

# **Parsing speech: applications and tools developed at the LIF for the syntactic analysis of spontaneous speech**

**Frédéric Béchet**

**Aix Marseille Université  
LIF-CNRS Laboratoire d’Informatique  
Fondamentale de Marseille**



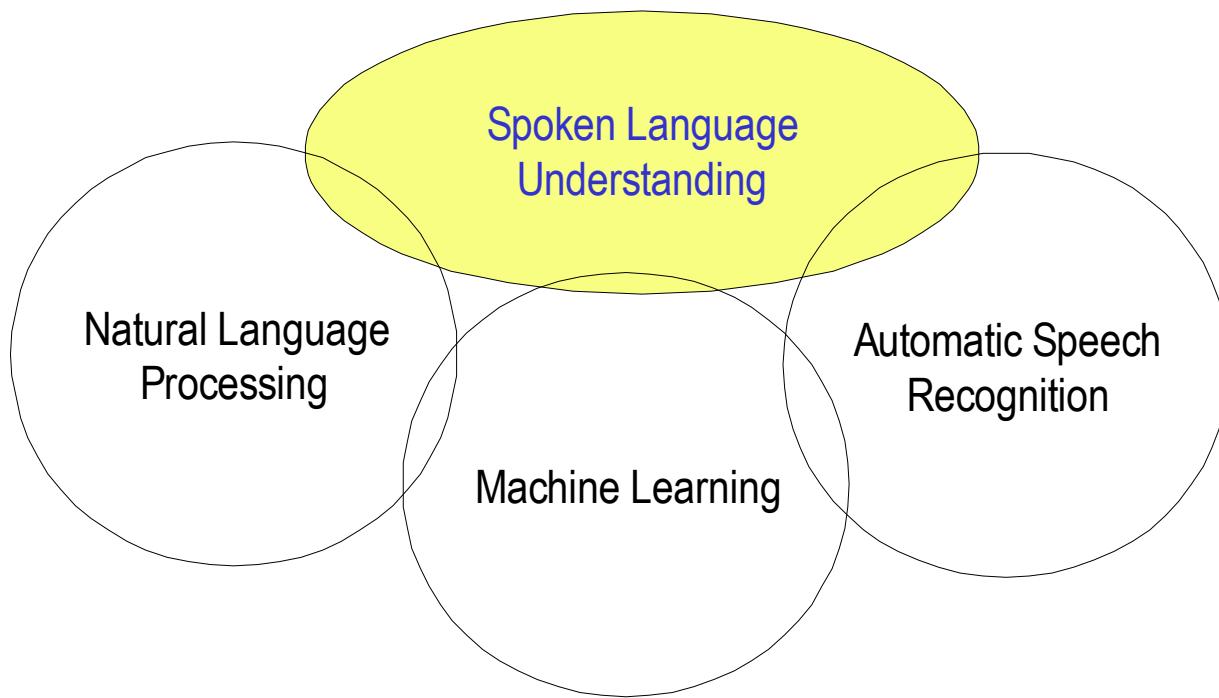
# Overview

- 1. Context of this study : Spoken Language Understanding**
- 2. Main issues in parsing speech**
- 3. Applications**
- 4. Tools & demo**

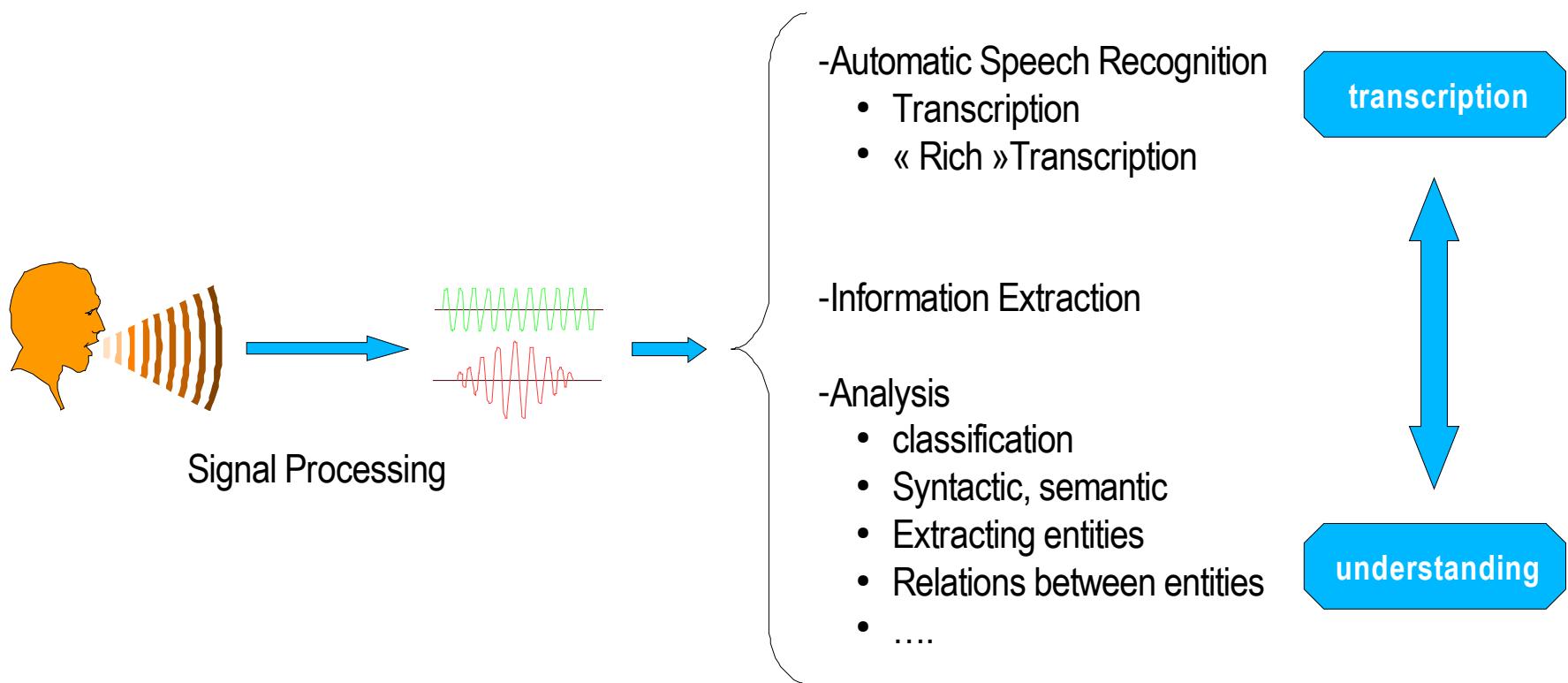


# 1. Context of this study

- **Parsing speech for Spoken Language Understanding (SLU) of spontaneous speech**



# Spoken Language Understanding



# Spoken Language Understanding

- **Two applicative frameworks**
  - Human-Machine spoken dialogue
    - Call-routing
    - Form-filling
    - Negotiation
  - Speech Analytics / Voice Search
    - Broadcast News
    - Large speech archives
    - Call centers
- **Main challenge : processing spontaneous speech**
  - Human/Machine conversation
  - Human/Human conversation



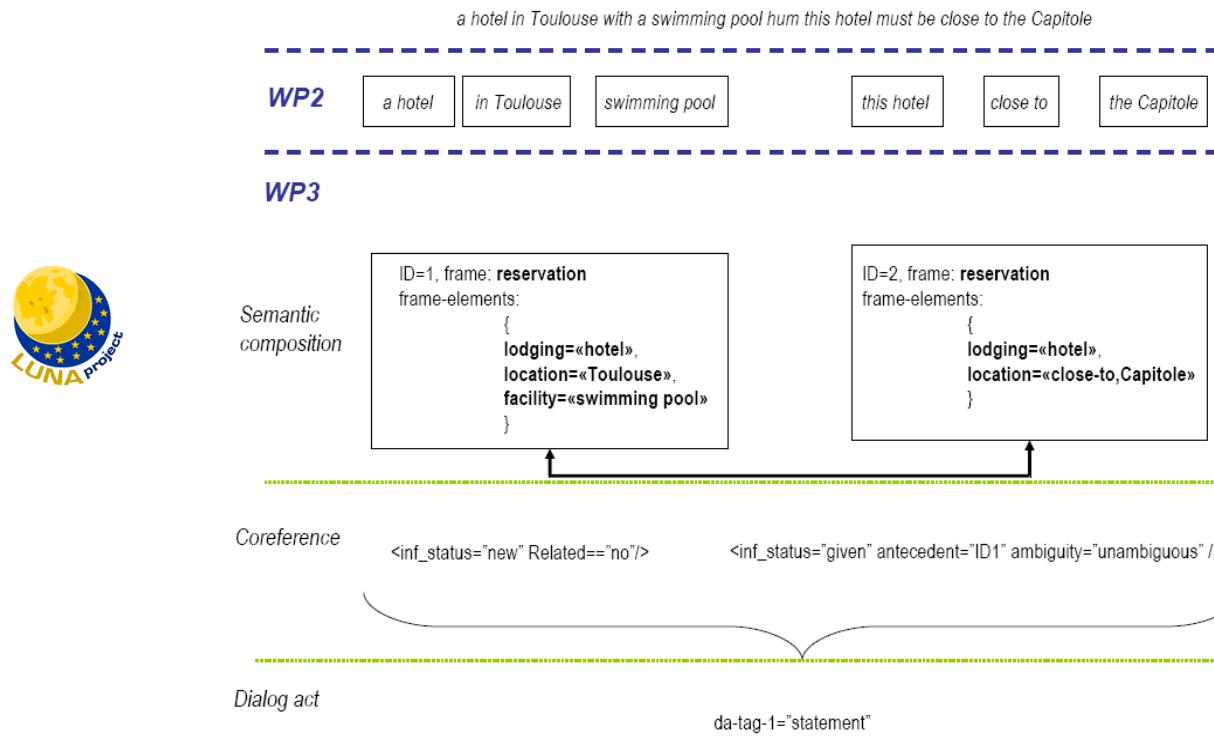
# Spoken Language Understanding

- **Choice of a Meaning Representation Language**
  - General view
    - « meaning » as the composition of basic constituents
    - Definition of constituents and relations independent from a given application
  - Application specific view
    - « interpretation » = « *A representation that can be executed by an interpreter in order to change the state of the system* » (Speech Communication 48 - SLU)
    - Goal of SLU from a system point of view → SYSTEM INTERPRETATION
- **Can be defined by the application framework**
  - In a call-routing application : Call-type
  - In a database query application : SQL query
  - In a directory assistance application : entry in the directory
  - In a speech analytics / voice search application
    - Distillation / Summarization
    - Behavioral analysis



# Spoken Language Understanding

- Can be a formal language
  - Flat representation → « concepts »
    - Named entities, Sequences of keywords, Verbs, ...
  - Structured representation
    - Logical formulae
    - predicate/argument structure (FrameNet) + Dialog act + reference resolution



# Spoken Language Understanding

- **Which models and algorithm to perform SLU?**
  - Different points of view
    - Understanding is a classification task !
    - Understanding is a translation task !
    - Understanding is a parsing task !



# Spoken Language Understanding

- **Understanding is a classification task**
  - Mapping a speech message to a label
    - Call-type, dialog act, ...
  - Direct system interpretation of a message
  - Classifiers
    - SVM, Boosting, decision tree
  - Examples
    - HowMayIHelpYou? (Boosting, SVM)
    - Semantic Classification Trees (on ATIS)
    - Dialog act tagging (CALO - meeting)



# Spoken Language Understanding

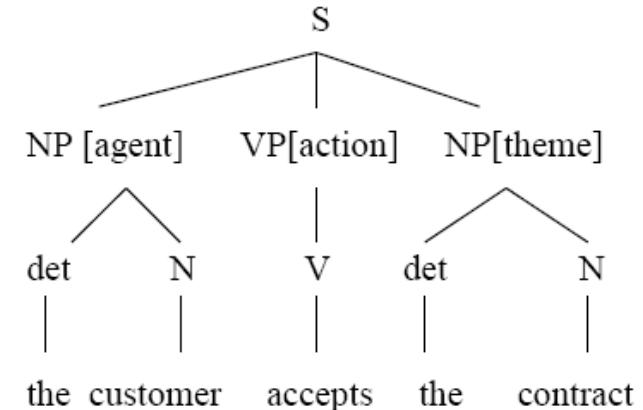
- **Understanding is a translation task**
  - Translation to a formal semantic language
  - Related to tagging approaches (e.g. POS tagging)
  - Main model: Concept decoding
    - Mapping a sequence of words to a sequence of attribute/value tokens
  - Main approaches
    - Hidden Markov Models, MaxEnt tagger, Conditional Random Fields



# Spoken Language Understanding

- **Understanding is a parsing task**

- syntactic/semantic parsing
- Structured representation
  - Constituent tree, dependency tree
- System interpretation
  - obtained directly from the parse tree

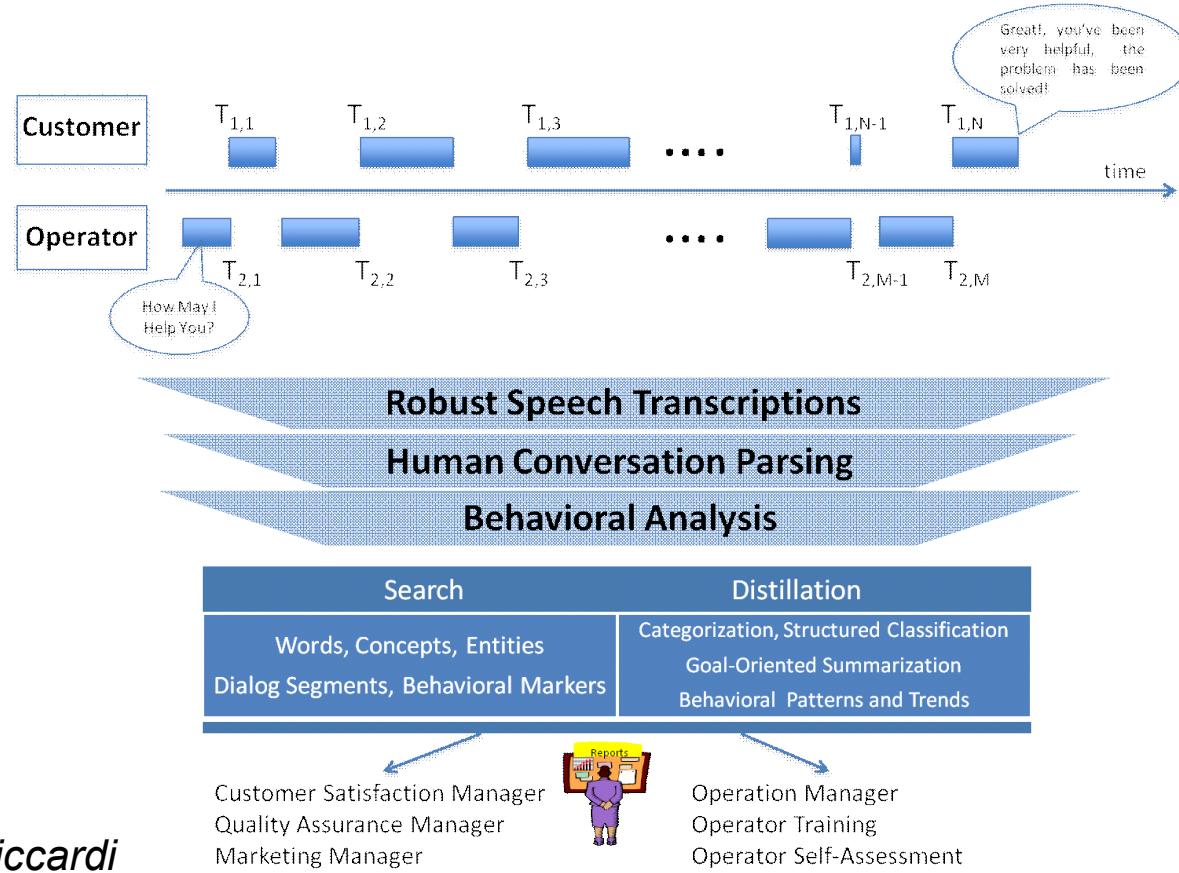


- Examples
  - Full parse : mapping syntactic trees to semantic trees
    - Deep Understanding (Allen, ACL 2007)
  - Shallow parsing
    - Robust parsing TINA (MIT)
  - Parsing + Semantic rôle labelling
    - SLU with parsing + SVM classifiers (Moschitti, ASRU 2007)



# Spoken Language Understanding

- **Understanding can be a combination of these three paradigms**
  - Ex: Complex tasks such as Human Conversation Analysis

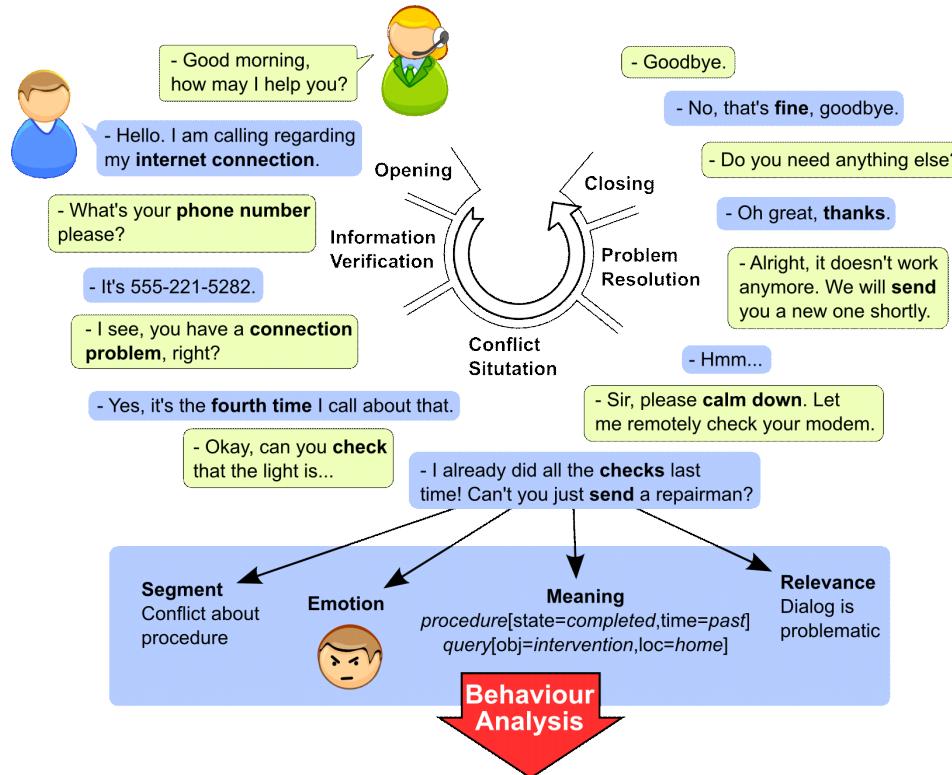


from G. Riccardi



# Spoken Language Understanding

- Understanding can be a combination of these three paradigms
  - Ex: Complex tasks such as Human Conversation Analysis



Answer complex questions  
about conversations:

- Was the operator empathetic to customer's concerns?
- Was the customer satisfied?
- How does the agent manage the argument?
- Did the agent achieve the goal of the call?
- Is the agent polite?

from B. Favre



# Spoken Language Understanding

- Why is syntactic parsing needed for SLU ?
  - Syntactic relations  $\Rightarrow$  semantic relations
    - From syntactic dependency to semantic dependency
      - Ex: *The CoNLL-2008 Shared Task on Joint Parsing of Syntactic and Semantic Dependencies*
  - Semantic disambiguation
    - By identifying the arguments of a verb and using semantic constraints on the arguments
    - Ex: in the French ESTER2 BN Named Entity evaluation program
      - Disambiguation between Organization / Location entities
      - “**France** has made a statement about ..”
      - “I’ll visit **France** next year”
  - Identifying the language register
    - Read speech / prepared speech / spontaneous speech
    - Speaker role identification



# Overview

## 1. Context of this study : Spoken Language Understanding

### 1. Main issues in parsing speech

### 1. Applications

### 2. Tools & demo



## 2. Main issues in parsing speech

- **Speech is not text !!**

- Text
    - « Complete » object: made to be read !!
    - Structured object
    - Designed according to strong specifications
  - Speech
    - « Dynamic» object: edit, repairs, repetition, hesitations
    - Structure = mostly at the acoustic level
    - Speaking style: read/prepared/spontaneous, dialogs

- **Automatic transcription (from ASR) is not text !!**

- ASR output
    - Partial information: all the acoustic dimension of speech is missing
    - Stream of words, no punctuations
  - ASR errors
    - Insertion/deletion/substitution
    - Each word can be weighted by a confidence score

Chapitre 1																		
Introduction																		
Sommaire																		
<table><tr><td>1.1 Avant-propos .....</td><td>11</td></tr><tr><td>1.2 Cadre de l'étude .....</td><td>13</td></tr><tr><td>    1.2.1 Modèles numériques .....</td><td>13</td></tr><tr><td>    1.2.2 Compréhension linguistique observable .....</td><td>13</td></tr><tr><td>    1.2.3 Vues d'ensemble des thèmes abordés .....</td><td>14</td></tr><tr><td>        1.3.1 Traitement automatique des langages et apprendre à parler .....</td><td>14</td></tr><tr><td>        1.3.2 Reconnaissance Automatique de la Parole .....</td><td>16</td></tr><tr><td>        1.3.3 Compréhension de la parole spontanée .....</td><td>17</td></tr><tr><td>        1.4 Plan du document .....</td><td>19</td></tr></table>	1.1 Avant-propos .....	11	1.2 Cadre de l'étude .....	13	1.2.1 Modèles numériques .....	13	1.2.2 Compréhension linguistique observable .....	13	1.2.3 Vues d'ensemble des thèmes abordés .....	14	1.3.1 Traitement automatique des langages et apprendre à parler .....	14	1.3.2 Reconnaissance Automatique de la Parole .....	16	1.3.3 Compréhension de la parole spontanée .....	17	1.4 Plan du document .....	19
1.1 Avant-propos .....	11																	
1.2 Cadre de l'étude .....	13																	
1.2.1 Modèles numériques .....	13																	
1.2.2 Compréhension linguistique observable .....	13																	
1.2.3 Vues d'ensemble des thèmes abordés .....	14																	
1.3.1 Traitement automatique des langages et apprendre à parler .....	14																	
1.3.2 Reconnaissance Automatique de la Parole .....	16																	
1.3.3 Compréhension de la parole spontanée .....	17																	
1.4 Plan du document .....	19																	
1.1 Avant-propos																		
Ce document présente nos travaux réalisés dans le cadre du Traitement Automatique de la Parole (TAP), et plus particulièrement ceux concernant la compréhension automatique de l'ensemble des langages humains. Il y a eu une dizaine d'années, à travers la collaboration avec le Professeur René de Mori, des travaux fondamentaux sur la modélisation du langage dans le cadre du dialogue oral. Celle-these effectuée par Yannick Estève aujourd'hui enseignant-chercheur à l'Université du Mans, c'est faite à l'occasion d'un premier partenariat avec le laboratoire de recherche de France Télécom à Lannion, le CNET, Lannion. C'est dans ce contexte que l'Institut de l'Informatique de Paris en France (et qui dépend actuellement de nom .....) a continué la première recherche d'être confronté à des données relatives de dialogue homme-machine, d'envisager des tâches dépassant le paradigme de la dictée vocale qui reste le modèle dominant en Reconnaissance Automatique de la Parole (RAP) et en effet d'approuver le problème de la compréhension de la parole spontanée. Ce travail sur les données, les corpus contenant de la parole complètement spontanée collectée grâce à des applications mises en service,																		
11																		



# Automatic transcriptions

- **Non-native text**

- *Generation* of a transcription  $W$  from a signal  $A$  :

$$P(\hat{W}|A) = \max_W P(W|A)$$

$$P(W|A) = \frac{P(A|W)P(W)}{P(A)}$$

- Therefore  $\Rightarrow$  *closed vocabulary*
    - All words are included in the generative model  $P(W)$
    - *Automatic transcriptions contain no unknown words !!*
  - Sentence segmentation
    - pause, fixed duration, a posteriori process based on prosodic/lexical features
  - Disfluencies
    - Badly recognized
  - Multiple hypotheses with confidence scores
    - Word lattices, word confusion networks, n-best lists



# Transcribing spontaneous speech ?

euh bonjour donc c' est XX à l' appareil je sais pas si vous savez très bien qui je suis euh donc par rapport au à la niveau de la satisfaction de ma satisfaction personnelle par rapport à votre service euh je dirai que dans l' ensemble je suis euh plutôt satisfait euh vous avez un très bon service clientèle qui sait écouter qui euh non qui j' ai pas grand chose à dire c' est c' est très très bien sinon ben juste par rapport au à ce que vous avez mis en place euh tout de suite justement c' est une très bonne idée justement d' une façon à ce qu' y ait un taux de réponse euh assez important maintenant c' est vrai qu' on est obligé de rappeler plusieurs fois et encore quand on prend le temps de rappeler pour euh pour euh pour euh pour répondre parce que quand on nous dit euh vous allez nous donner vous allez donner euh votre euh vos idées euh vos vos suggestions et ben on n' a rien en tête donc c' est pour ça que j' ai été obligé de raccrocher et de réfléchir à ce que je vais vous dire on ça c' est pas je pense euh que ça c' est le point collectif ou c' est le point négatif et sinon dans l' ensemble je suis très satisfait sinon y a une chose que j' ai à noter euh j' ai deux comptes chez vous euh je trouve ça un peu embêtant de pouvoir euh de pas pouvoir accéder euh aux deux par la même personne quand j' appelle mon service clientèle donc ça je trouve ça un peu dommage que je soit obligé de dépenser en plus parce que faut que je c' est pas le même type euh c' est pas la même personne qui s' occupe de mon dossier donc ce qui aurait été bien c' est quand même regrouper les deux dossiers sous euh euh sous un seul quoi de façon à ce que quand on appelle on puisse accéder aux deux dossiers séparément bien sûr mais les deux dossiers donc voilà euh sinon ben je vous remercie en tous cas pour euh pour votre gentillesse et votre amabilité vos conseillers clientèle sont très très gentils et très à l' écoute et donc je vous en remercie au revoir bonne journée bonne soirée

# Transcribing spontaneous speech ?

euh bonjour donc c' est XX à l' appareil je sais pas si vous savez très bien qui je suis euh donc par rapport au à la niveau de la satisfaction de ma satisfaction personnelle par rapport à votre service euh je dirai que dans l' ensemble je suis euh plutôt satisfait euh vous avez un très bon service clientèle qui sait écouter qui euh non qui **j' ai pas grand chose à dire** c' est c' est très très bien sinon ben juste par rapport au à ce que vous avez mis en place euh tout de suite justement c' est une très bonne idée justement d' une façon à ce qu' y ait un taux de réponse euh assez important maintenant c' est vrai qu' on est obligé de rappeler plusieurs fois et encore quand on prend le temps de rappeler pour euh pour euh pour euh pour répondre parce que quand on nous dit euh vous allez nous donner vous allez donner euh votre euh vos idées euh vos vos suggestions et ben on n' a rien en tête donc c' est pour ça que j' ai été obligé de raccrocher et de réfléchir à ce que je vais vous dire on ça c' est pas je pense euh que ça c' est le point collectif ou c' est le point négatif et sinon dans l' ensemble **je suis très satisfait** sinon y a une chose que j' ai à noter euh j' ai deux comptes chez vous euh je trouve ça un peu embêtant de pouvoir euh de pas pouvoir accéder euh aux deux par la même personne quand j' appelle mon service clientèle donc ça je trouve ça un peu dommage que je soit obligé de dépenser en plus parce que faut que je c' est pas le même type euh c' est pas la même personne qui s' occupe de mon dossier donc ce qui aurait été bien c' est quand même regrouper les deux dossiers sous euh euh sous un seul quoi de façon à ce que quand on appelle on puisse accéder aux deux dossiers séparément bien sûr mais les deux dossiers donc voilà euh sinon ben je vous remercie en tous cas pour euh pour votre gentillesse et votre amabilité vos conseillers clientèle sont très très gentils et très à l' écoute et donc je vous en remercie au revoir bonne journée bonne soirée

# Parsing automatic transcriptions

- **What kind of syntactic parsing ?**
  - “traditional” approaches based on CFG not suitable for processing spontaneous speech
  - Approaches to parsing based on dependency structures and discriminative machine learning techniques (McDonald et al., 2007) are much easier to adapt to speech:
    - they need less training data
    - the annotation with syntactic dependencies of speech transcripts is simpler than with syntactic constituents
    - partial annotation can be performed when the speech is ungrammatical or the ASR transcripts are erroneous
  - The dependency parsing framework also generates parses much closer to meaning which eases semantic interpretation



# Parsing automatic transcriptions

- **Main issues**

- Dealing with speech disfluencies and conversational speech
  - “yes so I hum ....hum I I wou would like”
  - 2 strategies
    - Adding disfluencies to the parsing models
    - Removing disfluencies prior to parsing
- Dealing with ASR errors
  - Using word confidence scores in the parsing process?
  - Insertion / substitution / deletion
- Integration into the ASR process
  - At the Language Model level?
  - On ASR output
    - word string? word lattice? confusion network? N-best list?



# Parsing automatic transcriptions

- **Approaches developed at the LIF**
  - Developing specific resources for parsing spontaneous speech
    - POS tagger / Chunker
    - DECODA project
  - Adapting dependency parsers on small amount of annotated speech
  - Using existing linguistic resources
    - Syntactic/semantic lexicons
  - Integration into the ASR process at the word lattice level
    - Building NLP tools taking weighted word lattices as input
      - Keeping alternative word string and word/sentence segmentation
      - Possibility to add paths removing disfluencies
      - For each analysis produced: possibility to go “back” to the signal
    - *N-best lists are not adequate !!*



# Overview

- 1. Context of this study : Spoken Language Understanding**
- 2. Main issues in parsing speech**
  - 1. Applications**
  - 1. Tools & demo**



### 3. Applications

- **Extracting dependencies on a set of selected target verbs**
  - SLU for spoken dialog systems (LUNA project)
  - Joint work with Alexis Nasr (LIF)
- **Detecting and characterizing interrogative sentences for speaker role detection**
  - Speech mining on Broadcast Conversations (EPAC project)
  - Joint work with Thierry Bazillon (LIF), Alexis Nasr (LIF), Benjamin Maza (LIA), Mickael Rouver (LIA)

# SLU for spoken dialog systems (LUNA project)

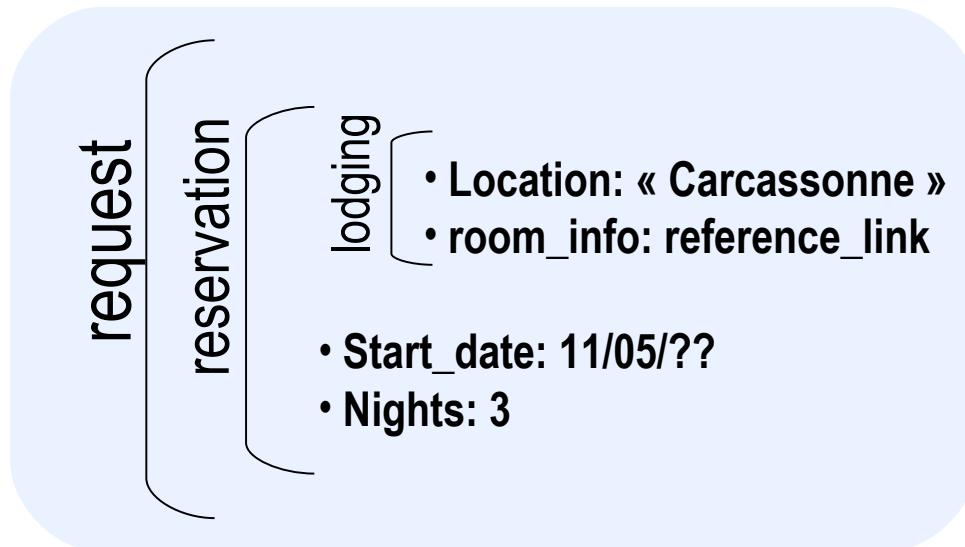
- **Context of this study**
  - Robust Spoken Language Understanding (SLU) for advanced Spoken Dialog Systems
    - FP6 LUNA Project
- **Goals**
  - Using dependency parsing to produce features used by the SLU component
  - Reducing the human effort needed to adapt the parser by using generic linguistic resources
  - Integrating the parsing process into the ASR/SLU pipeline
- **Experiments**
  - MEDIA-LUNA corpus
    - Spoken dialog corpus (hotel booking) annotated with conceptual units and semantic frames



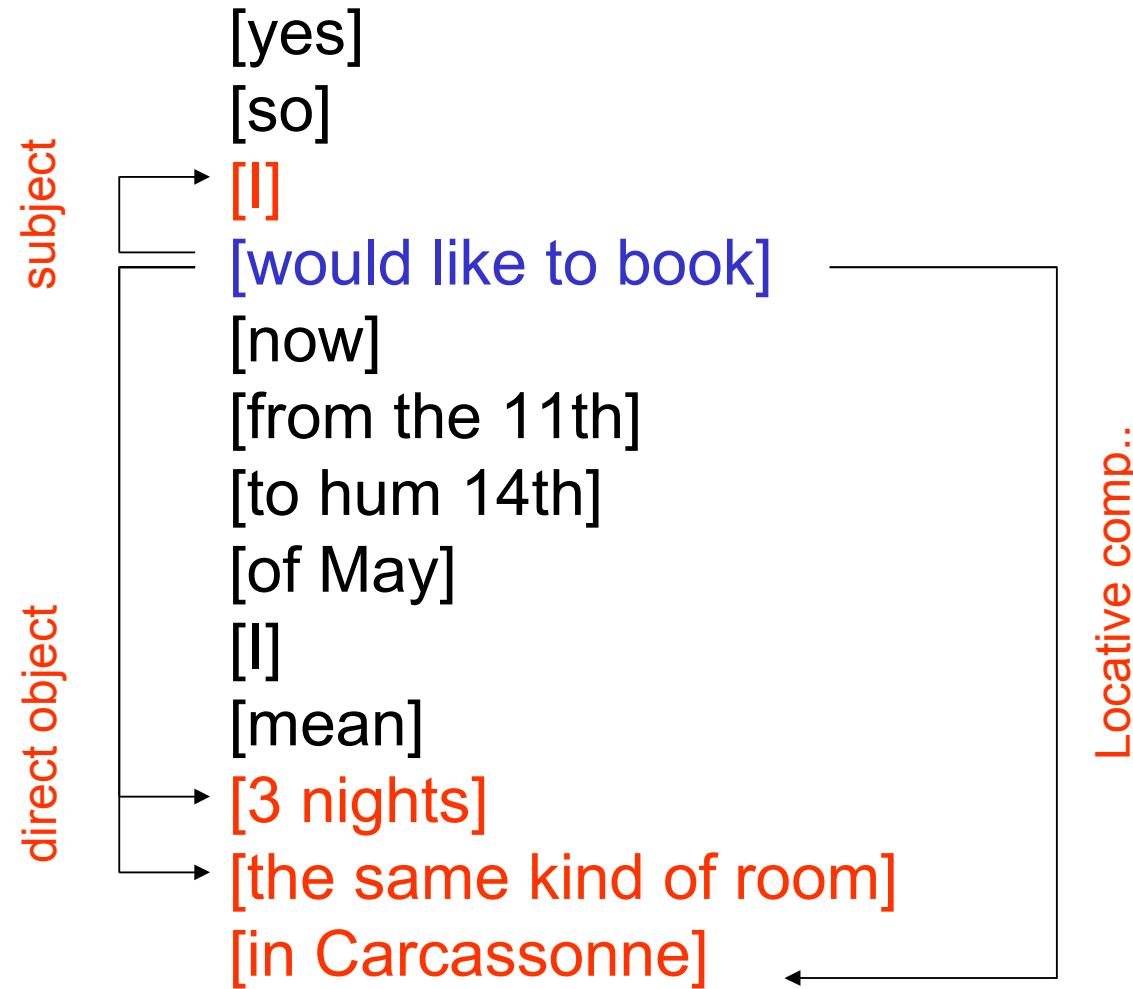
## Example

« oui donc j' aimerais réserver maintenant du 11 au euh  
14 mai c'est-à-dire 3 nuits la même prestation à Carcassonne »

« yes so I would like to book now from the 11th to hum 14th of May I  
mean 3 nights the same kind of room in Carcassonne »



# Example



# Dependency parsing and speech recognition

- **Solutions?**
  - Grammar adaptation + standard parsing process
    - Writing a grammar specific to an application
      - Semantic domain, interaction type
      - Robustness to disfluencies and ASR errors
    - Training a parser on a transcribed corpus with syntactic annotations
  - Generic linguistic resources + parsing process adaptation
    - Processing multiple inputs with confidence scores
      - word lattice, sausage, n-best list
    - Extraction of the dependencies relevant for a given application
      - No full parse needed



# Dependency parsing and speech recognition

- **Solutions?**
  - Grammar adaptation + standard parsing process
    - Writing a grammar specific to an application
      - Semantic domain, interaction type
      - Robustness to disfluencies and ASR errors
    - Training a parser on a transcribed corpus with syntactic annotations
  - Generic linguistic resources + parsing process adaptation
    - Processing multiple inputs with confidence scores
      - word lattice, sausage, n-best list
    - Extraction of the dependencies relevant for a given application
      - No full parse needed



**Limited adaptation to a specific application**

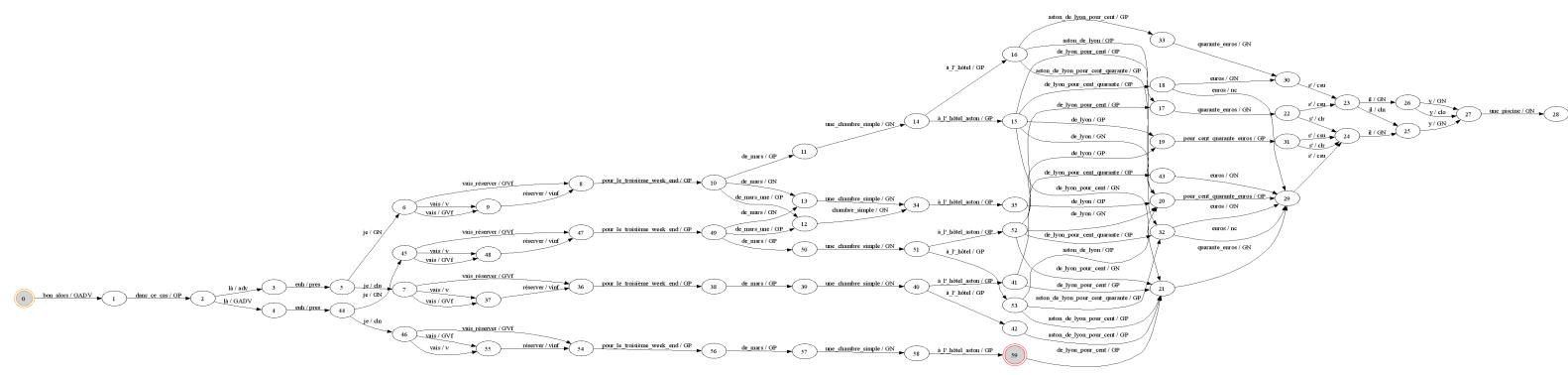
# Generic linguistic ressources

- Shallow parsing toolkit                    **MACAON toolkit**
- Syntactic lexicon                        **Dicovalence**
- Grammar                                  **FXMG grammar**



# Generic linguistic ressources

- **Shallow parsing toolkit: MACAON**
    - HMM based POS tagger
    - Syntactic chunk grammars
    - Finite-state-Machine approach
    - Take as input text represented as weighted FSMs
      - Any ASR output
    - Output FSM



## Syntactic chunk lattice

# Generic linguistic ressources

- **Syntactic lexicon: Dicovalence (P. Mertens, Leven)**
  - Description of syntactic sub-categorisation of 3700 French verbs
    - *Subcat frames*
  - 3700 verbs, 8000 entries, 311 frames
  - Very limited syntactic coverage
- **Grammar: FXMG**
  - TAG grammar generated from a meta grammar (Crabbe, Paris VII)
  - 7600 syntactic trees grouped in 92 clusters
- **Combining FXMG with Dicovalence**
  - Association of a tree cluster to each entry in Dicovalence
  - Each tree is a possible syntactic realization of a verb-subcategorization



# Dependency parsing strategy

- **Extracting the possible arguments of a verb from a lattice of syntactic chunks**
  - Defining a set of « target verbs » relevant for the dialog domain (e.g. “to book”)
  - For each occurrence  $v$  of these verbs in a chunk lattice
    - For each subcat frame  $\Psi$  of  $v$ 
      - Looking for all the possible syntactic realisations of  $\Psi$  in the chunk lattice
  - Reranking of all the possible dependency hypotheses according to the target domain



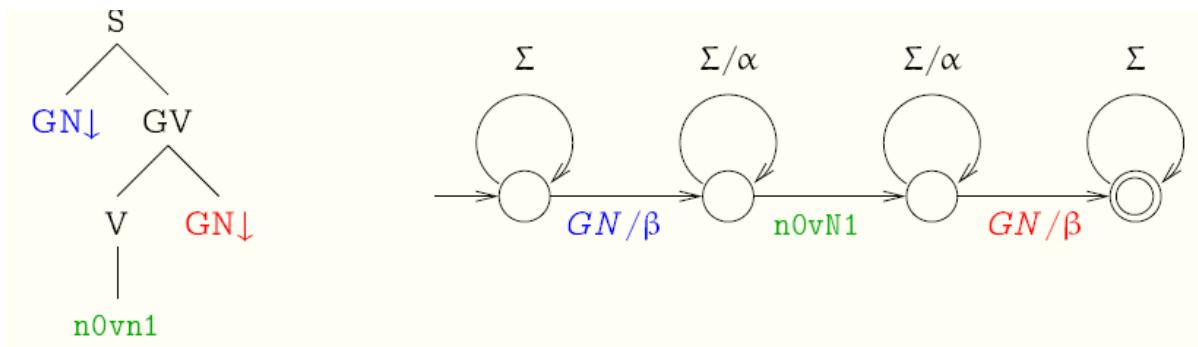
# Dependency parsing strategy

- **No full syntactic parse needed**
  - Strong syntactic constraint relaxation for dealing with speech disfluencies
    - Relative order of the arguments only
      - A dependency can “jump” over any chunk or sequence of chunks
  - Taking as input chunk lattices for dealing with ASR errors
    - Dealing with insertion and substitution errors
- **Adaptation effort**
  - Limited to the reranking process
  - No adaptation of the linguistic resources



# Extracting dependencies from a chunk lattice

- **Finite State Machine approach**
  - Building automata from the tree clusters
    - Each elementary tree is turned into an automaton

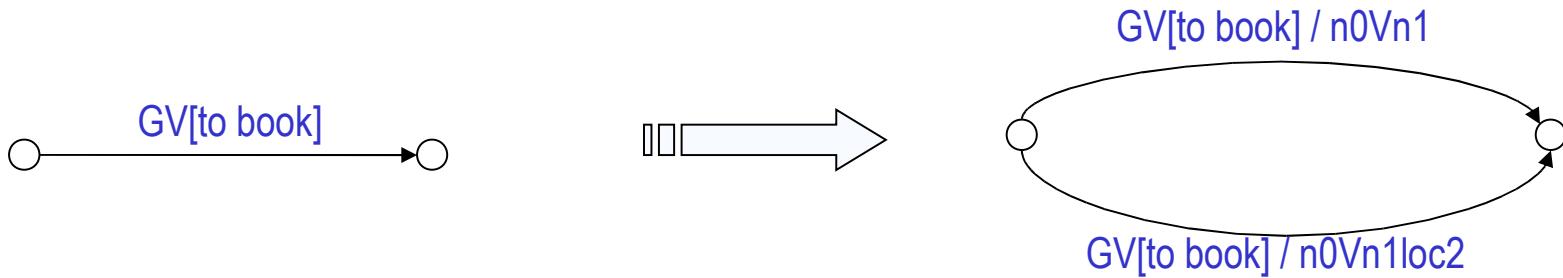


- Union of all the automata for each cluster:  $T_{F1} \dots T_{Fn}$
- Syntactic constraint relaxation with the  $\Sigma$  transitions
  - Dramatic increase of the ambiguity !!
- Heuristic: prefer close dependencies rather than distant ones
  - Weights:  $\alpha > 0 \quad \beta < 0$

# Extracting dependencies from a chunk lattice

- **Process**

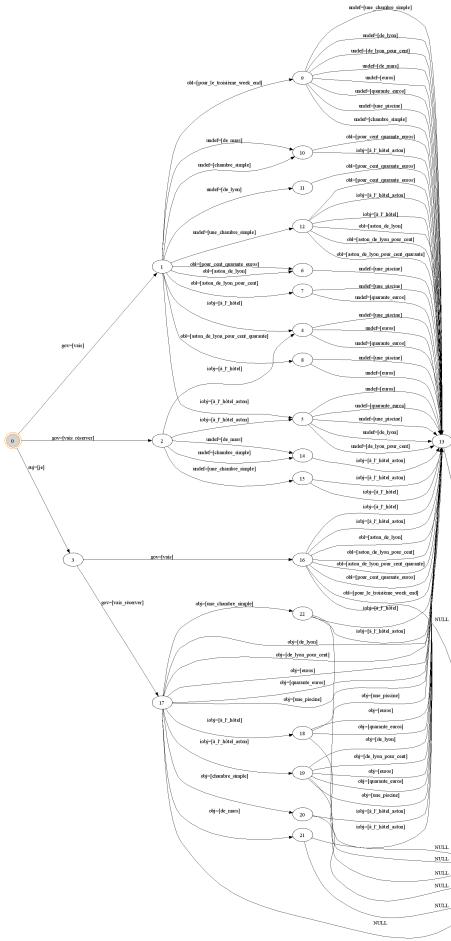
- Producing a lattice of chunks  $T_s$  from an ASR word lattice
- For each target verb  $v$  found in the lattice
  - All the arcs labeled with  $v$  are replaced by a set of arcs  $v/\psi_i$  where  $\psi_i$  are all the subcat frames associated with  $v$  in Dicovalence/FXMG



- Compose  $T_s$  with all the  $T_{Fi}$  automata  $\Rightarrow$  *dependency lattice*
- Produce the n-best list of dependency hypotheses

# Extracting dependencies from a chunk lattice

- Ex: dependency lattice and n-best list of dependency hypotheses



N	Subcat	Subject	GV	Direct object	Locative comp.
1	n0V	j'	aimerais réservoir	NULL	NULL
2	n0Vn1	j'	aimerais réservoir	quatorze mai <b>stade</b>	NULL
3	n0Vn1	j'	aimerais réservoir	trois nuits	NULL
4	n0Vn1loc2	j'	aimerais réservoir	quatorze mai <b>stade</b>	à carcassonne
5	n0Vn1loc2	j'	aimerais réservoir	la même prestations	à carcassonne
6	n0Vn1loc2	j'	aimerais réservoir	trois nuits	à carcassonne
7	n0Vn1	j'	aimerais réservoir	la même prestations	NULL
8	n0Vloc1	j'	aimerais réservoir	NULL	à carcassonne
9	n0Vn1	j'	aimerais réservoir	stade	NULL
10	n0Vn1loc2	j'	aimerais réservoir	la même	à carcassonne

Score=combination of ASR, POS tagger,  
Chunker and dependency heuristic scores

# Reranking dependency n-best lists

- **Method**

- Training a classifier for labeling each hypothesis as correct/incorrect
- Using the classification scores for reranking the n-best list

N	Subcat	Subject	GV	Direct object	Locative comp.
1	n0V	j'	aimerais réservoir	NULL	NULL
2	n0Vn1	j'	aimerais réservoir	quatorze mai <b>stade</b>	NULL
3	n0Vn1	j'	aimerais réservoir	trois nuits	NULL
4	n0Vn1loc2	j'	aimerais réservoir	quatorze mai <b>stade</b>	à carcassonne
5	n0Vn1loc2	j'	aimerais réservoir	la même prestations	à carcassonne
6	n0Vn1loc2	j'	aimerais réservoir	trois nuits	à carcassonne
7	n0Vn1	j'	aimerais réservoir	la même prestations	NULL
8	n0Vloc1	j'	aimerais réservoir	NULL	à carcassonne
9	n0Vn1	j'	aimerais réservoir	stade	NULL
10	n0Vn1loc2	j'	aimerais réservoir	la même	à carcassonne



« oui donc j' aimerais réservoir maintenant du onze au euh quatorze mai c'est-à-dire trois nuits la même prestation à carcassonne »

# Reranking dependency n-best lists

- **Training process**
  - Processing the reference transcriptions of a development corpus with the dependency parser
    - n-best list of dependency hypotheses for each dialog turn
  - For each dialog turn, a human annotator marks the correct hypotheses in the n-best list
  - “light” manual annotation
    - Limited to target verbs
    - Yes/No annotation
  - Training a classifier on the annotated corpus
    - Each hypothesis represented by a set of lexical/syntactic features
    - Boosting-based classifier
      - **IcsiBoost** (<http://code.google.com/p/icsiboost/>)



# Evaluation

- **Corpus**
  - MEDIA: hotel booking dialogs (WOZ+users)
    - Reference transcriptions
    - ASR word lattices
  - First experiment: target verb = « to book » (réserver)

<b>Corpus</b>	<b>Train</b>	<b>Dev.</b>	<b>Test</b>
# dialogue	727	79	208
# turn	12988	1265	3524
# target verb	659	73	187
WER	14.5	25.3	27.4



# Evaluation

- **Evaluation 1 on the whole dependencies**

- A dependency hypothesis is correct iff
  - The chunk containing the target verb is correct
    - words + chunk label
  - The subcat frame is correct
  - All the dependency constituents (chunks) are correct
    - words + chunk label + syntactic role

P=precision  
R=recall  
F=F-measure

Corpus	ref. trans.			ASR trans.		
	P	R	F	P	R	F
results						
Oracle	90.0	90.0	90.0	63.2	63.2	63.2
Baseline	42.8	38.3	40.4	34.7	29.2	31.6
Rerank.	59.9	70.8	64.9	45.9	53.1	49.2

- Comments
  - Good coverage of the generic linguistic resources (90% Oracle)
  - Strong impact of ASR errors
  - Need of an application specific adaptation: baseline very poor



# Evaluation

- **Evaluation 2 on the constituents**

- A constituent hypothesis is correct iff
  - The words, chunk label and syntactic role are correct

P=precision  
R=recall  
F=F-measure

Corpus	ref. trans.			ASR trans.		
	P	R	F	P	R	F
results	88.8	76.6	82.4	78.2	63.5	70.1
Baseline	91.1	90.2	90.6	80.4	76.0	78.2

- Comments
  - **Good results of the reranking strategy with a limited human effort**



# Summary on this application

- **Combining rich generic linguistic ressources and « light » manual adaptation**
- **Parsing process integrated into the SLU pipeline**
  - Word lattice  $\Rightarrow$  POS lattice  $\Rightarrow$  chunk lattice  $\Rightarrow$  dependency lattice
- **Reranking process can integrate various information sources**
  - Lexical, syntactic, semantic, contextual
- **Dependency hypotheses can be used as features for the automatic semantic frame annotation process**



### 3. Applications

- **Extracting dependencies on a set of selected target verbs**
  - SLU for spoken dialog systems (LUNA project)
- **Detecting and characterizing interrogative sentences for speaker role detection**
  - Speech mining on Broadcast Conversations (EPAC project)



# Detecting and characterizing interrogative sentences for speaker role detection

- **Context of this study**

- Speech Data Mining (EPAC project)
- Information about the type of audio stream and the role and identity of speakers is crucial to allow complex queries such as:
  - ``seek debates on X ,```` find all the interviews of Y", etc.
- Initial intuition
  - the form of the questions uttered is a signature of the role of the speakers in the conversation (anchor, guest, expert, etc.)
- Goal
  - Improving speaker role classification accuracy by adding syntactic features related to interrogative utterances.



# Detecting and characterizing interrogative sentences for speaker role detection

- **Validation of the initial intuition**

- French radio broadcast show
  - “le téléphone sonne” (France Inter)
  - Debates with a main animator, experts, callers and journalists
- Characterizing questions
  - A speech corpus transcribed and annotated with question labels
    - 20 hours / 32 shows
    - 17 different kinds of questions
    - About 1500 questions
- Characterizing speaker role
  - Animator (“anchor”)
  - Experts
  - Callers
  - Journalists repeating callers/internet questions

Type de questions	Nombre d'occurrences	Fréquence (%)
directe partielle intonation	404	25,98
directe totale est-ce que	214	13,76
directe partielle adverbe	198	12,73
directe totale intonation	178	11,45
directe totale inversion	154	9,9
directe partielle complexe	117	7,52
directe partielle pronom	88	5,66
directe partielle déterminant	76	4,89
indirecte totale si	46	2,96
indirecte partielle adverbe	26	1,67
indirecte partielle complexe	17	1,09
indirecte partielle déterminant	17	1,09
directe alternative inversion	10	0,64
directe alternative intonation	5	0,32
indirecte partielle groupe nominal	3	0,19
indirecte partielle pronom	1	0,06
directe alternative est-ce que	1	0,06



# Detecting and characterizing interrogative sentences for speaker role detection

- Validation of the initial intuition: statistics

Rôle	Nombre de questions	Fréquence (%)
Présentateur	791	50,87
Expert	323	20,77
Auditeur	307	19,74
Rapporteur	134	8,62
<b>TOTAL</b>	<b>1555</b>	<b>100</b>

Type question	Présentateurs (%)	Experts (%)	Auditeurs (%)	Rapporteurs (%)
directe partielle intonation (404)	98,51	1,24	0,25	0
directe totale est-ce que (214)	28,5	34,58	33,64	3,27
directe partielle adverbe (198)	18,69	43,94	21,72	15,66
directe totale intonation (178)	80,9	16,3	2,81	0
directe totale inversion (154)	33,12	7,14	21,43	38,31
directe partielle complexe (117)	21,37	48,72	19,66	10,26
directe partielle pronom (88)	34,1	35,23	19,32	11,36
directe partielle déterminant (76)	44,74	19,74	19,74	15,79
indirecte totale si (46)	4,35	2,18	93,48	0
indirecte partielle adverbe (26)	0	3,85	88,46	7,69
indirecte partielle complexe (17)	5,88	11,76	82,35	0
indirecte partielle déterminant (17)	0	11,76	88,24	0
directe global (1445)	54,46	21,94	14,46	9,13
indirecte global (110)	3,64	5,45	89,1	1,82

# Detecting and characterizing interrogative sentences for speaker role detection

- **Preliminary experiments**

- Adding “question features” to the speaker role classifier in addition to words + duration features
  - Question : yes/no
  - Type of question

Speaker	#	N-gram+duration	+ question	+ question type
caller	220	88.2	89.6	90.4
expert	118	83.9	83.0	84.4
animator	35	66.7	67.4	68.8
journalist	27	45.7	45.7	74.4
Total error	400	17.8%	17.3%	15.0%

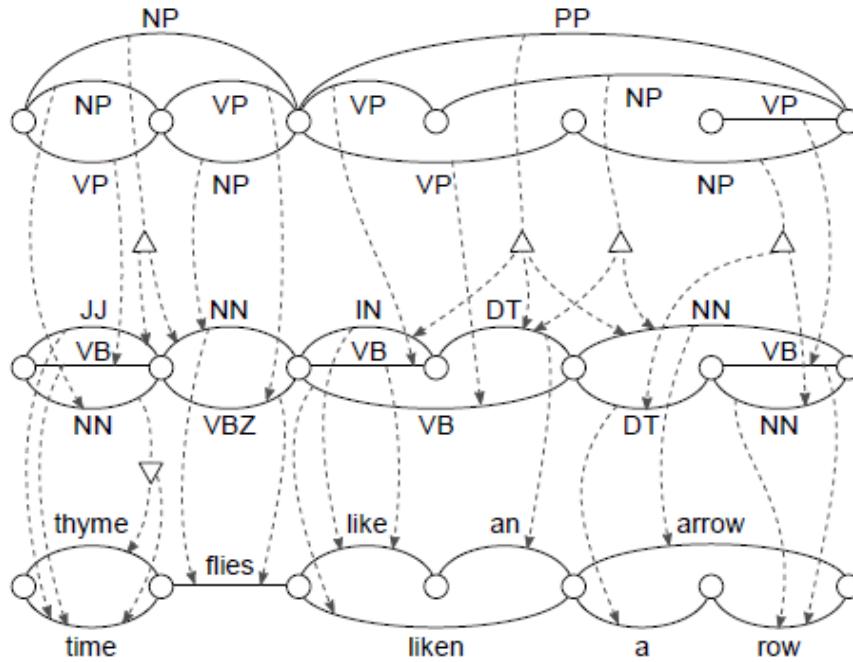
# Overview

- 1. Context of this study : Spoken Language Understanding**
- 2. Main issues in parsing speech**
- 3. Applications**
- 1. Tools & demo**



# Tools and demo

- **MACAON: A Word-Lattice Oriented NLP Tool Suite**
  - <http://macaon.lif.univ-mrs.fr/>
    - Alexis Nasr, Jean-Francois Rey, Frederic Bechet



# Tools and demo

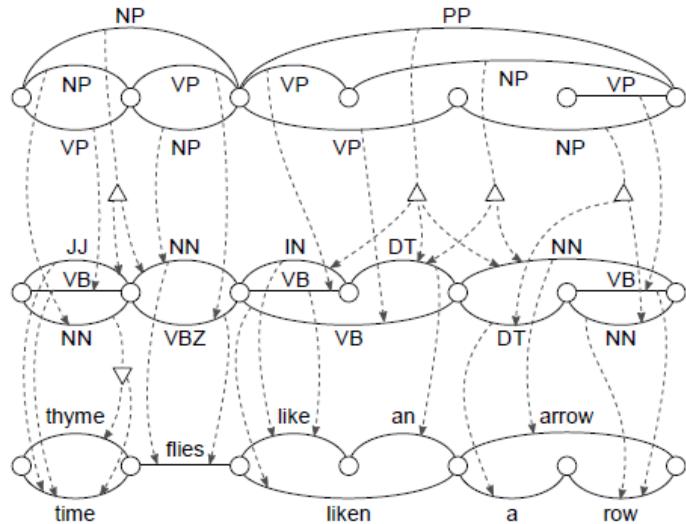
- **Description of the MACAON format**
  - Segments
    - A segment refers to a segment of the text or speech signal that is to be processed, as a sentence, a clause, a syntactic constituent, a lexical unit, a named entity . . .
    - A segment can be equipped with attributes that describe some of its aspects
    - A segment is made of one or more smaller segments
  - Sequence of segments
    - A sequence of segments covering a whole sentence for written text, or a spoken utterance for oral data, is called a segmentation
    - Such a sequence can be weighted.
  - Annotations
    - An annotation level groups together segments of a same type, as well as segmentations defined on these segments
    - Two relations are defined on segments
      - the precedence relation that organises linearly segments of a given level in order to constitute segmentations
      - the dominance relation that describes how a segment is decomposed in smaller segments either of the same level or of a lower level.



# Tools and demo

- Precedence relation

```
<fsm n="9">
  <start n="0"/>
  <accept n="6"/>
  <ltrans>
    <trans o="0" d="1" i="11" w="-7.23"/>
    <trans o="0" d="1" i="12" w="-9.00"/>
    <trans o="1" d="2" i="13" w="-3.78"/>
    <trans o="2" d="3" i="14" w="-7.37"/>
    <trans o="3" d="4" i="15" w="-3.73"/>
    <trans o="2" d="4" i="16" w="-6.67"/>
    <trans o="4" d="5" i="17" w="-4.56"/>
    <trans o="5" d="6" i="18" w="-2.63"/>
    <trans o="4" d="6" i="19" w="-7.63"/>
  </ltrans>
</fsm>
```



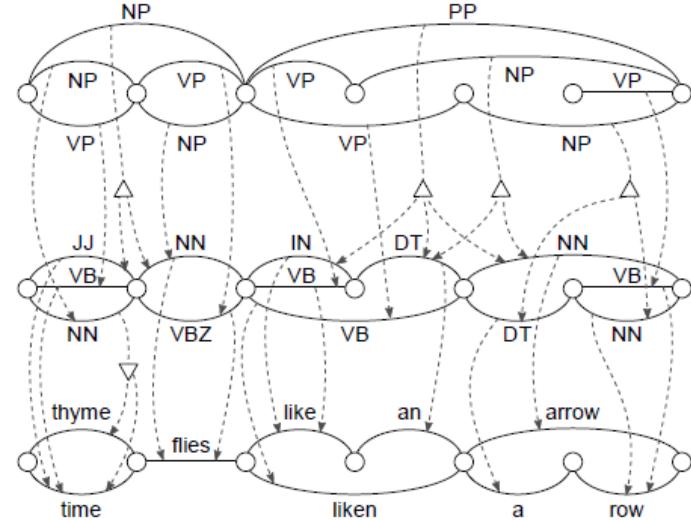
# Tools and demo

- **Dominance relation**

```
<segment type="chunk" stype="PP" id="c1">
  <sequence>
    <elt segref="p1"/>
    <elt segref="p2"/>
    <elt segref="p3"/>
  </sequence>
</segment>
```

- Dominance relation with ambiguity

```
<segment type="cat" style="NN" id="c1">
    <sequence>
        <elt segref="l1" w="-3.37"/>
    </sequence>
    <sequence>
        <elt segref="l2" w="-4.53"/>
    </sequence>
</segment>
```



# Demo MacaViz

