On measuring multiple lexical activation using the cross-modal semantic priming technique

Esther Janse Hugo Quené

Utrecht institute of Linguistics OTS

Utrecht University

The Netherlands

manuscript resubmitted to

Behaviour Research Methods, Instruments and Computers

date: December 2003

running title: Cross-modal semantic priming technique

word count: 7758 (50 pages in total)

Address for correspondence:
Esther Janse
Utrecht institute of Linguistics OTS
Trans 10
NL-3512 JK Utrecht
The Netherlands
E-mail: Esther.Janse@let.uu.nl

Abstract

Cross-modal semantic priming with partial auditory primes seems a good technique to assess spoken-word recognition, because it allows tracing the activation of multiple word candidates. However, previous research using this technique has found inconsistent results. First, a priming experiment is reported that addresses this technique's validity. Results show that semantic priming is not observed with partial auditory primes, but only with full primes, that is, after competition between multiple words has left only one word active. Secondly, Monte Carlo simulations are reported of a previous study that found partial priming effects; the simulations show that the particular design in that study yields a high risk of a Type-I error. In conclusion, the semantic priming technique cannot be used to investigate activation of multiple word candidates, and its use for that purpose should be discontinued.

key words:

cross-modal semantic priming, spoken-word recognition, multiple activation, Monte Carlo simulation

1. Introduction

Among current models of spoken-word recognition, there is remarkable agreement about two basic stages or processes. In the first or “access” stage, multiple word candidates in the listeners’ mental lexicon are activated that roughly match the available auditory input. In the second or “selection” stage, activation of these lexical candidates increases or decreases when more auditory input becomes available, or when semantic context starts to influence the selection (Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986; Norris, 1994; McQueen & Cutler, 2000).

Evidence for multiple activation of word candidates in the “access” stage comes from several studies that employed the cross-modal semantic priming paradigm. The cross-modal semantic priming technique in its most common form was first introduced by Swinney (1979). The technique is based on spreading of activation from one lexical element to other semantically or associatively related elements (Collins & Loftus, 1975). Semantically related items in the mental lexicon are interconnected via facilitating links: an increase in the activation of one item leads to an automatic increase in the activation of semantically related items.

If listeners are required to make a lexical decision on a visually presented target word (e.g., MONEY) after hearing an auditory prime word (e.g., salary;

throughout this paper, visual target words will be represented in roman uppercase and auditory prime words in lowercase italics), then they react faster if the visual target word and the auditory prime word are semantically or associatively related, in comparison to a control condition in which these words are not related (e.g., MONEY after unrelated control prime piano). These auditory prime words can be cut off before their acoustic offset (e.g., sala-…), at a point where the acoustic information is not sufficient to identify the intended word uniquely. This is referred to as partial priming. Supposedly, at the point in time where the prime is cut off, multiple word candidates are still active. By presenting a visual target related to one of these candidates immediately following the presentation of a partial prime fragment, one can measure the activation of that word candidate. This implies that supposedly, even before spoken words are completely recognised, they have already sent a detectable amount of activation to their semantic associates.

Zwitserlood (1989) used the cross-modal semantic priming task to investigate the activation of multiple word candidates, using partial priming. Her study was set up to investigate during which stage context affects the activation of lexical candidates. The various models of auditory word recognition make different predictions with respect to the relative weights of sentence context and bottom-up acoustic information during word processing. These differences mainly concern the moment at which sentence context starts to influence processing. In the autonomous models (Forster, 1976; 1979), lexical access and selection proceed on the basis of bottom-up acoustic information alone. Sentence context only plays a role in word perception if the speech input is ambigous between lexical solutions. By contrast, in a typical interactive model, such as TRACE (Elman & McClelland, 1985, McClelland & Elman, 1986), sentence context would have an effect on lexical candidates <u>at any moment</u> during word recognition, even before an acoustic stimulus is presented. The results of the Zwitserlood (1989) study showed two important things. First, multiple lexical candidates are accessed on the basis of partial auditory information: even when only a fragment of an auditory prime word is presented (e.g., <u>sala-…</u>), at a point where the acoustic information is not sufficient to identify the intended word uniquely, activation can be found for both word candidates <u>salaris</u> ('salary') and <u>salami</u> ('salami'), as witnessed by the semantic priming effect for both the semantically related visual targets GELD ('MONEY') and WORST ('SAUSAGE') (the original Zwitserlood study was conducted in Dutch; English translations will be provided in brackets). Second, Zwitserlood (1989) showed that sentence context can affect the activation of lexical candidates, thus providing evidence for a hybrid model of word recognition. However, after the Zwitserlood (1989) study, semantic priming studies have yielded inconsistent results, if any. The validity of this task is doubtful, and any effects are small and inconsistent, especially in sentence context (Gaskell & Marslen-Wilson, 1996; Jongenburger, 1996). Chwilla (1996) found no partial priming effects, even though part of her material was identical to that of Zwitserlood (1989). Zwitserlood & Schriefers (1995) found that a short prime fragment only yielded a priming effect when extra processing time was available (by inserting a time interval between prime fragment offset and presentation of the visual target). Hence the "selection" stage of the recognition process may have been already concluded by the time the listener responded. Using auditory stimuli that were phonetically ambiguous with respect to the voicing value of the initial consonant (e.g., between dip and tip), Connine, Blasko, & Wang (1994) observed multiple activation effects before the isolation point. The reported average lexical decision times are relatively long, however, which again raises questions about the on-line nature of these effects. Moss, McCormick, & Tyler (1997) reported only a weak semantic priming effect of 10 ms at the isolation point. Note that in two of these studies (Zwitserlood & Schriefers, 1995; Moss, McCormick, & Tyler, 1997), activation was only measured for the actual prime word, and not for other candidates that competed with the actual prime word. Consequently, these studies bear only little evidence on the matter of multiple activation.

So, whereas robust semantic priming effects have been reported with full primes (Meyer & Schvaneveldt, 1971; Neely, 1977; Swinney, 1979), it is important to know whether this technique also gives reliable and robust results when partial primes are presented. Importantly, the Distributed Cohort Model

(Gaskell & Marslen-Wilson, 1997; 1999; 2002) provides an explanation why semantic priming effects obtained with partial primes may not be as robust as some earlier results suggest. In the Distributed Cohort (henceforth DC) model, the process of speech perception is modelled as a recurrent neural network. Lexical units are points in a multidimensional space, represented by vectors of phonological and semantic output nodes. The speech input maps directly and continuously onto this lexical knowledge. As more bottom-up information becomes available, the network moves towards a point in lexical space corresponding to the word under consideration. Activation of a word candidate is then inversely related to the distance beteen the output of the network, which represents a point in the multidimensional lexical space, and the word representation in this space. In connectionist models such as this, multiple representations must interfere with each other if they are active simultaneously. This was also modelled in two older models, TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1984), in which lateral inhibition between activated word candidates is employed to reduce multiple activations. Before the uniqueness point of the to-be-recognised word, semantic activation depends stronly on the number of candidates that match the input sofar and their relative frequency.

Gaskell & Marslen-Wilson (1999) explain why the effects of phonological priming (also termed repetition, candidate or identity priming) are generally much stronger than those of semantic priming if partial primes are presented. In phonological priming, the relation between the prime-target pair is such that the visual target (e.g., PORT or PORK) is fully or partially identical to the auditory prime (e.g., por-…) in terms of its acoustic phonetic form. In the DC model, priming for a certain candidate occurs of a lexical representation is more similar to the target representation than to an unrelated baseline. Phonologically, the word candidates are obviously coherent, but the semantic representations of the different candidates often have no meaning overlap at all. "In repetition priming, the target lexical representation is related to the prime representation in all dimensions, so recognition of te target can take advantage of overlap on both semantic and phonological nodes (…). By contrast, semantic priming relies on overlap in the semantic nodes alone" (Gaskell & Marslen-Wilson, 1999, p. 452). Empirical results (Gaskell & Marslen-Wilson, 2002) support their claim that the phonological effects of partial priming are much stronger than the semantic effects. In their experiment, primes were presented either complete, or in two cut-off conditions. Semantic priming occurred only after the moment that the prime has become unambiguous onwards. By contrast, significant phonological priming effects were found at all cut-off points. In summary, partial priming (activating multiple candidates) necessarily leads to weak activation of the candidates' semantic associates, because the multiple candidates do not share common semantic properties. While multiple candidates are active, their disparate semantic properties provide only weak semantic priming of other words, if any. Activation of their phonetic associates is stronger, because all candidates share common phonetic properties. With full priming, only a single candidate remains, and its coherent semantic properties provide strong semantic priming of its related words.

Our understanding of spoken-word perception crucially depends on the concepts of automatic spreading of activation to semantic associates and of multiple activation of word candidates, as well as the time courses of these processes. The discrepancy among semantic priming studies cited above undermines our understanding of spoken-word perception.

The present study attempts to clarify this situation, in two ways. Its first aim is to provide decisive empirical evidence about partial semantic priming. This is done by means of a modified replication of the cross-modal semantic priming study by Zwitserlood (1989), which has provided the strongest evidence in favour of partial priming. In our "heteromethod" replication (Campbell, 1969), the original design and stimulus materials were slightly modified, to improve the chances of detecting partial priming effects. If the effect of semantic priming after the presentation of partial primes is robust enough, then we should also be able to find it with a design that is different from the original study. Nevertheless, results from this experiment indicate that priming occurs only after a single word candidate remains, i.e., no partial priming was observed. The second, methodological aim of this study is to illustrate the use of Monte Carlo simulations for post-hoc evaluation of experimental designs. Such simulations provide realistic estimates of the chances of Type I and Type II errors, even for complicated repeated-measures designs. Hence, they provide a relatively easy alternative to formal power analyses. Simulation results indicate that in the original study by Zwitserlood (1989), the chance of a Type I error was far greater than the stated level of significance.

Even if partial primes do not yield reliable semantic priming effects, they may very well activate multiple word candidates. It is quite likely that there is activation of multiple word candidates, but that it cannot be measured reliably via semantic priming. One explanation could be that the time course of the spreading of activation to semantically related items is somehow not immediate. Either the spreading of activation itself is delayed, or ativation of a word candidate may have to surpass a certain threshold level before it passes on to semantically related items. This latter explanation agrees with the idea of Gaskell & Marslen-Wilson (1999) that semantic priming cannot be reliably detected as long as the semantic representations of the different candidates have no meaning overlap at all. Therefore, the purpose of this study is not to argue against multiple activation. Instead, its main point is to show that after partial priming (activating multiple word candidates), the activation levels of these candidates' semantic associates are too low to be measured reliably.

2. Experiment

2.1 Design

Since Zwitserlood's sentences and test words have yielded the clearest partial priming effects to date, her (Dutch) materials were used whenever possible in the present experiment (also with Dutch listeners). The main difference here is in our experimental design. Due to her large number of conditions, Zwitserlood (1989) was forced to use an incomplete between-subjects design in which the effects of listeners and of conditions were confounded. The present experiment focuses on multiple activation, and it ignores the sentence context factor, yielding fewer conditions. A within-subjects design is therefore possible in the present experiment: each subject is presented with all conditions, but balanced over the different word items such that multiple presentation of the same word item (beit in different conditions) can be avoided.

Although a strict replication of Zwitserlood's experiment might seem a logical step, it may not be very informative. Whatever the results of this replication might be, it would not provide us with decisive information whether reliable semantic priming effects can be found after the presentation of partial primes. If we find a partial priming effect using the same design and set-up, this finding might again be related to the infelicitous design. If we do not find an early priming effect, this finding may result from the activation of the word candidates still being too small to be detected reliably. If the effect of semantic priming after partial primes is robust enough, then we should also be able to find it with a design that is different from Zwitserlood's.

Three factors are varied in the present cross-modal priming study. First, two word candidates are investigated for each auditory prime word: e.g., both the word salaris 'salary' and its closest competitor salami 'salami'. This Candidate factor involves cohort members: for partial primes, both the actual (intended) spoken word (e.g., salaris) and its competitor (salami) should prime each other's visual targets equally, hence priming is expected for both GELD ('money') and for WORST ('sausage'). For full primes, priming is expected only for the semantically related target (e.g., only for GELD, not for WORST).

The second factor varies the length of the auditory prime: either full (the prime word is presented in full, and the visual target is presented at the prime word's offset) or partial (the prime word is cut off at a point where both the actual prime word and its competitor are still compatible with the partial prime, and the visual target is presented at this cut-off point). Thirdly, priming effects are obtained by comparing a visual target (e.g., GELD 'money') that is semantically related to an auditory prime word (e.g. salaris 'salary') against the same visual target that is not related to the auditorily presented word (e.g., control piano 'piano'). The related (test) and unrelated (control) versions need to be presented to different listeners, because familiarity effects may arise if participants see the same visual target repeatedly. Consequently, the priming effect can only be determined indirectly, as an additional Relatedness factor (evaluated between listeners within item, or between items within listener).

A full factorial design of these three factors (Candidate, Prime Length, Relatedness) yields 2×2×2= 8 experimental conditions. Given that there are 24 stimulus items in Zwitserlood (1989; details are given below), and given that a balanced within-subject design is desirable, the 8 conditions would be counterbalanced over the 24 items on 8 different experimental lists, yielding 3 items per listener per condition. This design can be reduced, because with only 4 (instead of 8) listeners groups, repeated presentation of the same visual targets and auditory primes can still be avoided. Two conditions can be presented to the same group of participants. For the salaris prime word, for example, both the salaris/GELD ('salary/money') combination and the piano/WORST ('piano/sausage') combination (unrelated control word piano combined with unprimed target WORST, which is the semantic associate of competitor salami) were presented to the same listener group. Similar combinations were possible for all 4 listener groups. Thus, the design is not strictly Latin square, but it is the most efficient within-subject design possible. A schematic overview of the design of this experiment is provided in Table 1.

--- INSERT TABLE 1 ABOUT HERE ---

2.2 Materials

Prime words

All 24 Dutch prime words from Zwitserlood (1989) were re-used here. Only the carrier phrase condition was used. A male native speaker of Dutch read these phrases containing the prime words at a normal speaking rate. This was recorded on DAT tape with a Sennheiser ME30 microphone.

Cut-off points and competitors

A gating experiment was set up, in order to determine competitor words and appropriate cut-off points for the prime words, corresponding to the multiple activation stage. The 24 prime words were presented in their carrier phrase to 12 listeners in increasingly longer portions. On each presentation listeners were asked to write down what they thought the word was going to be. The isolation point, defined as the average gate at which listeners first come up with the intended prime word, without changing their response at later gates (Grosjean, 1980), served as the partial cut-off point. This procedure was similar to that in the Zwitserlood (1989) study.

For each prime word, its appropriate competitor was chosen from the responses mentioned in the gating task. This procedure ensures that competitors are indeed true competitors in the mental lexicon of participants, rather than artificial competitors determined arbitrarily by dictionary look-up.

Visual targets

Zwitserlood (1989) used the same targets repeatedly for multiple items in multiple conditions (which she presented to different participants). To avoid such duplicates in our within-subject design, new semantic associates were required. An association test was carried out to find appropriate associates to replace these duplicates, to find associates for new competitors, and to verify the associations for other words. Fifty-two participants were asked to give three associates (in the order in which they came to mind) to each of the actual prime words and their competitors. Whether a certain response was given as first, second or third associate determined its weight factor. This weight factor, together with the number of participants "voting" for an associate, were used as criteria to select the optimal associate.

Control words

Auditory non-priming control words were chosen, matching the prime words with respect to number of syllables, stress pattern, phonological structure, and word frequency (CELEX 1990). Control words, embedded in the same carrier phrase, were either presented in full, or cut off at approximately the same length as the corresponding partial test prime words.

The resulting materials are listed in Appendix I.

2.3 Design and Procedure

The 4 test conditions and the 4 corresponding control conditions were rotated and counterbalanced over 4 stimulus lists, as explained above. Each presentation consisted of an auditory prime or control word embedded in a carrier sentence. The visual target was presented immediately at the acoustic offset of the prime or control stimulus and remained visible for 50 ms. Listeners were seated in front of a computer screen in a sound-insulated cabin. Listeners were instructed to listen carefully to the audio material that was presented over closed earphones, and they were asked to give a lexical decision response to the visual target as fast and as accurately as possible. A button box was used to register lexical decisions: subjects were asked to rest their dominant hand (e.g. their index finger) on the YES button and their non-preferred hand was to rest on the NO button. Once they had pressed either button, the response could not be corrected. The main part of the experiment consisted of 48 test and control items, randomly mixed with 58 filler items. The total material set was balanced for visual words and nonwords. In addition, each run started with a practice session involving 12 items, after which feedback and additional instruction was possible. The inter-stimulus interval was 4 seconds. The order of items was randomised for each participant.

Reaction times (henceforth RTs) of lexical decisions were measured from the onset of the visual target presentation until one of the response buttons was pressed. Non-responses and responses with RTs exceeding 3000 ms were counted as misses. Incorrect lexical decisions were also counted as misses because one cannot be certain whether the subject misperceived the visual target or made a mistake in selecting the correct response button.

2.4 Participants

To each of the 4 experimental lists, 15 subjects were randomly assigned, none of whom had taken part in the gating test. They were students at Utrecht University, and received a moderate sum (NLG 10) for their participation.

2.5 Results

In total, 80 out of 2880 observations were missing (3%). Misses were evenly distributed over listeners, but not over items ($\chi^2(23)=106.6$, $p<.001$). Participants had had difficulty with one particular item (the infrequent target WIMPEL ('pennant') which is the semantic associate of vlaggen ('flags');

WIMPEL is also highly confusable with the more frequent Dutch word wimper ('eye lash')). This difficulty is translated into a high miss rate (19 out of 120), by a high average RT, an by a high standard deviation of RT. This item was therefore discarded from the data set. In order to balance conditions over the 4 stimulus lists, 3 other items were discarded as well. The remainign data set contained 2400 observations, of which 51 (2%) were missing responses. Remaining misses were distributed evenly over experimental conditions [$\chi^2(7)=0.46$, n.s.].

--- INSERT FIGURE 1 ABOUT HERE ---

The 51 missing observations were replaced by the 5% trimmed grand mean of 467 ms. Data were fed into repeated measures analyses of variance, over listeners (yielding F_1 ratios) and over items (yielding F_2 ratios). Fixed within-subject factors were Relatedness (semantically related vs. unrelated), Candidate (prime word vs. closest competitor), and Prime Length (partial vs. full auditory prime). Because all these factors are binary, univariate analyses of variance were chosen. (For designs having factors with more than 2 levels, violations of the sphericity assumption may necessitate multivariate analyses. However, if the sphericity assumption is not violated or if it is irrelevant, as it is here, then univariate analyses have higher power; cf. O'Brien & Kaiser, 1985; Maxwell & Delaney, 2004).

All three main effects were significant: average RTs were slightly faster for primed, semantically related probes (473 ms) than in unrelated control conditions (484 ms); the main effect of this Relatedness factor yields $\underline{F}_1(1,56)=7.05$, $\underline{p}=0.010$; $\underline{F}_2(1,19)=5.66$, $\underline{p}=0.028$. Average RTs were faster with actual prime words (468 ms) than with their competitors (489 ms); the main effect of this Candidate factor yields $\underline{F}_1(1,56)=24.2$, $\underline{p}<0.001$; $\underline{F}_2(1,19)=13.1$, $\underline{p}=0.002$. As expected, average RTs were significantly faster in conditions with full primes (467 ms) than with partial primes (490 ms); the main effect of Prime Length yields $\underline{F}_1(1,56)=25.8$, $\underline{p}<0.001$; $\underline{F}_2(1,19)=9.56$, $\underline{p}=0.006$.

As expected, the two-way interaction of Relatedness by Prime Length was also significant [$\underline{F}_1(1,56)=4.2$, $\underline{p}=0.045$; $\underline{F}_2(1,19)=9.9$, $\underline{p}=0.005$]: the difference in RT between related test and unrelated control conditions is smaller in the partially primed (490–490= 0 ms) than in the fully primed condition (478–457= 21 ms). Other two-way interactions were not significant.

With increasing auditory information, visual targets that are related to the intended auditory prime are predicted to show priming effects, whereas visual targets that are either unrelated or related to the prime's competitor should show no such effects. This predicted three-way interaction between all fixed factors was in fact not significant [$\underline{F}_1(1,56)<1$, n.s. $\underline{F}_2(1,19)<1$, n.s.], although the expected pattern is clearly visible in Figure 1. The analyses of variance were also carried out separately for each level of the Candidate factor. The two-way interaction between Relatedness and Prime Length was significant in the sub-analysis for intended auditory primes [$F_1(1,56)=4.6$, $p=0.037$; $F_2(1,19)=6.0$, $p=0.024$], but it was not significant in the sub-analysis for the prime's competitor [$F_1(1,56)=1.3$, n.s.; $F_2(1,19)=1.6$, n.s.]. This pattern of results confirms the predictions outlined above.

The absence of partial priming is an important finding in this study, as will be discussed below. But does this absence of a partial priming effect reflect the true absence of this effect? Or is it perhaps due to low power in detecting such an effect? This latter possibility is quite improbable, for two reasons. First, let us inspect one very robust priming effect, viz. priming of the visual probe by a semantically related auditory prime that is fully audible. This priming effect is also present in our data (472-444= 28 ms). Post-hoc power analyses for this separate contrast indicate that the present experiment had ample power in detecting this contrast, viz. .973 (for ANOVA by subjects) and .988 (for ANOVA by items). Second, post-hoc power analyses indicated that the relevant two-way interaction of Relatedness and Prime Length (for the sub-analyses for intended auditory primes, see above) was detected with adequate power, viz. .558 (by subjects) and .644 (by items). Taken together, these power analyses indicate that our amendments in the design and in the stimulus materials by Zwitserlood (1989) have not reduced the power of our study. Power appears to be sufficient to detect partial priming if this had indeed been present.

2.6 Discussion

First, the results obtained in this experiment show reliable priming if the full auditory prime is presented. This agrees with previous research using the same task (Swinney, 1979; Chwilla, 1996), which in turn lends credibility to results of the present experiment. Second, our results clearly show that there is no priming when partial primes are presented.

Third, results show that RTs were shorter for full primes than for partial primes – not only for intended auditory primes, but for their competitors as well. This is not due to priming of the Competitor words, because RTs were similar for related test and unrelated control conditions. This decrease in RT as auditory information increases, can be explained as an effect of lexical competition in general; responses to any visual target can be faster if there is less competition between lexical candidates. A similar pattern of results was reported and explained by Mattys & Clark (2002) who employed a pause detection task. In their study, RTs for early-unique words were shorter than for late-unique words, for which lexical competition persists longer. Hence, the absence of a Candidate by Prime Length interaction effect in the present study, and the absence of the three-way interaction effect, are in agreement with their explanation. This pattern of results therefore suggests that there is indeed lexical competition in partial-priming conditions, even though this multiple activation does not produce detectable priming of the intended lexical candidate nor of its competitor.

The reported absence of partial priming agrees with several studies (cited in the Introduction) that did not find partial priming either, but it contradicts two other studies reporting partial priming effects. What are the possible reasons for this discrepancy in experimental results? If semantic priming effects are indeed too weak to be detected while multiple word candidates are active, as suggested by several experiments including the present, then a new question arises. Why was this multiple activation observed in two studies using partial auditory primes? Our answers are as follows. One study (Moss, McCormick, & Tyler, 1997) was somewhat similar to the present study in experimental design, but competitor candidates were not investigated directly. This property renders that study irrelevant to address the issue of multiple activation of word candidates. Second, the other study reporting partial priming effects (Zwitserlood, 1989) suffered from an incomplete between-subjects design (Cochran & Cox, 1957), which was necessary because of its mis-proportion of number of conditions (32) and number of items (24). In such a design, an interaction effect between listeners and conditions cannot be separated from the main effect of conditions (Cox, 1958, Chapter 11; Bailey, 1982). To rely on the assumption that any effect of interest would be equal across all participants is in fact quite dangerous. Participants vary in their cognitive behaviour, and these differences obviously persist when they process spoken or written language (Connine, Blasko, & Wang, 1994; Plaut & Booth, 2000). Likewise, it is entirely conceivable that priming effects only show up in some listeners, perhaps the ‘fast’ ones or the verbally proficient, and not in others. These inter-listener differences amount to an interaction between listeners and conditions. In an incomplete between-subjects design, such interactions between listeners and conditions are pooled with the main effect of conditions, thus inflating the significance of the latter. Also note that comparisons between conditions must of necessity assume sphericity, that is, differences among conditions are assumed to have equal variances across listeners. This assumption is nowadays often regarded as dangerous and not warranted (O’Brien & Kaiser, 1985; Max & Onghena, 1999).

Could this property of experimental design indeed be responsible for the positive evidence reported by Zwitserlood (1989)? Even though the original data of that study are no longer available for re-investigation, it is still possible to address this question by means of Monte Carlo simulations, as illustrated in the following section.

3. Monte Carlo simulations

In a so-called Monte Carlo simulation (Hammersley & Handscomb, 1964), a data set from an imaginary experiment is generated at random, with statistical properties programmed in the simulation. The appropriate test statistic, e.g. an F ratio, is calculated from each simulated data set. This process is then repeated a large number of times. For example, we could generate many sets of RT data, in which there are no “true” differences between related test and unrelated control conditions (i.e. in accordance with H_0 which claims that priming effects are absent). In realistic simulations, of course, there are also random variance components associated with items, with listeners, and perhaps with other random sources of variance.

In our simulations, each data set was generated as follows. First, the difference between auditory prime words and their competitors was ignored; all simulations were run for actually intended prime words only. This yields 2 x 2 treatment conditions, defined by the Relatedness and Prime Length factors. All four treatment effects were set to zero. Each observation was generated by adding an arbitrary grand mean, plus the treatment effect, plus several random effects. Values for these random effects were drawn from separate gaussian distributions having mean zero, and having the following standard deviations: experimental lists (or listener group) 30 ms, listeners 100 ms, items 100 ms, and within-cell observations 100 ms. These standard deviations correspond roughly to the random variance components observed in other cross-modal priming studies in our laboratory.

--- INSERT TABLE 2 ABOUT HERE ---

For each observation, the appropriate values of these random effects were used, in order to simulate the actual design of the experiment presented above. Crucially, RT observations were generated both with and without the controversial listener-by-condition interaction as an additional random effect. This random variance component was drawn from a gaussian distribution, with mean zero and standard deviation 75 ms. The latter is a conservative estimate, based on the observed variance due to the corresponding interaction of Relatedness by Prime Length by Experimental List in the experiment above (MS=8007, s=89 ms). Each data set, or simulated experiment, consisted of 24 test items, and 4 groups of 15 listeners each, as in the actual experiment. For each design (with and without interaction), 5000 simulations were performed.

Each of these data sets was fed into two repeated measures univariate ANOVAs (over listeners aggregated over listener groups, and over items). Results of these Monte Carlo simulations take the form of 5000 F_1 ratios (by listeners) and 5000 F_2 ratios (by items) corresponding to the two-way interaction of Relatedness by Prime Length. Most relevant for our purposes is the proportion of these F ratios exceeding the appropriate critical F value. This corresponds to the probability of a ‘positive outcome’, i.e. of rejecting H_0, and of concluding that a priming effect exists. Since a priming effect is known to be absent in these simulations: rejecting H_0 amounts to a Type I error here.

Figures 2 and 3 give the distributions of $\underline{F}_1$ and $\underline{F}_2$ of this interaction, respectively, along with this probability of a positive outcome.

--- INSERT FIGURE 2 HERE ---

--- INSERT FIGURE 3 HERE ---

The Monte Carlo results for $\underline{F}_1$ and $\underline{F}_2$ show similar tendencies, and will be discussed together here. First, we see that if an interaction between listeners and conditions is present, then the chance of a Type I error is inflated somewhat (right) relative to the no-interaction case (left), even in this within-subject design. Note that there is a very low probability of Type I error, i.e. of finding spuriously significant effects, in case the priming effect is indeed absent.

Next, let us compare these findings with Monte Carlo simulations of the original experiment by Zwitserlood (1989). To this end, we changed the experimental design in the simulations from a complete within-subjects design to an incomplete between-subject design. In the original study, each listener participated in 24 out of 32 conditions. In our simulation of that study, each listener participated in 3 out of 4 conditions, corresponding to 3 out of 4 treatment conditions as defined above. Listeners and items were rotated evenly across conditions. Four listener groups were used, each consisting of 6 listeners (for further design details, see Zwitserlood, 1989). Again, the candidate factor (actual prime word vs. competitor) was ignored for practical purposes. Further details of these Monte Carlo simulations were identical to those above. The incomplete between-subjects design precludes any ANOVA by listeners, because a single listener only participated in 3 out of 4 treatment conditions. This Monte Carlo simulation therefore only yields F_2 ratios, which are summarised in Figure 4 below.

--- INSERT FIGURE 4 HERE ---

First, we see that a Type I error is indeed highly probable, if an interaction component between listeners and conditions is forced to be present in the data sets (right panel). If listeners vary in their susceptibility to priming, then this would indeed increase the chance of finding a ‘significant’ priming effect. The priming effect, although absent in these simulated data, is incorrectly reported to be ‘significant’ in about half of these simulations. This finding raises serious doubts about the validity of the simulated experiment. Second, we see that even without this disputed interaction in the data (left panel), the chance of finding a spurious significant effect, a Type I error, is dangerously high. This itself indicates that the reported priming effect may well have been spurious.

Zwitserlood (1989) also realised that listeners and conditions were confounded in her incomplete design. Differences among test conditions are ‘contaminated’ by differences among listeners’ averages. A normalisation procedure removes this contamination part, but analysis of the resulting data still requires the assumption that there is no interaction between conditions and listeners. The listener’s average RT was subtracted from each observation (and then the grand mean was added). The resulting normalised RTs are still contributed by different listeners in different conditions. Thus, individual priming differences are still likely to exist.

In order to investigate the effects of this normalisation procedure, the Monte Carlo simulations of the original experiment were repeated, with Zwitserlood’s normalisation procedure inserted after random generation of the data sets, before statistical analysis. In all other respects, the latter simulations were equal to those performed on the raw RT data. The resulting F_2 ratios are summarised in Figure 5 below.

--- INSERT FIGURE 5 HERE ---

First, we see that the presence of interaction between listeners and conditions in the normalised data inflates the probability of a Type I error to 62% (right panel). This is the probability of reporting a “positive outcome”, a significant priming effect, if in fact such an effect is forced to be absent in the data set, and if individual differences in priming susceptibility are forced to vary among listeners. Obviously, this high probability of a Type I error raises strong doubts about the validity of the experiment that is simulated here. As before, even without this disputed interaction component in the data (left panel), the chance of finding a spurious significant effect, a Type I error, is apparently inflated by the normalisation procedure.

In summary, these Monte Carlo simulations suggest that the original experiment by Zwitserlood (1989) does not warrant valid and reliable conclusions.

4. General Discussion and Conclusion

As argued in the Introduction, an absence of reliable partial semantic priming effects does not imply that multiple word candidates have not been activated. Although the results from the present experiment did not show multiple lexical activation, there is a large body of evidence supporting the concept of early multiple activation of lexical candidates (with subsequent competition among these candidates). This evidence has been collected using various experimental tasks: phonological priming (Slowiaczek, McQueen, Soltano, & Lynch, 2000), word identification (Luce, Pisoni, & Goldinger, 1990), word spotting (Cutler & Norris, 1988; Norris, McQueen, & Cutler, 1995), phoneme classification

(Borsky, Tuller, & Shapiro, 1998), phoneme monitoring (Gaskell & Marslen-Wilson, 1998; Vroomen & de Gelder, 1999), tracking of eye movement (Dahan, Magnuson, Tanenhaus, & Hogan, 2001), and pause detection (Mattys & Clark, 2002). Thus, multiple word candidates are indeed activate in the early stages of spoken-word recognition, but the cross-modal semantic priming technique is not suitable for tapping into these multiple activations.

Spreading of activation to semantically related items has always been seen as an automatic process: it is fast, of short duration, does not require attentional awareness, and presupposes no or only minimal demands on resource capacity (cf. Posner & Snyder, 1975). It seems therefore that spreading of activation is not immediate. The activation of a word candidate may have to surpass a certain threshold level before it passes on to semantically related items, or before it has detectably affected the activation of semantic associates. The amounts of activation of the to-be-recognised word and its semantic relatives need not be equal: perhaps only a certain proportion of the intended word's activation spreads to its relatives. Therefore, as long as multiple candidates are activated, activation of their relatives may not be high enough to be reliably detected. When one candidate remains, only then activation of that candidate will be high enough to spread a detectable amount of activation to its semantic relatives.

Within the Distributed Cohort model (Gaskell & Marslen-Wilson, 1997; 1999; 2002), semantic priming is explained in a somewhat different way. In a connectionist network, multiple representations must interfere with each other if they are active simultaneously. The semantic representations of the activated word candidates are often not coherent. As long as multiple candidates are active, the lack of overlap in the semantic nodes translates into small or inconsistent semantic priming effects, if any (cf. Gaskell & Marslen-Wilson, 2002). Perhaps future research can shed more light on the way activation spreads to related items. At this point it is impossible to judge between the spreading of activation account originally put forward by Collins and Loftus (1975) and the DC model. It is crucially important, however, that both theoretical accounts predict that tapping into multiple activation via semantic priming is inherently difficult because the effects are small.

This prediction is verified, not only by the results of our experiment, but also by several other failures to find partial-priming effects (Chwilla, 1996; Jongenburger, 1996; as well as several unpublished studies).

A recent article has shed more light on the psychological reality of the recognition point in spoken-word processing. Whereas the Cohort model (Marslen-Wilson, 1993), proposes a recognition point as the point where the word diverges from the other members of its word-initial cohort, the Shortlist model (Norris, 1994) does not make predictions as to when words presented in isolation will be recognised. The uni-modal repetition priming study by Bölte & Uhe (in press) investigated the influence of sensory information following the recognition point of the prime. The recognition point was established in a gating experiment. Repetition priming effects were then studied at the recognition (RP cut-off) point, at a cut-off point after the recognition point (RP-plus) and following the presentation of the complete prime. Bölte & Uhe found that the priming effect at the RP-plus condition was somewhat larger than at the recognition point (although insignificantly), but that the priming effect in the Complete condition was significantly larger than at the two cut-off conditions. These results provide counterevidence against a strong formulation of the recognition point, in which lexical activation does not increase any further from the recognition point onwards. Bölte & Uhe therefore argue that the recognition point is the "moment at which the word recognition system makes a commitment to a certain lexical representation. Further information is used (1) to distinguish between, for instance, morphological alternatives and (2) to raise the lexical activation of matching lexical representations rather gradually. Still, a word is not selected at this moment." (Bölte & Uhe, -in press-, p. 13). Importantly, this shows that at the recognition point, even the phonological priming effect has not reached its maximum. It is therefore no surprise that semantic priming effects show up only after the recognition point.

In conclusion, we have employed Monte Carlo simulations to investigate the chances of Type I and Type II errors in one key study (the data of which were no longer available). These simulations indicate that this key study had a very high chance of finding spurious effects of partial semantic priming (i.e., of a Type I error). Partial priming effects were not observed in the experiment reported here, nor in several similar studies.

However, there is overwhelming experimental evidence that multiple word candidates are activated during spoken-word recognition. But we have to conclude that the cross-modal semantic priming technique does not provide valid and reliable insight into this multiple activation. It is therefore advisable to discontinue its use for that purpose.

References

Bailey, R. A. (1982). Confounding. In S. Kotz (Ed.), Encyclopedia of Statistical Sciences (Vol. 2, pp. 128-134). New York: Wiley.

Bölte, J. & Uhe, M. (in press). When is all understood and done? The psychological reality of the recognition point. Brain and Language.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat": The effect of acoustic and semantic information on phoneme categorization. J. Acoustical Society of America, 103, 2670-2676.

Campbell, D. T. (1969). Prospective: Artifact and control. In R. L. Rosnow (Ed.), Artifact in Behavioral Research (pp. 351-382). New York: Academic Press.

CELEX. (1990). CELEX Dutch Database [electronic database]. CELEX, Dutch Centre for Lexical Information, Nijmegen, The Netherlands. Retrieved 13 June, 2001, from the World Wide Web: http://www.kun.nl/celex/ and http://morph.ldc.upenn.edu/Catalog/

Chwilla, D. J. (1996). Electrophysiology of word processing: the lexical processing nature of the N400 priming effect. Unpublished doctoral dissertation, University of Nijmegen, Nijmegen.

Cochran, W. G., & Cox, G. M. (1957). Experimental Designs. New York: Wiley.

Collins, A. M., & Loftus, E. F. (1975). A spreading activation theory of semantic processing. Psychological review, 83, 407-428.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. Perception and Psychophysics, 56, 624-636.

Cox, D. R. (1958). Planning of Experiments. New York: Wiley.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. J. Experimental Psychology: Human Perception and Performance, 14, 113-121.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes, 16, 507-534.

Forster, K.I. (1976). Accessing the mental lexicon. In: R.J. Wales, & E.C.T. Walker (eds.), New Approaches to Language mechanisms: a cross section of of psycholinguistic studies. Amsterdam: North Holland. North Holland Linguistics series 30, 257-287.

Forster, K.I. (1979). Levels of processing and the structure of the language processor. In: W.E. Cooper, & E.C.T. Walker (eds.). Sentence processing: Psycholinguistic studies presented to Merill Garrett. Cambridge, MA: MI Press, 27-85.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. J. Experimental Psychology: Human Perception and Performance, 22, 144-158.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: a distributed model of speech perception. Language and Cognitive Processes, 12 (5/6), 613-656.

Gaskell, M. G., & Marslen-Wilson, W. D. (1998). Mechanisms of phonological inference in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 24(2), 380-396.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition and blending in spoken word recognition. Cognitive Science, 23(4), 439-462.

Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. Cognitive Psychology, 45, 220-266.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. Perception and Psychophysics, 28, 267-283.

Hammersley, J. M., & Handscomb, D. C. (1964). Monte Carlo Methods. London: Methuen.

Jongenburger, W. (1996). The role of lexical stress during spoken word processing. Unpublished doctoral dissertation, Leiden University, Leiden.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), Cognitive Models of Speech Processing. Cambridge, MA: MIT Press.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken-language understanding. Cognition, 8(1-71).

Mattys, S.L., & Clark, J.H. (2002). Lexical activity in speech processing: evidence from pause detection. Journal of Memory and Language, 47, 343-359.

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. Jounal of Speech, Language and Hearing Research, 42, 261-270.

Maxwell, S.E., & Delaney, H.D. (2004). Designing experiments and analyzing data: A model comparison perspective (2nd ed.), Mahwah, NJ: Lawrence Erlbaum Associates.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1-86.

Meyer, D., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. Journal of Experimental Psychology, 90, 227-234.

Moss, H. E., McCormick, S. F., & Tyler, L. K. (1997). The time course of activation of semantic information during spoken word recognition. Language and Cognitive Processes, 12(5/6), 695-731.

Neely, J.H. (1977). Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading of activation and limited-capacity attention. Journal of Experimental Psychology: General, 106, 226-254.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. Cognition, 52, 189-234.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and Segmentation in Spoken-Word Recognition. Journal of Experimental Psychology: Learning Memory and Cognition, 21(5), 1209-1228.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. Behavioral and Brain Sciences, 23(3), 299-370.

O'Brien, R.G., & Kaiser, M.K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. Psychological Bulletin, 97(2), 316-333.

Plaut, D. C., & Booth, J. R. (2000). Individual and developmental differences in semantic priming: Empirical and computational support for a single-mechanism account of lexical processing. Psychological Review, 107(4), 786-823.

Slowiaczek, L. M., McQueen, J. M., Soltano, E. G., & Lynch, M. (2000). Phonological representations in prelexical speech processing: Evidence from form-based priming. Journal of Memory and Language, 43(3), 530-560.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. Journal of Verbal Learning and Verbal Behaviour, 18, 645-659.

Vroomen, J., & de Gelder, B. (1999). Lexical access of resyllabified words: Evidence from phoneme monitoring. Memory & Cognition, 27(3), 413-421.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. Cognition, 32, 25-64.

Zwitserlood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken word recognition. Language and Cognitive Processes, 10, 121-136.

Appendix I

List of stimulus materials: actual prime word and competitor (in lowercase) serving as auditory primes, with their respective semantically related words (in uppercase) serving as visual targets, and auditory control word (quasi-prime, in lowercase).

nr	auditoryprime		visual target		auditory control
	actual prime	competitor	actual prime's associate	competitor's associate	
1	klanten 'customers'	klappen 'slaps'	WINKEL 'shop'	SLAAN 'smack'	bergen 'mountains'
2	schapen 'sheep'	schaar 'scissors'	WOL 'wool'	KNIPPEN 'cut'	fruit 'fruit'
3	herberg 'inn'	hert 'deer'	SLAPEN 'sleep'	BOS 'forest'	pink 'little finger'
4	haven 'harbour'	hamer 'hammer'	BOOT 'boat'	SPIJKER 'nail'	beker 'cup'
5	krant 'newspaper'	kramp 'cramp'	NIEUWS 'news'	PIJN 'pain'	plank 'shelf'
6	bloed 'blood'	bloem 'flower'	ROOD 'red'	GEUR 'smell'	trein 'train'

7	kaas	kaarsen	BROOD	LICHT	heup
	’cheese’	’candles’	’bread’	’light’	’hip’
8	salaris	salami	GELD	WORST	piano
	’salary’	’salami’	’money’	’sausage’	piano’
9	orkaan	orkest	WIND	MUZIEK	beton
	’hurricane’	’orchestra’	’wind’	’music’	’concrete’
10	pilaren	piloten	KERK	VLIEGTUIG	seizoen
	’pillars’	’pilots’	’church’	’plane’	’season’
11	koffer	koffie	REIS	THEE	titel
	’suitcase’	’coffee’	’trip’	’tea’	’title’
12	dorst	dorp	DRINKEN	KLEIN	gang
	’thirst’	’village’	’drink’	’small’	’corridor’
13	knopen	knoken	GAT	BOT	verf
	’buttons’	’knuckles’	’hole’	’bone’	’paint’
14	salaris	salami	ZEE	RIJK	sigaret
	’captain’	’capital’	’sea’	’rich’	’cigarette’
15	kraan	kraag	WATER	JAS	trap
	’tap’	’collar’	’water’	’coat’	’stairs’
16	vrede	vrees	OORLOG	ANGST	proef
	’peace’	’fear’	’war’	’fear’	’test’
17	been	beest	LOPEN	DIER	kleur
	’leg’	’animal’	’walk’	’animal’	’colour’
18	kozijnen	kozakken	RAAM	RUS	dolfijnen
	’frame’	’cossacks’	’window’	’russian’	’dolphins’
19	vlammen	vlaggen	BRAND	WIMPEL	blokken
	’flames’	’flags’	’fire’	’banner’	’blocks’
20	draak	draad	VUUR	NAALD	plein

	’dragon’	’thread’	’fire’	’needle’	’square’
21	scherven	schelpen	GLAS	STRAND	krukken
	’splinters’	’shells’	’glass’	’beach’	’stools’
22	stallen	stad	PAARD	DRUK	douche
	’stables’	’city’	’horse’	’busy’	’shower’
23	broek	broer	PIJP	ZUS	gras
	’pants’	’brother’	’trouser leg’	’sister’	’grass’
24	spelden	spek	NAAIEN	VARKEN	knal
	’pins’	’bacon’	’sew’	’pig’	’crack’

Authors' Note

Esther Janse, Hugo Quené, Utrecht institute of Linguistics, Utrecht University, The Netherlands.

Portions of this research were presented at a workshop on Spoken Word Access Processes (SWAP), 29-31 May 2000, Max Planck Institute of Psycholinguistics, Nijmegen, The Netherlands. We thank Sieb Nooteboom, Pienie Zwitserlood, Frank Wijnen, Saskia te Riele, Willy Jongenburger, Anne Cutler, James McQueen, Arthur Samuel and Huub van den Bergh, for valuable comments and suggestions, and Theo Veenker for technical assistance.

Correspondence concerning this article should be addressed to Esther Janse, Utrecht institute of Linguistics OTS, Utrecht University, Trans 10, NL-3512 JK Utrecht, The Netherlands, e-mail <esther.janse@let.uu.nl>.

Tables

Table 1. Schematic design of experiment.

Prime Type	Candidate	Length	Auditory prime [a]	Visual target	Listener group
Test	Actual	partial	sala-	GELD ‘money’	1
Control	Actual	partial	pia-	GELD	3
Test	Actual	full	salaris ‘salary’	GELD	2
Control	Actual	full	piano ‘piano’	GELD	4
Test	Competitor	partial	sala-	WORST ‘sausage’	3
Control	Competitor	partial	pia-	WORST	1
Test	Competitor	full	salaris	WORST	4
Control	Competitor	full	piano	WORST	2

a The auditory prime was always preceded by the Dutch carrier phrase Het volgende woord is... (‘The next word is...’).

Table 2.

Variance components (in standard deviations, in ms) for various random effects in the Monte Carlo simulations of the present experiment, in which the main effect of Prime Type was absent. Simulations were performed with and without the interaction component printed in italics.

experimental list (or listener group)	30
listener	100
item	100
unique	100
listener × condition	75

Figures

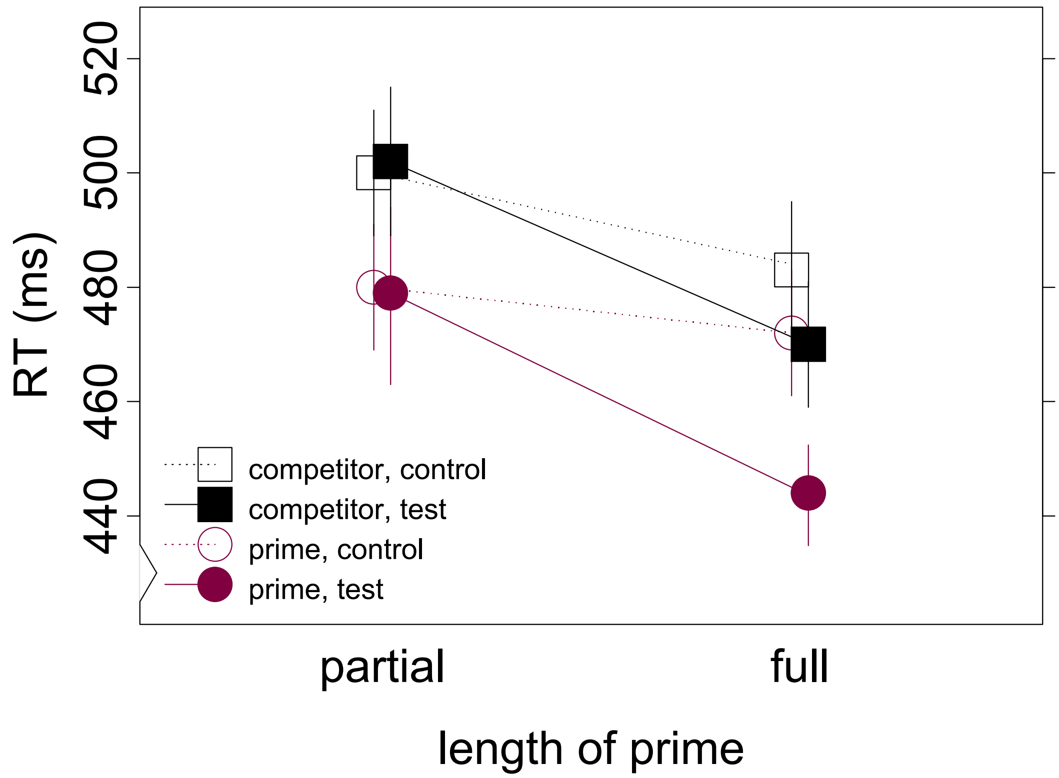

Figure 1.

Average lexical decision times (in ms) obtained in the present experiment, with associated 95% confidence intervals, broken down by related test and unrelated control items (filled and open symbols, respectively), by prime words and competitor words (Candidate factor; circles and squares, respectively), and by partial primes and full primes (Prime length factor; on the abscissa).

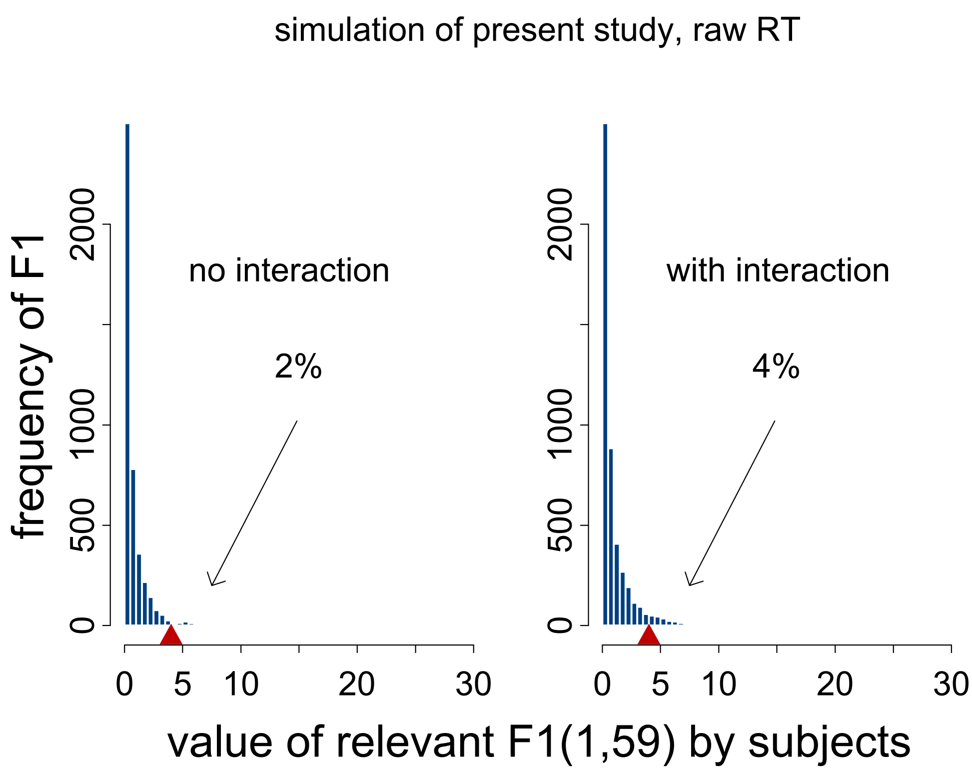

Figure 2.

Distributions of $\underline{F}_1$ ratios (with 1 and 59 degrees of freedom), with each distribution obtained from 5000 simulated data sets. Distributions are given for simulations with a listener by condition <u>interaction</u> effect either absent (left) or present (right) in the data sets. The critical <u>F</u> value for α=.05 is marked along the abscissa, and the proportion of 'positive outcomes' (<u>F</u> exceeding its critical value) is given with each distribution.

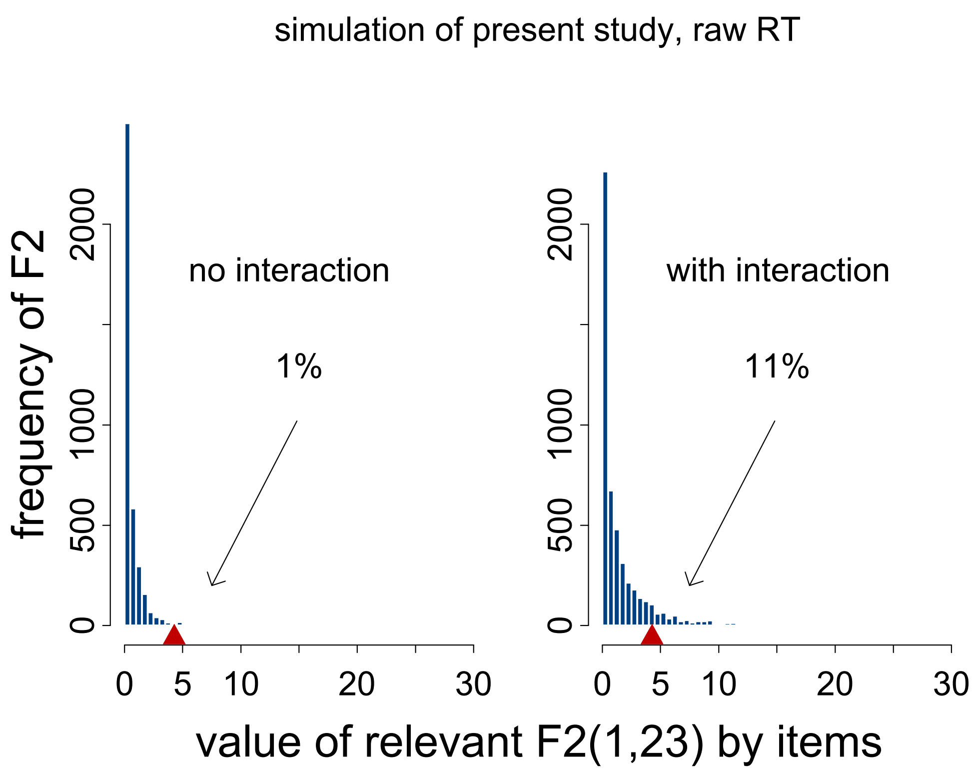

Figure 3.

Distributions of $\underline{F}_2$ ratios (with 1 and 23 degrees of freedom), with each distribution obtained from 5000 simulated data sets. Distributions are given for simulations with a listener by condition interaction effect either absent (left) or present (right) in the data sets. The critical $\underline{F}$ value for $\alpha=.05$ is marked along the abscissa, and the proportion of 'positive outcomes' ($\underline{F}$ exceeding its critical value) is given with each distribution.

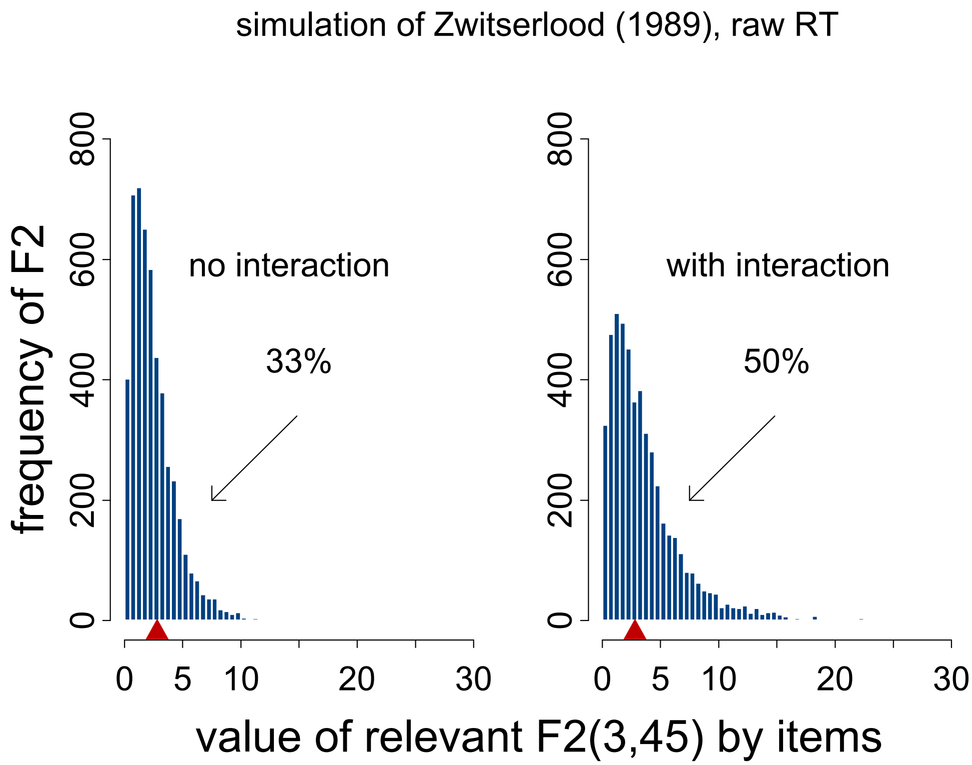

Figure 4.

Distributions of $\underline{F}_2$ ratios (with 3 and 45 degrees of freedom), with each distribution obtained from 5000 simulated data sets. Distributions are given for simulations with a listener by condition interaction effect either absent (left) or present (right) in the data sets. The critical value for α=.05 is marked along the abscissa, and the proportion of 'positive outcomes' ($\underline{F}$ exceeding its critical value) is given with each distribution.

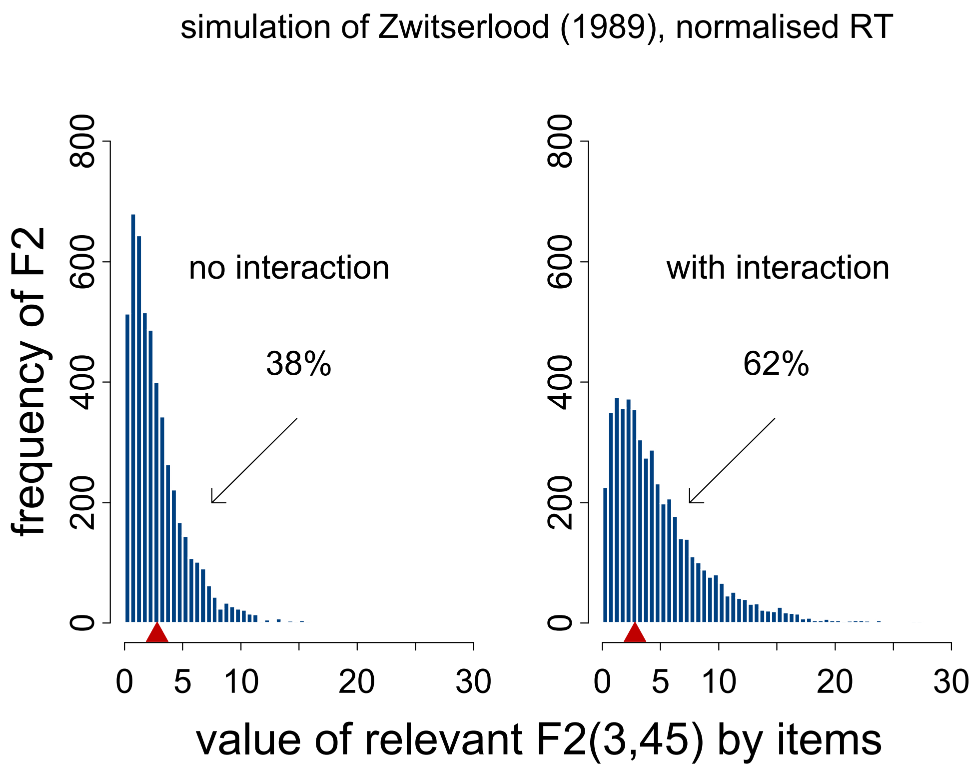

Figure 5.

Distributions of $\underline{F}_2$ ratios (with 3 and 45 degrees of freedom), with each distribution obtained from 5000 simulated data sets, after normalisation of each data set according to Zwitserlood (1989). Distributions are given for simulations with a listener by condition interaction effect either absent (left) or present (right) in the data sets. The critical $\underline{F}$ value for α=.05 is marked along the abscissa, and the proportion of ‘positive outcomes’ ($\underline{F}$ exceeding its critical value) is given with each distribution.