Gaëtanelle Gilquin and **Sylvie De Cock** (eds.). *Errors and disfluencies in spoken corpora*. Amsterdam: John Benjamins. 2013. 172 pp. ISBN 978-90-272-0271-0. Reviewed by **Susan Nacey**, Hedmark University College, Norway.

Errors and disfluencies in spoken corpora results from a pre-conference workshop held in 2009 at the 30th ICAME conference at Lancaster University, dealing with issues that had arisen during the compilation of the Louvain International Database of Spoken English Interlanguage (LINDSEI). The workshop focused primarily on the distinction between errors and disfluencies, the practicalities of mark-up and annotation of spoken learner language, the possible functions of hesitation, together with various pedagogical implications of corpus research into errors and disfluencies. This volume includes five individual articles, a mix of papers presented at that workshop as well as invited contributions from other scholars working in the field. The overall goal is to shed additional light on certain phenomena restricted to spoken language, such as fillers, silent pauses, speech rate and error rate.

Gaëtanelle Gilquin and Sylvie De Cock's introduction sets the scene by first attempting a definition of both 'error' and 'disfluency': "errors are defined as forms that deviate from a given native-speaker norm and disfluencies cover phenomena that are generally seen to reflect speakers' online planning and encoding difficulties" (p. 5). They are best viewed as a continuum where 'errors' lie at one outer pole and 'disfluencies' at the other. The study of errors and disfluencies has been revolutionized by the advent of corpus linguistics, showing just how frequently they occur despite often going unnoticed. Gilquin and De Cock go on to take up particular challenges in the annotation of spoken data and then summarize a few major earlier studies into errors and disfluencies. The introduction also touches on the limitations of both spoken corpora and other data types used to investigate errors and disfluencies in oral language and then provides an overview of possible applications of such investigations in various fields such as foreign language teaching, clinical linguistics, etc. This section closes with a fairly comprehensive bibliography that constitutes nearly a third of the entire introduction – in and of itself almost worth the price of the book.

Gunnel Tottie presents a preliminary investigation into the fillers *uh* and *um* in British English through examining their occurrence in two sub-corpora of the British National Corpus: BNC-CG with recordings from the domains of business, education, leisure and public institutions, and BNC-DEM containing informal impromptu speech. Her particular focus concerns differences in their frequency in the two subcorpora, together with choice of filler (nasalized *um* versus non-nasalized *uh*). Tottie presents the data gathered from the transcriptions of the spoken data broken down in three different ways, according to gender, age and socio-economic class of the informants. In addition, she emphasizes the need for the availability of sound files (rather than transcriptions only), so that researchers would no longer be completely dependent upon the transcription policies of the individual corpus compilers, and would be able to conduct more qualitative investigations. This is especially important when studying items such as *uh* and *um* which some may not consider words, and where the length of the utterance may reveal something about its function.

Christoph Rühlemann, **Andrej Bagoutdinov** and **Matthew Brook O'Donnell** investigate the role of pauses in conversational narratives from the Narrative Corpus, extracted from the BNC. Their data is taken from 150,000 words across 279 stories involving more than 600 speakers. They first compare the frequency of four pause types – the fillers *er* and *erm* (the only ones noted in the BNC), along with both short and long silent pauses – in narratives and general conversation, finding that all but long pauses occur more often in narratives. This finding leads them to investigate the distribution of pause frequency in initial, medial and final stages of narratives for any patterns there. In addition to reporting on frequencies alone, however, Rühlemann *et al*. also examine frequent collocates of pauses as well as discourse patterns, to flesh out the function pauses play in storytelling, as well as immediately before and after the narrative sequence. They convincingly argue through careful analysis and thoughtful interpretation that pauses are not disfluencies per se, but rather 'windows on the mind' providing evidence of cognitive processes in telling narratives.

Karin Aijmer compares the use of the pragmatic marker *well* in the spoken English of both Swedish students in LINDSEI and young native speakers of British English in the Louvain Corpus of Native English Conversation (LOCNEC). After uncovering a comparative overall overuse of the term in the Swedish L2 English, she then broadens her study to examine its occurrence with respect to both position and category. In doing so, she creates a functional typology of *well* with two main categories – speech management and attitudinal functions – and finds significant differences in the functions of *well* when uttered by the Swedish learners when compared to its use by native speakers of English.

Such contrasting patterns have clear implications for the foreign language class-room: "One cannot use *well* 'anywhere' and expect to sound English" (p. 112). Aijmer offers specific suggestions to help learners acquire a more strategic use of pragmatic markers, and also proposes numerous possibilities for further, related research.

Christiane Brand and **Sandra Götz** present a pilot study looking into whether there is any observable correlation between accuracy and fluency in the spoken English of native German speakers, and if so, how these variables relate to each other. They obtain their primary learner data from the error-tagged German version of LINDSEI, and add an extra element by using data from LOC-NEC as a baseline by which to compare learner fluency. Brand and Götz's primary aim, however, is the presentation of a multi-method approach for the investigation of such a correlation, involving three main components: 1) a quantitative analysis of error rate and temporal fluency variables (speech rate and frequency of filled and unfilled pauses) of all 50 learners in the LINDSEI subcorpus, 2) a detailed qualitative analysis of the output of five selected learners with the aim of comparing their accuracy and fluency rates to identify possible trends for a correlation, and 3) having 50 native speakers of English rate the overall degree of oral proficiency of the five learners. This last step allows for the investigation of possible links between learners' accuracy, fluency and perceived proficiency. The value of this article is thus two-fold. The preliminary results of the study open numerous avenues for future research and the multi-method approach demonstrated provides promising tools for such research.

John Osborne investigates the relationship between temporal fluency and informational content in both spoken L1 and L2 production, looking at characteristics of fluency across languages. He investigates the language of L1 and L2 speakers of French and English, having retrieved his data from the *PAROLE* corpus. This article is thus notable through broadening the scope of learner corpus research beyond the focus on English so typical of the field. Osborne develops a 'fluency index' – a composite measure of speech rate, quantity of pausing and mean length of run between hesitations – and looks at the scores of the L1 and L2 speakers in combination with measurement units for syntactic units, information units and utterance boundaries. His work has important ramifications for the development of less subjective criteria for fluency than those now found in guiding documents such as the Common European Framework of Reference for Languages.

This volume was a pleasure to read, from beginning to end, and researchers interested in language performance – how language is really used in spoken discourse – will benefit from all five articles as well as the informative introduc-

tion. I would, however, like to raise two issues deserving of more attention that came to mind while reading the book. The first concerns the rationale for the various comparisons in several of the articles between the production of native and non-native speakers. Justification for such comparisons is left unwritten, presumably because this type of research has become the norm. This is partially thanks to the original version of the Contrastive Interlanguage Analysis (CIA) approach designed to, among other things, facilitate the identification of overuse or underuse of linguistic phenomena in the language of non-native learners when compared to the language of native speakers. A possible implication then becomes that the native speaker language is superior (and always the target), something that is being questioned more and more often especially with respect to English, global language that it is. This issue has important ramifications for an understanding of what constitutes error, particularly given the definition offered in the introduction (cited above), which Gilquin and De Cock consider a 'fairly non-committal' use of the term. Defining error in terms of a native-speaker norm, however, might be less non-committal than is presupposed here. While comparison of native and non-native language varieties is certainly legitimate and valuable, explicitly addressing this issue in some way would be advisable. In her keynote presentation at the 2013 Learner Corpus Research conference (LCR2013), for example, Granger suggested modifying CIA through dropping the terms 'NS variety' and 'NNS variety' in favor of 'Reference Language variety' and 'Interlanguage variety'.

Second, corpus linguists tend to be either 'lumpers' (e.g. Aijmer) who treat the quantitative data gathered from their informants as a whole, or 'splitters' (e.g. Brand and Götz) who treat the contributions of individual informants separately. The lumpers generally have a particular challenge when it comes to statistical analysis, especially where the use of certain deceptively simple and commonly performed statistical tests is concerned. By way of example, Tottie (pp. 41–42) reports that 377 male informants accounted for 14,568 occurrences of *uh* and *um* in the 1,454,344 words in the BNC-DEM, whereas 418 female informants uttered these fillers a total of 18,405 times in 2,264,094 words; for easier comparison these frequencies are normalized per 1000 words. Tottie then reports that a chi-square test of independence indicates significance at the level of $p<0.0001$, being careful to remark in a footnote that all statistical calculations were carried out using the observed frequencies instead of the normalized frequencies. The upshot seems to be that men use these fillers more often than women.

This type of reporting has become fairly standard in the world of corpus linguistics, where researchers often work with data generating lots of numbers. It is

only natural to want to discover whether these numbers are significant in statistical terms, or merely the result of chance. Employing a test such as chi-square is particularly tempting, seeing as how there is any number of online calculators at our disposal. Plugging in the observed frequencies quickly reveals significance, or lack thereof. The problem, however, is that every statistical test has conditions that need to be met in order for the calculated results to be valid; for example, the chi-square test is only appropriate in the case of independent observations. In the instance noted above, unless there were 14,568 (or more) informants and each person uttered a filler no more than once, the observations are not independent. In short, the many *um*'s uttered by a single 'ummer' are linked; once someone starts umming, we should not be surprised if they keep umming. A statistical test may make it appear that the data at hand 'clearly support' a particular hypothesis, but if the conditions of that test were not met, its results are necessarily invalid.

Picking upon an example from one single researcher is admittedly a bit unfair, as this is a general problem in corpus linguistics. Indeed, this particular article was presumably subject to both double-blind peer review and editorial review without the issue having been raised. This problem with validity, however, becomes all too easy to spot in numerous pieces of corpus research once one first becomes aware of it (many thanks due to Bård Uri Jensen for his keynote address *A chi-square test showed that – or did it really?* at LCR2013). The responsibility for a solution belongs to individual researchers, editors, and peer reviewers, along with supervisors, universities, and that rarest of breeds: statisticians, especially those who double as linguists. They are perhaps among the first to roll their eyes at such dubious statistics, but have not yet managed to help many of us understand how to choose and apply appropriate statistical measures. Statistics are not witchcraft; surely they can be understood by the linguists who try to use them, if only statisticians could train themselves to speak 'human' in order to help us mend our ways. But this of course presupposes that we know enough to realize that we have a serious problem and need to seek help.

To sum up, this volume represents a significant contribution to the study of oral production of both L1 and L2 speakers, clearly demonstrating that so-called errors and disfluencies in spoken language provide valuable evidence about cognitive processing and should not be disregarded as merely performance or competence mistakes. The research here also demonstrates the importance of spoken corpus evidence, without which it would be difficult to investigate (or perhaps even notice) fillers and silent pauses. Perhaps most valuable is the wealth of ideas for further investigation as well as the clearly explained and tested methods that will undoubtedly be explored by future researchers.