Entrainment in Supreme Court Oral Arguments

Sarah Ita Levitan, Dr. Julia Hirschberg

Department of Computer Science, Columbia University

sarah.i.levitan@gmail.com, julia@cs.columbia.edu

Abstract

In conversation, people tend to become similar to their dialogue partner by adopting lexical, acoustic, prosodic, and syntactic characteristics of the interlocutor's speech. Research shows that this phenomenon, known as entrainment, is associated with task success and dialogue quality. We studied entrainment patterns in the Supreme Court corpus, and examined relationships between trial success and entrainment between lawyers and justices. We use Amazon Mechanical Turk (AMT) to preprocess the data and excise noisy areas in the audio files that skew the analysis process. Our initial results show that lawyers entrain more than justices, supporting the theory that the less dominant interlocutor is more likely to entrain to the more dominant speaker.

1. Introduction

Entrainment is prevalent in dialogue at many levels of production, including lexical, syntactic, acoustic and prosodic. For example, people tend to adopt specific word choices of their interlocutor [1], or adjust their amplitude to match that of their speaking partner [2]. This phenomenon is important because of its connection to dialogue success and quality. Research shows that conversations with greater entrainment are perceived as more natural and are strongly correlated with task success [3,4].

Studies in this area generally use corpora involving dialogue between two people. Our research is unique in that it uses the Supreme Court Corpus, which includes conversation between one lawyer and multiple justices. Instead of focusing on entrainment between two speakers, our study examines multiparty interactions in the corpus.

Our long-term goal is to study patterns of entrainment in Supreme Court oral arguments and analyze when and why entrainment is occurring. This corpus is useful in that we are able to quantify the success of a dialogue because we have knowledge of the case outcome, including individual justice votes. This is helpful in evaluating the comparative effects of entrainment or disentrainment. Some of the questions we are interested in answering are:

- Do justices entrain more to the lawyers whom they eventually side with?
- Does entrainment depend on other factors, such as justice gender, ideology, or investment in the case?
- Do more successful lawyers entrain more?
- Do lawyers entrain dynamically, adjusting their speech to each of their multiple interlocutors?

In answering these questions, we can explore theories of dominance and other social mediators of entrainment.

2. The Supreme Court Corpus

The SCOTUS Corpus consists of audio recordings of 76 United States Supreme Court sessions. These sessions are from the 2001 term, and the OYEZ project [5] identified speakers and manually word-aligned the recordings to their transcriptions. Each oral argument is approximately one hour long and consists of turns between the nine justices and two or more lawyers.

There are many advantages to using this corpus to study entrainment. The speaker identification provided by the OYEZ project is very helpful in studying entrainment patterns of a particular lawyer or justice, and the word alignment can be useful in researching lexical entrainment. More importantly, for each session we have records of the case outcome—which lawyer won the case, and how each justice voted. This gives us a clear way to quantify the success of a dialogue, and consequently examine how entrainment relates to the outcome.

Another useful aspect of the corpus is the distinction in rank in the courtroom. The justices are clearly the authority in the conversation, while the lawyers are the subordinate group. This clear division allows us to easily identify the dominant speaker in a given dialogue, and explore theories of dominance related to entrainment. For example, is the less dominant speaker more likely to entrain to his interlocutor than the more dominant speaker? Or, in our case, are lawyers more likely to entrain to justices?

Finally, the SCOTUS corpus is interesting because it contains multi-party conversations. The lawyers often interact with multiple justices in one conversation. This feature allows us to explore the idea of multi-party entrainment. For example, do people entrain differently depending on who their speech partner is? Do lawyers have different models for entrainment depending on which justice they are addressing?

One challenge associated with using this corpus is that sessions were not recorded in an ideal setting, and there is a lot of background noise in the recordings. Rustling papers and coughing are pervasive, and skew the calculation of various speech features that we extract in order to identify entrainment. To solve this problem, we employ Amazon's Mechanical Turk (AMT) [6] to identify noisy areas in the recordings, and excise these areas before analyzing the data.

3. Amazon Mechanical Turk

Amazon's Mechanical Turk (AMT) is a marketplace for tasks that require human intelligence. In AMT, a requester loads HITs (Human Intelligence Tasks) for turkers/workers to complete, and pays them a small sum for their work. AMT is an invaluable research tool, as it provides and on demand workforce of a wide range of people, with fast results at minimal cost.

In our case, we used AMT to identify noisy areas in the Supreme Court sessions. We divided each session into inter-pausal units (IPUs), where each unit is defined as a pause-free segment of speech from one speaker. We then concatenated the IPUs into segments of at least two seconds each. Each of our HITs requires the workers to listen to a set of 20 such segments (each segment has a duration ranging from 2-10 seconds), and determine if each segment contains noise or not (Fig. 1). They then select a checkbox labeled "Yes noise" or "No noise." The instructions define noise as coughing, rustling papers, laughter or a gavel bang, and include examples of each in order to train the worker. The worker is paid US $0.25 for each HIT successfully completed.

To control the quality of the AMT results, we set qualifications regarding which workers can

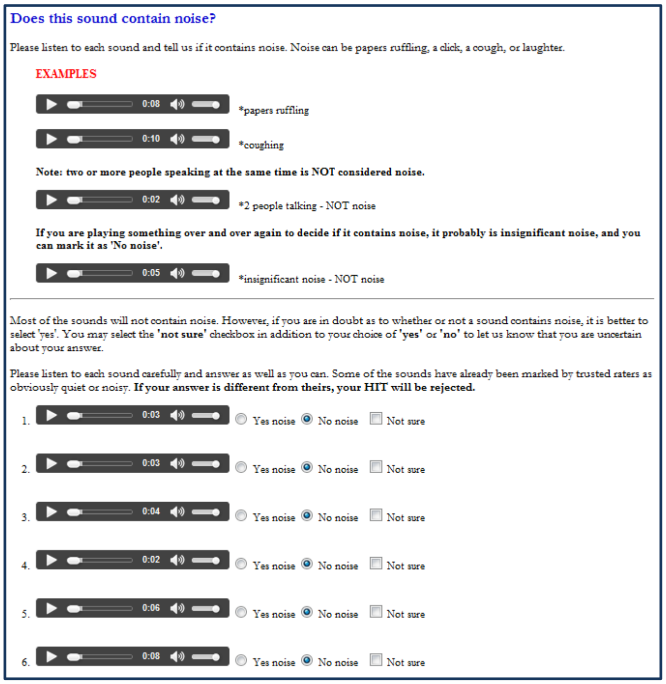

Fig. 1. Sample HIT

complete our HITs. Only workers who are in the United States and have a 90% acceptance rate on previous AMT HITs are accepted for our HITs. We initially created a small qualification exam that turkers were required to pass before being accepted to do our HITs. However we found that this discouraged workers from completing the HITs, even after passing the exam. Therefore we removed that requirement, and added *gold standard* questions to our HITs (segments which we had already identified as noisy or clean), and automatically rejected workers who answered those questions incorrectly.

Our noise identification HITs are still in progress; with 3 complete sessions (about 130 HITs) completed. We hope to build a classifier to identify noisy segments using the AMT results as training data.

4. Methods

4.1 AMT HITs

We used Amazon's Command Line Tools (CLT) to create and load HITs to AMT. These are essentially prewritten scripts that allow the requester to perform operations including loading HITs, retrieving results, and approving or rejecting work. , We wrote several Python scripts to generate input, question and properties files which the CLT accepts as parameters.

Since our HITs involve media elements, which AMT's restrictive question format does not handle well, we use an external question file, where the data remains on our machines, and Amazon redirects the worker to an external web page. Consequently, our HITs are part of dynamic webpages, and a specific HIT with a unique set of IPUs is generated when the worker accepts the task. We use a Python script along with CGI (Common Gateway Interface) to generate these dynamic HITs.

4.2 Identifying Entrainment

Once the HITs are completed and the results are retrieved, we convert the data to a form readable by Praat [7]–Text Grids (Fig 2).

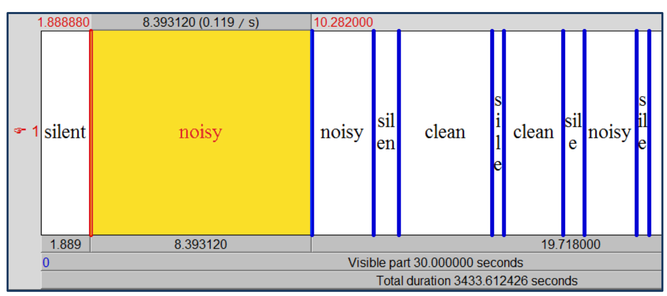

Fig. 2. Text Grid with noise identification

We then use Praat to extract features from the speech, skipping over the IPUs that are labeled as noisy. These features, including intensity, speaking rate, and voice quality, are used to identify points in conversation when entrainment is taking place. For example, we calculate the mean intensity at the beginning and end of turns in conversation, using the speaker identification tags to determine when turns begin and end. We then get a measure of local entrainment by calculating the differences in mean intensity to see at which points people are entraining. We use R to perform a series of statistical analyses on the data, and consider a result to be significant when its *p*-value is lower than 0.05.

We found significant differences between intensity values extracted from all the segments and values extract only from the segments that were found to be clean (t=-188.87, df=844.43, $p \approx 0$), reinforcing the importance of removing the noisy data.

5. Results

Our analysis shows smaller intensity differences between lawyers and justices than between justices and lawyers (t=-7.92, df=17622, p=2.57e-15, mean_lawyer=3.59, mean_justice=3.94). This indicates that lawyers entrain more than justices, which supports the theory that the less dominant interlocutor is more likely to entrain to the more dominant speaker.

We did not find a significant difference in entrainment between male and female lawyers (t=1.29, df=2205.1, p=0.20, mean_male=3.61, mean_female=3.50).

We were excited to find that differences between justices and petitioners are significantly smaller when the justice sides with the petitioner (t=-2.14, df=294.86, p=0.03, mean_petitioner=3.71, mean_respondent=4.18). This supports our hypothesis that justices entrain more to the lawyer that they eventually side with. However, we also found that differences between justices and respondents are also significantly smaller when the petitioner wins the case (t=-2.53, df=217.9, p=0.01,mean_petitioner=3.68,mean_respondent=4.26). In other words, justices entrain more to both lawyers, whenever the petitioner wins the case.

We plan to extract more speech features as more corpus data is preprocessed through AMT.

6. Acknowledgements

I would like to thank Professor Julia Hirschberg of Columbia University for mentoring me in this project, and DREU (Distributed Research Experience for Undergraduates) for this summer research opportunity.

7. References

[1] S.E. Brennan. 1996. Lexical entrainment in spontaneous dialog. *ISSD'96*.

[2] R. Coulston, S. Oviatt, and C. Darves. 2002. Amplitude convergence in children's conversational speech with animated personas. *ICSLP'02*.

[3] Daniel Goleman. 2006. *Social Intelligence*. Bantam.

[4] A. Nenkova, A. Gravano, and J. Hirschberg. 2008. High frequency word entrainment in spoken dialogue. In *Proceedings of ACL/HLT*.

[5] http://www.oyez.org

[6] http://www.mturk.com

[7] http://www.praat.org