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

Esther Janse and Hugo Quené

Utrecht University

Abstract

Cross-modal semantic priming with partial auditory primes seems a good technique to assess spoken-word recognition, because it allows tracking 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. 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.

1 Introduction

Among current models of spoken-word recognition, there is remarkable agreement about one basic stage or process, namely that of initial multiple activation of word candidates in the listeners' mental lexicon that roughly match the available auditory input. In a later stage, activated candidates compete and each candidate's activation 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 comes from several studies that employed the cross-modal semantic priming paradigm. This paradigm 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 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*), then they react faster if the visual target word and the auditory prime word are semantically or associatively related, than if these words are not related. In *partial priming*, the auditory prime words are 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. At that cut-off point, multiple word candidates are supposed to be still active. By presenting a visual target related to one of these candidates, immediately following the prime fragment's offset, one can measure the activation of that word candidate. This implies that, 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 (cf. Forster, 1976; 1979; McClelland & Elman, 1986). 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., *sala*-…), activation was found for both compatible word candidates *salaris* ('salary') and *salami* ('salami')[1]. Second, sentence context affects 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 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 (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 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 competing with the actual prime word. Consequently, these studies bear only little evidence on *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 the word under consideration. Activation of a word candidate is then inversely related to the distance between the output of the network (a point in the multidimensional space), and the word representation in this space. In connectionist models, 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): lateral inhibition between activated word candidates is employed to reduce multiple activations. Before the uniqueness point, semantic activation depends strongly 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 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

[1] Priming effects were found for both related visual targets, GELD ('MONEY') and WORST ('SAUSAGE').

(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 if 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. While multiple candidates are active, their disparate semantic properties provide only weak semantic priming of other words, if any. 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 present study therefore attempts to clarify this discrepancy among semantic priming studies, 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. Due to space limitation, the design and results of this replication experiment will be discussed only very briefly. 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.

2 Experiment

Since Zwitserlood's sentences and test words have yielded the clearest partial priming effects to date, her (Dutch) materials were used here whenever possible. The main difference here is in our experimental design. The present experiment ignores the sentence context factor, such that a within-subjects design is possible. For a more detailed description of the materials, design, procedure, and for more detail on the results, the reader is referred to Janse (2003).

The results of the replication experiment can be summarised as follows: the presentation of the partial prime fragment (prime cut off at isolation point) did not yield semantic facilitation, not for the target related to the actual prime word, nor for the target related to the competitor. The presentation of the full actual prime word yielded a significant 32 ms priming effect for the prime word's related target. The reaction time data were fed into repeated measures

ANOVAs, with Relatedness (related vs. unrelated/control condition), Candidate (target related to either prime word or closest competitor), and Prime Length (partial/full) as fixed factors. With increasing auditory information, visual targets 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 three-way interaction was in fact not significant [$F_1(1,59)=2.8$, n.s.; $F_2(1,23)=2.8$, n.s.]. The ANOVAs were also carried out for Actual Primes and Competitors separately. The Relatedness × Prime Length interaction was significant in the sub-analysis for intended auditory primes [$F_1(1,59)=9.3$, $p=0.003$; $F_2(1,23)=9.4$, $p=0.005$], but not for the prime's competitor [$F_1(1,59)<1$; $F_2(1,23)<1$].

The absence of partial priming is an important finding in this study. But could this perhaps be due to low power in detecting such an effect? This is quite improbable, for two reasons. First, let us inspect one very robust priming effect, viz. priming of the visual target by a semantically related auditory prime that is fully audible. This priming effect is also present in our data (32 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 Candidate × 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 analyses indicate that our amendments in design and stimulus materials by Zwitserlood (1989) have not reduced the power of our study.

3 Discussion

First, reliable semantic priming is observed in this experiment, if full auditory primes are presented. This agrees with previous research (Swinney, 1979; Chwilla, 1996), which lends credibility to the present results. Second, no priming was observed when partial primes are presented. Third, RTs were shorter for full primes than for partial primes: not only for intended auditory primes, but also for their competitors, *and* for unrelated control conditions as well. Hence, this decrease in RT as auditory information increases is not due to priming, but can be explained as an effect of lexical competition in general; responses to any target can be faster if there is less competition between lexical candidates. A similar pattern of results was found by Mattys & Clark (2002) using 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 Relatedness × Prime Length interaction, and of the three-way interaction effect, are in agreement with their explanation. This pattern of results suggests that there is indeed lexical competition in 'partial' conditions, even though this does not lead to semantic priming effects.

But how can we explain the discrepancy between these replication results and those of the original study? Our answer is that the original study (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 participants is in fact quite dangerous. Participants vary in their cognitive behaviour, and these differences obviously persist when they process spoken or written language (e.g. Connine, Blasko, & Wang, 1994; Plaut & Booth, 2000). Thus, it is entirely conceivable that priming effects show up in *some* listeners, perhaps the 'fast' or verbally proficient ones, but 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). Even though the original Zwitserlood (1989) data are no longer available for investigation, it is possible to address this design issue by means of Monte Carlo simulations.

4 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 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 × 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.

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 *SD* 75 ms. The latter is a conservative estimate, based on the observed variance due to the corresponding interaction of Relatedness × Prime Length × 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. 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 × 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, rejecting H_0 amounts to a Type I error here. Figures 1 and 2 give the distributions of F_1 and F_2 of this interaction, respectively, along with this probability of a positive outcome. The Monte Carlo results for F_1 and F_2 show similar tendencies, and will be discussed together.

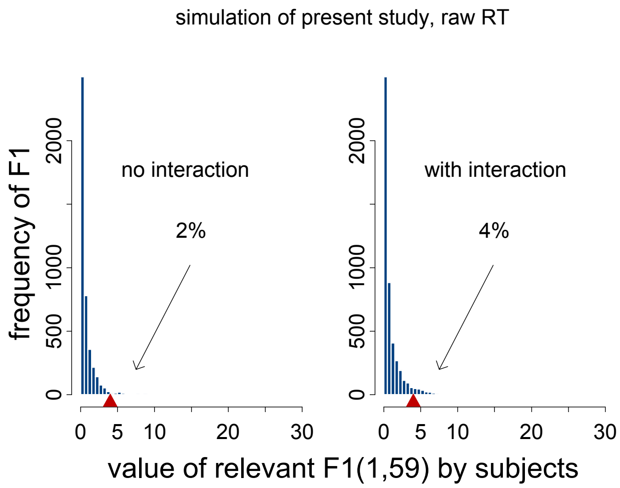

Figure 1. Distributions of 5000 F_1 ratios. Simulations were done with a listener × condition interaction effect either absent (left) or present (right) in the data sets. The critical F_1 value (α=.05) is marked along the abscissa, and the percentage of 'positive outcomes' is given with each distribution.

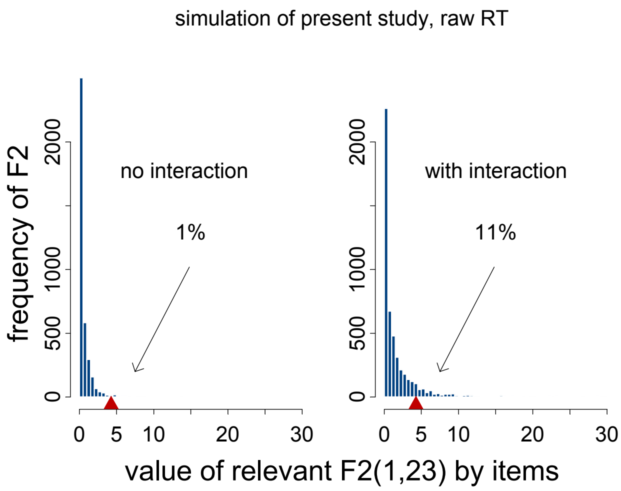

Figure 2. Distributions of 5000 F_2 ratios.

First, 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 (cf. Zwitserlood, 1989). Again, the Candidate factor (actual prime 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 given in Figure 3 below.

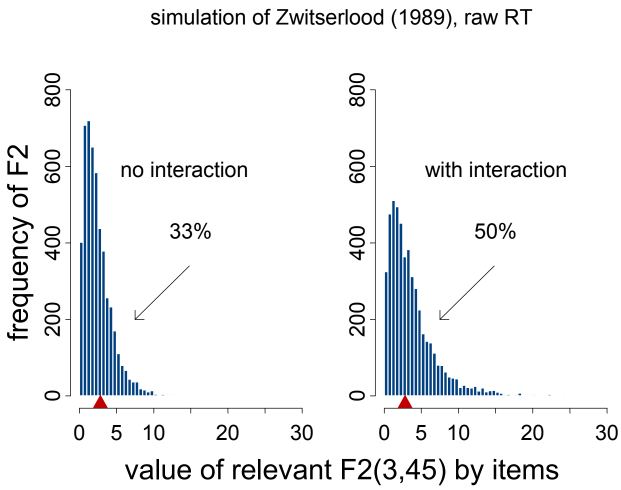

Figure 3. Distributions of 5000 F_2 ratios.

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

Zwitserlood (1989) also realised that listeners and conditions were confounded in her 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 effect of this normalisation, 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 simulations were equal to those on the raw RT data. The resulting F_2 ratios are summarised in Figure 4.

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 significant priming effect, if in fact such an effect is forced to be absent, 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.

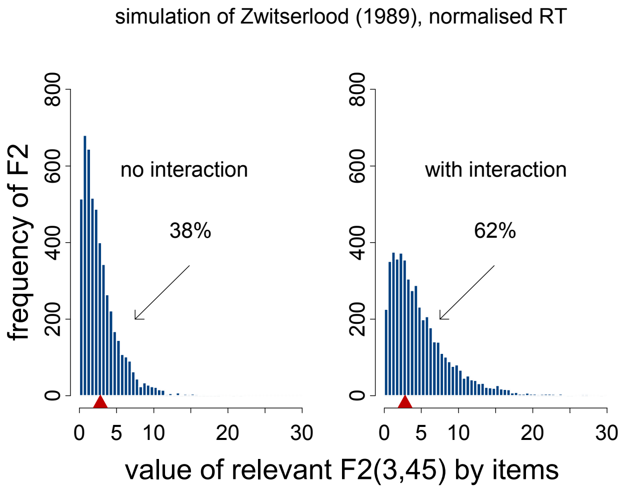

Figure 4. Distributions of 5000 F_2 ratios, after normalisation of each data set.

5 Discussion and conclusion

An absence of reliable partial semantic priming effects does not imply that multiple word candidates have not been activated. 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, spoken-word processing indeed involves activation of multiple word candidates, but cross-modal semantic priming is not suitable for tapping into these multiple activations.

It seems that 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. Only when one candidate remains, activation of that candidate will be high enough to spread activation detectably to its semantic relatives.

Within the DC model (Gaskell & Marslen-Wilson, 1997, 1999, 2002), as in all connectionist networks, multiple representations must interfere with each other if they are active simultaneously. 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 how activation spreads to related items. At this point it is impossible to choose between the original spreading-of-activation account (Collins & Loftus, 1975) and the DC model. However, both accounts predict that tapping into multiple activation via semantic priming is inherently difficult because the effects are small. This prediction is verified by several other failures to find partial-priming effects (Chwilla, 1996; Jongenburger, 1996).

A recent article has investigated the psychological reality of the recognition point in spoken-word processing. 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) however does not predict when a word presented in isolation will be recognised. The uni-modal repetition priming study by Bölte & Uhe (2004) investigated the influence of sensory information following the recognition point of the prime. Repetition priming effects were studied at the recognition (RP cut-off) point, at a later cut-off point (RP-plus), and at the offset of the complete prime. Bölte & Uhe found that the priming effect at the RP-plus condition was slightly larger than at the recognition point, but that the priming effect in the complete-prime 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 (2004: 145) 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." Importantly, 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. These simulations indicate that this study had a high chance of finding spurious effects of partial priming (of a Type I error). Partial priming effects were not observed in the replication experiment, nor in several similar studies. However, there is overwhelming experimental evidence that multiple word candidates are activated during spoken-word recognition. 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. (2004). When is all understood and done? The psychological reality of the recognition point. *Brain and Language*, *88* (1), 133-147.

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.

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* (pp. 257-287). Amsterdam: North Holland. North Holland Linguistics series; 30.

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* (pp. 27-85). Cambridge, MA: MIT Press.
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. *J. 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.
Hammersley, J. M., & Handscomb, D. C. (1964). *Monte Carlo Methods*. London: Methuen.
Janse, E. (2003). *Production and perception of fast speech*. Doctoral dissertation, Utrecht University, Utrecht, pp.175-197.
Jongenburger, W. (1996). *The role of lexical stress during spoken word processing*. Doctoral dissertation, Universiteit 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. *J. 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. *J. 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. *J. 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. *J. 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. *J. 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. *J. Memory and Language*, *43* (3), 530-560.
Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *J. 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.