Processing spontaneous speech in deployed Spoken Language Understanding systems: a survey

Frédéric Béchet

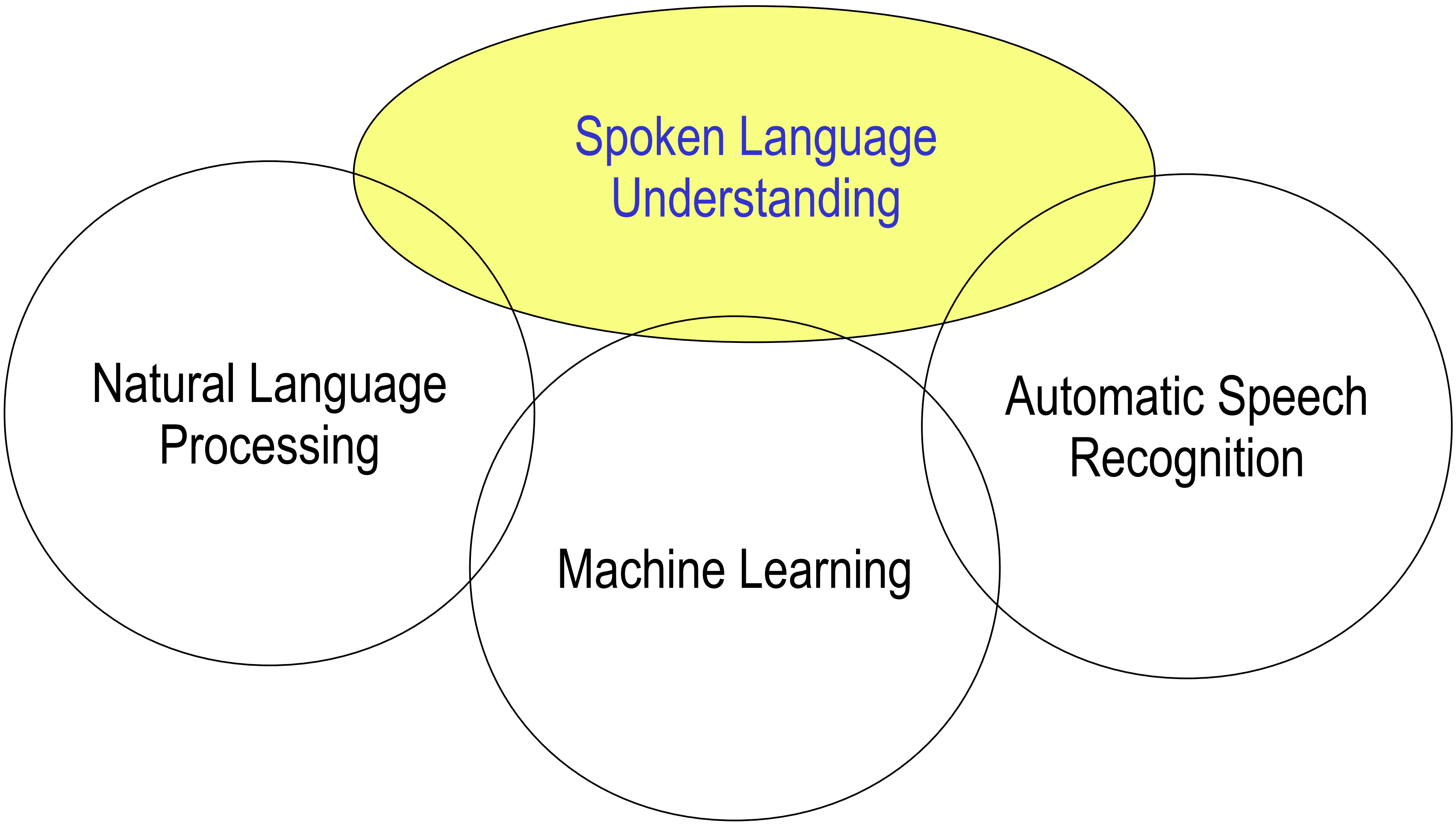
LIA, Université d'Avignon

LABORATOIRE INFORMATIQUE D'AVIGNON

UNIVERSITÉ D'AVIGNON

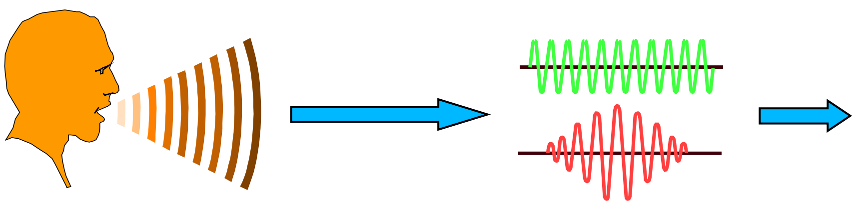
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

Framework



Spoken Language Understanding
Natural Language Processing
Automatic Speech Recognition
Machine Learning

Spoken Language Understanding



Signal Processing

-Automatic Speech Recognition
- Transcription
- « Rich »Transcription

-Information Extraction

-Analysis
- classification
- Syntactic, semantic
- Extracting entities
- Relations between entities
- ….

transcription

understanding

Spoken Language Understanding

- Three applicative frameworks
  - Human-Machine spoken dialogue
    - Call-routing
    - Form-filling
    - Negotiation
  - Speech mining / Voice Search
    - Broadcast News
    - Large speech archives (e.g. Malach, INA, etc.)
    - Call centers
  - *Speech to Speech Translation*

Problem formulation

- Four main issues:
  - How to represent the meaning of a spoken message?
    - *Choice of a Meaning Representation Language*
  - What kind of models and algorithms can we use to perform understanding?
    - *Choice of an interpretation paradigm*
  - Which corpus and wich annotations to develop/train/evaluate an SLU system?
    - *Choice of an applicative framework*
  - Evaluating SLU systems?
    - *Global vs. Modular evaluation*

Problem formulation

- Processing text *vs.* Processing speech messages
  - SLU = Natural Language Processing on ASR transcriptions?
  - Integrating ASR and SLU processes
- Dealing with deployed SLU systems
  - Robustness issues
  - Real corpus = real life issues
    - *« Expecting the inexpected »*

How to represent the meaning of a spoken message?

Choice of a Meaning Representation Language

- Two different point of views
  - 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

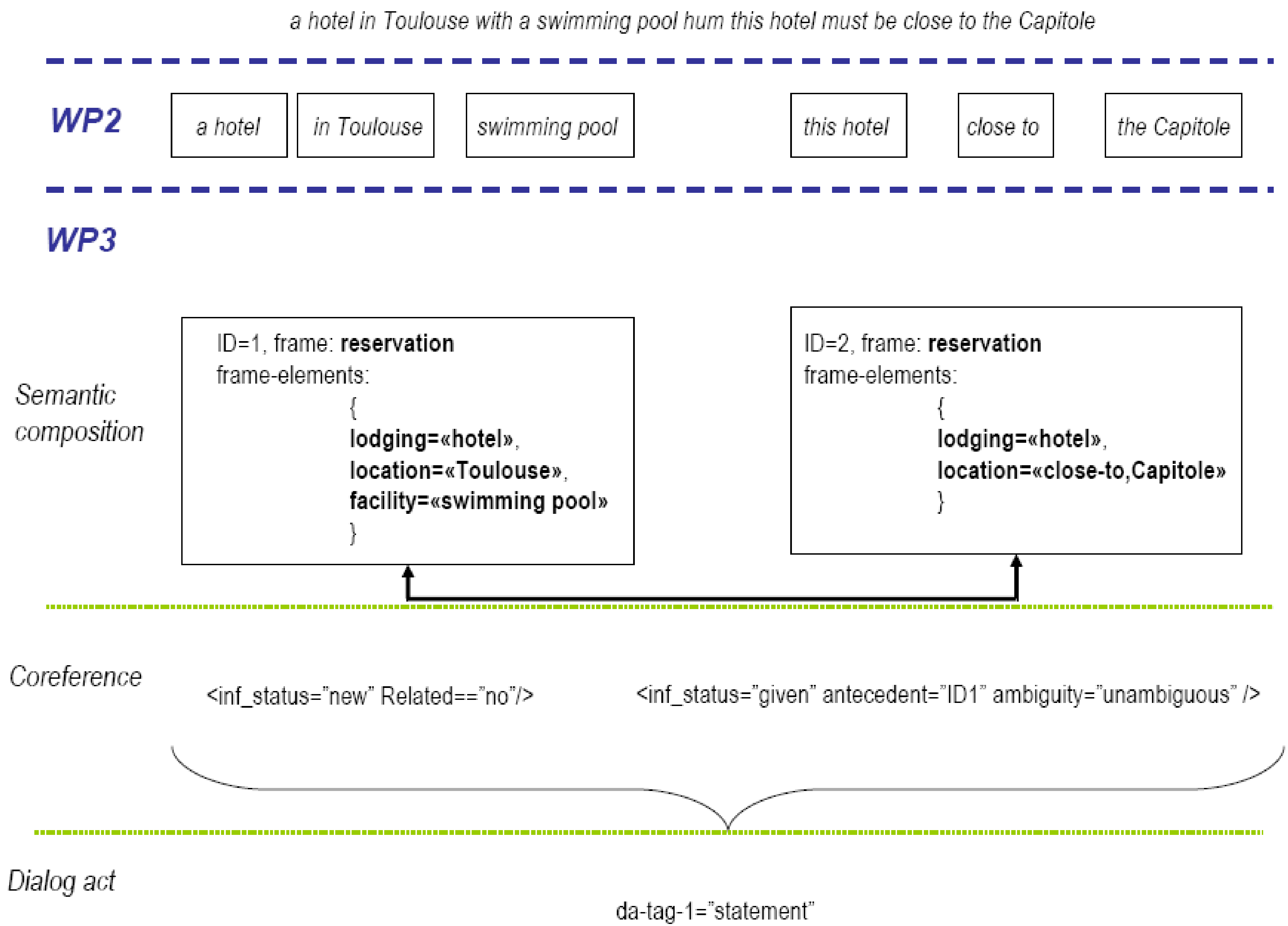
      → ***SYSTEM INTERPRETATION***

Choice of a Meaning Representation Language

- 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
    - The entry in the directory
  - In a speech mining / voice search application
    - Distillation
    - Request: topic, genre, enitites, ...

Choice of a Meaning Representation Language

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



a hotel in Toulouse with a swimming pool hum 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"

What kind of models and algorithms can we use to perform 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 !

Which models and algorithm to perform SLU?

- **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)

Which models and algorithm to perform SLU?

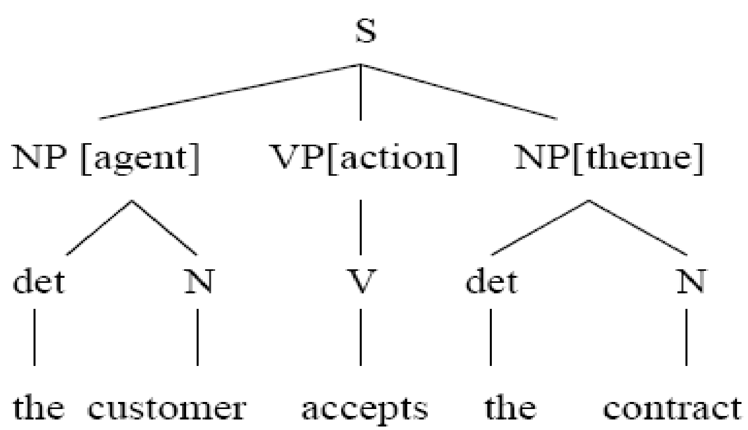
- **Understanding is a translation task**
  - Translation to a formal semantic language
  - Related to tagging approaches (e.g. POS tagging)
  - Main model: Concept decoding
    - Mapping a sequence of words to a sequence of attribute/ value tokens
  - Main approaches
    - Hidden Markov Models
    -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

Which models and algorithm to perform SLU?

- **Understanding is a translation task**
  - System interpretation
    - Concept composition
  - Examples
    - Chronus (Pieracini & Levin)
      - HMM based model on ATIS
    - Comparison generative/discriminative methods
      - Raymond & Riccardi (Interspeech 2007)
      - Hahn et al. (Interspeech 2008)

Which models and algorithm to perform SLU?

- **Understanding is a parsing task**
  - syntactic/semantic parsing
    - Concepts = leaves in a parse tree
    - System intepretation
      - obtained directly from the parse tree



Which models and algorithm to perform SLU?

- **Understanding is a parsing task**
  - 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)

Corpus based / Knowledge based methods

- Generik or application specific?
  - Some generik corpora available
    - Tree Bank, Frame corpora, Named Entity, ...
    - *But limited coverage and mostly for English*
  - Some generik linguistic ressources available
    - WordNet, FrameNet, PropBank, …
    - *But limited coverage and mostly for English*
  - But in both cases
    - Need for application specific corpora/knowledge ressources
    - Need to collect examples in both cases

Corpus based / Knowledge based methods

- Corpus based methods
  - Collecting examples to build a corpus
  - Annotation thanks to an annotation guide
  - No need of handcrafted rules defined by experts but..

    → need a semantic model anyway
    - To *define the annotation model*
    - *To write the annotation guide*
    - *Experts needed to train the annotators*
  - Incremental process when examples are collected
    - Improving the semantic model and the annotation guide

Corpus based / Knowledge based methods

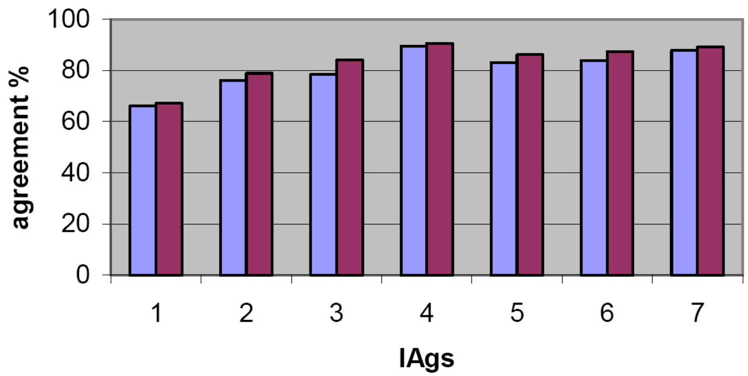
- Knowledge based methods
  - Hand crafted rules and grammars
    - To detect, extract and compose semantic entities
  - Can be seen as a procedural version of the annotation guide
  - Bootstrap with a few examples
  - Incremental process
    - Rules and grammars are improved as new examples are collected
    - e.g. Example-based learning

Corpus based / Knowledge based methods

- Example on the French MEDIA corpus
  - Relation between the training corpus size and the concept error rate

Corpus size	25	50	100	200	400	600	720
CER	46.0	43.7	34.6	31.9	26.3	24.2	24.3
correct	61.9%	63.3%	70.9%	74.7%	79.6%	80.8%	81.3%

  - Inter-annotator agreement according to the amount of dialog annotated



Which corpus and wich annotations to develop/train/evaluate an SLU system?

Corpus for SLU

- « artificial corpus »
  - Collected through evaluation program (Ex: ATIS, MEDIA)
  - Manual annotations
  - Limited size
  - Application domain
    - Spoken dialogue systems, question answering, speech doc. retrieval
- « real life corpus »
  - Spoken archives (e.g. INA, log files from call centers)
  - Collected from real users of a speech-service
    - Ex: AT&T How May I Help You?, France Telecom Voice Services
  - Annotations = automatic/manual/none
  - Unlimited size
  - Application domain
    - Call-centers, Audio messages, Deployed SDS

Corpus for SLU

- Main differences
  - Artificial corpus
    - controlled conditions
    - cooperative speakers
    - => little “out-of-domain” data
  - Real life corpus = real life issues !!
    - Very spontaneous speech
    - Very large variability
      - Speech: accents, language
      - Usage: different classes of users (new and regulars)
    - Unpredictable behaviors
      - Comments, incoherence

French BN corpus with Named Entities annotations: ESTER

<ENAMEX TYPE= "PERSON"> **Patricia Martin** <\ENAMEX>, que voici, que voilà!
oh, bonjour <ENAMEX TYPE= "PERSON"> **Nicolas Stoufflet** <\ENAMEX>.
<ENAMEX TYPE="ORG"> **France-Inter** <\ENAMEX> , <TIMEX TYPE=TIME> **7 heures** </TIMEX>.
le journal, <ENAMEX TYPE= "PERSON"> **Simon Tivolle** <\ENAMEX>. bonjour!
<TIMEX TYPE=DATE> **lundi 7 décembre** <\TIMEX TYPE=DATE>. deux incendies <TIMEX TYPE=TIME> **cette nuit** </TIMEX> en région parisienne, dans une maison de retraite de <ENAMEX TYPE=LOCATION> **Livry-Gargan en Seine-Saint-Denis** <\ENAMEX>, 7 personnes ont péri dans les flammes.
et puis dans <ENAMEX TYPE=LOCATION> **le neuvième arrondissement de Paris** <\ENAMEX>,
le feu a pris dans un immeuble d'habitation et 3 personnes ont été tuées.
le sculpteur <ENAMEX TYPE= "PERSON"> **César** <\ENAMEX> est décédé
<TIMEX TYPE=DATE> **hier** </TIMEX> à l'âge de <COMPTE TYPE=TEMPS> **77 ans** </COMPTE>,
il était devenu célèbre grâce à ses compressions et à ses oeuvres réalisées avec de la ferraille.
<ENAMEX TYPE="ORG"> **France-Inter** <\ENAMEX> ouvre son antenne à <ENAMEX TYPE="ORG"> **la chaîne de l'espoir** <\ENAMEX> pour aider les enfants malades, qui n'ont pas les moyens de se soigner.
<ENAMEX TYPE="ORG"> **la chaîne de l'espoir** <\ENAMEX>, association de médecins et de bénévoles, a déjà sauvé 5000 enfants.
le quarante-quatrième congrès de la <ENAMEX TYPE="ORG"> **CFDT** <\ENAMEX> s'ouvre <TIMEX TYPE=DATE> **aujourd'hui** </TIMEX> à <ENAMEX TYPE=LOCATION> **Lille** <\ENAMEX>, un congrès de la décrispation.
et puis ouverture aussi aujourd'hui des discussions à la <ENAMEX TYPE="ORG"> **SNCF** <\ENAMEX> entre les syndicats et la direction.
10 mois après l'assassinat du préfet <ENAMEX TYPE= "PERSON"> **Érignac** <\ENAMEX> en <ENAMEX TYPE=LOCATION> **Corse** <\ENAMEX TYPE=LOCATION> , nous ferons le point sur l'enquête.

Spoken survey corpus with opinion annotations: SCOrange (France Telecom)

oui c'est monsieur NOMS PRENOMS j'avais appelé

donc le service client ouais

<seg label=accueil,pos> j'ai été très bien accueilli </seg>

<seg label=efficacité,pos> bons renseignements </seg>

sauf que

<seg label=efficacité,neg> ça ne fonctionne toujours pas </seg>

donc je sais pas si j'ai fait une mauvaise manipulation ou y a un problème enfin voilà sinon

<seg label=efficacité,pos label=accueil,pos> l'accueil était et les conseils très judicieux </seg>

même si

<seg label=efficacité,neg>} le résultat n'est pas n'est pas là </seg>

merci au revoir

Spoken Dialogue corpus (WOZ): MEDIA

n	W^{c_n}	c_n	*mode*	*spécifieur*	*valeur*
1	*je vais réserver*	command-tache	+		reservation
2	*dans cet hôtel hôtel Richard Lenoir*	nom-hotel	+		richard lenoir
3	*six*	nombre-chambre	+	reservation	6
4	*chambres individuelles*	chambre-type	+		simple
5	*pour le trente et un mai*	temps-date	+	reservation	31/05
6	*deux jours et deux nuits*	sejour-nbNuit	+	reservation	2

n	W^{c_n}	c_n	référence	*mode*	*spécifieur*	*valeur*
1	ils	lienRef	guillermo champ-mars pullman	+	coRef	pluriel
2	proches d'	loc-distanceRelative		?	hotel	proche
3	un parc	loc-lieuRelatif		?	general-hotel	parcJardin

Concept + Reference + Semantic Frame annotations

Corpus 3000 France Télécom

- « standard request »
  - désactiver mon transfert
  - transfert d'appel
  - payer mon facture
  - messagerie vocale
- « difficult requests »
  - je vous appelle à propos de d' une facture qui n' a pas été réglée et qui a été réglée alors je voudrais avoir quelqu' un pour m' expliquer avec
  - oui je voudrais un renseignement qui est très important s'il-vous-plaît
  - pourquoi j' ai plus de tonalité
  - je voudrais savoir si j' ai reçu la facture que je dois payer ou si je n' ai pas reçu de facture ou si je vous dois de l' que je rentre de l' et je vais y retourner c' est tout ce que je veux savoir mon numéro c' est le je sais plus du tout où j' en suis je suis très fatiguée et je retrouve pas la facture si je dois vous payer ou si je vous ai payée je sais rien du tout alors savoir si vous m' avez envoyé une facture le montant et quand je dois vous la payer
- « out of domain » requests
  - je comprends pas pourquoi on me dit j' entends la sonnerie et puis quand je vais pour décrocher il y a il y a plus rien je comprends pas ça à plusieurs fois à plusieurs ça fait ça
  - allo je voudrais connaître le numéro d' un d' un pressing qui a changé de propriétaire et et de numéro de téléphone évidemment je ne sais pas si je suis bien reliée

Evaluating SLU systems?

Why evaluation ?

- Main goal = comparing different approaches
- For example :
  - SLU as a parsing process
  - SLU as a tagging process
  - SLU as a classification process
- Therefore ..
  - ***The Semantic Model used for evaluation must be as independent as possible from a given approach***

SLU vs. ASR evaluation

- Evaluation ?
  - Comparing a reference to a prediction
- Main issue
  - On which «objects» this comparison should be performed ?
- Automatic Speech Recognition
  - « standard » objects that can be observed (more or less)
    - phoneme, syllable, character, words, etc.
- Spoken Language Understanding
  - No standard !! Nothing can be directly observed

→ ***RELY ON A SEMANTIC MODEL !!***

What can we evaluate ?

- Choice of an evaluation level
  - Global evaluation on the system performance
    - **Positive**
      - **That's what matter to the users !!**
      - **Evaluate the whole process, the interconnection between the different levels**
    - **Negative**
      - **Difficult to put in place and reproduce for a dialogue application**
      - **« black box » evaluation**
  - Multi-level evaluation
    - **Positive**
      - **Modularity**
      - **Can be done in batch mode**
    - **Negative**
      - **What is the real impact on the global performance of an improvement obtained on a given level**
      - **Risk of performing « artificial » evaluation (e.g. word disambiguisation)**

Evaluation measures

- Sequence evaluation
  - Let:
    - N = # of reference symbols, N' = # of hypothesized symbols
    - i = # insertion ; s = # substitution ; d = # deletion
  - We have:
    - Correct rate : :

$$T_{cor} = \frac{(N' - i - s) \times 100}{N}$$

    - 3 error rates :

$$T_{ins} = \frac{i \times 100}{N} \quad T_{sub} = \frac{s \times 100}{N} \quad T_{sup} = \frac{d \times 100}{N}$$

    - Token Error Rate – TER:

$$T_{TER} = T_{ins} + T_{sub} T_{sup}$$

    - Token accuracy, acc: :

$$T_{acc} = 100 - T_{TER}$$

Evaluation measures

- Global evaluation

$$Précision = \frac{nombre\ de\ décisions\ correctes \times 100}{nombres\ de\ décisions\ produites}$$

$$Rappel = \frac{nombre\ de\ décisions\ correctes \times 100}{nombre\ de\ choix\ dans\ la\ référence}$$

$$F = \frac{2 * Rappel * Précision}{Rappel + Précision}$$

- Oracle rates
  - For a set of hypotheses $L = \{H_1, \ldots, H_N\}$ and an error rate TE

$$Oracle_{TE}(L) = \min_{1<=i<=N} TE(H_i)$$

Multi-level evaluation : an example

a hotel in Toulouse with a swimming pool hum 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"

Multi-level evaluation : Genericity of the annotations ?

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

WP2

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

WP3

OK

Rather independent from any semantic model

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

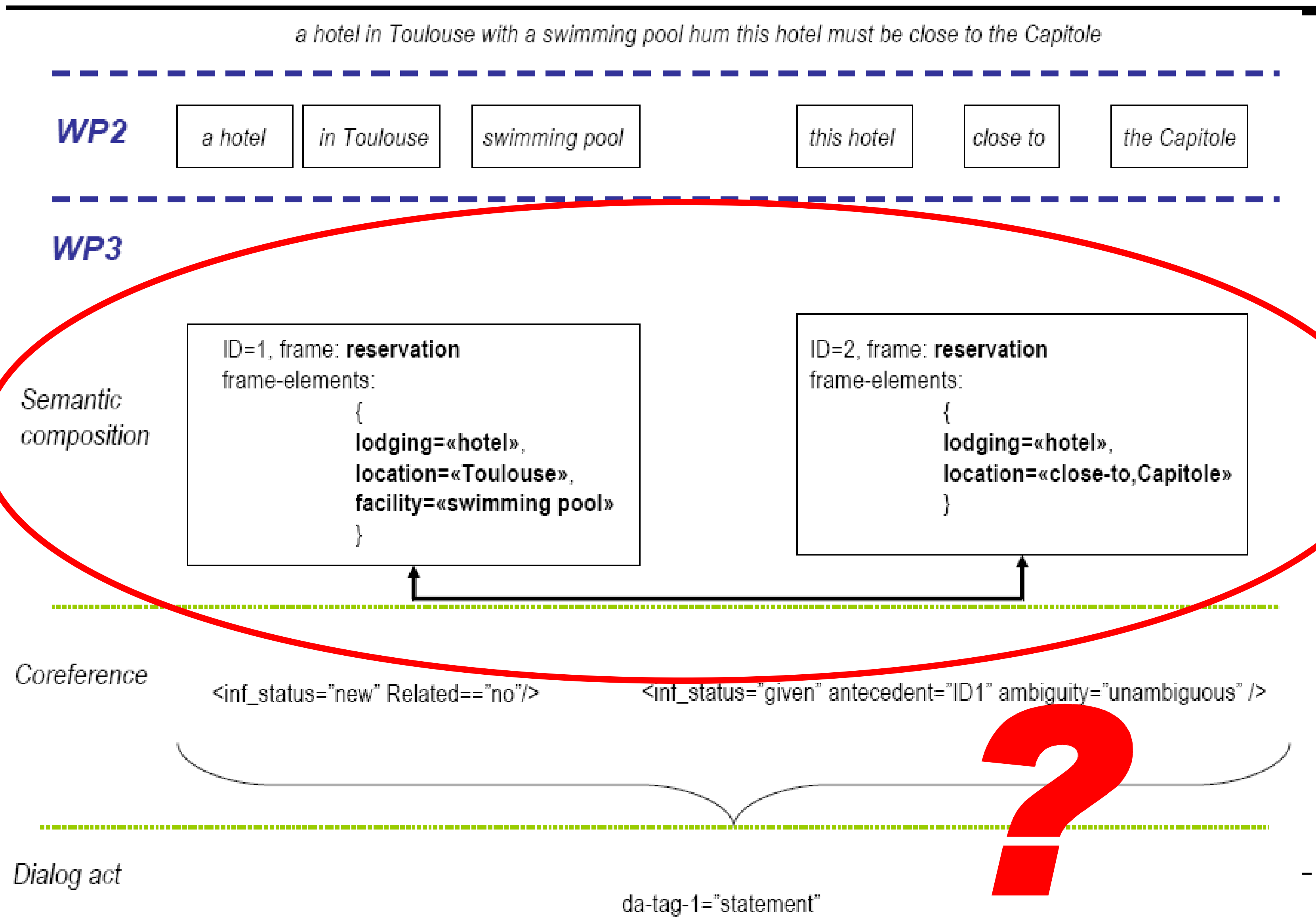
<inf_status="new" Related=="no"/>

<inf_status="given" antecedent="ID1" ambiguity="unambiguous" />

Dialog act

da-tag-1="statement"

Multi-level evaluation : Genericity of the annotations ?



Semantic annotation relying heavily on a specific semantic model

The Technolangue MEDIA experience

- Evaluation on « traces » of understanding
- Goal
  - Being as independent as possible of a semantic model
  - Going further than the attribute/value level:
    - Mode (+,-,?,~)
    - Semantic specifier (« traces » of hierarchical structures)
    - Co-reference
- Pros
  - Evaluation of 5 very different SLU systems
  - Multi-purpose corpus
- Cons
  - Biased toward tagging/classification approaches
  - Artificial evaluation ?

The Technolangue MEDIA experience

- Example of attribute/value+mode+specifier+reference evaluation
  - Prompt
    - « je vous propose trois hôtels hôtel Guillermo hôtel du champ de mars hôtel Pullman »
  - Message
    - « et j' aimerais savoir s' ils sont proches d' un parc »

n	W^{c_n}	c_n	référence	*mode*	*spécifieur*	*valeur*
1	ils	lienRef	guillermo champ-mars pullman	+	coRef	pluriel
2	proches d'	loc-distanceRelative		?	hotel	proche
3	un parc	loc-lieuRelatif		?	general-hotel	parcJardin

Processing text vs. Processing speech messages

SLU vs. Text processing

- SLU = ASR + text processing ?
  - Text documents vs. Speech utterances
  - Automatic transcripts
    - ASR issues
      - Uncertainty, misrecognition, unknown words
    - Partial information
      - All prosodic information missing
    - No structure = stream of words
  - Text
    - “finite” object
    - Text + structure + “graphical” information

SLU vs. Text processing

- Main issues
  - Text
    - “open world”
    - Capacity of handling new phenomenon
      - Words, compounds, entities
    - Need: Generalization capabilities of the models
  - ASR transcript
    - “closed world”
    - ASR lexicon+Language Model define this “world”
    - No unknown words (just misrecognitions !!)

      → no generalization needed
    - Need: robust detection of the expected information
      - Confidence estimation $P(W|A)$

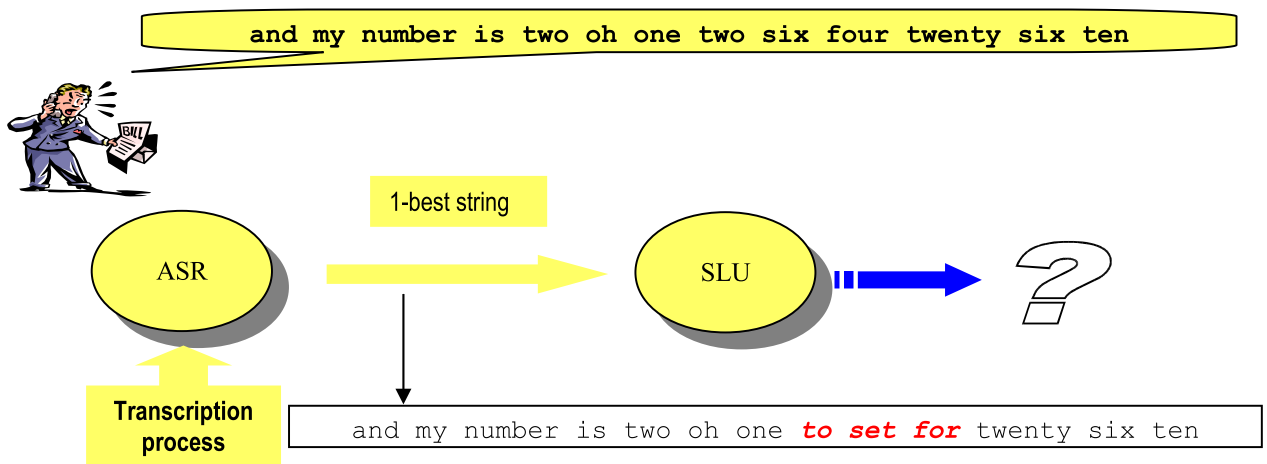
Example

- Transcription ?

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 annoter 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 sois 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

Integrating ASR and SLU processes

- « sequential approach »
  - ASR => SLU => Manager
    - ASR module produces a text document
    - SLU module processes this text document
    - Manager = exploits SLU output



Integrating ASR and SLU processes

- « integrated approach »
  - ASR ⇔ SLU ⇔ Manager
  - All 3 processes should collaborate
    - Definition of a context (local + global)
    - ASR+SLU+Manager: tuning according to the context
    - ASR output = multiple hypothesis (word lattice)
    - SLU = from a word lattice to an « interpretation lattice »
    - Manager = decision strategy on multiple hypothesis output

Example (global context)

I wanna know why I was charged on **September sixth 11 dollars 63 cents** for calling **8 5 6 2 1 6 5 5 2 1** **Clementon New Jersey** for **1 minute**

PHONE BILL SEPTEMBER 2001

DATE	PHONE#	DURATION	PLACE	AMOUNT
09062001	8562165521	01:00	Clementon, NJ	11.63
….	….	….	….	….
….	….	….	….	….

Exemple: AT&T How May I Help You? tm

Example (local context)

```
system>  in Marseille I propose the Hotel la Fanette
         and the Hotel du Port
user>    where is the Hotel la Fanette?
ASR>     where is the Hotel Lafayette
```

Integrating ASR and SLU processes

- An example of integrated strategy:
  - The LUNA project

LUNA

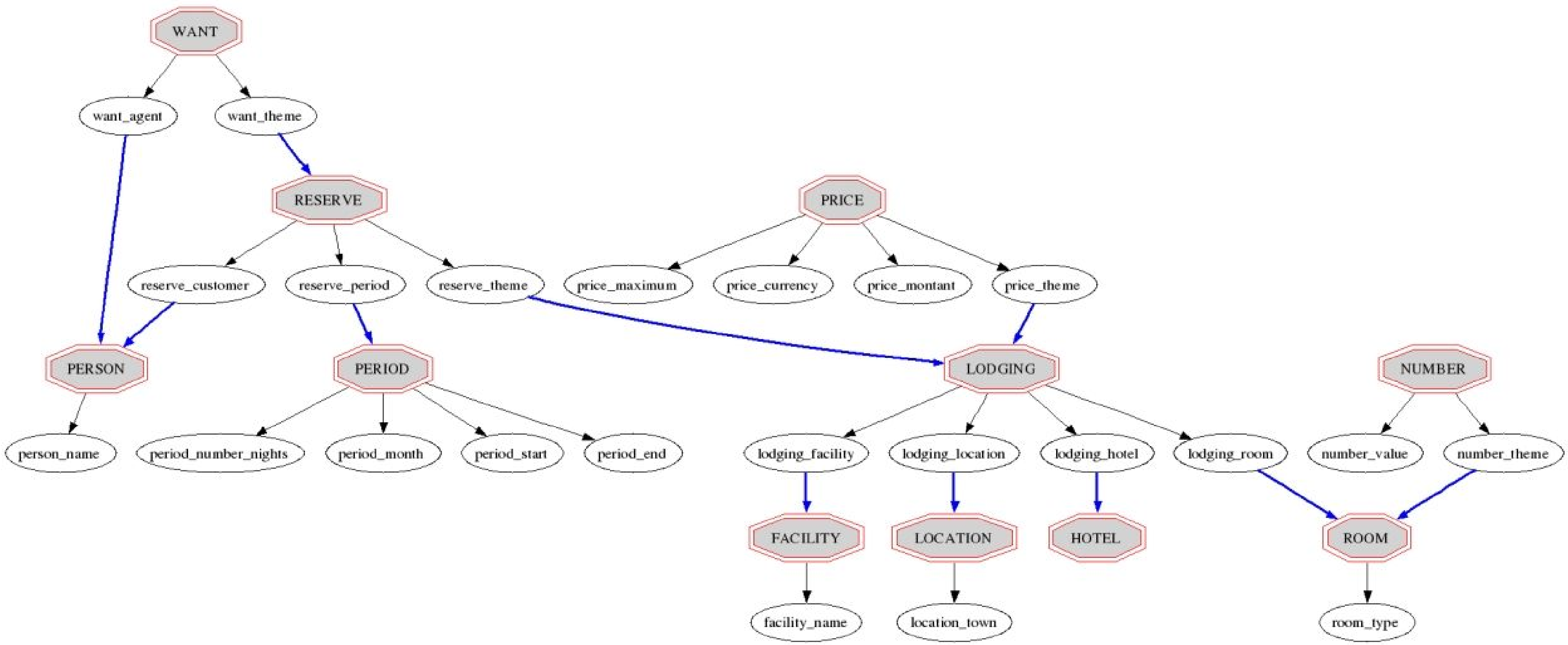
- FP6 European project: LUNA
  - spoken Language UNderstanding in multilinguAl communication systems
  - September 2006
- Goal
  - Build robust multilingual SLU strategies
  - Five main objectives
    - Language Modelling for Speech Understanding;
    - Semantic Modelling for Speech Understanding;
    - Automatic Learning (including Active and On-Line Learning);
    - Robustness issues for SLU;
    - Multilingual portability of SLU components.
- Partners
  - Loquendo, RWTH Aachen, University of Trento, University of Avignon, France Telecom R&D, CSI-Piemonte, Polish-Japanese Institute of Information Technology, Institute of Computer Science - Polish Academy of Sciences

SLU models in LUNA

- Multi level semantic representation
  – Concept decoding: from words to concepts
  – Semantic composition: from concepts to interpretations
  – Coreference / Anaphoric relation resolution
  – Speech acts
- Corpus annotation on these levels
  – Concepts
    - word+POS tag+chunk+ Ontology in OWL
  – Interpretations
    - Framenet-like approach
  – Reference resolution
    - ARRAU framework
  – Speech acts
    - Subset of DAMSL
  – DEMO (MEDIA)

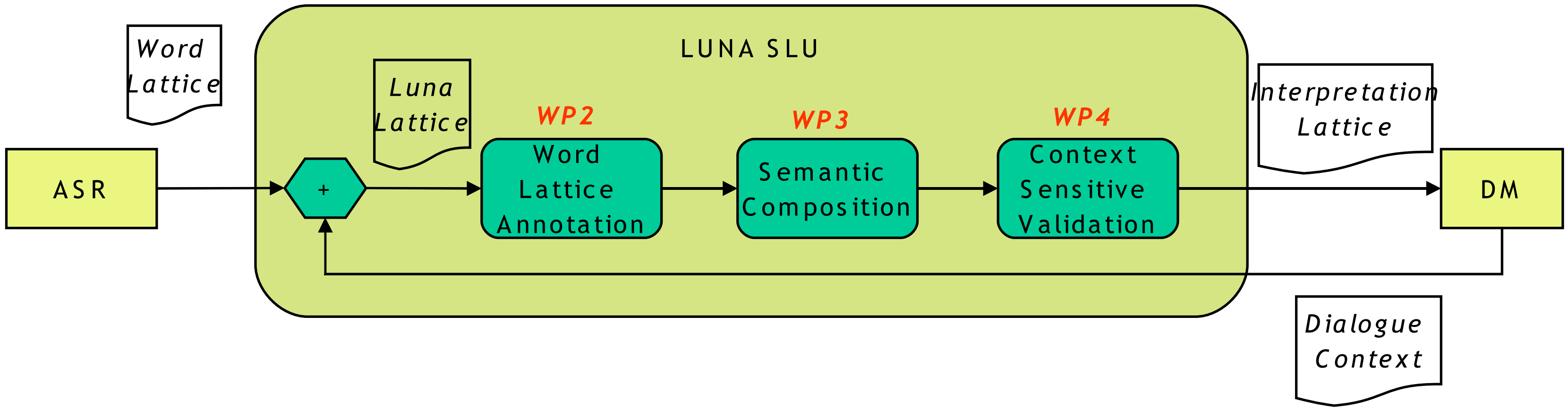
Example of LUNA corpus

- On the MEDIA corpus

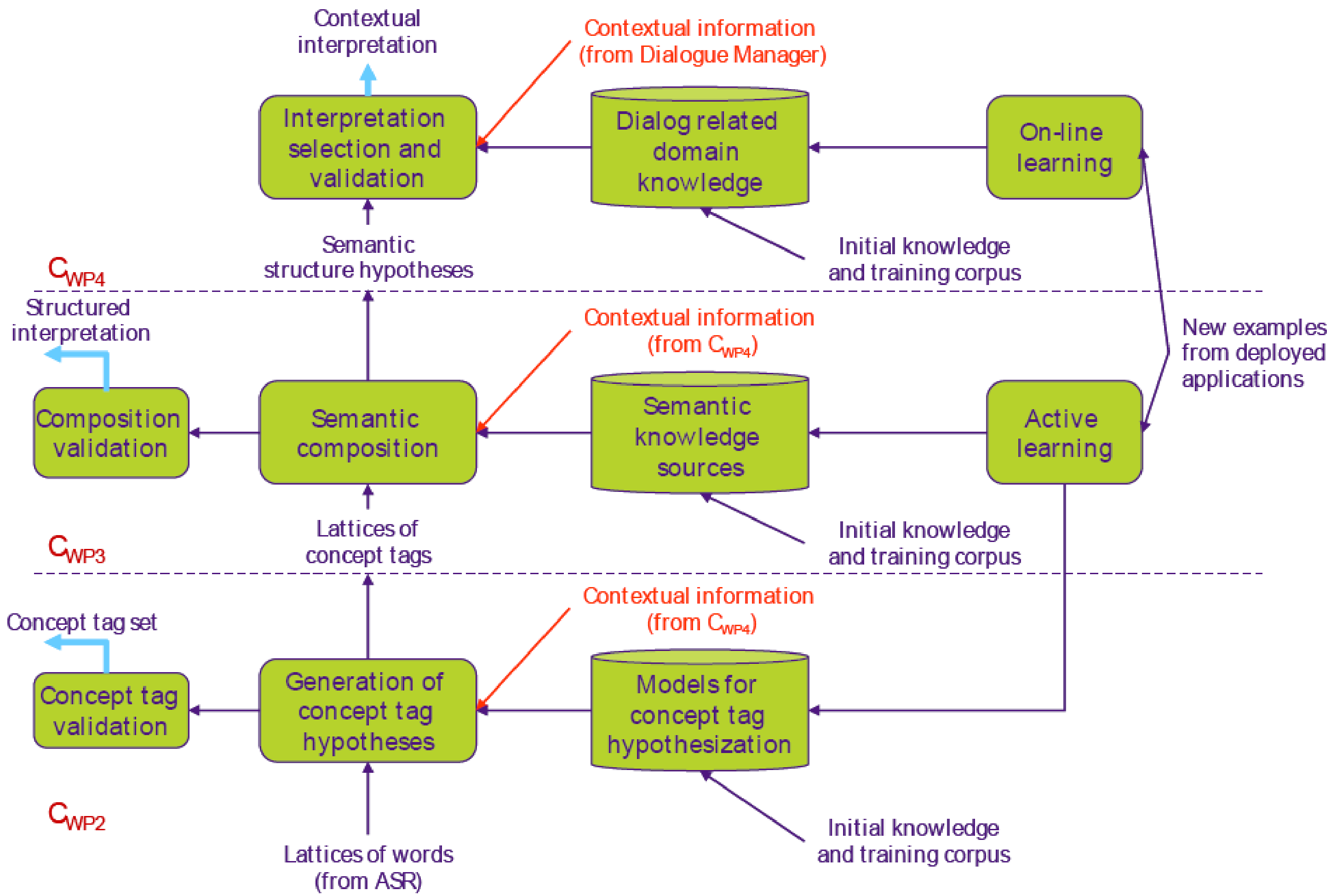


LUNA: an integrated approach

- Process
  - From a word lattice to an entity lattice
  - From an entity lattice to an interpretation lattice
  - With references, with speech acts
  - Each level using contextual information
    - A priori information on the application context
    - Dynamic information provided bt the dialog manager
- Corpus based + knowledge based methods

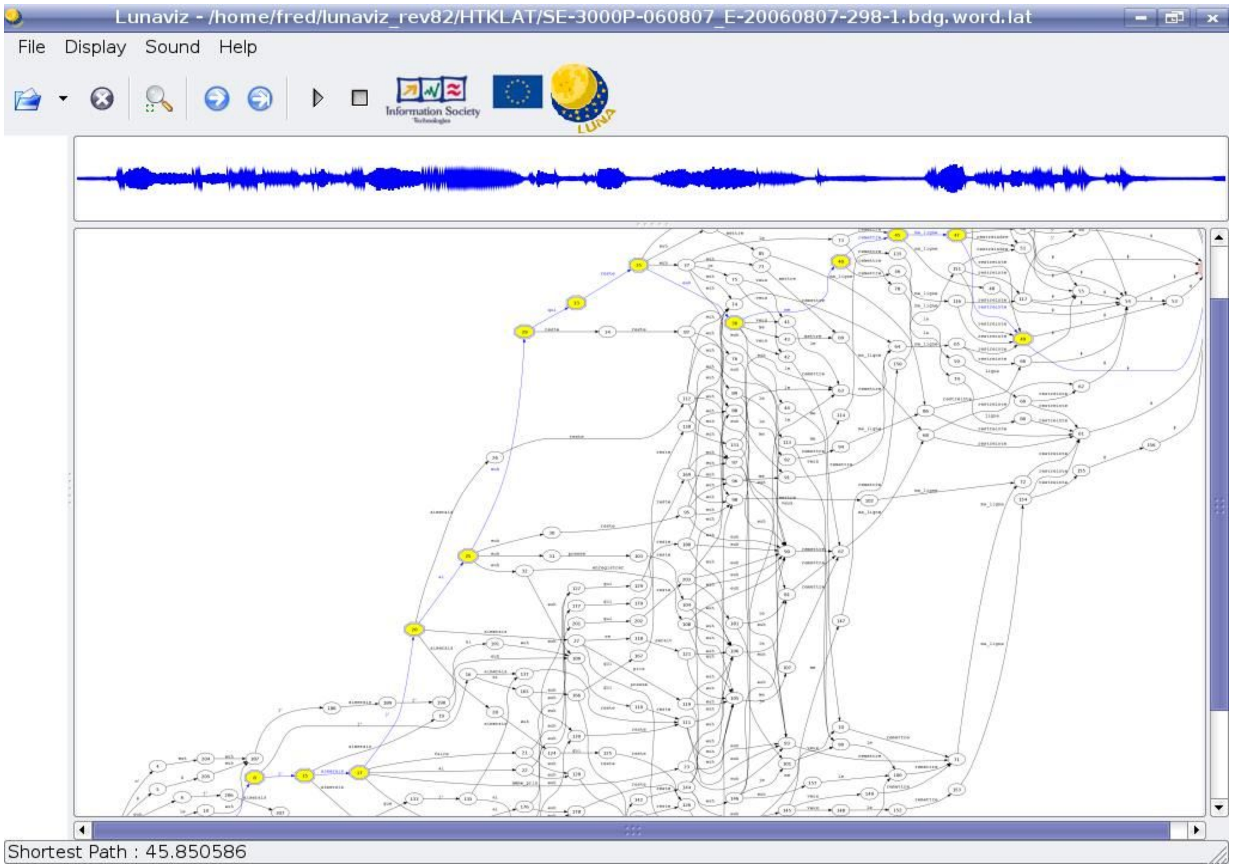


LUNA architecture



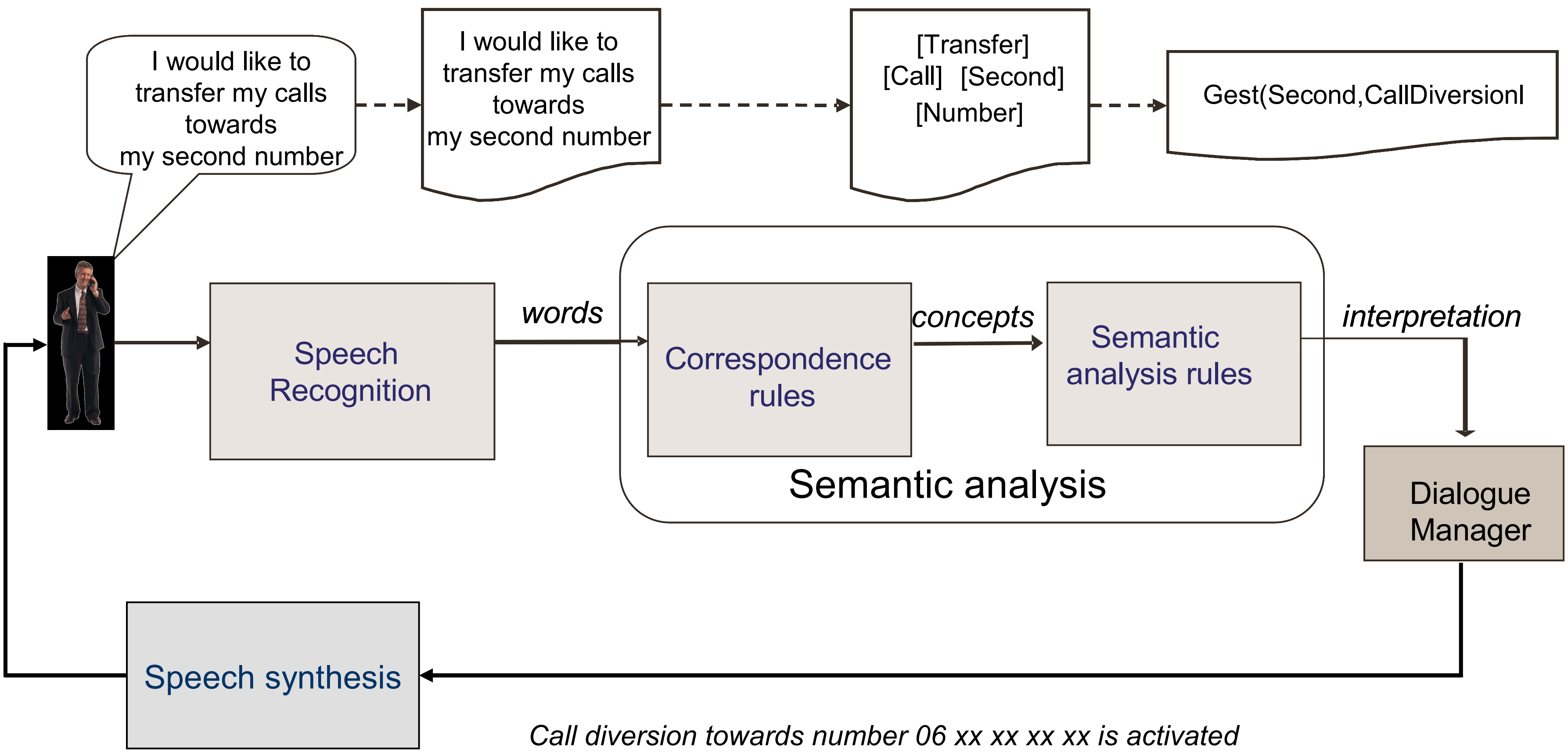
DEMO

- FT3000 (Tool: LunaviZ)



Dealing with deployed SLU systems

An example of deployed application: France Telecom FT3000



FT3000 dialogue corpus

- Lexicon: 2,2k words
- Training corpus 42k utterances
- Test corpus 4.5k utterances
- Translation of words into concepts
  - 400 concepts
  - 1200 hand written regular grammars
- Translation of concepts into interpretation
  - 2030 different possible interpretations
  - A set of 3200 hand written inference rules
  - Interpretations expressed by an attribute/value pair

LABORATOIRE INFORMATIQUE D'AVIGNON

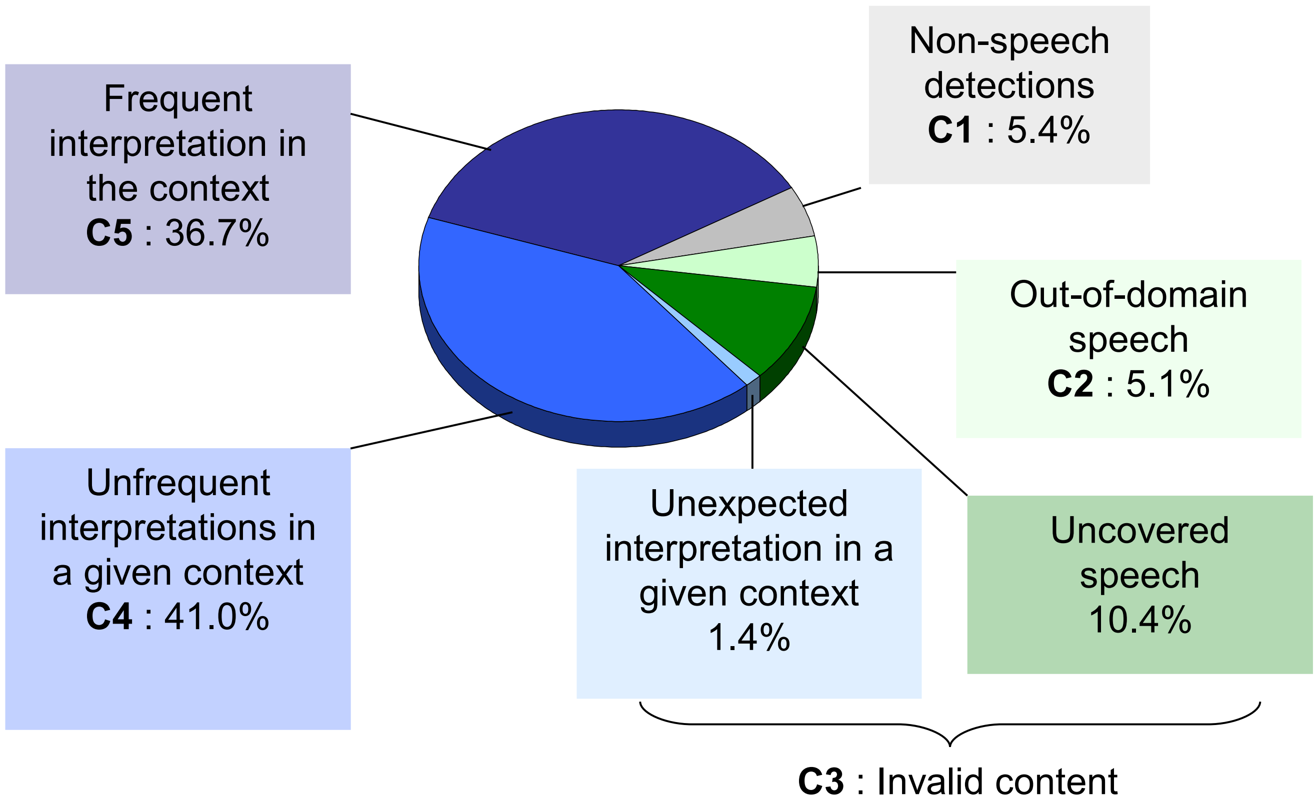
UNIVERSITÉ D'AVIGNON
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE

Utterance characterization

	#	%	example
non-speech C1	246	5.4%	biiiiip
out of domain C2	231	5.1%	bon qu'est-ce que je dois dire là euh *well what am I supposed to say now*
invalid content C3	536	11.8%	Not covered by interpretation rules on m'appelle tout le temps et y a personne je sais pas qui c'est *I receive calls all the time and there's no one I don't know who it is* Not expected in a given dialogue phase oui / *yes* as an answer to an open question
non-frequent C4	1870	41.0%	j'appelle pour avoir un renseignement euh pour euh avoir le service du réveil *I'm calling in order to have information about urh for having the wake-up service*
frequent C5	1671	36.7%	payer ma facture *pay my bill*

Utterances variety in real corpora

- Real conditions → variety of users / utterances



Different users, different behaviors

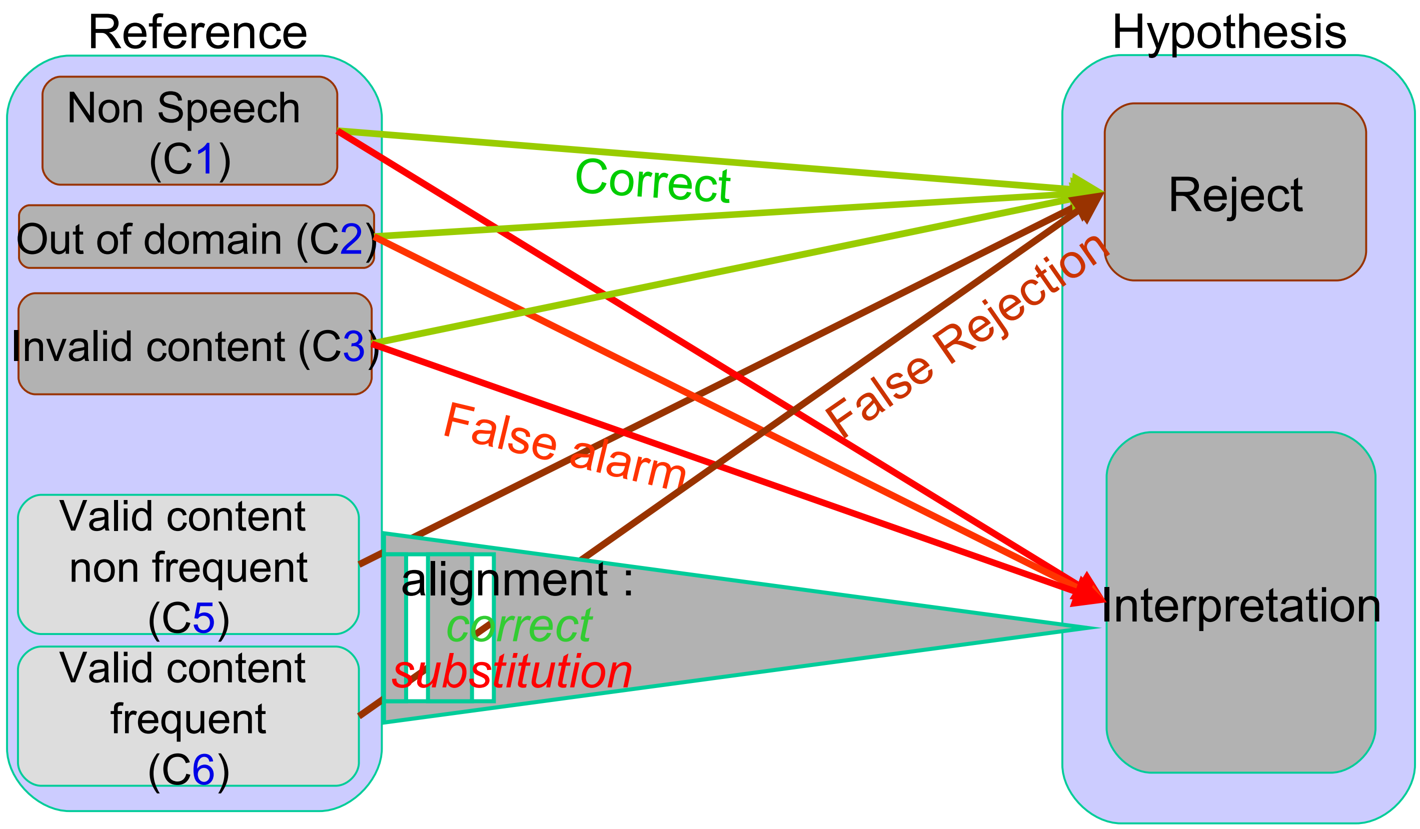
- User profiles: experienced vs. new users

	New users	Registed users
# dialogues	350	467
# utterances	1288	717
# words	4141	1454
av. dialogue length	3.7 turns	1.5 turns
av. utterance length	3.2 words	2.0 words
OOV rate	3.6%	1.9%
Disfluency rate	2.8%	2.1%
OOD rate	10.6%	3.3%
Dialogues with OOD	14.3%	3.3%

OOV=out of vocabulary word
OOD=out of domain utterance

Experienced users prefer keywords and don't make comments !!

Dialogue-oriented characterization and evaluation



$$IER = \frac{nb(False\ Alarms) + nb(False\ Rejections) + nb(Substitutions)}{nb(Valid\ Content)}$$

A few comments as a conclusion ..

Comment & Conclusion

- SLU is not speech dictation !
  - Can't have always a good match between training data and field data
  - Must integrate ASR and SLU processes
- Processing speech transcripts is not processing text !
  - Dealing with uncertainty
  - Specificities of speech transcripts

Comment & Conclusion

- A speech message can only be understood with its context of utterance
  - Local + global context
- A speech message is not just made with words
  - Acoustic and prosodic information are crucial
    - *Voice quality*
    - *Emotion*
    - *Genre*
    - *...*

Thank you

LABORATOIRE INFORMATIQUE D'AVIGNON

UNIVERSITÉ D'AVIGNON
UNIVERSITÉ D'AVIGNON ET DES PAYS DE VAUCLUSE