Composition of complex numbers:

Delineating the computational role of the left anterior temporal lobe

Esti Blanco-Elorrieta[1*] and Liina Pylkkänen[1,2,3]

1. NYUAD Institute, New York University Abu Dhabi, Abu Dhabi, UAE
2. Department of Linguistics, New York University, New York, NY, USA
3. Department of Psychology, New York University, New York, NY, USA

Please address correspondence to:
Esti Blanco-Elorrieta
New York University Abu Dhabi
Neuroscience of Language Lab
NYUAD Saadiyat Campus, A2
P.O. Box 129188, Abu Dhabi, UAE
estibaliz.blanco@nyu.edu

Highlights

- We studied the involvement of the left ATL in the construction of complex numbers.
- Productions of three types of combinatorial phrases were compared to list controls.
- Complex numbers and adjectival modifications engaged the LATL, quantifiers did not.
- The LATL supports the composition of complex concepts, including complex numbers.
- The AG was recruited for all phrases requiring the construction of a plural noun.

ABSTRACT

What is the neurobiological basis of our ability to create complex messages with language? Results from multiple methodologies have converged on a set of brain regions as relevant for this general process, but the computational details of these areas remain to be characterized. The left anterior temporal lobe (LATL) has been a consistent node within this network, with results suggesting that although it rather systematically shows increased activation for semantically complex structured stimuli, this effect does not extend to number phrases such as 'three books.' In the present work we used magnetoencephalography to investigate whether numbers in general are an invalid input to the combinatory operations housed in the LATL or whether the lack of LATL engagement for stimuli such as 'three books' is due to the quantificational nature of such phrases. As a relevant test case, we employed complex number terms such as 'twenty-three,' where one number term is not a quantifier of the other but rather, the two terms form a type of complex concept. In a number naming paradigm, participants viewed rows of numbers and depending on task instruction, named them as complex number terms ('twenty-three'), numerical quantifications ('two threes'), adjectival modifications ('blue threes') or non-combinatory lists (e.g., 'two, three'). While quantificational phrases failed to engage the LATL as compared to non-combinatory controls, both complex number terms and adjectival modifications elicited a reliable activity increase in the LATL. Our results show that while the LATL does not participate in the enumeration of tokens within a set, exemplified by the quantificational phrases, it does support conceptual combination, including the composition of complex number concepts.

Keywords: Conceptual combination, Magnetoencephalography, Left anterior temporal lobe, Numerical Quantification, Plural representation.

1- INTRODUCTION

Understanding the brain basis of linguistic creativity is a fundamental goal for the cognitive neuroscience of language: what is the neurobiology of our ability to create an infinity of conceptual representations from the basic building blocks of language? Large networks of brain areas have been proposed to partake in the brain's "semantic network" (Binder et al., 2009; Binder & Desai, 2011) including the left inferior frontal cortex (e.g., Hagoort & Indefrey, 2014), the superior temporal gyrus (e.g., Friederici, 2011), the angular gyrus (e.g., Price et al., 2015) and the left anterior temporal lobe (LATL). While the computational details of these various network nodes are still mostly not understood, a broad methodologically diverse and internally consistent body of work strongly implicates the LATL as a basic site for semantic combination. Core evidence for this include hemodynamic and neuropsychological research proposing that this brain area acts as a 'semantic hub' in which conceptual representations are bound together and processed by a common set of neurons (Bright, Moss, & Tyler, 2004; Clarke, Taylor, & Tyler, 2011; Clarke, Taylor, Devereux, Randall, & Tyler, 2013; Gauthier, Anderson, Tarr, Skudlarski, & Gore, 1997; Grabowski et al., 2001; Rogers et al., 2006; Tyler et al., 2004) as well as sentence processing studies showing that structured sentences elicit greater LATL activity than meaningless sentences or word lists (Friederici, Meyer, & von Cramon, 2000; Humphries, Binder, Medler, & Liebenthal, 2006, 2007; Mazoyer et al., 1993; Pallier, Devauchelle, & Dehaene, 2011; Rogalsky & Hickok, 2009; Stowe et al., 1998; Vandenberghe, Nobre, & Price, 2002; Xu, Kemeny, Park, Frattali, & Braun, 2005). More recently, magnetoencephalography (MEG) studies on minimal combinations of two words have demonstrated that this activity relates to very basic combinatory operations as opposed to sentence-level phenomena (Bemis, & Pylkkänen, 2013; Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis, & Blanco Elorrieta, 2014).

While this large dataset is still compatible with many definitions of "semantic processing", the robustness of these findings and their generality across multiple methodologies presents an opportunity for a systematic investigation of the computational details of this activity. One step towards sharpening our understanding involves recent MEG results on language production (Del Prato & Pylkkänen, 2014), where the modification of object denoting nouns with color adjectives (*blue cups*) engaged the LATL, while numerical quantification of the same nouns (*two cups*) did not. Given that both of these combinations involve semantic composition, these data are incompatible with a general semantic composition account of the LATL. Instead, they suggest a narrower computation, perhaps better characterized as a type of "conceptual combination," a label employed in the concepts and categories literature for a host of cases where, intuitively, the combination of two concepts serves to form a more complex one, typical examples being adjective-noun and noun-noun combinations. Given that in phrases such as *two cups*, *two* does not add a feature to the concept denoted by *cup* but rather enumerates the number of tokens in a set of cups, such cases would, by hypothesis, fall outside the definition of conceptual combination that is relevant for the LATL. Related evidence for the conceptual nature of the LATL include the sensitivity of its combinatory response to conceptual specificity (Westerlund & Pylkkänen, 2014) and the correlation between the LATL activation elicited by specific concepts like *boy* and the product of the activations for their constituent concepts (i.e., *male* and *child*) (Baron & Osherson, 2011).

The purpose of the current experiment was to further characterize which input elements and specific computations constitute the "conceptual combinations" which drive activity within the LATL**.** Specifically, our study was designed around the question of whether complex number terms, such as *thirty-two*, would elicit combinatory activity in the LATL, despite its insensitivity to numerical quantification. Since this study builds on the results of Del Prato and Pylkkänen (2014), which was conducted in production, the current study is also a production study. Importantly, previous MEG research (Pylkkänen, Bemis & Blanco-Elorrieta, 2014) has shown that functionally similar combinatorial mechanisms operate in both production and comprehension, including effects in the LATL. Further, as described in Methods, our production paradigm allowed us to keep the physical stimulus almost completely constant across conditions (cf., Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis & Blanco-Elorrieta, 2014), which was particularly useful given that confounding low level factors are often an issue in language studies. In general, our research on the production of minimal combinatory phrases takes advantage of the fact that the syntactic and semantic planning of small two-word phrases is thought to occur entirely prior to the onset of articulation (and accompanying motion artifacts) and thus with a technique capable of capturing these planning stages millisecond-by-millisecond, we are able to begin characterize the detailed convolution of the processes leading from picture onset to articulation (Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis & Blanco-Elorrieta, 2014).

Behavioral research on conceptual combination has classically been quite focused on one particular domain; the modification of nouns (e.g., Medin & Shoben, 1988; Murphy, 1990; Wisniewski, 1996; Hampton, 1997). Given that the LATL is at least activated by the core cases of conceptual combination, as evidenced by the many studies on adjective-noun combinations (Bemis, & Pylkkänen, 2013; Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis, & Blanco Elorrieta, 2014; Westerlund & Pylkkänen, 2014), it now becomes possible to concretely test what types of semantic combinations drive this activity. In other words, what is the brain's definition of "conceptual combination"?

Number words are a particularly interesting test case for this purpose as they are a very multifaceted word class in terms of the position and semantic functions they can fulfill in a sentence (Hurford, 1975). The most widely spread view states that (simplex) cardinals such as 'one', 'two' and 'three' are determiners (Barwise & Cooper 1981; Bennett 1975; Montague 1974; Scha 1981) and they have traditionally been treated either as generalized quantifiers (Montague, 1974; Barwise & Cooper 1981) or restrictive modifiers (Link, 2002) when they precede the noun. However, according to Hurford (1975, 1987, 2001, 2003) and Ionin and Matushansky (2006): "when not acting as modifiers, the vast majority of simplex cardinals are singular nouns and belong to one or another open lexical class available in a language". Therefore, number words do not fall clearly in either open or close class word categories and can interestingly occupy the place of both in a noun phrase. This unique feature provides the opportunity to create different combinations and investigate to which extent the conceptual details of the input elements matter by creating a number of instinctively different combinations, while keeping the input elements constant.

The purpose of this experiment was to develop some understanding of the bounds and generality of the computations performed in the LATL regarding exactly what types of representations it combines. Particularly, as numerical quantification did not elicit conceptual combination in the LATL (Del Prato and Pylkkänen, 2014), our focus was on assessing whether this was because the LATL does not perform quantificational operations – which was Del Prato and Pylkkänen's interpretation – or because numbers in general are not a valid input to the LATL's combinatory mechanism. As a critical test case, we employed complex number terms such as *thirty-two*, which at least intuitively, may be instances of conceptual combination with numbers as the input. If such combinations engage the LATL while numerical quantifications do not, this would be evidence that it is the nature of the combinatory operation as opposed to the nature of the input items that matters for the LATL.

Like Del Prato & Pylkkänen (2014), our study employed a production paradigm where subjects named perceptually parallel displays in different ways, depending on task instruction. In all, our design included three combinatory conditions: complex number terms, numerical quantifications, and adjectival modifications, all of which were compared to non-combinatory list controls. We aimed for minimal lexical differences in the produced utterances, and thus, given that complex number terms involve number words in both first and second position (*thirty two*), we designed the numerical quantifications to also have this property (e.g., *three twos*) while adjectival modifications involved a combination of a color adjective and a number term (*green twos*). As a primary non-combinatory control, we used lists consisting of two single-digit numbers (*two*, *three*), but also included lists consisting of a decade number and a single-digit number (*thirty*, *two*), given that lexically, this yields a form identical to the complex number term. However, given that decade numbers are themselves potentially complex, this latter control was not obviously non-combinatory, and thus could have been predicted to pattern somewhere between the combinatory conditions and our single-digit list condition. In fact, this is what we observed and thus decided to treat the straightforward number list (*two*, *three*) as the main non-combinatory control.

As depicted in Fig. 1, all our stimuli consisted of rows of four numbers, some of which were colored. In the Color Modification condition, participants were asked to name the color and the identity of the colored numbers as utterances such as “green twos”. Although the cardinal name is not a prototypical noun, it is thought to act as a noun when placed as the head of the Noun Phrase (Hurford, 1975, 1987, 2001, 2003; Ionin & Matushansky, 2006). Our aim was to replicate the prior finding that adjectival modification of nouns engages the LATL (Bemis & Pylkkänen, 2011, 2013; Del Prato & Pylkkänen, 2014; Westerlund & Pylkkänen, 2014). Crucially, this condition assessed whether a non-canonical open class word such as a number word could ever elicit LATL-relevant conceptual combination.

In the Numerical Quantification condition, participants were asked to name aloud the number of colored digits and the name of those digits (e.g. “Three twos”). In this condition, the cardinal number which acts as a modifier characterized the exact cardinality of sets. Thus, in this case, ‘three’ meant “having exactly three members” (Bale & Khanjian, 2011; Ionin &

Matushansky, 2006). In contrast, in the Complex Number condition, participants were asked to name aloud the two-digit complex number formed by the two colored digits onscreen (e.g., "thirty-two"). Crucially, both this condition and Numerical Quantification combined two number words, but the combinatory operation is different (quantifying the number of numbers vs. creating a complex number). On the basis of Del Prato & Pylkkänen's (2014) findings, we expected no LATL involvement in the processing of the numerical quantifications. If complex numbers patterned similarly, this would suggest that numbers in general cannot function as the "additional feature" whose incorporation constitutes LATL-relevant conceptual combination. In contrast, if complex numbers do engage the LATL, this would indicate that the combinatory operation housed in the LATL does not necessarily require conceptually rich input items, but instead, will also operate on featurally impoverished concepts such as numbers.

So far, we have assumed that at least potentially, the naming of complex numbers involves combinatory operations, but in fact, in the number literature, there has been significant debate about whether two-digit numbers are represented compositionally (i.e., each digit pair is processed as a decade digit and a unit digit separately; e.g. "thirty-two" is the combination of the concept "thirty" and the concept "two") (Grossber & Repin, 2003; Levelt, Roelofs, & Meyer, 1999; McCloskey, 1992; McCloskey, Sokol, & Goodman, 1986; McCloskey, Caramazza, & Basili, 1985; Nuerk, Weger, & Willmes, 2001; Verguts & De Moor, 2005) or holistically (i.e., each digit pair is processed as a single concept: "thirty-two" is stored as a single mental representation) (Brysbaert, 1995; Dehaene, Dupoux, & Mehler, 1990; Dotan, Friedmann, & Dehaene, 2014; Reynvoet & Brysbaert, 1999). Evidence in favor of holistic representations has included studies reporting that priming effects are constant regardless of whether they involve priming by a single or two-digit number (e.g., the priming effect of 7 on 9 is the same as the priming effect of 11 on 9; Reynvoet & Brysbaert, 1999), studies finding that reaction times for deciding whether a two-digit number is larger or smaller than 65 show no significant discontinuities at decade boundaries (i.e., subjects respond "smaller" more slowly to 59 than to 51, although the 5 in the decades position already indicates that both of these numbers are smaller than 65; Dehaene et al., 1990), and studies showing that number magnitude, frequency of the number and sometimes the syllable length of the number name (but not differences in decade) influenced number reading times (Brysbaert, 1995). The main argument in favor of decomposition has, however, been that it is easier to judge whether a complex number is smaller than another when both the units and the decade digit are smaller (e.g. 67 and 52) than when only one of them is (e.g. 62 and 47) even if controlled for overall numerical distance, which suggests that subjects pay attention to the value of each digit, not the whole number (Nuerk et al., 2001; for more recent evidence see Verguts & De Moor, 2005; for a review see Grossberg & Repin, 2003). Additionally Zhou, Chen, Chen and Dong (2008) suggest that whether complex numbers are processed holistically or compositionally depends on the stage of processing. In sum, then, whether complex numbers involve combinatory processing to begin with is an unsettled empirical question, and one that our dataset should be able to shed novel light on. It should be mentioned that even if the holistic vs decomposed dispute emphases on magnitude judgment, the "triple-code model" assumes that in addition to modality-specific symbolic codes in the visual Arabic and auditory verbal domain, there is also a supramodal abstract "number sense" that conveys semantic, information (Dehaene, 1992). Thus, even if the task itself does not explicitly require processing of numerical magnitude this abstract "number sense" remains activated.

A final noteworthy aspect of our design was that both our Color Modification and Numerical quantification conditions involved the production of plural noun phrases (i.e., "blue threes" and "two threes"), whereas the complex number condition did not ("twenty-three"). Previous studies have found increased activation in the left angular gyrus when contrasting plural nouns that were morphosyntactically marked for the number feature (-s) with singular nouns (Domahs et al., 2012) and when linking back two-sentence discourses to plural rather than singular subjects (Boiteau, Bowers, Nair, Almor, 2014). Given that the left angular gyrus (AG) has also been identified as the most likely candidate for number case agreement violations (Carreiras, Carr, Barber & Hernandez, 2010) and proposed to support the manipulation of numbers in verbal form together with other left perisylvian areas (Dehaene, Piazza, Pinel & Cohen, 2003), we analyzed whether the angular gyrus would show effects of morphosyntactic plurality. For completeness, in addition to regions where LATL combinatory effects have localized in prior studies (Brodmann areas 38, 20 and 21), our analysis also included the ventromedial prefrontal cortex, a common though somewhat less consistent locus of basic combinatory effects (e.g., Bemis & Pylkkänen, 2011; Brennan & Pylkkänen, 2012) as well as the left inferior frontal gyrus (LIFG), given its traditional association with language production, though not, at least in MEG, basic combinatory processing either in comprehension (Bemis & Pylkkänen, 2011; 2012) or production (Pylkkänen, Bemis, Blanco Elorrieta, 2014).

2- MATERIALS AND METHODS

2.1- Participants

18 right-handed, native English speakers participated in this experiment (10 female 8 male, 5 years average 5.15 sd). All were neurologically intact, with normal or corrected-to-normal vision and all provided informed written consent.

2.2- Stimuli and experimental design

The experiment consisted of 480 trials in which participants were presented with four digits in a row. The numbers were presented on a dark gray background and some of the digits were colored in pink, blue or green while the rest were presented in light gray (e.g., participants would see “3218” where “3” and “2” were green and “1” and “8” were light gray). The stimuli were kept constant across all conditions, thus assuring that there was no perceptual variation amongst them. Participants were required to name the numbers colored in pink, blue or green aloud; the specifics of the naming task varying upon instruction. Participants were asked to either name the colored digits individually (‘three, two’), the quantity of colored digits and the digits that were colored (‘three twos’), the color of the digit and the digits that were colored (‘green twos), the complex number that they formed (‘thirty-two) or the complex number and the units separately (‘thirty, two’) (Fig. 1). The quantity of colored digits and their location on the number varied in a controlled fashion between conditions and all numbers were composed by four digits to assure that participants could perceive all the digits of each number at one glance (Kaufman, Lord, Reese, & Volkmann, 1949; Saltzmann & Gamer, 1948).

In order to assure that participants were performing a genuine quantification task in the Numerical Quantification condition, we aimed for the base number word (which corresponded to the described colored number) to be bimorphemic (one morpheme in root, and –s). As a consequence, ‘one’ was excluded as the first-position word and only ‘two’, ‘three’ or ‘four’ were included. The color words for the color modification condition were chosen such as to match the quantifier number words as closely as possible while also being maximally visually distinctive from each other (if the latter constraint was not met, then the condition involving color naming could have turned out harder than number naming). Consequently, we required the color words to be monosyllabic like the number words. For the rest of lexical-level variables, we chose English Lexicon Project naming times as a summary statistic (Balota, et al., 2007). The color words ‘blue’, ‘pink’ and ‘green’ were chosen as optimally matching these number words. The English Lexicon Project naming times of these color words were somewhat faster (mean = 577ms) than those of the Number words (mean = 605ms), but this difference was not significant (p = .13). In addition, only monosyllabic number words from 1 to 9 were included in the design as base numbers. We excluded disyllabic number words (e.g. “seven”) to avoid effects related to larger number of syllables (such as delayed naming latencies and greater motor preparation for more syllabled words). Therefore, only eight numbers were used as base numbers (‘one’, ‘two’, ‘three’, ‘four’, ‘five’, ‘six’, eight’, ‘and ‘nine’).

All base number words were combined with the three possible first position words (either number words or colors) eliciting 24 combinations for each condition which formed each of the experimental blocks. Each block was repeated four times during the experiment, eliciting 96 trials per condition (480 trials in total). All pictures were presented foveally using Presentation (Neurobehavioral System Inc., California, USA) and subtended in a range from 55.16° height and 33.36° width on a screen ~85cm from the subject.

2.3- Procedure

Before recording, each subject’s head shape was digitized using a Polhemus dual source handheld FastSCAN laser scanner (Polhemus, VT, USA). MEG data were collected in the Neuroscience of Language Lab in NYU Abu Dhabi using a whole-head 208 channel axial gradiometer system (Kanazawa Institute of Technology, Kanazawa, Japan) as subjects lay in a dimly lit, magnetically shielded room. Vocal responses were captured with an MEG compatible microphone (Shure PG 81, Shure Europe GmbH).

In all conditions, trials began with a fixation cross (300 ms), followed by the presentation of the stimuli. The picture remained onscreen until speech onset (1400 ms timeout), and participants were allowed 1200 ms to finish their speech before the fixation cross for the following trial would appear. The entire recording lasted ~25 min.

2.4- Data acquisition and preprocessing

MEG data were recorded at 1000 Hz (200 Hz low-pass filter) and epoched from 200 ms before to 700 ms after picture onset. During our blink and artifact rejection routine, all trials exceeding the absolute threshold of 2.5 pT in amplitude after noise reduction were rejected. Trials containing any remaining blinks were identified by visualizing the topographies of any sudden, stark increases of activity. If the magnetic field pattern had the characteristic frontal distribution of a blink, that trial was rejected. This procedure resulted in the exclusion of more than 85% of the trials for three participants due to excessive artifacts in their recordings (caused by construction works effectuated next door at the time of the experiment). Therefore, these three participants were excluded from further analyses. For the participants included in the analysis, the artifact and blink rejection routines resulted in the exclusion of 28.75% of the trials (7.7% sd), leaving 342.6 trials on average per subject (37.41 sd). A 200 ms interval (-200 0 ms) was used for baseline correction and data were averaged for each condition and subject. Averages were low-pass filtered at 40 Hz. Due to the noise-conditions of this lab, no high pass filtering was required.

2.5- Statistical analysis

L2 Minimum norm estimates of source activity were created for each average using the default parameters of BESA 6 (MEGIS Software GmbH). Regions of interest were defined in terms of Brodmann areas (BAs), which were isolated with the Tailarach daemon (Lancaster et al., 1997; Lancaster et al., 2000) from the BESA source space, consisting of 1500 regional sources evenly distributed in two shells (750 regional sources each shell) along a smooth cortex adjusted to the participants' digitized head shape information.

As the main goal of the current study was to compare combinatory effects in color modification as opposed to either complex number combinations or number quantification, we conducted a main analysis in the areas of the LATL (BAs 38, 20 and 21) that have previously been implicated in conceptual combination (Bemis & Pylkkänen, 2011; Pylkkänen, Bemis, Blanco Elorrieta, 2014; Westerlund & Pylkkänen, 2014). Although BA 20 and 21 stretch to more posterior regions of the temporal lobe, they were included in this analysis in order to cover anterior temporal cortex outside of the temporal pole (i.e., BA 38). In prior MEG studies, LATL combinatory effects have centered both around the pole (e.g., Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis, Blanco Elorrieta, 2014) as well as more laterally (Westerlund & Pylkkänen, 2014; Westerlund et al., 2015). Crucially, we complemented our ROI-analyses with liberally thresholded whole brain contrasts capable of revealing the centers of activity within the ROIs.

In addition to the hypothesis-driven LATL analysis, we ran a separate more explorative analysis in areas that do not constitute the main focus of the study but could be sensitive to the current experimental manipulations. This second analysis included the angular gyrus (AG), the ventromedial prefrontal cortex (vmPFC) and the left inferior frontal gyrus (LIFG). The AG was included since we wanted to identify a possible locus for numerical quantification and plural composition, and this area has been previously reported both for number word processing (Dehaene, Piazza, Pinel & Cohen, 2003) and for plural representations (Boiteau, Bowers, Nair, Almor, 2014; Carreiras, Carr, Barber, & Hernandez, 2010; Domahs, Nagels, Domahs, Whitney, Wiese & Kircher, 2012). The vmPFC was included as previous studies (Bemis & Pylkkänen, 2011, 2013; Pylkkänen, Bemis & Blanco-Elorrieta, 2014) have found the vmPFC to be involved to some extent in basic composition. Following such studies, left and right BA11 were collapsed into a single ROI due to spatial adjacency along the midline. Lastly, although previous MEG studies on language production have not found composition effects in the LIFG (Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis & Blanco-Elorrieta, 2014) we also included this area in the analysis; given its general prominence in research in production (e.g., Haller, Radue, Erb, Grodd, & Kircher, 2005; Indefrey et al., 2001; Menenti, Gierhan, Segaert, & Hagoort, 2011). Due to the small number of regional sources within BA 44-45, they were also collapsed into a single ROI. BA 39 was used for the angular gyrus.

For the time-course data of each region, a non-parametric cluster permutation test (Maris & Oostenveld, 2007) with 10,000 permutations was used to identify temporal clusters during which the localized activity differed significantly between conditions, corrected for multiple comparisons over time. For initial cluster selection, we adopted the parameters of prior studies: 10 adjacent time points showing an effect at an alpha level of $p < 0.3$, (e.g., Bemis & Pylkkänen 2011; Bemis & Pylkkänen 2012; Bemis & Pylkkänen 2013; Del Prato & Pylkkänen 2014; Leiken & Pylkkänen 2013; Pylkkänen, Bemis & Blanco Elorrieta, 2014; Westerlund & Pylkkänen, 2014). Then, for each cluster surviving these thresholds, a test statistic was constructed that was equal to the summed *t*-values of the point-by-point test-statistics over the selected cluster interval and finally, the cluster with the largest summed test statistic was chosen for further computations. Due to the last step, this test is only capable of identifying one effect within any given analysis interval and thus in order to be able to characterize potential earlier and later effects, all analyses were conducted both in a mid (300 - 450 ms) and a late (450 - 600 ms) time window, defined to avoid early perceptual components and late motion artifacts. For the largest cluster within an interval the corrected p-value ($p < .05$) was calculated as the ratio of permutations yielding a test statistic greater than the actual observed test statistic. Since only increases for combinatory conditions overs lists were interpretable in light of our hypotheses and since our study in general was based on prior evidence that combinatory conditions elicit stronger activity than list conditions (Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis, Blanco Elorrieta, 2014), all permutation t-tests were one-tailed. Finally, to moderately protect our analysis against false positives across multiple regions within the same analysis while still maintaining power, a false discovery rate (FDR; Benjamini & Hochberg, 1995; Genovese, Lazar, & Nichols, 2002) of 0.1 was used throughout. This rate was kept somewhat liberal given that our main research question had no explorative component and the combinatory effects in our main dependent measure, LATL amplitude as measured by MEG, have already been replicated in numerous studies (Bemis & Pylkkänen, 2011, 2012, 2013; Westerlund & Pylkkänen, 2014; Del Prato & Pylkkänen, 2014; Pylkkänen, Bemis, Blanco Elorrieta, 2014; Leffel et al., 2014; Westerlund et al., 2015; Zhang & Pylkkänen, 2015).

Given that two of our LATL ROIs, BA 20 and 21, covered not only anterior but also posterior temporal cortex, as a final step in our analysis we visualized the centers of any obtained LATL effects with liberally thresholded uncorrected full brain analyses focused on the statistical peaks of the ROI effects in each comparison (Fig. 3). Full brain contrasts were run at the time points of the highest uncorrected statistic in the ROI cluster and in the visualization, individual sources were plotted as red (indicating an increase for a combinatory condition) or blue (indicating a decrease for a combinatory condition) when they and at least two of their adjacent spatial and temporal neighbors showed an uncorrected significance of $p < .05$. As emphasized above, the purpose of these analyses was simply to address any potential spatial ambiguity in the ROI analyses.

3- RESULTS

3.1- Behavioral Results

Reaction times were submitted to a one way ANOVA with five levels. The results showed a main effect of condition [$F(4,72) = 72.54$, $p<.0001$]. Participants were the slowest naming the numbers in Color modification condition ($M = 987$ ms; *SD* = 224ms) and planned t-tests showed that this delay was significant when compared to the other two experimental conditions: numerical quantification [$t(18)=3.62$, $p = .001$] and complex number naming [$t(18)=9.86$, $p<.0001$]. Additionally, Numerical Quantification was also significantly slower than Complex number naming [$t(18) = 10.23$, $p < .0001$]. There was no significant difference between our two control conditions; complex number list and number list [$t(18)=0.91$, $p=.3$], although number list was slightly slower on average ($M = 804$ ms; $SD = 185$ for number list vs $M = 793$ ms $SD = 162$ for complex number list) (Fig. 2).

Accuracy in all conditions was at ceiling, with each participant making an average of 7 errors in the course of the whole experiment (0.016%). For this reason we forewent analyses of accuracy and rejections of incorrect trials from the MEG data in this task.

3.2- MEG results

In order to find the best control condition for our combinatorial effects, we ran a one way ANOVA with three levels (Complex number, Complex number list and Number list) in the LATL. Although non-significant, the result pattern showed a clear layered effect of composition, with Complex number eliciting the greatest activity followed by Complex number list, and Number list being the one eliciting the least activity (Additional Figure 1). This pattern suggested that Complex number list condition may have elicited combinatorial activity to some extent. For this reason, we used Number list condition as the baseline control condition to assess combinatorial effects in our experimental manipulations.

The results of the pairwise comparisons conducted in the LATL (BA 38, 20, 21) revealed increased activation for Color modification over Number lists in all the analyzed areas (Fig. 3). The effect was reliable in BA21 [300-450 ms; $p = .04$] and marginally reliable in BA20 [331-450 ms; $p = .06$]. Additionally, a cluster of activity was also identified in BA38, although it did not reach significance [392-423 ms; $p = .1$]. The pattern of results was parallel for Complex number condition, as this condition also elicited significantly greater activity than Number lists in the three areas. Specifically, this increase in activity was reliable in BA21 [450- 600ms; $p = .03$] and marginally reliable in BA20 [450-590 ms; $p = .06$]. In addition, a non-significant cluster of activation was observed in BA38 [506-583 ms; $p = .1$]. In contrast, the comparison between Numerical Quantification and Number List, did not elicit significant differences in any of the analyzed areas (no clusters of activation were found in BA20 and BA38, and the cluster located in BA21 did not reach reliability [450-551 ms; p = .1]). To further contrast the differences between Complex number and Numerical Quantification, we ran an additional direct comparison between them and the results showed that Complex number elicited marginally significant increased activity in BA38 [516-600 ms; p= .09], and non-reliable increased activity in BA20 [540-594 ms; p = .16] and BA21 [552-561 ms; p = .3] (Fig. 3).

Thus in results both complex number formation and color modification modulated the LATL, while numerical quantification did not affect it. When the center of this effect was visualized in the full brain contrasts, it localized in the anterior half of ventrolateral temporal cortex for both color modification and numerical quantification (Fig. 3), though clearly not as anteriorly as the temporal pole. This conforms to the ROI results where the temporal pole only showed marginal effects. Also, this localization is within the range of previously reported effect centers in prior LATL composition studies: although it is somewhat more posterior than, e.g., the production results of Del Prato & Pylkkänen (2014), it is consistent with for instance the comprehension results of Pylkkänen & Westerlund (2014). Thus there is not yet any clear generalization as regards the precise locus of combinatory effects within the LATL, and given the somewhat fuzzy spatial resolution of MEG, it is also unclear whether such genralizations could be obtained with MEG.

The analyses on the AG aimed to test sensitivity for plurality in this area. Therefore, conditions containing plural NPs (Color modification and Numerical quantification) and singular NPs (Complex number and Number list) were contrasted. The results showed significant differences for Color modification over both Complex number [447-600 ms; p = .05] and Number list [408-600 ms; p = .01]. The same pattern replicated when comparing Numerical Quantification with Complex number [414-600 ms; p = .03] and Number list [416-600 ms; p = .01]. Additionally, no clusters of activation were found when comparing the two singular or the two plural conditions. The vmPFC was reliably engaged when performing a color modification in comparison to naming a number list [389-562 ms; p = .04] but not when performing a Numerical Quantification [457-551 ms; p = .17] or naming a complex number [500-549 ms; p =.23]. Lastly, no comparisons elicited reliable clusters of activation in the LIFG (Fig. 4).

4- DISCUSSION

4.1- Complex number composition in the LATL

In this study we aimed to develop some understanding into the specifics of the combinatorial processes performed in the LATL and the nature of the elements that can enter these computations. Particularly, we tried to elucidate if the lack of LATL engagement during numerally quantified phrases such as *two cups* in previous studies (Del Prato & Pylkkänen, 2014) was because number words are not a valid input element for the LATL or because the LATL is not involved in the computations underlying quantification in particular. To achieve this, we tested complex number expressions (*thirty-three*) where the combinatory mode intuitively resembles conceptual combination, already demonstrated to drive the LATL (Baron & Osherson, 2011; Bemis & Pylkkänen, 2011, 2013; Del Prato & Pylkkänen 2014; Pylkkänen, Bemis & Blanco-Elorrieta, 2014; Westerlund & Pylkkänen, 2014), and compared this to numerical quantification. According to our results, the reason for the lack of LATL effects in numerical quantifications is not the numerical input but rather the mode of composition: when numbers compose into complex concepts as in *thirty-two*, LATL activity is increased in comparison to both non-combinatory lists (*three*, *two*) as well as to quantificational phrases (*three twos*). Thus our findings suggest that the LATL is not a general purpose combiner of meanings but rather specializes in some version of conceptual combination, potentially delimited to situations where one combining element characterizes a property of the other.

Additionally, the finding of combinatorial activity for our complex number condition conforms to theories where complex numbers undergo a composition process before being produced (Deloche & Seron, 1987; Levelt et al., 1999), as opposed to being processed holistically (Brysbaert, 1995; Dehaene et al., 1990; Reynvoet & Brysbaert, 1999). In other words, our results suggest that in order to produce “twenty-three”, the concept of “twenty” and the concept of “three” are retrieved and combined. This proposal is consistent with McCloskey et al.’s (1986) model where complex number production involves the generation of a syntactic frame that specifies each to-be-retrieved word in terms of a number-lexical class (ones, teens…) and a position within that class. Moreover, it conforms to proposals within theoretical linguistics where complex number composition follows the standard principles of semantic and syntactic combination (Ionin & Matushansky, 2006).

Despite the anatomical overlap in regions activated by complex number naming and color modification, we did observe a difference in the time course of these effects: Composition effects for complex numbers occurred around 100 ms later than the adjectival modification effects. However, this timing contrast may simply have resulted from the difference in the number of syllables in the two conditions, with complex numbers consisting of three and adjectival modifications of two syllables. Also, our adjectival modification effect onset somewhat later, at around 300 ms, than parallel effects in the two prior production studies that the current study directly followed up on, both of which found increases for adjectival modification over two-word lists by 250 ms (Pylkkänen, Bemis & Blanco-Elorrieta, 2014; Del Prato & Pylkkänen, 2014). A potential explanation for the later effects in the current study lies in the number of possible production responses associated with a given visual stimulus type. In the prior studies, combinatory and list productions were elicited by different types of visual displays whereas here, the nature of the visual display was kept constant across all five experimental conditions, with only task instruction separating the conditions from each other. Thus participants always had to select the appropriate naming task out of five options without the stimuli cuing the correct response. This increase in the possible utterances for each stimulus could have delayed the retrieval of the appropriate individual concepts and consequently the composition process in this experiment.

4.2- Contributions of the vmPFC, AG and LIFG

The AG was included in our analysis to test whether we would replicate the finding by Domahs et al. (2012) that plural nouns elicit reliably greater AG activity than singular nouns. Our results straightforwardly mirrored this as the two conditions that involved naming a plural NP (Color modification and Numerical Quantification) elicited greater activity in this region than the conditions including a singular NP (Complex number and Number list). Thus our result extends Domahs et al.'s comprehension finding to production. However, there is a divergence between our and Domahs et al.'s results. One of their findings was that mass nouns also increased activity in the AG as compared to singular nouns, leading them to suggest that the left AG was involved in the semantic interpretation of different kinds of non-singularity. In contrast, our results suggest that there is at least a limitation to that interpretation of non-singularity, as complex number representations are non-singular and did not engage this area, tentatively because the NP was singular. In the behavioral data, we also observed longer naming latencies in exactly the two conditions that involved naming a plural NP (Color modification and Numerical Quantification) as compared to the three conditions that involved naming singular NPs. This "plural effect" conforms to prior findings in comprehension showing delayed responses to plural as compared to singular NPs (Tucker, Idrissi, & Almeida, 2015; Wagers, Lau, Phillips, 2009). In sum, the angular gyrus seems to be recruited both for syntactic number marking (Domahs et al. 2012, and this study) and for general number magnitude processing, with future work hopefully elucidating the relationship between them.

Another possible interpretation of our AG pattern could be in terms of combinatorial activity, given that two of our combinatory conditions elicited an AG increase as compared to the list condition. This would conform to proposals of the AG as a type "integrative" site (Lau et al. 2008) or as a locus of combinatorial semantics (Price, Bonner, Peelle & Grossman, 2015). However, given the strong prior evidence that morphological plurality drives the AG, an explanation in terms of this factor appears more straightforward, since under this account the lack of an AG effect in the complex number condition directly follows. In conclusion though, both types of accounts remain logical possibilities.

Our results also revealed a significant increase in vmPFC activity for color modification in contrast to number lists, replicating previous studies where this area has shown effects of adjective noun combination both in comprehension (Bemis & Pylkkänen, 2011) and production (Pylkkänen, Bemis & Blanco-Elorrieta, 2014). Clusters of increased activity were also identified in both numerical quantification and complex number naming, though these effects did not survive correction for multiple comparisons. In the more general neuroscience literature, the vmPFC has been proposed as a general integrative site (Roy, Shohamy, & Wager, 2012) particularly active when responses are shaped by conceptual information about specific outcomes. Given that in our results, the vmPFC trended towards larger amplitudes in all combinatory conditions, any account in terms of general integrative processing should be able to explain this data pattern.

Finally, consistent with prior MEG studies on basic combinatory phrases both in comprehension (Bemis & Pylkkänen, 2011; 2012) and production (Del Prato and Pylkkänen, 2014; Pylkkänen, Bemis & Blanco-Elorrieta, 2014), we did not observe combinatory effects in the LIFG. Thus to the extent that the LIFG may be involved in linguistic composition (e.g., Hagoort, 2005), the nature of its contribution is not a time-locked evoked response of the sort that would be revealed by the type of source modeling of averaged data as performed here.

5- CONCLUSION

In this work we set out to delineate the combinatory computation housed in the LATL, with a focus on number words and their composition. We found that the LATL composes number words but only if the composing items combine into a more complex concept (*thirty two*) and not when one word enumerates the number of tokens of the other (*three twos*). Our findings suggest that the engagement of the LATL is determined by the computations underlying the performed combinatorial process as opposed to the nature of the input items. Additionally, we found support for the angular gyrus as a leading candidate for the processing of plural representations.

Acknowledgements

This research was supported by the National Science Foundation grant BCS-1221723 (LP) and grant G1001 from the NYUAD Institute, New York University Abu Dhabi (LP).

References

Bale, A. C. & Khanjian, H. (2011). Cross-linguistic representations of numerals and number marking. In *Proceedings of SALT* (Vol. 20, pp. 582-598).

Balota, D.A., Yap, M.J., Hutchison, K.A., Cortese, M.J., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445-459.

Baron, S.G. & Osherson, D. (2011). Evidence for conceptual combination in the left anterior temporal lobe. NeuroImage, 55 (4), 1847–1852.

Barwise, J., & Cooper, R. (1981). Generalized quantifiers and natural language.*Linguistics and philosophy*, *4*(2), 159-219.

Bemis, D.K. & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. The Journal of Neuroscience, 31 (8), 2801.

Bemis, D.K. & Pylkkänen, L. (2012). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. *Cerebral Cortex, 23(8),* 1859-1873.

Bemis, D.K. & Pylkkänen, L. (2013). Basic linguistic composition recruits the left anterior temporal lobe and left angular gyrus during both listening and reading. Cerebral Cortex, 23 (8), 1859–1873.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 289-300.

Bennett, M. R. (1975). *Some extensions of a Montague fragment of English* (Vol. 129). Indiana University Linguistics Club.

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.

Binder, J.R., Desai, R.H. (2011). The neurobiology of semantic memory. Trends in Cognitive Sciences 15 (11), 527–536.

Boiteau, T. W., Bowers, E., Nair, V. A., & Almor, A. (2014). The neural representation of plural discourse entities. *Brain and language*, *137*, 130-141.

Brennan, J. & Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *NeuroImage, 60(2),* 1139-48.

Bright, P., Moss, H., & Tyler, L. (2004). Unitary vs multiple semantics: PET studies of word and picture processing. Brain and Language, 89 (3), 417–432.

Brysbaert, M. (1995). Arabic number reading: On the nature of the numerical scale and the origin of phonological recoding. Journal of Experimental Psychology: General, 124, 434–452.

Carreiras, M., Carr, L., Barber, H. A., & Hernandez, A. (2010). Where syntax meets math: Right intraparietal sulcus activation in response to grammatical number agreement violations. *Neuroimage*, *49*(2), 1741-1749.

Clarke, A., Taylor, K.I., & Tyler, L.K. (2011). The evolution of meaning: Spatio-temporal dynamics of visual object recognition. Journal of Cognitive Neuroscience, 23 (8), 1887–1899.

Clarke, A., Taylor, K.I, Devereux, B., Randall, B., Tyler, L.K. (2013). From perception to conception: How meaningful objects are processed over time. Cerebral Cortex, 23 (1), 187–197.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1), 1-42.

Dehaene,S., Dupoux, E. & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. Journal of Experimental Psychology: Human Perception and Performance, 16, 626–641.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive neuropsychology*, *20*(3-6), 487-506.

Deloche, G., & Seron, X. (1987). *Numerical transcoding: A general production model*. Lawrence Erlbaum Associates, Inc.

Del Prato, P., & Pylkkänen, L. (2014). MEG evidence for conceptual combination but not numeral quantification in the left anterior temporal lobe during language production. *Frontiers in Psychology 5(524).*

Domahs, F., Nagels, A., Domahs, U., Whitney, C., Wiese, R., & Kircher, T. (2012). Where the mass counts: common cortical activation for different kinds of nonsingularity. *Journal of cognitive neuroscience*, *24*(4), 915-932.

Dotan, D., Friedmann, N., & Dehaene, S. (2014). Breaking down number syntax: Spared comprehension of multi-digit numbers in a patient with impaired digit-to-word conversion. *Cortex*, *59*, 62-73.

Friederici, A. D. (2011). The brain basis of language processing: from structure to function. *Physiological reviews*, *91*(4), 1357-1392.

Friederici, A.D., Meyer, M., & Von Cramon, D.Y. (2000). Auditory language comprehension: An event-related fMRI study on the processing of syntactic and lexical information. Brain and Language, 74 (2), 289–300.

Gauthier, I., Anderson, A.W., Tarr, M.J., Skudlarski, P., & Gore, J.C. (1997). Levels of categorization in visual recognition studied using functional magnetic resonance imaging. Current Biology, 7 (9), 645–651.

Genovese, C.R., Lazar, N.A., & Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage, 15*(2), 870-878.

Grabowski, T.J., Damasio, H., Tranel, D., Ponto, L.L. B., Hichwa, R.D., & Damasio, A.R. (2001). A role for left temporal pole in the retrieval of words for unique entities. Human Brain Mapping, 13 (4), 199–212.

Grossberg, S., & Repin, D. V. (2003). A neural model of how the brain represents and compares multi-digit numbers: spatial and categorical processes. *Neural Networks*, *16*(8), 1107-1140.

Hagoort, P. (2005). On Broca, brain, and binding: a new framework. *Trends in cognitive sciences*, *9*(9), 416-423.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual review of neuroscience*, *37*, 347-362.

Haller, S., Radue, E.W., Erb, M., Grodd, W., & Kircher, T. (2005). Overt sentence production in event-related fMRI. Neuropsychologia, 43, 807–814.

Hampton, J. A. (1997). “Conceptual combination,” in *Knowledge, Concepts, and Categories,* eds. K. Lamberts and D. Shanks (London: UCL Press), 133–159.

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.

Hurford, J. (1975) The linguistic theory of numerals. Cambridge University Press. Cambridge.

Hurford, J. (1987) Language and Number: The Emergence of a Cognitive System. Blackwell. Oxford.

Hurford, J. (2001) ‘Numeral systems’. In N. J. Smelser & P. B. Baltes (eds),International Encyclopedia of the Social and Behavioral Sciences. Pergamon. Amsterdam. 10756–10761.

Hurford, J. (2003) The interaction between numerals and nouns. In F.Plank (ed), Noun Phrase Structure in theLanguages of Europe. Typology of Languages in Europe 7. Mouton de Gruyter. The Hague.

Indefrey, P., Brown, C.M., Hellwig, F., Amunts, K., Herzog, H., & Hagoort, P. (2001). A neural correlate of syntactic encoding during speech production. Proceedings of the National Academy of Sciences, 98, 5933–5936.

Ionin, T., & Matushansky, O. (2006). The composition of complex cardinals.*Journal of Semantics*, *23*(4), 315-360.

Kaufman, E.L., Lord, M.W., Reese, T.W., Volkmann, J. (1949). The discrimination of visual number. American Journal of Psychology, 62, 498-525.

Lancaster, J. L., Rainey, L. H., Summerlin, J. L., Freitas, C. S., Fox, P. T., Evans, A. C., ... & Mazziotta, J. C. (1997). Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Human brain mapping*, *5*(4), 238.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., ... & Fox, P. T. (2000). Automated Talairach atlas labels for functional brain mapping. *Human brain mapping*, *10*(3), 120-131.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, *9*(12), 920-933.

Leffel, T., Lauter, M., Westerlund, M., & Pylkkänen, L. (2014). Restrictive vs. non-restrictive composition: a magnetoencephalography study. *Language, Cognition and Neuroscience.*

Leiken, K. & Pylkkänen, L. (2013). MEG evidence that the LIFG effect of object extraction requires similarity-based interference. *Language and Cognitive Processes*.

Link, G. (2002). The logical analysis of plurals and mass terms: A lattice-theoretical approach. Formal semantics: The essential readings, 127-146.

Maris, E. & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of Neuroscience Methods, 164 (1), 177–190.

Mazoyer, B.M., Tzourio, N., Frak, V., Syrota, A., Murayama, N., Levrier, O. et al. (1993). The cortical representation of speech. Journal of Cognitive Neuroscience, 5 (4), 467–479.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition*, *44*(1), 107-157.

McCloskey, M., Caramazza, A., & Basili, A. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and cognition*, *4*(2), 171-196.

McCloskey, M., Sokol, S. M., & Goodman, R. A. (1986). Cognitive processes in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General*, *115*(4), 307.

Medin, D. L., & Shoben, E. J. (1988). Context and structure in conceptual combination. *Cognitive Psychology*, *20*(2), 158-190.

Menenti, L., Gierhan, S.M.E., Segaert, K. & Hagoort, P. (2011). Shared language: Overlap and segregation of the neuronal infrastructure for speaking and listening revealed by functional MRI. Psychological Science, 22, 1173–1182.

Montague, R. (1974). Formal philosophy; selected papers of Richard Montague.

Murphy, G. L. (1990). Noun phrase interpretation and conceptual combination. *Journal of memory and language*, *29*(3), 259-288.

Nuerk, H.C., Weger, U., & Willmes, K. (2001) Decade breaks in the mental number line? Putting tens and units back into different bins. Cognition, 82, B25–B33.

Pallier, C., Devauchelle, A.D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. Proceedings of the National Academy of Sciences, 108 (6), 2522.

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging Evidence for the Neuroanatomic Basis of Combinatorial Semantics in the Angular Gyrus. *The Journal of Neuroscience*, *35*(7), 3276-3284.

Pylkkänen, L., Bemis, D. K., Blanco Elorrieta, E. (2014). Building phrases in language production: An MEG study of simple composition.*Cognition*, 133, 371-384.

Reynvoet, B. & Brysbaert, M. (1999). Single-digit and two-digit arabic numbers address the same semantic number line. Cognition, 72, 191–201.

Rogalsky, C. & Hickok, G. (2009). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. Cerebral Cortex, 19 (4), 786.

Rogers, T.T., Hocking, J., Noppeney, U., Mechelli, A., Gorno-Tempini, M.L., Patterson, K. et al.(2006). Anterior temporal cortex and semantic memory: Reconciling findings from neuropsychology and functional imaging. Cognitive, Affective, & Behavioral Neuroscience, 6 (3), 201–213.

Roy, M., Shohamy, D., & Wager, T. D. (2012). Ventromedial prefrontal-subcortical systems and the generation of affective meaning. *Trends in cognitive sciences*, *16*(3), 147-156.

Saltzman, I.J., Gamer, W.R. (1948). Reaction time as a measure of span of attention. American Journal of Psychology, 25, 227-241.

Scha, R. (1984). Distributive, collective and cumulative quantification. *Truth, Interpretation, and Information, GRASS*, *2*, 131-158.

Stowe, L.A., Broere, C.A. J., Paans, A.M. J., Wijers, A.A., Mulder, G., Vaalburg, W. et al. (1998). Localizing components of a complex task: Sentence processing and working memory. NeuroReport, 9 (13), 2995.

Tucker, M.A., Idrissi, A. & Almeida, D. (2015). Representing number in the real-time processing of agreement: Self-paced reading evidence from Arabic. *Frontiers in Psychology,* 6, 314.

Tyler, L.K., Stamatakis, E.A., Bright, P., Acres, K., Abdallah, S., Rodd, J., et al. (2004). Processing objects at different levels of specificity. Journal of Cognitive Neuroscience, 16 (3), 351–362.

Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550-560.

Verguts, T. & De Moor, W. (2005). Two-digit Comparison: Decomposed, Holistic, or Hybrid? Experimental Psychology, 52(3), 195-200.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Representations and processes. *Journal of Memory and Language*, *61*(2), 206-237.

Westerlund, M., Kastner, I., Al Kaabi, M., & Pylkkänen, L. (2015). The LATL as locus of composition: MEG evidence from English and Arabic. *Brain and Language, 141*, 124-134

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, *57*, 59–70.

Wisniewski, E. J. (1996). Construal and similarity in conceptual combination. *Journal of Memory and Language*, *35*(3), 434-453.

Xu, J., Kemeny, S., Park, G., Frattali, C. & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. NeuroImage, 25 (3), 1002–1015.

Zhang, L. & Pylkkänen, L. (2015). The interplay of composition and concept specificity in the left anterior temporal lobe: An MEG study.*Neuroimage.*

Zhou, X., Chen, C., Chen, L., & Dong, Q. (2008). Holistic or compositional representation of two-digit numbers? Evidence from the distance, magnitude, and SNARC effects in a number-matching task. *Cognition*, *106*(3), 1525-1536.

Figure legends:

Figure 1. Experimental design. Participants were asked to describe the colored digits onscreen by (A) naming the color and the numbers that were colored (B) counting the number of colored digits and naming the digits that were colored, (C) naming the complex number the colored digits formed, (D) naming the complex number on the left and the units number on the right individually or (E) naming the colored digits individually in a list like fashion.

Figure 2. Mean reaction times as a function of the performed naming task.

Figure 3. ROI results for pairwise comparisons in the LATL, activation averaged across subjects. On the waveform plots, the shaded regions indicate that the difference in activity between the two tested conditions was significant at a p = .05 value (corrected), while the boxed region indicates marginally significant effects. Significance was determined using a non-parametric, permutation test (Maris & Oostenveld, 2007) performed from 300 to 450 and 450 to 600ms (10,000 permutations). The point-by-point t-statistic is also plotted in grey, with a red line indicating uncorrected significance at the $p < .05$ level. Finally, the right most panel visualizes the activity centers of the LATL effects obtained in the ROI analysis by plotting a source-by-source full brain comparison at the time of the statistical peak of the ROI effect. Individual sources are plotted as red when they and at least two of their adjacent spatial and temporal neighbors showed an increase for the combinatory condition at $p < .05$.

Figure 4. ROI results for pairwise comparisons in the vmPFC, AG and LIFG, activation averaged across subjects. On the waveform plots, the shaded regions indicate that the difference in activity between the two tested conditions was significant at a p = .05 value, while the boxed region indicates marginally significant effects. Significance was determined using a non-parametric, permutation test (Maris & Oostenveld, 2007) performed from 300 to 450 and 450 to 600ms (10,000 permutations).

Additional figures:

Additional figure 1. ROI results for a one-way ANOVA in the LATL, activation averaged across subjects. No region showed significant differences at an alpha value of $p = .05$. Significance was determined using a non-parametric, permutation test (Maris & Oostenveld, 2007) performed from 300 to 450 and 450 to 600ms (10,000 permutations).