Hesitation phenomena in the language production of bilingual speakers: The role of working memory

Carol Fehringer & Christina Fry
Newcastle University

This paper is an empirical investigation of the use of hesitation phenomena, specifically filled pauses (ums and ers), automatisms (sort of, at the end of the day), repetitions and reformulations, in both the mother tongue (L1) and second language (L2) of highly proficient adult bilingual speakers (English and German). Its purpose is to ascertain: i) whether speakers who are highly proficient in L2 produce an approximately similar amount of hesitation phenomena in both languages; and ii) whether the production of such elements (in both languages) is linked to working memory capacity. Results show that: i) despite high proficiency, speakers produced a higher overall rate of hesitation phenomena in their L2, indicating that there was an additional cognitive load imposed by working in L2; and ii) in each language there was an underlying negative relationship between memory capacity and the production of hesitation phenomena, implying that speakers with lower memory ability rely more heavily on such time-buying devices. Furthermore, it was shown that the individual types of hesitation phenomena produced by speakers in their L1 were carried over into their L2, which suggests that a speaker’s planning behaviour is mirrored in both languages.

Keywords: bilingual, hesitation, prefabricated utterance, speech production, working memory, L2

1 We are grateful to three anonymous Folia Linguistica reviewers for valuable comments.

ISSN 0165–4004, E-ISSN 1614–7308 © Mouton de Gruyter – Societas Linguistica Europaea
1. Introduction

One of the ways in which individuals differ from one another is in how successfully they can “manage” the language they speak (Fillmore 1979: 88, 91, 95). Some speakers are able to convey their messages clearly and succinctly, while others are more hesitant, prone to ums and ers, automatisms (often referred to as “filler” words and phrases), repetitions and corrections. Indeed, in extreme cases, such dysfluencies in the speaker’s output almost overwhelm the message that he/she is trying to convey, placing a very heavy computational cost on the listener.

It has been pointed out in the literature on language production that hesitation phenomena serve an important purpose in on-line processing, namely to buy time for the speaker to plan the next part of the utterance (Raupach 1980: 270), although there is much debate as to whether these elements are “automatic” (Levelt 1983) or deliberate “performance additions” (see Clark & Fox Tree 2002 for further discussion). What is interesting to bilingualism researchers is that such time-buying devices are thought to be particularly necessary when speaking a foreign language (Faerch & Kasper 1983: 216). Clearly, some speakers make more use of these devices than others, both in their native language (henceforth L1) and in their second language (L2), and it has been suggested that this is linked to working memory (henceforth WM) capacity: i.e. speakers with a low WM capacity need to buy more time and therefore use more hesitation phenomena than those with a high WM capacity (Temple 1997). However, no empirical evidence specifically demonstrating a link between WM and the use of hesitation phenomena has been put forward in the L2 literature. For L1, Daneman & Green (1986) and Daneman (1991) found evidence for a link between a high WM capacity and a greater degree of fluency in production, yet these studies looked not only at the use of hesitation phenomena but also at speech errors such as spoonerisms as a measure of general verbal fluency.

In the fields of second language acquisition and second language production much work has been done on comparing the use of pauses (both silent and “filled” pauses) in L1 versus L2, with most studies concluding that this is more prevalent in L2 (Wiese 1984, Pouliisse 1997). Similarly, the use of prefabricated utterances, which includes automatisms as well as more general idiomatic phrases and collocations, has been shown to be more widespread in L2 than in L1 (Bygate 1988, Wray & Perkins 2000). It should
be noted, however, that most of these studies concentrate on L2 learners at secondary school or undergraduate university level rather than on speakers of higher proficiency; therefore the increased use of hesitation phenomena in L2 is not surprising.

The purpose of this paper is twofold: firstly, to investigate the production of hesitation phenomena in the L1 versus the L2 of highly proficient bilingual speakers in order to ascertain whether, as demonstrated in the L2 learner studies, L2 speakers who are beyond the learner stage still make more extensive use of such elements in their L2 than in their L1, or whether the high proficiency of a speaker is reflected in his/her native-like use of hesitation phenomena. 2

Secondly, the paper provides an empirical test of the hypothesis that the use of hesitation phenomena (specifically ums and ers, automatisms, repetitions and reformulations) in spontaneous L1 and L2 production is linked to WM capacity, an idea that has been proposed for L1 by Daneman (1991) as part of the more general concept of verbal fluency, and has been implicit in some L2 studies although not explicitly formulated or supported by experimental data. Specifically, if hesitation phenomena are used to buy processing time, one would expect speakers with a low WM capacity to rely more heavily on the use of such phenomena than those with a high WM capacity. In other words, there should be a negative correlation between the use of hesitation phenomena and WM ability.

2. Working memory

WM is a cognitive system that handles the temporary storage and processing of information. There are many different models of WM (see Miyake & Shah 1999), yet particularly influential is Baddeley’s (1996) model which consists of a CENTRAL EXECUTIVE controlling attention, and two slave systems: a PHONOLOGICAL LOOP to hold and manipulate speech-based information, and a VISUOSPATIAL SKETCHPAD functioning in the same way for visual images. The Central Executive is a fractionable system that is involved more

2 The term “bilingual” is used here to refer to the fact that the speakers are now highly proficient in their L2 (see §4 for further details). They did not, however, acquire their L2 as a child in a family context but rather learned it in school as a foreign language.
in encoding than in retrieval (Baddeley 2001: 117). A fourth WM component, the Episodic Buffer, has recently been proposed (Baddeley 2000) as a store capable of operating beyond the timescale assumed for the slave systems. This component can temporarily hold and manipulate information such as that involved in the comprehension of a prose passage, which involves the activation of existing structures in long-term memory (Baddeley 2000: 420). With the length of WM in the form of the Phonological Loop being presumed to be approximately 1 to 2 seconds (Baddeley 1986: 93), and speech being delivered at the rate of two to three words per second (Levelt 1999: 112), it may be seen that the number of words that would be expected to be held in WM corresponds closely to Miller’s (1956) magical number of seven plus or minus two.

WM has been shown to play an important role in language processing, particularly in the area of comprehension: see, for instance, Daneman & Carpenter’s (1980) study which demonstrates that the ability to comprehend complex syntactic constructions is correlated with WM capacity, as measured by reading span tests. In the L2 literature, most work has concentrated on the relationship between WM and reading comprehension abilities in language learners (Harrington & Sawyer 1992, Hummel 2002, Walter 2004), and has argued that reading comprehension performance is often positively correlated with reading span tests.

Fewer empirical studies have investigated the relationship between WM and language production. For L1, Daneman & Green (1986) and Daneman (1991) argue that the facility with which a speaker can produce an appropriate lexical item on-line is related to that speaker’s ability to coordinate the processing and temporary storage functions of WM. Similarly, for L2, Temple (1997) appeals to WM as an important constraint on fluency, although her study does not test WM capacity empirically. Kroll et al. (2002) argue that WM is not as important as linguistic proficiency in determining lexical fluency in L1 and L2, but WM does play a role in tasks that involve greater computation and processing, such as translation, for example. Indeed, many of the L2 studies involving WM in language production focus on the relationship between WM capacity and the performance of professional simultaneous interpreters, arguing that although WM is a critical subskill for simultaneous interpreting (Christoffels et al. 2006), it is word knowledge, rather than WM capacity, that is more directly related to the ability of interpreters to comprehend and produce simultaneously (Padilla et al. 2005).
3. Hesitation phenomena in language production

Under generally accepted models (e.g. Garrett 1980, Levelt 1999), language production is assumed to be incremental, allowing limited parallel processing to occur across stages, with higher levels delivering information concerning only part of the element under construction piecemeal to levels lower in the hierarchy, before the whole representation of that element is complete at the higher level (Berndt 2001: 379). In order to plan and organise output, information must be retrieved from long term memory, and integrated in real-time with other information passing through WM (Olson 1973: 156). As Levelt points out (1999: 112), there is no more complex cognitive-motor activity than speaking, since a speaker must undertake a considerable amount of concurrent processing, simultaneously formulating several elements at different levels (e.g. monitoring and repairing content, grammaticality, and articulation; maintaining coherence and cohesion; avoiding repetition; adhering to the topic; taking into account the needs of the listener; and obeying the discourse requirements of his/her culture). It is hardly surprising then that the speaker needs to buy some time now and then in order to continue with his/her utterance effectively. The elements that the speaker uses to do this are outlined in the following subsections.

3.1. Automatisms

The designation AUTOMATISM is used here as an umbrella term to cover expressions which are habitual, formulaic and generally semantically vacuous. These include idioms, proverbs, expletives and recurring phrases of relatively fixed forms (see van Lancker 1993, 2004). Some common examples are at the end of the day, all’s well that ends well, good heavens!, and expressions such as you know, sort of, actually, used as fillers. Alternative terms that have been used to describe these forms are FORMULAIC SEQUENCES or PREFABRICATED UTTERANCES (Wray & Perkins 2000). Although it is advantageous for a person to have a large repertoire of ready-made responses to a wide range of situations, over-reliance on such formulaic language is seen as a disadvantage, implying that the speaker is unable to respond creatively (Fillmore 1979: 94). Indeed, such stereotyped language is a common feature of language disorders, as demonstrated by Lum & Ellis (1999), whose aphasic subjects had less difficulty in processing set idioms.
such as *don’t beat about the bush* than similar non-idiomatic phrases (e.g. *don’t dig behind the bush*).

Although it is accepted that automatisms can often fulfil a sociolinguistic and/or pragmatic function, for instance they may demonstrate politeness or signal social cohesion with the listener (Wray & Perkins 2000), it is the time-buying function of such elements that is of interest here. Indeed, the use of automatisms in this way has also been recognised in some of the psycholinguistic work on hesitation phenomena. For example, Möhle (1984: 46) observes that German speakers use hesitations strategically to gain time, and this is particularly effective if they combine them with what she refers to as “parenthetic remarks”, such as *also* ’well’, *wie sagt man* ‘how do you say’. It must be noted, however, that Möhle does not include these automatisms in the data analysis of her experiments, which largely concentrate on the occurrence of filled and silent pauses, as is the case with many of the hesitation studies of the 1980s (Deese 1980, Wiese 1984). One exception is Raupach (1984), who investigates a certain type of automatism, namely the overuse of the standardised modifiers *assez* ‘quite’ and *vraiment* ‘really’ in German learners of French, and comes to the conclusion that the use of these modifiers relieves the speaker from having to resort to subsidiary hesitation phenomena such as filled and unfilled pauses (Raupach 1984: 132). He asserts that this integration of formulaic sequences in order to avoid non-fluent devices is a form of planning behaviour comparable to that of native speakers and therefore gives the impression of greater idiomaticity. Thus, these “islands of reliability” (Raupach 1984: 135) not only facilitate processing in L2 but also make the learner sound more native-like.

Finally, in addition to recognisable, culture-wide automatisms there also exist idiolectal automatisms, or VERBAL TICS, by which individuals can achieve economies in computation (Olson 1973: 155). Some examples from our participants include the overuse of *perhaps*, *only just* and *frequently*. Such elements also need to be taken into account in any study of hesitation phenomena.

3.2. Mazes

*Maze* is the general term used in this paper to cover instances of repetitions, corrections or expansions of words or longer strings, and is exemplified by the following utterance taken from one of the participants in the present study: *I guess it— it must be a—, you know, a crucial er a— a crucial factor, a—*
bu–, you know, I just don’t understand tho– those sort of– er those sort of issues.

Mazes require the hearer to expend extra memory resources to construe, compute, and then reject each of the mazes, while holding on-line that which has gone before, in order to marry it up with the repaired version. Similarly, utterances which are simply abandoned by the speaker part-way through waste the listener’s computational effort on phrases or clauses which are destined never to be finished.

Mazes, including abandoned utterances, can be construed as evidence of the speaker’s insufficiency in utterance planning, in which WM plays an important role. Self-monitoring takes place constantly, with two main functions: checking that what has been said matches with what was intended; and detecting speech errors or syntactic mistakes, while maintaining the appropriate prosodic aspects of speech (Levelt 1983: 49–50). This view of self-monitoring makes the assumption of a perceptual loop, which allows speech to be checked, and problems to be detected and quickly corrected: the centrally-controlled nature of self-monitoring and repair, then, means that it is liable to WM limitations (Levelt 1983: 50).

3.3. Filled pauses

The function of FILLED PAUSES, in English usually er or um, is similar to that of automatisms in that they buy the speaker time to plan his/her utterance.\(^3\) As with other hesitation phenomena, the excessive use of filled pauses can place a considerable strain on the hearer, as exemplified by this sentence from a participant in the present study: er er action now will er um mean that you don’t have to er er work your knackers off later on!

In sum, the elements examined in this paper all have in common that they buy time for the speaker, yet, at the same time, impose a computational cost on the hearer. All these elements form a part of normal language

\(^3\) We are using the traditional term “filled pauses” here, although it is recognised that, at present, there is much debate about the status and function of um and er in spontaneous speech. Clark & Fox Tree (2002) assert that they should not be seen as errors or even filled pauses but as proper English words, as interjections, which are deliberately used to buy time. By contrast, O’Connell & Kowal (2005) reject the time-buying function of um and er, and their status as interjections, although it should be noted that O’Connell & Kowal’s work is largely based on written rather than on spoken corpora.
production, but there is seemingly a threshold over which their use is regarded as pathological. Indeed, Montgomery (1992: 523) demonstrates that the presence of filled pauses, mazes, and automatisms such as you know, I mean reliably differentiates the performance of adolescents with mild to moderate language learning disability (LLD) from that of controls.

In the L2 literature, many studies on hesitation phenomena have also looked at temporal variables such as silent or “unfilled” pauses, speech rate, speech-to-silence ratio etc. as a measure of a speaker’s general fluency. The problem with pause studies, however, is that there has been disagreement in defining how long an unfilled pause should be in order to be counted as such (Raupach 1983: 201), particularly as unfilled pauses allow the speaker to breathe, and occur naturally at certain constituent boundaries. With regard to the hearer’s encoding of speech, silent pauses are largely ignored, whereas filled pauses, false starts, corrections and repetitions (i.e. mazes) are “difficult and unpleasant to listen to” (Deese 1980: 79–80). This is due to the fact that the articulation of initial sounds, whether these be words as in the case of mazes, or simply ums and ers, activates a host of potential representations in the hearer’s mind (see Marslen-Wilson & Tyler’s 1980 cohort model of spoken word recognition). Such red herrings can be an irritation to the hearer, and it is in the speaker’s interest to avoid these if he/she wants to be well received. Indeed, courses in public speaking train speakers to avoid automatisms and to replace ums and ers with silent pauses (Clark & Fox Tree 2002: 98). With silent pauses, the cost to the hearer would only be significant if they were extremely long and/or frequent, which may well be observed in L2 learners of low proficiency but was certainly not the case with our participants.

3.4. Hesitation phenomena in L1 versus L2

If, as hypothesised above, the use of hesitation phenomena is at least partially related to WM capacity, it is likely that speakers with the lowest WM capacity will produce the greatest amount of hesitations, automatisms, repetitions etc. in their spontaneous speech. This relationship between memory and hesitation phenomena should be the case regardless of the language being

---

1 Controls are participants who are closely matched to the experimental group, but who are not exposed to the variable under investigation.
spoken, as WM resources are generally assumed to be a global cognitive attribute.

Naturally, questions of L2 proficiency also come into play, and numerous studies have shown that L2 learners rely more heavily on various time-buying elements than do native speakers. This has been shown to be the case with automatisms or prefabricated utterances (Bygate 1988) and with filled pauses and mazes (Lennon 1990).

Yet what about L2 speakers who are no longer learners, at least in the sense of classroom learning, but who have reached near-native competence? One might expect that the closer to native ability an L2 speaker’s production came, the more similar their behaviour would be in the two languages with respect to hesitation phenomena. Having said that, one must take into account that the task of producing one’s L1 is not wholly identical to that of producing one’s L2, even at the highest levels of proficiency. For instance, a proportion of WM capacity will necessarily be taken up by the inhibition of interference from L1, possible differences in lemma retrieval, and structural differences between the syntax of L1 and L2 (Green 1986). Thus, it is hypothesised that a speaker’s performance in L2 should exhibit a slightly greater rate of hesitation phenomena than that shown in the mother tongue, as he/she will need to buy a little more processing time in L2 than in L1.

In other words, across speakers, it is predicted that there will be a continuum of hesitation phenomena in L1 (i.e. the rate of occurrence of elements such as automatisms, filled pauses and mazes), and a continuum of hesitation phenomena in L2, and that the total L2 rate will be slightly higher than that of L1.

4. Method

This analysis draws upon a corpus of data collected from a group of bilingual speakers. The memory tests and all language stimuli were presented aurally, in order to reflect naturalistic language interaction. The subjects, data collection, memory tests, and data analysis are described below.

4.1. Participants

The study elicited spontaneous speech from 20 participants drawn from among postgraduates and staff of Newcastle University. The data were
collected at the university, during 2005. There were 10 males and 10 females, with a mean age of 34.1 years (SD 10.6). The ages of subjects ranged from 22 to 63 years. The lower bound was such as would ensure that almost all native language acquisition had been achieved, and the higher bound was chosen so that there should be no (or very little) effect of the memory decrements associated with normal ageing.

Ten participants had L1 German and L2 English, the other 10 had L1 English and L2 German. All of them had learned, rather than acquired, their L2. That is, they had spent their childhood in the country of their L1 and had learned their L2 at school as a foreign language. The English native speakers all had a First Class university degree in their L2 and were either undertaking or had completed postgraduate work in their L2. The German native speakers all had an IELTS (International English Language Testing System) score of 9 for their L2, which classes them as ‘Expert users, having a fully operational command of the language: appropriate, accurate and fluent with complete understanding’ according to the test’s descriptors. The IELTS scores were chosen as a proxy measure, owing to the lack of comparability between English and German university degree results. This high level ability group was targeted in order to provide (as near as possible) a homogeneous set of subjects, thereby minimising the variability endemic to earlier stages of L2 learning.

4.2. Data collection

The procedures and stimuli used to elicit data, which were previously piloted in studies of monoglot speakers (Fry 2002), are given in Appendix I. With the exception of the pictures used to elicit story telling, all stimuli were presented entirely in the oral/aural modality, to reflect the demands of normal interpersonal communication. Spontaneous speech was elicited in a dyadic interview: Fehringer conducted the interview in German, and Fry in English. All subjects were exposed to the same stimuli, first in their L1 and then in their L2, with the double interview session taking approximately 40

5 It is recognised, however, that the homogeneity of the group cannot be absolute, as there will always be subtle individual differences in the speakers’ proficiency that are difficult to measure; not to mention differences arising from relative L2 exposure (for instance, the German native speakers are living in their L2 environment while the English are not).
Hesitation phenomena in the language production of bilingual speakers

to 45 minutes in toto. Care was taken to minimise the anxiety of the participants (see Dewaele 2001), by chatting informally to put them at their ease and ensuring them that this was not a test of L2 proficiency or correctness. Unlike the language tasks of Daneman (1991), and indeed of many other studies, there was no time-limit placed on our participants’ responses beyond the normal pragmatic expectations, thus maximising the naturalness of the interaction.

All the language stimuli were translated, and therefore parallel versions were available in each language. All participants were presented with the items from version A in their L1, and from version B in their L2. Stimuli consisted of two sets of five questions; the first set, the WHAT AND WHY QUESTIONS, centring round general world knowledge (e.g. why is it good to vote in secret?), and the second set, the SPECULATION QUESTIONS, being more conjectural (e.g. if you won a million pounds on the lottery, how would it change your life?). In addition, in order to elicit a more narrative style, we included a productive story-telling task in which subjects were invited to pretend that a stimulus picture was a still from a film and to tell the story of the film. The stimulus was a photograph, which participants chose freely from among six unrelated pictures. The story-telling task was entirely spontaneous and unconstrained, since the participants were free to create any story that they chose, and no guidance was given to them regarding schemas, characters or plot development.

Elicited speech consisted of responses by each subject, in reply to the above-mentioned stimuli, the aim being to represent as near as possible normal speech patterns. The experimenters’ input was limited to phatic communication, nodding and smiling, rather than entering into a full conversation, in order to maximise the comparability between subjects’ output, which would unquestionably have been compromised in free-flowing conversation.

The entire double interview was recorded on audio tape for later transcription and memory test scoring. Transcription was orthographic. Filled pauses were indicated using M (for variants of um) and R (for variants of er/äh), and mazes were signalled by a dash following the last element of the maze item (e.g. he knocked at– on the door). Fry transcribed the English data and Fehringer the German data, and each checked the others’ transcriptions. Any disagreements were resolved by joint discussion.
4.3. Memory tests

In order to compare participants’ performance across different memory test types, as well as between languages, we used two tests of memory: AURAL WORKING MEMORY SPAN (Fry 2002), henceforth referred to as WM span, and the STORY RETELLING SUBTEST from the Adult Memory and Information Processing Battery (AMIPB) (Coughlan & Hollows 1985), each of which takes place over a different timescale, ranging from some 8 seconds up to about 40 seconds (plus recall time). This is considerably longer than the 1½ to 2 seconds that is normally considered to be the length of WM, specifically the Phonological Loop component, which a number of L2 studies have tested (see e.g. Brown & Hulme 1992, Service 1992, Chee et al. 2004). Although, as these studies show, the Phonological Loop is clearly implicated in lexical access and the production of individual clauses, it cannot contain sufficient material to account for the production of connected discourse. As the focus of this paper is at the level of discourse, memory measures restricted to the Phonological Loop were not thought to be appropriate.

The rationale for choosing the WM span and story retelling tests was that each is used in different spheres and for different purposes. The WM span task is based on the paradigm of Daneman & Carpenter’s (1980) reading/listening span and Baddeley et al.’s (1985) WM span tests, which have been used in psychological studies on reading skill, comprehension and working memory. The AMIPB Story Retelling subtest (henceforth referred to as AMIPB) is a standard neuropsychological test which is widely used in clinical practice. It is designed to assess immediate registration of verbal information, and to reflect the memory demands of everyday life (Coughlan & Hollows 1985: 19). Although the task is not generally used in psychological research, it was included in the present study because it operates on a longer timescale than is commonly attributed to WM, and we take this as the time frame over which the production of spontaneous discourse must function.

The WM span task involves the participant being presented orally with sets of sentences. At the end of each sentence, the subject performs an oral judgement on the truth or falsity of that sentence; while at the end of a set, he/she recalls aloud, in any order, the final word of each sentence within that set. The sets are of increasing length (from two up to five sentences) and three sets are presented at each length. Sentences consist of short declarative statements concerning world knowledge, with truth/falsity balanced within sets and across the test overall. The truth/falsity judgements are not scored,
being a distraction task to prevent subvocal rehearsal and to ensure conceptual manipulation. The WM span score represents the longest span that the participant can consistently achieve (e.g. 4.3 indicates recall of one set of five final words). The procedure is identical for L1 and L2, except that different sentences are used in each language.

In the AMIPB test, the participant hears a story read aloud to him/her, and immediately recounts that story orally in as much detail as he/she can recall. The task of story recall is a supraspan task which, by design, overloads the participant’s memory with data (Lezak 1995: 456). In order to overcome this, the participant must therefore organise the data into schemata, which ensures manipulation. In accordance with the scoring guidelines in the AMIPB battery, any correctly recalled idea (or an accurate paraphrase thereof) is awarded two points, any vaguely or partially recalled ideas receive one point, and points are then totalled. The AMIPB exists in two forms: in this study, subjects heard one of these stories presented in English, and the other story in German. The format of the test is identical in both languages, with different (yet comparable) stories used for each. The scores for the two different stories can be compared using Z scores (i.e. measures of the distance in standard deviations of a sample from the mean).

4.4. Data analysis

All the output resulting from the language stimuli for each subject was transcribed, yielding a total corpus of 27,886 words (13,646 in the L1s, 14,240 in the L2s). The instances of automatisms, filled pauses and mazes were counted, and the number per hundred intelligible words was computed.

Examples of filled pauses in the data are primarily forms of er and um (with the addition of some idiosyncratic sounds, e.g. fff). These are common to both English and German. Drawls (the elongation of a syllable as a form of hesitation), which occur in languages such as French (Möhle 1984), are not a standard feature of either English or German production and therefore did not occur in the recorded data.

Mazes subsume exact repetitions (e.g. I don’t have the- the knowledge), expanded repetitions (e.g. earn money to support the family- to help support the family) and amended repetitions (e.g. if you- if I had to choose). Each

I.e. covert, silent “speech” that occurs when a participant is attempting to maintain an experimental stimulus in active memory.
repetition or reformulation of a string within a functional category (CP, IP, DP) was counted as one maze. Abandoned utterances were also counted under this heading (e.g. *I think there are too many of those reality shows but I’m not…. On the other hand there are one or two good ones*).

The elements that were the most problematic to categorise were the automatisms, as it was difficult to determine at which point a particular word or phrase should be classified as such. The most straightforward cases were the semantically vacuous fillers, e.g. *you know, like, I mean, stuff like that*. As the literature on formulaic language points out, the difference between formulaic and non-formulaic language is best represented as a continuum rather than a dichotomy (Pawley & Syder 1983: 192, van Lancker 1993: 217, 2004: 4,), which means that any attempt to establish a cut-off point must necessarily be arbitrary.

For instance, included under the heading of formulaic language are common collocations, roughly defined as the habitual co-occurrence of individual lexical items, such as a *very big place* or a *good Catholic education* (Yorio 1989: 67). These phrases have a clear (i.e. semantically transparent) meaning, in contrast to semantically opaque idioms such as, for example, *beat about the bush* and proverbs, e.g. *a stitch in time saves nine*, which must be analysed as wholes. It is not clear, however, at what point a certain combination of words actually becomes a collocation, and the difficulty in providing a clear definition of the term, or of similar concepts such as LEXICALISED SENTENCE STEMS, has been pointed out in the literature (Pawley & Syder 1983: 212). For this reason, the present study makes a distinction between collocations on the one hand, and idioms/proverbs on the other, and counts only the latter as automatisms, although it is recognised that this restriction may lead to the number of automatisms being somewhat under-represented (see also Sinclair 1991: 115–121).

Indeed, if we were to include collocations under our heading of automatisms, we would run into a further methodological problem: phrases that are classified as collocations in the English studies, e.g. *summer holidays* (Yorio 1989: 67), *bullet point* (Wray & Perkins 2000: 16) are often expressed as compound words in German, e.g. *Sommerferien, Nummerierungspunkt*. Similarly, some studies (e.g. Yorio 1989) count English verb + preposition combinations, e.g. *to bring up*, as collocations, most of which have one-word equivalents in German. If we counted such collocations in English as automatisms, the English output of our speakers would appear to contain inherently more automatisms than their German output, which means that
any statement about the relative rate of automatisms in L1 versus L2 would
be rendered meaningless.

Thus, our definition of the category automatism is restricted, firstly, to
semantically vacuous fillers, e.g. you know, as it were, kind of thing (German
halt, also, irgendwie) and, secondly, to idioms, proverbs and other clichéd
predictable strings, such as at the end of the day, when all’s said and done
(German im Großen und Ganzen). In addition, it was often the case in our
interviews that individual speakers overused a particular adverb or adverbal
phrase so that it became an idiolectal automatism. Examples are the overuse
of häufig ‘often, frequently’ in the story telling task by a German native
speaker, the use of vielleicht ‘perhaps’ by another German native in response
to most of the interview questions, and extensive use of sehr ‘very’ by an
English speaker of German. Similarly, I think, I guess, I suppose (German ich
glaube, ich denke) were often overused. If a speaker used such a word or
phrase three times or more, it was counted as an idiolectal automatism.

5. Results

All data were examined for normality and for the presence of outliers (i.e.
 extreme deviations from the mean). Three participants produced one score
that was at the extreme of the distribution (for one variable each, and in one
language only). However, these were not sufficiently severe to be classified as
outliers, and all participants were therefore included in the data set.

As anticipated, across speakers, there was a notable amount of variation
in the rate of total hesitation phenomena used, calculated per hundred
intelligible words, as shown in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hesitation phenomena in L1</td>
<td>13.95</td>
<td>4.48</td>
<td>8.69</td>
<td>17.92</td>
</tr>
<tr>
<td>Hesitation phenomena in L2</td>
<td>17.39</td>
<td>5.28</td>
<td>11.39</td>
<td>31.25</td>
</tr>
</tbody>
</table>

\[n = 20\]

Table 1. Descriptive statistics for total hesitation phenomena per 100 words in L1 and L2

The production of total hesitation phenomena in L1 correlated significantly
with that in L2 \(r = .75, p = .000\), two-tailed).

There was a significant difference between the rate of hesitation
phenomena used in the two languages \(t = -4.32, df = 19, p = 0.000\), two-
tailed), with L2 showing a significantly higher rate than L1, as demonstrated in the bar chart below.

![Bar chart showing mean rate of hesitation phenomena in L1 versus L2](image)

*Figure 1. Mean rate of hesitation phenomena in L1 versus L2*

The amount of hesitation phenomena produced also varied within the individual categories, as demonstrated in Table 2.

<table>
<thead>
<tr>
<th>Category</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatisms L1</td>
<td>5.17</td>
<td>2.57</td>
<td>.83</td>
<td>11.09</td>
</tr>
<tr>
<td>Automatisms L2</td>
<td>6.09</td>
<td>2.28</td>
<td>2.33</td>
<td>11.72</td>
</tr>
<tr>
<td>Filled pauses L1</td>
<td>5.11</td>
<td>1.94</td>
<td>1.11</td>
<td>9.39</td>
</tr>
<tr>
<td>Filled pauses L2</td>
<td>6.45</td>
<td>4.06</td>
<td>.56</td>
<td>19.74</td>
</tr>
<tr>
<td>Mazes L1</td>
<td>3.68</td>
<td>2.32</td>
<td>1.89</td>
<td>9.78</td>
</tr>
<tr>
<td>Mazes L2</td>
<td>4.85</td>
<td>2.3</td>
<td>2.6</td>
<td>10.4</td>
</tr>
</tbody>
</table>

[n = 20]

*Table 2. Descriptive statistics for categories of hesitation phenomena per 100 words*

An investigation of the individual hesitation categories shows a strong correlation between their use in L1 and L2, as demonstrated in Table 3.
Hesitation phenomena in the language production of bilingual speakers

### Table 3. Correlation matrix for individual categories in L1 and L2

<table>
<thead>
<tr>
<th></th>
<th>Automatisms in L1</th>
<th>Automatisms in L2</th>
<th>Mazes in L1</th>
<th>Mazes in L2</th>
<th>Filled pauses in L1</th>
<th>Filled pauses in L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automatisms in L1</td>
<td>1.000</td>
<td>.573**</td>
<td>.079</td>
<td>.120</td>
<td>.318</td>
<td>.206</td>
</tr>
<tr>
<td>Correlation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sig.</td>
<td></td>
<td>.008</td>
<td>.742</td>
<td>.614</td>
<td>.172</td>
<td>.383</td>
</tr>
<tr>
<td>Automatisms in L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.573**</td>
<td>1.000</td>
<td>-.072</td>
<td>-.210</td>
<td>.413</td>
<td>.170</td>
</tr>
<tr>
<td>Sig.</td>
<td>.008</td>
<td></td>
<td>.763</td>
<td>.373</td>
<td>.070</td>
<td>.473</td>
</tr>
<tr>
<td>Mazes in L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.079</td>
<td>-.072</td>
<td>1.000</td>
<td>.661**</td>
<td>.013</td>
<td>.197</td>
</tr>
<tr>
<td>Sig.</td>
<td>.742</td>
<td>.763</td>
<td></td>
<td>.002</td>
<td>.956</td>
<td>.405</td>
</tr>
<tr>
<td>Mazes in L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.120</td>
<td>-.210</td>
<td>.661**</td>
<td>1.000</td>
<td>-.259</td>
<td>-.005</td>
</tr>
<tr>
<td>Sig.</td>
<td>.614</td>
<td>.373</td>
<td>.002</td>
<td></td>
<td>.270</td>
<td>.983</td>
</tr>
<tr>
<td>Filled pauses in L1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.318</td>
<td>.413</td>
<td>.013</td>
<td>-.259</td>
<td>1.000</td>
<td>.723**</td>
</tr>
<tr>
<td>Sig.</td>
<td>.172</td>
<td>.070</td>
<td>.956</td>
<td>.270</td>
<td>.</td>
<td>.000</td>
</tr>
<tr>
<td>Filled pauses in L2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correlation</td>
<td>.206</td>
<td>.170</td>
<td>.197</td>
<td>-.005</td>
<td>.723**</td>
<td>1.000</td>
</tr>
<tr>
<td>Sig.</td>
<td>.383</td>
<td>.473</td>
<td>.405</td>
<td>.983</td>
<td>.000</td>
<td>.</td>
</tr>
</tbody>
</table>

[n = 20] ** Correlation is significant at the 0.01 level (2-tailed)

All the individual categories in L1 correlate with their L2 equivalents at the .01 level of significance, but there are no significant correlations across categories. The proportions of the individual hesitation categories remain remarkably stable across languages, as shown in the pie charts below.

![Figure 2. Categories of hesitation phenomena in L1.](image-url)
There was no significant difference between the amount of automatisms produced in L1 and L2 ($t = -1.81$, $df = 19$, $p = .085$), nor between the amount of filled pauses in L1 and L2 ($t = -2.01$, $df = 19$, $p = .059$). However, mazes did differ significantly between those produced in L1 and in L2 ($t = -2.76$, $df = 19$, $p = .012$), and examination of the data showed the greater number to be produced in L2.

There was a significant difference between languages in each of the genres (What and why questions: $t = -3.34$, $df = 19$, $p = .003$; Story telling: $t = -3.1$, $df = 19$, $p = .006$; Speculation questions: $t = -3.02$, $df = 19$, $p = .007$). In each case, a higher number of hesitation phenomena was produced in L2, as shown in Table 4.

<table>
<thead>
<tr>
<th>Genre</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>What &amp; why L1</td>
<td>13.02</td>
<td>5.98</td>
<td>6.14</td>
<td>29.72</td>
</tr>
<tr>
<td>What &amp; why L2</td>
<td>16.26</td>
<td>6.68</td>
<td>7.73</td>
<td>32.99</td>
</tr>
<tr>
<td>Story telling L1</td>
<td>12.83</td>
<td>5.25</td>
<td>2.80</td>
<td>24.62</td>
</tr>
<tr>
<td>Story telling L2</td>
<td>14.94</td>
<td>6.80</td>
<td>3.03</td>
<td>30.30</td>
</tr>
<tr>
<td>Speculation L1</td>
<td>15.04</td>
<td>4.71</td>
<td>9.14</td>
<td>25.49</td>
</tr>
<tr>
<td>Speculation L2</td>
<td>18.57</td>
<td>4.96</td>
<td>13.16</td>
<td>29.41</td>
</tr>
</tbody>
</table>

[n = 20]

Table 4. Descriptive statistics for different genres
Investigation of the rate of hesitation phenomena in the three genres within L1 shows no significant difference ($\chi^2 = 2.7$, df = 2, p = .259). There is a significant difference in the rate of hesitation phenomena in the three genres within L2 ($\chi^2 = 8.4$, df = 2, p = .015), and examination of the data shows that the greatest number is produced in the Speculation questions in L2.

WM scores varied between speakers, as shown in Table 5. The AMIPB results are shown as Z scores, because of the need to compare different stories in the two versions.

<table>
<thead>
<tr>
<th>Test</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMIPB in L1</td>
<td>0.07</td>
<td>1.02</td>
<td>-2.08</td>
<td>1.68</td>
</tr>
<tr>
<td>AMIPB in L2</td>
<td>0.03</td>
<td>1.03</td>
<td>-2.64</td>
<td>1.48</td>
</tr>
<tr>
<td>WM span in L1</td>
<td>3.69</td>
<td>0.71</td>
<td>2.6</td>
<td>5</td>
</tr>
<tr>
<td>WM span in L2</td>
<td>3.27</td>
<td>0.65</td>
<td>2.3</td>
<td>5</td>
</tr>
</tbody>
</table>

[n = 20] Table 5. Descriptive statistics for memory test scores

The scores for AMIPB in L1 and L2 correlated significantly, as did those for WM span in L1 and L2. There was no significant correlation between AMIPB and scores on WM span.

<table>
<thead>
<tr>
<th>AMIPB in L1</th>
<th>AMIPB in L2</th>
<th>WM span in L1</th>
<th>WM span in L2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation</td>
<td>.654**</td>
<td>.192</td>
<td>.380</td>
</tr>
<tr>
<td>Sig.</td>
<td>.002</td>
<td>.418</td>
<td>.099</td>
</tr>
<tr>
<td>Correlation</td>
<td>.192</td>
<td>.057</td>
<td>.629**</td>
</tr>
<tr>
<td>Sig.</td>
<td>.418</td>
<td>.811</td>
<td>.003</td>
</tr>
</tbody>
</table>

[n = 20] **Correlation is significant at the 0.01 level (2-tailed) Table 6. Correlations between measures of memory in L1 and L2
There was no significant difference between WM scores in L1 and in L2 as measured by AMIPB (t = 0.21, df = 19, p = 0.839, two-tailed), but the difference between scores in L1 and L2 on WM span was significant (t = 3.15, df = 19, p = 0.005).

The correlation coefficients between the two memory tests in each language and hesitation phenomena in L1 and L2 are shown in Table 7.

<table>
<thead>
<tr>
<th>Memory test</th>
<th>Hesitation phenomena in L1</th>
<th>Hesitation phenomena in L2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson correlation</td>
<td>significance</td>
</tr>
<tr>
<td>AMIPB in L1</td>
<td>- 0.383</td>
<td>0.096</td>
</tr>
<tr>
<td>AMIPB in L2</td>
<td>- 0.238</td>
<td>0.312</td>
</tr>
<tr>
<td>WM span in L1</td>
<td>- 0.032</td>
<td>0.892</td>
</tr>
<tr>
<td>WM span in L2</td>
<td>- 0.473*</td>
<td>0.035</td>
</tr>
</tbody>
</table>

[n = 20] *Correlation is significant at the 0.05 level (2-tailed)

Table 7. Correlations between WM and hesitation phenomena

AMIPB in L1 demonstrates a significant correlation with hesitation phenomena in L2, and a trend with those in L1, while with AMIPB in L2 a slight trend is observable in both languages, as is illustrated in the scattergram below.

Figure 4. Scattergram of AMIPB and hesitation phenomena in L1 and L2
Correlations between hesitation phenomena in each language and WM span are all negative, but those with the measure in L1 are negligible, while those with WM span in L2 show significance or a trend towards it.

6. Discussion

The results above demonstrate that the use of hesitation phenomena does indeed vary from speaker to speaker, and that there is a significant correlation between the amount produced in L1 and L2 ($r = .746$, $p = .000$). There is, however, a significant difference between the rate of hesitation phenomena between languages, namely, there is significantly more produced in L2 ($t = -4.317$, $df = 19$, $p = .000$).

In second language research there has been much emphasis placed on the notion of the L2 speaker’s “cognitive deficit” (Paradis 1994), arguing that L2 speech production needs more WM, with attention directed at virtually every stage of processing. It is unclear, however, what the exact nature of this deficit is (Dewaele 2002: 232). In the case of L2 LEARNERS, particularly at the lower levels of proficiency, one could argue, as Poulisse (1997: 202–203) does, that L2 knowledge is not complete: learners have fewer words at their disposal, and their L2 grammatical knowledge is underdeveloped, which means that they often use compensatory strategies to solve lexical and grammatical problems (one of which might be the use of hesitation...
phenomena to buy time). When L2 proficiency improves, however, the compensatory strategies are less evident. Indeed Lennon (1990) demonstrated, in his study of four advanced German learners of English, that the number of repetitions, and filled and silent pauses, produced in the L2 gradually decreased as the learners’ linguistic proficiency improved.

The fact that the participants in our study produced more hesitation phenomena in L2 than in L1, despite having long since passed the learner stage and being highly proficient speakers of L2, suggests that differences in fluent speech between L1 and L2 are not simply a function of proficiency. One can reach an extremely high level of ability in L2 grammar and syntax, for example, yet it is still not quite one’s native language. Our speakers have demonstrated, at least with regard to the use of hesitation phenomena, the correctness of the commonly-held view that L2 users rarely reach a level of fluency approaching that of the native speaker (Harrington 1992: 127).

In addition to the rate of occurrence of hesitation phenomena as a whole in L1 and L2, we also considered whether the type of elements used differed between the two languages. From inspection of our data it became apparent that, in their L1, many speakers had a preference for some types over others: e.g. some speakers who produced a considerable number of filled pauses had few automatisms and mazes, whereas others produced a preponderance of automatisms and very few filled pauses (see Appendix II for each participant’s scores). Our results also demonstrate that the individual categories that comprise hesitation phenomena as a whole correlated significantly across languages within each category (automatisms: $r = .57$, $p = .008$; filled pauses: $r = .72$, $p = .000$; mazes: $r = .66$, $p = .002$). However, they did not correlate significantly across categories.

These findings largely support Raupach’s (1983: 207–208) observation that many factors that constitute a learner’s fluency in his/her L1 are liable to occur, in one form or another, in the learner’s L2 performance, and that there is a general tendency for language learners to transfer their “planning dispositions” (Möhle & Raupach 1989: 210) to the second language. Similarly, Wiese (1984: 22), in his analysis of temporal variables in advanced German learners of English, argues that between language production in L1 and L2 there is only a quantitative rather than a qualitative difference. It must be noted, however, that English and German speakers make use of the same kinds of hesitation phenomena. If we were dealing with a language such as French, for example, whose speakers tend to use draws rather than filled pauses, we would be likely to see a change in the phonological realisation of
the elements in question, as was the case with Möhle’s (1984: 44) German students of French who progressed from filled pauses to drawls as an approximation of the French native speakers’ behaviour.

The similarity of planning dispositions in each of their languages is shown by our participants (see the pie charts above), where the proportions of the different hesitation phenomena categories are very similar across languages. Nonetheless, the actual numbers of elements in each category do differ from one language to another. Mazes differ significantly in L1 and L2 (t = -2.76, df = 19, p = .012), and filled pauses are approaching significance (t = -2.01, df = 19, p = .059). By contrast, automatisms are not significantly different between languages. The importance of this will depend, to a large extent, on how one construes the relative functions of each of these categories. For instance, mazes and filled pauses could be thought to indicate extra planning demands when speaking L2. By contrast, the use of automatisms could be argued to represent an attempt at idiomaticity, which, as Raupach (1984: 135) maintains, is a form of planning behaviour comparable to that of native speakers.

Turning to the question of memory, many researchers (e.g. Harrington & Sawyer 1992, Hummel 2002) argue that resources for L1 and L2 might be different and therefore measure memory in each language separately. We also did this, and found that scores were higher in L1 than in L2, and that WM span was significantly different between languages (t = 3.15, df = 19, p = .005), although AMIBP did not differ significantly.

It might be questioned why behaviour on one memory test differs from that on the other. One reason could be the nature of the span test format, which requires memorising sets of semantically unrelated words, and is therefore an unnatural task which in no way reflects normal interpersonal interaction. Moreover, it is extremely demanding of attention, and the slightest change in strategy can result in failure. By contrast, one of the stated aims of the AMIBP battery was to reflect the memory demands of everyday life (Coughlan & Hollows 1985: 19). It is self-evident that recalling a story is closer to the demands of normal language use, e.g. in recounting one’s experiences, anecdotes, gossip, etc., where the speaker needs to remember and manipulate varying amounts of detail. Such recall necessitates the synthesis of multiple schemata, vocabulary items and other elements from long term memory; and it was precisely to account for prose recall that Baddeley (2000) proposed the Episodic Buffer, as a multi-modal integrational component. Such integration necessarily operates over a longer
timescale than the 2 seconds of classic WM, and was found, in the present study, to occur over periods typically upwards of 40 seconds.

This is not to say, however, that the act of recalling a story that one has previously been told is analogous to an act of creative language production, such as the spontaneous story-generation task used in our experiment. Although the two tasks might be thought to be similar, we maintain that there is a clear distinction to be drawn between the MEMORY test of story recall and the PRODUCTIVE linguistic task of story telling de novo. In the memory test, the recalled concepts were most frequently given using the same words and the same syntactic structures as were presented in the original stimuli. Thus, it appears that very little spontaneous generation of language takes place during the memory test. By contrast, the productive story-creation task did require spontaneous generation, where participants were left completely free to invent their own characters, plots etc. with no guidance or restriction from the experimenters.

The results of the correlations between memory tests and hesitation phenomena per 100 words (both measured in both languages) show consistent negative relationships, although not all reach significance. One reason for this could be the relatively low statistical power of our study, as there were only 20 participants. Ideally, a larger-scale study should be undertaken, yet there is a major obstacle in finding sufficient English native speakers who have near-native competence in German, learned as an L2. Nonetheless, there is a clear trend indicating that lower measures of memory are connected with greater production of hesitation phenomena, and higher memory ability is associated with lesser amounts of such phenomena.

There are a number of other factors that have been claimed to contribute to individual differences in the use of hesitation phenomena, such as tiredness, anxiety (Dewaele 2001) and individual speaking style. Clearly, some people place great importance on speaking “correctly”, both in their L1 and in their L2, while others are more concerned with keeping the conversation flowing, or with the content of their message and its interest value. In his study on the nature and distribution of filled and silent pauses in L1 and L2, Raupach (1984: 117) argues that much is determined by the speaker’s personal style. The same can be said for the use of automatisms, with some L2 speakers relying on formulaic sequences not only to bypass the difficulties of processing but also to add textual bulk (Wray & Perkins 2000: 18). Although, owing to practical considerations, such psychological and sociolinguistic variables were not measured in this study, it is nonetheless
necessary to bear in mind that they could be confounding factors in the relationship between hesitation phenomena and WM capacity.

Another factor which has been shown to influence the use of hesitation phenomena is the type of speech situation in which speakers find themselves, and indeed their perceptions of this situation, with the consequent demands placed on their performance. For instance, Pawley & Syder (1983: 200–201) demonstrate how a young man, who under normal circumstances is an experienced public speaker, produced extremely hesitant language when defending his PhD thesis. Pawley & Syder suggest that the higher incidence of hesitation phenomena in this context was due to the student’s need to grapple with the subject matter and choose his words carefully in a situation that required exactness rather than fluency.

In our study, it is unlikely that situational factors had any influence on the speech production of our participants, as they were in a physical environment with which they were familiar, and in conversation with fellow academics. However, the higher number of hesitation phenomena produced during answers to the Speculation questions in L2 is noteworthy. One might ask whether this has to do with the nature of the questions per se. Unlike the highly complex intellectual concepts with which Pawley & Syder’s (1983) participant was contending, while defending his thesis, our speculative questions were relatively unchallenging (e.g. If you had to be an animal, which animal would you choose to be and why?). Moreover, in L1, there was no significant difference between hesitation phenomena produced across the three genres (What and why questions, Story telling, and Speculation questions). This indicates that the increased production of hesitation phenomena in the Speculation questions in L2 could be attributable to the simple fact that these questions were presented at the very end of the entire interview. This suggests two possible confounding factors: firstly, that participants, having by this time been tested for some 40 minutes, might therefore be expected to be somewhat tired. Secondly, it is not unlikely that at least some participants could have been feeling more relaxed and expansive by that stage. Either, or indeed both, of these factors could be at play here.

Some studies of hesitation phenomena have suggested that there might be a link between the use of hesitant speech and syntactic complexity. Deese (1980: 71) suggests that “in exerting great effort to produce well-formed syntactic structures, it is at the expense of producing a high density of filled and unfilled pauses.” Similarly, Pawley & Syder (1983: 203) observe that
keeping to a clause-chaining style is more effective in terms of fluency than an integrating style where the speaker produces a lot of embedded clauses, which could lead to hesitation and syntactic breakdown. However, these studies do not substantiate their claims by empirical evidence, nor do they give an exact definition of embedding. Embedded clauses could be in the form of either Inflectional Phrases functioning as arguments of the verb, or Complementizer Phrases functioning adverbially. As such IPs are obligatory, whereas the CPs are optional, there will necessarily be different demands made on a speaker’s mental resources in their production, with the CP embeddings being more demanding (for further discussion see Fehringer & Fry 2007).

Examination of our data shows that there is a significant negative correlation between hesitation phenomena in L1 and the production of optional CPs in L1 ($r = -0.474$, $p = 0.033$). In L2, the equivalent correlation is also negative, although not significant ($r = -0.341$, $p = 0.142$). This indicates that greater production of optional CPs is connected with fewer hesitation phenomena, which appears to argue against the claims of Deese (1980) and Pawley & Syder (1983). It would seem, therefore, that some third factor is at play, and we would suggest that this is in the form of memory constraints. We found that the amount of optional CPs produced in each language is very similar (and is not significantly different: $r = 0.191$, $p = 0.42$), which might be a function of high proficiency in L2. Nevertheless, the amount of hesitation phenomena is significantly greater in L2, indicating that a speaker might be at a high level of L2 proficiency, and be able to produce complex syntax, yet might not have sufficient resources left to suppress the production of hesitation phenomena, which are not the focus of the speaker’s attention.

It is also highly probable that interference from L1 plays an important role in constraining the native-like performance of our L2 speakers. If we accept that the bilingual’s languages are organised into different subsystems, each of which can be selected for use, i.e. selecting one language while simultaneously suppressing the other (Green 1986, 1998), then it is likely that speakers with poorer WM resources will find it more difficult to “control” their languages, and the language that is supposed to be suppressed might interfere with the selected language in use. Consequently, the speaker will have to work harder to display an ease and fluency in L2 similar to that in L1 (Dewaele 2002: 222).
7. Conclusions

The production of hesitation phenomena varied considerably from person to person, across what might be expected to constitute a relatively homogeneous group of participants. The overall rate produced in L1 and L2 correlated significantly, and the proportions of the individual categories (filled pauses, mazes and automatisms) remained remarkably stable across languages, which provides support for the claim by Raupach (1983) and Möhle & Raupach (1989) that speakers’ planning dispositions in L1 are regularly carried over into L2. Our results show that such dispositions are applicable not only to temporal variables (speech rate, use of filled and silent pauses, which were the main focus of Raupach and his co-workers), but also apply to automatisms and mazes, and thence to hesitation phenomena in the more general sense.

Although the pattern and proportions of the individual categories of hesitation phenomena were similar in L1 and L2, the overall amount per hundred words was significantly greater in L2, and this was most noticeable in mazes. This indicates that, even at high levels of proficiency, the L2 speaker faces extra computational demands, as shown by the increased necessity for reformulation in L2.

Similarly, performance on the memory tests varied considerably between participants, yet scores were compatible with those found in the AMIPB norms or in previous psychological studies for WM span tasks. On the whole, a negative relationship was found to exist between memory ability, as measured in both tests, and the amount of hesitation phenomena produced per 100 words. However, it is recognised that the statistical power of this study, comprising only 20 participants, is relatively low for the purposes of inferential statistics, and future, larger-scale, studies are needed in this area. In the present preliminary study, although not all of the correlations reached significance, the overall tendency could nonetheless be observed that higher memory ability is related to fewer hesitation phenomena, and lower memory ability is associated with the production of more filled pauses, mazes and automatisms.

Moreover, any correlation between memory and hesitation phenomena will necessarily be less than perfect, as there are many confounding factors affecting the production of hesitation phenomena, irrespective of the language being used. Many of these factors (e.g. tiredness, anxiety, pragmatic ability etc.) are difficult, if not impossible, to measure on anything other than
a subjective scale. Furthermore, one can never rule out, in a correlational study, the possibility of there being some unknown third variable which affects both memory and language production ability.

The present study could have wider implications. For instance, as our results indicate that the planning dispositions in both languages are very similar, this could have a possible practical application: namely, to improve the performance of L2 learners, by training them to become aware of their production of hesitation phenomena in the mother tongue. This might lead to a diminution in the number of such phenomena produced in L2, on the basis of the standard speech and language therapy practice of bringing problem elements to the client’s attention in order to effect remediation.

We would also make the prediction that even greater amounts of hesitation phenomena would be produced in both languages when participating in free (i.e. normal) conversation, because of the demands of interpersonal interaction (e.g. monitoring, repairs etc.), and especially where speakers are vying for possession of the conversational floor. Such interactional requirements would necessarily need additional resources, thus reducing those available for linguistic computation. This ties in with Segalowitz’s (1997: 105–106) observation that individual differences in WM capacity play more of a role when the task demands are heavy than when they are light.

Another implication of this preliminary study might be to underline the need to assess verbal WM across a number of test formats which employ different timescales. Although WM is classically assumed to be 1½ to 2 seconds (the length of the Phonological Loop), we would suggest examining the longer timescale, such as that measured by the AMIPB (or similar story-re-telling tasks). Our results suggest that this timescale might be more appropriate to the assessment of language at the larger discourse level, where the requirements of topic maintenance, cohesion, floor holding, repair etc. come more to the fore.

**Appendix I: Instructions and test information**

The stimuli given below constitute the L1 interview (English version, for illustrative purposes), in the order in which they were presented, together with the instructions (italicised). The German L1 interview is a direct
translation of the English L1 interview, except that the two proverbs are replaced by similar German ones.

This L1 interview was immediately followed by the L2 interview, with equivalent, but different, stimuli (i.e. the same tasks, but with different questions, different story to recall, and different lexical items in the memory tests), presented in the same order, as set out below. The same procedure was followed consistently throughout the experiment, with all participants.

1. What & why questions

*These questions have no right or wrong answers. I'd just like you to tell me what you think about them.*

Why should we avoid getting into debt?
What does this saying mean? “One swallow doesn’t make a summer.”
Why are people who are born deaf usually unable to talk?
What does this saying mean? “A stitch in time saves nine.”
Why is it good to vote in secret in elections?

2. AMIPB Story recall subtest form 2 (Coughlan & Hollows 1985)

*I’m going to tell you a little story, and when I’ve finished, I’ll ask you to tell it back to me.*

Mr Peter Williams, who died last month, has left two hundred thousand pounds to a charity that provides seaside outings for the children of refugees. His younger brother, who lives in Canada, will inherit his house, his yacht, and his Rolls-Royce car. Mr Williams came from a poor family but he was determined to do well. He worked extremely hard and everyone liked him. His first job was as a butcher’s boy but he earned extra money by doing night-work in a laundry. When he was thirty he bought a van and started a removals business. However he eventually made his fortune selling paintings and antique clocks.

3. Story telling (preliminary)

*I’d like you to look through these pictures, and then choose one. In a few minutes I’m going to ask you to pretend that it’s a still from a film, and tell me the story.*
4. **Aural Working Memory Span (Fry 2002)**

*I’m going to give you some sentences. After I’ve said a sentence, you say whether it’s true (T) or false (F). Then I’ll give you another sentence and you say “true” or “false”. And then I’ll ask you to tell me the last word of each of the sentences. The number of sentences will go up, but I’ll tell you when it does. We’ll have a practice first.*

**Practice:**

- you cook with gas T
- babies drink coffee F

*Okay? Here we go then.*

2/1 houses have doors T
- water is purple F
2/2 grandfathers are young F
- you cut wood with a saw T
2/3 children go to school T
- butterflies eat cheese F

3/1 chairs have legs T
- the Earth is a planet T
- you put milk in the oven F
3/2 you open a lock with a key T
- people with lots of money are poor F
- trees have leaves T
3/3 you see with your eyes T
- cats eat metal F
- you sleep in a bath F

4/1 swimming makes us wet T
- Spain is in Asia F
- you play tennis with a dog F
4/2 frogs live in a pond T
- it’s hot in the winter F
- you buy things in a shop T
- monkeys are bigger than lions F
4/3 New Year’s Eve comes in June F
- horses have wheels F
you make holes with a drill T
the sea contains salt T

5/1 you cut bread with a spoon F
diamonds are shiny T
little girls like dolls T
you wear shoes on your head F
camels like snow F
5/2 all boys are sons T
newspapers have pages T
all cars are yellow F
two times five is ten T
birds have roots F
5/3 a square has four corners T
you find trains at a station T
perfume has no smell F
goats play the drums F
spiders sing songs F

5. Story telling (story)

Have you decided on a picture?
You’re telling me the story of the film: what’s happened already, what’s going on in the picture, and what’s going to happen afterwards.
[The participant then invents a story]

6. Speculation questions

These are some more questions that ask your opinion. Please, just tell me what you think.

If you won £1 million on the lottery, how would it change your life?
Why do people seem happier in the summer than in the winter?
How did gardening and cookery programmes on television come to be so popular?
What makes the ideal friend?
Appendix II: Individual categories of hesitation phenomena used by participants (scores per 100 words)

<table>
<thead>
<tr>
<th></th>
<th>automatisms</th>
<th></th>
<th>automatisms</th>
<th></th>
<th>mazes</th>
<th>L1</th>
<th>mazes</th>
<th>L2</th>
<th>filled pauses</th>
<th>L1</th>
<th>filled pauses</th>
<th>L2</th>
<th>total L1</th>
<th></th>
<th>total L2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Eng</td>
<td>0.83</td>
<td>7.07</td>
<td>5.47</td>
<td>3.76</td>
<td>4.81</td>
<td>8.62</td>
<td>11.11</td>
<td>19.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>4.41</td>
<td>7.64</td>
<td>2.20</td>
<td>5.01</td>
<td>3.90</td>
<td>5.01</td>
<td>10.51</td>
<td>17.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>3.07</td>
<td>2.93</td>
<td>4.36</td>
<td>5.21</td>
<td>3.72</td>
<td>7.82</td>
<td>11.15</td>
<td>15.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>11.09</td>
<td>7.30</td>
<td>9.78</td>
<td>10.40</td>
<td>5.75</td>
<td>10.71</td>
<td>26.61</td>
<td>28.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>3.44</td>
<td>8.30</td>
<td>1.89</td>
<td>3.21</td>
<td>6.01</td>
<td>2.45</td>
<td>11.34</td>
<td>13.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>4.45</td>
<td>7.05</td>
<td>7.20</td>
<td>8.63</td>
<td>6.15</td>
<td>7.38</td>
<td>17.80</td>
<td>23.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>2.56</td>
<td>2.65</td>
<td>3.07</td>
<td>8.27</td>
<td>3.07</td>
<td>5.93</td>
<td>8.69</td>
<td>16.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>5.19</td>
<td>5.24</td>
<td>4.26</td>
<td>8.61</td>
<td>1.11</td>
<td>0.56</td>
<td>10.56</td>
<td>14.42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eng</td>
<td>1.15</td>
<td>2.33</td>
<td>2.31</td>
<td>4.19</td>
<td>5.48</td>
<td>7.21</td>
<td>8.93</td>
<td>13.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>8.29</td>
<td>6.83</td>
<td>2.07</td>
<td>3.32</td>
<td>8.29</td>
<td>9.75</td>
<td>18.65</td>
<td>19.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>3.81</td>
<td>4.79</td>
<td>3.21</td>
<td>2.97</td>
<td>3.01</td>
<td>3.63</td>
<td>10.02</td>
<td>11.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>5.90</td>
<td>5.59</td>
<td>1.97</td>
<td>4.16</td>
<td>6.55</td>
<td>4.40</td>
<td>14.41</td>
<td>14.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>8.87</td>
<td>11.72</td>
<td>2.41</td>
<td>3.69</td>
<td>6.72</td>
<td>5.73</td>
<td>18.00</td>
<td>21.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>3.20</td>
<td>4.33</td>
<td>8.53</td>
<td>4.72</td>
<td>6.18</td>
<td>6.30</td>
<td>17.91</td>
<td>15.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>5.36</td>
<td>5.40</td>
<td>2.22</td>
<td>3.65</td>
<td>4.44</td>
<td>4.23</td>
<td>12.01</td>
<td>13.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>4.92</td>
<td>4.49</td>
<td>2.15</td>
<td>3.83</td>
<td>5.69</td>
<td>7.15</td>
<td>12.77</td>
<td>15.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>6.25</td>
<td>6.49</td>
<td>2.02</td>
<td>2.60</td>
<td>3.13</td>
<td>2.41</td>
<td>11.40</td>
<td>11.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>7.34</td>
<td>7.66</td>
<td>2.39</td>
<td>2.69</td>
<td>3.58</td>
<td>2.96</td>
<td>13.31</td>
<td>13.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ger</td>
<td>6.35</td>
<td>5.08</td>
<td>3.61</td>
<td>5.50</td>
<td>5.15</td>
<td>6.91</td>
<td>15.12</td>
<td>17.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References


Van Lancker, Diana. 2004. ”When novel sentences spoken or heard for the first time in the history of the universe are not enough: Toward a dual-process model of language”. International Journal of Language and Communication Disorders 39: 1–44.

Authors’ addresses
Dr Carol Fehringer
School of Modern Languages
Newcastle University
NE1 7RU, United Kingdom
e-mail: Carol.Fehringer@ncl.ac.uk

Dr Christina Fry
Speech Section
School of Education, Communication and Language Sciences
Newcastle University
NE1 7RU, United Kingdom
e-mail: C.S.Fry@ncl.ac.uk

received: 12 July 2006
revision invited: 29 September 2006
revised version accepted: 5 November 2006