

Glissando Dialogs: a Corpus for the Analysis of Entrainment in Phone Services

David Escudero¹, Lourdes Aguilar², Juanma Garrido³

¹Dpt. Computer Science, Universidad de Valladolid, Spain

²Dpt. of Spanish Philology, Universidad Autónoma de Barcelona, Spain

³Dpt. of Translation and Language Sciences, Universidad Pompeu Fabra, Spain

descuder@infor.uva.es, juamaria.garrido@upf.edu, lourdes.aguilar@uab.es

Entrainment in speech is commonly defined as a speaker's adaptation to the speech of his interlocutor [1]. This paper presents a corpus that contains a collection of dialogs that simulate real phone services. The preliminary tests that have been done on the prosodic features of the dialog turns evidence that the entrainment phenomenon is clearly observed. We defend the future use of this corpus for analyzing and modeling this phenomenon in technological applications.¹

The corpus comprises two distinct data-sets, a news subcorpus and a dialogue subcorpus, the latter containing either conversational or task-oriented speech. More than twenty five hours were recorded by twenty eight speakers per language (Catalan and Spanish). Among these speakers, sixteen were professional (four radio news announcers and four advertising actors). The entire material presented here has been transcribed, aligned with the acoustic signal and prosodically annotated. All material contained in the corpus is provided under a Creative Commons Attribution 3.0 Unported License. This poster focuses on the task-oriented dialogs subcorpus and its application on the analysis of entrainment in conversational speech.

Our aim in the collection of this subcorpus is to offer a set of recorded interactions between two speakers oriented towards obtaining a specific goal. In all cases, speakers request for information, though for three different purposes: a) for traveling, b) for an exchange-university course, and c) for a tourist route.

These scenarios were selected following the diverse interests of research in areas such as (a) the development of speech technology dialogue systems (the design of automatic travel information systems, machine learning systems and tourist guides) and (b) linguistic studies that investigate the effect of variations derived from different communicative conditions on the speech of a given speaker.

The speakers who participate in each dialogue solve their task according to strict protocols, about which they were informed prior to their recording session, so that they would understand clearly what was expected from them (the protocols can be found in the technical report [2]). A relationship of cooperation is established between the participants, since both speaker and listener are involved in the execution of the task, and they both want to complete it with the maximum possible communicative success. These recordings stand as samples of intentional speech, similar to other kinds of intentional speech found in natural contexts, but obtained in a laboratory environment.

Due to our research interests on speech technology, all conversations were simulated to take place on the phone. For each

conversation, one of the speakers plays the role of instruction-giver and the other, the role of instruction-follower. In order to avoid long silences or unnatural hesitations, both participants were provided with the information necessary to solve each task, and it was made sure that they would read it, and become familiar with each scenario before the recording started.

Travel information is the most formal task, since the scenario consists in a telephone conversation between an operator and a customer who requests for price information and time schedules of a specific route. A graph facilitated to the instructions-given to solve the task (see [3] for an explanation of the methodology).

Information request for an exchange-university course. This dialogue takes place between a staff member of a university administration office who provides information about a course at a foreign university, and a student who requests for it.

The information request for a tourist route is a type of interaction inspired by the Map Task [4]). Nevertheless, the description of the situation and the type of task are different. In the Map Task corpus, subjects are required to cooperate in order to reproduce on the follower's map the route printed on the giver's map, and the success of the communication is quantified by the degree of coincidence of both routes. In our corpus, however, one of the speakers plays the role of somebody who is planning a trip to the Greek island of Corfu, and telephones a colleague who has lived for five years in Greece, in order to request for specific information concerning the route on the island. There is no specific route to reproduce; there is only an initial and a final point of the trip, and some places to visit on the way.

Finally we recorded 72 task oriented dialogues which durations goes from 4 to 16 minutes. 24 of the informants recorded also free-speech dialogues. 4 of them participated also in the read aloud news subcorpus. The recording was done in studio conditions simulating the phone interaction. All the material has been transcribed, phonetically aligned and enriched with prosodic information.

- [1] R. Levitan, A. Gravano, and J. Hirschberg, "Entrainment in speech preceding backchannels," in *Annual Meeting of the Association for Computational Linguistics (ACL/HLT)*, 2011, pp. 113–117.
- [2] D. Escudero, C. V. L. A. Valentín Cardeñoso, C. de la Mota, J. Garrido, O. Larrea, and E. Rodero, "Proyecto glissando: Grabación de corpus prosódico de noticias y diálogos en español." Departamento de Informática, Universidad de Valladolid, Tech. Rep. IT-DI-2010-3, Decembre 2010.
- [3] M. F. McTear, "Spoken dialogue technology: enabling the conversational user interface," *ACM Comput. Surv.*, vol. 34, pp. 90–169, March 2002.
- [4] A. Anderson, M. Bader, E. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, H. Thompson, and R. Weinert, "The hcrc map task corpus," *Language and Speech*, no. 34, pp. 351–366, 1991.

¹This research has been funded by three research grants awarded by the Spanish Ministerio de Ciencia e Innovación, namely the *Glissando project* FFI2008-04982-C003-01,2,3 FFI2011-29559-C02-01,2