Dialogue Strategy Modelling for Human-Agent Interaction

MIND seminar – LITIS – INSA Rouen

G. Dubuisson Duplessis

July 13th, 2017

ISIR-UPMC-CNRS, Paris, France

gdubuisson@telecom-paristech.fr

http://www.dubuissonduplessis.fr

Outline

Parcours professionnel

2005-2010 Diplôme d'ingénieur de l'INSA de Rouen, département Génie Mathématique

2009-2010 Master 2 MATIS, spécialité Modélisation, Interactions et Systèmes complexes à l'Université du Havre

2010-2014 Doctorat en informatique (CNU 27), spécialité intelligence artificielle au LITIS (équipe MIU/MIND)

2013-2014 ATER à l'Université Lille 1 (équipe SMAC)

2014-2016 Postdoctorant sur le projet CHISTERA Joker (LIMSI-CNRS)

2017- Postdoctorant sur le projet européen ARIA-VALUSPA (ISIR, UPMC, CNRS)

Parcours d'enseignement (360h éq. TD)

Lieux et niveaux

- IUT d'Orsay (bac +1)
- Université Lille 1 (L1, L3, M1)
- INSA de Rouen (bac +1, +2, +3)

Matières

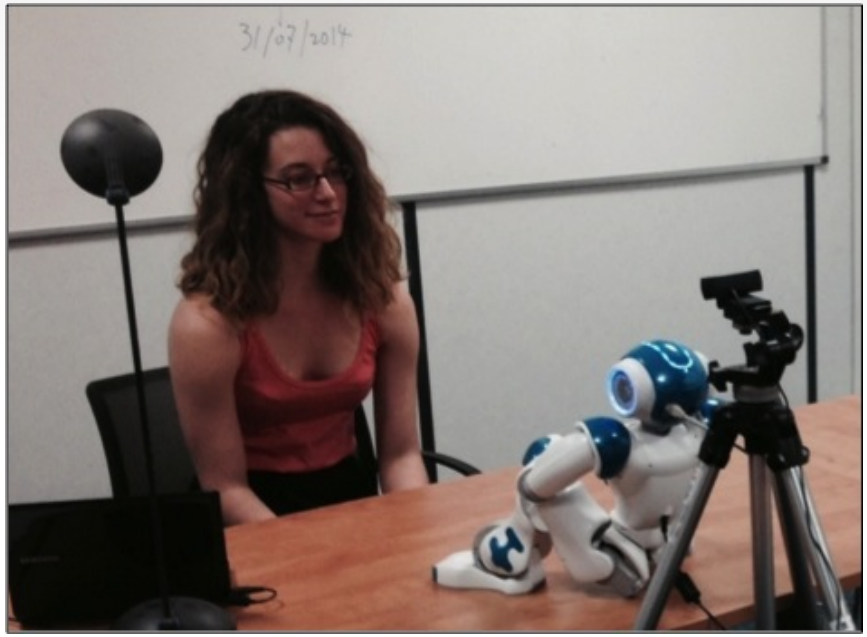
- Structure de données et algorithmes (introduction, avancé)
- Programmation en langage C, C++
- Génie logiciel (UML, gestion de projet)
- Programmation orienté objet (java, patron de conception)
- C2I
- Encadrement de projet
- Encadrement de stage

Outline

Selection-based Approach to Dialogue Management: JOKER Project

Main Features

- Multimodal spoken interaction
- Social dialogue
- Constraints: face-to-face, real-time



Devillers, L.; Rosset, S.; **Dubuisson Duplessis, G.**; Sehili, M. A.; Béchade, L.; Delaborde, A.; Gossart, C.; Letard, V.; Yang, F.; Yemez, Y.; Turker, B. B.; Sezgin, M.; El Haddad, K.; Dupont, S.; Luzzati, D.; Estève, Y.; Gilmartin, E.; Campbell, N., **Multimodal Data Collection of Human-Robot Humorous Interactions in the JOKER Project**, 6th International Conference on Affective Computing and Intelligent Interaction (ACII 2015), 2015, pp. 348-354

Goal

Problem

Maintaining human participation in dialogue when occur **unexpected** and **open-domain** utterances

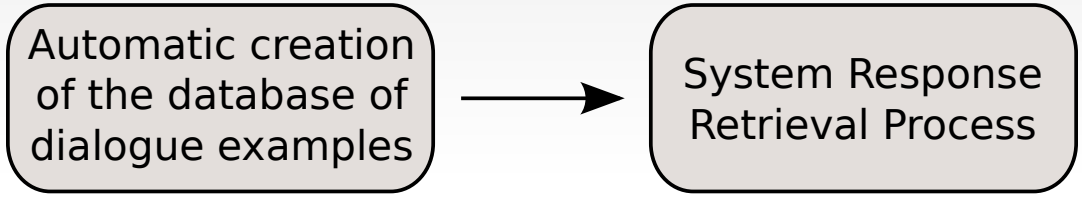
Real out-of-domain human utterances

- "what's up?"
- "what do you do all day?"
- "are you a machine?"

Goal

Producing appropriate system responses to open-domain utterances

Selection-based Approach to Dialogue Management (1/3)



Features

- Example-based dialogue modelling [Lee et al., 2009]
- Automatic, unsupervised
- Corpus-based

Related work: [Gandhe and Traum, 2007, Gandhe and Traum, 2013], [Banchs and Li, 2012], [Nio et al., 2014], [Ameixa et al., 2014]

Selection-based Approach to Dialogue Management (2/3)

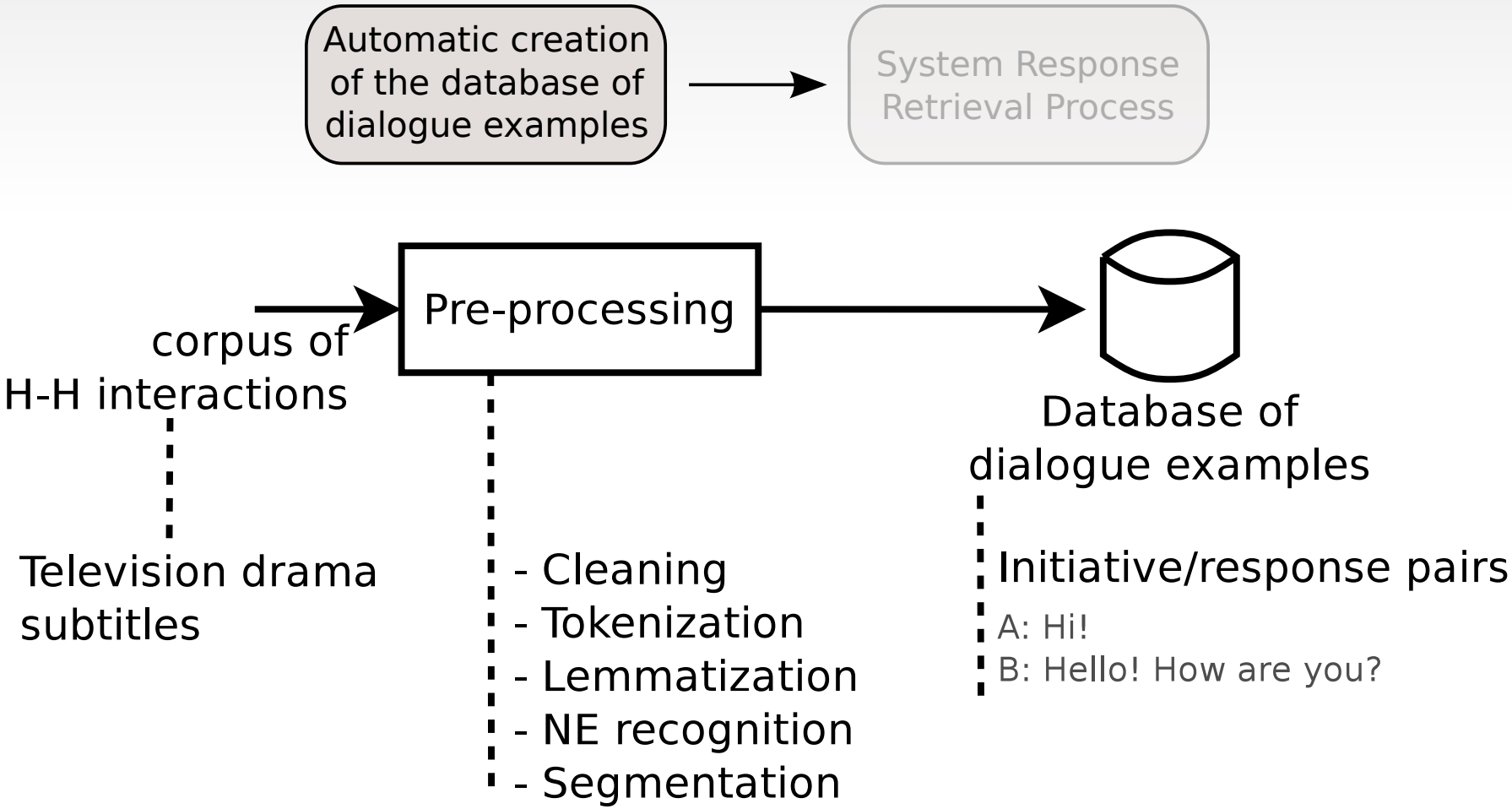
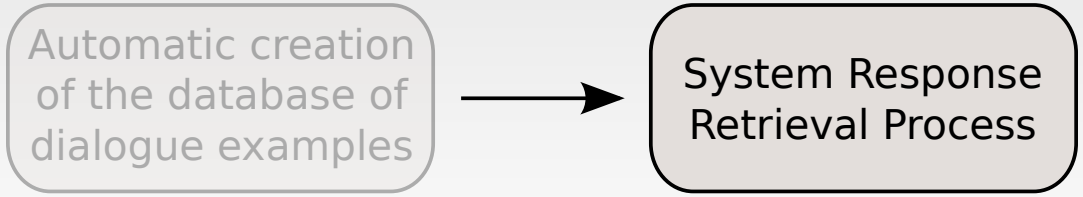
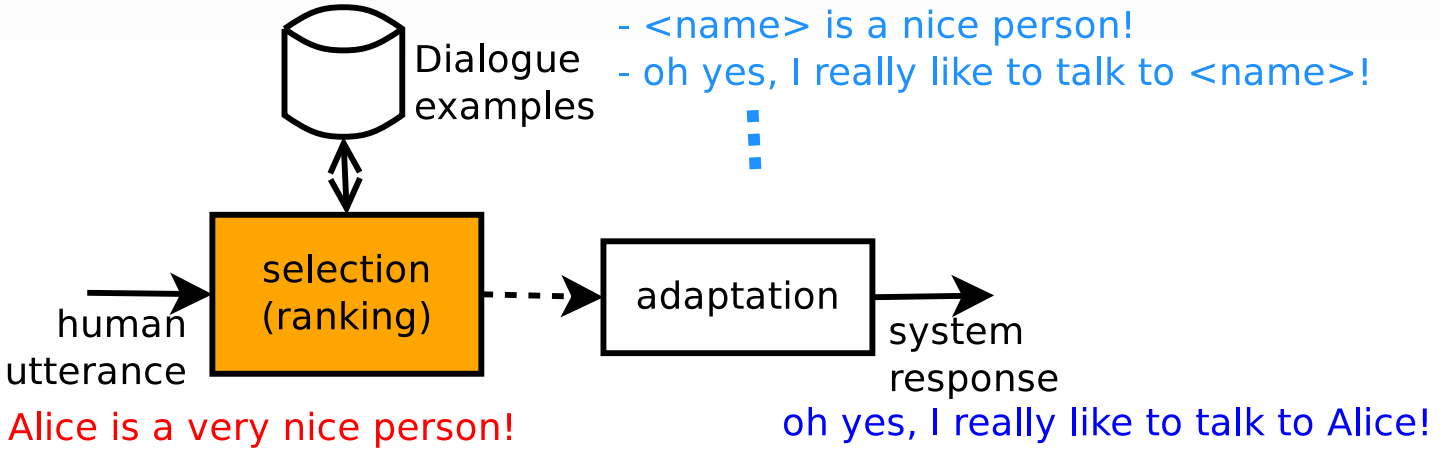


Figure 1 : NLP Pipeline: from a noisy corpus to a dialogue example database

Task: Ranking of Dialogue Examples





Data

- Open-domain utterances
- Unlabelled dialogue data
- $>$ 3 million dialogue examples from OpenSubtitles2016 [Lison and Tiedemann, 2016]

Proposition: RSTP-based Selection Model

Patterns of Language Use [Allwood, 1994, Clark, 1996]

- **Questions**: "How do you . . . ", "What are you . . . ", "May I . . . ", ". . . , isn't it?", . . .
- **Agreement**: "Yes , . . . ", "No , . . . ", "I do . . . ", "I do not . . . ", . . .
- **And many others**: "Let me tell you that . . . ", "I would like to . . . ", "My name is . . . ", . . .

Idea and Contributions

Exploiting Recurrent Surface Text Patterns of language use to **represent**, **index** and **compare** open-domain dialogue utterances for a retrieval task

Dubuisson Duplessis, G.; Charras, F.; Letard, V.; Ligozat, A.-L.; Rosset, S., **Utterance Retrieval based on Recurrent Surface Text Patterns**, 39th European Conference on Information Retrieval (ECIR), pp. 199–211, 2017

RSTP and Extraction

Definition of a RSTP

Contiguous sequence of tokens that appears in at least two utterances

Some Utterances and RSTP

- u_1: "How do you usually introduce yourself ?"
- u_2: "How do you know ?"
- u_3: "Hi !"

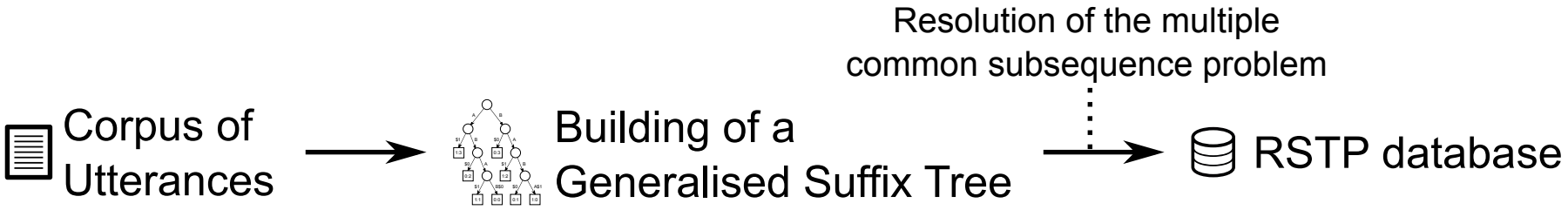


Figure 2 : Extraction of RSTP: resolving the MCSP [Gusfield, 1997]

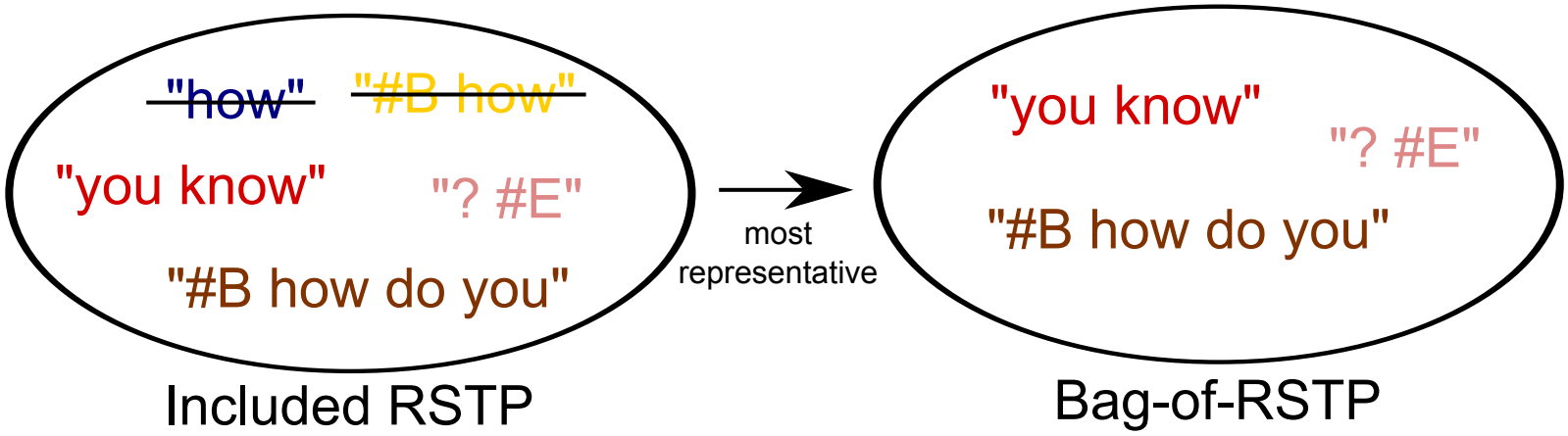
RSTP-based Representation of Utterances

Utterance Representation

Utterance = bag-of-RSTPs

D = {“how”, “you know”, “? #E”, “#B how”, “#B Hi ! #E”, “#B how do you”}

utterance = **“How do you know ?”**



GVSM Model based on RSTP

Retrieval Model

Generalised Vector Space Model (GVSM)

- Terms = RSTP
- Query = Human utterance
- Document = first utterance of a dialogue example

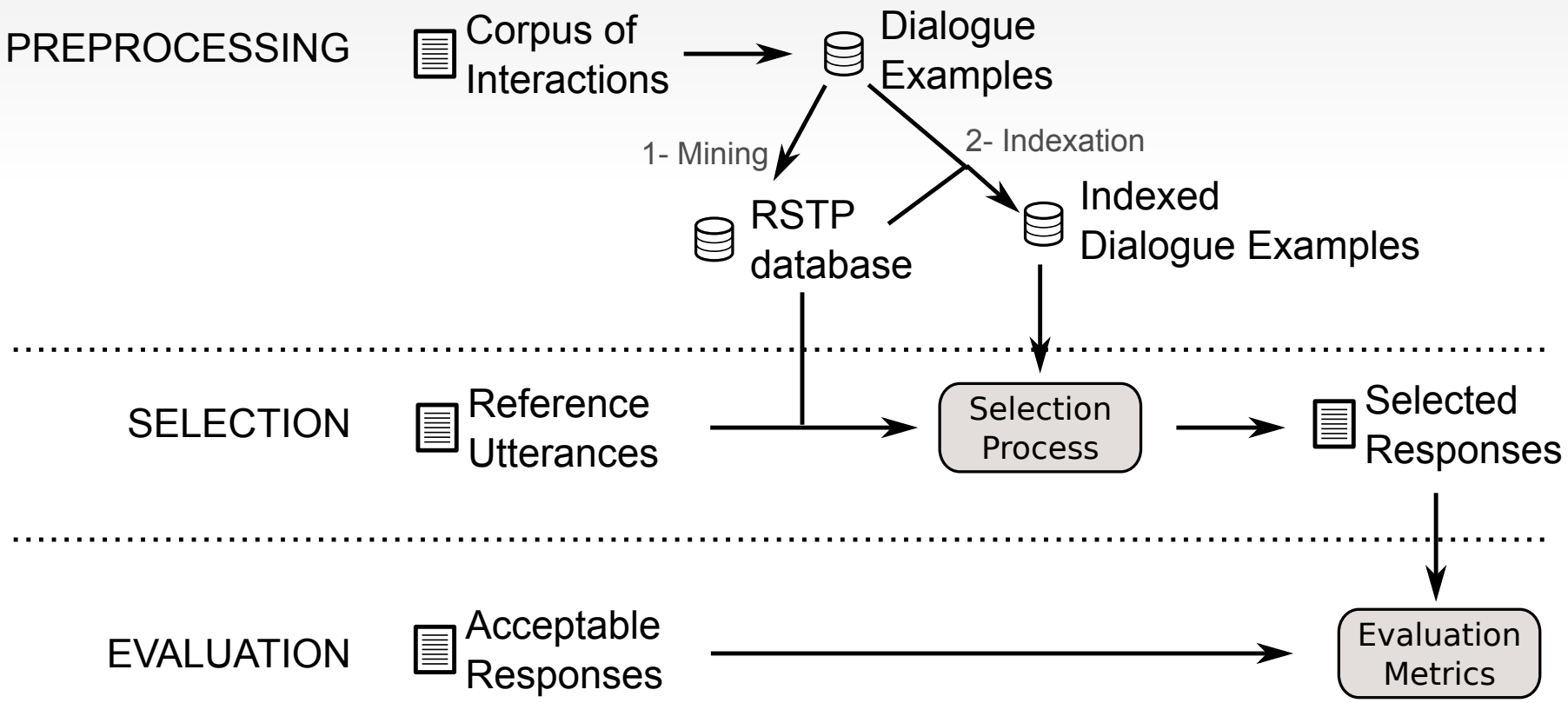
$$\vec{d}_\alpha \cdot \vec{q} = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{\alpha i} q_j \vec{t}_i \cdot \vec{t}_j \tag{1}$$

$$\vec{t}_i \cdot \vec{t}_j \approx \frac{|lgcs(t_i, t_j)|}{|t_i| + |t_j| - |lgcs(t_i, t_j)|} \text{ (Jaccard Index)}$$

Features

- Relatedness between RSTPs
- RSTP frequency and IDF

Experimentation Setup



Features of the Approach (1/2)

Does the size of the database of RSTP explode?

Corpus

Subset of the OpenSubtitles 2016 corpus (approx. 3 million unique utterances)

RSTP database	Full	Used
Size	5,776,901	3,846,956
Tokens per RSTP		
. . . avg/median	4.77/4.0	4.57/4.0
. . . std, min/max	2.23, 1/157	1.96, 1/157

Features of the Approach (2/2)

How does the RSTP method compare to n-gram models?

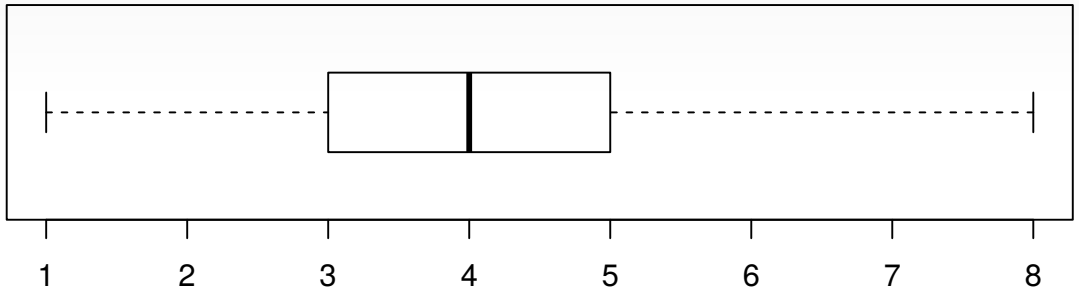


Figure 3 : Distribution of the size of the RSTP effectively used to represent the initiative utterances (in tokens, including begin and end markers).

Sparsity

Number of RSTP per utterance is in average 3.09 (std=3.24, median=3.0, min=1, max=582)

Automatic Evaluation (1/3)

Goal

Comparison of selection methods on a task of response selection

Reference utterance: "**can I help you?**"

Method	Response
Random	a had accomplices.
TF-IDF	we'll get her anyway
trigram	we'll get her anyway
doc2vec	what are you doing?
RSTP	yeah

Automatic Evaluation (2/3)

Protocol (comparison of response selection models)

- Set of 1000 reference utterances along with a predefined set of acceptable responses
- Evaluation of responses produced by each model via the "Translation Error Rate" (TER)

$$\text{TER} = \frac{\#\text{ of edits}}{\text{average }\#\text{ of reference words}}$$

Reference utterance: "**can I help you?**"

Method	Response	TER Score
Random	a had accomplices.	0.15
TF-IDF	we'll get her anyway	0.44
trigram	we'll get her anyway	0.44
doc2vec	what are you doing?	0.29
RSTP	yeah	0.0

Automatic Evaluation (3/3)

	Random	TF-IDF	Trigram	doc2vec	RSTP
TER	0.632	0.537	0.549	0.566	**0.505**

	Random	TF-IDF	Trigram	doc2vec	RSTP
Random	–	***	**	*	***
TF-IDF	***	–	ns	ns	ns
Trigram	**	ns	–	ns	*
doc2vec	*	ns	ns	–	**
RSTP	***	ns	*	**	–

Paired Wilcoxon test; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; ns: not significant

Application: Chatbot Dialogue System

International Shared Task

- Participation to an international shared task
  - **collecting** chat-oriented dialogue data that can be made available for research purposes
  - developing a framework for the **automatic evaluation** of chat-oriented dialogue
- URL: `http://workshop.colips.org/wochat/`
- Data: `http://workshop.colips.org/wochat/data`

H : would you like some wine?

Joker : no

H : beer maybe?

Joker : it's hard to get drunk on beer

Joker : i 'm Kurt Cobain .

H : come as you are

Joker : as you were

H : as I want you to be

Joker : as a friend

H : as a friend

Joker : as an old enemy

Selection-based Approach to Dialogue Management

- Recurrent Surface Text Pattern (RSTP) model
  - Extraction of RSTP from a corpus of written dialogue utterances
  - Representation of utterances as bag-of-RSTPs
  - Features: corpus-based, unsupervised, parameterless, exploiting regularities understandable from a human perspective.
- Retrieval Model: GVSM where terms are RSTPs
- Automatic Evaluation: based on the "Translation Error Rate" (TER)
- Application: open-domain response selection embeddable in conversational agent (fallback strategy, chatbot)
- Perspectives
  - Subjective evaluation
  - Taking into account dialogue history during selection
  - Trying other open-domain dialogue corpora
  - Toward automatic evaluation metrics ((RE-)WOCHAT workshops)

Outline

H2020 European Project: ARIA VALUSPA

URL: `http://aria-agent.eu/`

Main Features

- Virtual agent
- Multimodal interaction (verbal/non-verbal behaviour)
- **Adaptation**
  - Unexpected situation
  - **Socio-emotional state of the user**
- Constraints: face-to-face, real-time

Convergence and Verbal Alignment (1/2)

Convergence and Interactive Alignment

- Communication Accommodation Theory [Gallois et al., 2005]
- Interactive Alignment Theory [Pickering and Garrod, 2004]

Loc.	Utterance
S_1	hi i'm sam , nice to meet you what is your name ?
H_2	alex
S_3	how are you doing ?
H_4	i am great
S_5	i really appreciate going fifty fifty with you on clearing out this locker.

Table 1 : Corpus H-A 311 neg1

Loc.	Utterance
H_1	hi
S_2	hi i'm sam , nice to meet you
H_3	nice to meet you i'm erica
S_4	how are you doing ?
H_5	i'm doing good how are you
S_6	pretty good
H_7	good
S_8	i really appreciate going fifty fifty with you on clearing out this locker.

Table 2 : Corpus H-A 376 neg1

Convergence and Verbal Alignment (2/2)

Loc.	Utterance
	[...]
S_1	deal
H_2	deal
S_3	thank you
H_4	thank you
S_5	nice doing business with you
H_6	it's a pleasure
S_7	until next time
H_8	have a good day
S_9	goodbye
H_{10}	bye

Table 3 : Corpus H-A 302 neg1

Loc.	Utterance
	[...]
S_1	deal
H_2	deal
S_3	thank you
H_4	thank you
S_5	it's a pleasure doing business with you
H_6	it's a pleasure doing business with you too
S_7	goodbye
H_8	goodbye

Table 4 : Corpus H-A 352 neg1

Why studying verbal alignment?

Lessons from H-H interaction

- Subconscious phenomenon that naturally occurs in H-H dialogues [Pickering and Garrod, 2004]
  - Speakers reuse lexical as well as syntactic structures from previous utterances [Reitter et al., 2006, Ward and Litman, 2007]
- Facilitates successful task-oriented conversations [Nenkova et al., 2008, Friedberg et al., 2012]

. . . and what about H-M interaction?

- Linguistic alignment occurs: users adopt lexical items and syntactic structures used by a system [Brennan and Clark, 1996, Stoyanchev and Stent, 2009, Parent and Eskenazi, 2010, Branigan et al., 2010]
- . . . but it is only one-way!

Research Direction

Goal

Provide a virtual agent with the ability to

- **detect the alignment behaviour** of its human interlocutor
- **align (or not) with the user**

Motivation

- Natural source of variation in dialogue
- Taking into account the socio-emotional behaviour of the user ("social glue")
- Adaptation without the need of extensive user profiling

Outcomes

- Enhancing agent's believability, likeability and friendliness
- Increasing interaction naturalness
- Maintaining and fostering user's engagement [Clavel et al., 2016]
- Improving collaboration in task-oriented dialogue

Proposition

Approach

Providing measures characterising verbal alignment processes based on the transcript of dialogue

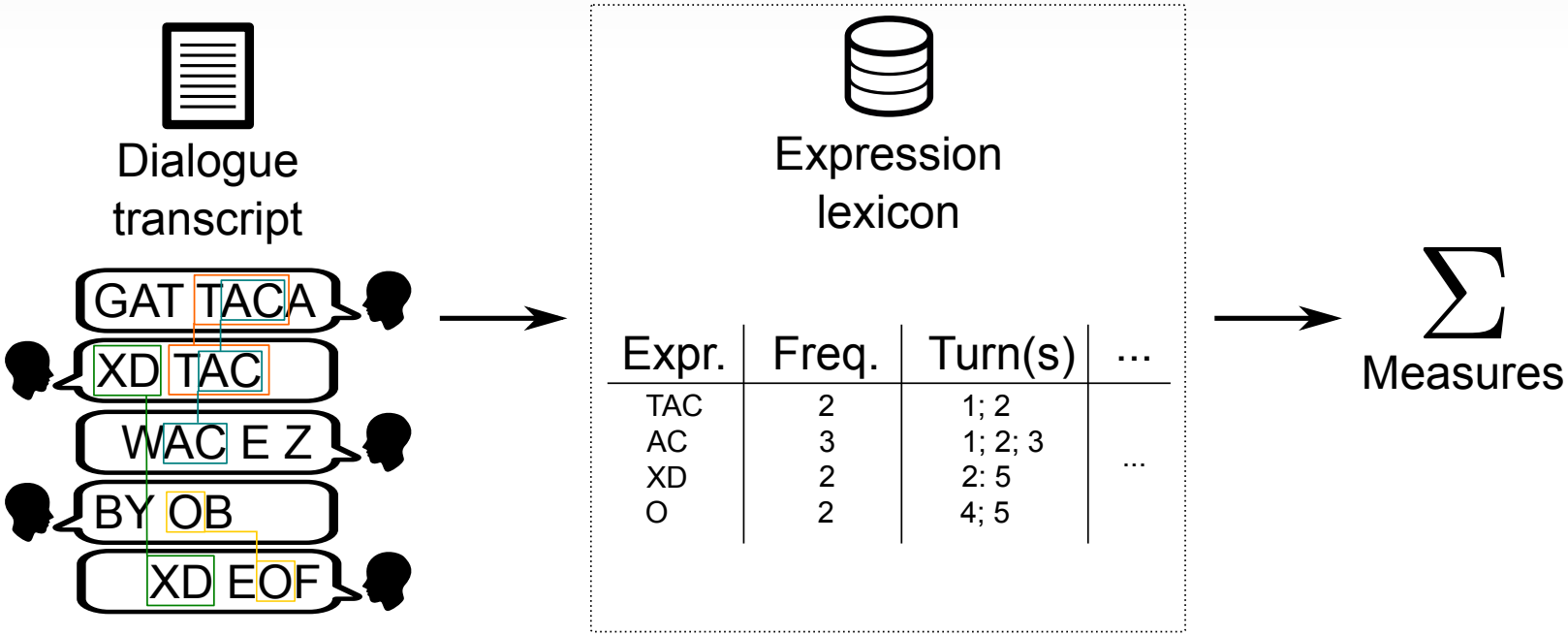


Figure 4 : Proposed framework

Dubuisson Duplessis, G.; Clavel, C.; Landragin, F., **Automatic Measures to Characterise Verbal Alignment in Human-Agent Interaction**, 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 11p. (to appear), August 2017

Expression Lexicon

Expression

A surface text pattern at the utterance level that has been produced by both speakers in a dialogue

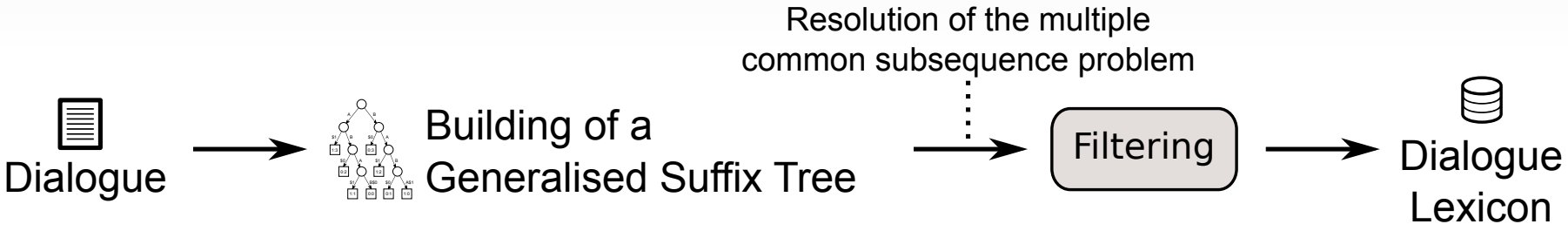


Figure 5 : Main steps to build the dialogue lexicon

A_1	well, that's an interesting idea. but no, that's not gonna work for me.
B_2	what will *work for* you?
A_3	*what* do *you* think about me getting two chairs and one plate and *you* getting one chair, one plate, and the clock?
B_4	*that's not gonna work for me*

Table 5 : H-A 329 neg2

Expr.	Freq.	Init.	...
that's not gonna work for me	2	A	...
work for	3	A	...
me	3	A	...
what	2	B	...
you	2	B	...

Table 6 : Expression Lexicon

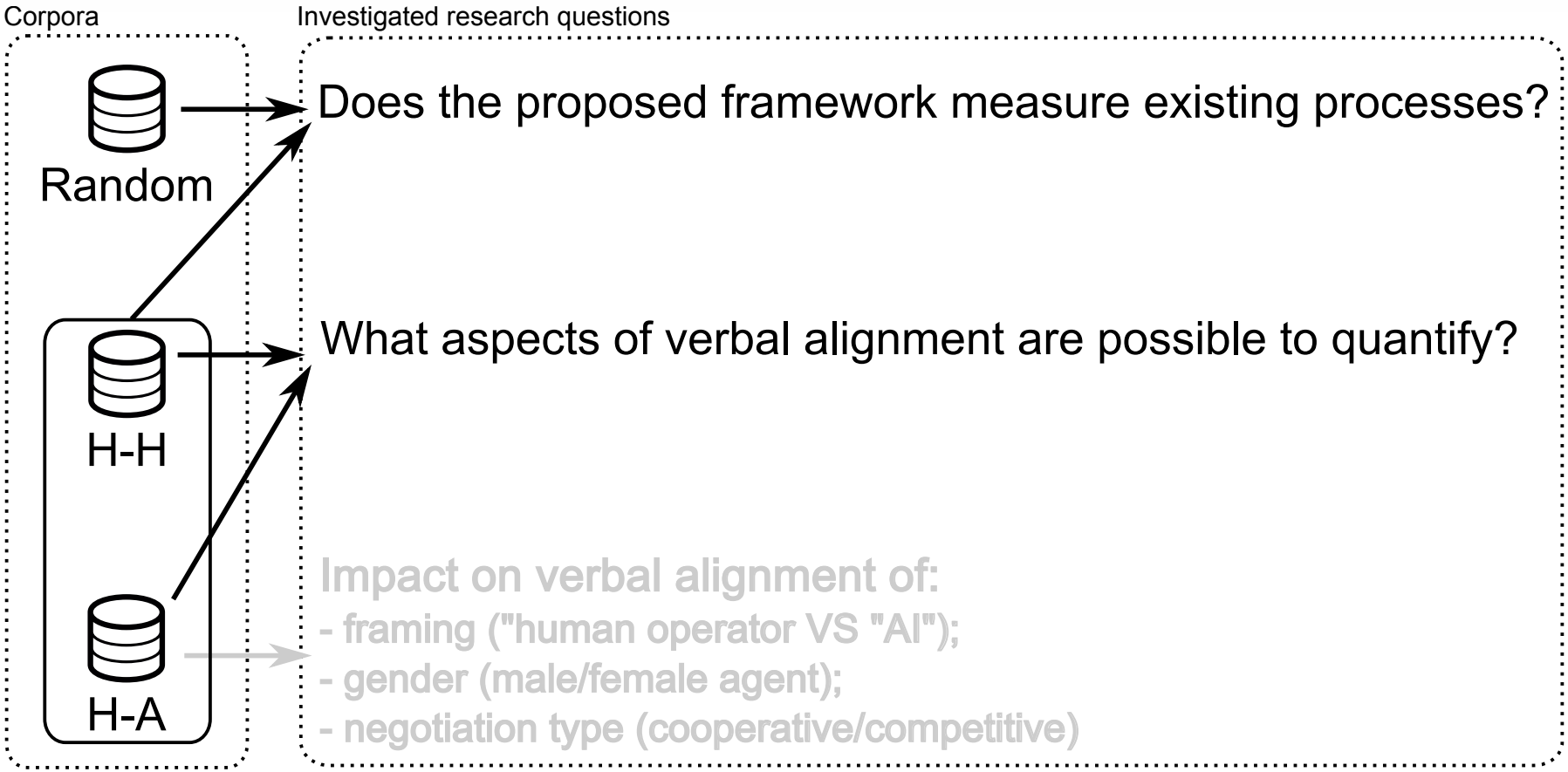
Measures Derived from the Expression Lexicon

Expr. Lexicon Size	Number of items in the expression lexicon (ELS)
Expr. Variety	$\mathrm{EV} = \frac{\mathrm{ELS}}{\#\mathrm{Tokens}}$
Expr. Repetition (S)	$\mathrm{ER}_S = \frac{\#\text{ Tokens from S in an established expr.}}{\#\text{ Tokens from S}}$ $\forall\ \mathrm{S},\ \mathrm{ER}_S \in [0, 1]$
Initiated Expr. (S)	$\mathrm{IE}_S = \frac{\#\text{ Expr. initiated by S}}{\mathrm{ELS}}$ $\forall\ \mathrm{S},\ \mathrm{IE}_S \in [0, 1]$

Experimentation Protocol

Protocol

Corpus-based study to assess the proposed framework and measures



Negotiation Corpora (1/2)

Figure 6 : H-A (Woz) Settings [DeVault et al., 2015, Gratch et al., 2016]

	H-H	H-A (Woz)
Dialogue	**84**	**154**
Utterance (unique) . . . avg (std)	10319 (7840) 122.8 (84.1)	17125 (6109) 111.2 (57.5)
Token (unique)	**79396** (2516)	**90479** (1335)
Tokens/Utterance avg/median (std) avg (std) min/max	 7.7/6.0 (7.4) 7.7 (7.4) 1/66	 5.3/4.0 (5.7) 5.3 (5.7) 1/154

Negotiation Corpora (2/2)

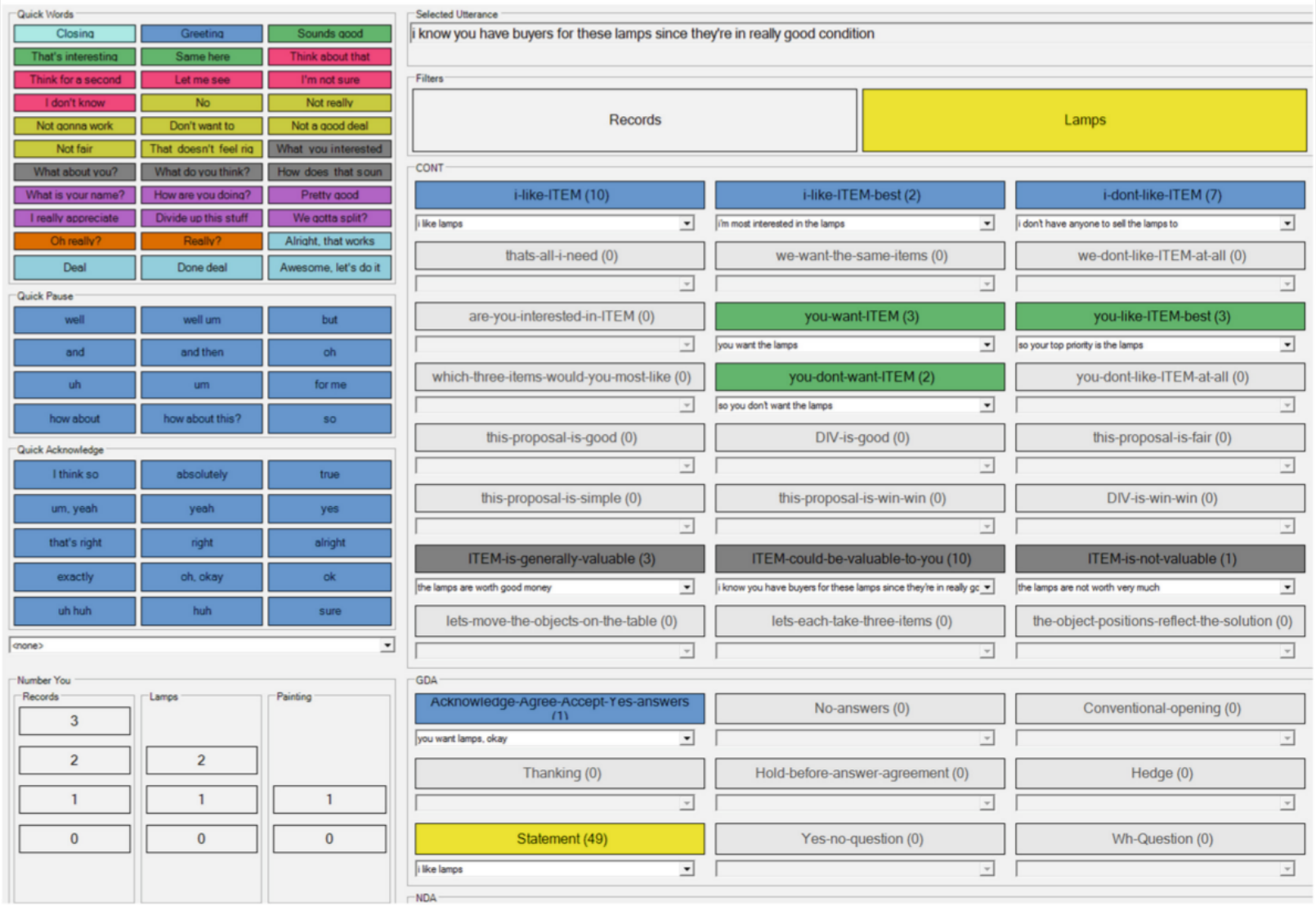


Figure 7 : The Woz system [DeVault et al., 2015]

Surrogate Corpora

Loc.	Real Utterance	Randomised Utterance
S_1	[...] deal	[...] deal
H_2	deal	okay well then i have a buyer for both for the albums how's that
S_3	thank you	thank you
H_4	thank you	okay great
S_5	it's a pleasure doing business with you	it's a pleasure doing business with you
H_6	it's a pleasure doing business with you too	we sure do
S_7	goodbye	goodbye
H_8	goodbye	well no do you

Table 7 : Corpus H-A 352 neg1 and one of its randomised version

Results: H-H/A VS Surrogate Corpora

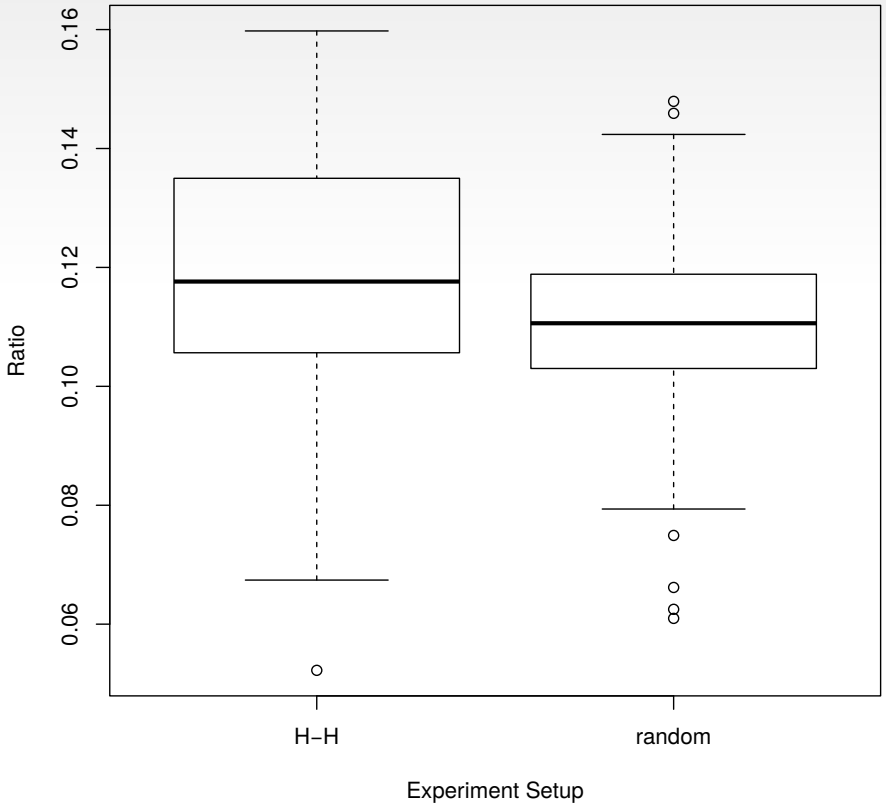


Figure 8 : H-H VS random. Expression Variety (EV). Difference is significant ($p < 0.001$).

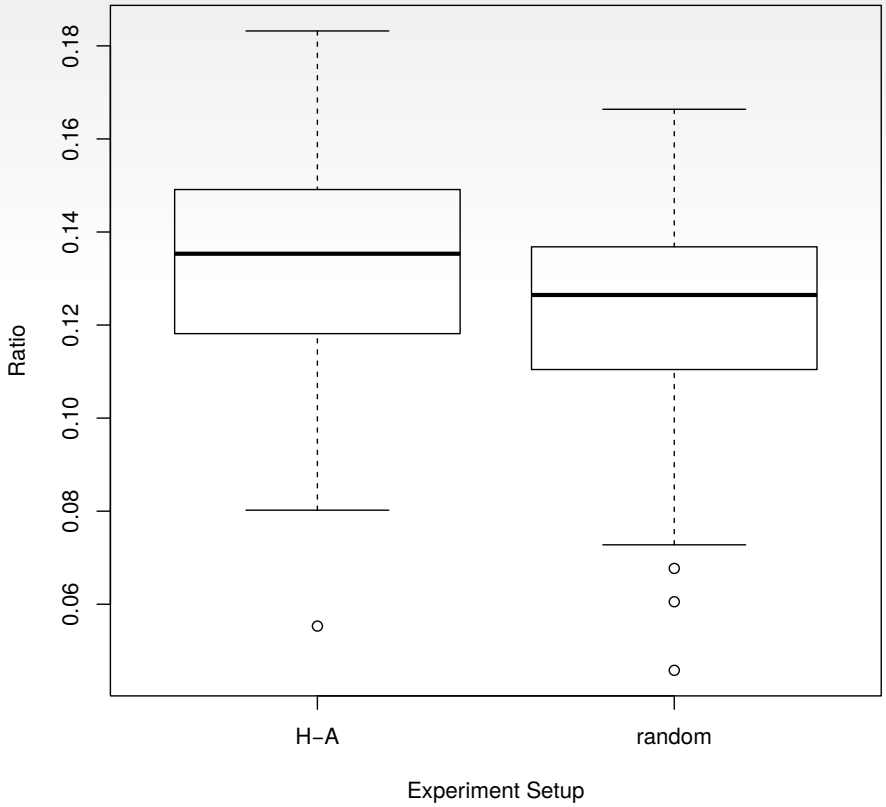


Figure 9 : H-A VS random. Expression Variety (EV). Difference is significant ($p < 0.001$).

Results: H-H VS H-A Corpora

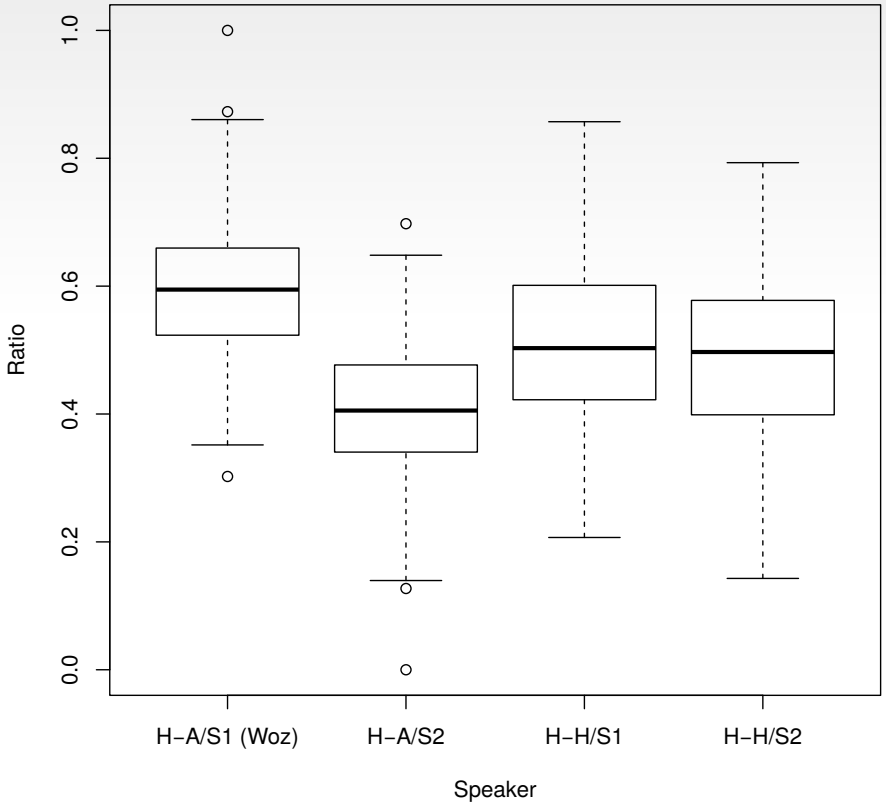


Figure 10 : Initiated Expressions (IE_S). Difference is significant for H-A ($p < 0.001$), not significant for H-H.

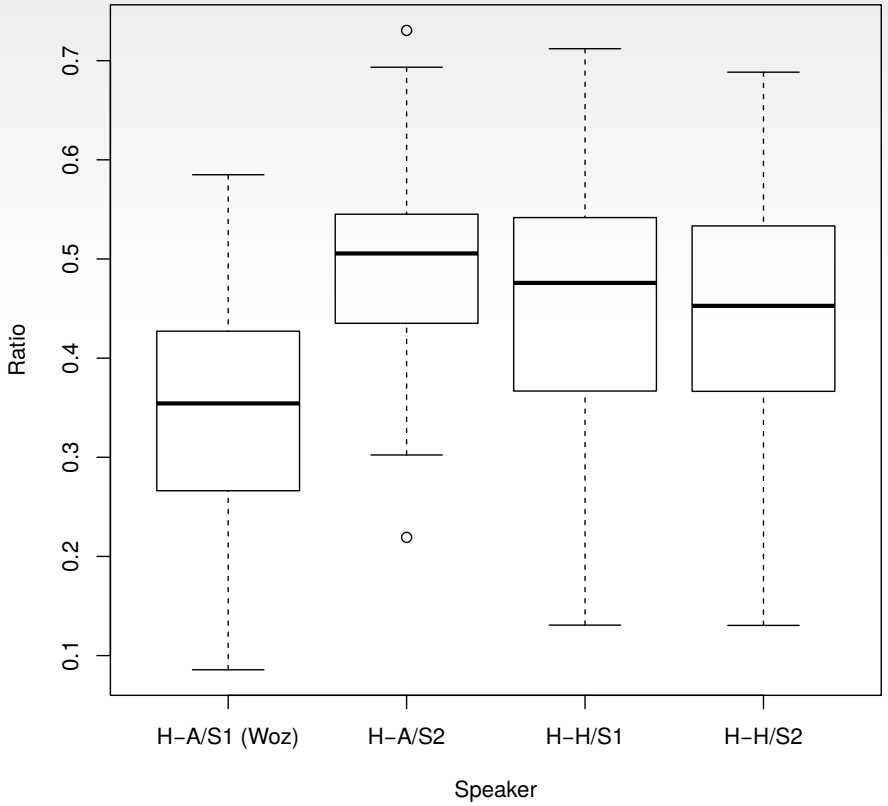


Figure 11 : Expression Repetition (ER). Difference is significant for H-A ($p < 0.001$), not significant for H-H.

Contributions

- Automatic and generic measures of verbal alignment based on the level of surface of text utterances characterising:
  - the routinization process;
  - the degree of repetition between dialogue participants;
  - the orientation of verbal alignment.
- Contrasting H-H and H-A verbal alignment (symmetry VS asymmetry)
  - Quantitative confirmation of predictions from previous literature regarding the strength and orientation of verbal alignment in Human-Machine Interaction [Branigan et al., 2010]
- Measures are based on efficient algorithms ($\Rightarrow$ online usage in a dialogue system)

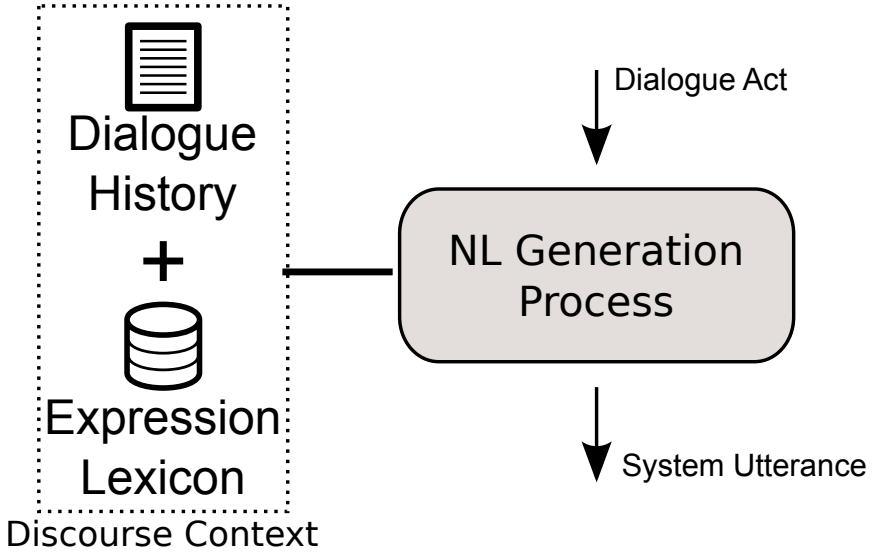
Perspectives: NLG and Evaluation

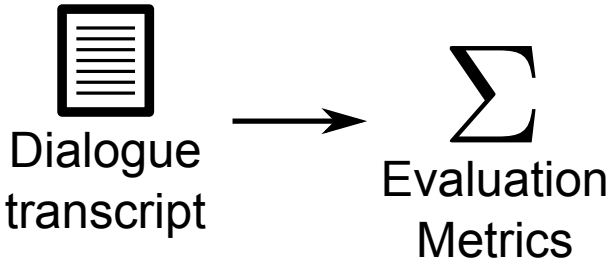
Verbal Alignment Strategy

Enabling verbal alignment in the NLG model of the agent

Automatic Evaluation

Studying the contribution of verbal alignment metrics to automatic evaluation procedures





Outline

Take Home Message

modelling and simulating human behaviour and language use

- observation, data collection, data analysis
- modelling, designing, implementing interaction models with and without an explicit task
- evaluation of interaction systems

Main research domains: artificial intelligence, human-machine interaction, dialogue, natural language processing

Dialogue Strategy Modelling for Human-Agent Interaction

Guillaume Dubuisson Duplessis – gdubuisson@telecom-paristech.fr
http://www.dubuissonduplessis.fr

G. Dubuisson Duplessis, V. Letard, A-L. Ligozat, S. Rosset.

Purely Corpus-based Automatic Conversation Authoring.

In proceedings of the 10th International Conference on Language Resources and Evaluation (LREC), 8p., May 2016.

G. Dubuisson Duplessis, F. Charras, V. Letard, A-L. Ligozat, S. Rosset.

Utterance Retrieval based on Recurrent Surface Text Patterns.

In proceedings of the 39th European Conference on Information Retrieval (ECIR), pp. 199–211, April 2017.

G. Dubuisson Duplessis, C. Clavel, F. Landragin.

Automatic Measures to Characterise Verbal Alignment in Human-Agent Interaction.

In proceedings of the 18th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), 11p., August 2017 (to appear).

Références I

Allwood, J. (1994).
Obligations and options in dialogue.
Think Quarterly, 3:9–18.

Ameixa, D., Coheur, L., Fialho, P., and Quaresma, P. (2014).
Luke, I am your father: dealing with out-of-domain requests by using movies subtitles.
In *Intelligent Virtual Agents*, pages 13–21. Springer.

Banchs, R. E. and Li, H. (2012).
IRIS: a chat-oriented dialogue system based on the vector space model.
In *Proceedings of the ACL 2012 System Demonstrations*, pages 37–42. Association for Computational Linguistics.

Branigan, H. P., Pickering, M. J., Pearson, J., and McLean, J. F. (2010).
Linguistic alignment between people and computers.
Journal of Pragmatics, 42(9):2355–2368.

Brennan, S. E. and Clark, H. H. (1996).
Conceptual pacts and lexical choice in conversation.
Journal of Experimental Psychology: Learning, Memory, and Cognition, 22(6):1482.

Clark, H. (1996).
Using language, volume 4.
Cambridge University Press.

Références II

Clavel, C., Cafaro, A., Campano, S., and Pelachaud, C. (2016).
Fostering user engagement in face-to-face human-agent interactions: a survey.
In *Toward Robotic Socially Believable Behaving Systems-Volume II*, pages 93–120. Springer.

DeVault, D., Mell, J., and Gratch, J. (2015).
Toward natural turn-taking in a virtual human negotiation agent.
In *AAAI Spring Symposium on Turn-taking and Coordination in Human-Machine Interaction. AAAI Press, Stanford, CA.*

D'Haro, L., Abu Shawar, B., and Yu, Z. (2016).
Shared task on data collection and annotation, re-wochat 2016–shared task description report.
In *Proceedings of the Workshop on Collecting and Generating Resources for Chatbots and Conversational Agents-Development and Evaluation Workshop Programme (RE-WOCHAT), LREC.*

Friedberg, H., Litman, D., and Paletz, S. B. (2012).
Lexical entrainment and success in student engineering groups.
In *Spoken Language Technology Workshop (SLT)*, pages 404–409. IEEE.

Gallois, C., Ogay, T., and Giles, Howard, H. (2005).
Communication accommodation theory: A look back and a look ahead.
W. Gudykunst (red.): Theorizing about intercultural communication. Thousand Oaks, CA: Sage, pages 121–148.

Gandhe, S. and Traum, D. R. (2007).
Creating spoken dialogue characters from corpora without annotations.
In *INTERSPEECH*, pages 2201–2204.

Références III

Gandhe, S. and Traum, D. R. (2013).
Surface text based dialogue models for virtual humans.
In *Proceedings of the SIGDIAL*.

Gratch, J., DeVault, D., and Lucas, G. (2016).
The benefits of virtual humans for teaching negotiation.
In *International Conference on Intelligent Virtual Agents (IVA)*, pages 283–294. Springer.

Gusfield, D. (1997).
Algorithms on Strings, Trees and Sequences.
Cambridge University Press, Cambridge, UK.

Lee, C., Jung, S., Kim, S., and Lee, G. G. (2009).
Example-based dialog modeling for practical multi-domain dialog system.
Speech Communication, 51(5):466–484.

Lison, P. and Tiedemann, J. (2016).
OpenSubtitles2016: Extracting Large Parallel Corpora from Movie and TV Subtitles.
In *10th edition of the Language Resources and Evaluation Conference (LREC)*, Portorož, Slovenia.

Nenkova, A., Gravano, A., and Hirschberg, J. (2008).
High frequency word entrainment in spoken dialogue.
In *Proceedings of the 46th annual meeting of the association for computational linguistics on human language technologies (ACL-HLT): Short papers*, pages 169–172. Association for Computational Linguistics.

Références IV

Nio, L., Sakti, S., Neubig, G., Toda, T., Adriani, M., and Nakamura, S. (2014).
Developing non-goal dialog system based on examples of drama television.
In *Natural Interaction with Robots, Knowbots and Smartphones*, pages 355–361. Springer.

Parent, G. and Eskenazi, M. (2010).
Lexical entrainment of real users in the let's go spoken dialog system.
In *INTERSPEECH*, pages 3018–3021.

Pickering, M. J. and Garrod, S. (2004).
Toward a mechanistic psychology of dialogue.
Behavioral and brain sciences, 27(02):169–190.

Reitter, D., Keller, F., and Moore, J. D. (2006).
Computational modelling of structural priming in dialogue.
In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (NAACL-HLT): Short Papers*, pages 121–124. Association for Computational Linguistics.

Stoyanchev, S. and Stent, A. (2009).
Lexical and syntactic priming and their impact in deployed spoken dialog systems.
In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (NAACL-HLT): Short Papers*, pages 189–192. Association for Computational Linguistics.

Références V

Ward, A. and Litman, D. J. (2007).
Automatically measuring lexical and acoustic/prosodic convergence in tutorial dialog corpora.
In *Speech and Language Technology in Education (SLaTE2007)*, pages 57–60.