
SPOKEN LANGUAGE UNDERSTANDING STRATEGIES ON THE FRANCE TELECOM 3000 VOICE AGENCY CORPUS

Géraldine Damnati¹

Frédéric Béchet²

Renato de Mori²

1 France Télécom R&D, France

2 LIA, Université d'Avignon, France



Context of this study

- Spoken Language Understanding
 - Spoken dialog systems
 - From word transcriptions to interpretations
 - Structure, theme, entities, etc.
 - Command in the dialog application
 - Corpus-based method = Need for observations
 - Direct observations
 - Linked to an action of the speaker
 - Indirect observations
 - Manual annotations of spoken message
 - Automatic annotations from deployed systems

Working with corpora from deployed SDS

- **Deployed system = real life issues !!**
 - **Users**
 - Very spontaneous speech
 - Very large variability
 - Speech: accents, language
 - Usage: different classes of users (new and regulars)
 - Unpredictable behaviors
 - Comments, incoherence
 - **System**
 - “unlimited” size of training corpus
 - Corpus collected daily
 - Using automatic annotations?
 - How building a new SLU?

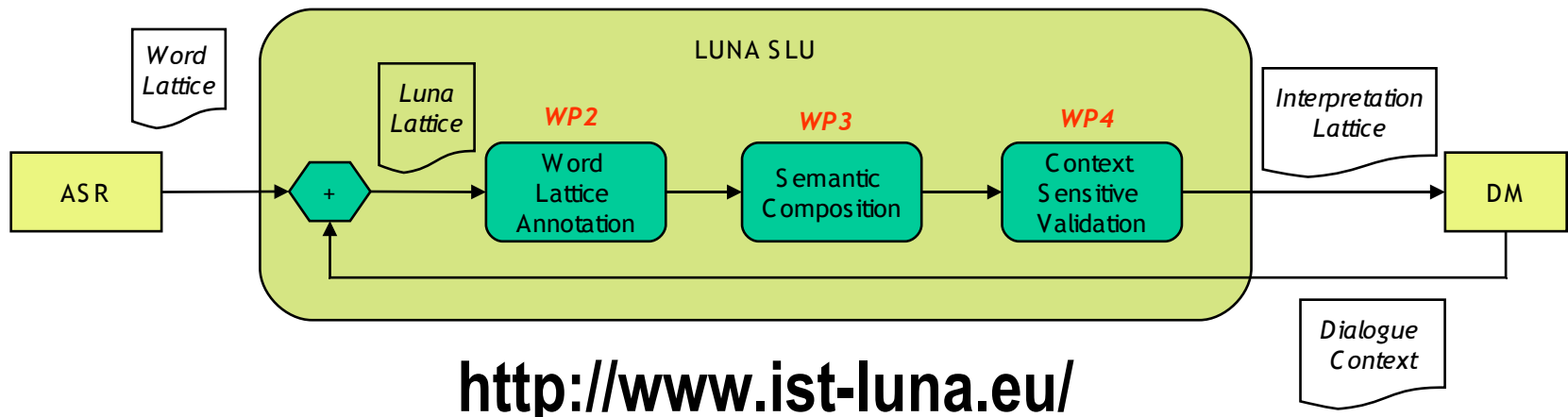
SLU Strategy proposed

- Integrated approach
 - ASR ↔ SLU ↔ Dialog Manager
 - All 3 processes should collaborate
 - Definition of a context
 - ASR+SLU+Dialog Manager: context adaptation
 - ASR output = multiple hypothesis (word lattice)
 - SLU = from a word lattice to an « interpretation lattice »
 - Manager = decision strategy on multiple hypothesis output
 - Contextual information used in each level
 - A priori information on the application domain
 - Dynamic information provided by the dialog manager



SLU Strategy: the LUNA project

- Developed through the FP6 European project: LUNA
 - Started in September 2006 (Aachen, Avignon, Trento, Warsaw, Loquendo, France Telecom, CSI Piemonte)
- Goal: Robust multilingual SLU strategies
- Multi level semantic representation
 - Concept decoding: from words to concepts
 - Semantic composition: from concepts to interpretations



FT 3000 Voice Agency service

- **Service**
 - obtain information about FT services
 - purchase almost 30 different services
 - access account
 - check consumption, pay bills
 - call forwarding, voice messaging
- **Deployed since October 2005**
- **Corpus collected daily**
- **Dealing with different kind of speech**
 - Speech/non speech
 - Speech out-of-domain/speech in domain
 - **Comments from the users**
 - Speech with a valid content/invalid content
- **Experienced vs. New users**



FT 3000 Voice Agency service

- Deployed system
 - Sequential approach
 - ASR 1-best => SLU => Dialog Manager
 - Semantic model
 - VERBATEAM
 - 2-level model
 - 1st level: word to concept
 - 2nd level: concept to interpretation
 - Non stochastic model
 - Hand written rules
 - Finite state dialog manager



FT 3000 Voice Agency service

- Semantic model

- 1st level: word to concept

- Concept = basic “bricks” on which a global interpretation of a spoken message can be built
 - 2 kinds of concepts
 - Concept = sequence of keywords representing services
 - ~100 concepts. Ex:
 - **illimités dix numéros** : [I10N]
 - **trente_et_un dix** : [AtoutPartout]
 - Concept = segments expressing a request, linked to a speech act
 - ~300 grammars. Ex:
 - **au fur et à mesure** : [Rapidement]
 - **comment diminuer** : [Limiter]

FT 3000 Voice Agency service

– 2nd level: concept to interpretation

- Logical rules on the concepts
- Ordered list: first match
- ~3000 rules

- Example:

```
( (Resilier|Annuler|Supprimer|Arreter|Plu)
```

```
# ((Appel|Appelle|Telephone|Telephoner) & Frequent &  
Domicile) )
```

```
=> {Gest(Resilier,Ambi(AtoutsPlus,HeureLocale,ForfaitLocal)) }
```

From a sequential to an integrated SLU

- **Method proposed**
 - ASR output = word lattice
 - Language Model = detection of out-of-domain segments
 - Concepts = local grammars + Concept tagger (HMM-based)
 - Interpretation rules
 - Encoded as transducers
 - Concept tags as input
 - Rule ID + rank in the rule database
 - Dialog states
 - Language model on the dialog states
 - All models implemented using a Finite State Machine approach
 - AT&T FSM & GRM Libraries



Detection of Out-of-Domain segments

- Modeling out-of-domain
 - Comments from the callers. Ex:
 - “I’ve already said that”
 - “what am I suppose to say now”
 - “I can’t believe it”
 - “you **** **”
- Specific 2-level language model
 - 1 general LM + 1 LM trained on the comment segments
 - Ex: **<s> w1 <comment> w2 w3 </comment> w4 </s>**

$$P^{G+OOD}(w_1, w_2, w_3, w_4) = P^G(w_1|start) \times P^G(_{OOD}_|w_1) \times P^{OOD}(w_2|start) \times P^{OOD}(w_3|w_2) \times P^{OOD}(end|w_3) \times P^G(w_4|_{OOD}_)$$

Stochastic model

SLU



$S = \{S_0, S_1, \dots, S_k\}$ Sequence of dialog states

$Y = \{Y_1, Y_2, \dots, Y_k\}$ Sequence of utterances

$\Gamma = \{\Gamma_1, \Gamma_2, \dots, \Gamma_k\}$ Sequence of interpretations

$C = c_1, c_2, \dots, c_n$ Basic concept string

$W = w_1, w_2, \dots, w_l$ Word string

$$P(S|Y) = \sum_{\Gamma} P(S\Gamma|Y)$$

Goal: obtaining the best sequence of states S over a sequence of spoken utterances Y

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

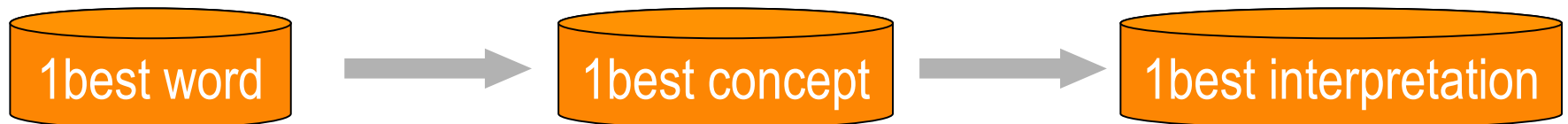
recursive calculation with: $H_k = \{S_{1,k-1}, \Gamma_{1,k-1}\}$

Strategy 1: sequential

- Best sequence of words $\hat{W} = \underset{W}{\operatorname{argmax}} P(W|Y)$
- Best sequence of concepts $\hat{C} = \underset{C}{\operatorname{argmax}} P(C|\hat{W})$
- Best interpretation $\hat{\Gamma} = \underset{\Gamma}{\operatorname{argmax}} P(\Gamma|\hat{C})$

Probability of a dialog state S at turn k:

$$P(S_k \hat{\Gamma}_k | H_k Y) \approx P(S_k | \hat{\Gamma}_k) P(\hat{\Gamma}_k | \hat{C}_k) P(\hat{C}_k | \hat{W}_k) P(\hat{W}_k | Y_k)$$



Strategy 2: integrated search

- Best sequence of concepts+words+interpretation

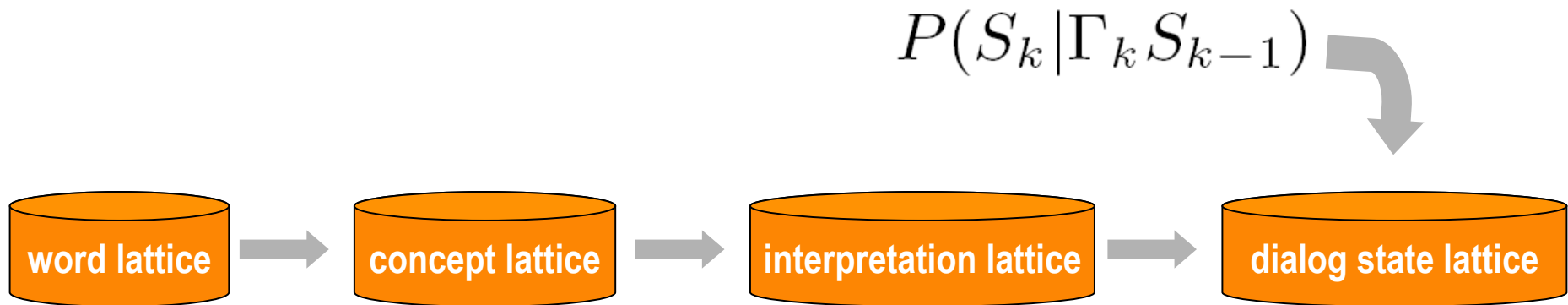
$$P(S_k \Gamma_k | H_k Y) \approx P(S_k | \Gamma_k) \times \max_{W_k, C_k} P(\Gamma_k | C_k) P(C_k | W_k) P(W_k | Y_k)$$

For estimating the probability of a dialog state and an interpretation: joint search for the best sequence of words and concepts



Strategy 3: integrated search + dialog context

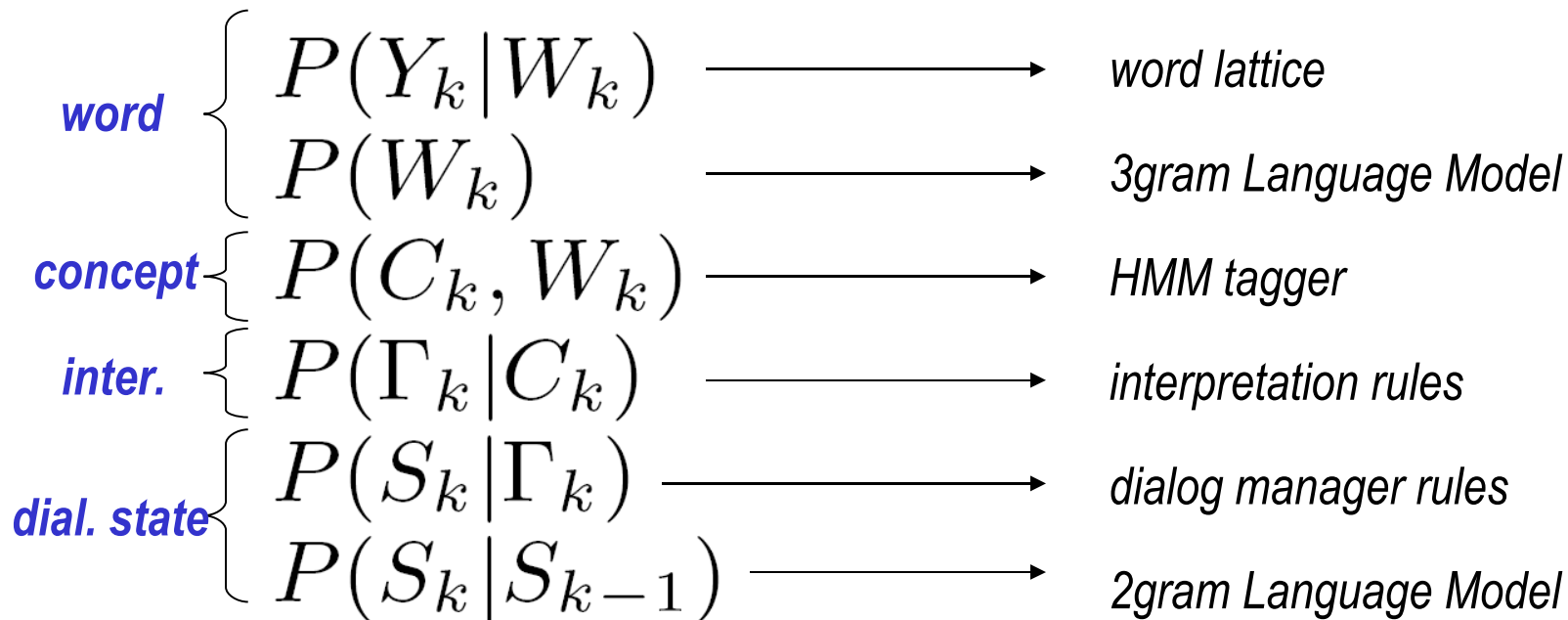
- Bigram Language Model in the dialog states:



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

Implementation with a Finite State Machine approach

- Language Models + Local grammars + rules
 - AT&T FSM + GRM Libraries



Corpus

- Corpus
 - Training
 - 44K utterances for LM (word and concept)
 - 7.4K dialogues (dialog state LM)
 - Test
 - 816 dialogues / 1950 utterances
- Two types of dialogues
 - "Transit"
 - Request for a specific task and rerouting towards the corresponding service
 - 80% of the calls, 60% of the utterances, registered users
 - "Other"
 - Request for information or new service purchase
 - Longer dialogs, more comments, often new users



Corpus

- User profiles: experienced vs. new users

	other	transit
# dialogues	350	467
# utterances	1288	717
# words	4141	1454
av. dialogue length	3.7	1.5
av. utterance length	3.2	2.0
OOV rate (%)	3.6	1.9
disfluency rate (%)	2.8	2.1

	other	transit
# dialogues	350	467
# utterances	1288	717
# OOD comments	137	24
OOD rate (%)	10.6	3.3
dialogues with OOD (%)	14.3	3.6

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

Results with the OOD Language Model

- OOD LM is very useful on the *other* dialogues, which contain most of the comments from the users
- No negative impact on the *transit* dialogues

IER	all	other	transit
size	1953	734	1219
LM^G	16.5	22.3	13.0
LM^{G+OOD}	15.0	18.6	12.8

- IER=Interpretation Error Rate (at the turn level)
- correct=all the attribute/value components must be correct

Example: {Gest(Resilier,Ambi(AtoutsPlus,HeureLocale,ForfaitLocal))}

Results according to the SLU strategy

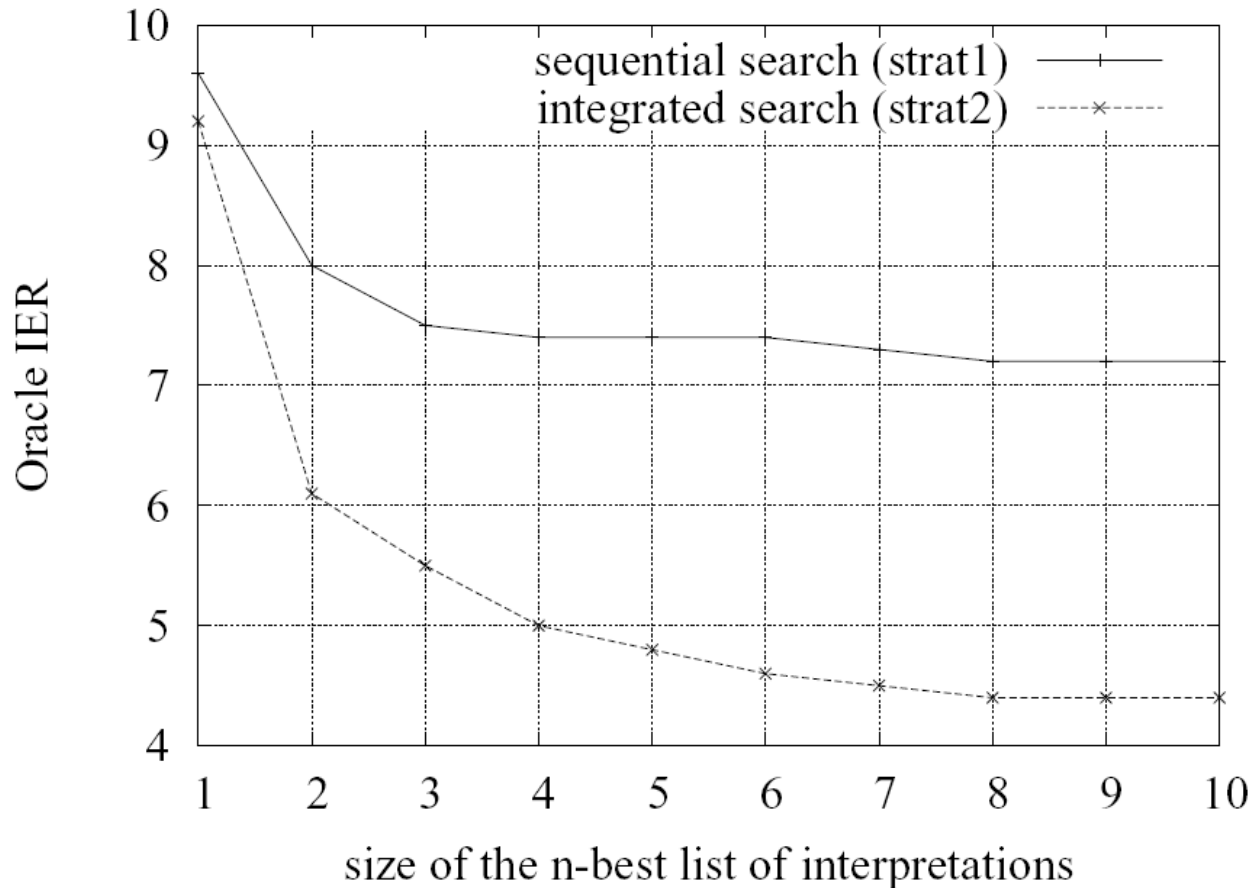
- Strat1=sequential
- Strat2=integrated
- Strat3=integrated + dialog context
- WER = Word Error Rate
- CER = Concept Error Rate
- IER = Interpretation Error Rate

<i>corpus</i>	all		
<i>error</i>	<i>WER</i>	<i>CER</i>	<i>IER</i>
strat1	40.1	24.4	15.0
strat2	38.2	22.5	14.5
strat3	38.3	22.5	14.7

- **Some gains obtained by using an integrated search (strat2)**
 - Higher correct values, but too many insertions (False Acceptance)
- **No improvement observed by using dialog context**
 - Very short dialogues (1.5 turns for Transit, 3.7 for Other)

Oracle measures

- sequential vs integrated strategy oracle error rates



Oracle measures

- Oracles in
 - Word lattice (WER)
 - Concept lattice (CER)
 - Interpretation lattice (IER)

<i>level</i>	<i>1-best</i>	<i>Oracle hyp.</i>
WER	33.7	20.0
CER	21.2	9.7
IER	13.0	4.4

- IER obtained on these Oracle hypotheses
 - IER on the Oracle 1-best word string = **9.8%**
 - IER on the Oracle 1-best concept string = **7.5%**
 - IER on the Oracle 1-best interpretation = **4.4%**

 *Joint search and optimization on the word/concept/interpretation levels*

SLU decision process

- Decision process based on multiple hypotheses output
- Example
 - Agreement on the different levels for detecting “expected” dialogs

IER	all		other		transit	
	IER	cover	IER	cover	IER	cover
1	15.0	100%	18.6	100%	12.8	100%
1^2	12.7	88.7%	15.1	86.4%	8.7	92.8%
1^2^3	12.0	87.6%	14.3	84.9%	8.3	92.3%

- Can be used to detect problematic dialogues

Conclusions

- For a better integration of the upstream and downstream processes
 - From a word lattice to an interpretation lattice
- Stochastic SLU can be trained on deployed Spoken Dialog System corpora
 - Manual + automatic annotations
- Need to take into account users profile and behaviors
 - Experienced vs. New users
 - Specific models for Out-Of-Domain segments
- Multiple hypotheses decision strategy
 - discriminant approaches
 - Context sensitive validation (**LUNA** project)



Perspectives

- Confidence measures and rejection strategies are crucial for processing “real” users’ utterances
 - Detecting as soon as possible «empty» utterances
 - Using «rich» search space only on reliable segments

