

Word- vs. sentence-based simulation effects in language comprehension

Barbara Kaup, Jana Lüdtke, & Ilona Steiner

1. Introduction¹

In the literature on language comprehension many authors nowadays assume that comprehenders understand language by mentally simulating the described objects, events and situations. These simulations are assumed to be experiential in nature as they are grounded in perception and action (Barsalou, 2008; Glenberg & Kaschak, 2002; Zwaan, 2004). More specifically, according to this simulation view of language comprehension, each interaction with the world leaves experiential traces in the brain. These traces are partially re-activated when people read or hear words referring to the respective entities. If words appear in larger phrases or sentences, the activated traces are presumably combined to yield simulations consistent with the meaning of the larger phrase or sentence (Zwaan & Madden 2005). There is a steadily growing body of evidence for this view. On the one hand there are neuroscience studies indicating a considerable overlap between the mental subsystems utilized in representing linguistically specified states of affairs and those utilized in direct experience. For instance, studies using brain imaging techniques have shown that the processing of linguistic materials referring to actions that are typically performed with certain effectors (e.g., *to lick*, *to kick*, *to grasp*) activates those sections of the premotor and motor cortex that are specific for actions with the respective effector (Hauk, Johnsrude, & Pulvermüller, 2004; Tettamanti et al., 2005). Similarly, studies using transcranial magnetic stimulation have found that motor evoked potentials recorded from hand and foot muscles are specifically modulated by listening to hand-action-related vs. foot-action-related sentences, respectively (e.g., *He threw/ kicked the ball*; Buccino et al., 2005; Glenberg et al., 2008).

Kaup, B., Lüdtke, J., & Steiner, I. (2012). Word- vs. sentence-based simulation effects in language comprehension. In B. Stolterfoht & S. Featherston (Eds.), *Empirical Approaches to Linguistic Theory: Studies in Meaning and Structure*. Berlin: Berlin: de Gruyter.

In addition, numerous behavioural studies have provided evidence that linguistic and non-linguistic cognition interact. A particularly elegant paradigm was introduced by Glenberg & Kaschak (2002). In a sentence-sensibility-judgment task, participants were presented with sentences that described an action involving a movement either towards or away from the body (e.g., *You opened / closed the drawer*). For half of the participants the correct response involved a movement towards their body, for the other half a movement away from their body. Thus, the movement implied by the sentence either matched or mismatched the required response movement. In line with the idea that comprehenders mentally simulate the described actions when understanding the sentence, reading times were significantly faster in the match than in the mismatch conditions. Similar effects have been found in studies presenting isolated words. For instance, processing words like *up* vs. *down* or *towards* vs. *away* is facilitated if the correct response requires a matching rather than a mismatching movement (e.g., Lindsay, 2007). Also, for words referring to entities typically encountered in the upper vs. lower part of the visual field (e.g., *hair* vs. *shoe*) processing is facilitated when correctly responding requires an up vs. down response (e.g., Borghi, Glenberg, & Kaschak, 2004; Lachmair, Dudschig, de Filippis, de la Vega, & Kaup, in press; see also Estes, Verges, & Barsalou, 2008). In addition to these studies (providing evidence for the simulations view of language comprehension with respect to motor aspects) there are many behavioural studies providing evidence with respect to *perceptual* aspects of described states of affairs. For instance, in a study by Stanfield and Zwaan (2001), participants read sentences referring to a particular target entity. The sentences either implied a horizontal or a vertical orientation of the target entity (e.g., (1) and (2) respectively). Responding to a subsequently presented picture of the target entity was facilitated if the picture matched the orientation implied by the sentence. Similar results were obtained for the shape of the entities mentioned in a sentence. Zwaan, Stanfield and Yaxley (2002) presented sentences such as (3) and (4), which depending on the last word in the sentence, implied different shapes of the target entity. Picture-recognition and picture-naming latencies were significantly faster if the depicted shape matched the implied shape (i.e., an eagle with wings outstretched for (3), drawn in for (4)) compared to when it mismatched. The results of these latter studies fit nicely with the idea that readers mentally simulate the described state of affairs when comprehending the sentences. Matching pictures are primed by the simulations that were activated during sentence

reading. This may well be the reason why latencies are faster in the match than in the mismatch condition.

- (1) *He hammered the nail into the wall.*
- (2) *He hammered the nail into the ceiling.*
- (3) *The ranger saw the eagle in the sky.*
- (4) *The ranger saw the eagle in the nest.*

As was mentioned above, the simulations view of language comprehension assumes that words activate experiential traces stemming from encounters with the entities they refer to. These traces are then presumably combined to yield simulations of the described state of affairs in case the words are part of a larger phrase, sentence or text. However, as of yet, not much attention has been devoted to this composition process. The mechanisms by which experiential traces are combined to yield simulations of more complex states of affairs are still unclear. We do not even know whether these mental simulations are created by a compositional process at all. The currently available evidence in the literature does not provide a good basis for reasoning about mental simulations and compositionality. One reason for this lies in the fact that for many empirical results it is not completely clear, which level of comprehension they reflect. To illustrate, let us return to the finding that listening to hand-action-related vs. foot-action-related sentences modulates motor evoked potentials recorded from hand and foot muscles, respectively. In principle this effect may be a word-based effect that is solely due to the verb in the sentence (e.g., *throw*; *kick*). Alternatively, the effect may be a sentence-based effect that reflects the fact that throwing a ball is performed with the arm and kicking a ball with the foot (instead of, for instance, *throwing a tantrum* and *kicking the bucket*). A similar argument can be made for many of the studies. For instance, the match/mismatch effects observed in the orientation/shape studies mentioned above are usually interpreted as being sentence-based. A picture of a horizontally oriented nail is easier to process after reading (1) than after reading (2) because (1) but not (2) describes a situation in which the nail is horizontally oriented. Admittedly, as the orientation is not mentioned explicitly in the sentences, an account that attributes the effect to an individual word of the sentences is not possible. However, the effect may still be word-based rather than sentence-based: In principle it seems possible that the word *nail* in combination with the word *wall* activates an experiential trace of a nail that is horizontally oriented simply because this combination

of words has occurred more often in situations in which the nail was horizontally oriented. The same may be true for the eagle. In combination with the word *nest*, the word *eagle* may activate a trace of an eagle with its wings outstretched, whereas in combination with *sky* it activates a trace of an eagle with its wings drawn in. If so, the effect would not be based on a simulation of the described situation (sentence-based effect) but on traces activated by the combination of words in the sentence (word-based effect). In what follows we will use the term *sentence-based effect* in case the relevant variable is the meaning of the sentence as a whole, or in other words, the state of affairs described by the sentence. We will use the term *word-based effect* in case the relevant variable is the bag of words that make up the sentence, with the syntactic relations between the words being irrelevant. Thus, for a word-based effect, the difference between sentence (5) and (6) should be irrelevant because the same words are mentioned. For a sentence-based effect, in contrast, the difference should be relevant as the two sentences describe a rather different situation. Moreover, a word-based effect should even work with a stimulus such as (7) that simply presents a list of words that do not make up a grammatical sentence.

- (5) At the dance Sarah wore a red dress and black shoes.
- (6) At the dance Sarah wore a black dress and red shoes.
- (7) At a wore shoes the black and dance Sarah dress red.

Whether the effects are word- or sentence-based has theoretical implications. For explaining sentence-based effects one needs to assume that comprehenders composed the meaning of the stimulus sentence and mentally simulated the described states of affairs. In contrast, such an assumption is not required for explaining word-based effects. It suffices to assume that combinations of words activate situation specific experiential traces of the referents they refer to, independent of sentence meaning. Sentence meaning in this case may be composed by a propositional mechanism that is linguistic in nature and independent of the modal systems. Experiential simulations in this case possibly only constitute an optional by-product of comprehension rather than a functional component. We will come back to this issue in the general discussion.

The aim of the experiments reported in this chapter was to find out whether the effects observed in the studies by Stanfield and Zwaan (2001) and Zwaan et al. (2002) reflect word- or sentence based effects. The logic underlying the experiments was as follows: If the effects are sentence-based then match/mismatch effects should depend on the orientation or shape of the target entity as implied by the sentence as a whole. In contrast, if the effects are word-based, match/mismatch effects should only depend on the particular words mentioned in the sentences, and prove independent of the sentence meaning or the particular state of affairs described by the sentence. In addition, match/mismatch effects should be observed in an experimental paradigm that does not even involve sentences as stimulus but rather requires the participant to process a list of relevant content words.

2. Experiment 1

In Experiment 1, we presented participants with sentences that mentioned a particular target entity and implied that this entity was in a certain orientation or shape (e.g., (8) - (11)). In the Zwaan studies the differences with respect to implied shape or orientation were achieved by using different nouns in the two sentence versions (e.g., *sky* vs. *nest* in (3) and (4)) above). According to the word-based hypothesis, these nouns are critical for the occurrence of the match/mismatch effect. In the following, we will therefore call these nouns *critical nouns*. In contrast to the Zwaan studies, we used the same critical words in both sentence versions in this experiment. Differences with respect to the implied shape or orientation resulted from differences in word ordering and syntax. If the match/mismatch effect with respect to orientation and shape is a sentence-based effect then we should be able to replicate it in the present experiment (sentence-based view). The sentences clearly implied a different shape or orientation depending on sentence version. In contrast, if the effect is word-based and due to the co-occurrence of the critical words (word-based view), then it should not replicate in the present study: Both critical words are present in both versions of the sentences. The two sentence versions thus do not differ with respect to the relevant content words, and an effect that is based on words can therefore not be obtained after the sentence has been read.

- (8) *Maria entdeckt den Pinsel im Wasserbecher neben dem Malkasten.*
‘Mary finds the paint brush in the water mug next to the paint box.’
- (9) *Maria entdeckt den Pinsel im Malkasten neben dem Wasserbecher.*
‘Mary finds the paint brush in the paint box next to the water mug.’
- (10) *Der Wanderer fotografiert den Adler im roten Abendhimmel über dem Nest.*
‘The hiker takes a picture of the eagle in the red evening sky above the nest.’
- (11) *Der Wanderer fotografiert den Adler im Nest vor dem roten Abendhimmel.*
‘The hiker takes a picture of the eagle in the nest in front of the red evening sky.’

2.1. Method

2.1.1. Participants

Fifty-two people participated in the study, all with normal or corrected to normal vision.

2.1.2. Materials

A total of 32 experimental sentence pairs were constructed. Each pair mentioned a particular target entity (e.g., paint brush). The pairs were constructed in such a way that they differed with respect to the implied shape or orientation of the target entity. For instance, (8) clearly implies a vertical orientation of the paint brush. In contrast, in (9) the implied orientation is clearly horizontal. Importantly, the two versions of each pair mentioned the same nouns and verbs. The differences in implied shape or orientation were achieved by syntactic manipulations, i.e., by changing the word order or by exchanging spatial prepositions. We deliberately used a variety of different sentence structures to prevent readers from applying strategies when processing the sentences. Thus, some sentences were of the structure exemplified in (8)–(11), others for instance employed relative clauses, as in (12)–(13).

(12) *Der Pfadfinder benutzt die Leiter, um über die Grube neben dem Apfelbaum zu kommen.*

‘The boy scout uses the ladder to get across the pit beside the apple tree.’

(13) *Der Pfadfinder benutzt die Leiter, um auf den Apfelbaum neben der Grube zu kommen.*

‘The boy scout uses the ladder to get onto the apple tree beside the pit.’

Also, the two critical nouns were mentioned in different orders across the experimental sentence pairs. The noun that was decisive for the orientation or shape of the target entity (i.e., water mug in (8), paint box in (9)) was mentioned prior to the other noun (i.e. paint box in (8) and water mug in (9) in some sentences. In others, the order was reversed, as in (14- (15).

(14) *Der Mann schlägt den Nagel dicht unter der Holzdecke in die Wand.*

‘The man pounds the nail close to the wooden ceiling into the wall.’

(15) *Der Mann schlägt den Nagel dicht an der Wand in die Holzdecke.*

‘The man pounds the nail close to the wall into the wooden ceiling.’

Two black-and-white images, depicting the target object in the two implied shapes or orientations, were also constructed to correspond to each experimental sentence pair². This yielded two sentences and two pictures for each of the 32 target objects. Each experimental sentence could be paired with a picture that matched or mismatched the implied shape or orientation of the target object, yielding four possible sentence–picture combinations. Participants were to see only one of these four possible combinations for each target object (see below). A total of 32 additional filler sentences were constructed. These filler sentences were followed by pictures of objects not named in any of the sentences. All pictures were scaled to occupy a 3-inch square on the screen. In addition, a total of 16 comprehension questions were constructed. Eight of these were presented following experimental items and the other 8 following filler items. Half of the questions required a ‘yes’-response and the other half a ‘no’-response.

In summary, each participant saw 32 experimental sentences that were paired with pictures that required a “yes” response. In addition, each participant saw 32 filler sentences that were paired with pictures requiring a “no” response. In half of the experimental trials the picture matched the

shape and orientation of the target object as implied by the sentence. In the other half the picture mismatched with respect to shape or orientation.

To make sure that the two sentences in each experimental sentence pair indeed differed with respect to the picture they matched or mismatched with, we gave four participants, who did not take part in the experiment proper, a list of all 64 experimental sentences together with a list of all 64 pictures. We underlined the word for the target entity in the sentences, and asked participants to select the picture that best matched the target entity as described in the sentence. The mappings of sentences to pictures were as intended. Differences occurred in less than 3% of the cases (i.e., in 7 out of 256 answers).

2.1.3. *Design and Procedure*

We created four lists that counterbalanced items and conditions. Each list included a different one of the four possible versions (2 sentences x 2 pictures) for each object. Each participant saw one of these lists. For two of the overall four versions the picture matched the shape and orientation of the target object described in the sentence, and for the other two the picture mismatched orientation or shape (see Figure 1). For the statistical analyses we combined the two former and the two latter conditions, resulting in a 2 (match/mismatch) design.

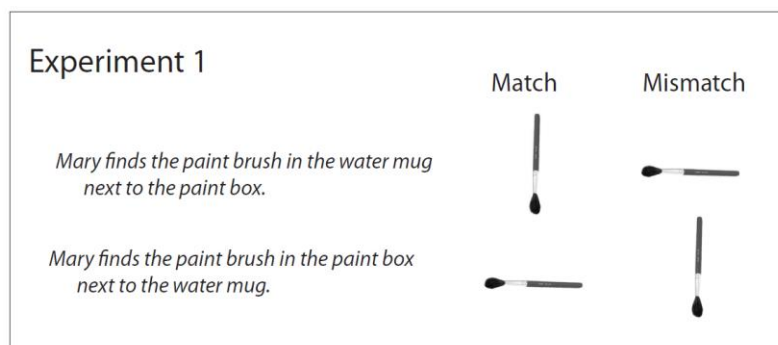


Figure 1. Sample Materials employed in Experiment 1.

Participants were instructed to read each sentence and then to decide whether or not the pictured object that followed had been mentioned in the preceding sentence. They were informed that reaction times and accuracy were being measured and that it was important for them to make the deci-

sions about the picture as quickly and accurately as possible. During each trial, participants first saw a sentence, left justified on the screen, which either did or did not mention the object that they would later see. They pressed the space bar when they had understood the sentence, and then a fixation point appeared at the centre of the screen for 1500 ms, followed by a picture. Participants then determined whether the pictured object had been mentioned in the preceding sentence, by pressing the appropriate key (.-key, marked with “J” for *Ja* (yes), x-key, marked with “N” for *Nein* (no)). On trials with a comprehension question, the question was presented next. Participants were asked to respond to the questions by pressing the “y” or “n” key, respectively. Participants were not given feedback on their responses. The comprehension questions were included in the procedure in order to make sure that participants were reading the sentences for comprehension, rather than only paying attention to the words. The experiment took approximately 20 minutes to complete.

2.2. Results and discussion

Response latencies of experimental trials were submitted to two paired-samples t-tests, one treating participants as random factor and one treating items as random factor. The latency analysis was performed on correct responses only. Responses longer than 6000 ms were omitted. In determining outliers within the remaining latencies, we took not only differences among the participants into account, but also differences among the items. We employed a two-step procedure: First, the valid latencies of each participant were converted to z scores. Then latencies with a z-score deviating more than 2 standard deviations from the mean z score of the respective item in the respective condition were discarded. This eliminated 6 % of the data. The mean latencies and accuracy scores in the match and mismatch conditions are displayed in Table 1, together with the 95% confidence interval for within participants designs (Masson & Loftus, 2003).

Overall participants responded with a mean latency of 1058 ms and a mean accuracy of 97 % to the picture-recognition task. The comprehension questions were answered with a mean accuracy of 84%. In contrast to the predictions of the sentence-based view, latencies in the picture-recognition task did not depend on whether the picture matched or mismatched the orientation/ shape implied by the sentence ($t_1(51) = -0.32$; $t_2(31) = -0.43$; both $ps >.60$). This also held true if we only included those participants who correctly responded to the comprehension questions in at least two third of the cases ($t_1(48) = -0.95$; $t_2(31) = -0.98$; both $ps >.17$, one-tailed).

When we analysed only those 28 participants who made two or less mistakes in the comprehension questions, picture-recognition times were nearly 30 ms faster in the match than in the mismatch condition (1076 vs. 1105ms), but this difference was still not significant ($t_1(27) = -1.013$; $t_2(31) = -0.73$; both p s $>.16$, one-tailed).

The different sentence versions clearly implied different shapes or orientations of the target object. Thus, if readers mentally simulated the sentence content and this then facilitated or hindered processing of a matching or mismatching picture (sentence-based view) then we should have observed a match/mismatch effect in this experiment. In contrast, the results *are* in line with the word-based hypothesis, according to which the match/mismatch effect results because reading the word for the target object (e.g., paint brush) activates different experiential traces depending on the other words in the context. If *mug* is present in the context, a trace with a vertical brush is being activated. If *box* is present, a trace with a horizontal brush is being activated. In this experiment, both of the critical words were present in the sentence, and thus both of the traces should be equally active according to the word-based hypothesis, leading to the observed null-effect.

Table 1. Mean latencies (in ms) and accuracy scores in the match- and mismatch conditions of Experiments 1 - 3. The size of the confidence interval was determined according to Masson & Loftus (2003)

Experiment	Match		Mismatch		CI _{95%}
	M	Acc	M	Acc	
Exp.1: both critical words	1054	.98	1059	.97	+/- 21.6
Exp.2: both critical words	1239	.98	1237	.97	+/- 42.4
Exp.2: one critical word	1115	.97	1191	.95	+/- 39.4
Exp.3: word lists	947	.83	1021	.80	+/- 17.25

Of course it is always difficult to draw conclusions from a null-effect. In principle it is possible that our experimental design was not powerful enough to unearth the effects, or that our stimuli were not adequate for obtaining match/mismatch effects. We therefore conducted Experiment 2, in which we directly compared the conditions of the present experiment with conditions for which both views predict a match/mismatch effect.

3. Experiment 2

In Experiment 2, we manipulated within one experiment whether in addition to the name of the target entity only the relevant critical word was being mentioned (as in the Zwaan experiments; e.g., (1) – (4)), or whether the other critical word was mentioned as well (as Experiment 1; e.g., (8) – (11)). If the word-based hypothesis is correct, then the match/mismatch effect should interact with the number of critical words mentioned in the sentence. If only the critical word is being mentioned that is relevant for the target entity's orientation or shape, then pictures in the match conditions should lead to faster latencies than those in mismatch conditions. If both critical words are mentioned, no latency difference should be observed.

3.1. Method

3.1.1. Participants

Thirty-two people participated in the study, all with normal or corrected to normal vision.

3.1.2. Materials

A total of 32 experimental items were constructed. Each experimental item was available in four versions. Two of the four versions were identical with respect to the nouns and verbs mentioned in the sentences but differed in sentential content. These versions mentioned both critical words. Differences with respect to the implied shape or orientation of the target entity resulted from differences in word order or spatial prepositions. These “both-critical-words”-versions corresponded to the versions employed in Experiment 1 (cf. (8)-(11) above). The other two versions differed with respect to sentential content as well as with respect to the words used in the sentences. The versions mentioned only one of the two critical words. Differences with respect to the implied shape or orientation of the target entity resulted from the use of different words. These versions correspond to the versions used in the original studies by Stanfield et al. (2001) and Zwaan et al (2002), and are exemplified by (16) - (19).

- (16) *Maria entdeckt den Pinsel im Atelier im Wasserbecher.*
'Mary finds the paint brush in the studio in the water mug.'
- (17) *Maria entdeckt den Pinsel im Atelier im Malkasten.*
'Mary finds the paint brush in the studio in the paint box.'
- (18) *Der Wanderer fotografiert den Adler im roten Abendhimmel im Park.*
'The hiker takes a picture of the eagle in the red evening sky in the park.'
- (19) *Der Wanderer fotografiert den Adler im Nest im Park.*
'The hiker takes a picture of the eagle in the nest in the park.'

As in Experiment 1, two black-and-white images, depicting the target object in the two implied shapes or orientations, were available for each experimental item. This yielded four sentences and two pictures for each of the 32 target objects. Thus, each experimental sentence could be paired with a picture that matched or mismatched the implied shape or orientation of the target object, yielding eight possible sentence–picture combinations. The four combinations involving the “both-critical-words”-versions were presented to one group of participants. The four combinations involving the “one-critical-word”-versions were presented to the other group of participants. For each target object, participants saw only one of these four combinations. As in Experiment 1, participants also saw 32 filler items with pictures depicting an object not mentioned in any of the sentences. Again, there were comprehension questions for 16 of the overall 64 trials in the experiment.

3.1.3. Design and Procedure

The design was the same as in Experiment 1, except that there were two groups of participants, one receiving the sentence-picture combinations with both critical words in the sentence, and the other receiving the combinations with only one critical word in the sentence. For each of the two groups four lists were created that counterbalanced items and conditions. Each list included a different one of the groups' four versions (2 sentences x 2 pictures) for each object. Each participant saw one of these lists. For two of the overall four combinations, the picture matched the shape and orientation of the target object as described in the sentence. For the other two the picture mismatched the target objects orientation or shape (see

Figure 2). For the statistical analyses we again combined the two former and the two latter conditions, respectively, resulting in a 2 (group: both-critical-words vs. one-critical-word) x 2 (match/mismatch) design.

3.2. Results and discussion

Response latencies of experimental trials in the picture-recognition task were submitted to two analyses of variance (ANOVAs), one treating participants as random factor and one treating items as random factor. To reduce error variance, we included the counterbalancing factor ‘list’ in the analyses, resulting in 2 (group: both critical words vs. one critical word) x 2 (match/ mismatch) x 4 (list) analyses with repeated measurement on match/mismatch in both the by-participant analysis and the by-items analysis. The latency analysis was performed on correct responses only. Outliers were eliminated according to the same procedure as in Experiment 1. This reduced the data set by less than 4%. One participant was excluded from further analyses because of too many errors in the picture-recognition task (>50%). The means of the remaining latencies and the accuracy scores in the match and mismatch conditions are displayed in Table 1, together with the 95% confidence interval for within participants designs (Masson & Loftus, 2003).

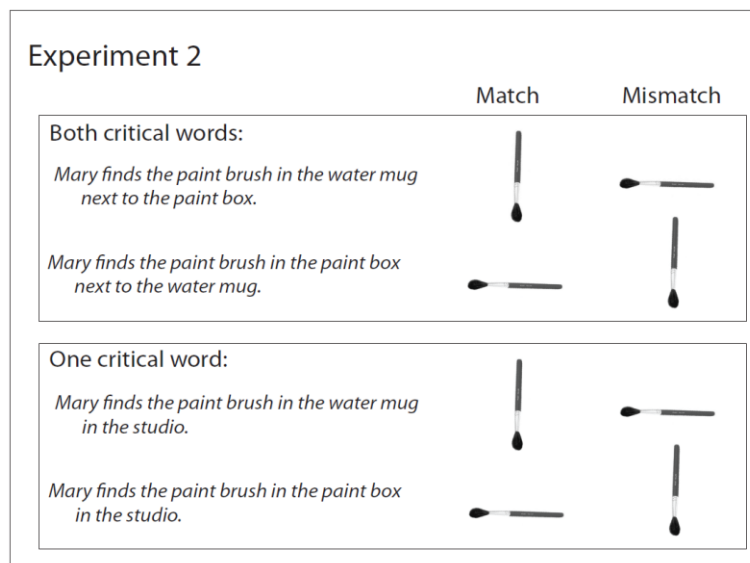


Figure 2. Sample materials employed in Experiment 2

Participants responded to the picture-recognition task with a mean latency of 1197 ms and mean accuracy of 97%. The comprehension questions were answered with a mean accuracy of 80%.

As in Experiment 1, participants' latencies in the match and mismatch conditions were nearly identical for conditions in which the sentence mentioned both critical words. In contrast, for conditions mentioning only one critical word, latencies in the picture-recognition task were 76 ms faster in case the picture matched the implied orientation and shape compared to when it mismatched shape or orientation. These differences were reflected in the statistical analyses: In the overall analyses, there was a main effect of group, which however was only significant in the by-item analysis, $F_1(1,23) < 1$, $F_2(1,28) = 7.0$, $p < .05$. There was no main effect of match/mismatch, $F_1(1,23) = 2.8$, $p = .11$, $F_2(1,28) = 2.8$, $p = .11$. The interaction between match/mismatch and group was only marginally significant in the by-participant analysis, $F_1(1,23) = 3.1$, $p = .09$, $F_2(1,28) = 1.2$, $p = .28$. According to the hypotheses, we nevertheless conducted separate analyses for the two groups. For the conditions with both critical words, there was no match/mismatch effect, $t_1(15) = .05$, $p = .96$, $t_2(31) = -0.01$, $p = .98$. In the conditions with only one critical word, however, the match/mismatch effect was significant in the by-participant analysis and marginally significant in the by-item analysis, $t_1(16) = -2.4$, $p < .05$, $t_2(31) = -1.4$, $p = .08$ (one-tailed).

At the first sight, the results correspond nicely to the word-based hypothesis. For conditions in which the sentence mentioned both critical words there was no match/mismatch effect. This presumably is due to the fact that the two critical words, in combination with the word referring to the target entity, activate conflicting orientations or shapes, therefore eliminating the match/mismatch effect. Put slightly differently, in these conditions the sentences were made up of the same nouns and verbs and should thus be equivalent according to the word-based hypothesis. The null-effect in these conditions replicates the null-effect observed in Experiment 1. The situation is different for the two additional conditions employed in the current experiment. In these conditions, the sentence mentions only one of the two critical words. In combination with the word referring to the target entity, this critical word presumably activates an experiential trace of the target entity in a particular orientation or shape. This orientation or shape matches the orientation or shape in the picture presented in the match-conditions, and mismatches the orientation or shape in the picture present-

ed in the mismatch-condition. The latencies in the picture-recognition task are therefore shorter in the match than in the mismatch conditions.

Maybe the match/mismatch effect was more pronounced in the one-critical word condition in the present experiment because participants in this group put more effort into reading the sentences compared to the both-critical words group? This hypothesis can be ruled out: When analyzing the comprehension questions that were the same for the two groups (i.e., the questions asked in filler trials) no difference in the accuracy scores between the two groups was observed ($t(29) = .33, p > .70$). Differences were observed, however, when analyzing the accuracy scores in experimental trials. Here the one-critical word group significantly outperformed the both-critical word group ($t(29) = -3.33, p < .01$). This probably reflects the fact that the experimental sentences presented in the both-critical word group were longer and more complex than those in the one-critical word group. Of course, this may also be a reason why the match/mismatch effect was more pronounced in the one-critical word condition. Future studies are needed that control for this confound before definite conclusions can be drawn. Moreover, when interpreting the results in favor of the word-based hypothesis, it should be kept in mind that the main task of the present experiment was word-based: In the picture-verification tasks, participants decided whether the depicted entity had been mentioned in the sentence or not. In principle it is possible that participants did not put too much effort into comprehending the sentences, and if so, it might come of less surprise that sentence-based effects were not observed. Indeed when analyzing only those participants who correctly responded to the comprehension questions in at least two thirds of the cases, the pattern of results looks quite different: For these participants we observe no match by group interaction (both $F_s < 1$) but a significant match effect in the overall analysis ($F_1(1,21) = 5.47, p < .05$; $F_2(1,28) = 6.55, p < .05$). As shown in Figure 3, both groups show a clear numerical advantage in the match compared to the mismatch condition, but a separate analysis for the both-critical word group did not reveal a significant match effect ($t_1(12) = -1.09, p = .15$; $t_2(31) = -1.428, p = .09$). It is difficult to interpret these post-hoc results. They could be taken to suggest that sentence-based effects may be observed provided that participants carefully read for comprehension. However, when interpreting the results in this way, it has to be kept in mind that no match/mismatch effect was observed in Experiment 1, not even for participants with high degrees of accuracy in the question-answering task.

Taken together, the results of this experiment combined with the results of Experiment 1 do not allow for definite conclusions with respect to the

word- and sentence-based explanation of the match/mismatch-effect. At the very least however, the results suggest that the usual explanation in terms of purely sentence-based processes may fall too short. Of course, the two explanations in terms of sentence- and word based process must not necessarily be contradictory. In principle it seems possible that the match/mismatch effect reflects a mixture of word- and sentence-based processes with sentence-based processes possibly being particularly pronounced when participants carefully read for comprehension and when the experimental task focuses on sentence-based rather than on word-based processes.

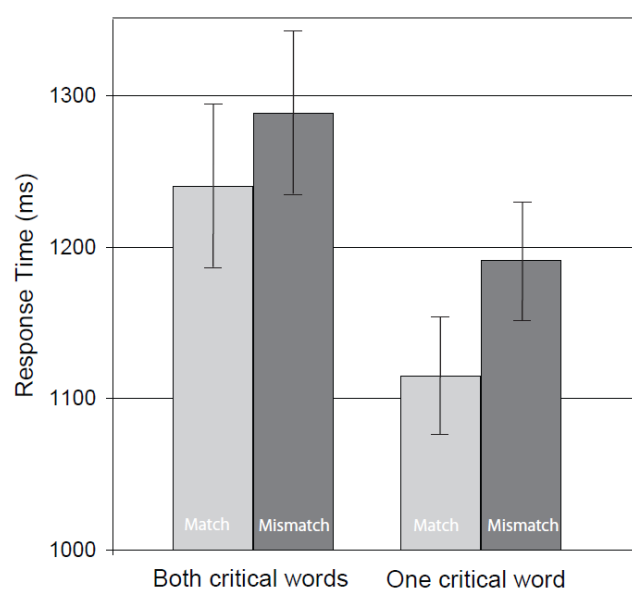


Figure 3. Match/Mismatch effect observed in Experiment 2 for participants with at least 66% accuracy in the comprehension-question task. Error bars represent the 95% confidence interval for within subject designs (Masson & Loftus, 2003)

In Experiments 1 and 2 the experimental task involved sentence reading, and the goal was to find evidence for a word- or sentence-based explanation of the match/mismatch-effect. In Experiment 3 we will focus exclusively on word-based processes in a task that does not involve sentence reading. The logic in this case is the following: Should word-based pro-

cesses contribute to the match/mismatch effect, then a match/mismatch effect should also be observed as long as the relevant words are being processed, even if they do not make up a larger phrase or sentence.

4. Experiment 3

Participants were presented with lists of words and non-words in a lexical decision task. After each sequence of six items a picture was presented, and participants decided as quickly as possible whether the depicted object had been mentioned in the sequence or not. In experimental trials, the sequence contained the relevant words of the ‘one-critical word’-conditions of Experiment 2, intermixed with non-words (cf. (20) and (21)). The picture presented in experimental trials depicted the target entity either in the orientation or shape that matched or that mismatched the words in the sequence (see Figure 4).

(20) *Pinsel / lorfen / entdecken / Tempe / Wasserbecher / Karumpe.*
‘paint brush / lorfing / finding / tempe / water mug / karumpe’

(21) *Pinsel / lorfen / entdecken / Tempe / Malkasten / Karumpe.*
‘paint brush / lorfing / finding / tempe / paint box / karumpe’

If word-based processes contribute to the match/mismatch effect then such an effect should be observed in this experiment because the combination of words in experimental sequences presumably activates a particular orientation or shape. If the presented picture matches the activated traces then picture-recognition latencies should be faster than when the presented picture mismatches the activated traces. In contrast, if the match/mismatch effect were solely due to sentence-based processes, no effect should be obtained in this experiment, in which participants are presented with lists of words, not with sentences.

4.1. Method

4.1.1. Participants

Thirty people participated in the study, all with normal or corrected to normal vision.

4.1.2. Materials

Eighty-five sequences of six strings of letters were constructed. Twenty eight of these sequences were experimental sequences. Each of these contained three words and three non-words, and was available in two versions. The three words in the two versions were taken from the experimental sentence-pairs employed in Experiment 2, namely the pairs in the ‘one-critical-word’ conditions. The three words in each experimental sentence consisted of the word referring to the target entity, the verb of the sentence (in infinitive form), and the critical word of the respective version. Two of the three non-words resembled nouns and the third non-word resembled a verb. The 28 experimental sequences were paired with the 28 corresponding experimental picture pairs from Experiment 2. Thus, each sequence was followed by one of two pictures of the target entity, one with matching and one with mismatching orientation or shape. Half of the 56 filler sequences consisted of four words and two non-words, the other half of two words and four non-words. For each filler sequence one of the words was a verb and one of the non-words resembled a verb. Fourteen of the filler sequences were paired with a picture of an entity mentioned in the sequence (i.e., requiring a ‘yes’-response in the picture-recognition task). The remaining 42 filler sequences were paired with a picture of an entity not mentioned anywhere in the list. Thus, overall, half of the trials required a ‘yes’- and half a ‘no’-response. Also, overall, one third of the trials consisted of three words and three non-words, one third of two words and four non-words and one third of four words and three non-words.

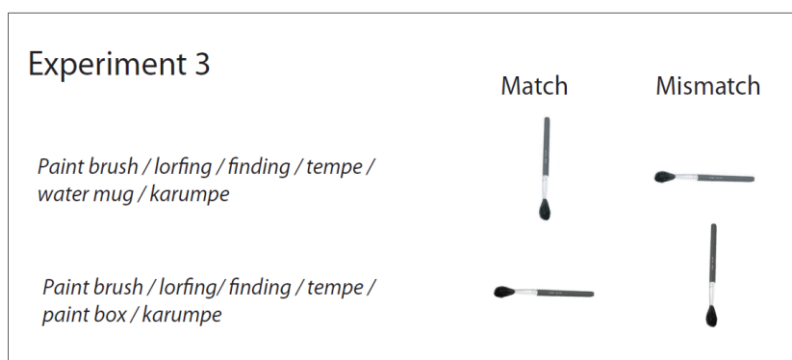


Figure 4. Sample materials employed in Experiment 3

4.1.3. Design and Procedure

The design was the same as in Experiment 1. We created four lists that counterbalanced items and conditions. Each list included a different one of the four possible versions (2 sequence versions x 2 pictures) for each object. Each participant saw one of these lists. For two of the overall four versions the picture matched the shape and orientation of the target object presumably activated by the words in the sequence, and for the other two the picture mismatched orientation or shape (see Figure 4). For the statistical analyses we combined the two former and the two latter conditions, resulting in a 2 (match/mismatch) design.

Participants were told that they were to see sequences of letter strings intermixed with pictures. For each string they were instructed to decide as quickly as possible whether or not it corresponded to a word of the German language, and for each picture they were to decide whether the depicted object had been mentioned in the preceding sequence or not. They were informed that reaction times and accuracy were being measured and that it was important for them to make the decisions about the picture as quickly and accurately as possible. During each trial, participants first saw a letter string left justified on the screen (black font). They decided as quickly as possible whether the string constituted a word or not by pressing the appropriate key (.key, marked with “J” for *Ja* (yes), x-key, marked with “N” for *Nein* (no)). Then a fixation cross came up for 400 ms (black font). Afterwards the next letter string appeared on the screen, and participants again decided whether it was a word or not. Four more letter strings followed in the same manner. After pressing the key in response to the sixth letter string, a red fixation cross appeared in the center of the screen for 400 ms. Then the picture came up. Participants decided whether the depicted object had been mentioned in the preceding sequence of letter strings or not, by pressing the appropriate key (again .key, marked with “J” for *Ja* (yes), x-key, marked with “N” for *Nein* (no)). Participants were not given feedback on their responses in the experiment proper. The experiment took approximately 20 minutes to complete.

4.2. Results and discussion

Outlier elimination was performed as in Experiment 1 with two exceptions. As the latencies in the picture-recognition task of this experiment (following word lists) were shorter than those in Experiment 1 (following sentenc-

es), we omitted responses longer than 4000 ms (rather than those longer than 6000 ms as in Experiment 1). Also, as error rates were higher in this experiment, and the number of observations per cell thus varied to a stronger degree, we did not use a fixed value of ± 2 standard deviations as a cutoff for determining outliers. Rather, we used different values depending on the number of observations per cell, as suggested by Van Selst and Jolicoeur (1994). This eliminated less than 3% of the data. The data of one item were discarded as more than 45% of the participants responded erroneously to this item. The means of the remaining latencies and the accuracy scores in the match and mismatch conditions are displayed in Table 1, together with the 95% confidence interval for within participants designs (Masson & Loftus, 2003). Participants responded with a mean of 984 ms, and an accuracy of 83% to the picture-recognition task in this experiment. Participants' responses to the picture were 74 ms faster if the picture matched the orientation and shape of the target entity as suggested by the combinations of words in the sequence than when it mismatched this orientation or shape (see Figure 5). This latency difference was significant in the by-participant analysis but just missed the usual significance level in the by-items analysis ($t_1(29) = -2.6$; $p < .01$; $t_2(26) = 1.6$; $p = .065$ (one tailed)).

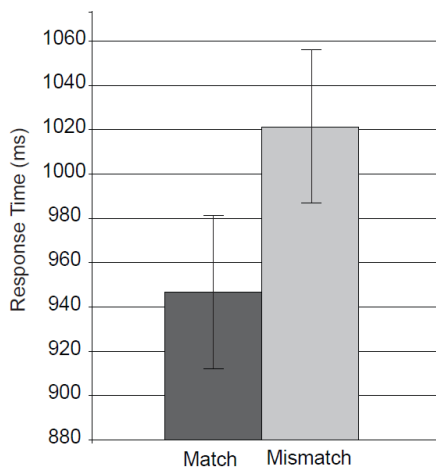


Figure 5. Match/Mismatch effect observed in Experiment 3. Error bars represent the 95% confidence interval for within subject designs (Masson & Loftus, 2003)

This result is in line with the hypothesis that word-based processes contribute to the match/mismatch effect and speaks against the view according to which the match/mismatch effect solely reflects sentence-based processes. An advocate of a pure sentence-based view could argue that participants might have mentally constructed a sentence from the words in the sequence. In this case the match/mismatch effect observed in this experiment would not reflect word-based processes but rather sentence-based processes, despite the fact that the experimental task in this experiment did not involve sentences. The present experiment was not designed to rule out this possibility. However, considering that the words were intermixed with the same amount of non-words in the experimental trials in this experiment we do not consider this alternative explanation to be very likely. In any case, future research is necessary which explicitly addresses this possibility. It would also be interesting to see whether the order in which the words appear in experimental trials makes a difference. In this experiment, the order was based on the order of appearance in the sentences employed in Experiment 2. In principle it seems possible that different results would be obtained if a different order was used.

5. General Discussion

In three experiments we investigated the question of whether the match/mismatch effects observed in the studies by Stanfield and Zwaan (2001) and Zwaan et al. (2002) reflect word- or sentence-based processes. In two experiments, participants in each experimental trial read a sentence referring to a particular target entity and subsequently responded to a picture of this entity. In both experiments, the depicted entity either matched the described state of affairs with respect to the orientation and shape of the target entity or mismatched this state of affairs. A clear match/mismatch effect was observed in conditions in which manipulating the implied orientation or shape went along with mentioning different content words in the sentence (i.e., *eagle* and *nest* vs. *eagle* and *sky*). In conditions in which the implied orientation was manipulated without changing the content words, the match-mismatch effect was less clear. In Experiment 1, no match-mismatch effect was observed under these conditions. In Experiment 2, the effect emerged for those participants who carefully read the sentences for comprehension, as indicated by their accuracy scores in the question-answering task. Furthermore, in Experiment 3 a match/mismatch effect was observed in a paradigm that does not involve sentence reading

but the processing of lists of individual words. These results suggest that the match/mismatch effect may at least partly be due to processes at the lexical level: Words activate experiential traces of the entities they refer to. Combinations of words activate a particular context-adequate subset of these traces. Thus, in isolation, the word *eagle* activates traces of all sorts of eagles, some with their wings outstretched and some with their wings drawn in. In the context of a word like *nest*, however, mainly those traces are being activated in which the eagle has its wings drawn in. This may be part of the reason why a sentence such as (3) activates an eagle with outstretched wings whereas (4) activates an eagle with wings drawn in. If the sentence mentions the target entity with both of these critical words (as in (10) and (11)) then the situation is similar to the isolated word case: Traces with both orientations/shapes are being activated, and accordingly a match/mismatch effect can not occur, unless sentence-based processes play a role as well. Whether sentence-based processes play a role cannot be unambiguously answered on the basis of the experiments reported in this chapter. The experiments neither provide clear evidence against nor clear evidence for the involvement of sentence-based processes in the match/mismatch effect.

The results reported in this chapter indicate that one needs to be careful when interpreting simulation effects observed in sentence comprehension tasks. Under certain conditions, these simulation effects may reflect word-based processes rather than sentence-based processes, even if the task involves sentence reading. As was already mentioned in the introduction, whether simulation effects are word- or sentence based has important theoretical implications: For explaining word-based effects, simple associative mechanisms suffice: Through co-occurrence in experience certain words or combinations of words get associated with certain experiential traces. These traces then are re-activated whenever the words are being encountered (cf. Zwaan & Madden, 2005). In accounting for word-based effects we do not need to assume a composition process that operates on these individual traces. Rather, we may stick to the assumption that the process responsible for composing sentence meaning operates on and results in meaning representations in a linguistic format (i.e., propositional representations; e.g., Kintsch, 1988; McKoon & Ratcliff, 1992; see also Chomsky, 1980; Fodor, 2000; Pinker, 1994). In contrast, for explaining sentence-based effects, we do need to assume sentence-based simulation processes. In our view there are two rather different potential accounts for sentence-based simulation effects. According to first account, experiential traces are the only kind of meaning representation utilized in language comprehen-

sion. Sentence meaning presumably is composed on the basis of the activated traces, and is assumed to result in an experiential simulation of the state of affairs described in the sentence. Obviously, with a radical account such as this one, much theoretical work is needed. Research would need to focus on the question of how lexically activated experiential traces can be combined to yield simulations consistent with the meaning of the larger phrase or sentence. The second account is less radical. According to this account experiential traces and simulations constitute only one of two kinds of meaning representations utilized in language comprehension. According to this account, words and word combinations activate experiential traces (see above). However, the composition process itself does not operate on these traces. Rather it operates on meaning components that are represented in a linguistic format, and its result is a meaning representation in a linguistic format. Once the composition process has determined the meaning of a sentence, the comprehender may simulate the corresponding state of affairs. Obviously, as was the case for the first account, more theoretical work would be needed for this account to be convincing. Research would need to determine how a propositional representation can be “translated” into an experiential simulation. Furthermore, it would be important to investigate whether the experiential simulations of the described states of affairs created after the composition process has taken place are functional for comprehension, or rather constitute an optional by-product of comprehension.

Currently it is not obvious how empirical studies in the field of language comprehension could distinguish between the two alternative accounts for sentence-based effects. Possibly, the temporal dynamics of simulation effects observed during language comprehension could be meaningful. In any case, first future research needs to find out which of the observed simulation effects are word-based and which are sentence-based. Only if there is clear evidence for sentence-based effects does research need to focus on investigating their theoretical basis.

Notes

1. We thank the students of the course “Experimental Methods in Linguistics”, especially I. Andris, H. Bischoff, B. Blankenhorn, J. Boegl, M. Fan, C. Hitzigrath, M. Joachim, S. Maile, A. Plätzer, K. Winkler und L. Riester for their help in material construction and data collection. We also thank Monica de Filippis and three anonymous reviewers for their very helpful comments on

an earlier version of this manuscript. The work reported in this chapter was supported by a grant from the German Research Foundation awarded to the first author (SFB 833; Project B4).

2. Many of the pictures employed in the present studies were employed in the original studies by Stanfield et al. (2001) and Zwaan et al. (2002). We are grateful to the authors of these studies for giving us access to those pictures.

References

- Barsalou, L. W.
2008 Grounded cognition. *Annual Review of Psychology*, 59, 617-145.
- Borghi, A.M., Glenberg, A.M., & Kashak, M.P.
2004 Putting words in perspective. *Memory & Cognition*, 32, 863-873.
- Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gailee, V., & Rizzolatti, G.
2005 Listening to action-related sentences modulates the activity of the motor system: A combined tms and behavioural study. *Cognitive Brain Research*, 24, 355-363.
- Chomsky, N.
1980 Rules and representations. New York: Columbia University Press.
- Estes, Z., Verges, M., & Barsalou, L.W.
2008 Head up, foot down. *Psychological Science*, 19, 93-97.
- Fodor, J.
2000 The mind doesn't work that way. Cambridge, MA: MIT Press.
- Glenberg, A. M. & Kaschak, M. P.
2002 Grounding Language in Action. *Psychonomic Bulletin & Review*, 9, 558-565.
- Glenberg, A. M., Sato, M., Cattaneo, L., Riggio, L., Palumbo, D., & Buccino, G.
2008 Language modulates motor system activity. *The Quarterly Journal of Experimental Psychology*, 61, 905-919.
- Hauk, O., Johnsrude, I., & Pulvermüller, F.
2004 Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, 41, 303-307.
- Kintsch, W.
1988 The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.
- Lachmair, M., De Fillipis, M., Dudschig, C., De la Vega, I., Kaup, B.
in press. Root versus roof: Automatic activation of location information during word processing. *Psychonomic Bulletin & Review*.
- Lindsay, S.
2007 The word action compatibility effect. Paper presented at ESP07. Saarland University.

- Masson, M. E. J., Loftus, G. R.
2003 Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology*, 57, 203-220.
- McKoon, G., & Ratcliff, R.
1994 Inference during reading. *Psychological Review*, 99, 440-466..
- Pinker, S.
1994 The language instinct. New York: Harper Collins.
- Stanfield, R. A. & Zwaan, R. A.
2001 The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science*, 12, 153-156.
- Tettamanti, M., Buccino, G., Saccuman, M.C., Gallese, V., Danna, M., & Scifo, P.
2005 Listening to action-related sentences actiates fronto-parietal motor circuits. *Journal of Cognitive Neuroscience*, 17, 273-281.
- Van Selst, M. V., & Jolicoeur, P.
1994 A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, 47 A, 631-650.
- Zwaan, R. A.
2004 The immersed experiencer. Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation, Vol 44* (pp. 35 -62). New York: Academic Press.
- Zwaan, R. A. & Madden, C. J.
2005 Embodied sentence comprehension. In D. Pecher & R. A. Zwaan (Eds.), *Grounding Cognition* (pp. 224-245). Cambridge: Cambridge University Press.
- Zwaan, R. A. & Stanfield, R. A., & Yaxley, R. H.
2002 Language comprehenders mentally represent the shapes of objects. *Psychological Science*, 13, 168-171.