

# Towards Choosing Better Primes for Spoken Dialog Systems

José Lopes<sup>1,2</sup>, Maxine Eskenazi<sup>3</sup>, Isabel Trancoso<sup>1,2</sup>

<sup>1</sup>INESC-ID Lisboa, Portugal

<sup>2</sup>Instituto Superior Técnico, Lisboa, Portugal

<sup>3</sup>Language Technologies Institute, Carnegie Mellon University, Pittsburgh, PA, USA

jose.david.lopes@l2f.inesc-id.pt

**Abstract**—When humans and computers use the same terms (primes, when they entrain to one another), spoken dialogs proceed more smoothly. The goal of this paper is to describe initial steps we have found that will enable us to eventually automatically choose better primes in spoken dialog system prompts. Two different sets of prompts were used to understand what makes one prime more suitable than another. The impact of the primes chosen in speech recognition was evaluated. In addition, results reveal that users did adopt the new vocabulary introduced in the new system prompts. As a result of this, performance of the system improved, providing clues for the trade off needed when choosing between adequate primes in prompts and speech recognition performance.

## I. INTRODUCTION

In human-machine communication, the demand for easier, more flexible interfaces promotes new challenges to the scientific community. And there are many open questions.

In order to enhance the quality of human-computer dialogs, researchers can learn from human-human dialogs. Brennan and Clark [1] argue that people have different strategies according to how confident they are that they will be correctly understood by others. In human-computer dialogs, generally speakers' confidence that the spoken dialog system (SDS) will understand them is lower than their confidence that another human will. Also in human-human dialogs, when the expectation of being understood is lower (eg. non-native listeners) people tend to use simpler words or words that have been mentioned earlier in the conversation in order to increase their confidence that they will be understood. This lexical adaptation is generally known in the literature as lexical entrainment [2].

Some speech researchers tried to reduce the confusability of the vocabulary in command language speech recognizers. Roe et al [3] created a metric to do so and compared their predictions about which words would have a higher degree of confusability and the actual outcome of the speech recognition.

More recently spoken dialog system researchers tried to explore the concept of entrainment and extend it to acoustic, prosodic and syntactic levels. Ward et al [4] show that human users of spoken dialog systems have a wide variety of convergent speech behavior in their choice of both lexical items and acoustic/prosodic elements [4]. As in [4], we will employ the concept of prime defined by Reitter et al [5] and others. They define a prime as a linguistic choice made by the

recipient that was influenced by the system. It can be either a grammatical structure or a lexical choice.

Later studies have focused on real users to study the effect entrainment can have in their utterances. Stoyanchev et al [6] studied lexical and syntactic priming in Let's Go! [7]. Using a small dataset, they stated that both lexical and syntactic choices and concept identification rates are influenced by the system choice of primes. Our work goes beyond the conclusions on verb usage in [6] by using a wider variety of primes, thus significantly modifying the prompts produced by the system. Parent et al [8] studied entrainment in real users of Let's Go! and found the users entrained more to some primes than others. The work concluded that more frequent words are more likely to be adopted by real users. We extend this work in effort toward automatic choice of primes by establishing a tradeoff between the primes chosen for the system and what we need to improve ASR performance.

To sum up this paper joins the best of both worlds. It can be seen as a loop where we are trying to remove the words that cause problems in system performance, either because they are likely to be confused with other words in the vocabulary or because they seem awkward to the users. Those primes are replaced by synonyms in the system prompts. Finally, the system will run with the new primes and the loop can start over from the beginning with observance of the user uptake. If a prime is quickly adopted and has good ASR rates it could be considered as a good prime. The approach used here was completely manual with the intention that, once we identify what makes a good candidate for prime, this choice will be automated.

In this study we use Noctívago, a dialog system similar to Lets Go!, a bus schedule information system operating in Pittsburgh, answering calls after human operators leave in the evening. The language of Noctívago is European Portuguese (EP) and the target is schedule information for night and weekend buses in the city of Lisbon.

The paper will start by describing the system architecture (section II). Then, it will describe a pilot study and its results (section III) that paved the way for the studies described in this paper (section IV). Results of the latter will be presented and discussed in section V, before the conclusion (section VI) and future work (section VII).

## II. SYSTEM DESCRIPTION

Our system was built using the Olympus publicly available framework for spoken dialog systems [9]. Since this framework was developed for English, porting it to European Portuguese required some changes. PocketSphinx was replaced by Audimus [10], our in-house ASR module, based on hybrid models. A version for SAPI of DIXI [11], a concatenative unit selection synthesizer based on the Festival model was used. The language understanding module, Phoenix, was not modified in its core as the understanding is based on the grammar. A domain-specific grammar was created for Portuguese. The module that converts time expressions was also modified. The dialog manager was developed based on RavenClaw. The back-end was modified to retrieve information about landmarks and give correct information of bus lines which run with different frequencies during weekends. New templates with Portuguese prompts were created for the language generation module. The other modules of the framework were not submitted to any modification.

## III. PILOT EXPERIMENTS

Early tests with Noctívago were carried out with people who volunteered to test the system under tightly controlled conditions, using a landline and a very limited set of scenarios. The users had to ask for specific bus schedules. In these early trials the system received 56 calls, which correspond to a total of 742 turns and 1019 words.

There were 143 different words in the system prompts, but not all of them could be considered as candidates for primes. Ward and Litman [12] argued that any word with a potential synonym in the system prompts can be used as a potential prime. Those that do not have an alternative synonym would not be candidates. For example, time expressions containing numerical data could not be used as primes. By looking at the past error rate for each prime candidate, we were able to find which primes were less likely to be correctly recognized by the ASR module and find synonyms for those. The prime error rate (PER) was computed as we wanted to evaluate recognition at the prime level and not at word level, as most of the primes are composed of more than a single word. Results for PER and frequency are shown in table I.

The primes for **next bus** (*próximo*, *próximo autocarro*), **new query** (*outro percurso*, *nova pesquisa*) and **now** (*agora*) were chosen for replacement. Besides the relatively high word error rate for each of them, another reason for modifying

Prime	PER (%)	Frequency
<i>próximo</i> (next)	19.0	21
<i>próximo autocarro</i> (next bus)	26.3	19
<i>anterior</i> (previous)	18.2	22
<i>autocarro anterior</i> (previous bus)	18.8	16
<i>outro percurso</i> (another route)	50.0	6
<i>nova pesquisa</i> (new request)	36.4	11
<i>agora</i> (now)	60.6	33
<i>não</i> (no)	88.6	35
<i>sim</i> (yes)	70.4	27

TABLE I  
PRIME ANALYSIS IN PILOT EXPERIMENTS.

these primes was the feedback from the users. Some of them reported that the synthesizer did not pronounce some words correctly. For instance, the word *pesquisa*, pronounced /p@Skiz6/, was sometimes pronounced as /piSkiz6/, depending on the context. Some users used the synthesizer erroneous pronunciation instead of the standard pronunciation that was in the lexicon of the speech recognizer. The new set of primes is shown on Table II.

Concept	Old prime	New prime
Now	<i>agora</i>	<i>imediatamente neste momento o mais brevemente possível o mais rápido possível</i>
New query	<i>nova pesquisa outro percurso</i>	<i>procurar novamente nova procura outra procura nova busca</i>
Next bus	<i>próximo</i>	<i>seguinte</i>

TABLE II  
LIST OF PRIMES TO BE ENTRAINED.

## IV. EXPERIMENTAL SET UP

During two weeks paid users participated on the experiments. The primes changed from the first to the second week as detailed in table II without informing the users. They were asked to call the system in both weeks completing the three different scenarios in each week. The scenarios involved asking bus schedule information for night buses in Lisbon. At the end of the scenarios they had to fill in a small form where they were supposed to say if they had received the correct information in each scenario and give their opinion about the system. They should also provide information about the device they used to complete the test. During the second week, users who called the system in both weeks compared the performance of the system in different weeks.

## V. RESULTS

During the first week the system received 45 calls corresponding to 131 valid sessions. In the second week, the number of callers decreased to 36 callers, corresponding to 125 valid sessions. 19 people called the system both weeks, 26 of them only called the system the first week, and 17 only called the system the second week.

In this evaluation, a session was considered successful if the system provided the user with the information he/she had asked for. All of the sessions were carefully analyzed at the utterance level, describing the type of error found in the incorrectly recognized utterances.

### A. System performance

Table III shows the number of calls for each week and the dialog success rate, according to the device used. The table does not show a correlation between the device and system performance despite the fact that the acoustic models used were developed for telephone speech. What the table does show is that dialog success rate increased significantly from the first to the second week, except for mobile phones. The number of calls in the first week using cellphones is not

significant enough to draw any conclusions about cellphone speech degrading the system performance.

	# Calls W1	# Calls W2	DSR W1 (%)	DSR W2 (%)
Landline	86	52	48.8	57.6
Cellphone	12	37	66.7	54.1
VoIP	24	27	54.1	59.3
Unknown	9	9	11.1	33.3
TOTAL	131	125	48.5	54.8

TABLE III  
DISTRIBUTION OF CALLS AND DIALOG SUCCESS RATE (DSR) ACCORDING TO THE DEVICE USED.

Table IV shows further details about the success rate of the system in both weeks. The evaluation measures used in the Spoken Dialog Challenge [13] enable us to compare the performance of our system with the systems that were run in the SDC on a similar task.

The standard defines two possible types of calls: 'No output' and 'Any output'. 'No output' means that the dialog did not provide any information to the user. 'Any output' means that it provided some sort of information to the user. The calls classified as 'Any output' are further divided in 'Acceptable output' calls and 'Incorrect output' calls. 'Acceptable output' when the correct information was provided. 'Incorrect output' when it did not provide the correct information.

	Week 1	Week 2
No output	31.3%	19.2%
Any output	68.7%	80.8%
Acceptable output	71.1%	68.3%
Incorrect output	28.9%	31.7%

TABLE IV  
SUCCESS RATE OF THE SYSTEM IN EACH WEEK.

Table IV shows a significant improvement from week one to week two in the percentage of calls which gave any sort of output to the user. Although the percentage of successful dialogs increased in the second week, the percentage of the acceptable outputs shows a small decrease. This may be related to the fact that in one of the scenarios the users had to ask for a bus for a specific weekday and departure time (eg.: "Saturday at 1:15 am"). This scenario was completed by 36 times in the first week, and 45 times in the second week. Users made a significant pause between uttering the weekday and the time of day. The ASR module would output the day of the week immediately, and the language understanding module would bind the day of the week to the travel time concept, setting a default departure time. Later the default inter word pause was modified to solve this problem.

During the detailed analysis of the sessions a few sources of errors were identified. The ASR was one of them, but other errors were caused either by the voice activity detector (VAD) or the parser for date and time expressions (Datetime). In some sessions we found that the telephone channel was left open a few seconds after the user abandoned the call and consequently all the audio was captured and held in the recognizing queue until the next call came up (Hung up, HU). Finally, some errors were caused by the loss of synchronism between the recognizer and the remainder modules of the SDS (Lost of

	Week 1 (%)	Week 2 (%)
Correct	32.3	41.1
ASR	52.8	44.2
VAD	4.0	4.0
HU	3.8	6.7
LS	6.3	6.4
Datetime	5.6	4.5

TABLE V  
ANALYSIS OF ERRORS AT THE TURN LEVEL.

synchronism, LS). The percentage of correctly perceived turns and percentage of turns wrongly recognized are presented in table V. The percentage of turns that were not correctly recognized is divided according to the source of the error.

The most important element to retain here is that the percentage of correctly recognized turns was 10% higher in the second week. The other sources of errors were not eliminated since they do not depend on lexical choice. The most recent version of the SDS eliminates the HU and LS source of errors.

### B. ASR results

The information available on the logs allowed us to compute the WER for the live tests. Later the off-line version of Audimus was applied to each session with the same language and acoustic models used in the on-line system. Since we only wanted to test the ASR module, the information of the time where to search for the word was given to the speech recognizer in the off-line tests. Results in Table VI show WER, deletions and insertions for the two weeks.

These results again confirm that the system performed better in the second week, at least in the on-line test. A 6% decrease in WER without changing anything other than the prompts is quite remarkable.

The difference found in the off-line recognition test was not significant, considering WER in the different weeks. Later it was found that the device used to call the system which had no major influence on the dialog success rate (as shown in Table III), had impact the off-line WER. The acoustic models used were trained mainly with landline telephone speech, down sampled broadcast news speech and very few examples of cellphone speech. VoIP speech presents a different encoding scheme where the bit rate is automatically adjusted to the bandwidth of the channel. When compared with the WER for landline telephone users, the WER for cellphone speech was about 8% higher, and for VoIP speech 14-17% higher. The number of sessions held with landline telephone speech decreased from 86 to 57 in the second week. This can help explain why the WER did not improve in the second week.

The results of the off-line test also suggest that the on-line VAD and ASR modules may need further tuning, since the on

	on-line		off-line	
	Week 1	Week 2	Week 1	Week 2
WER	58.3%	52.3%	30.0%	30.6%
insertions	99	177	137	192
deletions	876	638	15	131
substitutions	553	528	520	463

TABLE VI  
WER FOR THE DIFFERENT WEEKS.

line system has far more deletions.

### C. Prime usage analysis

This subsection analyzes the user behavior concerning the use of the primes described in table II. Table VII presents the frequency of the primes in the data in both weeks and PER for each prime in both weeks.

The results presented in Table VII show that the callers entrained to the system, incorporating the system primes in their speech. The only word which appears with similar frequency in both weeks is *agora*. This is a very frequent word in Portuguese and in addition the word was explicitly written in one of the scenarios, thus biasing callers towards that word.

Words like *agora* (especially in week 2) and *seguinte* have very significant error rates. One possible source is the corpus generated to train the language model. The algorithm used to generate the corpus tried to incorporate the more variety as possible for each concept described on the grammar. Thus, the more possibilities a concept has, the less frequently each one will appear in the corpus. This problem can be solved by modifying the corpus generation algorithm.

Another interesting result is obtained if we sum the results for all of primes used for **now** and **new query**. In both cases the performance of the new set of primes is better than that of the prime used in the first week. Some primes did contribute to this more than others. The phonetic distance seems to be a good metric to choose prompts with the same meaning considering only the ASR module performance. For instance, when the prime used was *mais rápido possível*, where the word *rápido* sounds very similar to *Rato*, a bus stop covered in our system, the ASR module never recognized it correctly. The same happened with *seguinte* for the **next bus** prime. This latter was recognized several times as *sim* or *sair* (quit).

Looking at the last column of table VII we have the Part of Speech (POS) of the primes. No evident correlation can be seen between our data and the entrainment that users showed to the prime.

In order to better understand why users are more likely to adopt some primes than others our data was analyzed in order to find which primes were uptaken immediately after being used in the system prompts. This conclusion comes from [4], [8] who concluded that primes are more likely to be used immediately after the system uses them. In addition, we wanted to understand whether a system prime has the most influences on the user's lexical choice.

Therefore, for every prime the two previous utterances were taken into account. Four different features were computed to analyze the users' entrainment to the proposed primes. The 'uptake' column shows the number of times that each prime was used after being invoked in the system's previous utterance. The 'no uptake' column corresponds to number of times that prime was used without being used by the system. The number between brackets represents the number of times that the prime was uptaken from a previous prime usage in the same session. The '(%) of usage' stands for the percentage of

user utterances that adopted that prime when it was present in the system prompts. Finally, the '# prompts' column shows the total number of system prompts in the corpus where that prime appears. Table VIII has an example of how these statistics were computed from our data.

The results are presented in table IX. We notice that in most cases the exposure to the prime is enormous compared to the usage. In most cases, the prime is included on the system prompt but it is not the only answer available. For instance, the prompts for the **now** concept are used in the context of asking for the time the user wants to travel. One can answer with a time-relative expression such as **now** or a specific hour. Again, the fact that the word *agora* was explicitly written in one of the scenarios is reflected in the results since the caller did not use it from the previous system prompt in most cases.

During the first week the differences in the (%) of usage found for the **new query** show a clear preference for *nova pesquisa*. A similar conclusion for the **next bus** prime is not obvious, since the level of exposure to the word *próximo* was nearly five times the exposure to the word *seguinte*.

Moving to the second week the variety of primes for each concept increases and the users have more options. The way the system chooses the prompts is random, although some primes are included in more than one prompt. The results for **now** are inconclusive, although it is interesting to observe that the word *imediatamente* is more used than the others when not included in the system prompts. Unlike the **now** concept, the primes used for **new query** are uptaken in the majority of the answers. The only prime which was used without being uptaken from the system was *nova procura*. Finally, *seguinte*, the prime the system adopted for **next bus** in the second week was quickly adopted by the users.

## VI. CONCLUSIONS

The previous section underscores reports in the literature showing that lexical entrainment occurs and could be used in an SDS in order to optimize ASR and consequently overall dialog system performance. According to [1], in human-human conversation a new term has to be repeated two or more times before uptake occurs and a conceptual pact is established. Apparently in human-computer dialogs it seems that humans generally start to use the terms introduced by the system immediately including those that are less frequent in daily use.

The questionnaire at the end of the scenarios gave a different perspective from the user point of view. 18 of the users who called the system in both weeks answered the questionnaire comparing both systems. 10 of them felt better understood by the system. 15 of them noticed some sort of evolution in the system. The comments valued positively the wider variety of prompts available in the second week. They also felt that the system was answering faster than before although it was not.

The users' feedback emphasizes that the two modules that they interact directly with, the ASR and the TTS, are of major importance in the users' perception of the system. The new prompts proved to be very effective, reducing the number of

Concept	Prime	# W1 (% W1)	# W2 (% W2)	PER W1 (%)	PER W2 (%)	POS
now	W1: <i>agora</i>	64 (100.0)	77 (59.7)	56.3	89.6	ADV
	W2: <i>imediatamente</i>	0	26 (18.1)	-	61.5	ADV
	W2: <i>neste momento</i>	0	15 (10.4)	-	20.0	PRO/ART <sup>1</sup> + N
	W2: <i>mais rápido possível</i>	0	5 (3.5)	-	100.0	ADV + ADJ + ADJ
	W2: <i>mais brevemente possível</i>	0	6 (4.2)	-	16.6	ADV + ADV + ADJ
	W2: <b>new primes together</b>	0	52 (40.3)	-	40.4	-
new query	W1: <i>nova pesquisa</i>	35 (85.4)	1 (3.1)	51.4	0.0	ADJ + N
	W1: <i>outro percurso</i>	6 (14.6)	0	16.7	-	ART + N
	W2: <i>procurar novamente</i>	0	15 (46.9)	-	0.0	V + ADV
	W2: <i>nova procura</i>	0	8 (25.0)	-	25.0	ADJ + N
	W2: <i>outra procura</i>	0	5 (15.6)	-	20.0	ADJ + N
	W2: <i>nova busca</i>	0	3 (9.4)	-	0.0	ADJ + N
	W2: <b>new primes together</b>	0	31 (96.9)	-	9.7	-
	next bus	W1: <i>próximo</i>	62 (83.8)	3 (4.8)	35.5	100.0
	W2: <i>seguinte</i>	12 (16.2)	59 (95.2)	100.0	45.8	N or ADJ

TABLE VII  
USE OF THE PRIMES AND ERROR RATE IN BOTH WEEKS.

Speaker	utterance	uptake	no uptake	(%) usage
System	<i>Podê pedir informações sobre outro percurso, saber qual o próximo autocarro ou o autocarro anterior.</i>	-	-	-
User	<b>Próximo.</b>	✓	×	1/1
System	<i>Podia repetir se deseja saber o horário do autocarro seguinte, do autocarro anterior ou se deseja fazer uma nova pesquisa?</i>	-	-	-
User	<b>Próximo autocarro.</b>	×	✓	-
System	<i>Deseja fazer uma nova pesquisa, saber qual o próximo autocarro ou o autocarro anterior.</i>	-	-	-
User	<b>Autocarro seguinte.</b>	×	×	1/2
Total	-	1	1	50 %

TABLE VIII  
EXAMPLE OF UPTAKEN STATS TAKEN FORM INTERACTION FOR THE PRIME *próximo*.

Primes	Week 1				Week 2			
	uptaken	no uptake	(%) of usage	# prompts	uptaken	no uptake	(%) of usage	# prompts
<i>agora</i>	-	64	-	0	21	56(6)	63.3	60
<i>imediatamente</i>	-	-	-	-	14	12(3)	41.2	67
<i>neste momento</i>	-	-	-	-	11	4(1)	45.8	65
<i>mais rápido possível</i>	-	-	-	-	5	1(1)	33.3	30
<i>mais brevemente possível</i>	-	-	-	-	6	0	60.0	21
<i>nova pesquisa</i>	34	1	86.2	167	-	1	-	0
<i>outro percurso</i>	3	3(3)	42.9	106	-	-	-	-
<i>procurar novamente</i>	-	-	-	-	15	0	94.4	86
<i>nova procura</i>	-	-	-	-	6	2	83.3	51
<i>outra procura</i>	-	-	-	-	4	0	80.0	54
<i>nova busca</i>	-	-	-	-	3	0	33.3	54
<i>próximo</i>	51	11(2)	100.0	198	2	1	100.0	38
<i>seguinte</i>	8	3	53.8	39	57	2(1)	96.5	105

TABLE IX  
ANALYSIS FROM THE UPTAKEN OF THE PRIMES.

misunderstandings and removing some incorrectly synthesized words.

To better understand what could be improved in the future, we compared the ASR hypothesis with the reference. We noticed that some of the mistakes were happening due to the lack of read data to train the language model. Therefore we trained a language model with an automatically generated corpus comprised of 30k sentences. Although this was a good starting point, some gender and number agreements were incorrect and later resolved.

The majority of the primes used in the second week improved the speech recognition. However, a prime should also be adapted to the human's partner use of the language. The combination of the results in tables VIII and IX provide information that can enable us to set a tradeoff between the speech recognition performance and the uptake of the primes introduced.

In the case of **now**, *agora* was more used than other primes in both weeks. Despite being written in one of the scenarios, it is a very frequent word and therefore it should be chosen instead of the new primes. The effort should be in improving the speech recognition performance for this prime by using a better corpus to create the language model. *Imediatamente* due the usage in the no uptake condition should also be included in the set of primes for **now**. *Mais brevemente possível* has a very good PER, however it is not very frequent in daily language. For **new query**, table IX shows that the users have made good use of *nova pesquisa* during the first week, despite the synthesis problem reported in section III. Among the options for primes available in the second week *procurar novamente* that has very good PER and *nova procura* which was twice used without uptake and has also a low PER, can be considered good primes. The primes tested for next bus were very frequent in the weeks when they were more often incorporated in the prompts. However, the PER is lower for *próximo*. In addition, the usage of *próximo* without uptaken was much higher in the first week comparing to the use of *seguinte* in the same conditions in the second week. This indicates that *próximo* is good prime for **next bus**, however taking into account that users valued the variety of prompts it might be a good option to maintain both primes.

According to the last paragraph, it seems that there is some correlation between the use of the prime that is not adopted and the prime quality, since those are the words that the users recall immediately when the system requests a given concept.

## VII. FUTURE WORK

The results shown in this paper are encouraging. There are many directions to follow in finding optimal primes for spoken dialog systems.

Since the test were held with paid users differences were found in the final results, because callers want to please the system more than real users do. Once data from real users will be available, both the language and acoustic models will improve since we will have much more than what we used for the tests in this paper. In addition, modifications can be

made to the dialog strategy to make it more suitable to the real user's needs.

The results showed that users entrain to the system very quickly. We want to understand what makes a good prime. In section V-C, the phonetic distance was hypothesized to be a good measure. Therefore, in the near future we could compute the phonetic distance between prime candidates and the existing words on the vocabulary to minimize the confusability.

Despite the fact that the phonetic distance is a very useful feature for choosing words that are not confusable to the speech recognizer, it does not necessarily mean that it is a good prime, therefore other features are needed. As in [8], no evidence was found that POS had an influence on entrainment. But this could be due, in our case, because the fact that the paid users adopted every prime the system introduced. In the future we want to explore the features that are inherent in good primes for spoken dialog systems. The results indicate that there is a correlation between the use of a prime that is not adopted by the user and the "quality" of the prime. This is something to explore in a future work. In addition, phonetic distance, POS and word frequency could also be used. To close the loop, the trials would confirm if the primes found do not increase the degree of confusability in the speech recognizer vocabulary.

## REFERENCES

- [1] Brennan, S. E., Clark, H. H., "Conceptual Pacts and Lexical Choice in Conversations", *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 1996, Vol. 22, No. 6, 1482-1493.
- [2] Garrod, S. and Anderson, A., "Saying what you mean in dialogue: a study in conceptual and semantic co-ordination", *Cognition*, 27, 181-218, 1987.
- [3] Roe, D. B., Riley, M. D., "Prediction of Word Confusabilities for Speech Recognition", *Proc. ICSLP 94*, Yokohama, Japan.
- [4] A. Ward and D. Litman, "Automatically Measuring Lexical and Acoustic/Prosodic Convergence in Tutorial Dialog Corpora", *Proc. SLATE 2007*, Farmington, Pennsylvania, USA.
- [5] Reitter, D., Keller, F., Moore, J. D., "Computational Modelling of Structural Priming in Dialogue", *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, New York, 2006.
- [6] Stoyanchev, S., Stent, A., "Lexical and Syntactic Priming and Their Impact in Deployed Spoken Dialog System", *Proceedings of NAACL HLT 2009*, Boulder, Colorado, USA.
- [7] A. Raux, B. Langner, D. Bohus, A. Black and M. Eskenazi "Let's Go Public! Taking a Spoken Dialog System to the Real World", *Proc. Interspeech 2005*, Lisbon, Portugal, 2005.
- [8] Parent, G., Eskenazi, M., "Lexical Entrainment of Real Users in the Let's Go Spoken Dialog System", *Proceedings of Interspeech 2010*, Makuhari, Japan.
- [9] Bohus, D. et al, "Olympus: an open-source framework for conversational spoken language interface research, Bridging the Gap: Academic and Industrial Research in Dialog Technology workshop at HLT/NAACL 2007.
- [10] J. Neto, H. Meinedo, M. Viveiros, R. Cassaca, C. Martins and D. Caseiro, "Broadcast News Subtitling System in Portuguese", *Proc. ICASSP 2008*, Las Vegas.
- [11] S. Paulo et al, "DIXI - a generic text-to-speech system for European Portuguese", *Proc. PROPOR'2008 - 8th International Workshop on Computational Processing of the Portuguese Language*. Curia, Portugal: LNAI, Springer-Verlag, Sep. 2008.
- [12] A. Ward and D. Litman, "Measuring convergence and priming in tutorial dialog. In Technical report TR-07-148", University of Pittsburgh, 2007.
- [13] Black, A. W., Burger, S., Langner, B., Parent, G., Eskenazi, M., "Spoken Dialog Challenge 2010", *Proceedings of SLT 2010*, Berkeley.