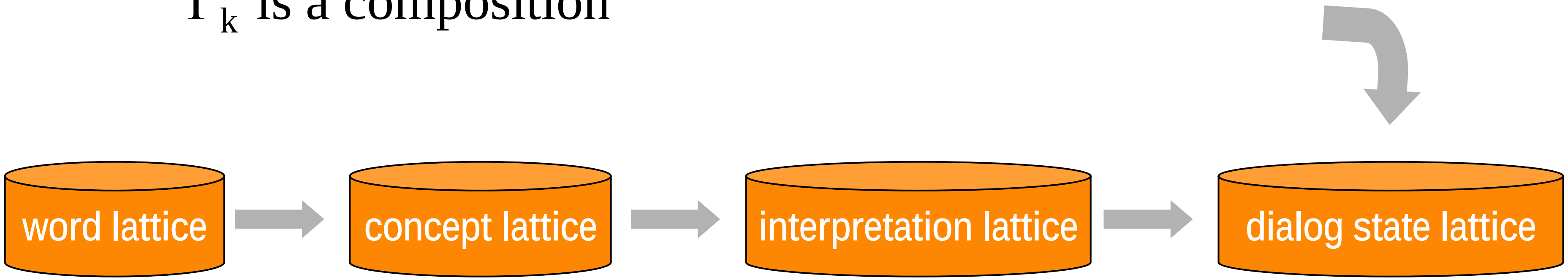
A lattice of words is thus translated into a set of states

FT/LIA System 3000

Béchet et al. ICASSP 2007

Γ_k is a composition

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



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

$$P(S_k\Gamma_k|H_kY) \approx P(S_k|\Gamma_k S_{k-1}) \times \max_{W_k, C_k} P(\Gamma_k|C_k)P(C_k|W_k)P(W_k|Y_k)$$

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)

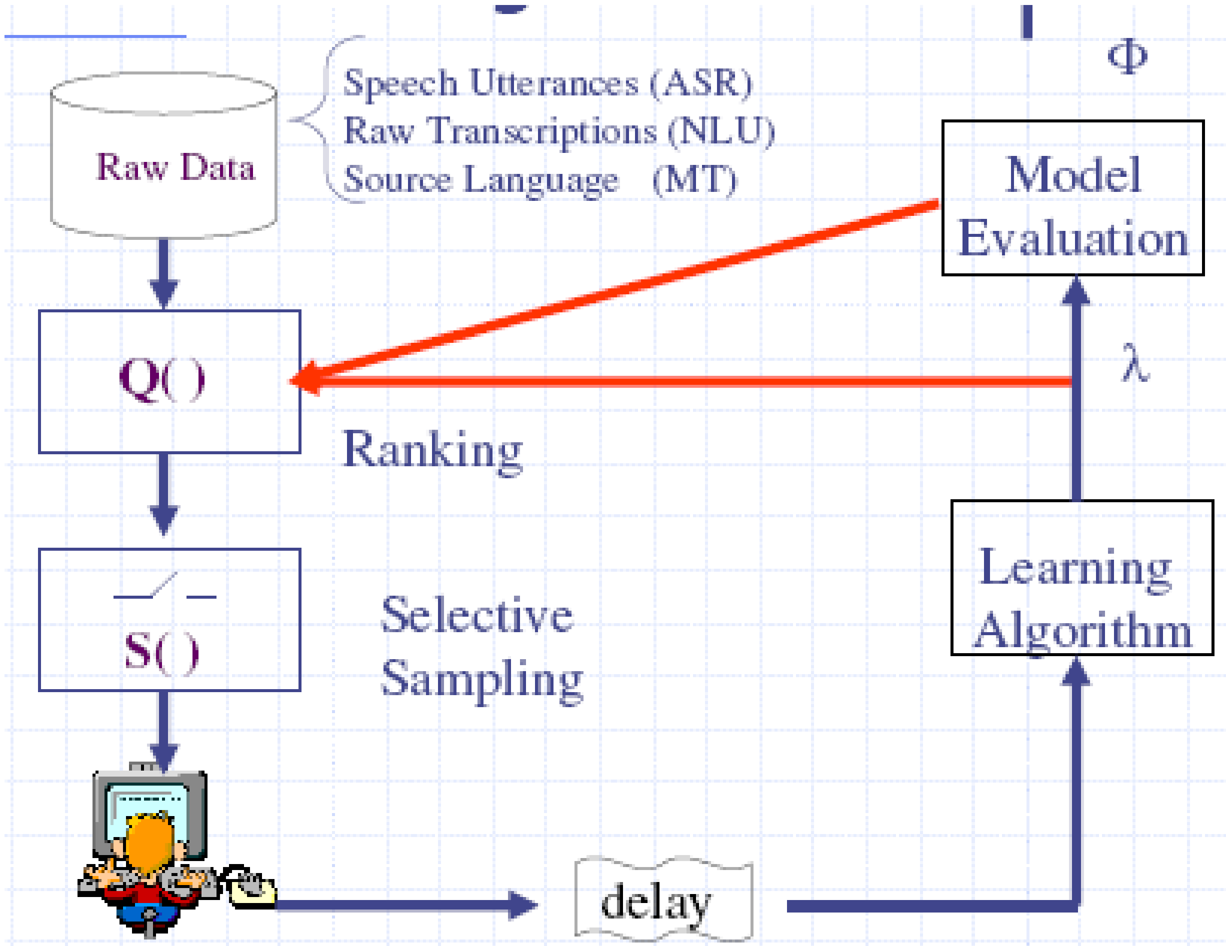
Initially make only the role assignments that are unambiguous according to a verb lexicon ((Kate and Mooney, 2007).

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

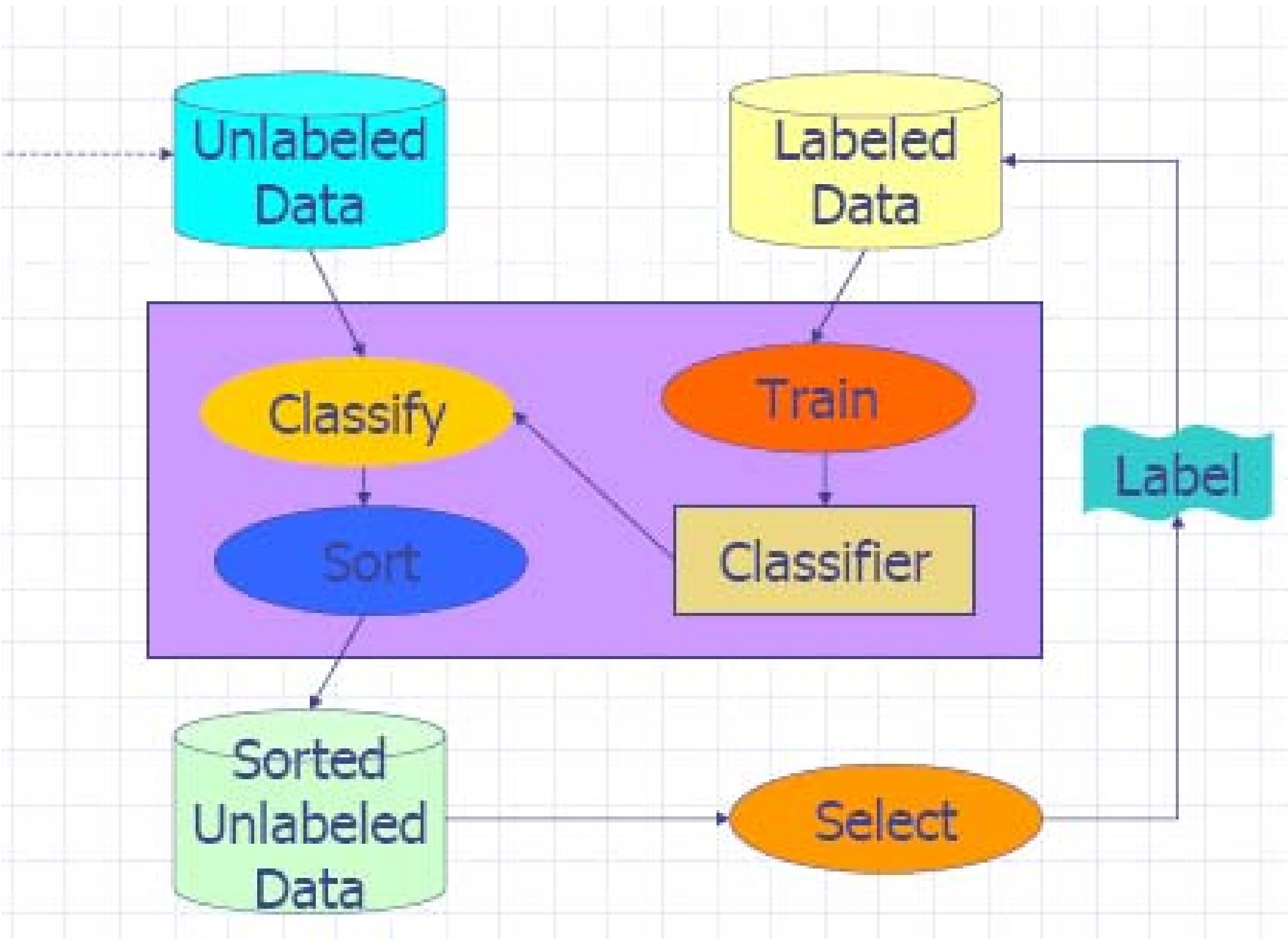
When unlabeled test examples are also available during training, a transductive framework for learning can further improve the performance on the test examples

Active Learning

Hakkani-Tür, Riccardi Gorin, 2002)

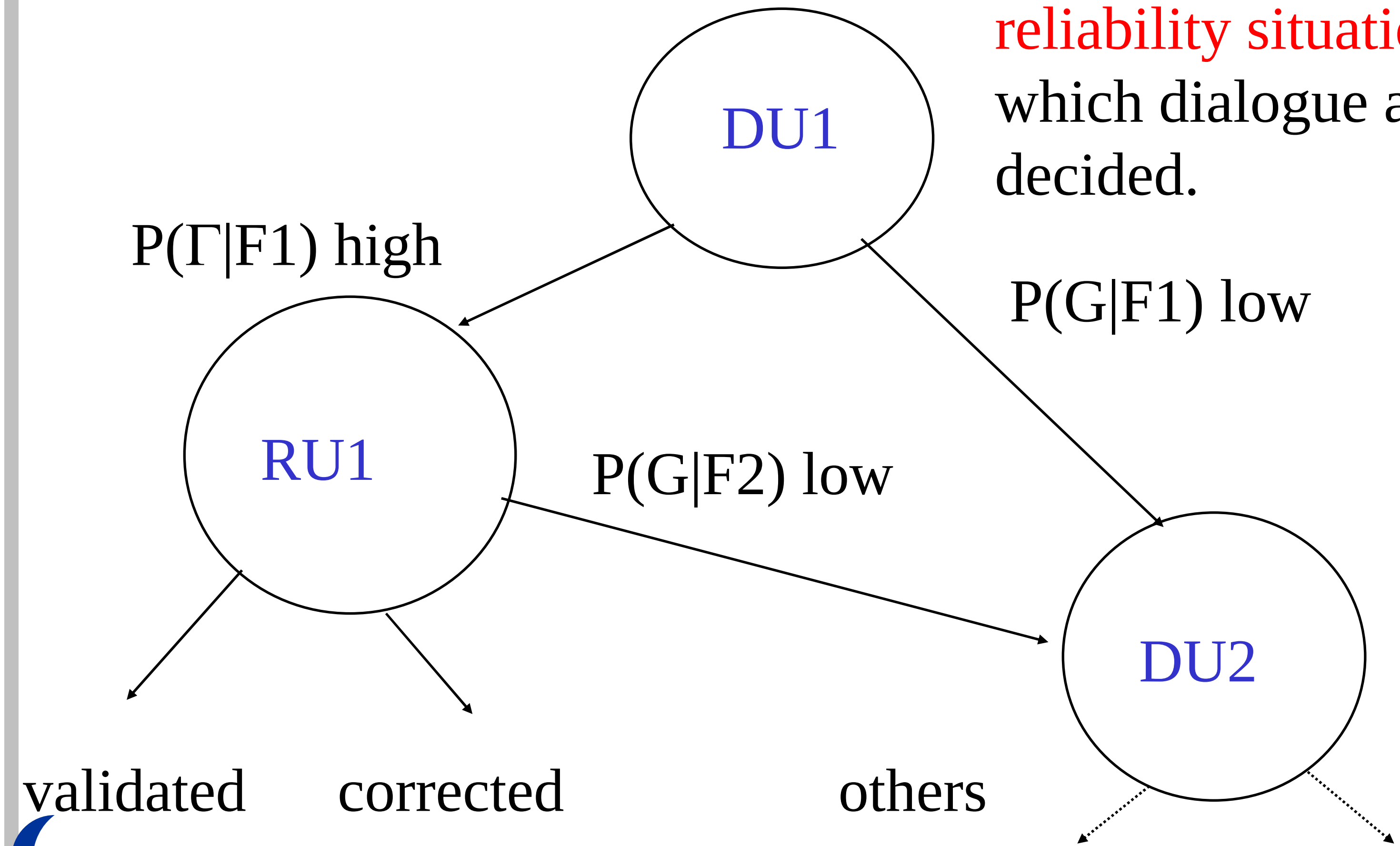


Certainty-Based Active Learning for SLU



Sequential decision using different features sets

Confidence is used to define reliability situations based on which dialogue actions can be decided.



Confidence

Evaluate confidence of components and compositions

$$P(\Gamma | \Phi_{conf})$$

Φ_{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.

Features for confidence

Many features are based on empirical considerations:

semantic weights

assigned to words,

uncovered word percentage,

gap number,

slot number,

word, word-pair and word-triplet

occurrence counts,

Features for confidence

Word counts in an N-best list, lattice density, phone perplexity, language model back-off behaviour, and posterior probabilities

Measures related to the fact that sentences that are grammatically correct and free of recognition errors tend to be easier to parse and the corresponding scores in the parse tree are higher than those of the ungrammatical sentences containing errors generated by the speech recognizer (IBM).

Other features for confidence

In (Lieb, 2005), during slot-value pair extraction, semantic tree node confidences are translated into corresponding slot and value confidences, using a rule-based policy.

In (Higashinaka et al., 2006) it is proposed to incorporate discourse features into the confidence scoring of intention recognition results.

Lin and Wang (2001) propose a concept-based probabilistic verification model, which exploits concept N-grams.

A confidence model is a kind of a classifier that scores or classifies words/concepts based on training data (Hazen, 2002)

Other features for confidence

Use of pragmatic analysis to score concepts uttered by the user (Ammicht et al., 2001).

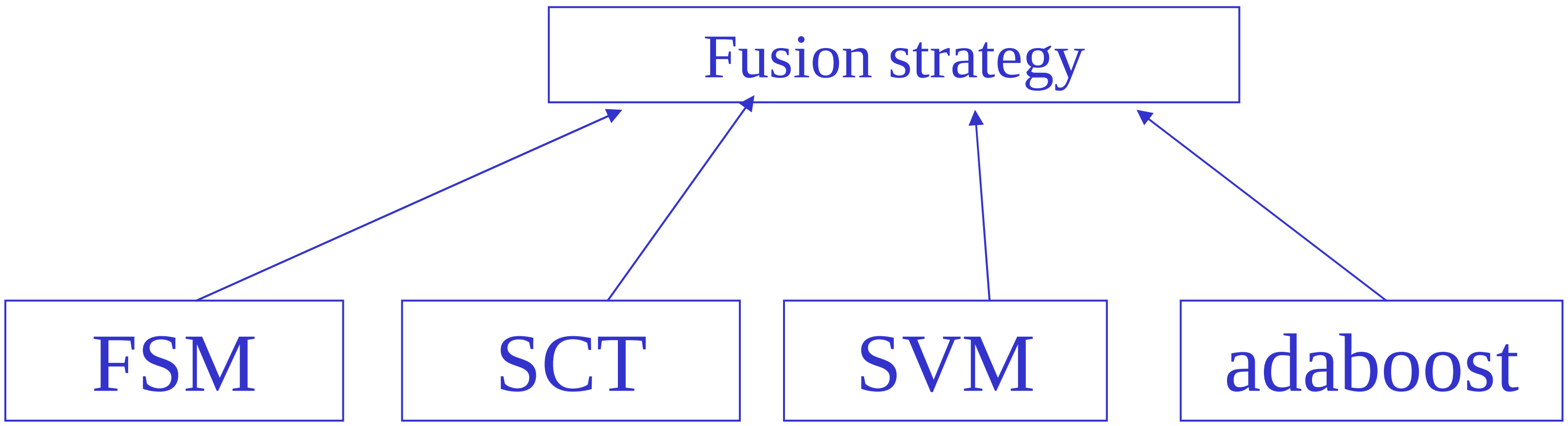
When an already recognized concept seems to have been implicitly confirmed, the confidence of that concept is augmented.

Hirschberg et al. (2004) introduce a number of prosodic features, such as F0, the length of a pause preceding the turn, and the speaking rate.

Combining Confidence Scores with Contextual Features (Purver et al. 2006)

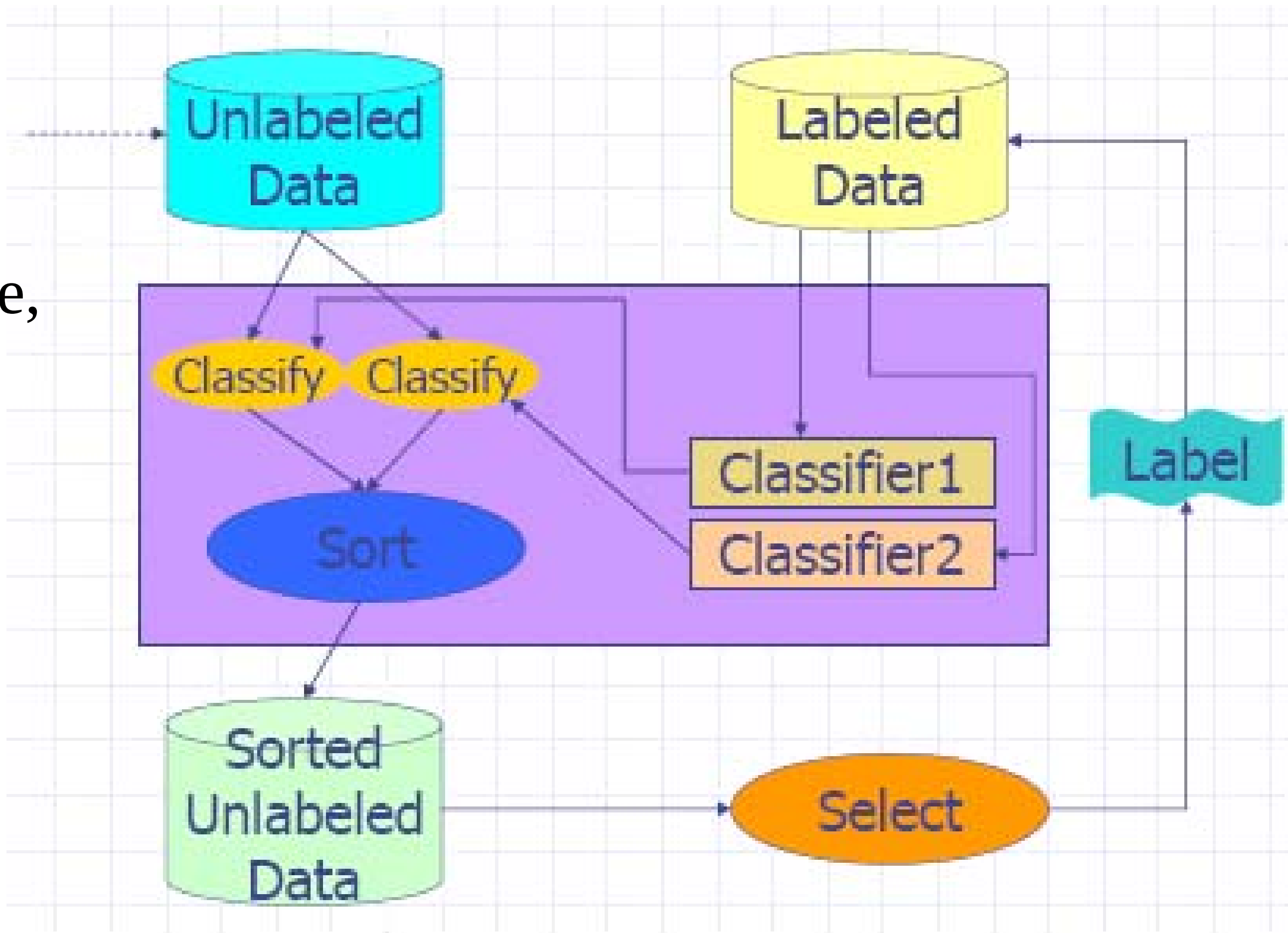
Define confidence-related situations

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)



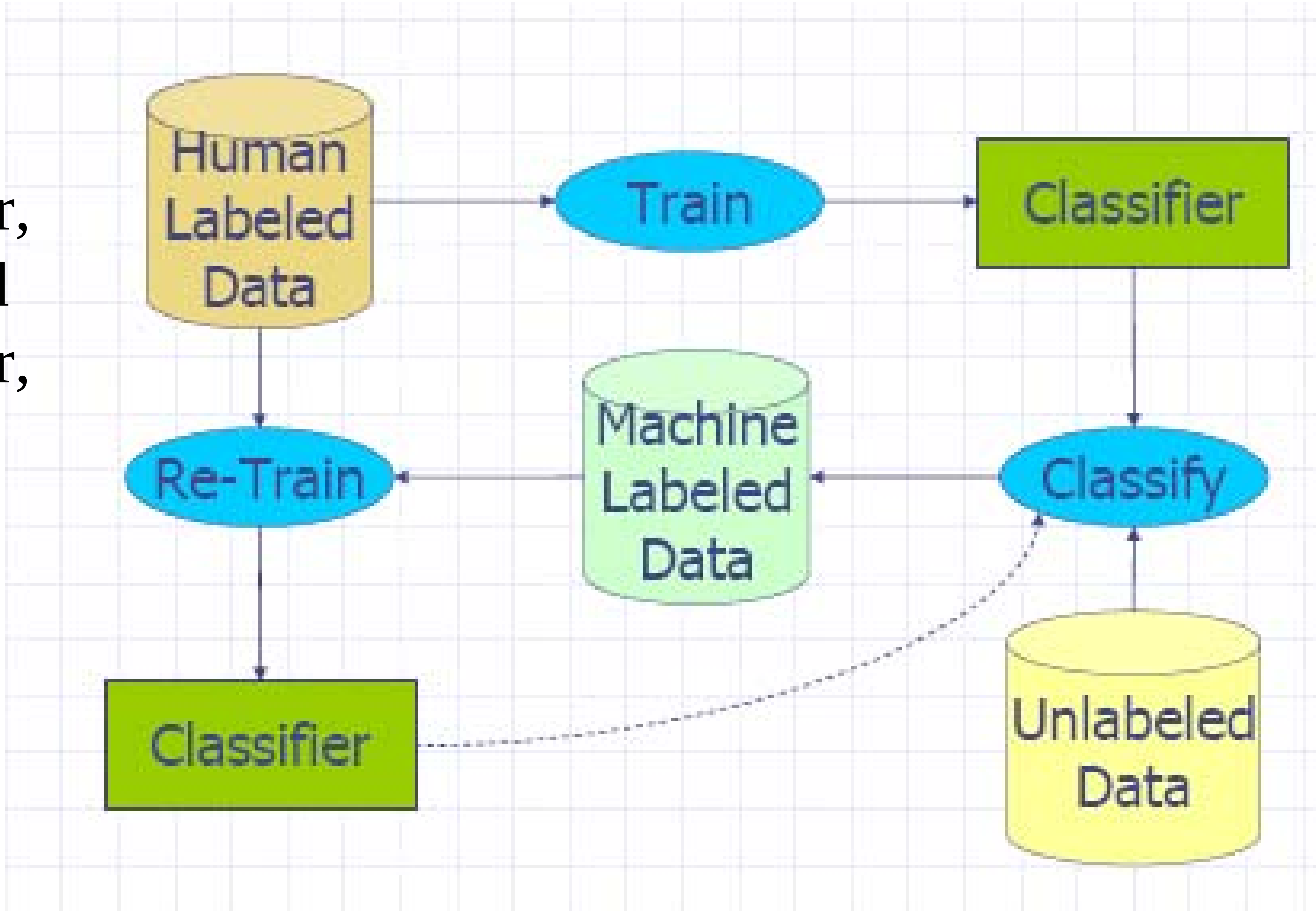
Committee-Based Active Learning

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

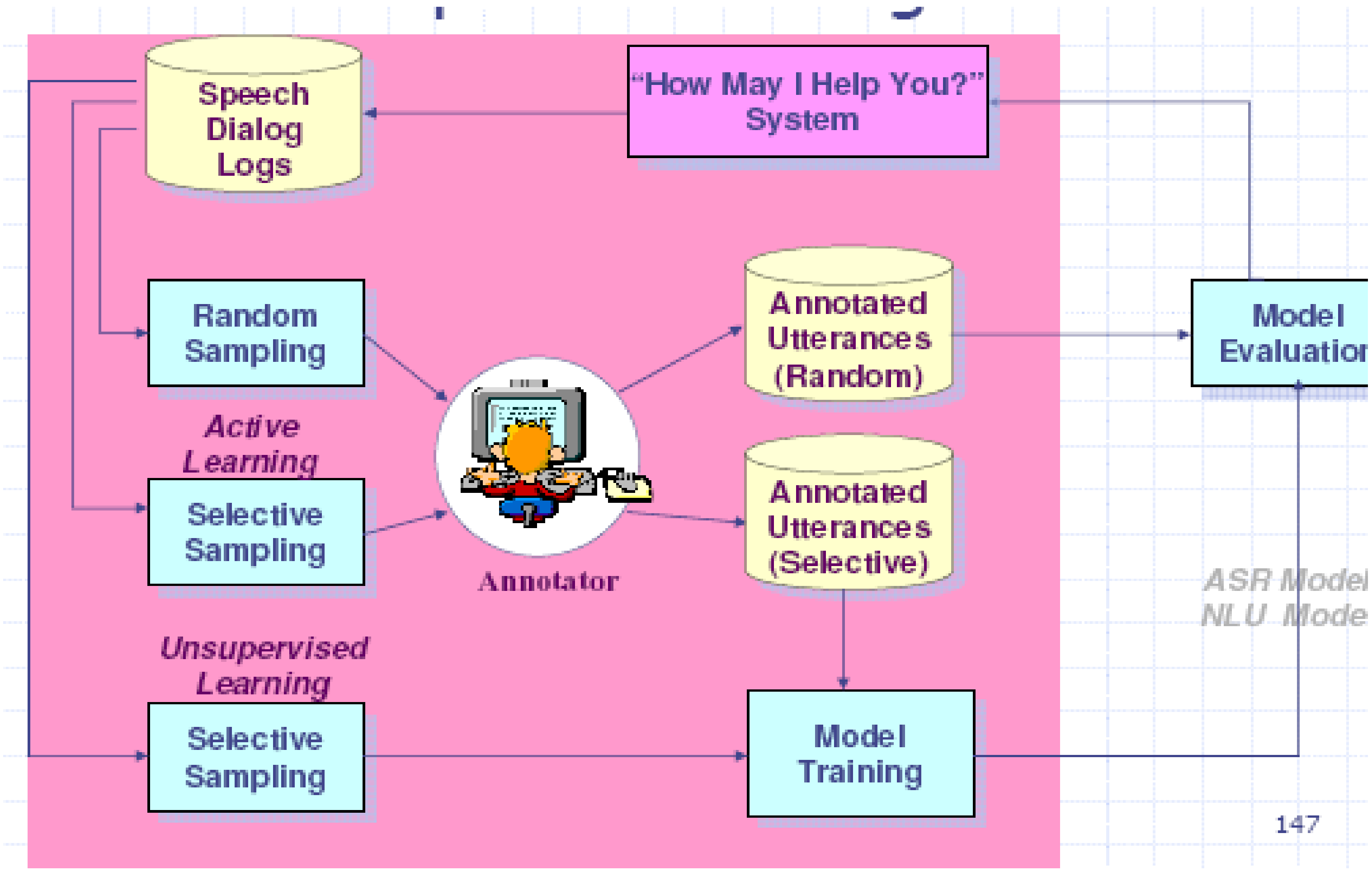
Label and add these selected ones to the training data and retrain

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