



## Affirmative constituents in European Portuguese dialogues: prosodic and pragmatic properties

Vera Cabarrão<sup>1</sup>, Ana Isabel Mata<sup>2</sup>, Isabel Trancoso<sup>3</sup>

<sup>1</sup> FLUL/CLUL; L2F, INESC-ID

<sup>2</sup> FLUL/CLUL

<sup>3</sup> L2F, INESC-ID; IST- Universidade de Lisboa

veracabarrao@gmail.com, aim@letras.ulisboa.pt, isabel.trancoso@inesc-id.pt

### Abstract

This paper investigates the correlation between the prosodic properties and pragmatic functions of affirmative constituents in adult-adult interactions in European Portuguese (CORAL corpus). 515 affirmative constituents produced in 460 answers, extracted from 11 dialogues between 12 speakers, were analyzed. Results show that: i) *sim* 'yes', ok and grunts are the most frequent affirmative constituents; ii) *sim* 'yes' is associated with all the communicative functions analyzed, agreement, auto positive and confirm, ok tends to occur with agreement, and grunts are mainly associated to auto positive; iii) affirmative constituents have different prosodic properties according to their pragmatic function: agreement and confirm show a similar behavior, being auto positive the most distinct function. Agreement and confirm are commonly uttered with (H+)L\* L%, whereas auto positive is commonly uttered with L\*+H / (L+)H\* H%. When affirmative constituents co-occur in the same answer, there are evidences of tone copying between them. Correlations between constituents were also found in the following parameters: energy, pitch mean, maxima and minima, as well as pitch range. As for context-answer pairs, a pitch concord effect is also found between the pairs instruct-agreement and propositional question-confirm, although expressed in different degrees.

**Index Terms:** *affirmative constituents, communicative functions, prosody, dialogues*

### 1. Introduction

This study aims at describing the prosodic properties of affirmative constituents in adult-adult interactions in European Portuguese (EP) map-task dialogues, in order to correlate them with their pragmatic functions.

Affirmative constituents are one of the most frequent structures in spontaneous speech, much due to their role as a mechanism for structuring conversations ([1], [2]). Even though these structures are a topic of growing interest in areas like automatic speech processing, they have not been the object of many studies in EP, unlike what happens, for instance, in English (e.g. [3], [4], [5], [6], [7]).

In EP, an affirmative answer can be produced in several ways: (i) with one or more affirmative constituents, like *sim* 'yes'; with ok or grunts, like *humhum*; (ii) with the total or partial repetition of the previous utterance; or (iii) with *é* (the frozen form of the verb *ser* 'to be'). Moreover, affirmative constituents can also be associated with different

communicative functions, being, in some contexts, ambiguous or vague. The same affirmative word can, for instance, signal agreement with what was said previously, or as a backchannel strategy, as a positive reinforcement that the speaker can continue his/her speech, and also as a way to take the floor, to start or to end an utterance (e.g., [8], [9], [5], [9], [10]).

The high frequency of affirmative constituents in spontaneous speech, their ambiguity or vagueness and the emerging need to automatically recognize them justifies the need to linguistically characterize these structures in order to find some distinctive properties amongst them. Recent studies showed that both lexical and prosodic cues could help to better describe their properties. Following this line of work (e.g., [4], [5], [11], [12]), this study also adopts a perspective oriented to the prosodic and discursive analysis of affirmative constituents, considering both the lexical choices made by the speakers, and the properties of the adjacent contexts.

Using the Columbia Game Corpus, [12] showed that, for English, affirmative words with a backchannel function tend to occur with high/rising nuclear contours (L+H\* H-H%), and are produced mainly after utterances also with high/rising contours. Affirmative words coded as agreement, on the other hand, tend to occur with the tone H\* and the boundary tones L-H% or H-L%. [10] and [13], using the same corpus, confirmed these evidences, and showed that the acoustic-prosodic contrast between affirmative words lies mainly in its intensity levels and tonal boundary events.

Considering that dialogues are collaborative acts between speakers ([14], [15]), we also analysed the prosodic and pragmatic properties of the adjacent contexts that elicited the affirmative answers. It has been established for English ([4], [5], [11]), and also for EP ([16]) that context plays an important role in identifying the communicative function of the answer. In [16], the authors showed that the prosodic patterns of children's answers in EP dialogues vary according to the pragmatic function of the questions. There were also evidences of correlations between nuclear  $f_0$  levels in question-answer pairs, considered by the authors as a manifestation of pitch concord between speakers. In the work of [6], [17], and [18] in dyadic conversations in English, the authors showed that when speakers adapt or entrain their speech with their interlocutor, the dialogue is more natural and more likely to be successful.

### 2. Methodology

This study uses the CORAL corpus (ISLRN 499-311-025-331-2) ([19], [20]), comprising 64 dialogs in map-task format between 32 speakers, with a total of 7 hours orthographically

transcribed (61k words). All the dialogues are produced between two speakers; one is the giver of information and the other is the follower. The giver's map has a route drawn and some landmarks and his/her task is to provide informations and directions for the follower to reconstruct the same route in his/her incomplete map.

The subset used comprises 11 dialogues between 12 speakers (7 female and 5 male). Their age range was similar. All the speakers play both the roles of giver and follower. We analyzed 515 affirmative constituents, extracted from 460 affirmative answers, corresponding to several linguistic structures, namely: *sim* 'yes', *okay*, *exacto* 'exact', *exactamente* 'exactly', *certo* 'certainly', *está bem* 'that's right', *está correcto* 'that's correct', *pronto, pois*; grunts (e.g., *humhum*); the frozen form of the verb *ser* 'to be'; and both the repetition of the previous utterance, as well as the repetition of the main or auxiliary verb. These constituents occur as isolated utterances or in the beginning, middle or end of longer utterances. In the corpus analyzed, answers have a maximum of three affirmative constituents. Considering also the cooperative principle of dialogues ([14]), as well as the influence of the previous utterance in affirmative answers, we also extracted their adjacent contexts.

Affirmative constituents and contexts were classified as proposed by the *Semantic Annotation Framework – Part 2: Dialogue Acts* ([21], [22], [23]), which is being developed as an ISO standard. The affirmative constituents were coded with three communicative functions: agreement (AG) – the speaker accepts the previous utterance as true (1); auto positive (AuP) – the signal was well interpreted and the listener is still paying attention (2); and confirm (CONF) – answers to yes-no questions and check questions (3).

- (1) A: *Ora, de frente pela direita.* 'Well, in front by the right.'  
B: *Sim, de frente pela direita.* 'Yes, in front by the right.'
- (2) A: *A descer e a virar pela esquerda.* 'Descending and turning right'  
B: *Humhum.* 'Huh-huh.'
- (3) A: *Charneca das patas chocas é para onde eu vou, não é?* 'Charneca das patas chocas is where I'm going, right?'  
B: *Charneca das patas chocas, sim.* 'Charneca das patas chocas, yes.'

The adjacent contexts were also classified according to this taxonomy, with the following communicative functions: instruct, propositional questions (Yes-no question), set question (Wh-question), check question (tag question), inform, disconfirm, turn assign, and interaction structuring.

The prosodic annotation of the affirmative constituents and adjacent contexts was done using the ToBI (Tones and Break Indices) system ([24]), applying the first proposal *Towards a P\_ToBI* ([25]). For each affirmative constituent and adjacent context, we also extracted the following features: energy mean, pitch maxima, minima and range, as well as duration, using the Praat tool [26]. All  $f_0$  values were converted to semitones (ST).

### 3. Results

#### 3.1. Distributional patterns

Results show that affirmative answers containing the constituent *sim* 'yes' and variations of *sim* 'yes', like *sim sim* 'yes yes' or *sim senhor* 'yes sir', are the most frequent in the corpus (40%). The majority of these structures (36%) correspond to only one *sim* 'yes' produced in an isolated utterance. This is also the only affirmative word to be produced by all the speakers. Following this constituent, the most frequent ones are *ok* (16%) and grunts (10%). Together, *sim* 'yes', *ok* and grunts cover 75% of the data analyzed, the remaining constituents being quite residual in the corpus.

As for distributional patterns, the affirmative constituents occur mainly as isolated utterances (62%), but also at the beginning (18%), in the middle (6%), and at the end of longer utterances (14%). This shows that speakers prefer short answers to longer utterances with more than one affirmative constituent or with additional information.

Looking at the communicative functions analyzed, results show that affirmative constituents expressing agreement occur more often (53%) than those interpreted as auto positive (26%) and confirm (21%). These results were expected according to the nature of the corpus. The map-task format implies that one speaker gives informations or instructions about the map, as the interlocutor tends to express agreement with what was said before. Regarding the distribution of the affirmative constituents according to its communicative functions, *sim* 'yes' is associated with all the functions analyzed (N=204: AG=88; AuP=76; CONF=40); *ok* tends to occur with agreement (N=84: AG=67; AuP=12; CONF=5); and grunts are mainly associated to auto positive (N=50: AuP=31; AG=12; CONF=7). Affirmative constituents that co-occur in the same answer share communicative functions, which allow us to hypothesize that the pragmatic function does not depend on the position of the constituent, but relies on its context.

As for the contexts that elicit an affirmative answer, we observed three main tendencies: (i) declarative utterances, classified with the communicative functions *inform* (32%), and *instruct* (27%), associated with the act of sharing information and giving instructions of what to do next in the map; (ii) interrogative utterances, coded as propositional question and check questions (both with 10%), used by the speakers to elicit information about where to go in the map; and (iii) affirmative constituents, classified as agreement and confirm (both with 7% of occurrences), that elicit also a positive feedback from their interlocutor. The first ones correspond to about 80% of the total of contexts analyzed.

As for the speakers, we found that the role they play influences the production of affirmative words. The same speaker naturally produces more affirmative constituents as follower than as a giver. Moreover, we also found that speakers tend to use the same affirmative constituents despite their role, which can be related with their own personal experiences or style. Thus, results show that there is a greater inter-speaker variation than an intra-speaker one in the selection of affirmative constituents. We did not find any correlation between the choice of affirmative constituents and the gender of the speakers.

### 3.2. Prosodic properties

In our data, the affirmative constituents that occur in the beginning, middle or end of an utterance tend to form an independent prosodic unit. Moreover, in utterances with two affirmative constituents, there are evidences of tone copying between them, as well as of prosodic correlations in the following features: energy ( $rs=.383$ ,  $p(\text{one-tailed})<.05$ ),  $f_0$  mean ( $rs=.815$ ,  $p(\text{one-tailed})<.01$ ),  $f_0$  max ( $rs=.780$ ,  $p(\text{one-tailed})<.01$ ),  $f_0$  min ( $rs=.554$ ,  $p(\text{one-tailed})<.01$ ), and pitch range ( $rs=.472$ ,  $p(\text{one-tailed})<.01$ ).

Looking at the affirmative constituents that occur isolated or turn-initially, results show that the latter are mainly associated with a low/falling nuclear tone (68%), and a low boundary tone (70%). Those that are produced as an isolated utterance present more variation between low/falling (52%) and high/rising (48%) nuclear tone and low (48%) and high (52%) boundary tone. A statistical analysis, namely a Mann-Whitney test, also revealed significant differences between the affirmative constituents in these two positions in the following features: duration ( $U=18232.5$ ,  $p<.001$ ), intensity ( $U=9365$ ,  $p<.001$ ),  $f_0$  max ( $U=17403$ ,  $p<.05$ ) and min ( $U=18232.5$ ,  $p<.001$ ), and pitch range ( $U=17648.5$ ,  $p<.05$ ). Overall, these results show that, unlike the affirmative words that co-occur in the same utterance, those that are isolated or occur turn-initially have a different prosodic behavior. Therefore, the following prosodic analysis will only focus on the isolated affirmative structures, which account for 320 instances.

The isolated affirmative constituents classified as agreement and confirm are commonly uttered with low/falling nuclear tone (AG: 61%; CONF: 60%) and low boundary tone (AG: 57%; CONF: 53%), specifically (H+)L\* L%, whereas those coded as auto positive are commonly uttered with high/rising nuclear tone (64%), and high boundary tone (67%) - L\*+H / (L+)H\* H%. The auto positive function is also the most distinctive one regarding the mean values of  $f_0$  maxima and minima of the nuclear and boundary tones (Table 1). Results show that this function has the highest  $f_0$  values in almost all of the features, while agreement and confirm presents the lowest ones.

Table 1. Mean values of prosodic features  $f_0$  max and min of the nuclear tone and boundary tone.

Communicative functions	Nuclear tone		Boundary tone	
	Low/Falling	High/Rising	Low	High
	$f_0$ min*	$f_0$ max	$f_0$ min*	$f_0$ max
AG	8.526	15.809	8.663	18.045
AuP	<b>12.224</b>	<b>16.395</b>	<b>11.969</b>	17.797
CONF	9.212	15.439	8.410	<b>18.408</b>

\*statistical significant differences ( $p<.05$ )

Regarding their acoustic properties, results also show that the three functions differ in varying degrees: (i) between agreement and auto positive, there are statistical significant differences ( $p<.05$ ) in the features  $f_0$  mean ( $U=10015.5$ ), max ( $U=10632.5$ ), and min ( $U=10559$ ); (ii) between agreement and confirm, there are only significant differences in intensity ( $U=5388$ ,  $p<.05$ ), and (iii) between auto positive and confirm, the statistically significant differences ( $p<.05$ ) are found in the following features: intensity ( $U=4213.5$ ),  $f_0$  max ( $U=2666.5$ ), and min ( $U=2532.5$ ). These results point out that, even though

all functions display a unique prosodic behavior, auto positive is the most distinctive one.

Results also show (Table 2) that words coded as agreement present higher duration, even though they have the lowest values in the remaining features, namely  $f_0$  max, min, and pitch range; those with the auto positive function present the highest values in almost all of the  $f_0$  features, namely, mean, max, and min, and those coded as confirm have the highest intensity values, as well as range. Again, it seems that agreement and confirm are more closely related with each other than with the auto positive function.

Table 2. Mean values of prosodic features duration, intensity,  $f_0$  mean, max and min, and range.

Communicative functions	Duration (s)	Intensity (dB)	$f_0$ mean (ST)	$f_0$ Max	$f_0$ Min	Range
AG	<b>0.45451</b>	65.5910	12.047	14.946	8.949	6.003
AuP	0.43744	65.5008	<b>13.583</b>	<b>17.417</b>	<b>11.143</b>	6.273
CONF	0.45305	<b>67.9389</b>	12.032	15.656	9.002	<b>6.647</b>

These results are in line with the findings made by Gravano and colleagues ([10], [13], [27], [28]), regarding the analysis of affirmative cue words in English. The authors found that words classified as backchannel (function comparable with auto positive) tend to present a high/rising final contour, while those with the agreement function tend to have a low/falling or plateau final contour. Moreover, we can also establish a comparison with the work of [29], and [30]. The authors claim that low or plateau boundaries are associated with the notion of closure, which we can find in the agreement function, as high boundary tones are consistent with the notion of openness, a behavior correlated with auto positive.

### 3.3. Affirmative constituent *sim* 'yes'

*Sim* 'yes' is the most frequent affirmative constituent in our data, the only one produced by all speakers, and also the only constituent associated with all the communicative functions, being, therefore, the most ambiguous one. In isolated position, it accounts for 136 occurrences, the object of this analysis.

When classified as agreement, *sim* 'yes' tends to occur with high/rising nuclear tone (69%) and high boundary tone (71%). As auto positive, it presents the same behavior (66%, and 69%, respectively), and as confirm, it shows greater variation, being produced mainly with low/falling nuclear tone (54%) and low boundary tone (54%). Despite these tendencies, we also observe that the acoustic correlates of *sim* 'yes' vary between the different types of nuclear contours. A Mann-Whitney test showed that this constituent, when produced with a low/falling nuclear tone, presents statistically significant differences ( $p<.05$ ) between auto positive and agreement – in the features duration ( $U=232$ ), intensity ( $U=260$ ),  $f_0$  mean ( $U=237$ ), max ( $U=236.5$ ) and min ( $U=252.5$ ), and  $f_0$  min of the nuclear tone ( $U=243.5$ ) – between auto positive and confirm – in the features duration ( $U=66$ ),  $f_0$  mean ( $U=75$ ), max ( $U=77$ ) and min ( $U=61.5$ ), and  $f_0$  min of the nuclear tone ( $U=60.5$ ) – but not between agreement and confirm. When produced with low boundary tone, there are also statistical differences ( $p<.05$ ) between the functions auto positive and agreement – intensity ( $U=217$ ),  $f_0$  mean ( $U=200$ ), max ( $U=201.5$ ) and min ( $U=212.5$ ), and  $f_0$  min of the boundary tone ( $U=215$ ) – and between auto positive and confirm – duration ( $U=62$ ),  $f_0$  mean ( $U=65$ ), max ( $U=69$ ), min ( $U=53.5$ ), and  $f_0$  min of the boundary tone

(U=47). Again, no statistically significant differences were found between agreement and confirm. As for the high/rising nuclear contours, the same test did not reveal any differences between communicative functions.

These results show that *sim* ‘yes’ follows the tendencies observed with the remaining affirmative constituents, the auto positive function being the most distinctive one.

### 3.4. Context-answer pairs

Another perspective investigated during this work was the correlation between the most frequent contexts (inform, instruct, check question, and propositional question) and the affirmative constituents of the answer. For this task, we compared the  $f_0$  maxima and minima of the nuclear tone and boundary tone between the context and the affirmative constituent produced in adjacency.

Results show evidences of prosodic correlation between the pairs instruct and agreement -  $f_0$  min of the low/falling nuclear tone ( $r_s=.455$ ,  $p(\text{one-tailed})<.05$ ), and  $f_0$  min ( $r_s=.465$ ,  $p(\text{one-tailed})<.05$ ) and max ( $r_s=.595$ ,  $p(\text{one-tailed})<.01$ ) of the low and high boundary tone, respectively – and the pairs propositional question and confirm -  $f_0$  min of the low/falling nuclear tone ( $r_s=.571$ ,  $p(\text{one-tailed})<.05$ ). Therefore, the correlation found is strong in the pairs instruct-agreement, less strong between propositional question-confirm, and inexistent in the following pairs: inform-agreement, inform-auto positive, instruct-auto positive, and check question-confirm. These results were expectable given the nature of the data, as the CORAL corpus is characterized by short utterances and sequenced interactions between speakers, where one gives instructions so that the route in the map can be completed and the other signals if they were received or not. The correlations found, even though expressed in different degrees, suggest evidences of pitch concord effects between speakers. In order to verify this entrainment hypothesis, we intend to make a similar analysis between the speakers in the contexts and those in the answer, regarding their  $f_0$  levels, intensity, jitter, shimmer, noise-to-harmonics ratio (NHR), and speaking rate.

We also observed the correlations between context and answer as for their timing patterns. Results show that, in most cases, the affirmative answer is given after a silent pause, after the previous speaker ended his/her turn (N=353). Looking at the duration of these pauses by type of affirmative constituents, results show that, in most cases, the mean of duration is higher after a declarative context than after an interrogative one (Figure 1), exception made for the constituents *exacto/exactamente* ‘exact/exactly’, verb repetition, and context repetition. With declarative contexts, pauses are longer when the answer starts with the affirmative constituents *estar+certo/bem/bom* ‘to be+right/fine/good’ and *certo* ‘certainly’. As for the interrogative context, the pauses are longer when the answer is the repetition of the context. A Kruskal-Wallis test showed that these differences are statistically significant ( $H(10)=29.337$ ,  $p<.05$ ) in what concerns the types of constituents but no significance was found regarding their communicative functions. On the other hand, there are only 107 occurrences of overlapped speech, and in 50% of these cases the constituent is an agreement.

Based on this preliminary approach, and following the work of [6], we intend, in a near future, to expand this analysis to include temporal and rhythmic alignment patterns between contexts and affirmative constituents to establish social relations of dominance or accommodation between interlocutors.

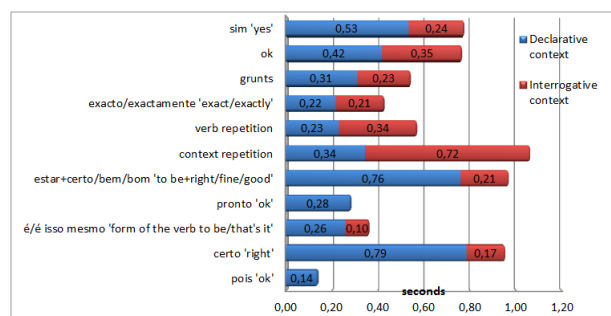


Figure 1: Mean duration of silent pauses between declarative and interrogative contexts and their affirmative answer.

## 4. Conclusions

This paper represented our first attempt to describe the acoustic-prosodic behavior of affirmative constituents in EP adult-adult interactions. The results show that affirmative constituents are related with different communicative functions: *sim* ‘yes’ is associated with all the functions, *ok* tends to occur as agreement, and grunts, as auto positive. In line with other works for English ([4], [11], and [5]), we can also conclude that the lexical selection of affirmative constituents *per se* is not enough to disambiguate their communicative functions in EP. Moreover, the affirmative constituents show different prosodic patterns according to their pragmatic function: agreement and confirm tend to occur with low/falling nuclear contour, showing significant differences between them in the intensity feature, as auto positive mainly occurs with high/rising nuclear contours, being significantly different from the remaining ones on almost any other features (e.g.  $f_0$  mean, maxima, and minima). We can hypothesize that this behavior is due to the fact that agreement and confirm imply an affirmative interpretation of the message, as in auto positive it is also implied that the speaker does not want to take the floor, rather wants to keep listening.

As for the most frequent affirmative constituent in the data, *sim* ‘yes’, it is associated with all the functions analyzed, showing much variation in the nuclear pitch accents and boundary tones with which is produced. There are also regularities in the prosodic behavior of this constituent that allow for the distinction between its functions. The analysis of the context-answer pairs show that there are correlations regarding pitch height between the functions instruct and agreement, on one hand, and propositional question and confirm, on the other, even though they are expressed in different degrees.

This preliminary analysis was aimed at being the baseline for further experiments in affirmative constituents in EP dialogues. With this knowledge of their linguistic and pragmatic properties, it will be possible, in a near future, to include them in automatic spoken dialogue systems. As future work, we also aim at a detailed investigation of the prosodic entrainment or accommodation features between speakers.

## 5. Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with references UID/LIN/00214/2013, and UID/CEC/50021/2013, under PhD grant SFRH/BD/96492/2013.

## 6. References

- [1] A. Schegloff, "Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences", in D. Tannen (Ed.), *Analyzing discourse: Text and talk. Georgetown University Roundtable on Languages and Linguistics 1981* (pp. 71-93). Washington, DC: Georgetown University Press, 1982.
- [2] B. Grosz and C. Sidner, "Attention, intention, and the structure of discourse", in *Computational Linguistics*, 12(3):175–204, 1986.
- [3] J. Allwood, J. Nivre, and E. Ahlsén, *On the semantics and pragmatics of linguistic feedback. Journal of Semantics*, 9(1):1–30, 1992.
- [4] D. Litman and J. Hirschberg, "Disambiguating cue phrases in text and speech" in *Proceedings of the 13<sup>th</sup> International Conference on Computational Linguistics*, pp. 251–256, Helsinki, 1990.
- [5] D. Jurafsky, E. Shriberg, B. Fox, and T. Curl, "Lexical, prosodic, and syntactic cues for dialog acts", in *Proceedings of ACL/COLING, Workshop on Discourse Relations and Discourse Markers*, pp. 114–120, Montreal, 1998.
- [6] S. Benus, A. Gravano, and J. Hirschberg, "Pragmatic aspects of temporal accommodation in turn-taking." *Journal of Pragmatics* 43, no. 12: 3001-3027, 2011.
- [7] N. G. Ward, *Possible Lexical Cues for Backchannel Responses. Interdisciplinary Workshop on Feedback Behaviors in Dialog*, 2012.
- [8] J. Carletta, A. Isard, S. Isard, J. Kowtko, G. Doherty-Sneddon, and A. Anderson, *HCRC dialogue structure coding manual. Technical Report HCRC/TR-82*, Human Communication Research Centre, University of Edinburgh, June, 1996.
- [9] J. Allen and M. Core, *Draft of DAMSL: Dialog act markup in several layers*. in <http://www.cs.rochester.edu/research/speech/damsl/RevisedManual/RevisedManual.html>, 1997.
- [10] A. Gravano, J. Hirschberg, S. Benus, "Affirmative cue words in task-oriented dialogue". *Computational Linguistics*. Vol. 38(1), pp. 1-39, 2012.
- [11] J. Hirschberg and D. Litman, "Empirical studies on the disambiguation of cue phrases". *Computational Linguistics*, 19(3):501–530, 1993.
- [12] S. Benus, A. Gravano, and J. Hirschberg, "The prosody of backchannels in American English", in *Proceedings ICPHS XVI*, Saarbrücken: 1065-1068, 2007.
- [13] A. Gravano and J. Hirschberg, "Backchannel-inviting cues in task-oriented dialogue" In *Proceedings of Interspeech*, pp. 1019–1022, Brighton, 2009.
- [14] H. P. Grice, "Logic and conversation" In Peter Cole, Jerry Morgan (Eds.) *Syntax and Semantics*, Vol. 3: *Speech Acts*, Academic Press, New York, pp. 41–58, 1975.
- [15] H. H. Clark and S. Brennan, "Grounding in communication", in L. Resnick, J. Levine, and S. Teasley, editors, *Perspectives on Socially Shared Cognition*, pp. 127–149. American Psychological Association (APA), Hyattsville, MD, 1991.
- [16] A. I. Mata, A. Santos, "Intonation of Early Answers to Confirmation-Seeking Questions in Spontaneous Speech" in N. Danis, K. Mesh & H. Sung (eds.), *Proceedings of the 35<sup>th</sup> Annual Boston University Conference on Language Development*, Somerville, Cascadilla Press, vol. 2: 440-452, 2011.
- [17] A. Nenkova, A. Gravano, and J. Hirschberg, "High frequency word entrainment in spoken dialogue.", in *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers*, pp. 169-172. Association for Computational Linguistics, 2008.
- [18] R. Levitan, J. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions", in *Proceedings of Interspeech 2011*. Florence, Italy, August 2011.
- [19] I. Trancoso, M. C. Viana, I. Duarte, G. Matos, "Corpus de Diálogo CORAL", in PROPOR'98 - III Encontro para o Processamento Computacional da Língua Portuguesa Escrita e Falada, Porto Alegre, Brasil, 1998.
- [20] M. C. Viana, I. Trancoso, I. Duarte, G. Matos, L. C. Oliveira, H. C. Campos & C. Correia "Apresentação do Projecto CORAL-Corpus de Diálogo Etiquetado", in Marrafa, P. & M. A. Mota (orgs) *Linguística Computacional: Investigação Fundamental e Aplicações*. Lisboa: Edições Colibri/APL, pp. 337-345, 1998.
- [21] H. Bunt, J. Alexandersson, J. Carletta, J. Chae, A. Chengyu Fang, K. Hasida, V. Petukhova, A. Popescu-Belis, C. Soria, and D. Traum, *Semantic annotation framework (SemAF), Part 2: Dialogue acts ISO Draft International Standard ISO 24617-2:2010*, ISO, Geneva, 2010.
- [22] H. Bunt, J. Alexandersson, J. Carletta, J. Chae, A. Chengyu Fang, K. Hasida, K. Lee, O. Petukhova, A. Popescu-Belis, L. Romary, C. Soria, and D. Traum, "Towards an ISO standard for dialogue act annotation.", in *Proceedings LREC 2010*, Malta, 2010.
- [23] H. Bunt, J. Alexandersson, J. Chae, A. Chengyu Fang, K. Hasida, O. Petukhova, A. Popescu-Belis, and D., "ISO 24617-2: A semantically-based standard for dialogue act annotation", in *Proceedings LREC 2012*, Istanbul, 2012.
- [24] K. Silverman, M. Beckam, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert & J. Hirschberg, "ToBI: a standard for labeling English prosody", in *Proceedings ICSLP*, Banff, vol. 2, 1992.
- [25] M. C. Viana, S. Frota, I. Falé, F. Fernandes, I. Mascarenhas, A. I. Mata, H. Moniz & M. Vigário, "Towards a P\_ToBI" In <http://www.ling.ohio-state.edu/~tobi/>, 2007.
- [26] P. Boersma, D. Weenink, *Praat: Doing phonetics by computer*, in <http://www.praat.org>, 2009.
- [27] A. Gravano, J. Hirschberg, "Turn-taking cues in task-oriented dialogue" in *Computer Speech and Language*. Vol. 25(3), pp. 601-634, 2011.
- [28] A. Gravano, J. Hirschberg, "Turn-taking cues in task-oriented dialogue", in *Computer Speech and Language*. Vol. 25(3), pp. 601-634, 2011.
- [29] J. R. Hobbs, "The Pierrehumbert-Hirschberg theory of intonational meaning made simple: Comments on Pierrehumbert and Hirschberg", in P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, MA, pp. 313–323, 1990.
- [30] J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse." in P. R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in Communication*. MIT Press, Cambridge, MA, pp. 271–311, 1990.