A Study of Laminar fMRI and Directed Connectivity During Sentence Processing: A Pilot Study

Lenno R. P. T. Ruijters[1,2]
Supervisors: Daniel Sharoh[1,2], Kirsten Weber[1,2]

[1] *Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands*
[2] *Max Planck Institute for Psycholinguistics, The Netherlands*

It is currently unclear how the language network in the brain evokes a concept that is the result of semantic composition. We conducted a pilot study laying the groundworks for a laminar functional magnetic resonance imaging (fMRI) study that aims to investigate this matter further. A paradigm was developed where target words were either primed through compositional semantics, identity primes, neither, or both. A behavioural experiment showed that the identity priming manipulation successfully reduced reaction times in a lexical decision task, which was the predicted behavioural effect. We predicted and found a modulated blood oxygen level-dependent (BOLD) response in the left inferior frontal gyrus and middle temporal gyrus to target words that were compositionally primed. This work will inform a subsequent laminar fMRI study, where the directed connectivity during sentence processing between these regions will be examined.

Keywords: sentence processing, compositional meaning, priming, laminar fMRI, pilot, IFG, MTG

Corresponding author: Lenno Ruijters; E-mail: l.ruijters@spinozacentre.nl

When using a statement like “It’s getting cold” as a request to close a window, one can see that an utterance can have meaning beyond the word-level. The example shows that it is possible to produce utterances that, as a whole, evoke a certain image unattributable to the individual words making up the sentence. A sentence such as, “The man followed the woman and lifted the wallet from her bag,” would likely evoke the image of a pickpocket. The distinction between the word-meaning and the sentence-meaning becomes clearer when the word order is switched around. In “The woman followed the man and lifted the wallet from her bag,” the idea of a pickpocket is largely irrelevant, even though the individual words in both sentences are identical. We will refer to this sentence-level meaning as the compositional meaning of the sentence, and the processes by which it is derived as composition.

Even though speakers constantly derive the compositional meaning of utterances they encounter, this topic has not received a great deal of attention in the neuroimaging and priming literature. Common psychological and neurobiological models of sentence processing acknowledge the presence of some form of semantic composition by including a step where listeners combine words into a coherent whole, but neglect to speculate how this compositional meaning is eventually retrieved (Ferreira & Lowder, 2016; Frazier & Clifton, 1996; 1997; MacDonald et al., 1994; Vosse & Kempen, 2000; Humphries et al., 2006; 2007; Friederici, 2011; Hagoort, 2005; 2013; 2016). If the syntactic and semantic components of an utterance are the recipe for the compositional meaning, then the current models would be akin to a cooking class that ended once eggs, flour and milk are mixed into a bowl. A critical subsequent step is missing to get to a fully baked conclusion. It is currently understudied how the networks responsible for the composition of language evoke a concept or image as a result of complex input.

The common language network

Computational and neurobiological models of sentence processing have both set out to explain the processes that give rise to sentence processing. Computational models set out to study what computations (e.g., statistical inferences or network behaviour) may lay at the root of sentence processing. Neurobiological models in the meantime try to understand which sections of the brain are involved, what the role of each section might be in sentence processing, and how they interplay. Research on the neurobiology of language has converged on a model that includes a large, mostly left-lateralised network. This network includes a large array of regions: from the visual word-form area on the occipito-temporal border, involved in reading (Dehaene & Cohen, 2011) and bilateral temporal cortex, which has been linked to speech sound processing (McGettigan & Scott, 2012), to the inferior frontal gyrus (IFG; Thompson-Shill et al., 1997) and angular gyrus (AG; Binder et al., 2009). We will focus on the properties of the left middle temporal gyrus (MTG), IFG, and AG, as these appear to play a more prominent role in semantic processing of both words and sentences (Friederici, 2011; Hagoort, 2005; 2013; 2016). Understanding this network is then crucial to understanding how compositional meaning is derived from utterances, and vice versa.

When it comes to developing a map of the language network, modern neurobiological models of language overlap in key aspects, for example the importance of the temporal lobe. Involvement of the MTG seems to be critical in semantic processing, lexical memory retrieval, and conceptualisation (Whitney et al., 2010; Snijders et al., 2010; Visser et al., 2012; Hagoort, 2013; 2016; Davey et al. 2016). Structurally, its spatial proximity to the hippocampus, as well as its pronounced structural connections to the occipital, frontal, and parietal lobes make it a strong candidate region for these functions (Turken & Dronkers, 2011). The connections between MTG and frontal cortex are further supported by Xiang et al. (2010), where a functional connectivity map of IFG based on resting state fMRI data showed links with MTG in the perisylvian language network. Here, functional connectivity shows how regions functionally communicate, beyond the structural pathways that may be in place. Due to its wide connections, the MTG is believed to be involved in conceptual representation processes (Snijders et al., 2010; Hagoort, 2013; 2016). Further frontotemporal connections are proposed by Friederici (2011), who lays out the structural pathways through which the temporal cortex communicates with the frontal lobe. The posterior superior temporal gyrus (STG) and BA44 are connected through a dorsal pathway, and the anterior STG is connected with BA45 and the frontal operculum through two ventral pathways (Friederici, 2011). These connections illustrate its potential for complex language processing.

Hagoort (2013; 2016) and Humphries et al. (2006) seem to agree that the MTG is involved in word-level semantic processing. Their claim is supported by, among others, Snijders et al. (2009), which tested the computational Unification Space model proposed by Vosse and Kempen (2000), and found that MTG activity is related to lexical retrieval, including word-level semantics. However, Humphries et al. (2006) also agree with Friederici (2011) in suggesting that the MTG (as well as the STG and the anterior temporal lobe regions) are additionally involved in semantic relation calculations, such as ambiguity resolution. The MTG's connectivity profile nominates it as a candidate for this early composition function, as it both centrally connects brain regions implicated in language processing and since it is sensitive to lexical retrieval demands. Despite the different interpretations of MTG's function in language processing, however, it is generally accepted that the MTG is involved in language processing in some capacity.

As the demand for integration of new input increases, so does IFG activity. It is considered critical to composition processing, be it for relation calculations (Friederici, 2011) or for composition directly (Hagoort, 2013; 2016). Hagoort (2016) provides a description of the connections between the MTG and the IFG, where the two regions are in continuous back-and-forth with one another during ongoing integration of a stream of input. Through feedback loops, context of the overall input is repeatedly adjusted (for a more detailed study of the connectivity profile of the IFG, see Xiang et al., 2010). Thompson-Shill et al. (1997) found that response selection on the basis of semantic relations involved the IFG, but simple retrieval processes did not. This suggests that the IFG is indeed involved when semantic properties of disparate items need to be evaluated (see also Roskies et al., 2006). Zhu et al. (2009) found that small semantic violations evoked larger IFG responses (corresponding with longer rejection times) than large violations (which had shorter rejection times). The differential IFG activity in semantic manipulations suggests that the IFG is involved in semantic processing (as opposed to exclusively syntactic processes), where a higher degree of successful integration leads to more activity. However, Zhu et al.'s (2009) results could also be interpreted as resulting from re-evaluation difficulties, where large violations simply are given up on, and small violations require more effort. Regardless, IFG activity appears to be related to the extraction of meaning from input that requires composition to interpret correctly. Studies on the time course of MTG and IFG activity provide converging evidence for a functional hierarchy between these two regions. Frontal activity during sentence processing follows temporal activity (100-300ms in Friederici, 2011; ~150ms in Hagoort, 2013), suggesting that frontal processes during sentence processing rely on input from the temporal cortex. Again, different models propose different functions of the IFG. Humphries et al. (2006) found IFG activity after both coherent and incoherent complex input, suggesting that the IFG is always active for complex input, and Friederici (2011) suggests that IFG activity in language processing is largely attributable to syntactic processing. Here we will assume that the IFG is the locus for compositional processes, following the Memory, Unification, and Control (MUC) model (Hagoort, 2013; 2016).

The model that we take as a starting-point in this study is Hagoort's (2016) MUC model, due to the fact that it makes clear predictions of feedback between its titular unification and memory components. Following this model, we expect that we can measure both the effect of input in the MTG through retrieval processes, and the effect of composition as a result of a neural feedback system in the MTG (Snijders et al., 2010; Baggio & Hagoort, 2011). The purpose of the present study is to determine whether the IFG communicates with the MTG in feedback fashion during compositional meaning processes, to evoke a unified representation. However, we deviate from the MUC model in one aspect. We would expect that compositional meaning elicitation could also involve other regions, like the AG, as some literature suggests (e.g., Humphries et al., 2006; Binder et al., 2009), which the MUC model proposes is involved in retrieval processes (Hagoort, 2016). Evidence for how or why the AG is involved in this process is currently still lacking, however. Overall, we will maintain an exploratory approach, in order to further develop a fully comprehensive model of sentence (or language) processing in the brain.

Laminar fMRI

The experiment discussed in this thesis will serve as a pilot study for subsequent work investigating the behaviour of laminar structures in relation to processing compositional meaning. It was as such designed with laminar resolution functional magnetic resonance imaging (fMRI) in mind.

Laminar fMRI refers to the acquisition of fMRI data at a high-enough resolution for the blood oxygen level-dependent (BOLD) -response in the grey matter to be separated into several bins over depth (conventionally voxel sizes <1mm^3). The responses within these bins are then looked at independently. Based on the assumption that layer-specific activity is attributable to top-down or bottom-up processes (e.g., Kok et al., 2016; Sharoh et al., 2019), it becomes possible to study how brain regions interact with one another at different levels. Here we define top-down processes as information being relayed from a hierarchically higher region to a hierarchically lower region (i.e., feedback). In more concrete terms, activity in the middle layer of isocortex is generally evoked by signals from brain regions that are involved at a lower level in a given process than a reference region (i.e., bottom-up processes); meanwhile activity in the superficial and deeper layers of isocortex are commonly attributed to input from higher order regions (i.e., top-down processes; Harris & Mrsic-Flogel, 2013; Lefort, et al., 2009; Brown & Hestrin, 2009).

Studying these top-down and bottom-up connections in vivo in this way was inaccessible until the advent of laminar fMRI (Koopmans et al., 2010; Olman et al., 2012; Kok et al., 2016; Sharoh et al., 2019). It is important to understand these connections, since a lot of the brain's functioning relies on attentional and world-knowledge restrictions that are imposed by higher-order areas onto, for example, sensory brain regions (e.g., Hagoort et al., 2004; Kok et al., 2016). One reason why feedback during sentence processing is currently understudied, is because there are notable difficulties in capturing this activity using non-invasive neuroimaging techniques. One way to study feedback processes in the brain would be to look at the functional time-course of different brain regions, and see how activity in higher-order regions temporally correlates to activity in lower-order regions. In fMRI, this is difficult because the low temporal resolution of the BOLD response, on which fMRI relies, makes it difficult to disentangle activity patterns with high temporal overlap. Additionally, the variations in the hemodynamic response make it difficult to compare the temporal relations between activity patterns (Aguirre et al., 1998; Kastrup et al., 1999). These properties of the signals that are measured in fMRI cause problems when inferring which underlying activity preceded which. Electroencephalography (EEG) and magnetoencephalography (MEG) in the meantime suffer from their own restrictions. Even with the high temporal resolution of M/EEG, which allows these methods to capture the fine temporal aspects of communication between the IFG and MTG, and the relatively high spatial resolution of MEG, it is potentially problematic to decide the causal relationship of temporally related activity patterns. This is especially the case in on-going processes such as sentence processing, where continuous activity causes temporal overlap (but see Lyu et al., 2019).

Laminar fMRI, in combination with specialised analysis methods, offers an alternative to the strictly temporally-based method. Laminar fMRI enables the study of directed connections on the basis of structural properties of the in vivo brain (Koopmans et al., 2010; Olman et al., 2012; Kok et al., 2016; Sharoh et al., 2019). By looking at layer-specific activity in laminar fMRI, the issue of low temporal resolution is circumvented when looking at interregional interactions, by looking at which layers of different regions respond to a given stimulus (e.g., Kok et al., 2016; Sharoh et al., 2019). Assuming different layers receive input from different hierarchical directions, by finding an interaction between different layers of two regions, it is plausible to infer how these two regions communicate, especially if there is evidence for a directed correlation between regions. By designing stimuli that differentially evoke bottom-up and top-down activity in a region(s) of interest (ROI), it becomes possible to disentangle which regions communicate with this ROI, in what order, and what proportion of the measured activity within a region relates to bottom-up and top-down processes.

Our manipulation

Behavioural priming is the psychological effect where, once a person is presented with a stimulus, their reaction to a similar stimulus is affected compared to if an unrelated stimulus were presented. An example of these priming effects in language is identity priming, where recognition or production of a word is enhanced if it has been presented before (e.g., Buckner et al., 2000; Raposo et al., 2006). There is also semantic priming, where processing of an item is once again enhanced if it is semantically related to a preceding stimulus (e.g., Hutchison, 2003; Rissman et al., 2003; Rossell et al., 2003; Wible et al., 2006), or syntactic priming, where sentence structures are comprehended or produced more easily if they have been presented earlier (Pickering & Branigan, 1999). Moreover, these effects have been found in neuroimaging experiments in the form of modulated BOLD or event-related potential (ERP) responses (for a review, see Segaert et al., 2013). Our assumption is that identity priming is a mix of lower order priming and semantic priming, and that we can evoke a top-down semantic priming effect through compositional primes. We expect that by manipulating the degree of different forms of priming, we can disentangle the top-down and bottom-up processes that lead to a compositional interpretation of linguistic input. We will further discuss identity and semantic priming, as these forms of priming were used for our manipulation.

At the neuronal level, these priming effects manifest as a modulated response of the neuron due to repeated activation. While this often goes paired with a behavioural effect as well (for a review on semantic priming, see Hutchison, 2003), we will focus on the neural response in light of this study. The most straightforward and strongest effect is that of identity priming. If an individual is stimulated in an identical fashion twice in a row, the neural response to the second stimulation is modulated by the first. Most commonly, this effect is found in the form of repetition suppression (Segaert et al., 2013). This effect has been found in visual priming (Summerfield et al., 2008), auditory priming (Bergerbest et al., 2006), motor priming (de C. Hamilton & Grafton, 2008), and linguistic priming (Buckner et al., 2000; Raposo et al., 2006). We expect that we can modulate the lower order representations of semantic concepts by repeating the lower level stimulation. It should be noted, however, that repetition priming has also been linked to repetition enhancement effects. Segaert et al. (2013) concluded in their review that this enhancement is likely due to feedback effects as well, where the direction of the effect is task dependant. We nonetheless expect part of this offset to be driven by bottom-up activity modulating the neural response to the presentation of an identity-primed target word.

Semantic priming is one step removed from this. Semantic priming is the pre-activation of semantic representations based on a non-identical, semantically related prime. Behavioural experiments have previously shown that reaction times are faster in a number of tasks when a target is semantically primed against a non-primed condition (for a review, see Hutchison, 2003). At the word-level, the semantic priming effect has been found in several fMRI studies, where the STG and the MTG showed stronger activity in the unrelated condition than in the related condition (Rissman et al., 2003; Rossell et al., 2003; Wible et al., 2006). Notably, the priming effect appears to be stronger for identity priming than for semantic priming (e.g., Rugg, 1985). Moreover, the priming effects of linguistic context and repetition appear to be additive (Den Heyer et al., 1985). While Den Heyer et al. (1985) only studied these priming effects in a behavioural study, it suggests that the effect of repetition priming is not at ceiling and can be enhanced with context priming. As such, the combined effect can be distinguished from the individual effects. In addition, if we can find this additive effect in behaviour, it stands to reason that these individual effects have their unique (albeit potentially overlapping) neural fingerprints. Whether the semantic priming effect is the result of feedback from higher regions, communication of neurons on the same hierarchical level, or both is currently uncertain. It is worth noting that expectancy, a top-down effect, can influence the size of the semantic priming effect (e.g., Keefe & Neely, 1990), but it is unclear how such effects inform the expected effects of a compositional prime.

As with identity priming, semantic priming may induce enhancement effects, depending on task or ROI (e.g., Rossell et al., 2003; for a review, see Segaert et al., 2013), but again we expect that semantically priming a target evokes a modulation nonetheless. If we assume that a (deliberately constructed) sentence can elicit a unique semantic representation, then this representation should semantically prime related targets (similar to the context effect in Den Heyer et al., 1985, but at the sentence level). We expect a modulation as a result of the IFG communicating the output of composition to the MTG in the case of these compositional primes, without lower order priming effects. In response to the compositional stimuli, we expect IFG activity to be higher compared to a non-compositional condition, and we expect this activity to be related to MTG activity. By manipulating the compositional semantic representations, we predict we are able to study the top-down effects of the IFG onto the linguistic memory network through semantic priming effects (be they in the form of suppression or enhancement). It should be pointed out that identity priming may include a semantic priming effect, since prime and target are semantically related, but it includes a lower order priming effect as well. We do not expect a correlation with IFG activity and the identity manipulation.

Our main interest is to see how the interpretation of utterances would neurally prime a target word. To do this, we want to look at the differences between priming of a target word by an identity token and by a complex, compositionally related token. We assume that priming effects from a compositional token would result from higher-order composition processes. These higher-order processes proceed to activate a combined conceptual representation that would be relayed back to the memory compartment of the language system. This feedback-induced priming effect would be identifiable by its laminar profile. In order to allow the study of the laminar activity profile of top-down processes in the future, we manipulate the bottom-up effect as well. This way, changes in the BOLD response between layers and conditions may be disentangled. We assumed that the priming effect of the identity prime includes a bottom-up effect in the MTG. Combining these factors, we end up with four conditions: a compositional prime (sentence) that includes an identity prime (i.e., a copy of the target) (+Comp/+ID); a compositional prime (sentence) where the identity prime is replaced with another word, but is otherwise identical to the +Comp/+ID prime (+Comp/-ID); a non-compositional prime, which is a scrambled list of words, one of which is the identity prime for the target (-Comp/+ID); a non-compositional prime (scrambled list of words) that does not contain the identity prime (-Comp/-ID).

Since this is a pilot study, our predictions will remain general. We predict that we find a priming effect in the left MTG for both the identity and compositional priming conditions. We expect that the IFG is more strongly involved in our compositional condition. We expect that MTG activity shows a repetition effect for the identity conditions, and a semantically primed pattern for the compositional conditions. We also might expect some AG activity in response to the compositional primes.

Methods

Design

We designed an experiment to identify potentially distinct language processing mechanisms through different priming methods. Our paradigm consisted of a 2*2 design with compositional priming (on vs. off) by identity priming (on vs. off) manipulations. Primes consisted of strings of words. Targets consisted of single words. The prime strings were either intact Dutch sentences (Sentence), or random strings of Dutch words (Scramble). The sentences were always designed to be compositionally equivalent to the target word (compositional primes), whereas the scrambles were designed not to be (non-compositional primes). Our goal was to semantically prime the target with the compositional primes. In addition, we manipulated whether the target word was present in the prime string (identity prime) or replaced by a word that did not prime that target by itself (non-identity prime). See Table 1 for a schematic overview of the study design. Target words never exceeded 13 characters in length to fit within the reading span limit imposed by Rayner (2009) and consisted only of nouns and (inflected) verbs. Scrambles were created so that there was no intrinsic semantic priming at both the compositional and word levels (see subsection scrambling for the scrambling procedure).

Stimulus creation

Stimuli were constructed in several steps outlined below. Two native Dutch speakers constructed 236 sentences for the compositional prime conditions. 126 of these sentences were adapted from Lai et al. (2015). Each sentence was constructed with a target in mind that semantically summarised the prime (e.g., PRIME: “The man followed the woman and lifted the wallet from her bag”; TARGET: “Pickpocket”). Additionally, the target had to be able

Table 1. The primes for the target “verdrinking” (drowning) in each condition. Original Dutch presented in black, followed by the grey English translation. The translation is intended to reflect the intended message of the sentence and may not accurately reflect the word-count or the semantic relations between individual items and the target of the Dutch item.

Prime-type	Compositional	Non-compositional
Identity	Ze probeerde haar hoofd boven het oppervlak te houden maar kon de **verdrinking** niet meer bevechten. She tried to keep her head above the surface, but could not fight the **drowning** any longer.	Door keek naar hadden had linker torso waren de opzocht achter rieten **verdrinking** de op niet. By looked to had had left torso were de searched behind reed **drowning** the on not.
Non-identity	Ze probeerde haar hoofd boven het oppervlak te houden maar kon de **stroming** niet meer bevechten. She tried to keep her head above the surface, but could not fight the **current** any longer.	Door keek naar hadden had linker torso waren de opzocht achter rieten **stroming** de op niet. By looked to had had left torso were de searched behind reed **current** the on not.

to replace an item in the sentence without altering the interpretation or interpretability of the sentence (e.g., PRIME: "The pickpocket followed the woman and lifted the wallet from her bag"; TARGET: "Pickpocket"). The stimuli based on Lai et al. (2015) contained a valence confound. We subsequently constructed our items so that half of the final set would contain negative valence composition, and the other half neutral valence composition. Valence was considered an orthogonal manipulation but was not included in the pilot analyses.

Nearest neighbour test. Following stimulus construction, we confirmed that the individual lexical items in the primes were not likely to semantically prime the target by themselves. The vector space model for semantic distance snaut was used for this purpose (Mandera et al., 2017) and rank nearest neighbour was taken as our distance metric (Lapesa and Evert, 2013; 2014). Lemma-level rather than word-level relations were used to avoid the influence of syntactic information. Nearest neighbour ranks of lemmas were based on the cosine distance between vector representations (Kenett et al., 2017; Ettinger et al., 2016; Lenci, 2018). Lexical items that fell outside the first 40 nearest semantic neighbours were considered non-priming. We verified both forwards and backwards distance values.

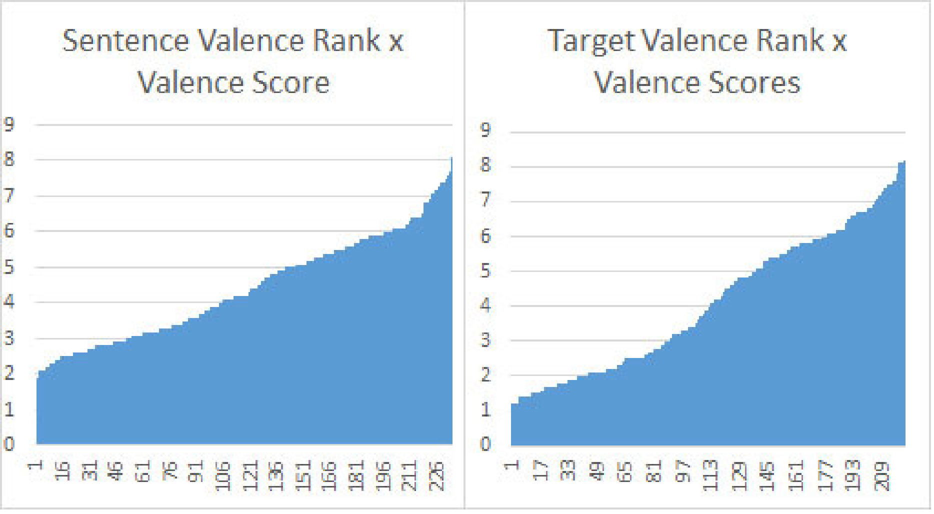
Frequency in the training data of snaut seemed to affect the reliability of the semantic distance values. Vector values that were based on insufficient data were taken to be unstable. We determined an item needed at least 500 exemplars in SoNaR-500 (Oosterdijk et al., 2013) . Word-level distance was considered if snaut reported no instances for the lemma-level distance, and if the frequency threshold was met in SoNaR-500. Where spotted and possible, if the item used in the stimuli had a low frequency (<500), a synonym with a higher frequency was used in snaut (e.g., kots instead of braaksel; both "vomit"). In case the target had a SoNaR-500 frequency of <500 and no reasonable alternative form was available, the target was replaced entirely. Two items were excluded after this step (resulting N = 234).

Independent prime-target associations. Following the nearest-neighbour assessment, one RA and two interns were asked to provide at least three associations for each prime independently. If at least two of these judges agreed on a target, the target was accepted for later steps (be it our target or a new target). If one agreed with our target and/or the rest gave associations close to our target, the target was checked and either changed accordingly, or left unchanged for later steps. If targets were changed, they were checked against snaut again. If there was no consensus on a target, or if implementing a new target would violate the other criteria for our stimuli, the entire item was removed from the stimulus list. 223 items remained in the stimulus list after this process.

Assessing stimulus frequency. A frequency analysis of the lemmas of our items was performed next using the subtitle corpus SubtLex-NL (Keuleers et al., 2010; see also New et al., 2007; Dimitropoulou et al., 2010). This was done to reduce the influence of item frequency on our measurements. Non-content words were deleted from the stimulus list, and remaining items were replaced with their uninflected forms. The raw lemma frequency for each content word was extracted and the 10log for each item was calculated (Howes & Solomon, 1951; but see also Murray & Forster, 2004). Next, the mean and standard deviation for targets (M = 2.74, SD = 0.712, n = 234) and prime content items (M = 3.60, SD = 1.10, n = 1606) were calculated separately. The outcome indicated that targets were on average 10 times less frequent than prime content items. This was to be expected since target items were often more specific (i.e., less underspecified) than the items they replace in the primes, and thus less frequent. We ensured that no items were more or less frequent than three times the standard deviation from the mean. This was done separately for target items and prime content items. Six items fell outside the 3SD range (one of 234 targets and five of 1606 sentence items). These were replaced and the semantic relations of these new items to the targets/primes were confirmed using snaut. A subsequent identical check was done at 2SD, to ensure that, while all items were within 3SD, the distribution was not weighted towards the tails. Here 52 items stood out (six of 234 targets, 46 of 1606 sentence items). This was considered to be within acceptable limits.

Valence check. As the stimuli contained a valence manipulation, we verified that independent participants replicated our valence labelling. A valence judgement survey was constructed for LimeSurvey (Limesurvey GmbH). Items had to be judged on a 1-to-9 scale (1 being negatively valent, 5 being neutral, 9 being positively valent), using the Self-Assessment mannequin (Bradley & Lang, 1994). Four separate surveys were developed totalling all individual lexical items in our stimuli (N = 1107, 223 Target items), and another survey for the compositional primes as a whole (N = 2342). Subjects were recruited using the Max Planck Institute for Psycholinguistics participant database until each list had exactly ten complete responses (total N = 57, male = 8). The study was conducted according to the institutional guidelines of the local ethics committees. The survey was conducted online and in an environment unknown to the experimenters. Subjects were encouraged to complete the survey on a laptop or desktop computer (mobile N = 4). Subjects could pause and recommence the survey at any time. Informed consent was collected through a confirmation screen at the start of the questionnaire, and subjects were reimbursed for 30 minutes of their time. Individual responses were excluded on the basis of the item-specific SD and mean, where responses > 2SD from the item mean were excluded. The mean and SD for each item was then recalculated. All three categories showed evidence for high linearity (sentences, targets, and sentence items). Rank correlation for prime valence was ϱ = 0.991 (Fig. 1A), for target valence ϱ = 0.985 (Fig. 1B), and for prime content item valence ϱ = 0.962. When comparing the pre-labelled valence groups (negative vs. neutral) on valence scores using a t-test assuming inequality between groups (negative = 115; neutral = 108), scores showed a significant difference between the negative-labelled items and the neutral-labelled items (P = 2.0e-53, ci = -2.64; -2.20, t-stat = -21.6 and DF = 193.15). Testing for a correlation between prime scores and target scores showed evidence for a correlation, after removing prime-target pairs with a difference in valence scores > 3 SD (= 3 * 0.76) from the mean difference (= 1.0) (Nremoved = 2). Due to an unidentified error, one item was missing from the resulting list. The resulting list included 220 items with a correlation coefficient of ϱ = 0.81 between prime valence and target valance (Fig. 2) at P = 3.7e-49. Prime content items had a mean valence value of 5.2, with an SD of 1.60 after removing outlier responses. Due to the high linearity in the valence scores, we decided to add the prime and target valence scores as regressors for each item for the analysis in the main study, rather than rely on the binary negative/neutral labels.

Semantic relatedness test. The remaining items were then assessed in terms of their semantic relatedness. A semantic relatedness judgment task was created using LimeSurvey (Limesurvey GmbH). Participants who participated in the valence survey were excluded from participating in this survey. Target items were judged on their semantic relatedness to a preceding sentence on a 1-to-5 scale. Sentences could not be revisited once the target was presented. Our stimuli were split into two counterbalanced lists. In each list, half the target items remained paired with their intended prime, and half the targets were swapped randomly between primes. These new pairs were then checked to make sure they were not judged as semantically related at the sentence level by one of the researchers. Next, two versions of both lists were created, where each version contained ~50% +Comp/-ID primes and ~50% +Comp/+ID primes, to ensure that the identity manipulation did not have a detrimental effect on the semantic relatedness judgment. This resulted in a total of four conditions (related vs. unrelated target * +Comp/-



A B

Figure 1. A. Valence scores ordered by rank for each prime. **B.** Valence scores ordered by rank for targets. Both primes and targets show high linearity between score and rank (ρ = 0.991 and ρ = 0.985 respectively).

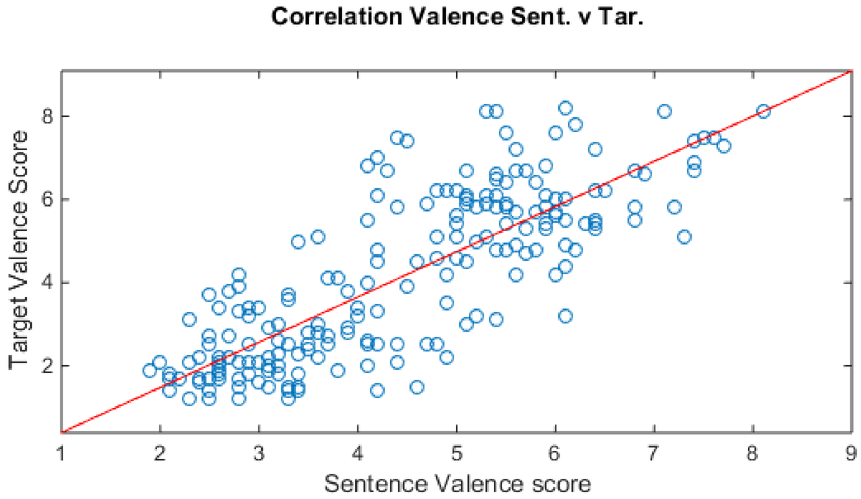


Figure 2. The correlation between the mean valence score for each sentence and their attributed target words, ρ = 0.81.

ID vs +Comp/+ID prime) for each stimulus across lists (stimuli per list = 221). Subjects were recruited using the Max Planck Institute for Psycholinguistics participant database until each list had a total of five complete responses (total N = 22). The study was conducted according to the institutional guidelines of the local ethics committees. The survey was conducted online and in an environment unknown to the experimenters. Subjects were encouraged to complete the survey on a laptop or desktop computer (tablet N = 1, mobile N = 1). Subjects could pause and recommence the survey at any time. Informed consent was collected through a confirmation screen at the start of the questionnaire, and subjects were reimbursed for 30 minutes of their time. All complete responses were included in the analysis. We chose an average relatedness score of >3 in the related conditions as a cut-off point to exclude items. This point coincided with 2SD from the mean of the +Comp/-ID primed targets. Thirteen items were removed from the stimulus list as a result (all on the basis of the +Comp/-ID prime relation), leaving a total of 208. Once these items were removed, no further statistics were performed on the results. Visualisation of the mean scores after exclusion indicated a clear separation between the related and unrelated conditions (Fig. 3).

Scrambling. After the two sentence lists (+Comp/+ID and +Comp/-ID primes), scramble primes for each target were generated (-Comp/+ID and –Comp/-ID). Two non-compositional prime lists were created, where one contained the identity prime and one did not. To create these lists, the +Comp/+ID prime lists were randomly divided into two groups. Each group was converted into a string array, where each prime was represented as a row vector. Individual strings in each array were first jumbled within columns. Gaps formed within rows. To repair these gaps, all strings were moved to the top of the array, and empty values were inserted between strings so that each row was of the same length as it was previously, matching the length of its associated compositional prime. Then all prime content items were jumbled within rows, and all strings were aligned to the left to remove any gaps. Identity tokens were swapped back to their original row and column, so that they matched the location in the original primes. Each resulting group was copied, and for each copy the identity token was replaced with the prime-matched token in the non-identity prime condition (e.g., verdrinking with stroming in Table 1). Some primes contained split verbs in the -ID condition but not in the +ID condition. Missing words in these non-identity scrambles were added to the prime at a location that matched the number of words between the prime-matched token and the target in the non-identity sentence. Each non-compositional prime was checked with snaut again (Mandera et al., 2017). In case of a violation (rank nearest neighbour <40), items were manually swapped between primes. The result was four non-compositional prime groups, each matching half of the compositional primes. These consisted of two -Comp/+ID prime groups, with two matching -Comp/-ID prime groups.

Next, all four prime conditions for each target were divided into presentation lists. Each list was created by combining non-compositional primes for one half of the targets with compositional primes for the other half. For both prime-types, half of the stimuli included an identity prime, and half did not,

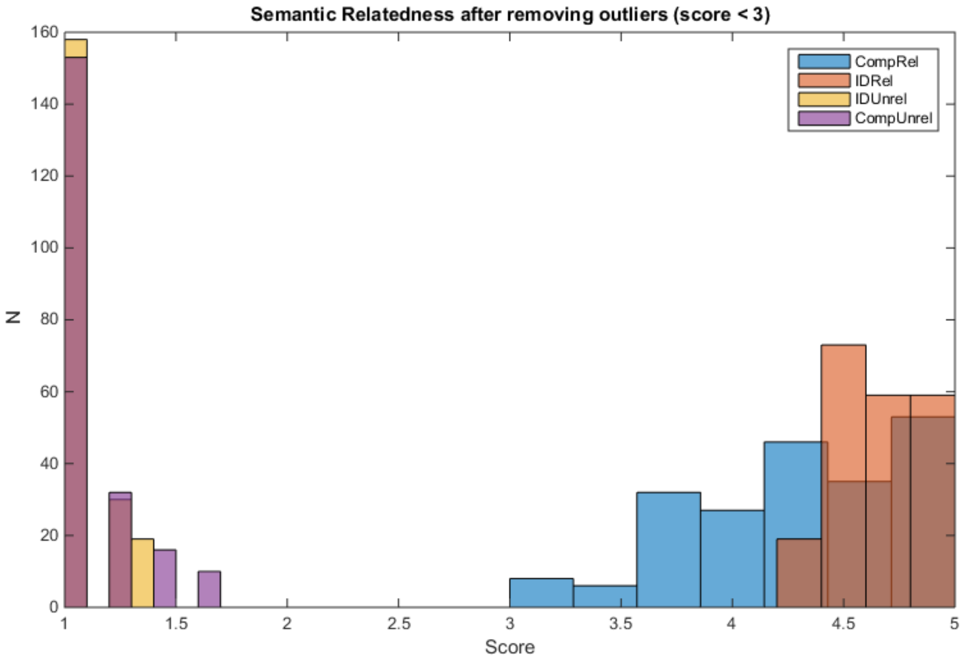


Figure 3. Histogram of the spread of mean semantic relatedness scores between sentences and targets. Purple and yellow indicate the condition in which an unrelated target was presented. Blue and orange show the Compositional/Non-identity and Compositional/+Identity relatedness scores respectively. Items in the relatedness condition with a score of < 3 were removed.

so that each subject saw all four conditions. This resulted in four unique, counter-balanced lists (see Supplementary information for the final lists used in the fMRI experiment). At this point, duplicate targets were still present in the stimulus list. Duplicate targets were removed to avoid between-trial identity priming. A final run-through of the stimuli revealed 15 items that still needed to be removed, leaving the final count at 193 prime-target pairs. The final number of Lai et al. (2015)-based stimuli was 102.

Behavioural experiment

Design. The behavioural experiment was still part of the pre-piloting, designed to see if a behavioural priming effect could be detected before attempting to find it in an fMRI experiment. The experiment consisted of a lexical decision task in order to provide participants with a task that did not explicitly draw attention to the different prime types. For this purpose, the main study design was extended to a 2*2*2 design, where a word-pseudoword factor was added. For each list, half the targets for all four conditions were replaced with a matched pseudoword. Pseudowords were constructed using Wuggy (Keuleers & Brysbaert, 2010). Verbal and written consonant-vowel structures as well as assumed stress patterns were maintained based on native speaker intuition. In addition, transitional probabilities were maintained where possible . This resulted in a total of eight unique, counter-balanced lists. Items were initially ordered randomly, before being reordered manually to avoid identity priming of targets by earlier unrelated primes. That is to say, specific primes would only appear once a particular target had already been presented. Some of these cases survived for the behavioural experiment but were filtered out in subsequent experiments.

The experiment was performed using Presentation® software (Version 20.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Primes were presented word-by-word in white text on a black background, with an on-time of 300ms and an off-time of 150ms. After each prime, a fixation cross was presented for 1000ms, before a 6000ms presentation of the target. Finally, a fixation cross was presented again for 1500ms before the next trial began. The task was divided into 13 blocks, between which short breaks were planned.

Data acquisition. Participants were recruited using the Max Planck Institute for Psycholinguistics participant database, and subjects received a reimbursement for 45 minutes of their time. The study was conducted according to the institutional guidelines of the local ethics committees. Written consent was collected before the start of the experiment. Participants were instructed to indicate upon target presentation whether the target was an existing Dutch word or not with a button press, and participants were informed that there was a time limit involved. A total of 33 healthy, right-handed native Dutch-speakers, aged 18-40 without diagnosed reading or language disorders, were recruited for this experiment. One subject was excluded due to a pre-screening error. One subject showed an error-rate well below chance (85-100% per condition), and their responses were taken to be inverted. This subject was still included in the analysis after reversing their responses. Consequently, we collected four complete responses for each stimulus list.

Analysis. For the analysis, responses were filtered to only include correct responses to real-word trials, leaving only the original two factors (+Comp prime vs. -Comp prime * +ID prime vs. -ID prime). Data were analysed in R (R core team, 2018) using a linear mixed effects (LME) model from the LME4 package (Bates et al., 2015). Our model included the fixed interaction effect between the two factors, as well as random effects of subject and stimulus/trial number. Contrasts for both factors were set to -0.5 and 0.5. This was the most complete model we could use for the analysis. Adding random slopes to the model prevented convergence and were therefore not included.

FMRI experiment

Design. For the fMRI experiment, we utilised the original 2*2 design (+Comp prime vs -Comp prime * +ID prime vs -ID prime). The pseudo-word condition was dropped, which meant that the amount of trials for the remaining conditions doubled, and thus the statistical power of the experiment was increased. The experiment was performed using Presentation® software (Version 20.2, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). Primes were presented word-by-word in white text on a black background, with an on-time of 300ms and an off-time of 150ms. After each prime, a fixation cross was presented for an amount of time that extended the prime duration to the next multiple of 650ms plus 650 or 1300ms. Afterwards, the target was presented for 1300ms. Finally, a fixation cross was presented again for 1000-2000ms (in steps of 50ms) before the next trial began. The task was divided into seven runs, between which short breaks were planned. During each break, a light-blue screen was presented to reduce participant fatigue. Catch trials were added every 6-8 trials. During catch trials, a question mark was presented, and participants had to indicate by means of button press whether the preceding prime was a sentence or a scramble. The motivation behind this task was to ensure participants read our primes as intended. Catch trial duration was intended to be 2600ms, but due to a programming error, catch trial duration was 3000ms for eleven of twelve subjects. This difference was accounted for in data analysis. Stimuli were projected with an EIKI LC-XL100 beamer at a resolution of 1024x768 (4:3 aspect ratio) and a refresh rate of 60Hz on a 451x338mm screen. Participants viewed stimuli at the back of the scanner through a mirror mounted on the head coil. Before the start of the experiment, subjects had an opportunity to adjust this mirror for full view of the projection. All four lists that resulted from the stimulus creation described above were included. Items were reordered again after target order concerns remained after the behavioural experiment.

Data acquisition. The experiment was conducted using a Siemens Prisma 3T MRI scanner at the Donders Institute for Brain, Cognition, and Behaviour. For each participant, we collected a 0.8mm isometric anatomical map using MP2RAGE (Marques et al., 2010) with a FoV of 256mm and a flip angle of 8°. A TR of 2400ms, TE of 2.22ms, and TI of 1000ms, as well as a GRAPPA acceleration factor of 2 meant a full anatomical image was acquired in just under seven minutes.

Functional data was not acquired with a laminar analysis in mind during the pilot. As such, only 1.3mm isometric fMRI data was collected. Our protocol entailed a 3D-EPI sequence based on Stirnberg et al. (2017) with a FoV of 210mm and a flip angle of 20°. The TR was 2600ms, TE1 32ms, multi-echo dTE 60ms, and TI 900ms. Acquisition was accelerated with CAIPIRINHA at a factor 6. SPAIR fat-saturation was enabled. FoV positioning was determined by means of a head-scout. The full scan-time was approximately 45 minutes, excluding breaks.

Heart rate and respiratory data were collected with a BrainProducts BrainAmp system and were added as regressors in the fMRI analysis. Heart rate was collected using a pulse sensor made MRI-tolerable, and respiratory data by means of a respiration belt. Physiological data were converted to workable files before they were checked and cleaned using HERA (Hermans, 2018), and finally converted to regressors using RETROIcorplus (Glover et al., 2000). The first through fifth order sine and cosine heart rate and respiratory regressors, as well as the first order sine and cosine interactions between heart rate and respiration were modelled. Eye-tracking was on during scanning to monitor participant drowsiness, but eye-movements were not recorded.

Thirteen healthy, right-handed native Dutch-speakers, aged 18-30 without diagnosed reading or language disorders, were recruited for this experiment. One subject was excluded due to a pre-screening error. The study was conducted according to the institutional guidelines of the local ethics committees, and subjects were reimbursed for 90 minutes of their time. Written informed consent was collected before the start of each experiment. For each stimulus list, three responses were taken into the analysis.

Pre-processing. Pre-processing was performed using AFNI tools (Cox, 1996; all AFNI programmes discussed are prefixed with 3d). Pre-processing consisted of realignment of the functional volumes using 3dvolreg to a base of the fifth volume of the first run. Motion correction parameters were also collected from this realignment. The anatomical images were skullstriped using 3dSkullStrip. A mask was applied to the base volume of each subject using 3dautomask, and the anatomical image was aligned to this mask using the 3dallineate tool with a quintic interpolation.

For one participant, FoV was shifted for the first run compared to the other runs. In this participant, the base volume for motion correction was taken from the second run. For five participants, the FoV was shifted for the last two runs compared to the rest. An extra pre-processing step was added for these runs. Transformation matrices were calculated by catenating the transformation matrices of an initial large realignment shift to bring the volumes into the same space as the base (AFNI 3dallineate) and a subsequent small motion correction shift (AFNI 3dvolreg). This matrix was then applied to the original NIfTIs using 3dallineate to reduce the number of interpolation steps by one. For all participants except one, this resulted in good realignment. Runs six and seven were excluded from analysis for the remaining participant. Heart rate data for run three in one participant were unusable. The design matrix for this participant was adjusted by only including the respiratory regressors for run 3, and adding the regressor arrays of runs 1-2, 3, and 4-7 separately.

For group-level comparisons, our data were aligned to MNI152 space. The anatomical scan of each subject was normalised using 3dUnifize in AFNI, and brain extraction was performed with the mri_watershed program in freesurfer (Ségonne et al., 2004; Fischl, 2012). For each subject, eleven brain-extracted images were generated where the height parameter of mri_watershed varied between ten and 20. For each subject, the best brain-extracted image was selected and manual improvements were applied as necessary. A non-linear transformation was then applied to these brain-extracted anatomicals to align them in MNI152 space using 3dQwarp. The align_epi_anat.py script in AFNI (Saad et al., 2009) calculated the affine transformation matrices for the automasked functional data to the brain-extracted anatomical, which were then catenated with the non-linear transformation parameters to bring the functional data to MNI152 space. The resulting images were then smoothed using an 8mm smoothing kernel in 3dmerge (Cox, 1996).

Analysis. Design matrices for the group-level analysis were modelled so that the first four columns matched conditions between subjects (column one: +Comp/+ID, column two: +Comp/-ID, etc.). These models were fitted within the GLM framework, and beta-maps and residual-maps were calculated for each subject. Fitting was performed with in-house MATLAB code and an open source fMRI analysis toolbox (https://github.com/TimVanMourik/OpenFmriAnalysis). 3dANOVA3 (Cox, 1996) was used to compare the resulting beta maps in a repeated measures ANOVA. To compare the +Comp/+ID condition to the -Comp/-ID activity patterns (the full priming effect vs. no priming) and the +Comp/-ID vs -Comp/+ID activity patterns (composition priming only vs. identity priming only), separate pairwise comparisons were performed using 3dttest++. To estimate the smoothness of the data, the spatial autocorrelations of the residual time series were calculated with 3dFWHMx and averaged over all runs and subjects. The resulting correlation parameter averages were entered into 3dClustSim to estimate the minimum cluster size required to exclude noise clusters with a likelihood of $\alpha < 0.05$. The uncorrected p-value threshold was set to $P < 0.001$.

Partial $\eta 2$ maps were calculated from the F-maps. These were calculated voxel-wise on the basis of voxel-specific F-values and the degrees of freedom (Cohen, 1973). Effect size-maps for pair-wise effects were also calculated using a version of Hedge's g (Hedge, 1981) that accounted for low sample sizes. Average beta-maps and standard deviation-maps for each condition across subjects were calculated using 3dMean (Cox, 1996). A voxel-wise calculation was performed over these maps based on voxel-wise betas and standard deviations per condition, and sample size. All effect size maps were then masked with a mask of the MNI template dilated by one level in 3dmask_tool to generate the final effect size maps.

Results

Behavioural experiment

The results from the LME analysis indicate a robust difference in lexical decision reaction time as a function of identity prime type compared to non-primed targets, and a potential difference as a function of compositional prime types. The LME showed a robust effect for subjects recognising identity primed words (+ID) 37.21 milliseconds faster than non-identity primed words (-ID) (SE = 7.71, t = 4.83). Additionally, the LME showed a potential effect where compositionally primed words (+Comp) were recognised 20.10 milliseconds faster than non-compositionally primed words (-Comp) (SE = 7.71, t = 2.61). There was no evidence for an interaction of ID-effect * Composition-effect ($\beta = 0.80$, SE = 15.42, t = 0.05). T-statistics were not subjected to significance testing owing to the challenges in assessing the degrees of freedom in LME analysis (Luke, 2017). T-to-z conversion is a common method to derive p-values for LMEs but is anti-conservative (as are most alternative methods; see ibid.). Not reporting p-values prevents marginal effects from being interpreted as significant, but the t-values still offer some insight into the robustness of the effects. See Figure 4 for a visualisation of the reaction time results (Allen et al., 2018; Wickham, 2009). See Table 2 for an overview of the main effects.

Subsequent pairwise comparisons were performed using the emmean package in R (Fig. 4; Lenth, 2019) in order to gain further insight into the potential additive effects of the compositional priming effect. The comparisons revealed a significantly faster reaction time of 37.6ms for identity-only primed (-Comp/+ID) targets compared to non-primed (-Comp/-ID) targets (SE = 10.9, $P < 0.005$), and a 36.8ms decrease in reaction times for identity and compositional primed (+Comp/+ID) targets versus compositional-only primed (+Comp/-ID) targets (SE = 10.9, $P < 0.005$). A significant advantage for identity-and-compositional primed (+Comp/+ID) words compared to non-primed (-Comp/-ID) words of 57.3ms was also found (SE = 10.9, $P < 0.001$). The direct comparison of these conditions did not inform us as to whether the additive effect of composition priming was significant, however, so this effect was not further explored behaviourally. Other pairwise comparisons did not show significant differences between conditions ($P > 0.2$). All p-values were adjusted using the Tukey method for comparing a family of four estimates (Tukey, 1949). For an overview of the pairwise comparison results, see Table 3.

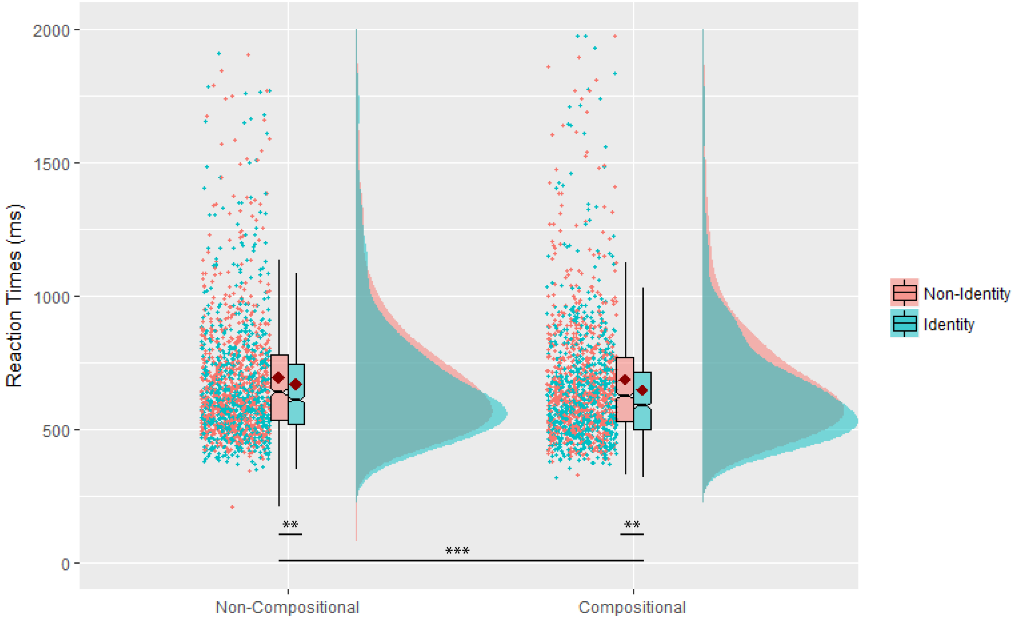


Figure 4. Reaction times are plotted against ID and Comp conditions. Significance bars indicate pairwise comparison effects (** $p < 0.005$; *** $p < 0.001$). Twenty data points were cut-off from the top for visibility purposes. Red dots indicate mean reaction times per condition.

fMRI experiment

A three-way ANOVA was used to assess the main effects of the priming conditions in this study. Composition priming and identity priming were modelled as fixed effects, and subject was modelled as a random effect. The ANOVA results were corrected for multiple comparisons with Puncorr = 0.001; α = 0.05. A cluster size > 703 voxels was determined to be significant. Functional data were mapped onto MNI space, and corresponding region labels are based on the Talairach-Tournoux Atlas (Talairach & Tournoux, 1988). Separate F-statistics were calculated to assess the main effects of Compositional and Identity priming. These were thresholded for multiple comparison correction to a value of F = 19.69 (1,11), Puncorr < 0.001. The left MTG, left superior and inferior frontal gyri (S/IFG), and left basal ganglia were sensitive to compositional priming manipulations (Fig. 5). Significant clusters sensitive to identity priming manipulations were found at the right middle frontal gyrus (MFG) and the right lentiform nucleus (see Fig. 6). There was no evidence for an interaction effect after correction. See Table 4 for an overview of the clusters in the main effects.

Group level t-statistics were calculated to assess significance of several additional contrasts. Due to the small sample size and associated Type II error rate inflation, it is challenging to observe small (ηp2 > 0.02) to medium-sized effects (ηp2 > 0.13; values based on Draper, 2011). When comparing the effect size maps to the ANOVA results, it becomes apparent that some large estimated effects (ηp2 > 0.26) were not found either under the current sample size (Fig. 7). Focus remained on the results that survived the initial correction at Puncorr < 0.001, but we endeavoured to improve our qualitative understanding of the results and potentially inform future work related to this pilot study by also exploring the results at less stringent P-value thresholds (Puncorr < 0.005 and Puncorr < 0.01). Differences in conditions found in this exploration might become statistically significant with a better-powered sample. The minimum cluster size was kept at 703 voxels at these lower P-thresholds. For an overview of all clusters that survived any of the corrections, see Table 5.

Identity-and-compositional versus non-primed. Significant differences in activation were found in response to identity-and-compositional primed versus non-primed targets (+Comp/+ID vs -Comp/-ID) at Puncorr < 0.001 (Fig. 8, orange). Clusters over left S/MTG and bilateral basal ganglia were significantly more active in the identity-and-compositional primed condition over the non-primed condition. At a correction of Puncorr < 0.005 positive clusters over left supramarginal gyrus and IFG also survived correction (Fig. 9, orange).

Compositional-only versus non-primed. Significant differences in activation were found in response to compositional-only primed versus non-primed targets (+Comp/-ID vs -Comp/-ID) at Puncorr < 0.001 (Fig. 8, green). A cluster over the left S/MTG showed significantly more activity in the compositional-only primed condition over the non-primed condition. At a correction of Puncorr < 0.005 a negative cluster over left SFG also survived correction, and the positive cluster over the left MTG extended over to the left IFG (Fig. 9, green). At a correction of Puncorr < 0.01 a positive cluster over the left basal ganglia survived (Fig. 10, green).

Compositional-only versus identity-only. Significant differences in activation were found in response to compositional-only primed versus identity-only primed targets (+Comp/-ID vs -Comp/+ID) at Puncorr < 0.001 (Fig. 11).

Table 2. Overview of the main effects of the mixed model analysis.

Main effect	Beta	SE	t-value
Identity (+ID *vs.* -ID)	-37.21	7.71	-4.83
Composition (+Comp *vs.* –Comp)	-20.10	7.71	-2.61

Table 3. Overview of the pairwise comparisons between the behavioural conditions. Ordered by significance. Italics indicate a significant difference between conditions.

Comparison	Beta	SE	z-ratio	p-value
+Comp/+ID vs. *-Comp/-ID*	*-57.3*	*10.9*	*-5.260*	*<0.0001*
-Comp/+ID vs. *-Comp/-ID*	*-37.6*	*10.9*	*-3.444*	*0.0032*
+Comp/+ID vs. *+Comp/-ID*	*-36.8*	*10.9*	*-3.382*	*0.0040*
+Comp/-ID *vs.* -Comp/-ID	-20.5	11.0	-1.871	0.2405
+Comp/+ID *vs.* -Comp/+ID	-19.7	10.9	-1.816	0.2658
+Comp/-ID *vs.* -Comp/+ID	17.1	10.9	1.568	0.3969

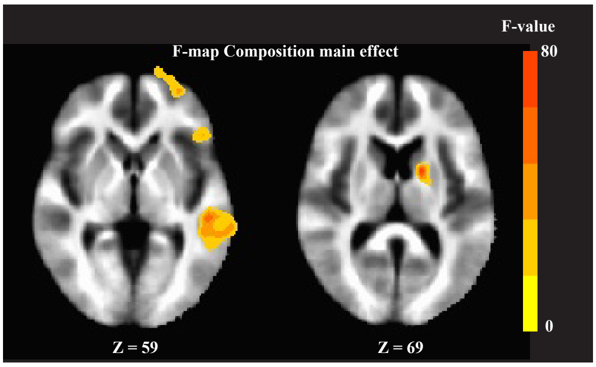


Figure 5. F-map of the Composition priming main effect. Clusters were corrected for multiple comparisons and were significant at Puncorr < 0.001, α = 0.05. Significant clusters after correction appeared over the left MTG, the left IFG, the left SFG (left), and the left basal ganglia (right).

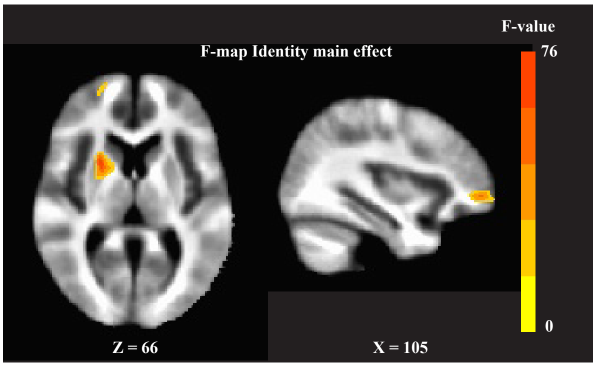


Figure 6. F-map of the Identity priming main effect. Clusters were corrected for multiple comparisons and were significant at Puncorr < 0.001, α = 0.05. Significant clusters after correction appeared over the right lentiform nucleus (left) and the right MFG (right).

Table 4. Overview of significant clusters of the main effects from the rmANOVA, and the locus region of each cluster based on the Talairach-Tournoux Atlas (Talairach & Tournoux, 1988). Hemisphere is given in the third column (L=left, R=right). Peak F-values are given in the fourth column. Peak xyz-coordinates are given in MNI space. No significant clusters were found for the interaction.

Main effect	Locus	L/R	F-value	x	y	z
Composition	Left MTG	L	70.33	-50.3	-21.7	-10.9
	Left SFG	L	39.14	-17.9	65.9	4.7
	Left IFG	L	37.65	-46.7	28.7	-3.7
	Left basal ganglia	L	79.23	-16.7	3.5	10.7
Idenity	Right MFG	R	48.02	-34.9	-53.9	-2.5
	Right lentiform gyrus	R	76.44	-22.9	-8.3	7.1

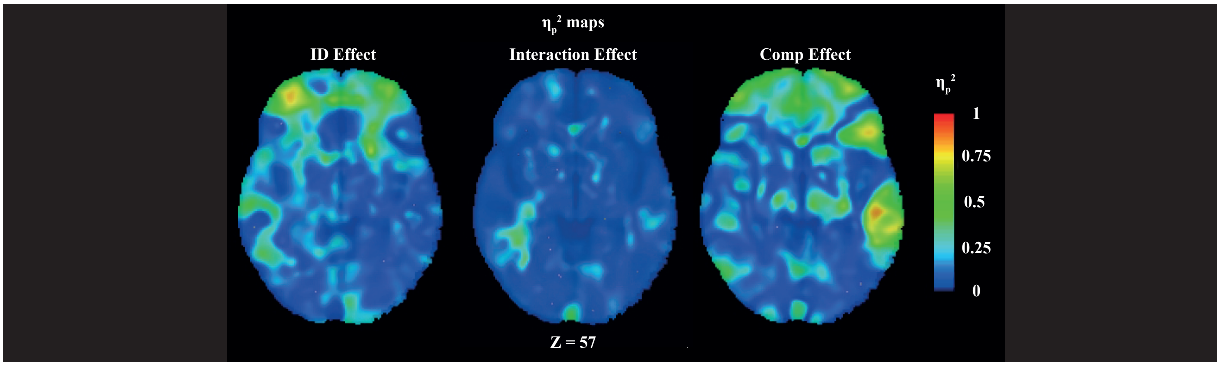


Figure 7. Partial η2 maps for the three main effects. Focussing on the left AG, we see that for the Identity priming main effect (left) and interaction effect (middle), we have a small effect size (~0.25). For the Composition priming effect, we see around a medium effect size (~0.5). N=12 is not enough to reliably detect a potential effect in this region.

Positive clusters over the left MTG and the IFG indicated significantly higher activity in response to compositional-only primed targets compared to identity-only primed targets. Negative clusters over the right AG and the bilateral S/MFG suggest higher activity in these regions in response to identity-only primed targets compared to compositional-only primed targets. At a correction of Puncorr < 0.005 negative clusters over the bilateral MFG dorsal to the significant clusters also survived correction (fig. 12, left). At a correction of Puncorr < 0.01 a negative cluster over the left inferior temporal gyrus (ITG) survived (fig. 12, right).

Identity-and-compositional versus identity-only. No clusters survived the initial correction when comparing activity in response to identity-and-compositional primed targets and identity-only primed targets (+Comp/+ID vs. -Comp/+ID). At Puncorr < 0.005, a positive cluster over the left

Table 5. Overview of surviving clusters in pairwise comparisons. Locus region of the cluster based on the Talairach-Tournoux Atlas (Talairach & Tournoux, 1988) is given in the second column and hemisphere in the third column (L=left, R=right). Peak t-values are given in the fourth column. Peak xyz-coordinates are given in MNI space. The uncorrected P-value thresholds to find the cluster are reported in the last column.

Comparison	Locus	L/R	t-value	x	y	z	P_{uncorr}
+Comp/+ID *vs.* -Comp/-ID	S/MTG	L	6.49	-51.5	-26.5	-2.5	0.001
	basal ganglia	L/R	7.23				0.001
	SMG	L	5.48	-59.9	-50.5	21.5	0.005
	IFG	L	6.11	-35.9	4.7	29.9	0.005
+Comp/-ID vs. -Comp/-ID	MTG	L	9.58	-53.9	-28.9	-2.5	0.001
	SFG	L	-4.67	-28.7	68.3	3.5	0.005
	IFG	L	-				0.005
	basal ganglia	L	5.84	-15.5	3.5	10.7	0.01
+Comp/-ID vs. -Comp/+ID	MTG	L	8.84	-52.7	-30.1	-3.7	0.001
	IFG	L	6.62	-46.7	28.7	-1.3	0.001
	AG	R	-6.35	45.7	-68.5	34.7	0.001
	S/MFG (ventral)	L/R	-5.95				0.001
	MFG (dorsal)	L/R	-6.08				0.005
	ITG	R	-4.46	64.9	-20.5	-19.3	0.01
+Comp/+ID *vs.* –Comp/+ID	MTG	L	5.36	-65.9	-36.1	-0.1	0.005
	SFG	L	-5.15	-16.7	65.9	3.5	0.01
+Comp/+ID *vs.* +Comp/-ID	MFG	R	4.07	37.3	51.5	-4.9	0.01
	PCC		4.78	-3.5	-34.9	23.9	0.01
-Comp/+ID *vs.* –Comp/-ID	No effects						

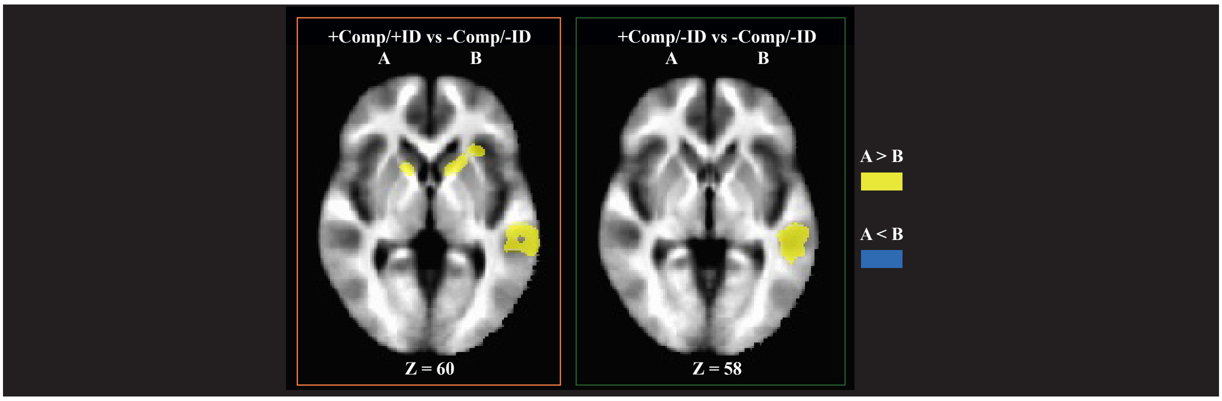


Figure 8. T-maps of the +Comp/+ID vs -Comp/-ID (left; orange) and +Comp/-ID vs -Comp/-ID (right; green) contrasts. Clusters were corrected for multiple comparisons and were significant at Puncorr < 0.001, α = 0.05. Significant clusters appear over the left MTG for both contrasts. In addition, there is a significant cluster over the bilateral basal ganglia in the +Comp/-ID vs -Comp/-ID contrast.

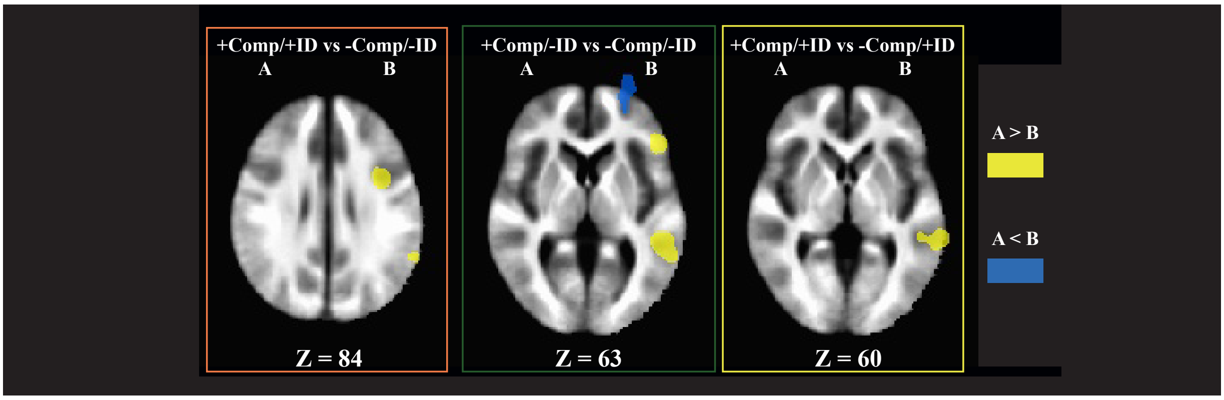


Figure 9. T-maps of the +Comp/+ID vs -Comp/-ID showing the clusters surviving correction at Puncorr < 0.005, α = 0.05 over the left IFG and the SMG (left; orange), +Comp/-ID vs -Comp/-ID showing clusters over the left SFG and the IFG (middle; green), and +Comp/+ID vs -Comp/+ID showing clusters over the left MTG (right; yellow) contrasts.

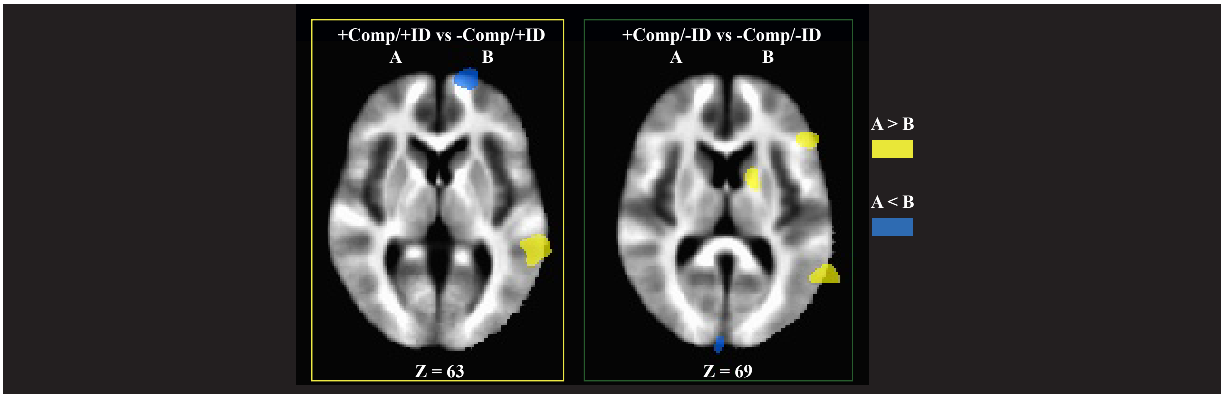


Figure 10. T-maps for +Comp/+ID vs -Comp/+ID (left; yellow) and Com/-ID vs -Comp/-ID contrasts (right; green) corrected at Puncorr < 0.01, α = 0.05. At the left, we see a surviving cluster over left SFG. At the +Comp/-ID vs -Comp/-ID contrast, we see a new cluster over the left basal ganglia.

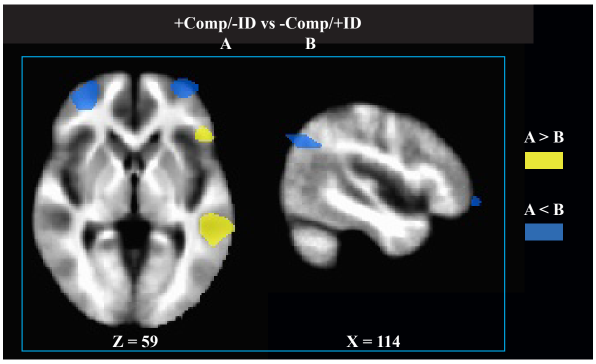


Figure 11. Transversal and sagittal view of the t-map for the +Comp/-ID vs -Comp/+ID contrast (blue). Clusters were corrected for multiple comparisons and were significant at Puncorr < 0.001, α = 0.05. Significant clusters appear over the left MTG, the left IFG, the bilateral STG (left), and the right AG (right).

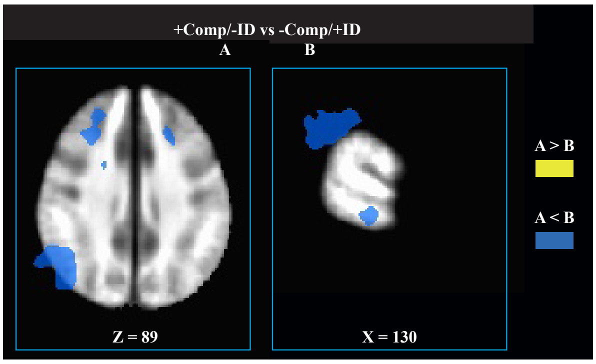


Figure 12. T-maps of +Comp/-ID vs -Comp/+ID contrast corrected at Puncorr < 0.005, α = 0.05 (left) and Puncorr < 0.01, α = 0.05 (right). At Puncorr < 0.005 we see bilateral S/MFG activity that was not present at Puncorr < 0.001 at a more dorsal Z-slice. At Puncorr < 0.01, we see a cluster over right ITG. In both images, we can also see the right AG cluster from Puncorr < 0.001.

MTG survived correction (Fig. 9, yellow), and at Puncorr < 0.01, a negative cluster over the left SFG survived (Fig. 10, yellow).

Identity-and-compositional versus compositional-only. No clusters survived the initial correction when comparing activity in response to identity-and-compositional primed targets and compositional-only primed targets (+Comp/+ID vs. +Comp/-ID). Only at Puncorr < 0.01 two positive clusters survived over the posterior cingulate cortex and the right MFG (Fig. 13).

Discussion

Behavioural experiment

In the lexical decision task we observed faster reaction times to items that were identity primed compared to items that were not identity primed, both as a main effect (Table 2) and as simple effects (Table 3). These results indicated that the paradigm was successful in eliciting a priming response consistent with previous literature (for an overview, see Segaert et al., 2013). A separate effect was found when comparing items primed by sentences to those primed by scrambled sentences, though this effect was less pronounced (see Table 2). The effect could indicate an effect of compositional priming, where the compositional meaning of the sentence semantically primed the target. As such, it shows promise of the expected BOLD response. However, this reaction time effect could also arise from processing difficulties in interpreting scrambles, rather than from the compositional priming effect. Reaction times might be shorter after any sentence than after any scramble. Scrambles may be harder to process, and any subsequent processes

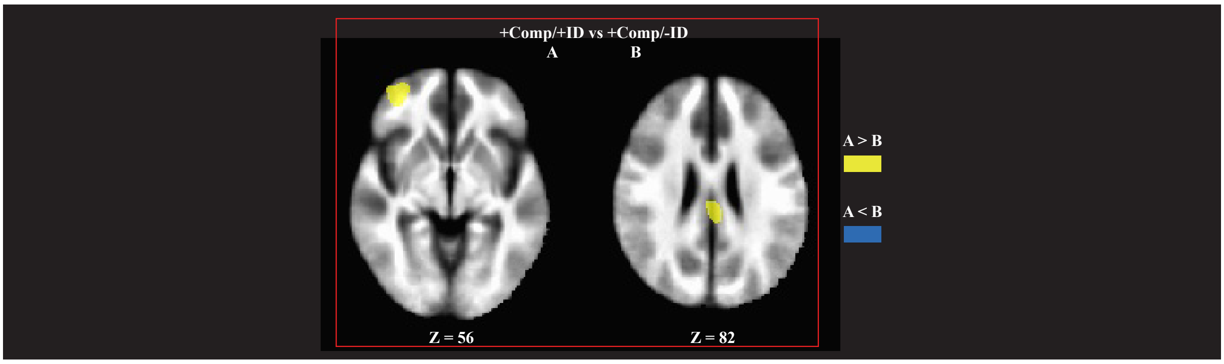


Figure 13. T-maps for +Comp/+ID vs +Comp/-ID corrected at Puncorr < 0.01, α = 0.05. Surviving clusters appear over right MFG and PCC.

could require more time. Further testing is required to comprehend the effects arising from processing hurdles of random word lists to subsequent word processing and compositional priming fully. We believe that, here, we have been able to lay the groundworks for a paradigm that can be utilised to compare the various priming effects that may occur between sentence primes and single-word targets, and could be expanded to study differences between sentence and word-list processing in general.

fMRI experiment

First, it should be noted that any conclusions drawn from this pilot experiment are preliminary and require support from the full experiment. Additionally, the results section discussed findings for corrections at both Puncorr < 0.001 and less stringent P-thresholds. For the discussion, we will primarily focus on the regions that showed significant differences in the pairwise comparisons with a correction at α = 0.05 and Puncorr < 0.001.

When looking at the compositional priming effect, we see that the left MTG strongly responds to the presentation of a target after a compositional prime compared to a non-compositional prime (Fig. 5, 7right, 8 & 11). This is accordance with our prediction that the MTG would be sensitive to compositional priming given its prominent role in the language network (as outlined in the introduction). Responses in the MTG were not sensitive to identity priming (see Table 5), suggesting that this result is more likely attributable to the compositional prime. What is interesting to note is the presence of an enhancement effect in response to compositionally primed targets, rather than a suppression effect. This effect may be similar to the repetition enhancement found in masked prime studies, as explained by Segaert et al. (2013). Schnyer et al. (2002) proposed that a masked prime may lead to enhanced fMRI signal if "activation from the masked prime continues to spread until the target is identified" (Segaert et al., 2013). Segaert et al. (2013) then propose that the response to the prime and the response to the target are additive, resulting in enhancement. A similar principle may be a at play here, where the spread of activation from the priming sentence increases the fMRI signal related to the target, in addition to the target activation itself, indicating successful compositional priming.

Somewhat surprisingly, we found little evidence for the identity priming effect in the left temporal cortex. There was no main effect in left MTG of +ID vs -ID (Table 4), or a pairwise effect (row 5 & 6 in Table 5). Since repetition priming is quite a well-established effect (e.g., Bergerbest et al., 2006; Schnyer et al., 2002; Elger et al., 1997), we expected that this response would be stronger in the identity comparisons than in the composition comparisons. Especially since the effect is absent from the -Comp/+ID with -Comp/-ID comparison, where no other priming effects should attenuate the identity priming effect, it seems that the current setup is unable to catch both the perceptual and semantic aspects of the identity priming effect. It could be that the task redirected the attention of the subjects away from the identity tokens, and as such, there were no differential responses to the identity tokens (for the effects of attention on visual priming, see Vuilleumier et al., 2006; Thoma & Henson, 2011). For example, Schnyer et al. (2002) found an identity priming effect at the word-level using a lexical decision task, which required lexical access to perform, whereas our task did not necessarily. If subjects had been asked during catch trials whether the target was present in the preceding string, perhaps a stronger identity priming effect would have been found. While a different task for the fMRI experiment should be considered for the main study, it should be pointed out that a task drawing attention to target words might consequently reduce attention to the unity within the primes, and as such potentially eliminate the compositional priming effect. A task like the lexical decision task would also reduce the power of the experiment by reducing the number of trials of interest. The challenge lies in finding a task that requires attention to both elements of each trials without disqualifying trials from subsequent analysis.

The left IFG also appeared to respond more strongly to the compositionally primed targets than the non-compositionally primed targets (Fig. 5; Table 4). If theleft IFG activity is related to the composition of linguistic input, and if the output of the IFG is subsequently relayed back to the left MTG, then it stands to reason that the IFG activates when a semantically related target to the complex prime is presented. No effect was expected for the identity vs non-identity comparisons in the IFG.

Somewhat surprising was the reduced response in the left S/MFG to compositionally primed targets vs non-compositionally primed targets (Table 5, rows 2-4). The S/MFG have been associated with working memory demands and cognitive control (Du Boisgueheneuc et al., 2006; Li et al., 2013; Emch et al., 2019). This could indicate that a reduced S/MFG activity in response to a compositionally primed target may be due to reduced demands to retrieve an item from working memory, as seen in all but one +Comp vs -Comp contrast (Table 5). Given that our task did not require overt attention to the presence of identity primes, it seems that this is an automatic response. In other words, if the prime is incoherent, the brain automatically allocates more resources to remembering its contents and retrieving it, whether task-relevant or not. Here, this effect is found to be bilateral and mostly seen in the compositional vs. non-compositional comparison. However, we also see this effect in the right MFG when comparing +Comp/+ID to +Comp/-ID. Here it appears that the right MFG responds stronger when an item may be present in working memory than when it is not, and may as such reflect a recollection process related to verbal working memory (Emch et al., 2019). The right MFG has also been associated with redirection of attention (Japee et al., 2015), which might indicate that the presentation of the target causes subjects to reprioritise the focus of the prime.

An unexpected result from this pilot was that when comparing the response to an exclusively compositionally primed target with the response to an exclusively identity primed target (+Comp/-ID vs. -Comp/+ID), we found evidence for the right AG and the ITG response (Fig. 11). While AG activity was expected in this condition, we expected it in the left hemisphere, as the left hemisphere is usually more strongly associated with linguistic processing. What is also surprising is that activity appears to be stronger for the identity primed condition than the compositionally primed condition. Since compositional priming has led to enhancement of the BOLD response in the other comparisons, it might be expected here. It is perfectly possible that different brain regions respond to the same prime differently, or it might be that this is an effect driven by the identity prime. Since this response does not show up for any of the other comparisons, it is difficult to tell to which condition the right AG is sensitive. It may also be, given the ‘pure’ priming conditions in our experiment, that the right AG responds to a condition only if the other priming stream is absent, and that once multiple paths of relating input become available, its involvement in the process is affected.

Right ITG activity may be a result of the predictability of the target word. Bonhage et al. (2015) found a correlation between the right ITG activity and word predictions in semantically rich contexts. Contrary to our results however, Bonhage et al. (2015) found that the right ITG was more active for a sentence condition compared to a jabberwocky condition where only word category could be predicted. It is unclear how these results may be reconciled with our findings. Right ITG activity may also be related to a type of repetition. Schweinberger et al. (2002) found that right ITG responses related to familiar face repetitions in an ERP study. Similarly, Li et al. (1993) found an inferior temporal response to familiarity of stimuli in rhesus monkeys. Perhaps this effect could be extended to familiar word repetitions, but this effect does not seem to be reported in linguistic contexts. It would be interesting to see if in a follow-up this effect might pop up in different conditions.

Basal ganglia activity was found bilaterally in the +Comp/+ID vs. -Comp/-ID and left lateralised in +Comp/-ID vs. -Comp/-ID. Basal ganglia involvement was expected, but no strong predictions were made because basal ganglia have been found to be involved in a number of linguistic (Copland et al., 2000; Bacon Moore et al., 2013) and non-linguistic tasks (Rao et al., 1997; Haaland et al., 1997). The effect that was found may be a verbal working memory process (Bacon Moore et al., 2013) involved in complex input processing. Interestingly, the main effects show that left and right basal ganglia are responsive to only compositional and identity priming respectively (Fig. 5 & 6). These main effects are not reflected in the pairwise comparisons, however. This makes it difficult to draw any conclusions about basal ganglia activity, as it is unclear which condition drives the basal ganglia effects in our analysis.

An alternative interpretation for our results arises not from a compositional priming effect, but rather simply due to participants processing sentence level information while reading the primes. We see that the language network responds in a fashion similar to on-line sentence processing (e.g., as modelled by Hagoort, 2016 or Friederici, 2011). The question is whether the response to the compositionally primed targets should be any different from the activity in on-line sentence processing. We presented participants with a one-word compositional summary of a sentence they previously read. Readers may relate this word back to the previous context, similar to how new words in an input are inserted into the previous context in Hagoort (2016) and Friederici (2011), involving comparable neural processes. For future analyses, it may be worthwhile to model prime presentation as a nuisance regressor to eliminate potential effects from the prime processing. We intend include this regressor in the study based on this pilot, but we do not expect this to have a detrimental effect on the results.

Through this pilot study, we were able to establish brain regions that likely play a role in semantic composition. The regions responsive to the manipulations in this study provide a starting point for a more thorough study on semantic composition processes. The left MTG and the IFG were expected candidates, but we also found bilateral S/MFG responses and potentially right AG and ITG activity related to processing primed targets. In the follow-up laminar study, these regions will hopefully show depth-dependant interactions with one another that would further increase our understanding of sentence processing. It remains to be discovered at what cortical depth the haemodynamic responses originate that drive the IFG and the MTG signal in this study, but the prediction that left IFG activity shows a correlation to deep-bin MTG activity still holds. If this correlation is found in the follow-up laminar fMRI study, it would be indicative of the predicted semantic composition process and top-down priming effect (as proposed by Hagoort, 2016), and would suggest that this stream of information is important for the conceptualisation of composed semantics. Any effect of right AG and bilateral S/MFG correlated to the compositional and especially identity priming effects would support a model for a bilateral language network that involves the right hemisphere at the word-level.

Conclusion

We have conducted a pilot study to see if it is possible to examine the processes involved in compositional meaning extraction of sentences. We used a priming paradigm in order to do so, where the target was either compositionally primed, identity primed, both, or neither. We found a strong effect of identity priming in a behavioural lexical decision task, as well as a smaller potential effect of compositional priming. We generated effect-size maps for each factor and pairwise contrast in our fMRI analysis, which can be utilised to estimate the sample needed for ROI selection in a follow-up laminar fMRI experiment. We found evidence for the existence of a neural compositional priming effect in left MTG, IFG and S/MFG. We also found an effect that could suggest a right-lateralised identity priming effect in S/MFG, AG and ITG. These findings will be taken as a starting point into a follow-up laminar fMRI experiment focussed on studying directed connectivity between these regions.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). The Variability of Human, BOLD Hemodynamic Responses. *NeuroImage, 8*(4), 360–369.

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. (2018). Raincloud plots: a multi-platform tool for robust data visualization. *PeerJ Preprints, 6*(e27137v1).

Bacon Moore, A., Li, Z., Tyner, C. E., Hu, X., & Crosson, B. (2013). Bilateral basal ganglia activity in verbal working memory. *Brain and Language, 125*(3), 316–323.

Baggio, G., & Hagoort, P. (2011). The balance between memory and unification in semantics: A dynamic account of the N400. *Language and Cognitive Processes, 26*(9), 1338–1367.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. E. (2006). Neural Correlates of Auditory Repetition Priming: Reduced fMRI Activation in the Auditory Cortex. *Journal of Cognitive Neuroscience, 16*(6), 966–977.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where Is the Semantic System? A Critical Review and Meta-Analysis of 120 Functional Neuroimaging Studies. *Cerebral Cortex, 19*(12), 2767–2796.

Bonahge, C. E., Mueller, J. L., Friederici, A. D., & Fiebach, C. J. (2015). Combined eye tracking and fMRI reveals neural basis of linguistic predictions during sentence comprehension. *Cortex, 68*, 33–47.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry, 25*(1), 49–59.

Brown, S. P., & Hestrin, S. (2009). Intracortical circuits of pyramidal neurons reflect their long-range axonal targets. *Nature, 457*, 1133–1136.

Buckner, R. L., Koutstaal, W., Schacter, D. L., & Rosen, B. R. (2000). Functional MRI evidence for a role of frontal and inferior temporal cortex in amodal components of priming. *Brain, 123*(3), 620–640.

Cohen, J. (1973). Eta-Squared and Partial Eta-Squared in Fixed Factor ANOVA Designs. *Educational and Psychological Measurement, 33*, 107–112.

Copland, D. A., Chenery, H. J., & Murdoch, B. E. (2000). Processing lexical ambiguities in word triplets: evidence of lexical-semantic deficits following dominant nonthalamic subcortical lesions. *Neuropsychology, 14*(3), 379–390.

Cox, R. W. (1996). AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages. *Computers and Biomedical Research, 29*(3), 162–173.

Dale, A. M. (1999). Optimal experimental design for event related fMRI. *Human Brain Mapping, 8*(2–3), 109–114.

Dale, A. M., Greve, D. N., & Burock, M. A. (1999). Optimal Stimulus Sequences for Event-Related fMRI. Presented at the 5th International Conference on Functional Mapping of the Human Brain., Duesseldorf, Germany.

Davey, J., Thompson, H. E., Hallam, G., Karapanagiotidis, T., Murphy, C., De Caso, I., … Jefferies, E. (2016). Exploring the role of the posterior middle temporal gyrus in semantic cognition: Integration of anterior temporal lobe with executive processes. *NeuroImage, 137*, 165–177.

de C. Hamilton, A. F., & Grafton, S. T. (2008). Repetition suppression for performed hand gestures revealed by fMRI. *Human Brain Mapping, 30*(9), 2898–2906.

Dehaene, S., & Cohen, L. (2011). The unique role of the visual word form area in reading. *Trends in Cognitive Sciences, 15*(6), 254–262.

den Heyer, K., Goring, A., & Dannenbring, G. L. (1985). Semantic priming and word repetition: The two effects are additive. *Journal of Memory and Language, 24*(6), 699–716.

Dimitropoulou, M., Duñabeitia, J. A., Avilés, A., Corral, J., & Carreiras, M. (2010). Subtitle-based word frequencies as the best estimate of reading behavior: the case of greek. *Frontiers in Psychology, 1*, 218.

Draper, S. (2011). Effect size [WWW document]. Retrieved from http://www.psy.gla.ac.uk/~steve/best/effect.html

du Boisgueheneuc, F., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehun, S., … Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: a lesion study. *Brain, 129*(12), 3315–3328.

Elger, C. E., Grunwald, T., Lehnertz, K., Kutas, M., Helmstaedter, C., Brockhaus, A., … Heinze, H. J. (1997). Human temporal lobe potentials in verbal learning and memory processes. *Neuropsychologia, 35*(5), 657–667.

Emch, M., von Bastian, C. C., & Koch, K. (2019). Neural Correlates of Verbal Working Memory: An fMRI Meta-Analysis. *Frontiers in Human Neuroscience, 13*, 180.

Ettinger, A., Feldman, N., Resnik, P., & Phillips, C. (2016). Modeling N400 amplitude using vector space models of word representation. *CogSci.*

Ferreira, F., & Lowder, M. W. (2016). Prediction, Information Structure, and Good-Enough Language Processing. *Psychology of Learning and Motivation, 65*, 217–247.

Fischl, B. (2012). FreeSurfer. *NeuroImage, 62*(2), 774–781.

Frazier, L., & Clifton, C. J. (1996). *Construal.* Cambridge, MA, USA: The MIT Press.

Frazier, L., & Clifton, C. J. (1997). Construal: Overview, Motivation, and Some New Evidence. *Journal of Psycholinguistic Research, 26*(3), 277–295.

Friederici, A. D. (2011). The Brain Basis of Language Processing: from Structure to Function. *Physiological Review, 91*, 1357–1392.

Geuter, S., Qi, G., Welsh, R. C., Wager, T. D., & Lindquist, M. A. (2018). Effect Size and Power in fMRI Group Analysis. *BioRxiv.*

Glover, G. H., Li, T.-Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic Resonance in Medicine, 44*(1), 162–167.

Haaland, K. Y., Harrington, D. L., O'Brien, S., & Hermanowicz, N. (1997). Cognitive–Motor Learning in Parkinson's Disease. *Neuropsychology, 11*(2), 180–186.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in Cognitive Sciences, 9*(9), 416–423.

Hagoort, P. (2013). MUC (Memory, Unification, Control) and beyond. *Frontiers in Psychology, 4*(416).

Hagoort, P. (2016). MUC (Memory, Unification, Control): A Model on the Neurobiology of Language Beyond Single Word Processing. In G. Hickok & S. L. Small (Eds.), *Neurobiology of Language* (pp. 339–347). Academic Press.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of Word Meaning and World Knowledge in Language Comprehension. *Science, 304*(5669), 438–441.

Harris, K. D., & Mrsic-Flogel, T. D. (2013). Cortical connectivity and sensory coding. *Nature, 503*,51–58.

Hedge, L. V. (1981). Distribution Theory for Glass's Estimator of Effect Size and Related Estimators. *Statistics, 6*(2), 107–128.

Hermans, E., & van de Laar, T. (2018). HERA: Heart Rate Analysis (Version 0.06 beta).

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and Semantic Modulation of Neural Activity during Auditory Sentence Comprehension. *Journal of Cognitive Neuroscience, 18*(4), 665–679.

Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2007). Time course of semantic processes during sentence comprehension: An fMRI study. *NeuroImage, 36*(3), 924–932.

Hutchison, K. S. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review, 10*(4), 785–813.

Japee, S., Holiday, K., Satyshur, M. D., Mukai, I., & Ungerleider, L. G. (2015). A role of right middle frontal gyrus in reorienting of attention: a case study. *Frontiers in Systems Neuroscience, 9*, 23.

Kastrup, A., Krüger, G., Glover, G. H., Neumann-Haefelin, T., & Moseley, M. E. (1999). Regional Variability of Cerebral Blood Oxygenation Response to Hypercapnia. *NeuroImage, 10*(6), 675–681.

Keefe, D. E., & Neely, J. H. (1990). Semantic priming in the pronunciation task: The role of prospective prime-generated expectancies. *Memory & Cognition, 18*(3), 289–298.

Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(9), 1470–1489.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods, 42*(3), 627–633.

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods, 42*(3), 643–650.

Kok, P., Bains, L. J., van Mourik, T., Norris, D. G., & de Lange, F. P. (2016). Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Current Biology, 26*(3), 371–376.

Koopmans, P. J., Barth, M., & Norris, D. G. (2010). Layer specific BOLD activation in human V1. *Human Brain Mapping, 31*(9), 1297–1304.

Lai, V. T., Willems, R. M., & Hagoort, P. (2015). Feel between the Lines: Implied Emotion in Sentence Comprehension. *Journal of Cognitive Neuroscience, 27*(8), 1528–1541.

Lapesa, G., & Evert, S. (2013). Evaluating Neighbor Rank and Distance Measures as Predictors of Semantic Priming. *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, 66–74.

Lapesa, G., & Evert, S. (2014). A Large Scale Evaluation of Distributional Semantic Models: Parameters, Interactions and Model Selection. *Transactions of the Association for Computational Linguistics, 2*, 531–545.

Lefort, S., Tomm, C., Floyd Sarria, J.-C., & Petersen, C. C. H. (2009). The Excitatory Neuronal Network of the C2 Barrel Column in Mouse Primary Somatosensory Cortex. *Neuron, 61*(2), 301–316.

Lenci, A. (2018). Distributional Models of Word Meaning. *Annual Review of Linguistics, 4*, 151–171.

Lenth, R. (2019). emmeans: Estimated Marginal Means, aka Least-Squares Means. Retrieved from https://CRAN.R-project.org/package=emmeans

Li, L., Miller, E. K., & Desimone, R. (1993). The representation of stimulus familiarity in anterior inferior temporal cortex. *Journal of Neurophysiology, 69*(6), 1918–1929.

Li, W., Qin, W., Liu, H., Fan, L., Wang, J., Jiang, T., & Yu, C. (2013). Subregions of the human superior frontal gyrus and their connections. *NeuroImage, 78*, 46–58.

Limesurvey GmbH. (n.d.). LimeSurvey: An Open Source survey tool (Version 3.19.2+191017). Hamburg, Germany: Limesurvey GmbH. Retrieved from http://www.limesurvey.org

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502.

Lyu, B., Choi, H. S., Marslen-Wilson, W. D., Clarke, A., Randall, B., & Tyler, L. K. (2019). Neural dynamics of semantic composition. *PNAS, 116*(42), 21318–21327.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703.

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language, 92*, 57–78.

Marques, J. P., Kober, T., Krueger, G., van der Zwaag, W., van de Moortele, P. F., & Greutter, R. (2010). MP2RAGE, a self bias-field corrected sequence for improved segmentation and T1-mapping at high field. *NeuroImage, 49*(2), 1271–1281.

McGettigan, C., & Scott, S. K. (2012). Cortical asymmetries in speech perception: what's wrong, what's right and what's left? *Trends in Cognitive Sciences, 16*(5), 269–276.

New, B., Brysbaert, M., Veronis, J., & Pallier, C. (2007). The use of film subtitles to estimate word frequencies. *Applied Psycholinguistics, 28*(4), 661–677.

Olman, C. A., Harel, N., Feinberg, D. A., He, S., Zhang, P., Ugurbil, K., & Yacoub, E. (2012). Layer-Specific fMRI Reflects Different Neuronal Computations at Different Depths in Human V1. *PLOS One.*

Oosterdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2012). The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In P. Spyns & J. Odijk (Eds.), *Essential Speech and Language Technology for Dutch. Theory and Applications of Natural Language Processing.* Berlin, Heidelberg: Springer.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences, 3*(4), 136–141.

R Core Team. (2018). R: A language and environment for statistical computing (Version R version 3.5.2 (2018-12-20), Eggshell Igloo) [X86_64, mingw32]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rao, S. M., Harrington, D. L., Haaland, K. Y., Bobholz,

J. A., Cox, R. W., & Binder, J. R. (1997). Distributed Neural Systems Underlying the Timing of Movements. *The Journal of Neuroscience, 17*(14), 5528–5535.

Raposo, A., Moss, H. E., Stamatakis, E. A., & Tyler, L. K. (2006). Repetition suppression and semantic enhancement: An investigation of the neural correlates of priming. *Neuropsychologia, 44*(12), 2284–2295.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506.

Rissman, J., Eliassen, J. C., & Blumstein, S. E. (2003). An Event-Related fMRI Investigation of Implicit Semantic Priming. *Journal of Cognitive Neuroscience, 15*(8), 1160–1175.

Roskies, A. L., Fiez, J. A., Balota, D. A., Raichle, M. E., & Peterson, S. E. (2001). Task-Dependent Modulation of Regions in the Left Inferior Frontal Cortex during Semantic Processing. *Journal of Cognitive Neuroscience, 13*(6), 829–843.

Rossell, S. L., Price, C. J., & Nobre, A. C. (2003). The anatomy and time course of semantic priming investigated by fMRI and ERPs. *Neuropsychologia, 41*(5), 550–564.

Rugg, M. D. (1985). The Effects of Semantic Priming and Word Repetition on Event-Related Potentials. *Psychophysiology, 22*(6), 642–647.

Saad, Z. S., Glen, D. R., Chen, G., Beauchamp, M. S., Desai, R., & Cox, R. W. (2009). A new method for improving functional-to-structural alignment using local Pearson correlation. *NeuroImage, 44*, 838–848.

Schnyer, D. M., Ryan, L., Trouard, T., & Forster, K. (2002). Masked word repetition results in increased fMRI signal: a framework for understanding signal changes in priming. *Neuroreport, 13*(3), 281–284.

Schweinberger, S. R., Pickering, E. C., Jentzsch, I., Burton, A. M., & Kaufmann, J. M. (2002). Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Cognitive Brain Research, 14*(3), 398–409.

Segaert, K., Weber, K., de Lange, F. P., Petersson, K. M., & Hagoort, P. (2013). The suppression of repetition enhancement: A review of fMRI studies. *Neuropsychologia, 51*(1), 59–66.

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., & Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *NeuroImage, 22*(3), 1060–1075.

Sharoh, D., van Mourik, T., Bains, L. J., Segeart, K., Weber, K., Hagoort, P., & Norris, D. G. (2019). Laminar Specific fMRI Reveals Directed Interactions in Distributed Networks During Language Processing. *PNAS, 116*(42), 21185–21190.

Snijders, T. M., Petersson, K. M., & Hagoort, P. (2010). Effective connectivity of cortical and subcortical regions during unification of sentence structure. *NeuroImage, 52*(4), 1633–1644.

Snijders, T. M., Vosse, T., Kempen, G., van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and Unification of Syntactic Structure in Sentence Comprehension: an fMRI Study Using Word-Category Ambiguity. *Cerebral Cortex, 19*(7), 1493–1503.

Stirnberg, R., Huijbers, W., Brenner, D., Poser, B. A., Breteler, M., & Stöcker, T. (2017). Rapid whole-brain resting-state fMRI at 3 T: Efficiency-optimized three-dimensional EPI versus repetition time-matched simultaneous-multi-slice EPI. *NeuroImage, 163*, 81–92.

Summerfield, C., Trittschuh, E. H., Monty, J. M., Mesulam, M.-M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience, 11*, 1004–1006.

Talaraich, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.

Thoma, V., & Henson, R. N. (2011). Object representations in ventral and dorsal visual streams: fMRI repetition effects depend on attention and part–whole configuration. *NeuroImage, 57*(2), 513–525.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *PNAS, 94*(26), 14792–14797.

Tukey, J. W. (1949). Comparing Individual Means in the Analysis of Variance. *Biometrics, 5*(2), 99–114. Retrieved from http://www.jstor.org/stable/3001913

Turken, A. U., & Dronkers, N. F. (2011). The Neural Architecture of the Language Comprehension Network: Converging Evidence from Lesion and Connectivity Analyses. *Frontiers in Systems Neuroscience, 5*(1).

Visser, M., Jefferies, E., Embleton, K. V., & Lambon Ralph, M. A. (2012). Both the Middle Temporal Gyrus and the Ventral Anterior Temporal Area Are Crucial for Multimodal Semantic Processing: Distortion-corrected fMRI Evidence for a Double Gradient of Information Convergence in the Temporal Lobes. *Journal of Cognitive Neuroscience, 24*(8), 1766–1778.

Vosse, T., & Kempen, G. (2000). Syntactic structure assembly in human parsing: a computational model based on competitive inhibition and a lexicalist grammar. *Cognition, 75*(2), 105–143.

Vuilleumier, P., Schwartz, S., Duhoux, S., Dolan, R. J., & Driver, J. (2005). Selective Attention Modulates Neural Substrates of Repetition Priming and "Implicit" Visual Memory: Suppressions and Enhancements Revealed by fMRI. *Journal of Cognitive Neuroscience, 17*(8), 1245–1260.

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2010). The Neural Organization of Semantic Control: TMS Evidence for a Distributed Network in Left Inferior Frontal and Posterior Middle Temporal Gyrus. *Cerebral Cortex, 21*(5), 1066–1075.

Wible, C. G., Han, S. D., Spencer, M. H., Kubicki, M., Niznikiewicz, M. H., Jolesz, F. A., … Nestor, P. (2006). Connectivity among semantic associates: An fMRI study of semantic priming. *Brain and Language, 97*(3), 294–305.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag. Retrieved from http://ggplot2.org

Xiang, H.-D., Fonteijn, H. M., Norris, D. G., & Hagoort, P. (2010). Topographical Functional Connectivity Pattern in the Perisylvian Language Networks. *Cerebral Cortex, 20*(3), 549–560.

Zhu, Z., Zhang, J. X., Wang, S., Xiao, Z., Huang, J., & Chen, H.-C. (2009). Involvement of left inferior frontal gyrus in sentence-level semantic integration. *NeuroImage, 47*(2), 756–763.

Supplementary information

Lists of the used stimuli can be found in the publication on the CNS Journal website: https://www.ru.nl/master/cns/journal.