Disentangling perceptual uncertainty and incrementality in syntactic processing: Click monitoring via ERPs

David J. Lobina, Josep Demestre & José E. García-Albea

Abstract

Past results with the click monitoring technique have pointed to an end-of-clause effect, but we here show that the issues at hand are a bit more nuanced —and more interesting. Firstly, by employing two types of simple, subject-verb-object sentences and three click positions, we report two experiments (1a and 1b) which show that reaction times (RTs) are affected by two factors: a) a strong perceptual effect we dub the position effect, involved in monitoring tasks in general, and which neutralises structural differences across experimental conditions; and b) the incremental processing the parser carries out means that more resources to respond to a tone are released as a sentence is presented, as evidenced in the tendency of RTs to decrease across a sentence. These two factors are then successfully discriminated and recorded by registering event-related brain potentials (ERPs) in a click monitoring task, with experiment 2 establishing that the amplitudes of the N1 and P3 components —the first component is associated with temporal uncertainty in perception and the second with processing effort in dual tasks— correlate with the RT pattern unearthed in the previous two experiments. In addition, it is argued that the P3 component is a useful metric with which to measure the cognitive load exercised by the syntactic parser when employing a dual-task experimental setting, and we believe this is the first time it is so employed in psycholinguistics. In experiment 3, we segregate these two factors by placing the last tone on the penultimate syllable of the experimental sentences, with the result that RTs to that position are found to increase greatly with respect to the immediately preceding tone, confirming a wrap-up effect (rather than an end-of-clause effect) in speech perception, and thereby both disrupting the position effect and highlighting purely structural factors. Finally, past results with the click-detection paradigm are reconsidered in terms of our data.

Keywords: Click monitoring; ERPs; Processing load; Position effect; Wrap-up.

1 The Past

Click monitoring originated from the click-location paradigm, the latter an experimental technique first employed in the mid 1960s as a way to show that the clausal hierarchies postulated by generative grammarians reflected how people in fact conceptualise them in performance. As Fodor and Bever (1965) tell us, theirs being the first study to employ click location in syntactic parsing, the click paradigm was at the time used by phoneticians as a means to probe perceptual consistency (ibid., p. 415), a phenomenon in which a processing unit can be found to resist interruption, hence constituting a "perceptual unit". Perhaps daringly, the click-location paradigm was subsequently employed by psycholinguists to figure out if there were any such perceptual units in syntactic parsing, an issue and a version of the click paradigm that will not feature much here. The monitoring, or detection, version of the click paradigm was being developed at around the time the location version was at its peak of popularity (or, more accurately, on its way out), and it is this version that we will employ here.

Granted, neither version of the paradigm is widely used these days, but as far as the monitoring version is concerned, this is to the loss of the psychology of language. In fact, and as we will discuss soon enough, recent studies using this type of monitoring task have yielded interesting data, and the work reported here provides further evidence for the usefulness of this technique in the study of the parsing process. Before proceeding to that, though, we offer a general outline of the paradigm itself, discussing, first of all, some of the results obtained with the location version in order to then move on to the detection/monitoring version, and therefrom to our study.

In general, the click paradigm consists in superimposing a short, extraneous sound —a click, a tone, or else— over some linguistic material, which is then played to subjects over headphones. In the location version Fodor and Bever (1965) ran, the participants would be asked to, first, write down the piece of auditory material they had just heard (in this case, a sentence) and then mark where they thought the click was placed. It was not a matter of probing the participants' accuracy in the task —they were indeed very inaccurate— it was instead an endeavour to map the errors subjects make, so that a comparison could be drawn between the objective position of the click and the position in which subjects subjectively perceive it.

Fodor and Bever (1965) reported that even though participants had a tendency to perceive a click before its objective position (a left bias, which was also reported by the developers of the click paradigm; ibid., p. 419), the overall majority of clicks, as subjectively perceived, were displaced towards clausal boundaries. Thus, in a sentence such as *that he was happy was evident from the way he smiled*, the click was perceived to be between the main and the subordinate clause; that is, between *happy* and the *was* to its right, when in fact the click would be placed either between the first *was* and *happy* or between the second *was* and *evident*. Plausibly, a biclausal sentence of these characteristics exhibits a certain complexity, as it contains various internal phrases and boundaries, and one may well wonder which frontier exactly the perception of a click would migrate to in such sentences. The results reported in Garrett, Bever and Fodor (1966) suggest that clicks only ever migrate to the deepest constituent boundary —the frontier between clauses.

Similarly, Bever, Lackner and Kirk (1969) concluded that within-clause boundaries do not appear to affect segmentation strategies. These results were taken as evidence that the clause is an important unit of syntactic processing, perhaps even constituting a perceptual unit (Fodor, Bever & Garrett, 1974). Furthermore, the clause-by-clause process these scholars believed to have unearthed appeared to be solely the effect of syntactic properties, as other factors were controlled for and did not seem to affect the results —some of these included pitch, intonation, response bias, the effect of writing down the sentence, the effect of extracting the sentence from memory before writing it down, etc. (Bever, 1973; Garrett et al., 1966). The latter point was contested by Reber and Anderson (1970), though; by employing much simpler sentences (mono-clausals such as *open roadside markets display tasty products*), they argued that the evidence suggested that a) a right bias was actually operative (which they claimed to be present in the results of the aforementioned papers as well), and b) extra-linguistic factors were clearly responsible for the errors participants made. They also reported a tendency for subjects to mislocate some clicks to the boundary between the subject and the verb, which might suggest that this break is also important for the processor, something that was explicitly denied in Fodor et al. (1974, p. 336).[1]

As mentioned, these now-classic publications employed the location version of the click paradigm, an offline experimental technique that is currently largely out of favour (see Levelt, 1978, for discussion of its flaws). Abrams and Bever (1969) developed the detection or monitoring version, an online technique that is much more reliable, and the one that is under the microscope here. In this version of the paradigm, participants are required to press (or depress) a button as soon as they hear the click, and so the analysis centres on the fluctuations in reaction times. The idea is that processing a sentence and monitoring for a click compete for attentional resources, and so reaction times ought to be greater at those junctures of a sentence that require more working memory resources; a correlation between reaction times and complexity, as it were.

The first studies employing this online version of the paradigm also employed biclausal sentences, with Abrams and Bever (1969) finding that clicks before the major break were reacted to more slowly than clicks *at* the major break or just *after* it in sentences such as *in addition to his wives, the prince brought the court's only dwarf* (a curious sentence to employ, to be sure). Similarly, Holmes and Forster (1970) found that reaction times (RTs) in the first half of a biclausal sentence are greater than in the second half. This phenomenon has been termed the end-of-clause effect (Bever & Hurtig, 1975); that is, the end of a major clause seems to involve some special processing, possibly due to the parser's attempt to close off the various syntactic nodes that are opened during the processing of a clause. In relation to what we will do here, the end of a clause may or may not coincide with

[1] Chapin, Smith and Abrahamson (1972) reported a similar result with biclausal sentences, but there are some problems with the materials they employed (see Fodor et al., 1974, pp. 336 et seq. for relevant discussion).

the end of a sentence, and therefore we will here use the term end-of-clause effect for when it doesn't and the term wrap-up effect for when it does —as a matter of fact, however, we believe these two effects to be separate and possibly very different (it is noteworthy, moreover, that none of the studies mentioned in this paragraph placed any tones at the end of sentences)

Surprisingly, the click-detection has not been employed as much as it perhaps deserves, given its apparent sensitivity to the different cognitive loads the parser goes through within and between clauses in complex sentences (see Levelt, 1978, this time for discussion of how the click monitoring doesn't share the flaws of the location version of the paradigm). After Flores d'Arcais (1978) successfully used it to show that main clauses are usually easier to process than subordinates (and that the main/subordinate order exerts less memory resources than the subordinate/main order), the 1980s and 90s hardly exhibit any other study employing this technique. It is not surprising, then, that Cohen and Mehler (1996) considered their work a "revisit" to the paradigm —making our study a further visit— when they reported that RTs to clicks at the boundary of reversible object relatives were greater than at structurally-identical subject relatives (or in other positions of a normal object relative; they also reported similar results with semantically reversible and irreversible sentences). Recently, the click-detection paradigm has been successfully employed in a word segmentation study (Gómez, Bion & Mehler, 2011), and it is hoped that the results we report here are further evidence for its usefulness in the study of language comprehension.

In particular, we report the results of four experiments with the detection version of the click paradigm, one in combination with a recording of event-related brain potentials (ERPs), which when put together demonstrate the following: a) RTs are affected by two factors: a strong perceptual effect operative in all monitoring tasks and the flow of linguistic information the processor successively receives (Experiments 1a and 1b); b) these two factors can be discriminated by taking a record of ERPs during the monitoring task (Experiment 2) as well as be successfully segregated in actual performance by placing the tone at the end of sentences, an especially taxing location for the parser (Experiment 3); and, c) the P3 component recorded in the ERP experiment can be usefully employed to measure processing effort in dual tasks such as the one administered here. In addition, these data provide the appropriate background for a re-analysis of past results obtained with this technique, and we shall provide such a discussion in the last section of the paper.

The paper is organised as follows. In the next section, we outline our aims and purposes, describe the sort of experimental materials we will employ in the four experiments reported here, and outline some of the predictions to be evaluated. This is followed by four experiments, organised in three sections, with the last section of the paper (sort of) putting it all together for discussion and contemplation.

2 Current Concerns

According to Cutler and Norris (1979), detection tasks such as phoneme- and word-monitoring generally exhibit a decrease in RTs across a sentence, a factor that we quip a "position effect" and one which Cutler and Norris deny to the click monitoring experiments of Abrams and Bever (1969) and Holmes and Forster (1970). This conclusion, however, is based on a less than careful analysis of the data that Abrams and Bever (1969), in particular, report. As mentioned, these authors established three different click positions in sentences such as *since she was free that day, her friends asked her to come* —that is, on the last word *before* the main clause break, *in* the clause break, and on the first word *after* the clause break, where the main clause break is to be found between *day* and *her*— and the following are the RTs they obtained, placed in the right order: 243 ms., 230 and 216. That's clearly a decrease in RTs, so why do Cutler and Norris (1979) conclude otherwise?

Somewhat curiously, Abrams and Bever (1969) exposed their participants to repeated presentations of the same material, and naturally, subjects' performance progressively improved. The RTs we have just provided were those of the first presentation, and it is only in the next two presentations wherein the linear decrease in RTs from the first to the third position disappears —and it is this pattern that Cutler and Norris (1979) focus on. Given that these participants were reacting to then-familiar sentences and click positions in the second and third presentations, those responses are obviously not comparable to those of the other monitoring tasks Cutler and Norris (1979) discuss. Moreover, Abrams and Bever (1969) also ran a version of their experiment in which the materials were artificially reconstructed (that is, words were recorded separately and then spliced together into coherent sentences) and the responses they obtained on the first presentation exhibited exactly the same pattern as in the first presentation of the normal exposure (the second and third presentations varied from the first in this case too). Clearly, Cutler and Norris (1979, p. 129) are too brusque in drawing a line between the memory resources involved in reacting to sentence-internal targets (such as those of phoneme- and word-monitoring tasks) and those at work in targeting sentence-external elements (as in a click-detection task) if their analysis is based on the successive presentations Abrams and Bever (1969) carried out. If we are correct in our analysis, the decrease in RTs is indeed well-established in the results of Abrams and Bever (1969), and this phenomenon is further evidenced in the data of Holmes and Forster (1970), who found that RTs to clicks placed on the first half of a biclausal sentence were greater than the RTs to clicks introduced in the second half.[2]

[2]Cutler and Norris (1979) also mention some of the data reported in Bond (1972), who found that clicks placed at the end of phonological phrases are reacted to more slowly than in previous positions, but that datum is not incompatible with the general point we are making. The end of a clause and/or of a sentence (or, in this case, a phonological phrase) is very likely the locus of special cognitive strain, but that doesn't mean there isn't a decreasing tendency in RTs across a sentence when not placed at the end of clauses/sentences/phonological phrases. We will come back to this

What we have here called the position effect did not feature in the discussions of Abrams and Bever (1969) and Holmes and Forster (1970), but given how robust this phenomenon appears to be in monitoring tasks, its potential role in explaining those data, and ours, ought to not be underestimated. Holmes and Forster (1970) come close to discussing this issue when they suggest, reasonably, that their participants must have been experiencing "maximal uncertainty" at the beginning of a sentence, something that is plausibly reflected in the high RTs for tones placed in the first clause. This maximal uncertainty makes reference to the expectations and predictions of the parser during processing, a slightly different type of factor than the position effect, which we take to be perceptual in nature. According to Holmes and Forster (1970), then, the processing load at the end of a clause ought to be minimal, given that 'structural information conveyed by the last few words would tend to be highly predictable' (p. 299). In other words, the cognitive resources exerted by the primary task (parsing a string) are much greater at the beginning of a sentence, while towards the end of a sentence the attentional mechanisms at work in the perception of a click have access to more resources —i.e., there is less competition for resources between the primary and the secondary tasks— and hence reactions to a click placed late in a sentence ought to be faster.[3]

This must certainly be true for the sort of biclausal sentences studies such as those of Abrams and Bever (1969) and Holmes and Forster (1970) employed, but the same point doesn't quite have the same import in phoneme- and word-monitoring tasks (as the materials Cutler & Norris, 1979, analyse show). Moreover, though Abrams and Bever (1969) and Holmes and Forster (1970) explain their data in terms of the processing load associated to the end of a clause, it is noteworthy that, in the case of Abrams and Bever (1969) at least, the tone these authors placed at the end of the major break also constitutes the end of a subordinate phrase (as in, recall, *in addition to his wives* or *since she was free that day*), and at that precise point the processor is obviously in a state of great uncertainty, for a significant amount of linguistic material is yet to come.[4] In other words, the pattern reported in these studies may not (only) be the result of an end-of-clause effect. In fact, if we take into consideration both the position effect and the changes in the parser's predictions/expectations, the claim that there is an end-of-clause effect at play in those studies —a claim that has, in addition, more recently being used in support of syntax-last models of processing (Townsend & Bever, 2001)— doesn't appear to be very well supported. It is to this constellation of issues that we are directing our attention here.

later on, where we will refine this point.

[3]We take this general state of affairs to follow from the incrementality of syntactic processing, an issue that we will discuss later on in the context of how this property relates to the end-of-clause and wrap-up effects.

[4]It is harder to evaluate the results of Holmes and Forster (1970), as we are not provided with exact RTs and length doesn't appear to have been controlled for. Indeed, these authors also found that RTs in a clause break were shorter than in positions wherein the click would interrupt a large number of constituents, but in some sentences the clause break would precede the no-clause break, while in others it was the other way around.

Taking our heed from the materials employed by Reber and Anderson (1970), mentioned supra, and contrary to what has generally been the case in other click-detection studies, and in most of the psycholinguistics literature for that matter, we will here use rather simple experimental sentences. Less roughly, no biclausal sentences, ambiguous constructions, or relative clauses will feature in what follows; instead, unambiguous, declarative, active sentences will form the basis of our investigation. If the two factors we have described are operative in complex sentences, wherein many other variables may partake in the parsing process, these ought to be also evident, and more clearly so, in monoclausal sentences —and that seems to us as the most appropriate place to start an investigation of the sort we are interested in.

The employment of these materials, moreover, will allow us to connect our work to some of the most robust findings in linguistics, the most relevant of which, achieved in the 1970s and beyond (see Moro, 2008, for relevant discussion), is the discovery that all syntactic phrases, including a sentence itself, are structured according to the same geometry: a specifier-head-complement(s) (SHC) schema. At some level of abstraction, then, it is reasonable to suppose that the parser recovers/builds a (macro) SHC structure (that is, the sentence) and its internal phrases, the latter SHC phrases themselves. Consequently, we will here focus on how the parser compiles the intrinsic structure of both a sentence and its internal phrases.

This is not an entirely novel way of describing what the parser must do at some level, but it nonetheless varies from similarly phrased approaches. The most relevant example here is that of Lewis (2000), who also considers the cognitive load exercised in the compilation of head-complement structures, albeit of a different kind to what we have in mind here. In order to be more precise, Lewis's main interest lies on rather more complex head-complement relations: those involved in locality and anti-locality effects in relative clauses and the like. We shall focus on rather simpler SHC structures, and the issue that motivates us is the apparent dynamic involved in such a way of conceptualising the parsing process. Namely, recovering the macro SHC structure and its internal SHCs would introduce a hierarchical layering (of operations and representations) at an early stage of parsing, as the internal SHC phrases may appear at either the S or the C position (of either a sentence or an internal phrase), and this ought to be reflected in the RTs recoverable with the click monitoring.

Consider, then, some of the general features of this point of view. The H position will always be occupied by a terminal element, while both S and C may be empty or occupied by other SHCs. Furthermore, we must distinguish between a macro SHC (a subject-verb-object sequence, basically) and the internal instantiations of this general scheme. A sentence, then, is an asymmetric structure composed of other asymmetric structures (NPs, VPs, etc.). In principle, this constitutes a complex problem (building a tense phrase (TP), the topmost node of a sentence according to some theories) that is reducible to simpler instances of the same problem (building NPs, VPs, PPs, in succession), making a recursive solution perfectly applicable. We will not evaluate that possibility here; rather, we will construct experimental materials in which complex phrases will appear at either the S or C position of a macro SHC structure —that is, to the left or right of the verb—, and the working hypothesis will be that RTs to clicks placed within complex phrases should be greater than to clicks at positions where no internal SHC phrase must be reconstructed.

Monoclausal, subject-verb-object Spanish sentences were constructed for the purposes of this investigation. Starting from a matrix *proposition* —that is, a predicate and its arguments— two different types of sentences were created. *Type A* sentences exhibited a complex subject but a simple object, while the reverse was the case for *type B* sentences. By a complex subject or object is meant a noun phrase (composed of a determiner and a noun) which is modified by another noun phrase (also composed of a determiner and a noun, but introduced by a preposition). A simple subject or object, on the other hand, would simply be composed of a determiner and a noun. The following are the experimental sentences to be used in all the experiments reported here, where the | symbol identifies the boundaries under study in the first three experiments (the fourth experiment will introduce some modifications, to be described in the corresponding section of the paper).

Type A: El candidato | del partido | se preparó | el próximo discurso.
'The party's candidate prepared his next speech'.

Type B: El candidato | ha preparado | un discurso | sobre la sanidad.
'The candidate has prepared a speech about the health service'.

While both sentences are macro SHC structures, they exhibit rather different internal SHC configurations. Namely, *type A* sentences are [NP-[PP-NP]-[VP-NP]] sequences, while *type B* sentences exhibit a different form, namely [NP-[VP-NP-[PP-NP]]]. Thus, there are internal SHCs to be constructed in these two types of sentences, but at different locations: either to the left-hand side of the VP (in type A sentences) or to its right (for type B). As mentioned, the | symbol indicates the relevant boundaries, but this is not the actual place of the click; as will be explained in the methodology sections below, the click was to appear two syllables after the relevant boundary. (Roughly, this is to allow the parser to readjust its operations as it proceeds so that we can be sure that it has completed the previous phrase fully and moved on).

More to the point, our contention will be that RTs to clicks placed on the left-hand side of the verb in type A sentences should be greater than to clicks on its right-hand side, and the same goes for the material on the right-hand side of the verb in type B sentences vis-á-vis a click placed on the left-hand side. Moreover, the position of the verb ought to be of some importance, and we expect to observe differences in RTs depending on whether the processor has processed the verb or not (this also applies to whether the parser has processed the verb's complement or not). Following up from that, we can draw comparisons between the two types of sentences, as the three boundaries here identified are equalled in length across sentence type, as described below. These comparisons allow us to formulate two further hypotheses, as we can predict that RTs to the same positions, measured in number of syllables, will be different across sentence type. To wit, we hypothesise that a) RTs to clicks in the first position of type A sentences should be greater than the RTs to clicks on the corresponding position in type B sentences, and b) RTs to clicks placed in the third position of type A sentences should be lower than those for the third position of type B sentences. Putting it all together, we expect the structural conditions of our experiments to mitigate the position effect and augment the role of the parser's expectations/predictions.

Before moving on to the experiments, we would like to advance some brief and general comments regarding what sort of architecture we envision for our parser. We suppose that our parser will attempt to resolve its operations at every stage of sentence processing, with no delay or deferment. That is, we suppose that our parser will attempt to compute, close, and discharge the chunks it receives by carrying out two operations widely discussed in the natural language processing literature (see, for instance, Berwick, 2011):

- Shift: keep incoming elements in memory until they can be combined.
- Reduce: merge words into phrases and close them from further operations.

Note that according to our predictions the window of the *shift* mechanism must be very short indeed, for the processor would not keep incoming elements in working memory until these can be appropriately packaged —the packaging must be almost immediate, if incrementality is so robust as it is supposed (long-distance dependencies notwithstanding). This is perfectly compatible with the proposition that the end of clauses and/or sentences constitutes a locus of particular operations (the end-of-clause effect in one case, the wrap-up effect in the other), without this involving the further claim, we believe not entirely warranted (see infra), that the parser at that point branches off all the nodes it opened during the processing of a sentence. The *compute-close-discharge* chain, then, would take place as soon as possible. Considering the type of materials we are manipulating in this study, and as we shall see, such a general type of parser more than suffices to explain the data.[5] Having outlined the general approach, we now turn to the experimental data.

3 Experiment 1

Two versions of the same experiment are reported in this section, both of them employing the same experimental materials but varying in terms of whether fillers and a comprehension task were used.

[5]Some colleagues have in the past pointed to Levy's (2013) Surprisal model and to the notion of Informativity (Chater, Crocker & Pickering, 1998) in order to account for our data, but neither of these two approaches appears to speak directly to what is involved in our experiments. As with much of the overall psycholinguistics literature, most models are proposed with more complex sentences in mind, and it's not always clear how they apply to simpler ones.

3.1 Experiment 1a

3.1.1 Method

Participants. 88 psychology students (20 male, 68 female) from the Rovira i Virgili University (Tarragona, Spain) participated in the experiment for course credit. The mean age was 20 years, and participants had no known hearing impairments. All were native speakers of Spanish.

Materials. Two variants of monoclausal, active, declarative, subject-verb-object Spanish sentences were constructed from 60 matrix *propositions*. Type A sentences exhibited an [NP-[PP-NP]-[VP-NP]] pattern in which a) the subject was a complex structure composed of a noun phrase (determiner + noun) modified by another noun phrase (always introduced by a preposition), b) the verb was either in perfect tense (with an auxiliary such as *ha*) or in reflexive form (and therefore always headed by the particle *se*) and c) the object was a simple noun phrase. In some cases, short and long modifiers were introduced in order to keep length constant. Type B sentences manifested a [NP-[VP-NP-[PP-NP]]] form in which a) the subject was a simple determiner-noun complex, b) the verb was of the same type as for type A and c) the object was a complex noun phrase modified by another noun phrase that was always introduced by a preposition. Type A and B are the structural conditions of the experiment. All sentences are unambiguous, composed of high- or very high-frequency words, according to the corpora and classification in Almela, Cantos, Sánchez, Sarmiento and Almela (2005) (which was cross-checked with Sebastián-Gallés, Martí, Carreiras and Cuetos 2000), and with a total length of 20 syllables. On average, the first boundary appeared after 4.2 syllables (standard deviation: 0.43), the second frontier after 9.1 (0.62) and the last juncture after 13.9 (0.72). The sentences were recorded in stereo with a normal but subdued intonation by a native, male speaker of the Spanish language using the Praat software on a Windows-operated computer. Special care was employed so that the intonational break between the subject and verb was not too marked, but no attempt was made to neutralise it (for some of the phonological factors that can affect click detection, see Bond, 1972).[6] The sentences were subsequently analysed with Praat (Boersma, 2001) in order to identify and eliminate any undesired noise. Three click positions per sentence were established, one for each of the boundaries. These are the three positional conditions of the experiment (1-2-3). It was decided that the clicks would be placed on the vowel of the second syllable following the relevant boundary, so that the processor could use the first syllable (usually a preposition,

[6] Prosody has been argued to play a rather central role in sentence comprehension, but this is usually framed in terms of syntactic attachment decisions, focus interpretation or the availability of contextual information in the resolution of lexical ambiguity (Schafer, 1997); that is, in terms of higher-order operations, which do not apply to the materials we are manipulating here —or not as much. In any case, Bond (1972) found evidence that suprasegmental factors influence subjects' performance in monitoring tasks in the following manner: a) participants react faster when the click is placed on a vowel that is unstressed, and b) intonational phrases appear to drive the segmentation process to a significant extent. According to this author, then, the first step in speech perception involves segmenting the string into phonologically-defined units, which is not implausible.

the beginning of a preposition, or the form heading the verb) to "disambiguate" the phrase it is processing at that moment, thereby updating the operation it needs to carry out at each stage (e.g., reconstructing a SHC in C position, or else). The software Cool Edit Pro (Version 2.0, Syntrillium Software Corporation, Phoenix, US) was employed to generate and superimpose tones with a frequency of 1000 Hz, a duration of 25 ms., and a peak amplitude similar to that of the most intense sound of the materials (around 80 dBs). Every sentence had one click only, and in order to make sure that every item went through every condition, three different copies of each experimental item were created, totalling 360 experimental sentences. A further 12 practice items were created, two items per experimental condition.

Procedure. The design of the experiment was a 2 (type of sentence factor) by 3 (click position factor) within-subjects, within-items factorial, and therefore six versions of the task were created, corresponding to six experimental groups. From a pool of 360 experimental items (six sets of 60 sentences in accordance to the six experimental conditions: type A and B sentences, and the three click positions), six lists were created with a total of 60 experimental items each and with a distribution of 10 items per experimental condition. Each version was arranged according to a Latin square (blocking) method so that the items were randomised within and between blocks. Consequently, every subject underwent every condition (but saw each sentence just once), and every matrix proposition also underwent every condition (two types of sentences, three click positions). Participants were randomly assigned to each experimental group. The experiment was designed and run with the DMDX software (Forster & Forster, 2003) and administered in a sound-proof laboratory with low to normal illumination in which a maximum of four subjects at a time would be tested. Participants were seated in front of a table containing a computer screen, a keyboard, a keypad, and a set of headphones. A list of instructions was placed on top of the keyboard for subjects to read before the start of the practice session. Once they had finished reading the instructions, the experimenter explained the task and answered any possible questions. As soon as this was done, participants were asked to put on the headphones in order to carry out a practice session while the experimenter was still in the room. The sentences were presented over the headphones binaurally and participants were instructed to hold the keypad with their dominant hand in order to press a button as soon as they heard the tone. They were told to be as quick as possible, but to avoid guessing. Once a sentence had finished, an instruction on the computer screen stated that the next sentence would be presented upon pressing the space bar, giving subjects control over the rate at which the sentences were presented. Each sentence would be played 500 milliseconds after pressing the space bar and would not last more than 5 seconds. This set-up ensured that participants had the dominant hand on the keypad and the other on the keyboard. Once the practice session was over, the experimenter clarified any final questions before leaving the experimental room. An experimental session of 60 items started immediately after and the DMDX software was used to measure and record reaction times. The whole session lasted around 20 minutes.

3.1.2 Results

The responses of 8 subjects had to be eliminated for a variety of reasons. Six of these were due to technical problems with the coding of the computer programme and/or the equipment, while the other two did not meet reasonable expectations regarding average performance. One of these two participants failed to record a single response (most likely because the task was carried out incorrectly), while the responses of the other suggest that this subject was not paying attention at all (the standard deviation of this participant more than doubled his/her rather high average response time).

The reaction times of the remaining 80 subjects were collected and polished with the DMDX programme. A response that occurred before the tone or 3 seconds after the tone was not recorded at all (in some cases, 3 seconds after the tone meant that the sentence had long finished), while responses deviating 2.0 SDs above or below the mean of each participant were eliminated (this affected 4.3% of the data). The resultant measures were then organised according to experimental condition. The analysis of reaction times was carried out with the SPSS package (IBM, US). Table 1 collates the RTs per condition.

Table 1: Experiment 1a. RTs per click position per sentence type (Mean RT with standard deviations in parentheses).

	Click Position		
Sentence Type	1	2	3
A	257.22 (59.1)	222.51 (41.0)	206.78 (40.1)
B	252.40 (52.0)	217.33 (43.9)	205.26 (44.3)

As can be observed in Table 1, RTs are greater in Position 1 and decrease thereon for each sentence type. A repeated-measures analysis of variance showed that the click position factor was significant for both the subjects and items analyses ($F_1(2,158) = 144$, $p < .001$; $F_2(2,118) = 295$, $p < .001$; $minF'(2,265) = 96.76, p < .001$), while the *sentence type factor* was only significant for the subjects analysis ($F_1(1,79) = 4.66$, $p < .05$; $F_2(1,59) = 2.48$, n.s.; $minF'(1,114) = 1.61$, n.s.). There was no interaction between the two experimental factors (all $Fs < 1$). Two-tailed t tests were carried out to compare pairs within sentence type (i.e., A1 vs. A2, etc.) and all were found to be significant ($ps < .001$).

3.2 Experiment 1b

In order to make sure that the results of the previous experiment did not reflect a possible "habituation" to the task, we ran a modified version that included filler sentences and a comprehension task. It was hoped that the variety of the fillers (mixed with the experimental sentences) would force the processor to parse each sentence anew (as it were), while the inclusion of the comprehension task was intended to probe whether participants were in fact paying attention to the meaning of the sentences they were hearing.

3.2.1 Method

Participants. 77 psychology students (8 male, 69 female) participated in the experiment for course credit. This was a different set of participants from Experiment 1a. The mean age was 22 years, and no subject had any known hearing impairment. All were native speakers of Spanish.

Materials. The experimental items were the same as in the previous experiment. 60 new sentences were now constructed to act as fillers. 24 of these fillers were long, biclausal sentences. Another 24 were monoclausal sentences with a different word order from the canonical subject-verb-object. The remaining 12 fillers were exactly like the experimental items (six of type A and six of type B) but did not carry a tone at all. This was also the case for 12 other fillers; namely, six biclausal and six non-canonical sentences did not contain a click. In total, 20% of the items did not have a click. A further six sentences were constructed in order to be included in the practice phase. All these sentences were recorded by the same speaker and using the same equipment. Regarding the comprehension task, 26 questions were constructed, 12 for the fillers, 12 for the experimental items, and 2 for the practice session. Thus, a question would appear in 20% of all items. The questions were rather simple in formulation and would query an uncomplicated aspect of either the subject, the object, or the verb of the corresponding items. The answer required was either a *yes* or a *no*. All other significant aspects of the task (generation and introduction of tones, etc.) remained unchanged from the previous experiment.

Procedure. The same as in the previous experiment, but with the addition of the fillers and the comprehension task. The fillers and the experimental sentences were randomised together for this version, which naturally included the questions some of these items were associated with. Regarding the comprehension task, each question appeared on the computer screen and the participants recorded their answers by pressing either the S key (for *sí*, that is, yes) or the N key (for, well, *no*). Upon entering the answer, a message on the computer screen would instruct the subject to press the space bar in order to hear the next sentence. Participants were also advised not to press a button if a sentence did not have a tone. The experimenter made sure that every subject understood the procedure during the practice session. The overall task was divided into three even blocks. During the break, the computer screen would turn white and subjects would be instructed to rest and relax, but to not disturb the others. The break would last two minutes, and at the end the screen would turn black in order to signal that the break had finished. The practice phase was now composed of 9 items and a practice sentence was placed at the beginning of each block. The practice session included at least an example of each new type, including two sentences without a tone and two questions. The session would restart as soon as the participants pressed the space bar. A third and final white screen indicated that the overall session had finished. In all other significant respects, the new task remained exactly the same as in the previous experiment. The session was now significantly longer, however, taking close to 40 minutes to complete.

3.2.2 Results

Ten participants were eliminated as they did not meet reasonable expectations regarding average performance. In particular, two subjects had an average response time that was close to 2 seconds, while another subject failed to record a single response. An analysis of the comprehension task showed that participants hardly made any errors, and apart from a participant who erred in 40% of the questions, everyone else was well under that figure. As we had settled on a 30% cut-off, only a further single subject was eliminated. The responses of the remaining participants were collected and polished following the same procedure of the previous experiments. Again, responses deviating 2.0 SD above or below the mean of each participant were eliminated; in this case, 4.0% of the data were affected.

As before, the reaction times, summarised in Table 2, were analysed with SPSS.

Table 2: Experiment 1b. RTs per click position per sentence type (Mean RT with standard deviations in parentheses).

	Click Position		
Sentence Type	1	2	3
A	340.71 (89.8)	290.86 (79.1)	283.00 (67.4)
B	335.42 (88.9)	296.54 (96.5)	291.25 (81.5)

As in Experiment 1a, and for each sentence type, RTs were greatest in the first click position and decreased thereon, suggesting that this pattern is very robust indeed. The analyses of variance with subjects and items as random factors once again showed that the *click position factor* was significant ($F_1(2,130) = 70.21$, $p < .001$; $F_2(2,118) = 36.61$, $p < .001$; $minF'(2,218) = 23.77, p. < 001$), while the *sentence type factor* did not prove to be significant in either analysis ($Fs < 1$). The interaction effect was also not significant ($F_1(2,130) = 1.5$, n.s.; $F_2 < 1$). As before, post-hoc pairwise comparisons were carried out. All within-sentence type pairs proved to be significant ($ps < .01$) except for A2-A3 ($t_1(65) = 1.5$, n.s.; $t_2 <$

1) and B2-B3 ($t < 1$ in both analyses).[7]

3.3 Discussion

As is clear from the results listed above, none of the hypotheses we had advanced were confirmed. Instead, and perhaps rather surprisingly, RTs were in fact greater at the very beginning and decreased from then on. Moreover, there was a regular decrease in RTs from the first to the last position in both sentence types, especially in Experiment 1a. This decreasing progression is rather robust, and the high significance of the (click) position factor is further confirmation. Prima facie, one could suggest that this is in line with the expectation that processing effort decreases as the sentence is presented —the least linguistic material to process, the easier it will be to respond to the tone— but this cannot be the whole story, as there were no differences in RTs across sentence type, and that needs to be explained. That is, each tone is placed in a rather different segment in each sentence type, and as such the parser cannot be computing the same expectations/predictions at each click position. This ought to be especially significant when it comes to whether the verb has appeared or not, the most central element of a sentence (and a proposition), and the one element the parser would be expecting the most. As the data show, though, the presence or absence of the verb doesn't appear to have had an effect.

We suspect that what is in addition at play here is the position effect, which we take to be a perceptual phenomenon involving a non-linguistic type of uncertainty. Let us elaborate a bit more now. On the one hand, and rather generally, it is to be expected that participants would be progressively better prepared to respond to the tone the more settled they are. On the other hand, and more importantly, as the sentence is presented, and if the tone hasn't yet appeared, a subject would increasingly be less surprised/uncertain when it finally does, and as such it ought not to be surprising that participants are faster in responding to the tone when it appears towards the end of sentences. Note that we are not conflating the processing effort involved in parsing and this perceptual phenomenon into one single factor (the position effect). Rather, we are suggesting, first of all, that the decrease in RTs is the result of the combination of both factors; in addition to this, however, the position effect appears to be so strong that differences across sentence type disappear altogether. As we will show below, however, the two factors can be discriminated and segregated, confirming our analysis. For now, we turn to a set of sub-analyses of the data of these two experiments.

Given the high significance of the click position factor, we conflated the two sentence types into one single condition in order to assess the significance of the click position when treated as a simple effect. That is, the RTs to position 1 in both sentence types were merged, an average calculated and the same process was

[7] A different version of this experiment, one with no breaks, was also run, but neither the RTs pattern (395, 355, 345 for Type A sentences; 394, 356, 337 for Type B) nor what transpired with the filler sentences or the comprehension task varied significantly. The data of this other experiment will be mentioned in the discussion.

applied to the other two positions. The overall averages for positions 1, 2, and 3 in the two experiments are represented in Table 3.

Table 3: RTs per click position treated as a simple factor (Mean RT with standard deviations in parentheses).

	Click Position		
Experiment	1	2	3
1a	254.81 (5.9)	219.92 (4.5)	206.02 (4.6)
1b	338.06 (10.7)	293.70 (10.3)	287.13 (8.8)

Pair comparisons show that the differences in RTs among the three positions were significant in the first experiment (all *ps* $< .01$), while the 2-3 pair did not prove to be significant in the second ($t_1(65) = 1.5$, n.s.; $t_2(59) = 1.0$, n.s.). Given the robust decrease in RTs, we decided to analyse the RTs to the filler sentences in Experiment 1b (and in the version mentioned in ft. 7) in order to check if the uncertainty effect played any role in the response pattern there too. Given that the clicks were introduced in a somewhat random manner in the construction of the fillers, a correlation analysis was conducted in which x stood for the number of syllables after which the click would appear in each item and y was the reaction time to the click. The Pearson's correlation was $r_{xy} = -.633$, $p < .01$ in Experiment 1b and $r_{xy} = -.681$, $p < .01$ for the footnoted experiment, indicating that the greater the number of syllables (that is, the deeper into the sentence the click is), the lower the reaction time to it. Plausibly, then, the factors we have identified are having a great effect in participants' performance in the overall task.

Another aspect of some interest is the different level of performance obtained in Experiment 1b compared to Experiment 1a, as the RTs in the former are significantly higher than in the latter (and the RTs were even higher in the experiment we ran without breaks).[8] This is of course not entirely surprising: a simpler experiment will yield a faster performance in participants. What a simpler experiment also yields, in this case at least, is a cleaner set of data. Indeed, the linear decrease in RTs is actually more pronounced in Experiment 1a than in the other two, as an analysis of the RTs subject by subject and item by item shows: 50% of subjects and 35% of items exhibit the decreasing tendency in Experiment 1a, for 22% of subjects and 8% of items in Experiment 1b and 27% of subjects and 13% of items in the experiment without breaks. Overall, then, our first experiment, which was

[8] The average response time in Experiment 1a was 240.68 ms., increasing to 313.67 in Experiment 1b, and are even higher in the experiment with no breaks (381.34). In fact, the RTs of the first experiment are in line with previous monitoring experiments, which by large used very few experimental sentences and usually no fillers or comprehension task, perhaps a by-product of the instruments then available.

only composed of critical items, yielded the most robust results.

Moving on, in the next experiment we change tack and proceed to attempt to discriminate the two factors we have postulated here but which are not easily discernible in the behavioural data so far obtained (the position effect and processing load). In order to do so, we combine click monitoring with the measurement of electrophysiological responses to external stimuli (i.e., ERPs).

4 Experiment 2

In this experiment, only type A sentences from the previous experiments were employed. As we aimed to probe and discriminate the position effect from the processing load the parser goes through as the sentence is presented, there really was no reason to employ both sentence types; the click positions, however, remain the same. Moreover, we chose type A sentences over type B sentences because it is easier to create type A sentences than type B ones, and the ERP experiment required a greater number of items than we had used in Experiment 1. In total, 120 sentences were used, but neither fillers nor a comprehension task were included in this version. We decided to not include fillers or a comprehension task for two main reasons. Firstly, their inclusion would have made the experiment a very long one indeed, and considering what is involved in an ERP experiment, we decided that this could negatively affect the performance of our participants. Moreover, and in consonance with what we defended in the previous section, we are unsure that tasks such as the click detection really necessitate them, and therefore we concluded that their exclusion would in fact be positive rather than negative.

There are two relevant predictions at hand here. It was, first of all, hypothesised that the N1 wave, a component that is associated with temporal uncertainty (Näätänen & Picton, 1987), would correlate with the RTs, and thus its amplitude would be highest at the first click position, the perceptual uncertainty of the participants being greatest at that point, and decrease thereon. This part of the experiment aimed to evaluate the significance of the position effect, and the N1 is a pertinent component for such a task, given that it tracks perceptual processes rather than (higher) cognitive ones. Regarding processing load, we decided to concentrate on the P3 (or P300), a component whose amplitude to a secondary task has been shown to be affected by the difficulty of the primary task in dual-task settings such as ours. Past results with dual-task experiments (e.g., Wickens, Kramer, Vanasse & Donchin, 1983) indicate that the P300 associated with a secondary task (in this case, reacting to the tone) will have a low amplitude if the primary task (here, parsing the sentence) is of considerable difficulty. In other words, there will be a correlation between the fluctuations in difficulty in a primary task and the amplitude of the P300 to a secondary task. In our experiment, as the primary task decreases in difficulty (as manifested by the linear decrease in RTs from the first to the third position), the amplitude of the P300 was predicted to increase from position 1 onwards. That is, as the sentence is being processed, the number of predictions and expectations the parser needs to fulfil is reduced, and thereby more resources can be allocated to responding to the click, something that should be reflected in the amplitude of the P300. There has been some discussion in the literature as to whether the P300 constitutes a good metric for memory load in cognitive resources, but doubts seem to be strongest regarding single task experiments; dual tasks appear to provide the right setting for the employment of the P300 as a reliable measure of processing effort (see Kok, 2001, for extensive discussion of these issues).

Crucially for our purposes, the biphasic pattern we are hypothesising is well established in the dual tasks literature. Both Sirevaag et al. (1993) and Wickens et al. (1983) report an N1-P300 pattern when an auditory probe is employed, and this is precisely what we are after here: an N1 wave tracking perceptual processes, and a P3 component tracking cognitive processes. To our knowledge, moreover, this is the first time that the P300 is employed in a study of syntactic processing as a metric of processing effort, and we hope our results constitute evidence for its general usefulness in psycholinguistics. Naturally, these two hypotheses hold if and only if the pattern in RTs obtained in the previous experiments does not vary, but we hypothesised that this would be the case indeed.

4.1 Method

Participants. 18 psychology students (2 male, 16 female) from a pool of different courses participated in the experiment. This was a different set of participants from the previous three experiments. The mean age was 22 years, and no subjects had no known hearing impairments. All were native speakers of Spanish.

Materials. The same as type A sentences from the previous experiments, but these now numbered 120 items. There were no fillers, nor a comprehension task was included.

Procedure. Participants were seated in a comfortable chair in a sound-attenuated, darkened, and dimly illuminated room. One individual was tested at each time. In this experiment, they were exposed to a total of 120 items, presented in three blocks. The pauses were of the same length as in experiment 3, and apart from the EEG measures that were undertaken, the task remained the same as in the previous experiments. The EEG was recorded continuously by 19 Ag/AgCl electrodes which were fixed on the scalp by means of an elastic cap (Electrocap International, Eaton, OH, USA) positioned in accordance with the 10-20 International system. In addition, one electrode was placed under the left eye in order to monitor blinking and vertical eye movements and another electrode was placed at the outer canthus of the right eye in order to monitor horizontal eye movements. Recordings were referenced to the right earlobe, whilst the left earlobe was employed as an active recording channel. Subsequently, ERPs were algebraically re-referenced to linked earlobes offline. Electrode impedances were kept below 5 $k\Omega$. All EEG and EOG channels were amplified using a NuAmps Amplifier (Neuroscan Inc., North Carolina, USA) and recorded continuously with a bandpass from 0.01 to 30 Hz and digitised with 2 ms. resolution. The EEG was refiltered off-line with a 25-Hz, low-pass, zero-phase shift digital filter. Automatic and manual rejections were carried out to exclude periods containing movement or technical artefacts (the automatic EOG rejection criterion was $\pm 50\ \mu$V).

4.2 Results

Behavioural Data

The reaction times of the 18 participants were collected and polished with the DMDX programme. As before, responses deviating 2.0 SDs above or below the mean of each participant were eliminated, which in this case affected 3.6% of the data. The statistical analysis was carried out with SPSS, the data being shown in Table 4.

Table 4: RTs per click position (Mean RT with standard deviations in parentheses).

	Click Position	
1	2	3
325.05 (64.7)	266.53 (36.5)	247.60 (31.0)

As expected, the RTs manifest the exact same pattern as in Experiments 1a and 1b: reaction times decrease from the first position onwards. A repeated-measures analysis of variance showed that the *click position factor* was significant for both the subjects and items analyses ($F_1(2,34) = 39$, $p < .001$; $F_2(2,238) = 93$, $p < .001$; $minF'(2,67) = 27, p < .001$). Regarding pair comparisons between the different click positions (1 vs. 2, etc.), the analyses showed that all comparisons were significant (all *ps* $< .01$).

Electrophysiological data

The data were processed using BrainVision Analyzer 2 (Brain Products, Gilching, Germany). Average ERPs were calculated per condition, per participant from -100 to 700 ms. relative to the onset of the click, and before grand-averages were computed over all participants. A 100 ms. pre-click period was used as the baseline. Only trials without muscle artefact or eye movement/blink activity were included in the averaging process. The nine relevant electrodes used in the analyses correspond to the International 10-20 Electrode System (Jasper, 1958) locations of Fz, Cz, and Pz along the midline; F3, C3, and P3 over the left hemisphere; and F4, C4, and P4 over the right hemisphere. We used this more restricted montage for two reasons: first, to simplify the exposition of the results, as using the entire montage would require multiple ANOVAs per epoch; and second, because the major effects to be reported were most clearly seen in these sites. Analysis involved a repeated-measures ANOVA with within-participants factors Click Position (position1, position2, position3), Location (anterior, central, posterior), and Laterality (left, middle, right). Omnibus ANOVAs were followed up with pairwise comparisons intended to discern whether there were differences among the three click positions. Based on both visual inspection of the results and prior reports, two time windows were selected for analysis of the mean amplitudes of the components of interest: the N1 component was analysed from 120 ms. to 200 ms. and the P300 component was evaluated from 230 ms. to 400 ms. The Greenhouse and Geisser (1959) correction was applied to all repeated measures having more than one degree of freedom in the numerator. In such cases, the corrected p-value is reported. In order to not clutter the presentation of our results, we only report the main effect of the Click Position factor and the significant interaction effects between this factor and the others.

Fig. 1 depicts brain potential variations in the linear derivation of a group of six electrodes (C3, Cz, C4, P3, Pz, P4). As can be seen in Fig. 1, the three click positions exhibit a clear biphasic pattern, with a first modulation in the N1 time window, which is followed by a second modulation in the P300 time window.

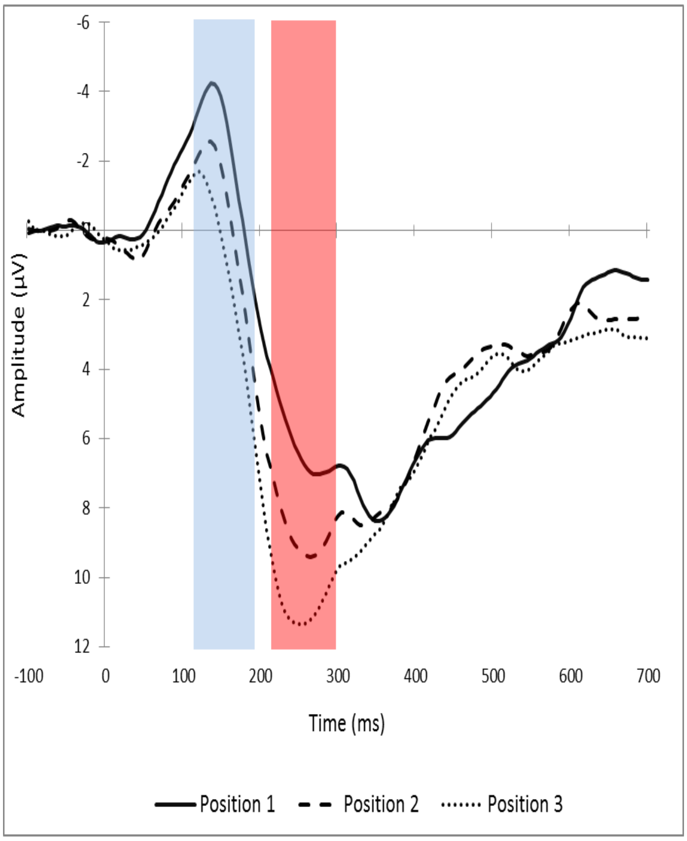
N1 epoch (120-200 ms). During the N1 epoch, there was a main effect of Click Position, $F(2,34) = 17.063$, $p = .000019$, $n_p^2 = 0.501$. Bonferroni-corrected pairwise comparisons showed that all three positions differ from each other significantly (all $ps < .025$), reflecting a more negative-going amplitude for position 1 relative to positions 2, and a more negative-going amplitude for position 2 relative to position 3. The two-way interactions between Click Position and Location and between Click Position and Laterality were not reliable. The three-way interaction between Click Position, Location and Laterality was not significant either.

P300 epoch (230-400 ms). During the P300 epoch, there was a main effect of Click Position $F(2,34) = 20.224$, $p = .000003$, $n_p^2 = 0.543$. Bonferroni-corrected pairwise comparisons showed that all three positions differ from each other significantly (all $ps < .025$), reflecting a less positive-going amplitude for position 1 relative to positions 2, and a less positive-going amplitude for position 2 relative to position 3. The two-way interactions between Click Position and Location and between Click Position and Laterality were not reliable. The three-way interaction between Click Position, Location and Laterality was not significant here either.

4.3 Discussion

As the behavioural data show, the prediction regarding the RTs pattern was confirmed; that is, RTs to the first click are slowest, and then become faster thereon. It is worth pointing out that the averages are closer to the RTs of Experiment 1a than to Experiment 1b, and this despite the fact that participants were exposed to a total of 120 experimental items. In any case, and as expected, all pair comparisons were significant. As such, this allows us to discuss the ERP data in the terms we

Figure 1: ERP waveforms for the three click positions shown from a 100 ms. before click presentation to a 700 ms. post-click presentation. The waveforms depict brain potential variations in the linear derivation of a group of six electrodes (C3, Cz, C4, P3, Pz, and P4). Colour boxes indicate significant differences elicited by Click Position in the N1 range (blue) and in the P300 range (red). Negative voltage is plotted up.



had devised.

The ERP data we obtained confirm the hypothesised amplitudes for both the N1 and the P300 components. Indeed, the N1 pattern indicates that Ss are indeed uncertain as to when the click is going to appear, and their uncertainty decreases as the sentence is being presented. This is of course unsurprising; as the sentence progresses and gets closer to the end, the chances that the click will finally appear increase, and the uncertainty thereby decreases. We had already stated that the linear decrease in RTs was due to a combination of two factors and the N1 data confirm that there is indeed a purely perceptual factor at play, what we called earlier the position effect and which was called upon in order to account for the very meagre differences between sentence types in Experiments 1a and 1b. As we will show in the next experiment and discuss in the last section of this paper, it is in fact possible to behaviourally disentangle the two aspects we have posited (the position effect and the parser's processing load).

Regarding the P300, its pattern can be explained in terms of task difficulty. As the amplitude of the P300 increases from position 1 onwards, and there is furthermore a negative correlation between RTs and the amplitude of the P300, this confirms that as the sentence is being processed, the parser's unfulfilled predictions and expectations decrease, and thereby more resources can be allocated to monitoring the click. We mentioned above that reacting to the click (that is, pressing a button) constituted the secondary task, but it is more appropriate to state that monitoring the click is the secondary task, while reacting to it is in fact a tertiary task. In principle, in fact, the P300 pattern should be obtained in the absence of this tertiary task altogether. As a case in point, Sirevaag et al. (1993) carried out a study of the resources flying pilots can devote to monitoring a tone (either high or low pitched) whilst communicating to air-traffic controllers (this experimental condition varied in terms of its frequency), and the P300 wave they measured varied alongside the very dimension we have focused on here. That is, the amplitude of the P300 associated with the secondary task (monitoring the tone) was greater when the primary task (communicating with air controllers) was easier, as in the low communication load condition these scholars had employed. Note that in that study the pilots did not have to react to the tone, all they had to do was pay attention to it; that is, they had to be aware that a tone would appear (in other words, they were monitoring it).

This is likewise the case in our experiment, we believe (recall, also, that the study just discussed employed an auditory probe —to wit, a tone— and that the same type of biphasic pattern was observed). To be sure, the motor response to a stimuli can also have an effect on ERP components, but we don't think that aspect of the task affects our explanation of the data. Clearly, we have here compared the amplitude of the P300, which is related to the monitoring of a tone, with the reaction times to a click, a motor response, and these are prima facie two different things. However, we have no option but to take the RTs as a measure of the monitoring itself, even though what we may well be dealing with here is a two-part process (monitoring plus reaction). This is certainly common ground in the field, and it is not clear what other behavioural measure could have been employed for the purposes at hand.

All in all, we have succeeded in discriminating —that is, recording— the two factors we had posited, and in the next experiment we will show how they can in addition be segregated in behaviour. In the last section of the paper, we shall have a few things to say about how these two factors may enter into an explanation of previously-obtained data.

5 Experiment 3

A number of factors and some experimental evidence suggest that the end of a sentence ought to exert a particular cognitive load in the parser. We are not referring to the end-of-clause effect reported in Abrams and Bever (1969) and Bever and Hurtig (1975); we are not entirely sure about those data, given what we have said about the two factors that seem to be operative in monitoring tasks —and, in any case, in those studies the end of the clause was never the end of a sentence. Rather, what we have in mind is the self-paced reading literature, wherein a wrap-up effect seems to be a very robust phenomenon, as evidenced in the tendency of reading times to be much higher towards the end of a sentence (Just, Carpenter & Woolley, 1982). This is prima facie a very natural result to obtain; it would be predicted to follow from the very general point that the more material the parser is inputted, the more strained working memory is likely to be. After all, the ceaseless inflow of material should result in a ever-greater number of open nodes, and these would only be "branched off" at the end of a sentence. Ambiguity resolution, the closing of the sentence node, and the construction of the underlying proposition are some of the other phenomena that further suggest a wrap-up effect —or at least this is what is assumed in a number of models of structural complexity and parsing.[9]

As mentioned, the wrap-up effect is related, but should not be confused with, the clause-by-clause strategy of parsing defended in, e.g., Fodor et al. (1974). Fodor et al. described this strategy in such a way as to suggest that semantic and pragmatic operations would only become operative once a clause had been completed, a view that was vigorously contested by, e.g., Tyler and Marslen-Wilson (1977). According to the data reported by the latter, the processor attempts to provide a meaningful representation to the material it receives *as soon as it receives it*, a phenomenon that has come to be known as *incremental* parsing. Because of its apparent incompatibility with incrementality, the clause-by-clause strategy is sometimes described with derision in some textbooks (e.g., Harley, 2001, p. 228), but incremental parsing and the importance of the end of a sentence (to adapt the clause-by-clause idea to our interests) in processing, what we call here the wrap-up effect, are entirely compatible phenomena once we abandon the view that semantic processes await the completion of a clause.

Having said that, it is certainly the case that the incrementality principle suggests some modifications regarding the manner in which we have just described the general issue now under study. That is, perhaps it is not all that reasonable to suppose that the parser opens up phrasal nodes that it can only close off at the end of a clause or a sentence. Instead, an incremental type of processing demands that a meaning/interpretation is constructed for each segment the processor receives, as it receives them to boot, and that would suggest that a syntactic object of some

[9]This point applies to a variety of models: Gibson's (1998) *storage-and-integration-costs* theory, Hawkins's (2004) *immediate constituents* account, and, perhaps more clearly, in Frazier and Fodor's (1978) *sausage machine*, as the latter postulates a syntactic process in which incoming elements are kept in working memory until they can be packaged into phrases.

sort is ipso facto constructed at that point too. That much is to be granted, but the importance of the end of a sentence as a locus in which the overall proposition is readjusted remains.

In any case, the wrap-up effect was not applicable in the previous experiments, as the last click there was nowhere near the end of the sentences. Here, we change the positions of the clicks and modify type B sentences from Experiment 1 in order to check if by placing a click at the end the strong tendency for RTs to decrease is disrupted. We only use type B sentences because these exhibit a complex noun phrase in the object position, and this is a better configuration for the purposes at hand, considering that the object position is naturally closer to the end of the sentence. Further, only one sentence type is necessary for our purposes, as we are only aiming to disrupt the decrease in RTs within a sentence —that is, we are interested in comparing click positions within a sentence type, not across sentence types.

Three click positions are maintained, one at the beginning of the sentence and two within the verb's complex complement; their precise position is specified in the following section. It is hypothesised that the wrap-up effect is indeed applicable at the end of a sentence and therefore the pattern in RTs should be different from the pattern observed in the previous experiments. In particular, we expect a V-shape pattern in which RTs to the first position are highest, descending significantly for the second position, but then raising for the third and last position, the postulated locus of the wrap-up.

5.1 Method

Participants. 48 psychology students (5 male, 43 female) from the Rovira i Virgili University (Tarragona, Spain) participated in the experiment for course credit. This was a different set of participants from all other experiments. The mean age was 22 years, and none of the subjects had any known hearing impairments. All were native speakers of Spanish.

Materials. Type B sentences from Experiment 1 were employed, but 14 of these had to be slightly modified or substituted by new ones. While preparing the experiment, it was noticed that in 46 of the 60 experimental sentences the parser could well carry out a wrap-up at the second boundary, as none of these sentences required any further material at that point. Consequently, it was decided that these 14 sentences be modified or replaced by new ones so that the experimental materials were as homogeneous as possible. The click positions were also modified in order to evaluate the wrap-up effect, and as before, were always placed on vowels. The first position was now placed on the third syllable, the second on the eleventh (and thereby far from the end of the internal phrase, thus avoiding the aforementioned possible wrap-up), and the third on the nineteenth, just one syllable before the end of the sentence (note that eight syllables separates each position). The following sentence shows the new click positions (where |, in this case, marks the placement of the tone):

(1) El candi|dato ha preparado un di|scurso sobre la sani|dad.

In all other respects, the task did not change.
Procedure. The same as in Experiment 1b, with the exception of the comprehension task, which in this case was not included. The session took around 25 minutes to complete.

5.2 Results

The responses of 11 participants had to be eliminated for a number of reasons: some failed to achieve a reasonable standard in performance, while others did not manage to register a single response. The reaction times of the remaining 37 subjects were collected and polished with the DMDX programme. As ever, responses deviating 2.0 SDs above or below the mean of each participant were eliminated, here affecting 3.8% of the data. The analysis of the reaction times, shown in Table 5, was carried out with the SPSS package.

Table 5: RTs per click position (Mean RT with standard deviations in parentheses).

	Click Position	
1	2	3
414.16 (51.3)	351.88 (33.9)	365.45 (37.4)

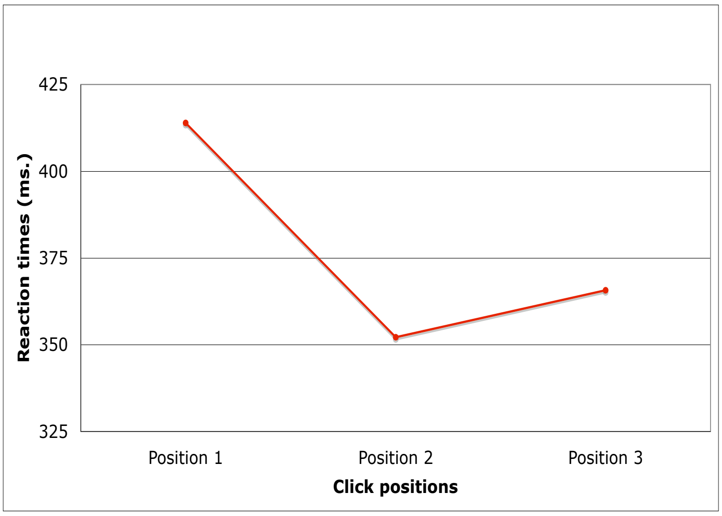
In this experiment, RTs were also greatest in the first position, but there was no decrease from the second to the third position; instead, there was a slight increase. A repeated-measures analysis of variance showed that the *click position factor* was significant for both the subjects and items analyses ($F_1(2,72) = 98$, $p < .001$; $F_2(2,118) = 110$, $p < .001$; $minF'(2,173) = 51.82, p < .001$). All post-hoc pairwise comparisons proved to be significant (all $ps < .001$).

5.3 Discussion

As predicted, the wrap-up effect was detectable with the click-monitoring task, and therefore it is a feature of speech processing as much as of reading. Moreover, the wrap-up effect disrupted the linear decrease in RTs somewhat, as can be seen in Fig. 2.

Indeed, even though RTs to the first position were greatest and there was a noticeable decrease from the first to the second position, the cognitive load associated with the wrap-up effect resulted in an increase in RTs from the second to the third position, in clear contrast with what was obtained in the previous experiments, and resulting in the V-shape pattern observed in Fig. 2. This would seem to indicate

Figure 2: RTs progression in Experiment 3



that the click paradigm is not entirely hostage to perceptual factors such as the position effect; a design can be found so that structural properties are brought out more clearly, resulting in the clear segregation of the two factors that have animated the whole discussion. If anything, this is behavioural confirmation of what was observed on the ERP record, vindicating the usefulness of click monitoring as a psycholinguistic technique.

Note, however, that RTs to the first position are still significantly greater than RTs to the last position, indicating that the processor is more strained there than at the end of sentences. Also, we are not at all certain as to what sort of operations precisely the wrap-up involves, only that something peculiar to that location does take place. It seems safe to suppose, given the evidence amassed in favour of incrementality (for a review, see Harley, 2001), that what doesn't take place is the branching off of the nodes the parser would have reputedly opened during sentence processing, as the *sausage machine* model (cited supra) would have it. This aspect of the wrap-up possibly merits more attention, but we here simply and tentatively suggest that what takes place is some sort of "putting it all together" process wherein the underlying proposition a sentence is said to express is completed and/or refined.

Be that as it may, whether this effect can be related to the end-of-clause effect apparently unearthed in previous experiments is not so clear. In those studies, and as already stated, the end of a clause was in fact the end of a subordinate clause within complex, biclausal sentences, with all the orbiting issues that arise therefrom. Moreover, the end-of-clause position is usually also the first click position of the sentences employed in those studies, pointing to the probable role of the position effect. We come back to some of those issues in the last section.

6 Part of the Past and Something for the Future

So this is what we have done here. First off, we employed two types of simple, declarative sentences to probe if the linear decrease in RTs observed in phoneme- and word-monitoring tasks is also a factor —and if so, how strong of a factor— in the monitoring of a tone during the parsing of a sentence. In order to mitigate the probable decrease in RTs, the two types of sentences we constructed varied along how much material appeared on either side of the verb in subject-verb-complement(s) constructions, where the tones would be placed (this manipulation yielded two structural conditions). The results show two things: a) a pronounced decrease in RTs for each individual sentence, which suggests that the parser's processing load decreases as the sentence is presented, and releasing more cognitive resources to monitor the click in doing so; and b) no significant differences in the RTs to the same click locations across sentence type, a datum that points to the purely perceptual properties involved in monitoring a tone. These two factors were then separately observed in an ERP experiment, and indeed separated in behaviour in a modified version of the main experiment.

The position effect seems to have gone entirely unnoticed in all previous studies of click monitoring, whilst the processing load factor (which we believe to follow from the incrementality property) has only been discussed by Holmes and Forster (1970). On the other hand, the originators of this technique, Abrams and Bever (1969), explained their data solely in terms of what they called the end-of-clause effect (see, also, Bever & Hurtig, 1975), but the two factors we have analysed here seem to be clearly operative there too, and that muddies their data significantly. Indeed, these scholars placed the first tone at the end of a clause, and that confuses and conflates perceptual and cognitive load factors with the structural effects they were looking for. In addition, their end-of-clause effect cannot be identified with our wrap-up effect —they belong to different parts of a sentences— and thus we really don't know how strong that effect is —if it is indeed there.

Putting it all together, we believe that the two factors we have identified here —the position effect and processing load— conspire to yield the RTs that can be obtained with the click monitoring technique, and as a result future experiments employing this technique, we advise, will need to take this contingency into consideration. In our study we have shown that the two factors can be teased apart, especially when one sets out to do so, but their interaction is also observable in the more complex sentences used in past experiments —one just has to select the right prism with which to analyse those data. Take the experiments in Cohen and Mehler (1996) as a case in point, which we now approach through an analysis that speaks to our concerns and certainly not to theirs; in particular, we will show that in this study too differences in RTs disappear across differing structural conditions when the position effect overcomes these experimental conditions, a factor that has gone unnoticed until now. It is hoped that this analysis further illuminates the main point of our study, and put an end to this paper to boot.

In the first three experiments reported in Cohen and Mehler (1996), length was controlled for across two types of sentences and different RTs were recorded in the same position across these sentence types, which naturally suggests purely structural effects. Cohen and Mehler (1996), however, used relative clauses, which are certainly more complex than the non-relative structures that both Abrams and Bever (1969) and Holmes and Forster (1970) employed. That the position effect appears to have been nullified in Cohen and Mehler (1996) may be the result of pushing memory resources to the limit, perhaps the key to eliciting structural effects in the click-detection task, and what we achieved ourselves in Experiment 3. A closer look at their materials will further illustrate.

In the first experiment, Cohen and Mehler (1996) compared reversible subject and transposed object relatives in French, a pertinent pair to compare given that in this condition the complementiser is the only differing element between the two sentences (*qui* in subject relatives, *que* in the object constructions). Consider the following pair, where the | symbol marks where the click was placed and the numbers within brackets indicate the RTs (the translation for these French sentences can be found in the original paper).

(2) Le savant (qui connait le docteur) t|ravaille. . . (218 ms.)

(3) Le savant (que connait le docteur) t|ravaille. . . (234)

Note that the RTs to a click placed right after the embedded clause indicates, in agreement with the literature (see some of the references mentioned in Gibson, 1998), that object relatives are harder to process than subject relatives. In a second experiment, these results were replicated (the RTs were 248 and 272, respectively) and then compared to a new click position: right before the end of the embedded clause.

(4) Le savant (qui connait le d|octeur) travaille. . . (249 ms.)

(5) Le savant (que connait le d|octeur) travaille. . . (250)

Interestingly, RTs to clicks before the end of a clause are not different across sentence type, and perhaps this is not unexpected, given that the adjustment processing a relative clause comports supposedly takes place once the internal phrase has been completed, and not before. If so, it seems that the greater cognitive load an object relative exerts is in fact operative *after* the embedded clause has been processed, but not during it. In the third experiment, though, the object relative was presented in its natural canonical order (i.e., it was not transposed), tones were placed after the embedded clauses, and the differences in RTs disappeared altogether:

(6) Le savant (qui connait le docteur) t|raivalle... (262 ms.)

(7) Le savant (que le docteur connait) t|raivalle... (264)

The last datum is interesting, as it suggests that object relatives in their normal manifestation are not more difficult to process than subject relatives —at least while participants monitor a tone.[10] We submit that when the differences in RTs disappear, this is due to the position effect, suggesting that in some cases the structural differences —the only factor Cohen & Mehler manipulated— are not as pronounced as they would prima facie appear to be. If so, the position factor has a great role to play in the explanation of the response patterns in these click-detection tasks too, but this has so far been ignored, or missed. More precisely, the design of a monitoring experiment appears to influence whether the position effect is operative or not —and to what extent it is.

The take-home message, then, is that reacting to a tone demands non-trivial attentional resources. We have shown that, on the one hand, two very different factors have a great influence upon response patterns in general; and on the other hand, that designing a click-monitoring experiment requires careful consideration if structural effects are being sought —the right balance needs to be struck between the quantity and complexity of the experimental materials and the memory threshold that is involved in processing them. As things stand, then, the results of the classic click-monitoring studies do not appear to us to be on very solid ground any more —is there really an end-of-clause effect?—, and a revisit to biclausal sentences and the like seems in order. We hope to contribute to that line of work in the near future.

7 Acknowledgements

The research presented here was funded, at least in part, by a Beatriu de Pinós fellowship awarded by the Catalan Research Council (AGAUR) to the first author (2011-BP-A-00127), and by an AGAUR research grant awarded to the Psycholinguistics Research Group at the Universitat Rovira i Virgili in Tarragona, Spain (2014-SGR-1444).

[10] A caveat. We do not have any data for the last possible combination: tones inside the embedded clause in the normal, canonical order. It could well be the case that the greater cognitive load of objective relatives is present after the embedded clause in the transposed condition but before it in the canonical order.

References

Abrams, K. & Bever, T. G. (1969). Syntactic structure modifies attention during speech perception and recognition. *The Quarterly Journal of Experimental Psychology*, *21*(3), 280-290.

Almela, R., Cantos, P., Sánchez, A., Sarmiento, R. & Almela, M. (2005). *Frecuencias del Español. Diccionario y estudios léxicos y morfológicos*. Madrid, Spain: Editorial Universitas.

Berwick, R. C. (2011). All you need is Merge: biology, computation, and language from the bottom up. In A. M. Di Sciullo & C. Boeckx (Eds.), *The biolinguistic enterprise* (p. 461-491). Oxford University Press.

Bever, T. G. (1973). Serial position and response biases do not account for the effect of syntactic structure on the location of brief noises during sentences. *Journal of Psycholinguistic Research*, *2*, 287-288.

Bever, T. G. & Hurtig, R. R. (1975). Detection of a nonlinguisic stimulus is poorest at the end of a clause. *Journal of Psycholinguistic Research*, *4*(1), 1-7.

Bever, T. G., Lackner, J. R. & Kirk, R. (1969). The underlying structures of sentences are the primary units of immediate speech processing. *Perception and Psychophysics*, *5*(4), 225-234.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9-10), 341-345.

Bond, Z. S. (1972). Phonological units in sentence perception. *Phonetica*, *25*(3), 129-39.

Chapin, P. G., Smith, T. S. & Abrahamson, A. A. (1972). Two factors in perceptual segmentation of speech. *Journal of Verbal Learning and Verbal Behavior*, *11*, 164-173.

Chater, N., Crocker, M. & Pickering, M. (1998). The rational analysis of inquiry: The case of parsing. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (chap. 441-468). Oxford, England: Oxford University Press.

Cohen, L. & Mehler, J. (1996). Click monitoring revisited: an on-line study of sentence comprehension. *Memory and Cognition*, *24*(1), 94-102.

Cutler, A. & Norris, D. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: psycholinguistic studies presented to Merrill Garrett* (p. 113-34). Lawrence Erlbaum.

Flores d'Arcais, G. B. (1978). The perception of complex sentences. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (p. 155-185). John Wiley and Sons.

Fodor, J. A. & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior*, *4*, 414-420.

Fodor, J. A., Bever, T. G. & Garrett, M. F. (1974). *The psychology of language*. London, England: McGraw-Hill.

Forster, K. I. & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, *35*, 116-124.

Frazier, L. & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*, 291-325.

Garrett, M. F., Bever, T. & Fodor, J. A. (1966). Active use of grammar in speech perception. *Perception and Psychophysics*, *1*, 30-2.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition*, *68*, 1-76.

Gómez, D. M., Bion, R. A. H. & Mehler, J. (2011). The word segmentation process as revealed by click detection. *Language and Cognitive Processes*, *26*(2), 212-23.

Greenhouse, S. W. & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95-112.

Harley, T. (2001). *The psychology of language*. Sussex: Psychology Press.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford, England: Oxford University Press.

Holmes, V. M. & Forster, K. I. (1970). Detection of extraneous signals during sentence recognition. *Perception and Psychophysics*, *7*(5), 297-301.

Jasper, H. H. (1958). The ten-twenty system of the international federation. *Electroencephalography & Clinical Neurophysiology*, *10*, 371-5.

Just, M. A., Carpenter, P. A. & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology*, *111*(2), 228-38.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557-77.

Levelt, W. J. M. (1978). A survey of studies in sentence perception: 1970-1976. In W. J. M. Levelt & G. B. Flores d'Arcais (Eds.), *Studies in the perception of language* (p. 1-74). John Wiley and Sons.

Levy, R. (2013). Memory and surprisal in human sentence comprehension. In R. P. G. van Gompel (Ed.), *Sentence processing* (p. 78-114). Hove, England: Psychology Press.

Lewis, R. L. (2000). Specifying architectures for language processing: process, control, and memory in parsing and interpretation. In M. W. Crocker, M. Pickering & C. Clifton Jr. (Eds.), *Architectures and mechanisms for language processing* (p. 56-89). Cambridge University Press.

Moro, A. (2008). *The boundaries of Babel*. Cambridge, MA.: The MIT Press.

Näätänen, R. & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375-425.

Reber, A. S. & Anderson, J. R. (1970). The perception of clicks in linguistic and nonlinguistic messages. *Perception and Psychophysics*, *8*(2), 81-89.

Schafer, A. J. (1997). *Prosodic parsing: the role of prosody in sentence comprehension.* PhD thesis, University of Massachusetts, Amherst.

Sebastián-Gallés, N., Martí, M. A., Carreiras, M. & Cuetos, F. (2000). *LEXESP. Léxico informatizado del español*. Barcelona: Edicions Universitat de Barcelona.

Sirevaag, E. J., Kramer, A. F., Wickens, C. D., Reisweber, M., Strayer, D. L. & Grenell, J. F. (1993). Assessment of pilot performance and mental workload in rotary wing aircraft. *Ergonomics*, *36*(9), 1121-40.

Townsend, D. & Bever, T. G. (2001). *Sentence comprehension*. Cambridge, MA.: The MIT Press.

Tyler, L. K. & Marslen-Wilson, W. D. (1977). The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 683-692.

Wickens, C., Kramer, A., Vanasse, L. & Donchin, E. (1983). Performance of concurrent tasks: a psychophysiological analysis of the reciprocity of information-processing resources. *Science*, *221*(4615), 1080-2.