Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 99–102.

Factors that determine the form and position of disfluencies in spontaneous utterances

Peter Howell, Jennifer Hayes* , Ceri Savage*, Jane Ladd*, Nafisa Patel**

* Department of Psychology, University College London, London, England

Abstract

This presentation reviews work on types of disfluency in the spontaneous speech of fluent speakers and speakers who stutter. Examination is made of factors that determine where disfluencies are located. It is concluded that the phonological, or prosodic, word provides a good basis for explaining the distribution of different types of disfluency in spontaneous speech.

1. Introduction

In the previous paper [13], a model that simulated the types of disfluency that occur within phonological word (PW) units was presented. PWs were defined in that paper as having a content word as nucleus and function words as optional satellites that can precede and follow the nucleus. Thus 'I sprang up' is a PW with one function word preceding and one function word following the content word.

The issues examined in section two of this paper are the types of disfluency and the different taxonomies that have been applied to them. All the disfluencies seen in fluent speakers are also seen in speakers who stutter, though they occur more frequently. Thus, taxonomies developed for stuttered speech can validly be applied across a range of speakers. In section three previous evidence and new preliminary data are reviewed that support the view that PW units are appropriate for analysis of disfluencies in spontaneous utterances and better than other contextual units (syntactic ones in particular). In section four, alternative ways of defining PW are considered and whether they improve predictions about disfluencies in spontaneous speech over content- and function-based definitions of PW.

2. Selected taxonomies of disfluency types

Speakers produce a range of different types of disfluency in spontaneous speech and these disfluencies can be classified in various different ways. The main division is between errors proper (where a phone occurs in the wrong position for an intended word) and fluency failures that do not show signs of speech error according to the preceding definition. This definition of error would include word-selection errors (e.g. 'left' for 'right'). Speech errors per se are not the focus of this article (though they feature indirectly in one of the accounts discussed below). Speech errors, as defined here, have been reported to be rare even in the spontaneous speech of fluent speakers [8].

The common types of fluency failure involve hesitation, production of whole or parts of words (as in false starts). Two observations about these fluency failures are necessary: First, all speakers show them though they occur more frequently in the speech of some speakers than others. For instance, speakers who stutter show a higher incidence which makes their speech convenient for collecting a large sample. Also, as shown in the previous paper, the balance between different types of fluency failure that changes over development in speakers who stutter could arise as a self-organizing change as the language production system matures. Thus, there is no inherent difference between the forms of fluency shown by speakers who stutter and fluent speakers, only an imbalance in the frequency with which the different types of event occur.

.Second, it is commonly recognized that the various identifiably different types of fluency failure need to be sub-divided. Some of the widely identified types of fluency failure are repetitions involving parts of words, single whole words or phrases, filled and unfilled pauses, prolongation of phones in a word (usually the first one) and a break within a word. This list was used by Ambrose and Yairi [1] with stuttered speech who also included revision and interjections. The revisions are not included as they involve errors as defined earlier (the word or words selected are different to those intended). Many interjections are essentially filled pauses and when so are included in that category. The way the remaining types of fluency failure are sub-divided varies between research groups and is not theoretically neutral. Yairi and Ambrose [26] use an empirical criterion to divide the events into two classes. One class is stuttering-like disfluencies (SLDs) which are more prevalent in the speech of people who stutter than in fluent controls. This class includes part- and whole-word repetitions, prolongations and broken words. The second class is other disfluencies (OD) which consists of phrase repetitions and the classes (not considered here) of interjections and repetitions. Though the assignment of disfluency types to SLD and OD is empirical, it has theoretical implications insofar as it implies there are some categories of disfluency that are associated with stuttering and others that are not.

A second way of dividing up the disfluency events uses Levelt's [19] repair framework. Repairs occur when the speaker makes an error and corrects it. An example that includes a comprehensive list of repair events (many of these are optional in actual speech) would be "Turn left at the, no, turn right at the crossroads". This would be excluded from the disfluent events being included here as 'left' does not include the correct phones for the intended word ('right'). However, some repairs do not include an error as in the related example 'Turn, turn right at the crossroads'. This is assumed to reflect an underlying error that is detected and corrected before output and is referred to as a covert repair. Pauses (filled and unfilled), word and phrase repetitions would all signify covert repair processes. Kolk and his students [18] have developed the covert repair hypothesis (CRH) which takes these events as signs of repair to underlying errors. Fragments of words (as in prolongations, word breaks and part-word repetitions) might be parts of a word produced in error [9].

The final division of disfluencies is that according to the EXPLAN model [10, 12, 14] which basically produces a similar division to CRH (pauses, word and phrase repetitions in one class and prolongations, word breaks and part-word repetitions in the other) but accounts for them in a different way.

Ambrose and Yairi's taxonomy is empirical whereas CRH's and EXPLAN's categorizations also involve theoretical assumptions about underlying causes of disfluency. The principle difference between CRH and EXPLAN is that CRH considers disfluency events are reflections of underlying errors whereas EXPLAN considers they are not, so they do not require an error-detection mechanism. According to EXPLAN, disfluencies should be divided into two classes because each class has a different role and the types of disfluency have their effect on different parts of utterances (in particular on different parts of PW).

The example used at the start of the article ('I sprang up') can be employed to explain the latter points about type and role of disfluency in more depth. 'Sprang' is the word most likely to be difficult to generate. The first class of disfluency involves the simpler function word preceding the content word 'sprang', which is usually either repeated or has a preceding pause. The role of pausing and word and phrase repetition involving the initial function word/s is to gain time to prepare the difficult content word. This class of disfluency has been referred to as stalling. The second class of disfluency involves the content word itself and usually involves disfluency on the first part of the word. This class of disfluency has been referred to as advancing. A way of contrasting the CRH and EXPLAN accounts is in terms of whether errors or timing are regarded as the paramount features that lead to fluency failure.

EXPLAN theory makes restrictive predictions about the positioning of each type of disfluency in PW which depends on the role attributed the respective type. The predictions and evidence about each type of disfluency are reviewed next.

3. Evidence on position and type of disfluency in spontaneous speech

3.1. Evidence for stalling

1. The view that word and phrase repetition and pausing represent an attempt to buy time (what is referred to here as 'stalling') has been proposed [5] and empirically tested [6, 20, 21] by several previous authors particularly those on child language.

2. Stalling at different position in an utterance. If word and phrase repetition serves a delaying role (as assumed to happen in stalling), these events should be located in early positions in utterances (in particular, prior to content word nuclei) and should be specific to function words. To examine whether position effects occur for function and content words, speech data from English and Spanish speakers who stutter were segmented into PWs. Both the English and Spanish data showed (1) Serial position functions for function words with higher disfluency percentage in earlier positions [2, 3]; (2) No such serial position effects occur for content words [2, 3].

3. Stalling prior to content words. As indicated under 2, stalling only works when it is done on function words that precede a content word. PW are ideal for these analyses as they can have function words prior to and after the single content word in such units. Thus, the function words that could serve a delaying role (those preceding the content word) and those that could not (those following the content word) can be coded unambiguously. The function words can then be examined to see whether it is only those that precede the content word that show the disfluencies we refer to as stallings. There are data form fluent speakers that support this prediction. Strenstrom and Svartvik [23] looked at repetition of subject ('he' in the phrase 'he hit him') and object pronouns ('him' in the phrase). There are two features to note about these examples: First, the units are a PW of the form FCF. Second, position relative to the content word is indicated by the different forms the words take. The prediction, if stalling applies, is that only the subject pronouns could serve a delaying role and so these are the only ones that should be repeated. This is exactly what Strenstrom and Svartvik found. Similar analyses have been made on English [2] and Spanish [3] using a wider range of PWs with essentially the same finding. Au-Yeung et al. [2] found significant effects for all the age groups they examined for English with higher rates with a high level of disfluency in pre-content function words and a low level on post-content function words. For Spanish [3], the difference in disfluency rate across these positions was significant for three age groups out of five.

4. Comparison of PW units for predicting position of stallings with other units. The serial and ordinal position effects indicated above may be mediated by units other than PW. Thus, syntactic units can be constructed from PW units, and then a syntactic unit will have a PW in initial position. This PW would lead to serial and ordinal position effects because they are in initial position in the syntactic unit, not because they are in initial position in a PW. Similar arguments would apply to units such as utterances [27]. To test between two alternatives (PW and utterances), utterances which contained two function words F1 (final in the utterance-initial PW) and F2 (initial in an utterance non-initial PW), […F1] .. [F2 …] …, were examined. F2 was stuttered more than F1 in spite of F1's earlier position in an utterance. This analysis shows that an utterance position effect cannot account for the effects of PW position.

5. Pause position in PW. Inserting a pause should also be a way of stalling. If so, pauses would be expected to be positioned at the start of a PW to play a delaying role. The majority (more than 50%) of pauses in stuttered speech occur at PW boundaries. One issue not examined to date is what happens to PWs that do not have initial function words (e.g. 'hit it'). Pauses could be used for delaying onsets when the content word is complex (in which case, such PW would have a higher incidence of pausing than other PW).

3.2. Evidence for Advancing

1. Words vary in difficulty and this should impact in different ways on words involved in stalling (function) and advancing (content) fluency failures. Stalling does not occur because the words themselves are difficult, but because an up-coming word is difficult. Advancings should occur on words that are difficult. Moreover, content words are not equally difficult and it should only be the difficult ones that attract advancing-type disfluency. To test these predictions, measures of word difficulty are required that can be applied to each class of word. Three metrics have been employed in work so far: The first was developed by Yairi's group and was an empirical measure that involved classifying words as to whether they contained a consonant string (CS), phonemes that were acquired late in language development (LEC) and whether the word was multisyllabic or not, MS [24]. They found little evidence of difficulty affecting stuttering rate for SLD and OD. Two facts are of note: First, they did not analyze the difficulty factors for word position and when that was done, effects were found particularly in older speakers who show more advancing-type disfluencies [15]. Second, as the OD and SLD classifications do not map directly onto stallings and advancings, their data may benefit from reanalysis using these categories.

The second measure, called the index of phonetic complexity or IPC for short, used an index based on babbled speech [17] that would have been thought to be particularly useful for younger speakers who stutter. As with the CS/LEC/MS metric, this was not useful with young speakers [25] but was with older speakers [7, 16]. It includes eight factors. The main drawback of this metric is that it is not word-position specific which is an important factor to include as most stuttering occurs in initial position.

The third metric is one based on non-linear phonology. Non-linear means, in this context, classification based on hierarchical schemes that divide the syllable into onset, nucleus and coda (thus, it captures position-dependencies in stuttered speech. The scheme has six syllabic and two phonetic factors. It still needs to be fully investigated for stuttered speech, but has been employed with children who are developing fluently (see 3 below).

A feature of note about all three schemes is that they all include CS and LEC factors and these factors are consistently found to be related to disfluency. For the CS/LEC/MS and IPC schemes which have been extensively evaluated in connection with their impact on stuttering, effects of difficulty are found on content words in older speakers.

The focus on phonological and phonetic factors is not intended to suggest that the only 'difficulty' leading to advancing are at these levels. It is conceivable that lexical factors (frequency of occurrence, age of acquisition, name agreement, etc.),m syntactic, lexical and prosodic factors also make a word difficult and hence prone to mis-timing.

2. Does stalling occur before difficult content words? In unpublished work, Howell and Ladd looked at the difficulty of word following stalling disfluencies using the IPC metric and compared this with difficulty of words not following stalling disfluencies. They found that the content words that followed stalling were significantly higher in difficulty than the words from fluent contexts. From this it would appear that the difficulty of the up-coming word determines whether the function words are repeated or not. Howell and Patel performed a similar analysis for young fluent children using ANOPHS to score word difficulty. Once again, the difficulty of the up-coming word was found to determine whether the function words are repeated or not

3. Priming of function and content words. Priming can be used to increase the speed of online speech production [22]. According to EXPLAN priming of content and function words should yield different effects on fluency. Savage and Howell conduced an experiment in which intransitive picture descriptions were elicited after priming of either function or content words (e.g. 'He is' or 'swimming' respectively) and compared performance for 12 children who stutter and 12 fluent age and gender matched controls (mean age of both groups was six years). Both groups produced significantly fewer disfluencies in their target responses after content word primes than function word primes and produced significantly more silent pausing after function word primes than content word primes. The effect of priming was significantly greater for CWS than for the fluent children and the target responses of CWS after a function word prime contained content words of significantly longer duration than did those of fluent children. These results are consistent with the view that advancing the point in time when a content word has to be produced (function word priming) causes more fluency problems than priming content words, they are pre-prepared for processing particularly in CWS.

3.3. Does producing one type of disfluency prevent the other?

1. Though there are two classes of fluency failure, they are both reflections of the same underlying problem (tackling a difficult content word either by repeating the words before or on the content word itself). If stalling is effective, then the speaker should not produce an advancing after a stalling and advancings should only occur when there is no stalling. Stuttering in PW has been examined to see whether it occurs in the either-or manner predicted. For English [2] and Spanish [3], less than 3% of disfluencies occurred on both the initial function words and the content word. An estimate of what would be expected by chance can be made by assuming function and content word disfluencies are independent. A related t-test across speakers showed the disfluency rates below 3% were significantly lower than expected on the assumption that function and content word disfluencies are independent.

2. Although disfluency in a PW is one sort or the other (e.g. stalling or advancing), the relative incidence of stallings and advancings can vary within or between individuals. A related observation is that disfluency in young speakers occurs on function words, but on content words in older speakers (which suggests that stalling is more prevalent at young ages and advancing at older ages. Both English and Spanish, speakers had high disfluency rates on pre-content function words and low disfluency rates on the content word. Disfluency on the function words dropped off with age and, as it did so, disfluency on the content words increased (termed an exchange function). In [13], an account of this change based on word frequency changes was presented.

4. Comparison with other units that could potentially account for these phenomena

The work in section two supported PW as a unit for analysis of disfluencies. However, alternative definitions of PW are possible that would be correlated with content-based segmentations for English. For instance, stressed words or low frequency words could act as nuclei. Lexical status (content versus function) can be dissociated from stress in Spanish where function words as well as content words are stressed. Segmentation of the same speech material into PW based either on content/function or stressed/unstressed criteria were made for Spanish material. Examination of PW that were segmented differently according to the two schemes showed that both produced ordinal and serial position effects and an exchange function [11]. Thus both would seem appropriate bases for defining PW. In recent analyses, Howell and Hayes have shown the two forms of PW do, however, show a difference when pauses were examined. More pauses occurred prior to PW that use stressed words as nuclei.

5. Conclusions and future work

The main conclusion is that PW appear to be a good unit for predicting the nature and position of disfluencies in speech. This is in stark contrast with Bernstein Ratner’s conclusion [4] that “we cannot actually tell whether ANY disfluency (stuttered or not) reflects problems with a specific *word*, the *clause* it is being embedded in, or a word three words *further down* in the utterance, etc.”. There is also slight evidence (that based on pauses) that stressed words may be a better basis for specifying the nuclei of PW. Work on fluency failure in PW in fluent speakers lags behind that in speakers who stutter and this needs to be rectified.

6. Acknowledgement

This research was supported by Wellcome Trust grant 072639.

7. References

[1] Ambrose, N.G. & Yairi, E. (1999). Normative disfluency data for early childhood stuttering. *Journal of Speech, Language and Hearing Research, 42,* 895-909.

[2] Au-Yeung, J. Howell, P. & Pilgrim, L. (1998). Phonological words and stuttering on function words.

Journal of Speech, Language and Hearing Research, 41, 1019-1030.

[3] Au-Yeung, J. Vallejo Gomez, I. & Howell, P. (2003) Exchange of disfluency from function words to content words with age in Spanish speakers who stutter. *Journal of Speech, Language and Hearing Research, 46*, 754-765.

[4] Bernstein Ratner, N. (in press). Is phonetic complexity a useful construct in understanding stuttering? *Journal of Fluency Disorders*.

[5] Clark, H. & Clark, E. (1977). *Psychology and Language: An Introduction to Psycholinguistics*. New York: Harcourt Brace.

[6] Clark, H.H. & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive Psychology, 37*, 201-242.

[7] Dworzynski, K. & Howell, P. (2004). Predicting stuttering from phonetic complexity in German. *Journal of Fluency Disorders, 29*, 149-173.

[8] Garnham, A.,Shillcock, R. C. Brown, G. D. A. Mill, A. I. D. & Cutler, A. (1981). Slips of the tongue in the London-Lund corpus of spontaneous conversation. *Linguistics, 19*, 805-817.

[9] Hartsuiker, R. J. & Kolk, H. H. J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113-157.

[10] Howell, P. (2002). The EXPLAN theory of fluency control applied to the Treatment of Stuttering by Altered Feedback and Operant Procedures. In E. Fava (Ed.), *Current Issues in Linguistic Theory series: Pathology and therapy of speech disorders* (pp.95-118). Amsterdam: John Benjamins.

[11] Howell, P. (2003). Is a perceptual monitor needed to explain how speech errors are repaired? *Gothenburg papers in Theoretical Linguistics, 90*, 29-32.

[12] Howell, P. (2004a). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communicative Sciences and Disorders, 39*, 122-139.

[13] Howell, P. & Akande, O. (in press). Simulations of the types of disfluency produced in spontaneous utterances by fluent speakers, and the change in disfluency type seen as speakers who stutter get older

[14] Howell, P. & Au-Yeung, J. (2002). The EXPLAN theory of fluency control and the diagnosis of stuttering. In E. Fava (Ed.), *Current Issues in Linguistic Theory series: Pathology and therapy of speech disorders* (pp.75-94). Amsterdam: John Benjamins.

[15] Howell, P. Au-Yeung, J. & Sackin, S. (1999). Exchange of stuttering from function words to content words with age. *Journal of Speech, Language and Hearing Research, 42*, 345-354.

[16] Howell, P. Au-Yeung, J, Yaruss, S. & Eldridge, K. (in press). Phonetic difficulty and stuttering in English.

[17] Jakielski, K. J. (1998). *Motor organization in the acquisition of consonant clusters*. PhD thesis, University of Texas at Austin. Ann Arbor Michigan.

[18] Kolk, H.H.J. & Postma, A. (1997). Stuttering as a covert-repair phenomenon. In R.F. Curlee, & G. Siegel (Eds.), *Nature and treatment of stuttering: New directions*.(pp. 182-203). Boston: Allyn & Bacon.

[19] Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

[20] MacWhinney, B., & Osser, H. (1977). Verbal planning functions in children's speech. *Child Development, 48*, 978-985.

[21 Rispoli, M. (2003). Changes in the nature of sentence production during the period of grammatical development. *Journal of Speech, Language and Hearing Research, 46*, 818-830.

[22] Smith, M. & Wheeldon, L. (2001). Syntactic priming in spoken sentence production - an online study. *Cognition, 78*, 123-164.

[23] Strenstrom, A-B. & Svartvik, J. (1994). Imparsable speech: Repeats and other nonfluencies in spoken English. In N. Oostdijk & P. de Haan (Eds.), *Corpus-based research into language*. Atlanta, GA: Rodopi.

[24] Throneburg, N.R. Yairi, E. & Paden, E.P. (1994). The relation between phonological difficulty and the occurrence of disfluencies in the early stage of stuttering. *Journal of Speech and Hearing Research, 37*, 504-509.

[25] Weiss, A. L., & Jakielski, K.J. (2001). Phonetic complexity measurement and prediction of children's disfluencies: A preliminary study. In B. Maassen, W. Hulstijn, R. Kent, H. F. M. Peters & P. H. M. M. van Lieshout (Eds.), *The 4th Nijmegen Conferences on Speech Motor Control and Stuttering.* Nijmegen: Uitgeverij Vantilt.

[26] Yairi, E. & Ambrose, N. G. (2004). *Early childhood stuttering*. Austin, TX: Pro-Ed.

[27] Yaruss, J.S. (1999). Utterance length, syntactic complexity, and childhood stuttering. Journal of Speech, Language and Hearing Research, 42, 329-344.