Vers la conception de robots conversationnels ("How to make talking robots")

Pierre Lison
Cognitive Systems @ Language Technology Lab
German Research Centre for Artificial Intelligence (DFKI GmbH)
Saarbrücken, Germany.

[pierre.lison@dfki.de]

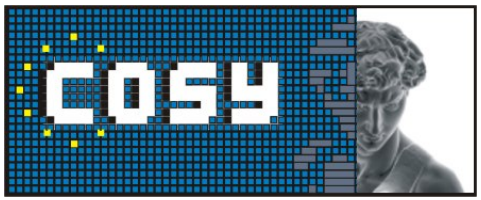
Weekly research seminar of CENTAL
Centre de Traitement Automatique du Langage
Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Deutsches Forschungszentrum für Künstliche Intelligenz GmbH

CoSy Project
"Cognitive systems for cognitive assistants"
EU FP6 Integrated project

Introduction

- The work we are going to present in this talk is part of the **CoSy** project (EU-funded integrated project).



- *Goal* : "Foundations of integrated, continuously developing cognitive architectures for embodied interactive agents".
- The emphasis here will be on **interaction/communication** : we seek to develop robots which are able to
  - operate in real-world, open-ended environments ;
  - *interact* with humans using (spoken) *natural language* to perform a variety of service-oriented tasks.

Introduction (cont'd)

- **Basic research question** : How do we make a robot understand *spoken*, *situated dialogue* ?
- $\Rightarrow$ Need to integrate a (rather sophisticated) **dialogue system** into the cognitive architecture.
- This dialogue system must encompass multiple processing stages, from speech recognition up to semantic interpretation.
- In this talk, we present (1) the spoken dialogue system we are currently using for our robotic platforms, as well as (2) some new techniques I developed (in my M.Sc. thesis) to make the dialogue system more **robust**.

Outline of the talk

Outline of the talk

Outline of the talk

Outline of the talk

Outline of the talk

Talking robots ?

- Our long-term aim :

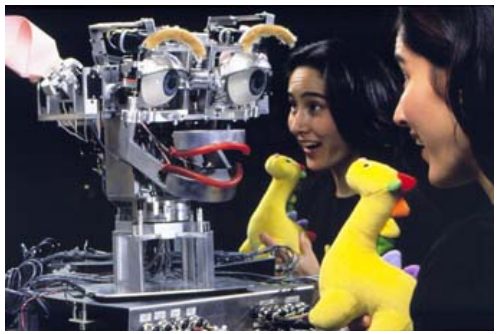
  « *Hi, I am C3-PO,Human Cyborg Relations.* »

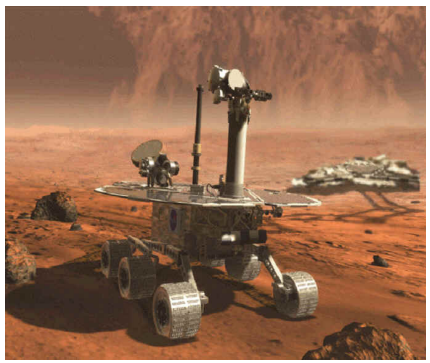
  (And he knows over 6 million languages...)

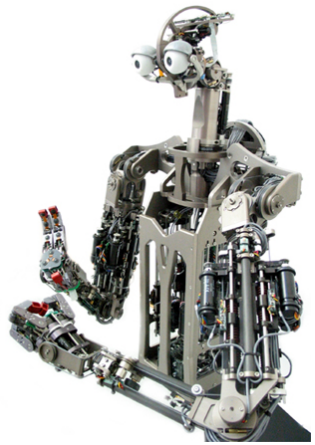
- For the time being, we'll obviously need to scale down our expectations...

Today's "state of the art"

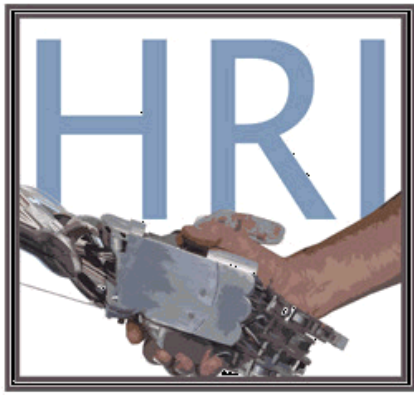
- Nevertheless, research in cognitive robotics is rapidly moving forward, and we are already able to do a few things :







Human-Robot Interaction



- How to make robots that actually *understand* what we say ? And that understand *why*, *when* and *how* they should say something ?
- Research in HRI seek to develop principles and techniques to allow *efficient* and *natural communication* between robots and humans
- *Interdisciplinary* research field : artificial intelligence, robotics, (computational) linguistics, & the social sciences – psychology, cognitive science, anthropology, etc.

Human-Robot Interaction (cont'd)

- HRI is always about *situated* interaction : Language often refers to reality, discusses actions & plans affecting that reality.
- **Situated dialogue understanding** is thus crucial for HRI.
  = Understanding and producing language, relative to a current or imaginable situation in which the agents are situated.
- It means that we *cannot* consider communication in isolution from the other modalities – we need to find meaningful ways to relate **language**, **action** and **situated reality**.

$\Rightarrow$ need to develop artificial **cognitive systems** able to integrate all these aspects into a common architecture.

Cognitive systems

What is cognition ?

- Cognition is more than intelligence...
- ... it is intelligence set in *reality*.
- "Cognition = perception + intelligence"

What is a cognitive system ?

- A cognitive system is a (artificial or biological) system able to actively *perceive* the environment it find itself in, *reason* about it and *achieve goals* through *plans* and *actions*.

Embodiment

- *Embodiment* modulates how a system sees, experiences, reality.
- "Cognition = embodiment[perception+intelligence]"
- Since they have very different "bodies" (perceptors and actuators), robots and human beings will experience and represent reality in very different ways.
- This difference of embodiment has profound implications for HRI : how can we "create a bridge" between two systems with wildly different conceptions of external reality ?

Philosophy and embodiment

- From dualism ..
  - Cogito ergo sum : the separation of mind and body
  - How do the mental and physical realms interact ?
- To behaviorism ...
  - Investigating (reactive) behavior linking mental and physical realms
  - What are the processes that underly cognition ?
- To contextualized phenomenology
  - Human activity is not context-free manipulation of representations, but contextualized experience of the body-environment system

R. Descartes
(1596-1650)

B.F. Skinner
(1904-1990)

M. Heidegger
(1889-1976)

Philosophy and embodiment (con'td)

- The question of development ...
  - Language is an inherently socially situated activity, and thus language use and development require embodiment
  - The body plays an essential role in development (interaction)
- And the phenomenology of mind
  - The way we perceive an object is determined by possibilities for bodily interaction

L.S. Vygotsky (1896-1934)

J. Piaget (1896-1980)

M. Merleau-Ponty (1908-1961)

Cognitive architectures

- Software architectures for "intelligent" robots are typically composed of several *distributed* and *cooperating* subsystems, such as :
  - communication ;
  - vision, perception ;
  - navigation and manipulation skills ;
  - deliberative processes (for planning, learning, reasoning).
- Our approach has been implemented as part of a *distributed cognitive architecture* for autonomous robots.
- In this presentation we focus only on the **communication** subarchitecture.

What is (situated) dialogue ?

- What is dialogue ?
  - spoken ("verbal") and, possibly, non-verbal **interactions** between two or more participants
  - The verbal and non-verbal interactions express **meaning**, which needs to be understood by all participants ("grounding") for the interaction to be successful
- What is *situated* dialogue ?
  - A form of dialogue in which the expressed meaning may (spatiotemporally) refer to a **physical environment**
  - *Grounding* situated dialogue thus involves being able to resolve the linguistic references to the (own) **situation awareness**, to yield a common ground in how to understand the environment

Spoken dialogue : Processing levels

- **Auditory** : speech recognition, intonation, etc.
- **Grammatical** : syntactic structure, semantic structure

  "A grammar specifies the relation between well-formed syntactic structures and their underlying (linguistic) meaning"
- **Discourse** : contextual reference resolution (anaphora, ellipsis), rhetorical relation resolution, etc.

  "Discourse interprets utterance meaning relative to the context, establishing how it contributes to furthering the discourse"

Open challenges

- **Robustness** in speech recognition :
  - noise, speaker independence, out-of-vocabulary words
  - poor performance of current ASR technology
- **Robustness** to ill-formed utterances :
  - partial, ungrammatical or extra-grammatical utterances
  - presence of various disfluencies (filled pauses, speech repairs, corrections, repetitions, etc.) in spoken dialogue.
- Pervasive **ambiguity** at all processing levels (lexical, syntactic, semantic, pragmatic)
- **Uncertainty** in contextual interpretation of utterances

Disfluencies in spoken dialogue : example

- Extract from a corpus of task-oriented spoken dialogue : *The Apollo Lunar Surface Journal*. [Audio file]

Example

Parker : That's all we need. Go ahead and park on your 045 <**okay**>. We'll give you an update when you're done.
Cernan : Jack is **[it]** worth coming right there ?
Schmitt : **err** looks like a pretty **go/** good location.
Cernan : okay.
Schmitt : We can sample the rim materials of this crater. **(Pause)** Bob, I'm at the **uh** south **uh** let's say east-southeast rim of a, **oh**, 30-meter crater - **err** in the light mantle, of course - up on the **uh** Scarp and maybe 300...**(correcting himself) err** 200 meters from the **uh** rim of Lara in **(inaudible)** northeast direction.

Communication subarchitecture

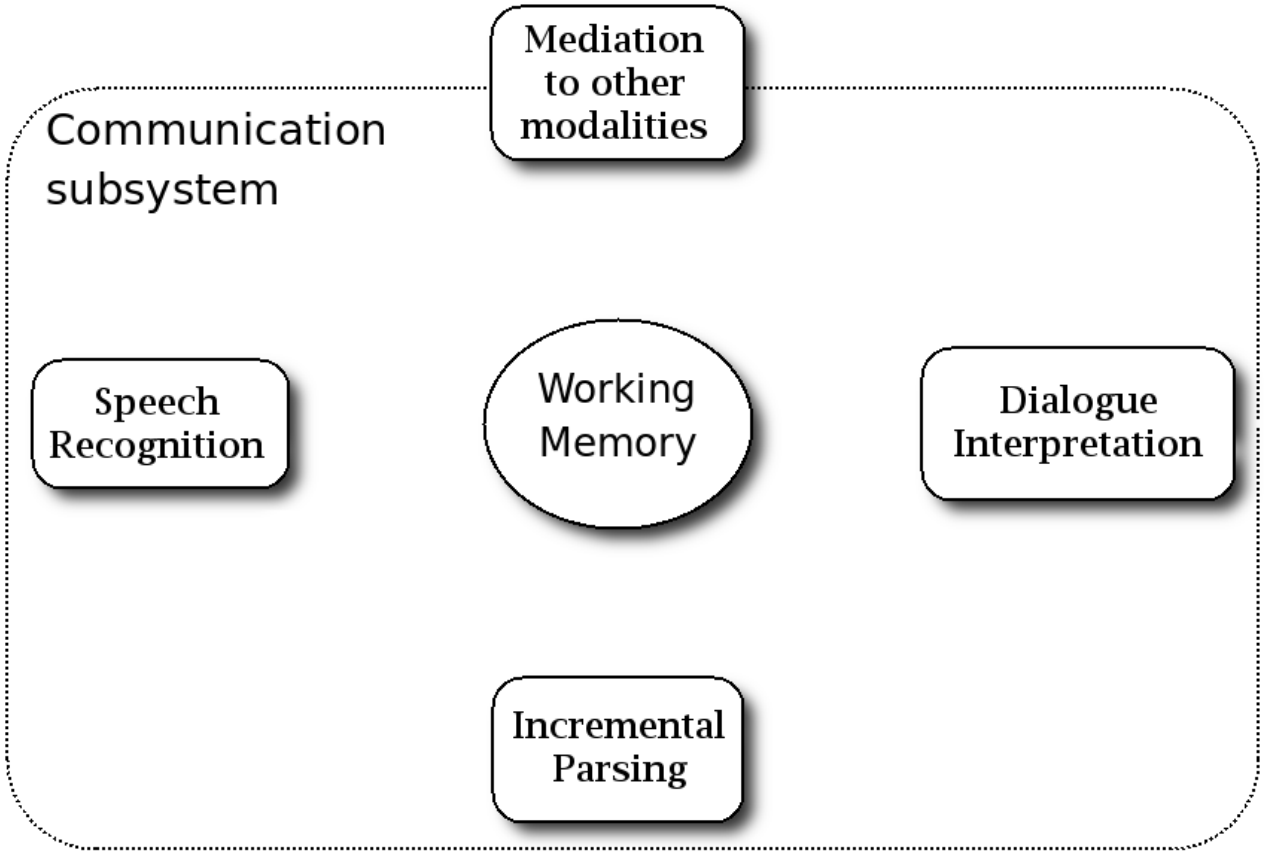


FIG.: Spoken dialogue comprehension

Communication subarchitecture

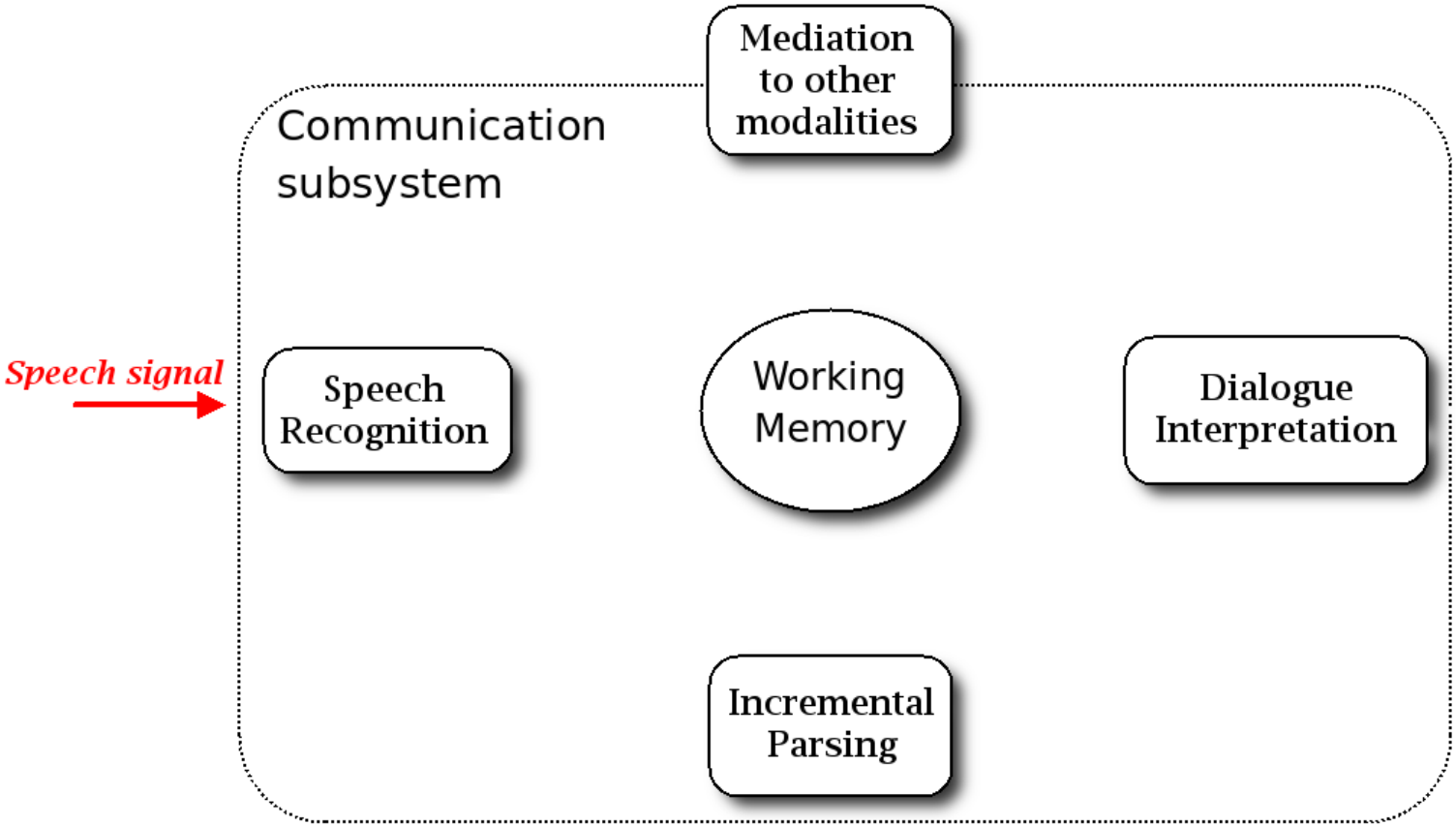


FIG.: Spoken dialogue comprehension : step 1

Communication subarchitecture

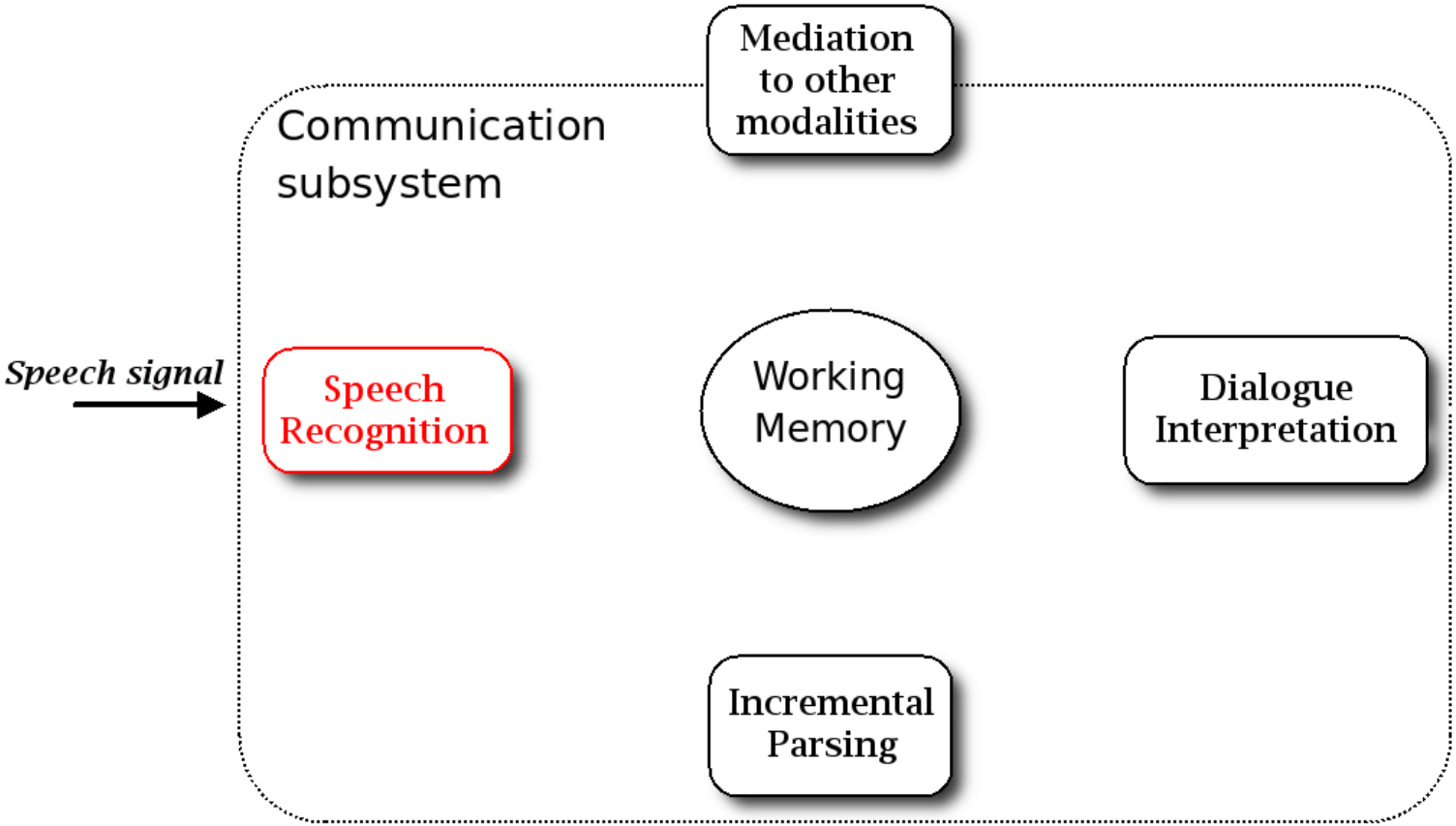


FIG.: Spoken dialogue comprehension : step 1

Communication subarchitecture

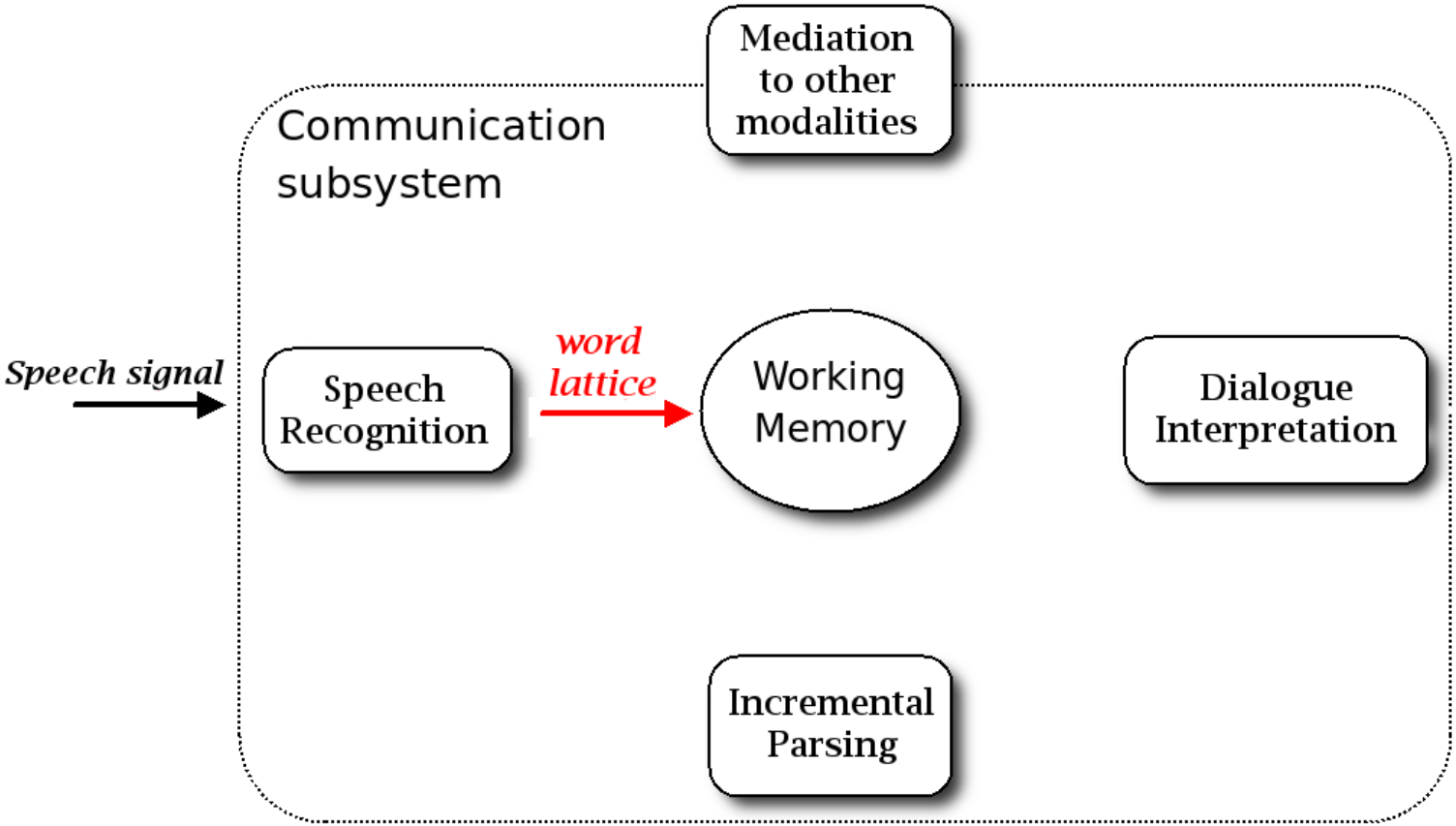


FIG.: Spoken dialogue comprehension : step 1

Communication subarchitecture

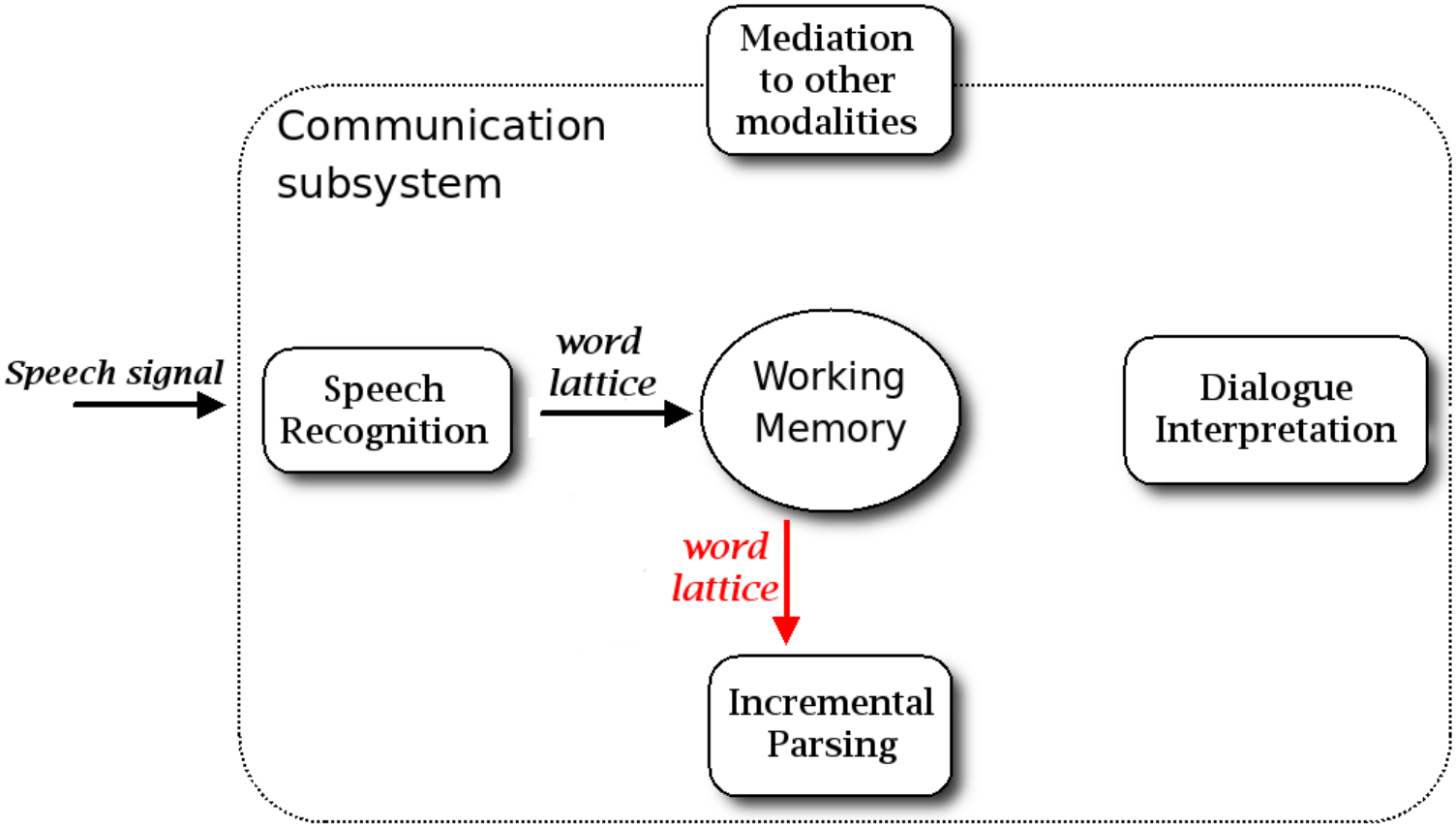


FIG.: Spoken dialogue comprehension : step 2

Communication subarchitecture

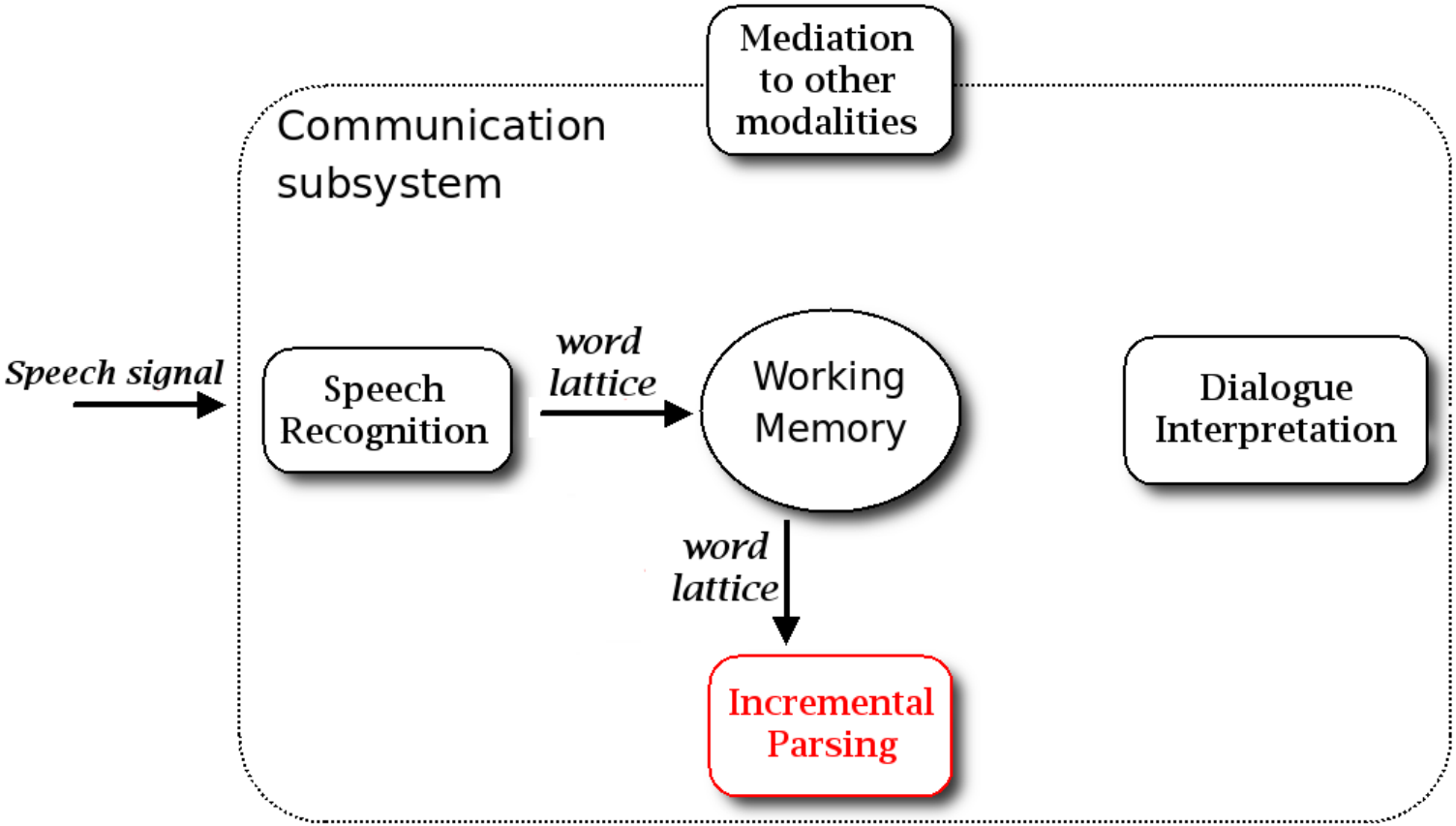


FIG.: Spoken dialogue comprehension : step 2

Communication subarchitecture

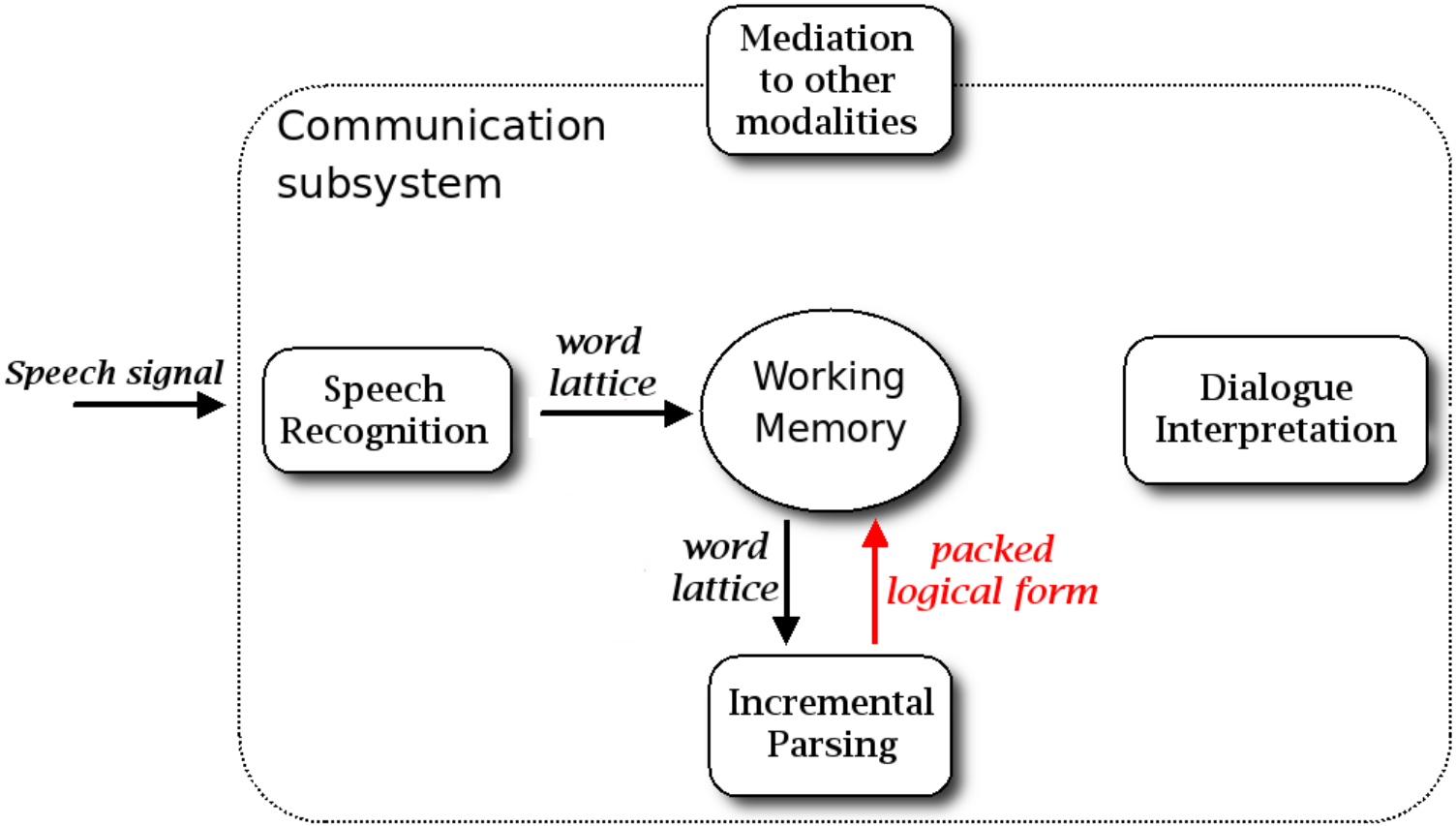


FIG.: Spoken dialogue comprehension : step 2

Communication subarchitecture

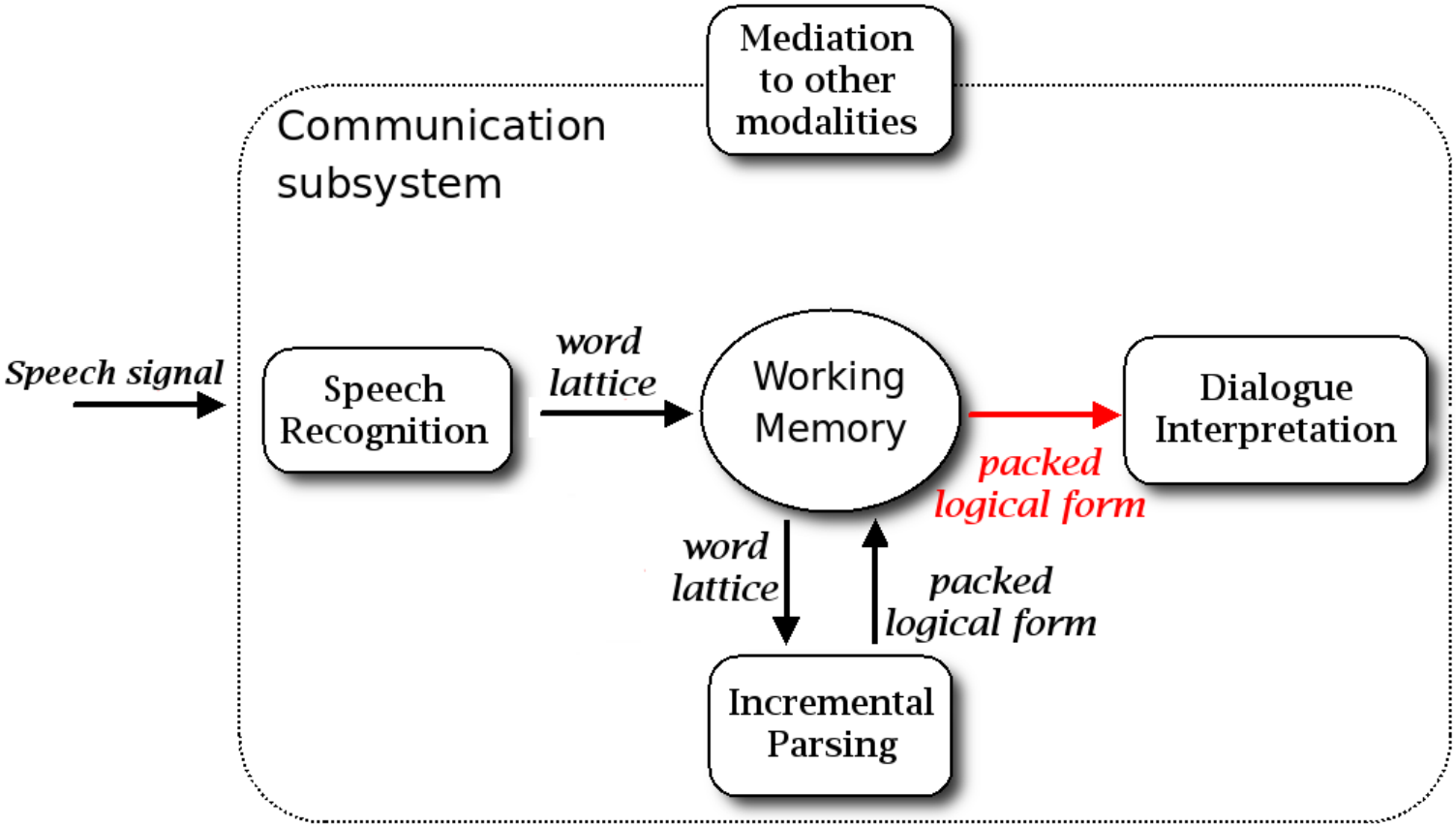


FIG.: Spoken dialogue comprehension : step 3

Communication subarchitecture

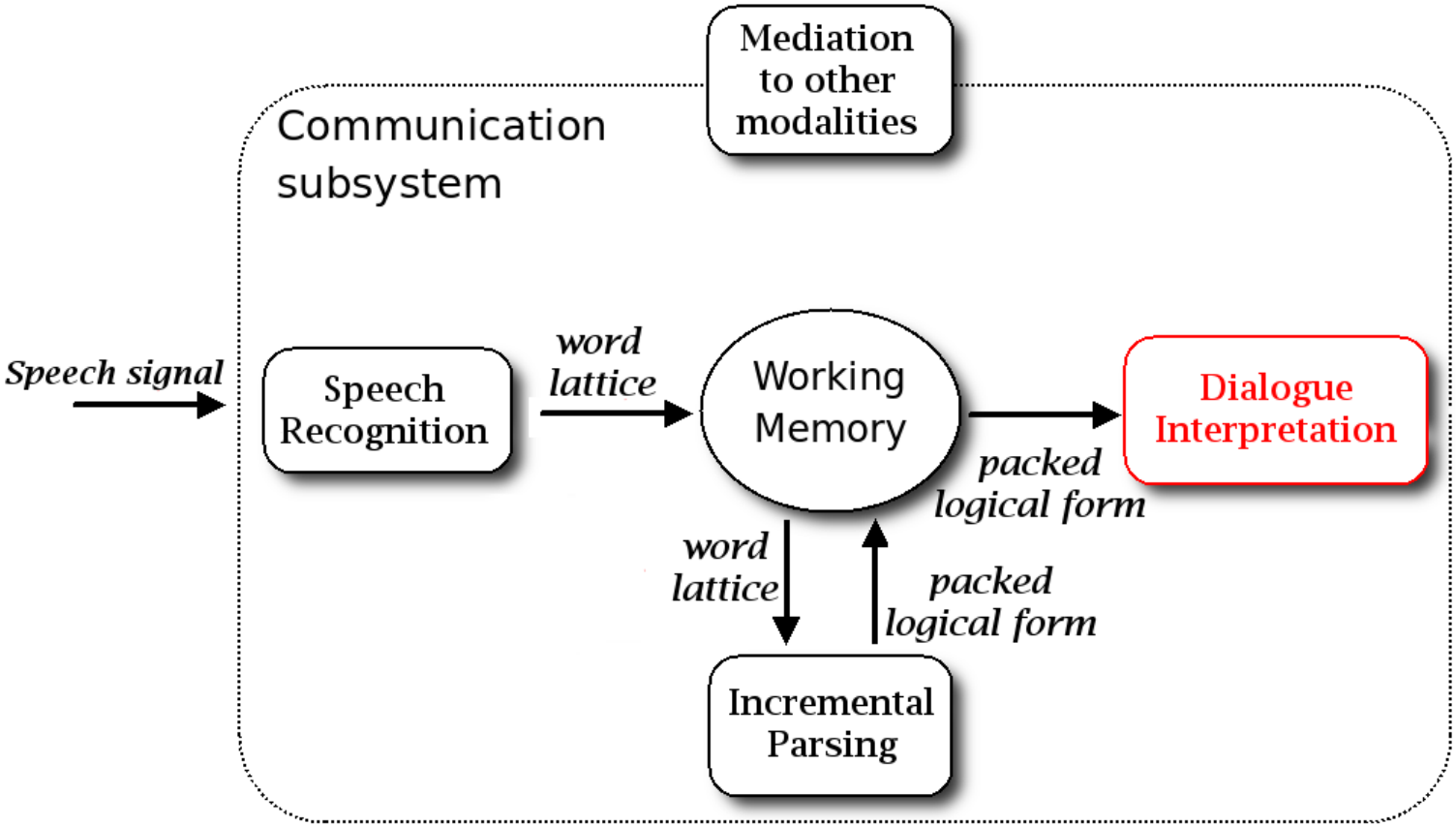


FIG.: Spoken dialogue comprehension : step 3

Communication subarchitecture

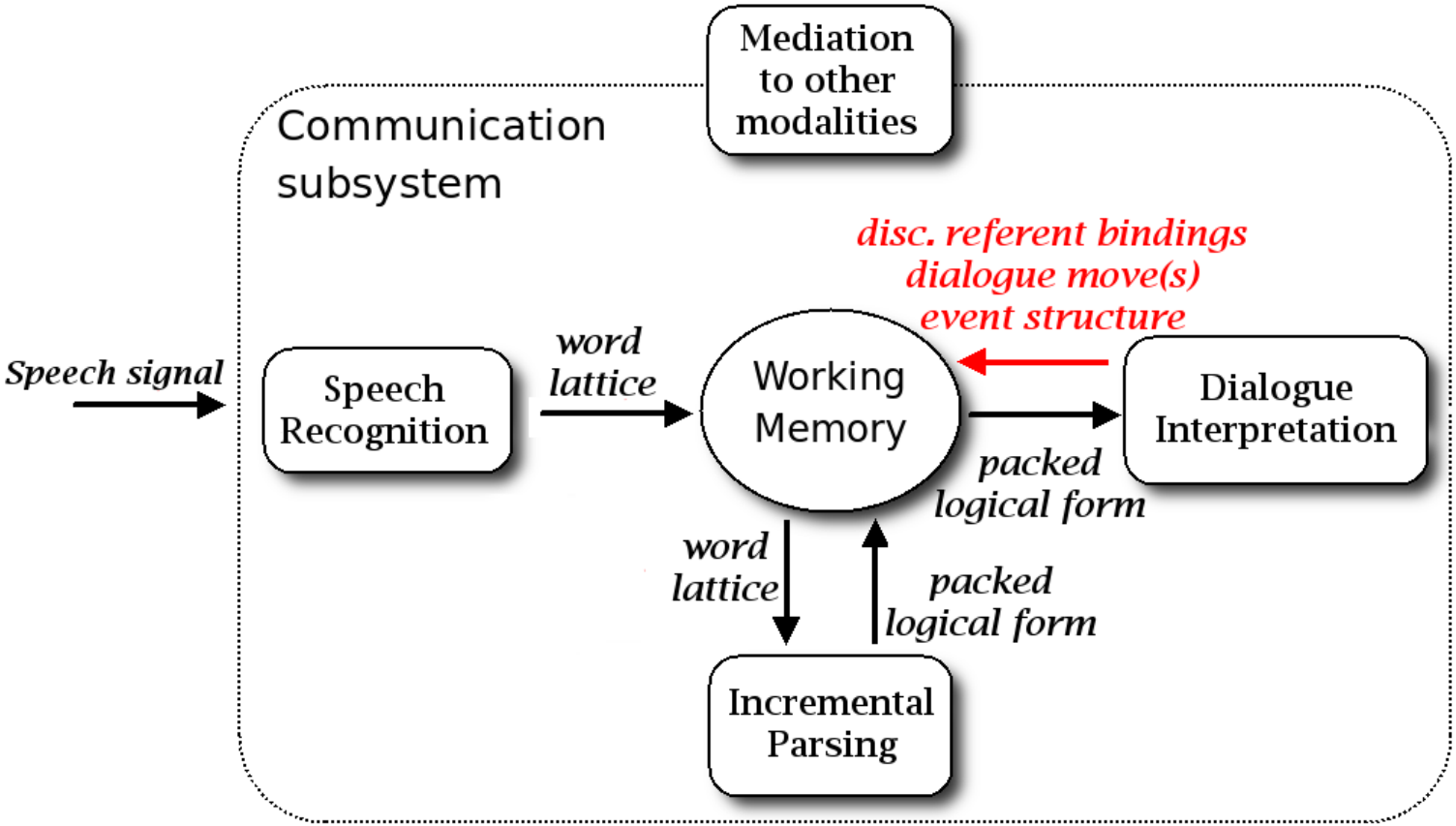


FIG.: Spoken dialogue comprehension : step 3

Communication subarchitecture

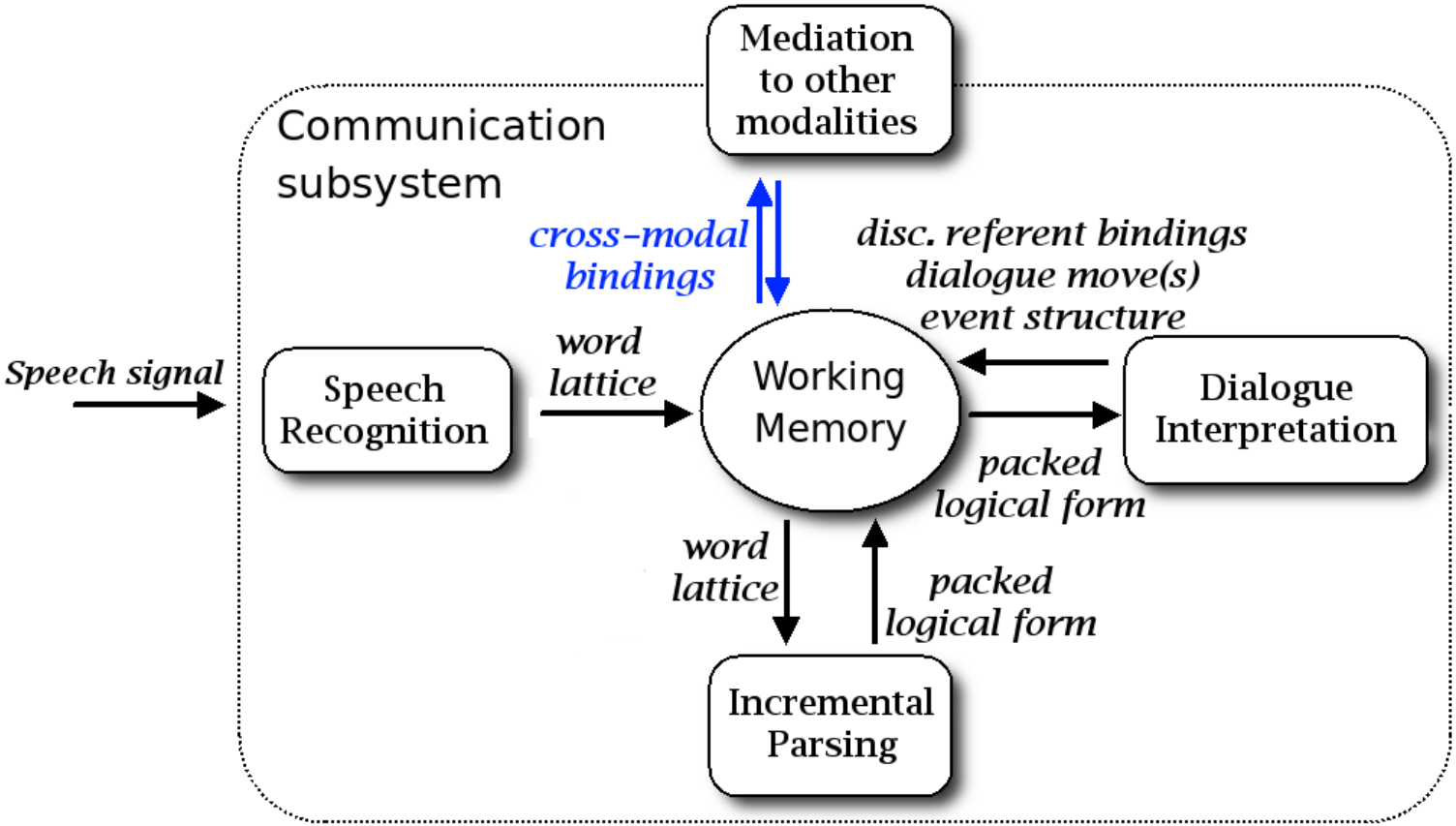


FIG.: Spoken dialogue comprehension : cross-modality

Grammatical analysis

- Why do a proper grammatical analysis ?
  - Proper analysis is better than keyword spotting !
  - Analysis reveals *structure*, shows how content is built up
  - Analysis does not need to be "brittle" : usefulness of partial analyses, possibilities for probabilistic inference, out-of-vocab
- What is the goal of such analysis ?
  - **Grammar relates form to meaning** ; the goal is to build a representation of the meaning(s) an utterance expresses
  - Conceive of linguistic meaning is *ontologically richly sorted, relational structures* to provide a basis for further interpretation

Parsing with CCG

- Bottom-up chart parsing, based on the CKY algorithm
- **Combinatory Categorial Grammar** (Steedman, 2000) :
  - Mildly *context-sensitive* grammar formalism ;
  - Fully *lexicalized* ;
  - Building structures using *combinatory rules* ;
  - Transparent semantics-syntax interface ;
  - Allows for incremental parsing and partial analyses ;
  - *Polynomial parseability* ;
  - Open-source development platform : OpenCCG [`http ://openccg.sf.net`].

Some issues in spoken dialogue processing

- Dialogue systems typically suffer from a lack of *robustness* and *adaptivity*.
- **Four issues** of particular importance :
  1. Difficulty of accommodating *spoken language phenomena* (disfluencies, fragments, etc.) in the dialogue system ;
  2. Pervasiveness of speech recognition errors ;
  3. Ambiguities arising at all processing levels ;
  4. Extra-grammaticality.

Some issues in spoken dialogue processing

- Dialogue systems typically suffer from a lack of *robustness* and *adaptivity*.
- **Four issues** of particular importance :
  1. Difficulty of accommodating *spoken language phenomena* (disfluencies, fragments, etc.) in the dialogue system ;
  2. Pervasiveness of speech recognition errors ;
  3. Ambiguities arising at all processing levels ;
  4. Extra-grammaticality.

Some issues in spoken dialogue processing

- Dialogue systems typically suffer from a lack of *robustness* and *adaptivity*.
- **Four issues** of particular importance :
  1. Difficulty of accommodating *spoken language phenomena* (disfluencies, fragments, etc.) in the dialogue system ;
  2. Pervasiveness of speech recognition errors ;
  3. Ambiguities arising at all processing levels ;
  4. Extra-grammaticality.

Some issues in spoken dialogue processing

- Dialogue systems typically suffer from a lack of *robustness* and *adaptivity*.
- **Four issues** of particular importance :
  1. Difficulty of accommodating *spoken language phenomena* (disfluencies, fragments, etc.) in the dialogue system ;
  2. Pervasiveness of speech recognition errors ;
  3. Ambiguities arising at all processing levels ;
  4. Extra-grammaticality.

Some issues in spoken dialogue processing

- Dialogue systems typically suffer from a lack of *robustness* and *adaptivity*.
- **Four issues** of particular importance :
  1. Difficulty of accommodating *spoken language phenomena* (disfluencies, fragments, etc.) in the dialogue system ;
  2. Pervasiveness of speech recognition errors ;
  3. Ambiguities arising at all processing levels ;
  4. Extra-grammaticality.

The strategy in three steps

1. Improve the performance of speech recognition by exploiting **contextual knowledge** about the *environment* and the *dialogue state*.
2. Allow for a **controlled relaxation** of the grammatical constraints to account for *spoken dialogue phenomena* and *speech recognition errors*.
3. Finally, apply a **discriminative model** on the resulting set of interpretations, in order to select the most likely one given the context.

The strategy in three steps

1. Improve the performance of speech recognition by exploiting **contextual knowledge** about the *environment* and the *dialogue state*.
2. Allow for a **controlled relaxation** of the grammatical constraints to account for *spoken dialogue phenomena* and *speech recognition errors*.
3. Finally, apply a **discriminative model** on the resulting set of interpretations, in order to select the most likely one given the context.

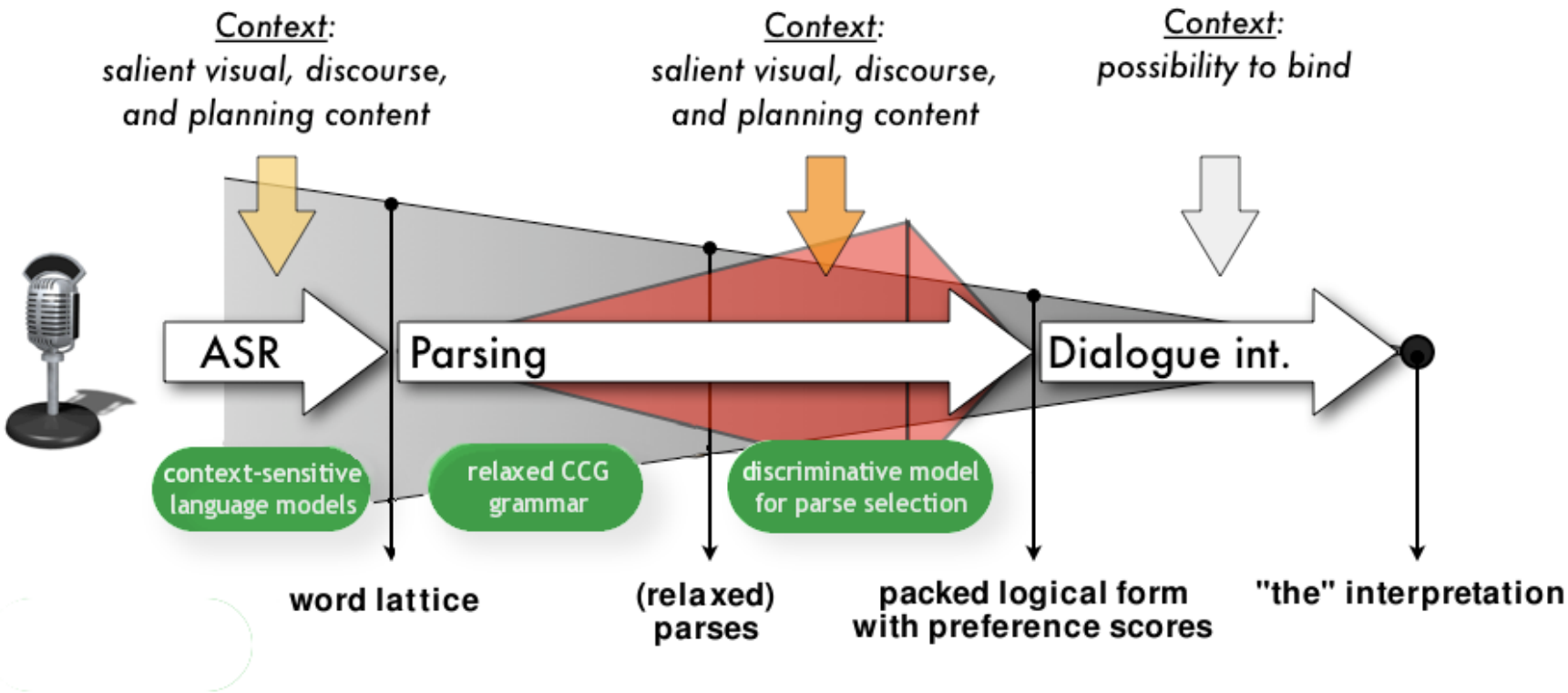
The strategy in three steps

1. Improve the performance of speech recognition by exploiting **contextual knowledge** about the *environment* and the *dialogue state*.
2. Allow for a **controlled relaxation** of the grammatical constraints to account for *spoken dialogue phenomena* and *speech recognition errors*.
3. Finally, apply a **discriminative model** on the resulting set of interpretations, in order to select the most likely one given the context.

The strategy in three steps

1. Improve the performance of speech recognition by exploiting **contextual knowledge** about the *environment* and the *dialogue state*.
2. Allow for a **controlled relaxation** of the grammatical constraints to account for *spoken dialogue phenomena* and *speech recognition errors*.
3. Finally, apply a **discriminative model** on the resulting set of interpretations, in order to select the most likely one given the context.

The strategy, graphically



Three defining characteristics

1. It is a **hybrid symbolic/statistical** approach
   - Combination of fined-grained linguistic resources with statistical models
   - Able to deliver both *deep* and *robust* dialogue processing
2. It is an **integrated** approach
   - Goes all the way from the speech signal up to the semantic & pragmatic interpretation
   - Interactions between various processing components
3. It is a **context-sensitive** approach
   - *Context* is used at every processing step to guide the processing
   - Both an *anticipation* tool and a *discrimination* tool

Three defining characteristics

1. It is a **hybrid symbolic/statistical** approach
   - Combination of fined-grained linguistic resources with statistical models
   - Able to deliver both *deep* and *robust* dialogue processing
2. It is an **integrated** approach
   - Goes all the way from the speech signal up to the semantic & pragmatic interpretation
   - Interactions between various processing components
3. It is a **context-sensitive** approach
   - *Context* is used at every processing step to guide the processing
   - Both an *anticipation* tool and a *discrimination* tool

Three defining characteristics

1. It is a **hybrid symbolic/statistical** approach
   - Combination of fined-grained linguistic resources with statistical models
   - Able to deliver both *deep* and *robust* dialogue processing
2. It is an **integrated** approach
   - Goes all the way from the speech signal up to the semantic & pragmatic interpretation
   - Interactions between various processing components
3. It is a **context-sensitive** approach
   - *Context* is used at every processing step to guide the processing
   - Both an *anticipation* tool and a *discrimination* tool

Three defining characteristics

1. It is a **hybrid symbolic/statistical** approach
   - Combination of fined-grained linguistic resources with statistical models
   - Able to deliver both *deep* and *robust* dialogue processing
2. It is an **integrated** approach
   - Goes all the way from the speech signal up to the semantic & pragmatic interpretation
   - Interactions between various processing components
3. It is a **context-sensitive** approach
   - *Context* is used at every processing step to guide the processing
   - Both an *anticipation* tool and a *discrimination* tool

- Original contributions of our M.Sc. thesis :
  1. A new model for **context-sensitive speech recognition**, which relies on the situated and dialogue context to dynamically adapt the ASR language model to the environment ;
  2. A new model for **robust parsing of spoken inputs**, based on a relaxed CCG grammar coupled with a discriminative model exploring a wide range of linguistic and contextual features.
  3. A fully working **implementation** for these two models, integrated into a cognitive architecture for autonomous robots. The implementation comes along with a complete set of training data and testing data.
- For more information about our approach, see the bibliography !

Conclusions

- **Human Robot Interaction** :
  - How do we make talking robots ?
  - Situated dialogue understanding requires *situated understanding*
- A **cognitive perspective** on HRI :
  - Cognition is *intelligence set in reality*
  - *Embodiment* modulates how a system sees, experiences, reality
- **Spoken dialogue processing** :
  - Different levels : auditory, grammatical, discourse
  - Various processing issues (speech recognition, disfluencies, sentence fragments, etc.)
  - Grammatical analysis using *Combinatory Categorial Grammar*
  - Steps towards improving the *robustness* of the dialogue system

Conclusions

- **Human Robot Interaction** :
  - How do we make talking robots ?
  - Situated dialogue understanding requires *situated understanding*
- A **cognitive perspective** on HRI :
  - Cognition is *intelligence set in reality*
  - *Embodiment* modulates how a system sees, experiences, reality
- **Spoken dialogue processing** :
  - Different levels : auditory, grammatical, discourse
  - Various processing issues (speech recognition, disfluencies, sentence fragments, etc.)
  - Grammatical analysis using *Combinatory Categorial Grammar*
  - Steps towards improving the *robustness* of the dialogue system

Conclusions

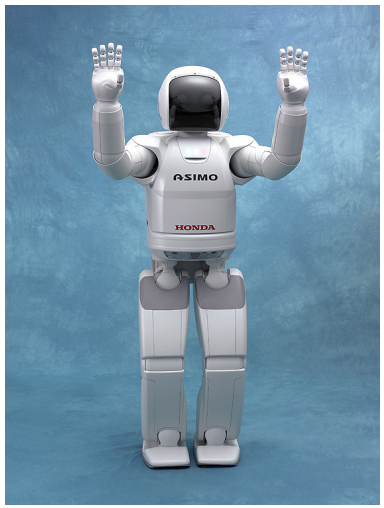
- **Human Robot Interaction** :
  - How do we make talking robots ?
  - Situated dialogue understanding requires *situated understanding*
- A **cognitive perspective** on HRI :
  - Cognition is *intelligence set in reality*
  - *Embodiment* modulates how a system sees, experiences, reality
- **Spoken dialogue processing** :
  - Different levels : auditory, grammatical, discourse
  - Various processing issues (speech recognition, disfluencies, sentence fragments, etc.)
  - Grammatical analysis using *Combinatory Categorial Grammar*
  - Steps towards improving the *robustness* of the dialogue system

Conclusions

- **Human Robot Interaction** :
  - How do we make talking robots ?
  - Situated dialogue understanding requires *situated understanding*
- A **cognitive perspective** on HRI :
  - Cognition is *intelligence set in reality*
  - *Embodiment* modulates how a system sees, experiences, reality
- **Spoken dialogue processing** :
  - Different levels : auditory, grammatical, discourse
  - Various processing issues (speech recognition, disfluencies, sentence fragments, etc.)
  - Grammatical analysis using *Combinatory Categorial Grammar*
  - Steps towards improving the *robustness* of the dialogue system

The end

Thank you for your attention ! !



⇒ **Questions, comments ?**

For more information, visit
http ://www.dfki.de/cosy

Bibliography I

C. Areces & B. ten Cate.
Hybrid Logics.
In P. Blackburn, F. Wolter & J. van Benthem, editeurs, Handbook of Modal Logics. Elsevier, 2006.

Nicholas Asher & Alex Lascarides.
Logics of conversation.
Cambridge University Press, 2003.

Bibliography II

J. Baldridge & G.-J. M. Kruijff.
Coupling CCG and Hybrid Logic Dependency Semantics.
In ACL'02 : Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pages 319–326, Philadelphia, PA, 2002. Association for Computational Linguistics.

Jonathan Ginzburg.
Semantics for conversation.
CSLI Publications, Stanford, 2009.
forthcoming in 2009.

Bibliography III

Michael A. Goodrich & Alan C. Schultz.
Human-robot interaction : a survey.
Foundations and Trends in Human-Computer Interaction, vol. 1, no. 3, pages 203–275, 2007.

N. Hawes, A. Sloman, J. Wyatt, M. Zillich, H. Jacobsson, G.J. M. Kruijff, M. Brenner, G. Berginc & D. Skocaj.
Towards an Integrated Robot with Multiple Cognitive Functions.
In AAAI, pages 1548–1553. AAAI Press, 2007.

Bibliography IV

H. Jacobsson, N.A. Hawes, G.J.M. Kruijff & J. Wyatt.
Crossmodal Content Binding in Information-Processing Architectures.
In Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction (HRI), Amsterdam, The Netherlands, March 12–15 2008.

P. Knoeferle & M.C. Crocker.
The coordinated interplay of scene, utterance, and world knowledge : evidence from eye tracking.
Cognitive Science, 2006.

Bibliography V

Geert-Jan M. Kruijff, Pierre Lison, Trevor Benjamin, Henrik Jacobsson & Nick Hawes.
Incremental, multi-level processing for comprehending situated dialogue in human-robot interaction.
In Proceedings of the Symposium on Language and Robots (LangRo'2007), Aveiro, Portugal, 2007.

Geert-Jan M. Kruijff, Pierre Lison, Trevor Benjamin, Henrik Jacobsson, Hendrik Zender & Ivana Kruijff-Korbayova.
Situated Dialogue Processing for Human-Robot Interaction.
In Henrik I. Christensen, Aaron Sloman, Geert-Jan M. Kruijff & Jeremy Wyatt, editeurs, Cognitive Systems : Final report of the CoSy project. MIT Press, 2009.

Bibliography VI

(in press).


George Lakoff & Mark Johnson.
Philosophy in the flesh.
Basic Books, 1999.


Pierre Lison.
Robust processing of situated spoken dialogue.
Master's thesis, Universität des Saarlandes, November 2008.


Pierre Lison.
A salience-driven approach to speech recognition for human-robot interaction.
In Proceedings of the 13th ESSLLI student session (ESSLLI 2008), Hamburg, Germany, 2008.

Bibliography VII


Pierre Lison & Geert-Jan M. Kruijff.
Salience-driven Contextual Priming of Speech Recognition for Human-Robot Interaction.
In Proceedings of the 18th European Conference on Artificial Intelligence (ECAI 2008), Patras, Greece, 2008.


Pierre Lison & Geert-Jan M. Kruijff.
An integrated approach to robust processing of situated spoken dialogue, 2009.
(in submission for EACL 2009).

Bibliography VIII

D. Roy.
Situation-Aware Spoken Language Processing.
In Royal Institute of Acoustics Workshop on Innovation in Speech Processing, Stratford-upon-Avon, England, 2001.

D.K. Roy.
Semiotic Schemas : A Framework for Grounding Language in Action and Perception.
Artificial Intelligence, vol. 167, no. 1-2, pages 170–205, 2005.

M. Steedman.
The syntactic process.
The MIT Press, 2000.

Bibliography IX

Mark Steedman & Jason Baldridge.
Combinatory Categorial Grammar.
In Robert Borsley & Kersti Börjars, editeurs,
Nontransformational Syntax : A Guide to Current Models.
Blackwell, Oxford, 2009.

G. van Noord, G. Bouma, R. Koeling & M.-J. Nederhof.
Robust grammatical analysis for spoken dialogue systems.
Journal of Natural Language Engineering, 1999.

Bibliography X

H. Zender, P. Jensfelt, Ó. Martínez Mozos, G.J.M. Kruijff & W. Burgard.
Conceptual Spatial Representations for Indoor Mobile Robots.
Robotics and Autonomous Systems, vol. 56, no. 6, June 2008. Special Issue "From Sensors to Human Spatial Concepts".

Luke Zettlemoyer & Michael Collins.
Online Learning of Relaxed CCG Grammars for Parsing to Logical Form.
In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pages 678–687, 2007.