Analysing Hesitation in the Map Task

Ilya Anisimoff, Robert Dale, and Rolf Schwitter

{Ilya.Anisimoff | Robert.Dale | Rolf.Schwitter}@mq.edu.au

Centre for Language Technology
Macquarie University
Sydney, Australia

Introduction

Spontaneous speech is produced in an incremental fashion; a speaker will initiate their utterance before they know exactly how they are going to finish. A consequence of this is that spontaneous speech production is an error-prone activity, and speakers can falter in numerous ways. These generally undesirable productions are referred to as *disfluencies*. This project probes the nature of a particular class of disfluency termed hesitation phenomena: unintentional spans of silence, *um*s and *uh*s in *filled pauses*, and drawn-out syllables known as *prolongations*. We aim to identify predictors of hesitations in annotated speech data, with the goal of determining the key characteristics that should be possessed by a cognitively-plausible computational model of spoken language production. Although there is a body of work in psycholinguistics that aims to determine the nature and causes of hesitation, much of that work relies on data collected under rather constrained laboratory conditions. In contrast, the present work describes some initial analysis of hesitations in a corpus of task-oriented dialogues.

Speech Errors, Disfluencies, Repairs, and Hesitations

Unintentional, erroneous, and non-fluent productions can be coarsely divided into speech errors and disfluencies. Speech errors are a class of phenomena which involve a mismatch between what the speaker intended to say and what they actually said. The classic example of a speech error is the Spoonerism (Garrett, 2001), where the sound segments of nearby words are exchanged:

(1) You have tasted the whole worm (wasted the whole term)

Speech errors are distinct from disfluencies, which are interruptions to the otherwise continuous flow of speech. While error-filled speech need not be disfluent, a speaker will usually be disfluent when she catches her error and corrects herself (often with an expression along the lines *oops, I mean ...*).

In its broadest sense, a disfluency is something that a speaker does or does not produce which interrupts their otherwise continuous stream of speech. Typical disfluencies include *um*s and *uh*s, self-corrections, and utterances which are abandoned then started anew. The terms 'disfluency' and 'self-repair' (or just 'repair') are often used to refer to the same types of phenomena, but they are not interchangeable. Strictly speaking, a repair occurs when a speaker interrupts themself and makes a modification to part of their preceding speech. So while a repair is a disfluency, a disfluency is not necessarily a repair. The following two examples illustrate this distinction:

(2) I should be there by uh... Sunday.
(3) I should be there by Sat- uh ... Sunday.

In the both cases the speaker is disfluent, but only in the second example does the speaker make a correction to their preceding speech. Hesitations are generally regarded as a category of speech disfluencies. Within this category, two types are most prominent in the literature: the periods of silence referred to as silent pauses, and the pauses containing a filler word such as *um* or *uh*. Prolongations are usually categorised as hesitations, although there is relatively little research on this type of phenomenon. Repeated sequences of words termed *repetitions* are sometimes categorised as hesitations, but this designation is not universally accepted. We remain uncontroversial by limiting the scope of our investigations to those universally accepted hesitations: filled and silent pauses.

Why Hesitate?

A number of explanations have been offered for why hesitation arises. These explanations are not necessarily mutually exclusive. One theory is that speakers have limited memory and attention spans, and therefore they are limited in how much speech they can plan. When they run out planned speech they will hesitate until they have planned some more. Another set of theories proposes that hesitations arise due to issues with the mental lexicon: either the speaker can not access the appropriate entry, or they have to choose from a selection of multiple appropriate entries. Others have suggested that some hesitations help coordinate turn-taking in conversation.

Speech Planning

There is a long history of using hesitations as evidence for planning units in speech production. The reasoning is that if a speaker hesitates once they have run out of planned speech, this hesitation will demarcate the boundary of a psychological unit of encoding. There have been reported increases in the incidences of filled pauses near the beginning of phrases (Maclay and Osgood, 1959), and discourse units (Swerts, 1998). Others have found silent pauses to be more frequent and longer at sentences boundaries than clauses (Goldman-Eisler, 1972; Holmes, 1988). So far there has yet to be an exhaustive analysis of units at multiple levels of abstraction, and studies tend to focus on only one or two units of interest. Superficially, it appears that a hesitation can occur just about anywhere, and its cause is attributed to the planning of the subsequent constituent.

Lexical Explanations

Early lexical explanations of hesitation claimed that hesitations appear at junctures where a word is relatively unlikely given the preceding context (Goldman-Eisler, 1958), even when word frequency has been controlled for (Beattie and Butterworth, 1979). More recent experimental evidence suggests that hesitation arises at points where there is more freedom of choice when selecting a lexical item (Hartsuiker and Notebaert, 2010). There is even some evidence that entire registers of speech are more hesitant than others because of varying degrees of restriction in vocabulary (Schachter et al., 1991).

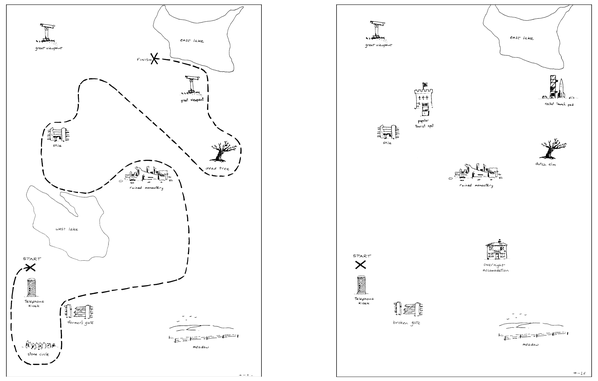
Conversational Signals

Some researchers have suggested that filled pauses serve as signals to coordinate conversation. Maclay and Osgood (1959) were perhaps the first to propose that filled pauses serve a floor-holding function which prevents the other speakers from seizing control of the conversation. Conversely, others have noted that sometimes filled pauses appear to be used by a speaker to signal uncertainty and invite assistance from her conversational partner (Bortfeld et al., 2001). Clark and Fox Tree (2002) claim that not only are filled pauses used by the speaker to signal an imminent delay, but there is also a functional difference between *um* and *uh*, which announce long and short delays respectively.

The Map Task Corpus

The HCRC Map Task (Anderson et al., 1991) is a collaborative task where two participants are seated facing each other and given nearly identical maps containing roughly 15 labelled landmarks. The participants cannot see each other's maps, and only one of the maps has route marked on it. The participant holding the map containing the route is designated the *Information Giver* (IG); their role is to describe the route to their partner, the *Information Follower* (IF), so she can draw it on her own map. Landmarks on the maps fall into one of four conditions:

Common: a landmark is identical on both maps;
Absent/Present: a landmark is present on one of the maps, but is absent from the other;
Name Change: a landmark is identical on both maps except in name; and
2:1: there are two instances of a landmark on one map, but only a single instance on the other.



The nature of the task means that hesitations are common; consider the following typical dialogue fragment.

(4)

GIVER erm ... we're start ... oop-- ... starting down ... that's the end of that ... erm starting above the telephone kiosk
FOLLOWER uh-huh
GIVER and we're going on the ... let's see left-hand side of that ... as I'm looking at it ... and come directly down
FOLLOWER uh-huh
GIVER below the stone circle ... and we come up
FOLLOWER I don't have a stone circle
GIVER you don't have a stone circle
FOLLOWER no
GIVER ah ... right okay

Deriving Hesitation Annotations

Although the Map Task corpus contains a disfluency annotation layer, this focuses on self-repair type disfluencies; hesitations are only marked when they occur within or adjacent to a repair. Since hesitations frequently occur outside of these contexts, we derived our own hesitation annotations by exploiting the annotations in the timed unit and part-of-speech annotation layers.

The timed unit layer marks stretches of speech as either a timed unit, which is a vocalisation that can be given an orthographic transcription; a period of silence; or an audible noise such as cough or an exhalation. Since such noises frequently occur interspersed through periods of silence, simply counting the periods of silences as silent pauses would result in an inflated hesitation count.

The Map Task was originally tagged using a modified version of the Brown Corpus tagset. Of particular utility for us is the FP tag, which was added to mark filled pauses.

	eh	...	*say*	*about*	*five*	*centimetres*	...	*on*	*the*	*map*
Unit Type	TU	Sil	TU	TU	TU	TU	Sil	TU	TU	TU
POS Tag	FP		VB	QDLT	CD	NNS		IN	AT	NN

We developed a simple grammar to identify hesitations. In the example above, two hesitations would be identified: an initial filled pause consisting of a filler and silence, and a second silent pause following *centimetres*. It could be argued that our initial filled pause is in fact two pauses, a filled pause and silent pause; however, we adopted a more conservative analysis to avoid potential inflation of pause counts. Our view is that a pause starts when a speaker ceases to be fluent and ends when she begins to be fluent once again. What distinguishes a filled pause from a silent pause is the presence of a filler, such as *um* or *uh*.

HesitantUnit := Silence ∨ Noise
∨ (TimedUnit.Tag = 'FP')
HesitantSequence := HesitantUnit+
FilledPause := (TimedUnit.Tag = 'FP')
∈ HesitantSequence
SilentPause := (TimedUnit.Tag = 'FP')
∉ HesitantSequence

Hesitations in the Corpus

Applying our algorithm to Map Task Corpus gave us the following results.

	FPs	SPs	Words	FPs/100 words	SPs/100 words
Corpus	1,948	35,767	150,705	1.293	23.733
Giver	1,516	13,356	103,440	1.466	12.912
Follower	432	22,411	47,265	0.914	48.330

What is immediately noticeable is the high proportion of silent pauses, especially in the information follower condition. This is likely to be a consequence of the high number of short responses participants gave when in this role.

We remarked earlier that hesitations can present themselves almost anywhere in an utterance, and a sceptic might argue that they occur randomly and are just attributed to planning whatever the following constituent is. For a preliminary analysis we decided to answer this most basic question: do hesitations occur randomly in speech? To answer this question we simplified POS tags for each word in the corpus and compared those that were preceded by a hesitation with those that were not.

	ADJ	ADV	CNJ	DET	EX	MOD	N	NP	NUM	P	PRO	TO	UH	V	VD	VG	VN	WH
Fluent	5,882	8,507	3,299	17,876	321	988	22,818	48	866	18,724	13,170	1,303	4,382	14,804	351	1,598	1,989	359
Hesitant	593	3,406	4,314	1,939	155	302	1,002	6	501	4,665	3,721	57	13,642	5,159	57	241	60	237

If hesitations do occur randomly, then we would expect the proportions to be the same. Instead we found that hesitation is associated with POS (X^2 = (17, n = 157342) = 38777, p < .001, ϕ_C = .496).

Conclusion and Further Research

We have developed tools for identifying hesitations in a corpus of task-oriented dialogues by exploiting existing annotation layers. Using these tools, we identified words which occurred in fluent and hesitant contexts and compared the POS. We have found an association between hesitancy and part-of-speech. These preliminary findings will serve as a launchpad for further research.

Our future work aims to shed some light onto the possible causes of hesitation. Initially, we will explore planning unit explanations of hesitation using this data. Where previous research has looked at the distribution of hesitations in relation to one or two units (e.g. a sentence or a clause), we will apply a bottom-up approach to the search for planning units in speech production. By using pauses to divide the speech stream into segments, and then seeing how well these segments align with syntactic units, we believe we can avoid a priori assumptions about the nature of speech units.

References

Anderson, A. H., Bader, M., Bard, E. G., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., and Weinert, R. (1991). The HCRC Map Task Corpus. *Language and Speech*, 34:352-366.

Beattie, G. W., and Butterworth, B. L. (1979). Contextual Probability and Word Frequency as Determinants of Pauses and Errors in Spontaneous Speech. *Language and Speech*, 22:201-211.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., and Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech*, 44:123-147.

Clark, H. H., and Fox Tree, J. E. (2002). Using Uh and Um in Spontaneous Speaking. *Cognition*, 82:73-111.

Garrett, M. F. (2001). The Psychology of Speech Errors. In N. J. Smelser and P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioral Sciences* (14864-14870). New York: Elsevier.

Goldman-Eisler, F. (1958). Speech Production and the Predictability of Words in Context. *The Quarterly Journal of Experimental Psychology*, 10:96-106.

Goldman-Eisler, F. (1972). Pause, Clauses, Sentences. *Language and Speech*, 15:103-113.

Hartsuiker R. J., and Notebaert, L. (2001). Lexical Access Problems Lead to Disfluencies in Speech. *Experimental Psychology*, 57:169-177.

Holmes, V. M. (1988). Hesitations and Sentence Planning. *Language and Cognitive Processes*, 3:323-361.

Maclay H., and Osgood, C.E. (1959). Hesitation Phenomena in Spontaneous English Speech. *Word*, 15:19-44.

Schachter, S., Christenfeld, N., Ravina, B., and Bilous, F. (1991). Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, 60:362-367.

Swerts, M. (1998). Filled Pauses As Markers Of Discourse Structure. *Journal of Pragmatics*, 30:485-496.