Original Article

QJEP

Quarterly Journal of Experimental Psychology
2018, Vol. 71(3) 759–778

DOI: 10.1080/17470218.2016.1276610
qjep.sagepub.com
SAGE

The role of executive control in resolving grammatical number conflict in sentence comprehension

André Vandierendonck, Maaike Loncke, Robert J. Hartsuiker and Timothy Desmet

Abstract

In sentences with a complex subject noun phrase, like "The key to the cabinets is lost", the grammatical number of the head noun (key) may be the same or different from that of the modifier noun phrase (cabinets). When the number is the same, comprehension is usually easier than when it is different. Grammatical number computation may occur while processing the modifier noun (integration phase) or while processing the verb (checking phase). We investigated at which phase number conflict and plausibility of the modifier noun as subject for the verb affect processing, and we imposed a gaze-contingent tone discrimination task in either phase to test whether number computation involves executive control. At both phases, gaze durations were longer when a concurrent tone task was present. Additionally, at the integration phase, gaze durations were longer under number conflict, and this effect was enhanced by the presence of a tone task, whereas no effects of plausibility of the modifier were observed. The finding that the effect of number match was larger under load shows that computation of the grammatical number of the complex noun phrase requires executive control in the integration phase, but not in the checking phase.

Keywords

Choice reaction time task; Dual-task paradigm; Executive control; Eye tracking; Sentence comprehension; Subject–verb agreement

Received: 19 March 2015; accepted: 12 December 2016

When reading a sentence like *Can you can a can as a canner can can a can?*, it becomes clear that the meaning of a sentence cannot just be extracted on the basis of the individual words. In order to comprehend a sentence, it is essential that its syntactic structure is analysed. Syntactic analysis involves the detection of the mutual relations and the dependencies between each of the words in the sentence and of their syntactic functions. One of the cues that readers rely on to identify syntactic relations is the agreement in grammatical number of constituents. For example, in the Dutch sentence "Die man hebben de vrouwen gekust" (lit. "That man have the women kissed"), the grammatical number of the finite verb "hebben" (plural) agrees with that of the noun phrase (NP) "de vrouwen" (plural), but not with that of the NP "die man" (singular). This pattern of agreement and disagreement signals to the comprehender that "de vrouwen", and not "die man", is likely to be the subject of the verb (Bates, McNew, Macwhinney, Devescovi, & Smith, 1982). However, the computation of the subject–verb relation on the basis of number agreement becomes more difficult when the subject NP is a complex one consisting of two or more nouns carrying a different grammatical number. For example, sentences like (1) are generally more difficult to process than sentences like (2).

1. The key to the cabinets is lost.
2. The key to the cabinet is lost.

In sentences like (1), the grammatical number of the verb "is" (singular) agrees with the number of the *head NP* "the

Department of Experimental Psychology, Ghent University, Gent, Belgium

Corresponding authors:
André Vandierendonck, Department of Experimental Psychology, Ghent University, Gent, Belgium.
Email: Andre.Vandierendonck@UGent.be

Robert J. Hartsuiker, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, B-9000 Gent, Belgium.
Email: Robert.Hartsuiker@UGent.be

Timothy Desmet is presently Managing Partner at Profacts, Eke, Belgium.

key" (singular) within the *complex subject NP* "the key to the cabinets", but not with the number of the *modifier NP* "the cabinets" (plural) within that complex subject NP. In sentences like (2), the grammatical number of the verb "is" agrees with that of both the head NP and the modifier NP (both singular). Typically, a processing cost is found for grammatical sentences with a complex subject NP containing a singular head NP and a plural modifier NP, like (1), as compared to sentences, like (2), in which the head NP and modifier NP carry the same grammatical number. This is confirmed in generally slower reading times as assessed in a self-paced moving-window technique and in eye tracking (e.g., Lago, Shalom, Sigman, Lau, & Phillips, 2015; Pearlmutter, Garnsey, & Bock, 1999), but also in judgments about the grammaticality of the sentence (e.g., Nicol, Forster, & Veres, 1997; Tanner, Nicol, & Brehm, 2014). Forced choices between the singular and plural form of a verb (Staub, 2009) are also slower or less accurate following a (singular) complex NP containing a plural modifier NP than following a singular modifier NP. Furthermore, in grammatical sentences number mismatch between head and modifier noun results in slower reading, whereas ungrammatical sentences like (3) are read faster than ungrammatical sentences like (4) (Pearlmutter et al., 1999).

3. *The key to the cabinets are lost.
4. *The key to the cabinet are lost.

In a grammaticality judgment task with French sentences, Kail and Bassano (1997) found that the ungrammaticality of the subject–verb agreement was generally detected later in sentences like (3) than in sentences like (4). In sentences like (3), the plural number feature of the modifier NP thus seems to obscure the ungrammaticality of the subject–verb agreement. Taken together, the findings on grammatical and ungrammatical sentences suggest that it is more difficult to compute the relation between the subject and the verb when the intervening modifier NP carries a different grammatical number than the head NP.

Further evidence regarding grammatical number matching was collected in a series of self-paced reading experiments by Wagers, Lau, and Phillips (2009). They varied several features of the sentences and observed a consistent slow-down for plural modifier nouns compared to singular ones. This effect occurred at the modifier noun region and the next region. By interspersing an adverb between the modifier and the verb, they were able to show that no number effect was observed at the verb region and beyond. Importantly, in the verb region and beyond an effect of grammaticality was present, and this effect interacted with modifier number.

Theories of subject–verb agreement computation in comprehension put forward two processing phases at which the mismatching number feature of the modifier NP could interfere and thus cause agreement computation difficulty—namely, the phase at which the head NP and the modifier NP within the complex subject NP are integrated, and the phase at which the syntactic relation between the subject and the verb is established. In the present paper we refer to these two phases as the integration phase and the checking phase, respectively; note that other labels or conceptualizations are also used in the literature. As is explained below, several authors assume that working memory is involved in order to avoid incorrect subject–verb bindings at the integration phase, at the checking phase, or at both.

The first phase during which the modifier and its number feature could cause interference is the point at which the head NP (e.g., "the key") and the modifier NP (e.g., "the cabinets") are integrated within the complex subject NP (e.g., "the key to the cabinets"; Nicol et al., 1997; Pearlmutter et al., 1999). This integration phase involves the computation of the grammatical number of the complex subject NP as a whole, on the basis of the number features of its parts. Normally, and according to the grammatical rules of English, the complex subject NP should take over the number feature of the head NP. However, some authors argue that the grammatical number of the modifier NP, instead of that of the head NP, sometimes erroneously "percolates" to the complex subject NP (Franck, Vigliocco, & Nicol, 2002; Nicol, 1995; Pearlmutter, 2000), but evidence against this proposal has also been reported (Lago et al., 2015; Tanner et al., 2014). When the modifier NP (e.g., "the cabinets") carries a different grammatical number (plural) than the head NP (e.g., "the key", singular), such percolation may lead to the assignment of an erroneous grammatical number to the complex subject NP (plural). Such percolation thus leads to a number conflict that needs to be resolved (and possibly even to an error that would need to be repaired). It is very much conceivable that such conflict comes with a processing cost.

The second phase at which a number mismatch might cause interference is during the checking phase when processing the finite verb. According to cue-based parsing models (e.g., Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; McElree, 2001; McElree, Foraker, & Dyer, 2003; Van Dyke & Lewis, 2003), when encountering the verb, the parser will have to integrate the verb into the syntactic representation built up so far and therefore will have to bind the verb to its subject. A source of interference during the checking phase would then be the case where the parser may attempt to integrate the verb with the wrong NP in working memory. That is, to bind the verb to the appropriate NP, the parser might compare the required subject features as specified by the verb (e.g., nominative case, Häussler, 2009) to those of representations in working memory of previously encountered NPs in the sentence. When the representations of both the head NP and the modifier NP match these features, the parser might erroneously try to establish a subject–verb relation between the verb and the modifier NP. In support of such an account is the observation that agreement processing is more difficult when both the head and modifier are plausible subjects of the verb than when only the head is plausible (Häussler, 2009). Thus, one might expect a particularly strong processing cost when the modifier is a plausible subject of the verb (facilitating an attempt at integration with the verb) but differs in number from the verb (resulting in that attempt to fail).

Note that percolation and retrieval accounts need not be mutually exclusive. According to a hybrid account (e.g., Häussler, 2006, 2009), a further source of interference during checking would be the case where the percolation process that we described above has led to an incorrect number specification of the subject NP, then an attempt to bind the verb to subject NP might fail because of an illusionary subject–verb number mismatch, thereby creating a processing cost. Importantly, such accounts would assume processing costs during both integration and checking.

Summarizing, accounts of agreement computation predict a processing cost during an early phase of integrating the modifier with the full NP (integration phase), a later phase of processing the verb (checking phase), or both. Explicitly or implicitly, theories of agreement computation (e.g., Badecker & Kuminiak, 2007; Häussler, 2009; Lewis et al., 2006) have assumed that this processing cost reflects a stronger involvement of working memory, at either of these phases.

The theories about interference effects of the modifier NP during subject–verb agreement computation in *production* are to a large extent parallel with those on agreement computation in comprehension. Eberhard, Cutting, and Bock (2005), for instance, presented a formal model of agreement computation in production. This model also has an integration phase ("morphing") and a phase where subject number is mapped onto the verb. In both natural and laboratory language production, instances in which agreement computation is flawed, like (3), are well documented (Bock & Cutting, 1992; Bock & Eberhard, 1993; Bock & Miller, 1991). Most theories of production hypothesize that the conflicting grammatical number of the modifier NP interferes during the integration phase (Eberhard et al., 2005; Vigliocco, Hartsuiker, Jarema, & Kolk, 1996; Vigliocco & Nicol, 1998). An important piece of evidence for that hypothesis is that effects of conflicting number of modifier NPs are modulated by the distance between the subject NP and the modifier NP within the phrase structure (i.e., "syntactic distance"), but not by the linear distance between these NPs in the sentence or between the modifier NP and the verb (e.g., Vigliocco & Nicol, 1998). Those findings thus suggest that conflict exerts an effect when the grammatical structure is pieced together.

Additionally, Badecker and Kuminiak (2007) found evidence consistent with the hypothesis that in production a checking phase might also be operating. They found an influence of the grammatical case of the head and modifier NP on subject–verb agreement in production in Slovak, suggesting that the production of subject–verb agreement requires reactivation of the subject NP(s) in working memory when producing the verb (see Hartsuiker, Schriefers, Bock, & Kikstra, 2003, for similar findings in German).

The involvement of working memory in agreement production was tested directly by Fayol, Largy, and Lemaire (1994) and Hartsuiker and Barkhuysen (2006). Fayol et al. investigated the involvement of working memory in written sentence production. In three dictation experiments, participants wrote down an orally presented sentence, while they conducted a secondary task (counting clicks), or while they were not conducting any secondary task at all. The head and modifier NP in the dictated sentences always mismatched in number. The verbs were orthographically, but not phonologically, marked for grammatical number. Hence, participants could not determine the correct inflection of the verb on the basis of phonological cues, but had to determine it by morphosyntactic analysis—namely, computation of subject–verb agreement. Participants produced many inflection errors on the verb in the dual-task conditions, and almost no such errors in the single-task condition. These dual-task effects suggest that working memory is needed to prevent agreement errors in written sentence production.

A second study, by Hartsuiker and Barkhuysen (2006), found that an extra-sentential word load and the speaking span of the participants (Daneman & Green, 1986) affected the number of subject–verb agreement errors they made in a spoken sentence completion experiment. More specifically, low-span participants committed more errors under a word load than under no word load, whereas the performance of high-span participants was not affected by the presence or absence of a word load. These results suggest that working memory is needed to avoid agreement errors in spoken sentence production as well. Recently, Allen et al. (2015) administered glucose to diabetic and non-diabetic subjects in order to create a state of euglycaemia or hypoglycaemia, which is known to affect working memory. The participants had lower reading spans and produced fewer correct sentence completions under hypoglycaemia. These results thus confirmed Hartsuiker and Barkhuysen's findings while manipulating working memory capacity within subjects.

Which phase of agreement computation in language production would demand working memory? Hartsuiker and Barkhuysen (2006) assumed that working memory is required in the integration phase in agreement production. In this phase, the head and modifier NP will compete to assign their number feature to the complex subject NP. Number assignment could be a rather automatic process, in which the number features of both NPs migrate in the tree representation towards the higher node of the complex subject NP. The one that reaches that node first will win the race. However, Hartsuiker and Barkhuysen suggested that if working memory is available, the conflict is resolved in a controlled way in order to avoid erroneous number assignment. Only when working memory is not available is the outcome of the conflict solely determined by the number feature that wins the race. If the mismatching number feature of the modifier NP wins the race, this will result in an agreement error.

A similar controlled process might take place during the integration phase in comprehension. Additionally, it is conceivable that working memory is also needed at the checking phase in comprehension to resolve the number conflict between an erroneously number-specified subject NP and the verb, or the erroneously reactivated modifier NP and the verb in a controlled way.

Experimental study

Studies about number agreement in comprehension have shown that there is a processing cost when the grammatical number of the head NP and the modifier NP in the complex subject NP do not match. This cost may originate at the integration phase or at the checking phase, or even in both. There is relatively little evidence to unequivocally localize the processing cost at any of these phases. Using self-paced reading, Wagers et al. (2009) did not observe any effects at the checking phase. Dillon, Mishler, Sloggett, and Phillips (2013) collected eye-tracking data at both phases and found no effects at the integration phase and no number match effect at the checking phase, but did find a Grammaticality × Number Match interaction at the checking phase. Further eye-tracking (Acuna-Farina, Meseguer, & Carreiras, 2014; Lago et al., 2015) and event-related potential (ERP) studies (Tanner et al., 2014) only analysed the checking phase (typically the verb and word following the verb). These studies only found number match effects in ungrammatical sentences.

The present study pursues the efforts at better understanding the processes that underlie the costs related to grammatical number agreement in *comprehension*, by using working memory loads as in the production studies and taking advantage of eye tracking to localize the phase at which any costs occur.

The focus of the present study was on creating a working memory load that taxes domain-general processes and that can be applied locally at a particular phase of sentence processing. Such domain-general processes involve a series of control processes (labelled attentional control, cognitive control, or executive control) and are assumed to be involved in situations involving planning, problem solving, overcoming habitual actions, and so on. (Botvinick, Braver, Barch, Carter, & Cohen, 2001; Burgess, 1997; Logan & Gordon, 2001; Miyake & Shah, 1999; Norman & Shallice, 1986; Vandierendonck, 2016). More specifically, executive control is exerted to manage the conflict between multiple, incompatible representations in memory and to select the appropriate one, so that it seems plausible that executive control is involved in the resolution of number conflict in either or in both of the hypothesized processing phases in agreement computation in comprehension. Therefore, in the present study, we investigated whether the comprehension of subject–verb agreement during syntactic processing involves executive control and, if so, whether it does in only one or both of the hypothesized processing phases.

Previous research has shown that in contrast to simple reaction time tasks (e.g., respond with a keypress when a sound occurs), choice reaction time tasks (e.g., press one key when a high tone occurs, and press another key when the tone is low) tax executive control (e.g., Szmalec, Vandierendonck, & Kemps, 2005) and therefore can be used to interfere with sentence processing at a specific processing phase. Many studies have shown that when more or more difficult choice reaction tasks have to be executed during the retention interval of a serial short-term memory task, recall is more impaired than when fewer or less difficult choice tasks are performed (Barrouillet, Bernardin, & Camos, 2004; Barrouillet, Bemardin, Portrat, Vergauwe, & Camos, 2007; Barrouillet & Camos, 2010; Barrouillet, Corbin, Dagry, & Camos, 2015; Barrouillet, Lépine, & Camos, 2008; Barrouillet, Portrat, & Camos, 2011; Barrouillet, Portrat, Vergauwe, Diependaele, & Camos, 2011). As these effects are observed both when the choice tasks are in the same modality and when they are in a different modality from the memory tasks, it is clear that these effects are not due to domain-specific overlaps (Vergauwe, Barrouillet, & Camos, 2009, 2010). Furthermore, Szmalec et al. (2005) demonstrated that a verbal fluency task, which is known to impose high executive demands (e.g., Phillips, 1997; Rende, Ramsberger, & Miyake, 2002), is impeded more by a choice reaction time (CRT) task than by a simple reaction time task, in which participants have to react to only one possible stimulus by a simple button press. This result suggests that the response selection process that is present in the CRT task but not in the simple reaction time task taxes executive control (see also Barrouillet et al., 2007). Also in studies with mental arithmetic (Deschuyteneer & Vandierendonck, 2005a, 2005b) and with voluntary but not with automatic eye movements (Vandierendonck, Deschuyteneer, Depoorter, & Drieghe, 2008), performance was impaired by the presence of a secondary CRT task but not by the presence of a secondary simple reaction time task.

In addition to requiring executive control for their execution, CRT tasks also have the advantage that they tax control processes only for a very short time. As their RT is typically about 450–600 ms (depending on the difficulty of the required judgment), the interval during which executive control is occupied by a CRT task is probably in the order of 250–300 ms (Pashler & Johnston, 1998). In view of these advantages, we decided to use a tone decision task in the present study. On each trial in this task, participants hear one of two possible tones, either a high tone or a low tone, and they decide whether that tone is the high or the low one. Whereas in the earlier studies of Hartsuiker and Barkhuysen (2006) and of Fayol et al. (1994), as in most studies imposing a working memory load, the secondary task was present for the entire duration of the primary task, we imposed the secondary tone discrimination task at a specific region of interest in the sentence. This enabled us to identify which particular phases in sentence processing require executive control. This technique was first introduced in Loncke (2012).

In the present study we used Dutch versions of sentences such as (5) and imposed the secondary tone discrimination task at either of two regions in the sentence comprehension task: either when the participants were processing the modifier noun ("soldier") or when they were processing the verb (died). At these regions, respectively the integration phase and the checking phase are supposed to be performed. If the process of conflict resolution in the integration phase requires executive control, we expect to see interference between the sentence comprehension task and the tone discrimination task at the modifier noun. Similarly, if the process of conflict resolution in the checking phase requires executive control, interference between tone discrimination and sentence comprehension is expected at the verb.

5. John mailed that the lieutenant of the soldier unfortunately died(singular) because of reckless behaviour

To ensure that the tone was presented exactly in the region of interest, we used eye tracking to record eye movements while participants were reading the sentences: The tone was sounded when the participant started to read one of the critical regions. Eye tracking has the additional advantage that it permits measures of processing times at the various regions in the sentence (Rayner, 1998), so that we could measure processing time at each of the two regions of interest—namely, the modifier noun and the verb. By presenting sentences in which the grammatical number of the modifier NP either matched or mismatched that of the head NP, eye tracking allows a comparative measure of reading time at either of the two regions of interest. Additionally, in half of the sentences, the secondary task was presented at the modifier NP. Because the secondary tone discrimination task is assumed to tax executive control processes, the presence of this task is expected to compete with processing of the modifier NP, but only if these processes also require executive control, so that the difference in processing duration of mismatching and matching sentences would be increased. In the other half of the sentences, the CRT task was presented at the verb. If checking duration at the verb is slower for mismatching than for matching sentences due to a larger call on executive control, the presence of the tone-discrimination task is expected to boost this processing difference. In short, inclusion of the tone-discrimination task at either of the two regions allowed us to test whether interference due to the secondary task was larger in mismatching than in matching sentences.

Because processing costs in the checking phase might be driven by an attempt to bind the verb to the modifier, and because such binding is also affected by semantic features (e.g., Fedorenko, Gibson, & Rohde, 2006; Gordon, Hendrick, & Levine, 2002; Lewis & Vasishth, 2005; Lewis et al., 2006; Martin, 2006; McElree et al., 2003; Van Dyke & Lewis, 2003), we further varied the plausibility of the modifier NP as a possible semantic subject of the verb. This is illustrated in Sentence 6. While in Sentence 5, the modifier NP (the soldier) is plausible as a subject of the verb (died), in Sentence 6 the modifier NP (the frontline) is not a likely subject of the verb (died). Therefore, it may be expected that at the checking phase, the parser will more often erroneously try to bind the verb to a local NP that is a plausible subject of the finite verb (died), like "the soldier", than a modifier NP that is an implausible subject of the verb, like "the frontline". Consequently, we expected the effect of the mismatch in grammatical number between the head NP and the modifier NP to be larger for plausible modifier NPs than for implausible modifier NPs. Such an effect would mirror the effects of plausibility on the number of subject–verb agreement errors in written French (Hupet, Fayol, & Schelstraete, 1998) and spoken English (Thornton & MacDonald, 2003) sentence production.

6. John mailed that the lieutenant at the frontline unfortunately died(singular) because of reckless behaviour

In summary, our first goal was to identify which of the two sentence-processing phases, integration or checking, shows a general processing cost when the (singular) subject noun mismatches in number with a (plural) modifier NP (see also Pearlmutter et al., 1999). If the processing cost is situated at the integration phase, we expected longer processing times (expressed in longer looking times) at the modifier noun when head NP and modifier NP mismatched in number than when they matched. Similarly, if the processing cost is situated at the checking phase, we expected longer processing times at the verb when the head NP and the modifier NP mismatched than when they matched. Moreover, we expected the latter effect to be modulated by the semantic plausibility of the modifier NP as the subject of the verb (local plausibility).

The second and more important goal of the present study was to investigate whether resolution of number conflict at the integration and/or checking phase requires executive

Table 1. Sample sentence quartet based on the Match × Plausibility variation with English word-by-word translation within brackets.

Condition	Sample sentence
Match: plausible	*John mailde dat de luitenant van de soldaat jammer genoeg sneuvelde door roekeloos gedrag* [John mailed that the lieutenant of the soldier unfortunately died(singular) because of reckless behaviour]
Match: implausible	*John mailde dat de luitenant aan de frontlinie jammer genoeg sneuvelde door roekeloos gedrag* [John mailed that the lieutenant at the frontline unfortunately died(singular) because of reckless behaviour]
Mismatch: plausible	*John mailde dat de luitenant van de soldaten jammer genoeg sneuvelde door roekeloos gedrag* [John mailed that the lieutenant of the soldiers unfortunately died(singular) because of reckless behaviour]
Mismatch: implausible	*John mailde dat de luitenant aan de frontlinies jammer genoeg sneuvelde door roekeloos gedrag* [John mailed that the lieutenant at the frontlines unfortunately died(singular) because of reckless behaviour]

control. To that end, we imposed the secondary tone discrimination task either at the integration or at the checking phase in the sentence comprehension task. When executive control is switched away from sentence comprehension towards the tone discrimination task, executive control is no longer available for sentence processing. Therefore, imposing the secondary task should be detrimental for the processes of integration and checking, at least if they require executive control. If they do require control, we expect to see an interaction between the presence of the tone discrimination task and the (mis)match of head and modifier NP on the reading times in the modifier NP region, such that with concurrent execution of the tone discrimination task, the mismatch effect is augmented. In the same vein, we expect an interaction between the presence of the tone discrimination task and plausibility of the modifier as a subject at the verb region, such that the plausibility effect is augmented under the presence of the tone task. If any effect of number matching would be evident at the checking phase, either on its own or in interaction with plausibility, it is expected that these effects will interact with the presence of the tone-discrimination task.

In sum, the dual-task experiment reported in the present study allows us to investigate whether processing of subject–verb agreement in comprehension taxes executive control and to pinpoint the exact processing phase(s) at which this control is needed. If it turns out that the choice reaction time task interferes with agreement computation in syntactic analysis, this will add to the evidence supporting the view that sentence comprehension, at least at phases that require conflict resolution, relies on domain-general memory systems (Fedorenko et al., 2006; Lewis et al., 2006; Loncke, 2012; Loncke, Desmet, Vandierendonck, & Hartsuiker, 2011; Swets, Desmet, Hambrick, & Ferreira, 2007).

Method

Participants. Forty-two students (32 female, 10 male) from Ghent University participated in exchange for course credit or payment. The mean age was 19.9 years ($SD=2.6$). All participants were native speakers of Dutch, had normal or corrected-to-normal vision, were reported to have normal reading skills, and were naive to the purpose of the study.

Design. Sentence quartets like the example in Table 1 were constructed by varying the match in grammatical number of the modifier NP (singular or plural) with the head NP (singular; match or mismatch), and the local plausibility of the modifier NP as subject of the verb [e.g., "the soldier(s)", plausible, or "the frontline(s)", implausible]. Each of these sentence variations was presented with a concurrent tone discrimination task: A high or a low tone was presented in either of two regions of interest—namely, the modifier noun (the soldier, in the example) or the verb (died). These three variations were crossed to obtain a 2 (match) × 2 (local plausibility) × 2 (tone region) repeated measures design. These factors were manipulated within participants and within sentences in such a way that each participant saw only one variant of each sentence (for more details, see below).

Materials. We developed 120 critical sentence quartets[1] and 160 filler sentences. All sentences were grammatical, Dutch sentences. The semantic plausibility of the modifier NPs as the subject of the verb within the sentence quartets was assessed in a separate plausibility rating study based on a further 39 native Dutch speakers (age $M=20.1$ years, $SD=1.5$). Participants rated the plausibility of the modifier NP as the subject of the verb on a 7-point Likert scale (cf. Thornton & MacDonald, 2003). Each participant rated only one of both modifier NPs within a sentence quartet. The NPs were presented in their singular form in order to match the number of the verb and thus to form a grammatical sentence. As expected, the NPs that we deemed as plausible were rated as more plausible ($M=6.15, SD=0.79$) than the ones we deemed as implausible ($M=2.35$, $SD=1.32$), $t(119)=26.34$, $p<.001$.

All critical sentences had the same syntactic structure—namely, a matrix clause comprising a subject (proper noun), a main verb, a relative pronoun, and a subordinate clause. The latter consisted of the head NP (article + noun), a preposition, the modifier NP (article + noun), an adverb, a finite verb, and an adverbial phrase or clause. Embedding the complex subject NP in a subordinate clause enabled us to insert an adverb (e.g., "jammer genoeg [unfortunately]" in Table 1) between the modifier NP and the verb, in order to create some distance between these two regions of interest (cf. Wagers et al., 2009). This way, spillover effects (Rayner, 1998) from one region of interest (i.e., the modifier noun) into the other region of interest (i.e., the verb) were made less likely. In view of Wagers et al.'s (2009) findings that no grammatical number effects occurred beyond the word after the critical region, inclusion of the adverb should suffice to block spillover. The finite verb was followed by an adverbial phrase or clause to avoid sentence wrap-up effects (Just & Carpenter, 1980; Rayner, Kambe, & Duffy, 2000) at the verb.

Some further issues were taken care of during the construction of the critical sentences. First, to avoid semantic interference in working memory between the subject of the main clause and the head and modifier NP of the subordinate clause, the subject of the main clause was always a proper name (e.g., "John" in Table 1). For the same reason, it was important that no object NP interfered between the complex subject NP and the verb of the subordinate clause. As relative clauses in Dutch have subject–object–verb (SOV) order, we selected intransitive verbs (e.g., "sneuvelde [died]" in Table 1). Second, the head NP in the subordinate clause (e.g., "de luitenant [the lieutenant]" in Table 1) was always singular, as agreement computation difficulties have mostly been observed with singular heads (and plural modifiers) and much less so with plural heads (and singular modifiers; e.g., Nicol et al., 1997; Pearlmutter et al., 1999; Wagers et al., 2009). Third, the gender of the head and modifier NP (e.g., "de luitenant [the lieutenant]" and "de soldaat [the soldier]" in Table 1) was always non-neuter; in Dutch, the article of non-neuter nouns ("de" in both singular and plural) does not give a clue for grammatical number as opposed to the article of neuter nouns (Antón-Méndez & Hartsuiker, 2010; Hartsuiker et al., 2003). Fourth, the verb of the subordinate clause was always in the simple past tense (e.g., "sneuvelde [died]" in Table 1), so that both the semantics and grammatical number of the verb were carried by a single word, allowing us to test for combined effects of these variables during the checking phase. Fifth, the verb of the subordinate clause was always at least two syllables long to minimize the chance that it was skipped during reading (Rayner, 1998).

Each participant only saw one sentence of a quartet with a tone in only one of the two tone regions. Thus, each participant saw 15 critical sentences within each of the eight conditions. We also developed 160 filler sentences, which were randomly mixed between the critical sentences (see Apparatus and Procedure) and were not analysed. The length (in number of characters, spaces included) of the filler sentences ($M = 84$) and critical sentences ($M = 87$) was comparable, $t(278) = 1.67$, $p = .096$. Seventy-five percent of the filler sentences (120 sentences) had the same structure as the critical sentences, but had a plural head NP and a singular modifier NP (60 sentences) or a plural head NP and a plural modifier NP (60 sentences). The remaining 40 filler sentences had a subject NP that consisted of only one NP, half of them singular, half of them plural. These simple NPs always contained one or two adjectives to make them comparable in length to the other filler and critical sentences. As in the critical sentences, the subject of the main clause of all filler sentences was always a proper name.[2]

In order to make tone presentation unpredictable during the experiment, a tone was presented at a pseudo-random moment in half of the fillers. Also, half of the fillers were followed by a comprehension question. Questions could concern any of the elements or relations between elements in the sentence (e.g., head NP–verb, modifier NP–verb, subject of the main clause–verb, head NP–PP, etc.). Thus, the questions did not focus participants' attention on the subject–verb relation in the sentences. Half of the comprehension questions required a yes-answer and half of them a no-answer. The presentation of a tone (tone/no tone) and the presentation of a comprehension question (question/no question) were crossed in the filler sentences, so that the presentation of a tone did not predict whether a comprehension question would be presented. The aim of having an unpredictable presentation of comprehension questions was to stimulate participants to read all sentences in the experiment to the end.

Apparatus and procedure. Eye movements were recorded using an SR Research Eyelink 1000 eye-tracking device (Mississauga, Ontario, Canada). Viewing was binocular, but eye movements were recorded from the right eye only. Fixation locations were sampled every millisecond. Each sentence as well as each comprehension question was presented on a single line, aligned left on the horizontal axis and in the centre on the vertical axis. Sentences and questions were presented in black on a white background in Courier New 10 font.

Before and after the main part of the experiment, a single-task tone discrimination task was performed (Szmalec et al., 2005). On each trial, one of two different tones with a frequency of either 262 (C1 note) or 524 Hz (C2 note) was presented. Each tone lasted 150 ms; the inter-stimulus interval was fixed at 650 ms. Participants hit either the right or the left button of an SR Research gamepad as fast as possible to indicate whether the pitch of the tone was low or high. They were instructed to keep their fingers on the buttons to avoid target-seeking movements between both keys. Moreover, they were asked to fixate a continuously presented fixation cross in the centre of the screen along the task and not to look down at their hands or to the gamepad. This way, they were trained to keep looking at the screen while performing the tone discrimination task. In both single-task tests, participants performed four practice trials and 60 critical trials.

The main part of the experiment was a dual-task test in which participants read a sentence for comprehension and judged the pitch of a tone when it was presented. At the beginning of each trial, participants fixated a point at the left of the screen to perform drift correction. Then, a sentence appeared. Participants were instructed to read the sentence carefully and with normal speed. In 71% of the trials (i.e., in all critical sentences and in half of the filler sentences), a high or a low tone (the same ones as in the single-task tests) was presented as soon as the participant 's eye position crossed the left boundary of the pre-specified region in the sentence for the first time[3] (see Materials). Participants were instructed to execute the tone discrimination task in exactly the same way as in the single-task tests. Participants were also instructed that after having read the whole sentence, they could self-pace to the next screen by pressing another pre-specified button on the gamepad (different from the tone buttons). Then, in all critical trials and half of the filler trials, the trial ended. In the other half of the filler trials, a comprehension question appeared on the screen, which required a yes- or no-response by pressing one of two pre-specified keys on the keyboard. Feedback following these answers was displayed for 500 ms. Overall accuracy on these questions was 86% ($SD=34$), indicating that participants read the sentences attentively. The 120 critical and 160 filler trials were presented in random order and were preceded by four practice trials. Calibration consisted of a standard 9-point grid. The whole session including camera set-up and calibration lasted about an hour.

Data analysis. One of the aims of the data analysis was to test whether and how sentence comprehension and secondary tone discrimination affected each other. Therefore, we analysed performance on both tasks. With respect to the tone discrimination task, the measures of interest were reaction time and discrimination accuracy in single- and in dual-task conditions. For the sentence comprehension task, the measures of interest were duration and frequency of fixations in the sentence regions of interest. In order to be able to localize secondary task effects at an early or a later stage of processing, we distinguished three different fixation, measures in both critical regions (the modifier noun and the verb). These eye-movement measures are illustrated in Figure 1 and refer to early and later gaze measures that covered the consecutive and non-overlapping time frames in the reading process of the region of interest.[4] The different measures were: (a) first fixations, (b) additional first-pass fixations, and (c) rereading fixations. First fixations concern the first fixation on the target

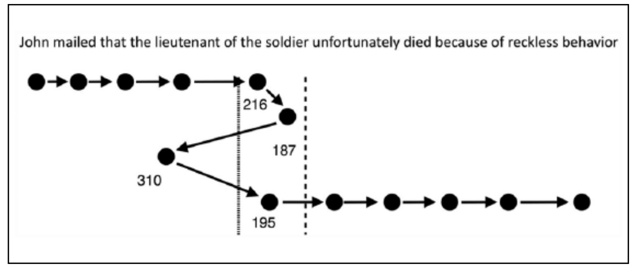

Figure 1. Hypothetical eye-movement record showing the different gaze measures.

region during the first passage over that region (Liversedge, Paterson, & Pickering, 1998). In Figure 1, the duration of the first fixation on the modifier noun "soldier" is 216 ms. Additional first-pass fixations were defined as the sum of all other fixations made in a region during the first passage over that region (i.e., until the point of fixation leaves the region either to the left or to the right) minus the first fixation. In Figure 1, the duration of the additional first-pass fixations on the modifier noun is 187 ms. Rereading fixations were defined as the sum of all fixations from the moment the eye leaves the target region to the left up to but excluding the first fixation to the right of the target region. In Figure 1, the duration of the rereading fixations on the modifier noun is 505 ms (comprising the fixations of 310 and 195 ms and thus excluding the duration of first-pass and additional first-pass fixations). These measures were analysed with respect to (a) their frequency, and (b) their duration.

Prior to the analysis, we excluded the data of a sentence quartet (no. 9 in the stimulus list, see Footnote 1), which, due to a clerical error, was not equally partitioned over the lists (loss of 1 data point per subject: 42 data points or 0.8%). On some trials, the critical region was skipped; this occurred on 202 trials (4.0%) of the modifier NP region, and on 148 trials (2.9%) of the verb region. As these trials were not included in the data analysis, 4796 data points were left for the analyses of the modifier NP region, and 4850 data points remained available for analyses of the verb region.

Data analysis was based on linear mixed effects (LME) modelling. The analyses were performed with the help of the R statistical package (R Core Team, 2015), using the lme4 package (Bates, Maechler, Bolker, & Walker, 2015) for R. Analyses of the continuous dependent variables (i.e., gaze durations and reaction times to the tones) were performed by fitting LME models using restricted maximum likelihood estimates (Pinheiro & Bates, 2000).[5] Following suggestions made by Baayen, Davidson, and Bates (2008) and Barr, Levy, Scheepers, and Tily (2013), first a model including all fixed effects and maximal random effects was fitted. As already explained, the fixed effects were the main effects and interactions following from the 2 (match) × 2 (plausibility) × 2 (tone region) factorial design. A Latin square design was used to counterbalance these eight conditions representing the fixed effects over the items and the subjects. In fact, these conditions were implemented as variants of the items, resulting in $8 \times 120 = 960$ different item variants. These 960 variants were partitioned into eight lists, which were randomly assigned to subjects, such that each list was presented to five or six subjects. The factors list (subsets of item variations) and group (the subjects presented with a particular list) are involved here merely for purposes of counterbalancing the materials over the subjects and are not considered to be part of the design. It follows then that the random factor item was orthogonally crossed with the conditions representing the fixed effects, and because of the usage of a Latin square it has to be assumed that the fixed effects do not interact with the factor item. For that reason, it suffices to include a random intercept for the factor item. It cannot be assumed, however, that the fixed effects would not interact with the factor subject. For that reason, in addition to a random intercept for the factor subject, also random slopes of the three fixed effects should be added. In order to obtain a "maximal" random structure, it could also be considered to include random slopes for all the interactions of the fixed effects. However, the lme4 package warns against the usage of such a complex random structures. Moreover, such complex structures often lead to a non-convergent model fit. Hence, in the notation used in the lme4 package, model fitting always started with the following model:

$$Y \sim \text{match} \times \text{plausibility} \times \text{tone} + (1 \mid \text{item}) \\ + (1 + \text{match} + \text{plausibility} + \text{tone} \mid \text{subject})$$

where Y is the dependent variable, and the fit is performed by using polynomial contrasts (as the fixed factors were considered to be ordered).

In addition to these random factors, which were imposed by the design, some ad hoc decisions had to be made concerning the best model structure. First, in the analyses of the data in the modifier NP region, variations in the length of the modifier NP may to some extent account for the subject's responses in processing this sentence part. For that reason it might be considered to add a random slope for this variable. Because these variations tended to correlate with the difference between matching (singular) and mismatching (plural) modifiers, it was decided to separately add the random effect (1|length:subject).

Furthermore, although care was taken in the development of the sentences to avoid spillover of processing from one critical region to the next region by including a buffer region immediately after the critical region, occasionally the response to the tone-discrimination task flowed over into the next processing phase. When the tone was presented in the modifier NP region, the response to the tone flowed over into the buffer region on 1127 trials (46.9% of the tone trials). The response to the tone did not flow over beyond this spillover region (see Results section). When the tone was presented in the verb region, overflow could occur from the verb region into the following spillover buffer region; 818 trials were affected (33.7% of the tone trials); in 722 of these trials the response occurred after the spillover buffer region, resulting in a response at the end of the sentence. As the presence of tone-response overflow might contaminate the data analysis, actions to avoid unwarranted conclusions were necessary. One possible option is to exclude all trials with a tone reaction time (RT) that exceeded the total gaze duration in the critical region. However, because there are so many of these trials, it was decided to keep these trials in the analysis but include the ad hoc factor overflow as an extra random effect. The values on this factor were either "present" (when there was a tone response that spilled over beyond the critical region) or "absent" (when there was no such event).

Obviously, overflow may be correlated with the fixed effects of match and plausibility,[6] and, therefore, the random intercept of overflow, as well as the slopes for match, plausibility, and their interaction can be added to the random effects structure. However, in order to avoid overparameterization of the model (Baayen et al., 2008), only effects that improved the model fit were included in the final model. If at any stage the model fit did not converge, the components with the smallest variance (see Baayen et al., 2008) were stepwise removed. After fitting the final model, the relative contribution of each of these random effects was estimated by comparing a model with one random factor excluded to the complete random model by means of a likelihood ratio test. The relative contribution of each of the fixed effects was estimated on the basis of the obtained model by using the anova() function of the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015), which yields an F-test with Satterthwaite approximation of the number of degrees of freedom.

For the frequency data, following suggestions of Jaeger (2008), generalized linear mixed modelling (GLMM) for binomially distributed data (logit) was applied by means of the glmer() function of the lme4 package. Again following Barr et al. (2013), the same strategy was used by first fitting the maximal random effects structure and then testing the effects by means of likelihood ratios. Probability measures of the fixed effects under the model were obtained by calculating Wald's z (Wald, 1943).

Results

First, we report statistical analyses of the eye movement data, separately for each of the two target regions (modifier NP and verb). Both duration and frequency of different types of fixations are reported. Because on some trials the tone response flowed over into the spillover buffer region, an analysis of the fixation durations in these buffer regions

Table 2. Means and standard errors of fixations durations in the modifier NP region of the critical sentences as a function of grammatical number match, semantic plausibility, and the presence of a tone in the region.

Tone condition	*N*	Match		Mismatch	
		Plausible	Implausible	Plausible	Implausible
First fixations	4796				
Tone		394 (14)	411 (14)	409 (14)	416 (14)
No tone		238 (4)	251 (5)	258 (4)	253 (4)
Tone spillover		316 (10)	304 (9)	322 (9)	328 (11)
Additional first-pass fixations	2046				
Tone		424 (26)	475 (27)	502 (32)	513 (26)
No tone		319 (15)	301 (12)	284 (12)	297 (13)
Tone spillover		357 (19)	369 (20)	349 (20)	339 (21)
Rereading fixations	1318				
Tone		930 (43)	902 (39)	966 (46)	963 (51)
No tone		500 (45)	549 (38)	499 (41)	536 (31)
Tone spillover		494 (63)	422 (28)	482 (54)	439 (35)
Spillover region	4161				
Tone		247 (6)	262 (6)	264 (6)	263 (7)
No tone		258 (4)	259 (5)	270 (5)	264 (6)
Tone spillover		287 (8)	287 (9)	305 (10)	309 (10)

Note: Means in ms. Standard errors in parentheses. NP = noun phrase; match = grammatical number match; plausibility = semantic plausibility. The table displays first fixations, additional first-pass fixations, rereading times, and fixations in the spillover region immediately following the modifier NP. The averages and standard errors displayed in this table were calculated by means of the method proposed by Bakeman and McArthur (1996) to estimate the within-subject variability corrected for between-subject variations.

will also be reported. Finally, performance on the tone discrimination task is also presented. Our presentation of the results will focus on the fixed effects that turned out to contribute to the maximal statistical model, but the random effect statistics are also reported.

Eye movements in modifier NP region

Gaze durations. Table 2 displays the averages and standard errors of the different types of gaze duration in the modifier NP region as a function of Match × Plausibility × Tone. Table 3 shows an overview of the effects that account for the gaze durations according to the LME modelling.

First-fixation duration. First fixations lasted for 308 ms on average and were affected by match and tone (see Tables 2 and 3). First-fixation durations were slower in the mismatch condition (M=314 ms) than in the match condition (M=301 ms). Presence of a tone in the region resulted in slower (M=365 ms) reading than when no tone was present (M=249 ms). Match and tone did not interact.

Additional first-pass fixations. The LME modelling based on additional first-pass fixations showed that when these fixations occurred (on 2046 trials), they lasted on average 366 ms. The fixed effect of tone and its interaction with match were both reliable, but the effect of match itself was not significant (see Tables 2 and 3). Sentences received longer additional first-pass fixations when a tone was present (M=431 ms) than when no tone was present (M=299 ms). Tone and match interacted, such that the effect of the presence of a tone was much larger in the mismatch (156 ms) than in the match (105 ms) condition.

Rereading fixations. Rereading occurred on 1318 trials. Rereading times added up to 727 ms on average. Only the effect of tone was significant (see Tables 2 and 3). Trials with a tone task were slower (M=834 ms) than without (M=523 ms).

Fixation duration in the spillover buffer region. This buffer region was fixated on 4161 trials. Average duration of these fixations was 271 ms. Only the fixed effect of match attained significance, with shorter fixations in the match (266 ms) than in the mismatch (275 ms) condition. Although sentences with a tone tended to have longer fixations, this difference did not attain significance.

Summary. Obviously, all durations in the modifier NP region, but not the fixations in the spillover buffer, were lengthened by the presence of a tone. Grammatical number match affected first-pass fixations and interacted with tone in the additional first-pass fixation and also affected the fixations in the spillover buffer, but these fixations did not vary with the presence of the tone task. Plausibility and its interactions with the other fixed effects failed to attain significance in the modifier NP region.

Table 3. Overview of the LME model of the gaze durations in the modifier NP region and its spillover buffer, specifying all the fixed effects on the basis of *F* tests with Satterthwaite approximations for the degrees of freedom and all the random effects on the basis of the likelihood ratio statistic.

Effect	Dependent variables			
	FF ($N = 4796$)	AF ($N = 2046$)	RR ($N = 1318$)	BF ($N = 4161$)
Match (M)	$F(1, 41) = 7.30$, $p < .01$	NA	$F(1, 59) = 0.38$, $p = .54$	$F(1, 41) = 5.52$, $p < .05$
Plausibility (P)	$F(1, 40) = 0.95$, $p = .34$	$F(1, 353) = 0.44$, $p = .50$	$F(1, 252) = 0.41$, $p = .53$	$F(1, 3949) = 0.16$, $p = .68$
Tone (T)	$F(1, 41) = 67.0$, $p < .001$	$F(1, 60) = 129.3$, $p < .001$	$F(1, 45) = 100.9$, $p < .001$	$F(1, 63) = 0.41$, $p = .52$
M × P	$F(1, 4534) = 0.41$, $p = .52$	$F(1, 1950) = 0.00$, $p = .97$	$F(1, 1245) = 0.05$, $p = .82$	$F(1, 3948) = 0.15$, $p = .70$
M × T	$F(1, 4536) = 0.53$, $p = .47$	$F(1, 981) = 12.56$, $p < .001$	$F(1, 1191) = 0.61$, $p = .43$	$F(1, 3945) = 0.10$, $p = .75$
P × T	$F(1, 4525) = 0.04$, $p = .84$	$F(1, 1950) = 0.49$, $p = .49$	$F(1, 1238) = 1.99$, $p = .15$	$F(1, 3938) = 0.15$, $p = .70$
M × P × T	$F(1, 4536) = 1.78$, $p = .18$	$F(1, 1961) = 2.47$, $p = .12$	$F(1, 1243) = 0.32$, $p = .57$	$F(1, 3942) = 0.36$, $p = .55$
Subject (S)	1 + M + P + T|S (10) 672, $p < .001$	1 + M + P + T|S (10) 179, $p < .001$	1 + M + P + T|S (10) 188, $p < .001$	1 + M + T|S (6) 317.2, $p < .001$
Item (I)	1|I (1) 10.8, $p < .001$	1|I (1) 19.4, $p < .001$	1|I (1) 3.20, $p = .07$	1|I (1) 34.0, $p < .001$
Spillover (O)	NA	1 + M|O (1) 28.8, $p < .001$	1|O (1) 71.6, $p < .001$	1|O (1) 54.6, $p < .001$

Note: χ^2 = likelihood ratio. LME = linear mixed effects; NP = noun phrase; FF, AF, RR, and BF refer to, respectively, first fixations, additional first-pass fixations, rereading fixations, and spillover buffer fixations; NA means not applicable. The fitted model represents a maximal random effects structure, as explained in the text, and all the fixed effects. The fixed effects are expressed as *F*-values with approximate degrees of freedom for the denominator (here rounded to the nearest integer); these are shown in the upper part of the table. Cells in this part containing "NA" indicate that it was not possible to calculate approximate degrees of freedom. The random effects are based on the likelihood ratio statistic, which follows a χ^2 distribution, and are shown in the lower part of the table. In this part, each cell contains a specification of the included intercept and slope(s). The notation 1 + M + P+ T|S indicates that the effect included a random intercept (1) and random slopes for match, plausibility and tone, whereas 1|S has only a random intercept. The number of degrees of freedom is shown between parentheses after the specification of the included slopes. The remainder of each cell displays the χ^2 value and its probability.

Table 4. Overview of the GLME model of the gaze frequencies in the modifier NP and the verb region, specifying all the random effects on the basis of the likelihood ratio statistic and all the fixed effects by means of Wald's z.

Effect	Modifier NP region		Verb region	
	AF	RR	AF	RR
Match (M)	$z = 2.78$, $p < .01$	$z = -0.09$, $p = .93$	$z = -0.14$, $p = .89$	$z = 2.16$, $p < .05$
Plausibility (P)	$z = 0.60$, $p = .55$	$z = 1.30$, $p = .19$	$z = 1.11$, $p = .27$	$z = -0.21$, $p = .84$
Tone (T)	$z = -0.02$, $p = .99$	$z = 4.36$, $p < .001$	$z = 4.07$, $p < .001$	$z = 4.80$, $p < .001$
M × P	$z = -0.41$, $p = .68$	$z = 0.13$, $p = .89$	$z = -.0.05$, $p = .96$	$z = -1.27$, $p = .20$
M × T	$z = -0.69$, $p = .49$	$z = 1.23$, $p = .22$	$z = -0.43$, $p = .67$	$z = -1.22$, $p = .22$
P × T	$z = 1.25$, $p = .21$	$z = -0.54$, $p < .59$	$z = -1.43$, $p = .15$	$z = -0.43$, $p = .67$
M × P × T	$z = -0.31$, $p = .76$	$z = -1.06$, $p = .29$	$z = 0.76$, $p = .45$	$z = 1.05$, $p = .30$
Subject (S)	1 + T|S + 1|L:S (4) 379, $p < .001$	1 + T|S + 1|P:S (4) 715, $p < .001$	1 + T|S (3) 308, $p < .001$	1|S + 1|T:S (2) 318, $p < .001$
Item (I)	1|I (1) 105, $p < .001$	1|I (1) 1.65, $p = .20$	1|I (1) 70.4, $p < .001$	1|I (1) 27.5, $p < .001$
Overflow (O)	NA	NA	NA	NA

Note: χ^2 = likelihood ratio. GLME = generalized linear mixed effects; NP = noun phrase; AF and RR refer to, respectively, additional first-pass fixations and rereading fixations; NA refers to "not applicable"; in all four analyses reported in this table, inclusion of the random effect of overflow resulted in a non-convergent model fit. The fitted model represents a maximal random effects structure, as explained in the text, and all the fixed effects. The fixed effects are shown in the upper part of the table as Wald's z (Wald, 1943) with an associated probability level. The random effects are shown in the lower part of the table based on the likelihood ratio statistic, which follows a χ^2 distribution. For the random effects, each cell contains a specification of the included intercept and slope(s). The notation 1 + T|S indicates that the effect included a random intercept and a random slope for tone, whereas 1|S has only a random intercept. The number of degrees of freedom is shown between parentheses after the specification of the included slopes. The remainder of each cell displays the χ^2 value and its probability.

Gaze frequencies in the modifier NP region

In 95% of the critical trials a first fixation occurred in the modifier NP region. Additional first-pass fixations occurred in 43% of the trials, and a rereading was present on 27% of the trials. The frequency of additional first-pass and rereading fixations were subjected to a GLMM analysis assuming a binomial distribution. As the information presented in the modifier NP region is not available at the time eye movements towards this region are performed, it is not meaningful to analyse the frequency of the first fixations. Table 4 displays Wald's z for the fixed effects and the likelihood ratios of the random effects of the model including a maximal random effects structure. Additional first-pass fixations occurred significantly more frequently in the mismatch (M=44.9%) than in the match condition (M=40.4%). This effect did not interact with any of the other factors.

In the modelling of the rereading frequencies, a main effect of tone was observed. When no tone was present, rereading was less frequent (M=18.9%) than when a tone was present (M=36.0%).

Eye movements in verb region

Gaze durations. Table 5 displays the means and standard deviations of the different types of gaze durations in the verb region. Table 6 displays details about the effects that contributed to the LME model of fixation durations.

First fixations. First fixations lasted on average for 312 ms; this was based on 4850 data points. The LME model showed a reliable effect of tone (see Table 6 for details), with longer fixations when a tone was present (M=366 ms) than without a tone (M=259 ms).

Additional first-pass fixations. Duration of additional first-pass fixations (1455 trials) was on average 319 ms. The LME modelling showed that these durations were only affected by tone (see Table 6). A tone prolonged fixation duration (M=383 ms vs. M=227 ms).

Rereading fixations. For the verb region, rereading fixations were present on 877 trials. Rereading lasted on average 723 ms. The LME model showed (see Table 6) that rereading was slower when a tone was present (M=786 vs. 579 ms).

Fixation duration in the spillover buffer region. This buffer region was fixated on 1789 trials. Fixation duration was on average 250 ms. The presence of a tone lengthened these fixations from 232 to 267 ms.

Summary. Similar to the modifier NP region, when a tone was present, reading time was lengthened for all types of fixations in the verb region. Presence of a tone also affected fixation times in the spillover buffer region. The presence of a tone did not interact with matching or plausibility, nor did any of these factors reliably affect fixation durations.

Gaze frequencies in verb region. First fixations were present on 96%, additional first-pass fixations on 29%, and rereading fixations on 17% of the trials. The GLMM of the additional first-pass frequency and the rereading frequency are shown in Table 4. In both models only the effect of tone was reliable. Both types of eye movements were more frequent when a tone was present (M=35.0% vs. 25.0% in additional first pass; M=25.5% vs. 11.0% in rereading).

Tone discrimination. As expected, discrimination RTs were faster in the single-task (M=414 ms, SD=71) than in the dual conditions (M=925 ms, SD=177). Because items do not matter in the single-task condition, an analysis of variance (ANOVA) was performed. The difference between these means was significant, $F(1, 41)=586$, $p<.001$, $\eta_p^2=.94$. In terms of accuracy, no differences were observed (means, respectively, .974 and .975 for the single- and the dual-task condition). For the analysis of RT as a function of Match × Plausibility × Tone Region, LME modelling was applied starting with the maximal random effects structure; only the random effect of subject with slopes for match, plausibility, and tone could be retained, $\chi^2(10)=242$, $p<.001$. None of the fixed effects were significant in this model (smallest $p=.29$). In the same analysis with accuracy of the tone task as dependent variable, only the random intercept for subject could be retained, $\chi^2(1)=55.5$, $p<.001$. None of the fixed effects attained significance (smallest $p=.25$).

General discussion

In a Dutch sentence comprehension experiment with eye tracking, we examined two main research questions with respect to the processing of sentences with a syntactic number mismatch between head and modifier NPs. First, we tested whether sentences with number mismatch show an integration cost (a processing cost at the modifier noun) and/or a checking cost (a processing cost at the verb) as compared to sentences with matching number. Second, we examined whether the cognitive processes at these phases (integration phase and checking phase) require executive control. To that end, we investigated the effect of imposing a secondary choice reaction time task, known to tax executive control, at either of these phases.

To determine at which phase the comprehension cost is localized, we tested how the match in grammatical number of the head and modifier NP affected the eye movements at the modifier noun and at the verb. At both regions, all gaze durations were lengthened by the presence of a tone. At the modifier noun, the first-fixation duration and the frequency of additional first-pass fixations were sensitive to number

Table 5. Means and standard errors of the mean of first fixation, additional first-pass fixations, rereading times, and fixation durations in the spillover buffer in the verb region of the critical sentences as a function of grammatical number match, semantic plausibility, and the presence of a tone in the region.

Tone	*N*	Match		Mismatch	
		Plausible	Implausible	Plausible	Implausible
First fixations	4850				
Tone		371 (11)	396 (12)	393 (13)	380 (11)
No tone		258 (5)	253 (4)	259 (5)	264 (5)
Tone spillover		333 (13)	328 (12)	326 (13))	326 (12)
Additional first-pass fixations	1455				
Tone		405 (22)	428 (21)	390 (23)	395 (18)
No tone		242 (16)	229 (12)	228 (13)	212 (11)
Tone spillover		296 (24)	337 (25)	322 (38)	348 (34)
Rereading fixations	877				
Tone		806 (82)	853 (67)	845 (47)	857 (59)
No tone		483 (48)	610 (74)	583 (60)	636 (78)
Tone spillover		627 (74)	468 (57)	497 (57)	340 (42)
Spillover region	1789				
Tone		254 (13)	247 (10)	246 (9)	260 (11)
No tone		240 (7)	234 (9)	221 (7)	233 (7)
Tone spillover		299 (22)	305 (17)	311 (23)	279 (17)

Note: Means in ms. Standard errors in parentheses. Match = grammatical number match; plausibility = semantic plausibility. The averages and standard errors of the mean displayed in this table were calculated by means of the method proposed by Bakeman and McArthur (1996) to estimate the within-subject variability corrected for between-subject variations.

Table 6. Overview of the LME model of the gaze durations in the verb region and its spillover buffer, specifying all the fixed effects on the basis of *F* tests with Satterthwaite approximations for the degrees of freedom and all the random effects on the basis of the likelihood ratio statistic.

Effects	Dependent variables			
	FF ($N=4850$)	AF ($N=1455$)	RR ($N=877$)	BF ($N=1789$)
Match (M)	$F(1, 88)=0.85$, $p=.36$	$F(1, 692)=1.86$, $p=.17$	$F(1, 199)=0.84$, $p=.36$	$F(1, 140)=0.90$, $p=.34$
Plausibility (P)	$F(1, 53)=0.11$, $p=.74$	$F(1, 916)=0.05$, $p=.82$	$F(1, 208)=0.24$, $p=.62$	$F(1, 345)=0.14$, $p=.71$
Tone (T)	$F(1, 41)=49.8$, $p<.001$	$F(1, 41)=98.7$, $p<.001$	$F(1, 48)=11.06$, $p<.01$	$F(1, 51)=6.84$, $p<.05$
M × P	$F(1, 4616)=0.24$, $p=.62$	$F(1, 1398)=0.61$, $p=.43$	$F(1, 786)=0.06$, $p=.81$	$F(1, 1695)=0.85$, $p=.36$
M × T	$F(1, 4619)=0.13$, $p=.72$	$F(1, 1394)=0.01$, $p=.92$	$F(1, 826)=0.60$, $p=.44$	$F(1, 1720)=0.60$, $p=.44$
P × T	F(1, 4617) = 0.06, $p=.80$	$F(1, 1393)=2.14$, $p=.14$	$F(1, 810)=1.01$, $p=.31$	$F(1, 1709)=0.02$, $p=.88$
M × P × T	$F(1, 4617)=2.88$, $p=.09$	F(1, 1393) = 0.22, $p=.64$	$F(1, 785)=0.11$, $p=.74$	$F(1, 1694)=0.66$, $p=.42$
Subject (S)	1 + M + P + T|S (10) 914, $p<.001$	1 + M + P + T|S (10) 128, $p<.001$	1 + M + P + T|S (10) 61.6, $p<.001$	1 + M + P + T|S (10) 108, $p<.001$
Item (I)	1|I (1) 31.1, $p<.001$	1|I (1) 1.02, $p=.31$	1|I (1) 1.26, $p=0.26$	1|I (1) 12.2, $p<.001$
Overflow (O)	NA	1|O (1) 11.4, $p<.001$	NA	1|O (1) 14.7, $p<.001$

Note: χ^2 = likelihood ratio. LME = linear mixed effects; FF, AF, RR, and BF refer to, respectively, first fixations, additional first-pass fixations, rereading fixations, and spillover buffer fixations; NA means not applicable. The fitted model represents a maximal random effects structure, as explained in the text, and all the fixed effects. The fixed effects are expressed as *F*-values with approximate degrees of freedom for the denominator (here rounded to the nearest integer); these are shown in the upper part of the table. The random effects are based on the likelihood ratio statistic, which follows a χ^2 distribution, and are shown in the lower part of the table. In this part, each cell contains a specification of the included intercept and slope(s). The notation 1 + M + P + T|S indicates that the effect included a random intercept and random slopes for match, plausibility, and tone, whereas 1|S has only a random intercept. The number of degrees of freedom is shown between parentheses after the specification of the included slopes. The remainder of each cell displays the χ^2 value and its probability.

match, with longer durations and more frequent fixations when the grammatical number of the modifier NP mismatched the number of the head NP. The expected interaction of number match and tone was present only at the additional first-pass fixation durations: The tone effect was larger in sentences with a mismatch than in sentences with a matching grammatical number between the head and the modifier NP. Rereading fixations, in contrast, showed no effects whatsoever of grammatical number match. On almost half of the trials, the response to the tone task flowed over into the buffer region. Although the match effect was present at this region, the tone task did not affect reading duration, and it never flowed over beyond this buffer region. Given systematic effects of the presence of the tone task on fixations within the modifier NP region and the absence of any effects of the tone task beyond that region, it is clear that the effects of the tone task were confined to the modifier NP region.

The findings clearly point at an *integration cost* at the modifier region for mismatching NPs, suggesting that the computation of the grammatical number of the complex subject NP takes longer when the head and modifier NP mismatch in grammatical number than when they match. The present findings of an integration cost in Dutch sentence comprehension add to the evidence for an integration cost in sentence comprehension in other languages, like English (Nicol et al., 1997; Pearlmutter et al., 1999; Wagers et al., 2009) and German (Häussler, 2009), and are also in line with integration costs found in production studies (Eberhard et al., 2005; Vigliocco et al., 1996; Vigliocco & Nicol, 1998). An alternative interpretation of the number mismatch effects at the modifier region is that plural modifiers require longer processing time than singular modifiers (Wagers et al., 2009). Indeed, as the critical sentences only had singular head nouns, all sentences with number mismatches between the head and modifier NP had a plural modifier NP. This issue is addressed later in this section.

Processing times at the verb showed a clear pattern of results with a consistent increase of duration and frequency of fixations due to the presence of the tone task. No other significant main effects or interactions were observed, except for an effect of match in the frequency of rereading fixations. As these fixations imply a possible revisit of the modifier noun, it is safer not to try to interpret this finding.

Considering the two regions together, the present results do not seem to support the hypothesis that plausibility of the modifier as a subject for the verb has an effect at the checking phase. In fact, at the verb region, in agreement with the findings of Wagers et al. (2009), the expected main effect of plausibility was not observed, whereas at the modifier NP region, we observed that grammatical number matches affected fixation durations but also no effects of plausibility.

We also investigated the observed costs (integration and checking) with respect to our second research question and examined whether these processes require executive control. For this investigation, we tested the effects of the presence of a concurrent tone-discrimination task in either of the two target regions, the modifier noun and the verb. The presence of such a task creates a dual-task situation, which results in a slowing of performance, usually on both tasks. This was also the case in the present study: All gaze durations were slower when the tone discrimination task was present, and the responses to the tones themselves were also slower than when the tone task was performed in single-task conditions. In order to be able to conclude that the tone discrimination task and the sentence reading task are calling on common control processes, an interaction must be found between a process that allegedly calls on executive control and the presence of the tone discrimination task. At the modifier noun, we observed that in the additional first-pass fixation duration, grammatical number match interacted with the presence of the tone task so that the difference between match and mismatch processing was augmented when a concurrent tone task was performed. Thus these findings support the hypothesis that resolution of number match conflicts calls on executive control. In particular, the presence of this interaction raises some doubt about the alternative interpretation based on the findings of Wagers et al. (2009) that the mismatch effect is simply due to longer processing times needed for plural nouns. Although plural processing slow-down may indeed account for the mismatch effect, it does not seem likely that simply processing a plural instead of a singular noun would require executive control and would thus be impaired by the presence of a tone discrimination task. A more plausible account, in our view, is that a number conflict between the head NP and the modifier NP has to be resolved and that this resolution process involves executive control.

Interestingly, the present study revealed no interactions of the tone-discrimination task with any other effects at the verb. The present findings thus suggest that the checking phase performed at the verb region basically runs off automatically, at least for the sentences used in this study.

Although many theories of language comprehension assume a role of working memory in sentence comprehension processes, only few studies have asked whether a grammatical number conflict results in an increased demand on working memory (Fedorenko et al., 2006; Lewis et al., 2006; Swets et al., 2007). The present study shows that it does, and it specifically reveals the involvement of domain-general processes related to executive control. This involvement is localized at the integration phase, and no evidence was found for such control processes at the checking phase. It would be interesting to explore in future research whether the working memory cost found in studies on production of subject–verb agreement (Fayol et al., 1994; Hartsuiker & Barkhuysen, 2006) also reflects the involvement of executive control at the integration phase and whether the checking phase does play a role in production, as has been suggested (Badecker & Kuminiak, 2007).

The dual-task interference effect in the present sentence comprehension study indicates that the parser relies on the domain-general cognitive mechanism of executive control for the integration process. This finding provides direct support for the account of Thompson-Schill (2005) and others (January, Trueswell, & Thompson-Schill, 2009; Kuperberg, 2007; Novick, Kan, Trueswell, & Thompson-Schill, 2009; Novick, Trueswell, & Thompson-Schill, 2005; Ye & Zhou, 2008, 2009), which postulates that many sentence comprehension processes rely on executive control. On a broader scale, the finding adds to the evidence that at least some processes in sentence comprehension (and production), but presumably not all of them, require domain-general mechanisms that are also useful in other cognitive domains, within language processing and even beyond (Fedorenko et al., 2006; Just & Carpenter, 1992; Lewis et al., 2006; Loncke et al., 2011; McElree et al., 2003; Swets et al., 2007).

Terms like "executive control" and "domain-general processes" cover rather broad areas of control mechanisms. It would be interesting, therefore, to know which specific control process is taxed in sentence comprehension. As the present study was confined to a comparison between processing times of sentences with or without a local dual-task load, it does not allow for empirically based conclusions about the kind of process involved. However, taking into account recent developments regarding executive control (e.g., Vandierendonck, 2016), it is possible to speculate on the possibilities. The tone presented at a particular region induces an interruption of the presently ongoing task—namely, sentence processing. If the current sentence-processing episode runs off automatically, there is no big problem because in that case the tone-discrimination task can be performed without interference. Nevertheless, some slowing of sentence processing must be expected, because another task set must be configured. Such slowing would be the same over all the variations of the same sentence, at least if the different versions of the sentences (match vs. mismatch, plausible vs. implausible) can be performed automatically.

In contrast, if the sentence-processing episode in one or more versions of the sentence requires executive control, interference is expected to occur. There are two options: Either sentence processing is interrupted immediately in favour of the tone-discrimination task or the tone is maintained in temporary memory to be executed after the current processing episode. In the former case, sentence processing is interrupted and kept on hold until the tone-discrimination task has been configured, and the appropriate response has been selected, after which the sentence-processing task is continued. Considering that sentences with a number mismatch between the head NP and the modifier NP require a selection of an appropriate action—namely, choose either the head NP or the modifier NP as the subject of the sentence—this process will be delayed because the secondary tone-discrimination task requires the same mechanism to select the appropriate response. Sentences with matching numbers between the two NPs do not require such control because no choice is required. Consequently, processing of number-matching stimuli may be slowed by the mere fact of the presence of a secondary task, as explained in the previous paragraph, but the slowing will be less than that of the mismatching sentence versions because in the latter, sentences are also delayed by the presence of the secondary task, and, in addition, they suffer from a delay due to the competition for the same selection process. This explanation is consistent with the findings presented in Table 2. Without tone task, there is almost no difference in fixation duration between matching and mismatching sentences (290 vs. 309 ms), but when a tone is present, matching sentences are 104 ms slower, while mismatching sentences are 156 ms slower.

As indicated, it is also possible that sentence processing continues and that the tone-discrimination task is postponed. This occurs when the response to the tone-discrimination task occurs after the end of the additional first-pass fixations. In this particular case, the fixation durations of matching and mismatching sentences are much shorter (354 ms) than when the tone task ends earlier (479 ms), but they are slower than when no tone is present (299 ms) because some processing must be performed to safeguard the tone in working memory.

In order to ascertain the scope of the present findings, future research should explore whether the domain-general mechanism of executive control is also involved in the comprehension of other sentence types that are known to bring about processing costs (Loncke et al., 2011). The gaze-contingent dual-task technique used in the present study could be particularly useful for doing so. The present study indicates that the technique is effective and hence promising to test whether and to which degree executive control is involved at particular processing phases. Put differently, this technique seems helpful to specify the periods of higher and lower density in the demands for executive control over the processing stages in the comprehension of sentences.

To conclude, using a gaze-contingent dual-task procedure, we observed a processing cost for sentences with mismatching head and modifier NPs in the integration phase (at the modifier noun) but not in the checking phase (at the verb). The integration cost suggests that computation of the grammatical number of the complex subject NP is more effortful when the head and modifier NP mismatch in grammatical number than when they match. The presence of a secondary tone discrimination task during these integration and checking processes amplifies the effect of number match, showing that this process does not occur automatically, or at least does not occur automatically all the time. Moreover, when rereading of sentence parts before the modifier NP region is required, local plausibility of the modifier noun seems to play a role, and this process also taxes executive control.

Notes

1. The complete list of sentences is available online at osf.io/u6xdp. This URL also contains the data obtained in the present study and the details of the data analysis.
2. That the majority of the filler sentences had a similar syntactic structure to the critical sentences may be a disadvantage of the present design. However, we considered it important that participants saw also an important number of sentences with a head NP that was not singular in order to avoid a reading strategy in which not much attention would be devoted to the head NP. Furthermore, it was important also to include the secondary task at other regions than the two regions of interest. Given this restriction and the need for a sufficiently large number of critical sentences, it was not possible to add more types of fillers.
3. Note that the boundary of the region of interest is crossed during the execution of a saccade, so that the tone co-occurred with the fixation. It is possible that the saccade would skip the region of interest so that the tone would occur in a later region. In such case, the trial was excluded from data analysis just like other trials where the region of interest was skipped.
4. Most studies of eye tracking in reading use a different scheme to distinguish between early or later fixations. In contrast to schemes typically used, the present scheme partitions the course of fixations at a region of interest into three non-overlapping time frames. This has the advantage that each time frame yields data that are independent from the data in the time frames that follow later.
5. As the dependent variables, gaze duration and reaction time, tend to deviate from normality by showing a rightward skewness, a correction seems necessary. The Appendix discusses the possibilities and reports on outcomes of alternative analyses.
6. By definition, overflow correlates with tone as overflow can only occur when a tone is present. For that reason, it is not possible to include a random slope for tone on overflow, because this could result in singularity of the variance–covariance matrix.

Acknowledgements

We would like to thank Denis Drieghe for eye-tracking advice, and Fernanda Ferreira, Els Severens, Sarah Bernolet, and Eva Van Assche for their valuable comments on previous versions of this manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the Research Foundation–Flanders [grant number G.0052.08 to André Vandierendonck, Timothy Desmet, and Robert J. Hartsuiker].

ORCID

André Vandierendonck http://orcid.org/0000-0002-9493-8862
Robert J. Hartsuiker http://orcid.org/0000-0002-3680-6765

References

Acuna-Farina, J. C., Meseguer, E., & Carreiras, M. (2014). Gender and number agreement in comprehension in Spanish. *Lingua*, *143*, 108–128. doi: 10.1016/j.lingua.2014.01.013

Allen, K. V., Pickering, M. J., Zammitt, N. N., Hartsuiker, R. J., Traxler, M. J., Frier, B. M., & Deary, I. J. (2015). Effects of acute hypoglycemia on working memory and language processing in adults with and without type 1 diabetes. *Diabetes Care*, *38*(6), 1108–1115. doi: 10.2337/dc14-1657

Antón-Méndez, I., & Hartsuiker, R. J. (2010). Morphophonological and conceptual effects on Dutch subject-verb agreement. *Language and Cognitive Processes*, *25*(5), 728–748. doi: 10.1080/01690960903513818

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi: 10.1016/j.jml.2007.12.005

Badecker, W., & Kuminiak, F. (2007). Morphology, agreement and working memory retrieval in sentence production: Evidence from gender and case in Slovak. *Journal of Memory and Language*, *56*(1), 65–85. doi: 10.1016/j.jml.2006.08.004

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods Instruments & Computers*, *28*(4), 584–589. doi:10.3758/bf03200546

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi: 10.1016/j.jml.2012.11.001

Barrouillet, P., Bemardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 570–585. doi: 10.1037/0278-7393.33.3.570

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. doi: 10.1037/0096-3445.133.1.83

Barrouillet, P., & Camos, V. (2010). Working memory and executive control: A time-based resource-sharing account. *Psychologica Belgica*, *50*(3–4), 353–382. doi: 10.5334/pb-50-3-4-353

Barrouillet, P., Corbin, L., Dagry, I., & Camos, V. (2015). An empirical test of the independence between declarative and procedural working memory in oberauer's (2009) theory. *Psychonomic Bulletin & Review*, *22*(4), 1035–1040. doi: 10.3758/s13423-014-0787-y

Barrouillet, P., Lépine, R., & Camos, V. (2008). Is the influence of working memory capacity on high-level cognition mediated by complexity or resource-dependent elementary processes? *Psychonomic Bulletin & Review*, *15*(3), 528–534. doi: 10.3758/pbr.15.3.528

Barrouillet, P., Portrat, S., & Camos, V. (2011). On the law relating processing to storage in working memory. *Psychological Review*, *118*(2), 175–192. doi: 10.1037/a0022324

Barrouillet, P., Portrat, S., Vergauwe, E., Diependaele, K., & Camos, V. (2011). Further evidence for temporal decay in working memory: Reply to Lewandowsky and Oberauer (2009). *Journal of Experimental Psychology: Learning Memory and Cognition*, *37*(5), 1302–1317. doi: 10.1037/a0022933

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Bates, E., McNew, S., Macwhinney, B., Devescovi, A., & Smith, S. (1982). Functional constraints on sentence processing: A cross-linguistic study. *Cognition*, *11*(3), 245–299. doi: 10.1016/0010-0277(82)90017-8

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*(1), 99–127. doi: 10.1016/0749-596x(92)90007-k

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language and Cognitive Processes*, *8*(1), 57–99. doi: 10.1080/01690969308406949

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*(1), 45–93. doi: 10.1016/0010-0285(91)90003-7

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*(3), 624–652. doi: 10.1037//0033-295x.108.3.624

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, *26*(2), 211–252.

Burgess, P. W. (1997). Theory and methodology in executive function research. In P. Rabbitt (Eds.), *Methodology of frontal and executive function* (pp. 81–116). Hove: Psychology Press.

Dag, O., Asar, O., & Ilk, O. (2016). AID: Estimation of Box-Cox Power Transformation Parameter. Retrieved from https://CRAN.R-project.org/package=AID.

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language*, *25*(1), 1–18. doi: 10.1016/0749-596x(86)90018-5

Deschuyteneer, M., & Vandierendonck, A. (2005a). Are "input monitoring" and "response selection" involved in solving simple mental arithmetical sums?. *European Journal of Cognitive Psychology*, *17*, 347–370. doi: 10.1080/09541440440000032

Deschuyteneer, M., & Vandierendonck, A. (2005b). The role of response selection and input monitoring in solving simple arithmetical products. *Memory & Cognition*, *33*, 1472–1483. doi: 10.3758/BF03193379

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language*, *69*(2), 85–103. doi: 10.1016/j.jml.2013.04.003

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*(3), 531–559. doi: 10.1037/0033-295x.112.3.531

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *The Quarterly Journal of Experimental Psychology Section A*, *47*(2), 437–464.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources★. *Journal of Memory and Language*, *54*(4), 541–553. doi: 10.1016/j.jml.2005.12.006

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject-verb agreement errors in French and English: The role of syntactic hierarchy. *Language and Cognitive Processes*, *17*(4), 371–404. doi: 10.1080/01690960143000254

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430. doi: 10.1111/1467-9280.00475

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*(1–3), 181–204. doi: 10.1080/01690960400002117

Hartsuiker, R. J., Schriefers, H. J., Bock, K., & Kikstra, G. M. (2003). Morphophonological influences on the construction of subject-verb agreement. *Memory & Cognition*, *31*(8), 1316–1326. doi: 10.3758/bf03195814

Häussler, J. (2006). *Disrupted agreement checking in sentence comprehension*. Paper presented at the 11th ESSLLI Student Session, Malaga, Spain. Retrieved from http://staff.science.uva.nl/~katrenko/stus06/images/haeussler.pdf.

Häussler, J. (2009). *The emergence of attraction errors during sentence comprehension*. (Doctoral thesis). University of Konstanz, Germany.

Hupet, M., Fayol, M., & Schelstraete, M. A. (1998). Effects of semantic variables on the subject-verb agreement processes in writing. *British Journal of Psychology*, *89*, 59–75. doi: 10.1111/j.2044-8295.1998.tb02673.x

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. doi: 10.1016/j.jml.2007.11.007

January, D., Trueswell, J. C., & Thompson-Schill, S. L. (2009). Co-localization of stroop and syntactic ambiguity resolution in broca's area: Implications for the neural basis of sentence processing. *Journal of Cognitive Neuroscience*, *21*(12), 2434–2444. doi: 10.1162/jocn.2008.21179

Just, M. A., & Carpenter, G. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. doi: 10.1037/0033-295X.99.1.122

Just, M. A., & Carpenter, P. A. (1980). A theory of Reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. doi: 10.1037/0033-295x.87.4.329

Kail, M., & Bassano, D. (1997). Verb agreement processing in French: A study of on-line grammaticality judgments. *Language and Speech*, *40*, 25–46. doi: 10.1177/002383099704000102

Kuperberg, G. R. (2007). Corrigendum to "neural mechanisms of language comprehension: Challenges to syntax" [brain Res. 1146 (2007) 23–49]. *Brain Research*, *1164*, 142–142. doi: 10.1016/j.brainres.2007.06.029

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in Linear Mixed Effects Models. R package version 2.0-29. Retrieved from http://CRAN.R-project.org/package=lmerTest.

Lago, S., Shalom, D. E., Sigman, M., Lau, E. F., & Phillips, C. (2015). Agreement attraction in spanish comprehension. *Journal of Memory and Language*, *82*, 133–149. doi: 10.1016/j.jml.2015.02.002

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419. doi: 10.1207/s15516709cog0000_25

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454. doi: 10.1016/j.tics.2006.08.007

Liversedge, S. P., Paterson, K. B., & Pickering, M. (1998). Eye movements and measures of reading time. In G. Underwood (Eds.), *Eye guidance in Reading and scene perception* (pp. 55–75). Oxford: Elsevier Science.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*, 393–434. doi: 10.1037/0033-295X.108.2.393

Loncke, M. (2012). *The role of executive control in sentence processing*. (Doctoral thesis). Ghent University, Belgium.

Loncke, M., Desmet, T., Vandierendonck, A., & Hartsuiker, R. J. (2011). Executive control is shared between sentence processing and digit maintenance: Evidence from a strictly timed dual-task paradigm. *Journal of Cognitive Psychology*, *23*(7), 886–911. doi: 10.1080/20445911.2011.586625

Martin, R. C. (2006). The neuropsychology of sentence processing: Where do we stand? *Cognitive Neuropsychology*, *23*(1), 74–95. doi: 10.1080/02643290500179987

McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *27*(3), 817–835. doi: 10.1037//0278-7393.27.3.817

McElree, B., Foraker, S., & Dyer, L. (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language*, *48*(1), 67–91. doi: 10.1016/s0749-596x(02)00515-6

Miyake, A., & Shah, P. (1999). *Models of working memory. Mechanisms of active maintenance and executive control*. Cambridge: Cambridge University Press.

Nicol, J. L. (1995). Effects of clausal structure on subject-verb agreement errors. *Journal of Psycholinguistic Research*, *24*(6), 507–516. doi: 10.1007/bf02143164

Nicol, J. L., Forster, K. I., & Veres, C. (1997). Subject-verb agreement processes in comprehension. *Journal of Memory and Language*, *36*(4), 569–587. doi: 10.1006/jmla.1996.2497

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwarts & D. Shapiro (Eds.), *Consciousness and self-regulation* (*Vol. 4*, pp. 1–18). New York: Plenum Press.

Novick, J. M., Kan, I. P., Trueswell, J. C., & Thompson-Schill, S. L. (2009). A case for conflict across multiple domains: Memory and language impairments following damage to ventrolateral prefrontal cortex. *Cognitive Neuropsychology*, *26*(6), 527–567. doi: 10.1080/02643290903519367

Novick, J. M., Trueswell, J. C., & Thompson-Schill, S. L. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive Affective & Behavioral Neuroscience*, *5*(3), 263–281. doi: 10.3758/cabn.5.3.263

Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Eds.), *Attention* (pp. 155–189). Hove: Psychology Press.

Pearlmutter, N. J. (2000). Linear versus hierarchical agreement feature processing in comprehension. *Journal of Psycholinguistic Research*, *29*(1), 89–98. doi: 10.1023/a:1005128624716

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language*, *41*(3), 427–456. doi: 10.1006/jmla.1999.2653

Phillips, L. H. (1997). Do "frontal tests" measure executive function? Issues of assessment and evidence from fluency tests. In P. Rabbitt (Eds.), *Methodology of frontal and executive function* (pp. 191–213). Hove: Psychology Press.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-effects models in S and S-plus*. New York: Springer.

R-Core-Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi: 10.1037/0033-2909.124.3.372

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology Section A*, *53A*(4), 1061–1080. doi: 10.1080/02724980050156290

Rende, B., Ramsberger, G., & Miyake, A. (2002). Commonalities and differences in the working memory components underlying letter and category fluency tasks: A dual-task investigation. *Neuropsychology*, *16*, 309–321. doi: 10.1037//0894-4105.16.3.309

Staub, A. (2009). On the interpretation of the number attraction effect: Response time evidence. *Journal of Memory and Language*, *60*(2), 308–327. doi: 10.1016/j.jml.2008.11.002

Swets, B., Desmet, T., Hambrick, D. Z., & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General*, *136*(1), 64–81. doi: 10.1037/0096-3445.136.1.64

Szmalec, A., Vandierendonck, A., & Kemps, E. (2005). Response selection involves executive control: Evidence from the selective interference paradigm. *Memory & Cognition*, *33*(3), 531–541. doi: 10.3758/bf03193069

Tanner, D., Nicol, J., & Brehm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, *76*, 195–215. doi: 10.1016/j.jml.2014.07.003

Thompson-Schill, S. L. (2005). *Dissecting the language organ: A new look at the role of Broca's area in language processing*. Paper presented at the twenty-first century psycholinguistics: Four cornerstones.

Thornton, R., & MacDonald, M. C. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*(4), 740–759. doi: 10.1016/s0749-596x(03)00003-2

Van Dyke, J. A., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, *49*(3), 285–316. doi: 10.1016/s0749-596x(03)00081-0

Vandierendonck, A. (2016). A working memory system with distributed executive control. *Perspectives on Psychological Science*, *11*(1), 74–100. doi: 10.1177/1745691615596790

Vandierendonck, A., Deschuyteneer, M., Depoorter, A., & Drieghe, D. (2008). Input monitoring and response selection as components of executive control in prosaccades and antisaccades. *Psychological Research*, *72*, 1–11. doi: 10.1007/s00426-006-0078-y

Vergauwe, E., Barrouillet, P., & Camos, V. (2009). Visual and spatial working memory are not that dissociated after all: A time-based resource-sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(4), 1012–1028. doi: 10.1037/a0015859

Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science*, *21*(3), 384–390. doi: 10.1177/0956797610361340

Vigliocco, G., Hartsuiker, R. J., Jarema, G., & Kolk, H. H. J. (1996). One or more labels on the bottles? Notional Concord in Dutch and French. *Language and Cognitive Processes*, *11*(4), 407–442. doi: 10.1080/016909696387169

Vigliocco, G., & Nicol, J. (1998). Separating hierarchical relations and word order in language production: Is proximity Concord syntactic or linear? *Cognition*, *68*(1), B13–B29. doi: 10.1016/s0010-0277(98)00041-9

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206–237. doi: 10.1016/j.jml.2009.04.002

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, *54*(3), 426–482. doi: 10.1090/S0002-9947-1943-0012401-3#sthash.BveKeG8f.dpuf

Ye, Z., & Zhou, X. L. (2008). Involvement of cognitive control in sentence comprehension: Evidence from ERPs. *Brain Research*, *1203*, 103–115. doi: 10.1016/j.brainres.2008.01.090

Ye, Z., & Zhou, X. L. (2009). Executive control in language processing. *Neuroscience and Biobehavioral Reviews*, *33*(8), 1168–1177. doi: 10.1016/j.neubiorev.2009.03.003

Appendix

When the distribution of the dependent variable is skewed to the right, the assumptions about the normality of the distribution may be violated. To obtain a more adequate analysis of the data, a number of possibilities can be considered. For reaction times, typically, the outliers are excluded (e.g., beyond 3 standard deviations of the mean). Because this strategy results in data loss, some researchers prefer a transformation of the data to bring them closer to the normal distribution. Frequently, a logarithmic transformation is used to achieve this, but any power transformation may be considered, and it is possible to select the best possible power transformation (Box & Cox, 1964). From a mathematical point of view this is probably the best strategy. Nevertheless, data transformation has a number of disadvantages. For one thing, transformation changes the scale of the dependent variable; the resulting scale is unnatural, and interpretation of the data becomes more difficult. Furthermore, application of linear statistics to nonlinearly transformed data makes it difficult to understand what the effect of a manipulation really means. In particular, some types of interactions, such as over-additive interactions, become much smaller and may fail to attain significance.

Clearly, researchers differ in their preferences for the treatment of data that deviate from normality. In order to allow an inspection of the data from different perspectives, in addition to the analyses of untransformed data as reported in the main body of the article, we also performed an analysis with exclusion of the outliers. These analyses yielded the same significant effects and are therefore not considered in more detail.

For the analyses based on transformed data, first the best value for the λ-parameter in the Box–Cox formula (Box & Cox, 1964) was estimated with the boxcoxnc function of the AID package (Dag, Asar, & Ilk, 2016) in R. In all analyses, the estimated value was near zero, indicating that a logarithmic transformation was most suitable. Table A1 shows the significant fixed effects obtained in the analyses of fixation in the respective regions of interest. Differences between this table and the results shown in the main body of the article are limited to the first-pass and rereading fixations in the modifier NP region, and the additional first-pass fixations in the verb region. More specifically, the log analysis revealed an interaction of Match × Plausibility (not expected) in the first-pass fixations of the modifier NP region, an interaction of Plausibility × Tone in the rereading fixations of the modifier NP region (in the absence of any plausibility effects, this is not expected), and also an interaction of Plausibility × Tone in the additional first-pass fixations of the verb region (again, in the absence of any main effect of plausibility, this is not expected). Not shown in the table, the log analysis of the RTs in the dual-task conditions revealed a main effect of region to which the tone was applied. All the other effects shown in the table were also present in the analysis of the untransformed data, which did not detect any effects that were not present in the log-transformed data analysis.

These small differences have no effects as to the interpretation of the data with respect to the research problem addressed in this article. The observation of a significant Match × Plausibility effect in the first-pass fixations of the modifier NP region is interesting, as it suggests that plausibility could play a role at the integration phase. However, in the absence of any further confirmation, this result should be considered with care. The observation of a

Plausibility × Tone interaction is not so easy to interpret. The fact that it occurred at rereading in the modifier NP region again suggests a role for plausibility in the integration phase; that plausibility interacts with the presence of the tone without a main effect of plausibility itself probably means that the presence of the tone may have required some rereading of the modifier with special attention to the meaning of the NP. The interaction of plausibility and tone in the additional first-pass fixations of the verb region may be taken to mean that plausibility plays a role in the checking phase; however, once more, without a main effect of plausibility, this interaction is difficult to interpret.

Table A1. Significant fixed effects in the log-analysis of the fixations in the regions of interest and in the buffer regions and of the tone RTs.

	FF	AF	RR	BUF
Modifier NP region				
M	$F(1,42) = 7.46$, $p < .01$			$F(1,41) = 5.45$, $p < .05$
T	$F(1.49) = 148.7$, $p < .001$	$F(1,44) = 132.9$, $p < .001$	$F(1,38) = 113.1$, $p < .001$	
M x T		$F(1,453) = 6.08$, $p < .05$		
M x P	$F(1,4566) = 3.85$, $p < .05$			
P x T			$F(1,453) = 6.08$, $p < .05$	
Verb region				
M		$F(1,363) = 4.37$, $p < .05$		
T	$F(1,44) = 72.66$, $p < .001$	$F(1,42) = 136.86$, $p < .001$	$F(1,44) = 24,76$, $p < .001$	$F(1,46) = 4.91$, $p < .05$
P x T		$F(1,1383) = 5.72$, $p < .05$		

Note. The following abbreviations are used in the table: for the regions, FF = first-pass fixation, AF = additional first-pass fixation, RR = rereading fixation, BUF = spill-over buffer; for the effects: M = match, P = plausibility, T = Tone.