A preliminary study of hesitation phenomena in L1 and L2 productions: a multimodal approach

Loulou Kosmala, Aliyah Morgenstern

▶ **To cite this version:**

Loulou Kosmala, Aliyah Morgenstern. A preliminary study of hesitation phenomena in L1 and L2 productions: a multimodal approach. Disfluency in Spontaneous Speech 2017, 2017, Stockholm, Sweden. hal-02360610

HAL Id: hal-02360610
https://hal.science/hal-02360610

Submitted on 2 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A preliminary study of hesitation phenomena in L1 and L2 productions: a multimodal approach

Loulou Kosmala and Aliyah Morgenstern
Sorbonne Nouvelle University

Abstract

This paper presents a preliminary study of vocal hesitations in L1 and L2 productions using a multimodal perspective. It investigates the use of vocal hesitations of French learners of English interacting in tandem with American speakers in semi-spontaneous speech. Several hesitation markers were analyzed (filled pauses, unfilled pauses, prolongations and non-lexical sounds) based on formal and functional features as well as their relation to gesture. Results do not show great differences in the frequency of vocal hesitations between L1 and L2 productions overall; however, we find differences in duration and combination complexity. Our study indicated that vocal hesitations mainly served planning functions and were very often accompanied with gaze aversion both in L1 and L2 productions. Moreover, speakers did not tend to gesture while hesitating. We conclude that hesitations mainly served planning strategies both in L1 and L2 speech, but with some differences in duration and complexity.

Introduction

One aspect of hesitation disfluency is that it is sequential (Crible, Degand & Gilquin 2017:71) – disfluency and hesitation markers do not only appear in isolation, but in combination. Merlo and Mansur (2004:496) showed that disfluencies were correlated (P-value of 0.01), and Rieger (2003:42) found combinations of filled pauses as well as combinations with other markers (such as 'a:and uh'). Therefore, some disfluencies can be 'complex' (Shriberg, 1994). Another aspect of hesitation disfluency is its duration—hesitation is manifested acoustically as speakers generally hesitate for a certain period of time (Barr 2003:461).

Indeed, time is required in conversation as speech flow can never be fully fluent. Planning discourse can be a demanding cognitive activity as speakers need to make processing decisions such as selecting the appropriate sentence, phrase, or lexical item. Hesitation markers, such as filled pauses, can be used by speakers to buy time in order to plan their utterances. (Jucker 2015; Holmes 1988; Fehringer & Fry 2007; Tottie 2014). Lengthenings can also serve as a cue to indicate that speakers are still currently planning their speech (Betz & Wagner 2016). In this sense, the main function of vocal hesitations is to provide thinking time to introduce a 'thought unit' (Kjellmer 2003:174) and to maintain contact with the interlocutor (Guaïtella 1993).

In L2 acquisition, such processing decisions are even more challenging and can lead to additional difficulties (Watanabe and Rose 2012). L2 speech is usually characterized by slower pace and frequent use of filled and unfilled pauses (Tavakoli 2011:71). Some studies show that speakers produce longer pauses in their L2 than in their L1 (Riazantseva 2001; Tavakoli 2011), and there is a significant correlation between the number of hesitation phenomena produced in L1 and L2 speech. (Fehringer & Fry, 2007:57).

Hesitation can also be expressed non-verbally. A number of studies have shown that speakers tend to avert their gaze while hesitating (Schober et al. 2012; Glenberg, Schroeder & Robertson 1998; Swerts & Krahmer 2005) and typically produce a 'thinking face' when they are searching for a word (Goodwin & Goodwin 1996). Though gestures do not often co-occur with filled pauses (Christenfeld, Schachter & Bilous 1991); some studies have shown that gesture suspensions (such as hold and retraction) tend to co-occur with disfluency (see Seyfeddinipur 2006).

Few studies have looked at the use of 'vocal hesitations' in relation to gesture. Vocal hesitations (Guaïtella 1993:128) include unfilled pauses (silence), filled pauses ('uh' and 'um') and lengthening (word or syllable prolongation). In this preliminary study, we focus on the use of vocal hesitations in L1–L1 and L1–L2 settings. They are analyzed based on duration and complexity, as well as their non-verbal features expressed in visual modalities.

Corpus and methods

The materials for the study were provided by the SITAF corpus (Spécificité des Interactions Verbales dans le Cadre de Tandems Anglais-Français) collected at Sorbonne Nouvelle University between 2012 and 2014 (Horgues & Scheuer 2015). The data include a 25-hour video-recorded corpus, comprising 21 pairs of undergraduate students, 21 native French speakers and 21 native English speakers. For this study, two tasks were focused on: (1) "Liar, Liar"—a storytelling task in which one participant has to tell a story and insert three lies

that the partner will have to identify; (2) "Like Minds" – a question-answering task in which both participants have to answer a controversial question and decide on their degree of agreement. Eight recordings were selected, four comprising L1-L1 pairings and four L1-L2. They involved four subjects: A03 and A07 (American speakers): F03 and F07 (French speakers). All participants were female, aged 18 to 21. The duration of our selected corpus is approximately 30 minutes.

Three types of hesitation markers were analyzed, focusing on vocal productions: (1) filled pauses—non-lexical autonomous vocalic fillers, i.e. 'uh', 'um', 'euh', 'eum'; (2) unfilled pauses–perceptible pauses exceeding 400ms; (3) prolongations–syllable, vowel or consonant lengthening that does not signal the end of an intonation unit. The duration and combination of these markers were also analyzed, and coded in three distinct categories: (1) brief hesitation—hesitation made of one marker that does not exceed 600 ms; (2) elongated hesitation—hesitation made of one marker which exceeds 600 ms; (3) complex hesitation (following Shriberg, 1994)—hesitation made of several hesitation markers. Our analysis of prolongations and the distinction between brief and elongated hesitations was done based on perception.

Table 1. Examples of hesitation categories

Category	Utterance	Duration	Marker(s)
Brief hesitation	**&euh** then we came back because we were really tired.	540ms	filled pause (FP)
Elongated hesitation	**a:and** yeah that was real long and we were really tired.	1.859ms	prolongation
Complex hesitation	**&um a:and &um** it was fun.	2.555ms	FP+prolong+ filler

We also coded the functions of hesitations, making two distinctions: (1) planning function–hesitation used to plan at the macro- or micro-level, i.e. plan a new utterance, continue planning, or plan a specific lexical item; (2) reformulating function–hesitations used to reformulate parts of the utterance, either by repeating, repairing or starting a new constituent; here the hesitation typically co-occurs with repairs, restarts and repeats.

Table 2. Examples of functions

Function	Utterance
Planning	**a:and** you know they're still part of society even if they're not living in it
Reformulating	**the:e** [/] the to the flower garden

The goal was also to see what happens non-verbally when speakers produce a hesitation (aligning verbal and non-verbal modalities) when there was a change in their non-verbal behavior. We observed the following features: (1) gaze direction–towards or away from the interlocutor (2) head movement–tilts, head shakes, head nods and downward head movement; (3) facial expressions–frowning, smiling or eyes closed (4) gesture phases (Seyfeddinipur 2006: 106)–rest position or return to rest position, preparation phase, hold, gesture unit, interruption of the gesture.

Results

A total of 330 hesitations were found in the data. Table 4 indicates that on average, speakers produced 11 hesitations per minute. It seems that French learners produced more hesitations in their L1 (16.2 per minute) than in their L2 (12.3 per minute), which does not support the view that L2 learners are more hesitant in their L2 (Fehringer & Fry, 2007:57). However, our results show great individual differences between the two French speakers, which is in favour of the idea that hesitation phenomena vary from speaker to speaker (Fehringer & Fry, 2007:57), and so no conclusions can be made at that point. No significant differences were found in the functions; our results show that hesitations were mostly used for planning.

The difference between L1 and L2 seem to be at the level of the structure of hesitations. As shown in Table 3, hesitations produced by the French speakers in their L2 are longer than in their L1. They are also longer than the ones produced by the English speakers in their L1. French speakers tend to use slightly more elongated hesitations and slightly fewer brief hesitations when speaking their L2. This could indicate that French learners need more time to plan their utterances in their L2 than in their L1. Their elongated hesitations are longer when using their L1, but only represent 17% of the hesitations. However, complex hesitations produced in their L2 are much longer than those produced in their L1 and those produced by the English speakers. Hesitations comprising a single form (brief and elongated) represent 65% of the hesitations (218 out of 330 in total). French speakers use more brief filled pauses in their L1 (48 out of 55) than in their L2 (30 out of 43), while English speakers use a higher number of brief prolongations. Our results suggest differences in duration in L1 and L2 speech.

Even though no striking differences were found in the number of complex hesitations in L1 and L2 speech (Table 3), we find differences in the complexity of hesitations (Table 4). We looked at all the different combinations in the complex hesitations produced by the French learners, and we found 14 combinations in total for L1 speech: 11 composed of two forms; two composed of three forms, and one composed of four forms.

Table 3. Overall results.

	Task 1 L1-L1						Task 1 & 2 L1-L2						Task 1 L1-L1 & 2 L1-L2						
	L1 French						L2 English						L1 English						Total
	F03		F07		Total		F03		F07		Total		A03		A07		Total		
Speaking time	1.74 mn		5.28 mn		7.02 mn		4.17 mn		4.26 mn		8.43 mn		7.14 mn		6.27 mn		13.43 mn		30 mn
No. hesitations	12		102		114		28		76		104		68		44		112		330
No. hesitations per minute	6.8		19.3		16.2		6.7		17.8		12.3		9.5		7		8.3		11
Average duration	878ms		786ms		796ms		663ms		1142ms		1014ms		969ms		917ms		948ms		916ms
No. Brief Hesitations	3	25%	52	50%	55	48%	14	50%	29	38%	43	41%	31	45%	13	29%	44	39%	142
No. Elongated Hesitations	3	25%	17	16%	20	17%	8	28%	20	26%	28	26%	10	14%	18	40%	28	25%	76
No. Complex Hesitations	6	50%	33	32%	39	34%	6	22%	27	35%	33	31%	27	39%	13	29%	40	36%	112
No. Planning functions	10	83%	79	78%	88	78%	20	71%	63	82%	82	79%	56	82%	42	95%	98	88%	268
No.Reformulating functions	2	7%	23	22%	25	22%	8	29%	13	18%	21	21%	12	12%	2	5%	14	12%	60

Two recurrent combinations were found: "FP+pause" and "prolongation+FP", which are the most frequent ones (produced 11 and 8 times). In L2 speech, however, our results include 18 combinations in total: 6 combinations of 2 forms, 7 combinations of 3 forms, 4 combinations of 4 forms and 1 combination of 5 forms. No recurrent combinations were found. Hesitations produced in the L2 have greater complexity. Table 5 shows that in 63% of cases, hands tend to be in rest position while the speaker produces the hesitation (209 out of 330). This suggests that speakers tend not to gesture when they hesitate. This is consistent with previous studies (Christenfeld, Schachter & Bilous, 1991).

However, in cases when they do gesture, they sometimes produce a gesture unit (13%) or their gesture tends to be held, interrupted or return to rest position at the same time as the hesitation (20%). Such interruptions indicate a suspension from the speaker both in verbal and non-verbal modalities. Speakers momentarily retreat from the interaction to gaze away (82% of the time) and think.

Figure 1 shows that the speaker produces a complex hesitation characterized by the prolongation of the vowel 'al' in the adjective 'traditional' and a non-lexical sound (a click). While she hesitates, her hands simultaneously return to rest position and she looks down. When she retrieves the noun phrase 'Christmas dinner', she opens her palms and gazes back at her interlocutor.

Table 4. Examples of complex combinations in L1 and L2

	Example	Combination
L1 French	e:et &euh	prolong+FP
	&eum (...)	FP+pause
L2 English	a:aand yea:ah &m (...) [click]	prolong+prolong+nl-sound+pause+nl-sound
	a:and the:e &um	prolong+prolong+FP

Table 5. Visual features accompanying hesitation.

	L1 Fr.	L2 En.	L1 En.	Total
gesture unit	7	19	18	44
hold	3	10	18	31
rest position	102	59	48	209
return to rest position	2	8	24	34
preparation phase	0	7	4	11
interrupted	0	1	0	1
head shake	2	1	2	5
tilt	1	2	2	5
head nod	0	1	0	1
head down	0	1	0	1
eyes closed	2	4	5	11
frowns	1	7	4	12
winces	2	0	5	7
Gaze away	90	82	101	273
Gaze toward interlocutor	24	22	11	57

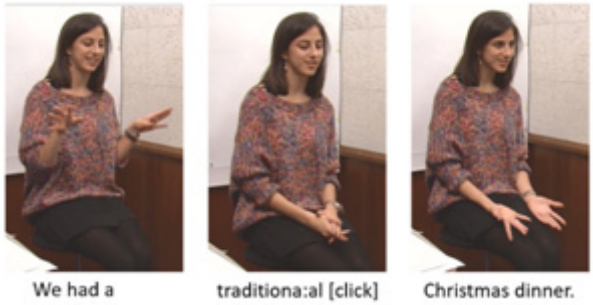


Figure 1. Multimodal activity during hesitation

Conclusions

This preliminary study was conducted in order to test new methods for analyzing vocal hesitations in L1 and L2 speech. Our results indicated that the frequency of hesitations did not differ greatly between L1 and L2 productions.

However, L2 learners seemed to require more time for planning in their L2 than in their L1, as their hesitations tended to be longer and showed greater complexity. This could suggest that hesitations did not result from speech processing difficulty, but rather helped speakers to structure their speech. Another interpretation would be that French speakers aligned their hesitations with American speakers as they produced longer hesitation markers in their L2 (see Finlayson, Lickley & Corley, 2010).

Speakers also often disengaged from interaction when hesitating, as they tended to avert their gaze from their interlocutor as a way to suppress the environment's control over cognition (Glenberg, Schroder & Robertson, 1998). They also did not tend to gesture often while hesitating, although a temporal connection between gesture suspension and speech suspension was found.

Our preliminary results are promising for conducting further analyses on vocal hesitations in L1 and L2 speech and their relation to gesture. Further research will be carried out on more subjects from the whole corpus in order to extend this analysis.

Acknowledgements

Many thanks to Céline Horgues, Maria Candea and Gaëtanelle Gilquin for their interest in our study and for reading a draft of this paper, and to our anonymous reviewers for their constructive input.

References

Barr, D. J. 2003. Paralinguistic Correlates of Conceptual Structure. *Psychonomic Bulletin & Review* 10 (2): 462–467.

Betz, S. & P. Wagner. 2016. Disfluent Lengthening in Spontaneous Speech. *Elektronische Sprachsignal-verarbeitung (ESSV) 2016*.

Christenfeld, N., S. Schachter & F. Bilous. 1991. Filled Pauses and Gestures: It's Not Coincidence. *Journal of Psycholinguistic Research* 20(1):1–10.

Crible, L., L. Degand & G. Gilquin. 2017. The Clustering of Discourse Markers and Filled Pauses. *Languages in Contrast* 17(1):69–95.

Fehringer, C. & C. Fry. 2007. Hesitation Phenomena in the Language Production of Bilingual Speakers: The Role of Working Memory. *Folia Linguistica* 41(1–2):37–72.

Finlayson, I., R. J. Lickley & M Corley. 2010. The influence of articulation rate, and the disfluency of others, on one's own speech. In *DiSS-LPSS Joint Workshop*, 25–26 September 2010, Tokyo, Japan, 119–122.

Glenberg, A. M., J. L. Schroeder & D. A. Robertson. 1998. Averting the Gaze Disengages the Environment and Facilitates Remembering. *Memory & Cognition* 26(4):651–658.

Goodwin, C. & M. H. Goodwin. 1996. Seeing as a Situated Activity: Formulating Planes. In D. Middleton & Y. Engestrom (eds.), *Cognition and Communication at Work*, edited by. Cambridge: Cambridge University Press.

Guaïtella, I. 1993. Functional, Acoustical and Perceptual Analysis of Vocal Hesitations in Spontaneous Speech. In *ESCA Workshop on Prosody*.

Holmes, V. M. 1988. Hesitations and Sentence Planning. *Language and Cognitive Processes* 3(4):323–361.

Horgues, C. & S. Scheuer, S. 2015. Why some things are better done in tandem? In J. A. Mompeán & J. Fouz-González (eds.), *Investigating English Pronunciation: Current Trends and Directions*. Basingstoke and NY: Palgrave Macmillan, 47–82.

Jucker, A. H. 2015. Uh and Um as Planners in the Corpus of Historical American English. *Developments in English: Expanding Electronic Evidence*, 162–77.

Kjellmer, G. 2003. Hesitation. In Defence of Er and Erm. *English Studies* 84(2):170–198.

Merlo, S. & L. L. Mansur. 2004. Descriptive Discourse: Topic Familiarity and Disfluencies. *Journal of Communication Disorders* 37(6):489–503.

Riazantseva, A. 2001. Second Language Proficiency and Pausing – A Study of Russian Speakers of English. *Studies in Second Language Acquisition* 23(04):497–526.

Rieger, C. L. 2003. Disfluencies and Hesitation Strategies in Oral L2 Tests. In R. Eklund (ed.), *Proceedings of DiSS 2003*, 5–8 September 2003, Göteborg University, Sweden, Gothenburg Papers in Theoretical Linguistics 90, 41–44.

Schober, M. F., F. G. Conrad, W. Dijkstra & Y. P. Ongena. 2012. Disfluencies and Gaze Aversion in Unreliable Responses to Survey Questions. *Journal of Official Statistics* 28(4):555.

Seyfeddinipur, M. 2006. *Disfluency: Interrupting Speech and Gesture*. MPI-Series in Psycholinguistics.

Shriberg, E. (1994). *Preliminaries to a Theory of Speech Disfluencies*. Ph.D. dissertation, University of California, Berkeley

Swerts, M. & E. Krahmer, E. 2005. Audiovisual Prosody and Feeling of Knowing. *Journal of Memory and Language* 53(1):81–94.

Tavakoli, P. 2011. Pausing Patterns: Differences between L2 Learners and Native Speakers. *ELT Journal* 65(1):71–79.

Tottie, G. 2014. On the Use of Uh and Um in American English. *Functions of Language* 21(1):6–29.

Watanabe, M. & R. Rose. 2012. Pausology and Hesitation Phenomena in Second Language Acquisition. *The Routledge Encyclopedia of Second Language Acquisition*, 480–483.

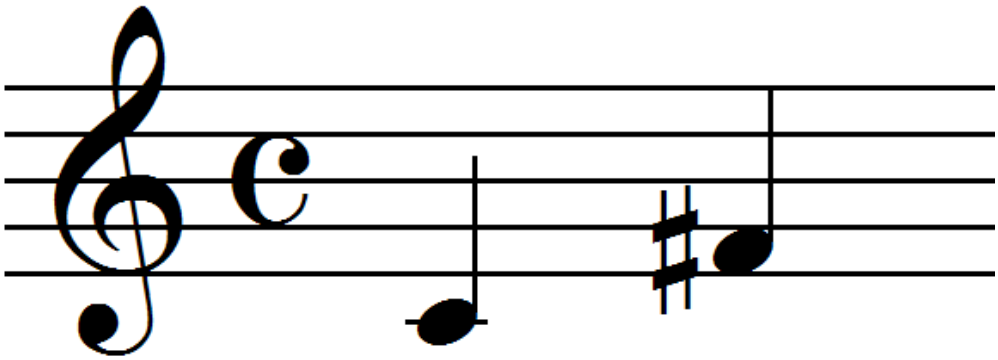
Proceedings of

DiSS 2017

The 8th Workshop on Disfluency in Spontaneous Speech

KTH Royal Institute of Technology
Stockholm, Sweden
18–19 August 2017

TMH-QPSR
Volume 58(1)



Edited by
Robert Eklund & Ralph Rose

Conference website: http://www.diss2017.org
Proceedings also available at: http://roberteklund.info/conferences/diss2017

Cover design by Robert Eklund
Graphics and photographs by Robert Eklund (except ISCA and KTH logotypes)

Proceedings of DiSS 2017, Disfluency in Spontaneous Speech
Workshop held at the Royal Institute of Technology (KTH), Stockholm, Sweden, 18–19 August 2017
TMH-QPSR volume 58(1)
Editors: Robert Eklund & Ralph Rose
Department of Speech, Music and Hearing
Royal Institute of Technology (KTH)
Lindstedtsvägen 24
SE-100 44 Stockholm, Sweden

ISSN 1104-5787
ISRN KTH/CSC/TMH–17/01-SE

© The Authors and the Department of Speech, Music and Hearing, KTH, Sweden