SEMANTIC NUMBER IN FRENCH AND CHINESE BRAINS

by

DONALD DUNAGAN

(Under the Direction of John Hale)

ABSTRACT

One aspect of natural language comprehension is understanding how many of what or whom. While previous work has documented the neural correlates of number comprehension and quantity comparison, we investigate semantic number from a cross-linguistic perspective with the goal of identifying cortical regions involved in distinguishing plural from singular nouns. We use two fMRI datasets in which Chinese and French native speakers listen to an audiobook of a children's story in their native language, selecting these two languages because they differ in their semantics. While Chinese nominals lack pluralization, French nouns are overtly marked for number. We find a number of known semantic processing regions in common in which cortical activation is greater for plural than singular nouns and posit a cross-linguistic role for number in semantic composition, the process by which individual concepts are combined to form complex meaning.

INDEX WORDS: [Neurolinguistics, Cognitive Neuroscience, Linguistic Typology, Linguistic Semantics, fMRI, GLM]

SEMANTIC NUMBER IN FRENCH AND CHINESE BRAINS

by

DONALD DUNAGAN

B.S., University of Georgia, 2019
A.B., University of Georgia, 2019

A Thesis Submitted to the Graduate Faculty of the
University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2020

SEMANTIC NUMBER IN FRENCH AND CHINESE BRAINS

by

DONALD DUNAGAN

Major Professor:	John Hale
Committee:	Margaret Renwick Tianming Liu

Electronic Version Approved:

Ron Walcott
Dean of the Graduate School
The University of Georgia
December 2020

ACKNOWLEDGMENTS

I would like to thank a number of collaborators without whom this project would not be possible. I would like to thank Shulin Zhang for her heroic effort preparing the Chinese portion of our corpus. I would like to thank Jixing Li for providing the Chinese fMRI dataset and for her willingness to provide annotations. I would like to thank Christophe Pallier for his insight, providing the French fMRI dataset, and reviewing my code. I would like to thank John Whitman for proposing the project topic. And I would like to thank John Hale for his guidance throughout.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

§ 1:1 Terminology Specifications

Before proceeding, we would like to make some clarifying comments related to terminology. A nominal is *singular* if its cardinality is one. A nominal is *plural* if its cardinality is more than one. Nominals with *general number* are underspecified for number: "one or more X." We use *semantic number* as whether a nominal's interpretation is singular or plural. We distinguish semantic number from *grammatical number* which is a grammatical category and which is expressed by morphological or syntactic means. Grammatical number expresses semantic number through grammar.

§ 1:2 Scientific Questions

One aspect of natural language comprehension is understanding how many of what or whom. While much work has been done to document the neural correlates of number comprehension and quantity comparison (Carreiras et al., 2010; Castelli et al., 2006; Dehaene et al., 2003; Kadosh & Walsh, 2009), the neural correlates of semantic number are less well understood. In the singular-plural opposition, at least in languages which mark a binary distinction, singular is taken to be the default form with plural as the non-default (de Swart & Farkas, 2010). The first question to be addressed is: "Do plural nouns elicit greater cortical activity than singular nouns?"

Portions of the prefrontal and parietal cortices, known as the multiple-demand network (MD) (Duncan, 2010), have been found to to be responsive to a wide variety of cognitive demands such as: verbal and spatial working memory, the Stroop task, the multi-source interference task, a memory-probe task, and most importantly to this paper, an arithmetic task (Fedorenko et al., 2013). On one hand, these regions have been shown to not track linguistic input as closely as language-selective regions (Blank & Fedorenko, 2017), and Fedorenko et al. (2011) find little or no overlap between cortical regions engaged in high-level linguistic processing and MD regions which respond to musical processing, general working memory, cognitive control, and most importantly here, mental arithmetic. On the other hand Carreiras et al. (2010) take the hypothesis that number in language will be subserved by the same neural mechanisms as number comprehension and quantity comparison in their analysis of grammatical number disagreement and find an increase in activation in parietal regions previously implicated in number processing (Dehaene et al., 2003) for stimuli with grammatical number violations.

Our next question, then, is: "If plural nouns elicit greater cortical activity than singular nouns, do these regions of increased activation align with regions known for either quantity and arithmetic processing or regions that that are known for linguistic processing?" We begin, as a point of departure, with the aforementioned hypothesis taken by Carreiras et al. (2010) and thus include number comprehension and quantity comparison in our literature review. As will be shown in the results section and expanded upon in the discussion, however, our results fail to support this hypothesis and we propose an explanation in the domain of the neural correlates of semantics.

Number in language is made a more interesting topic because languages can differ in their number semantics. We analyze parallel Chinese and French data, selecting these two languages specifically because of the differences in their semantics (Chierchia, 1998). Chinese lacks nominal pluralization and bare nouns (nouns without an overt determiner or quantifier) have a number interpretation which is general and includes the plural (1). In contrast, in French, count nouns are explicitly marked for grammatical number via determiners (2) which take different plural versus singular and definite versus indefinite forms. The following examples are from Rothstein (2017, pp. 147–148) and the theoretical details will be discussed in Chapter 2.

(1) a. wǒ kànjiàn gǒu le
I see dog PART
'I saw a dog/dogs, the dog(s).'

(2) a. J'ai vu #(un) chien.
I AUX saw a dog
'I saw a dog.'

While neural, cross-linguistic differences have been found in domains such as phonological access in a reading task (Paulesu et al., 2000), pitch contour processing (Gandour et al., 2003), and nominal and verbal representation (Li et al., 2004), similarities have been found for syntactic processing (see Obleser et al., 2011 for German results and Pallier et al., 2011 for French results) and comprehending linguistic content (Honey et al., 2012). Our final question, then, is: "Although they differ in their number semantics, if French and Chinese display increased activation for plural nouns over singular nouns, does that activation occur in the same or different regions and what implications does that have?"

§ 1.3 Contributions

In contrasting neural activation between plural and singular nouns, we observe several common regions of increased activation between the two languages: the left pars orbitalis, the left angular gyrus, and the left parahippocampal gyrus with additional regions occurring in the French results that do not occur in the Chinese results: the left middle temporal gyrus, the left pars triangularis, the left fusiform gyrus, left posterior cingulate cortex, and left dorsomedial prefrontal cortex.

In the context of targeting conceptual knowledge storage, these regions have previously been implicated in a distributed network for semantic processing (Binder et al., 2009), but usually only one region or one region group at a time, depending on the specific stimulus. In addition, because our stimuli are not word-pair or concept-pair contrasts, but words embedded within naturalistic sentences where there is opportunity for continuous meaning composition, and because our results are consistent with previous fMRI results for semantic composition (Graessner et al., 2020; Graves et al., 2010; Husband et al., 2011), we posit a cross-linguistic role for semantic number in semantic composition, the process by which individual concepts are combined to form complex meaning.

§ 1.4 Thesis Organization

The rest of this thesis is organized as follows: Chapter 2 reviews the relevant background literature and work related to number processing, quantity comparison, grammatical agreement, and number semantics, Chapter 3 details our datasets and methodology, Chapter 4 presents our results, Chapter 5 provides a discussion of the results, and Chapter 6 concludes.

CHAPTER 2

BACKGROUND AND RELATED WORK

§ 2:1 Number Processing and Quantity Comparison

Humans regularly use numbers and quantity, performing tasks such as counting items, evaluating arithmetic statements, and deciding whether there are more or less of one thing than some other thing. Numbers are used to understand the world. Dehaene (2001) says that humans have *number sense*, "a short-hand for our ability to quickly understand, approximate, and manipulate numerical quantities" (p. 2). Natural numbers are thought to be represented as analog magnitudes along a *mental number line* (Moyer & Landauer, 1967; Restle, 1970), with supporting evidence coming from effects for *symbolic distance* and *size*. In a task where participants determine which of two natural numbers is greater, Moyer and Landauer (1967) find that participants decide more quickly when the two numbers are farther apart (3 versus 23) than when they are closer together (3 versus 5). In a similar task, Parkman (1971) finds that when the two numbers are kept a fixed distance apart, participants take longer to decide which of two numbers is greater when the pairs are larger (23 versus 25) than when they are smaller (3 versus 5). Dehaene (1992) proposes a tripartite account of number sense in the human brain. In this *triple-code* model, three portions of the parietal lobe perform different roles in number processing (Dehaene et al., 2003). The horizontal segment of the intraparietal sulcus serves as a core quantity system which is augmented by an angular gyrus verbal system and a posterior, superior parietal visual and attentive system.

Because of its implication in number processing tasks, regardless of medium, the triple-code model posits that the horizontal segment of the intraparietal sulcus (HIPS), bilaterally, serves as the mental number line responsible for number representation and keeping track of number size and distance. The HIPS is more activated while mental arithmetic operations like multiplication or subtraction are performed than just reading numerical symbols (Chochon et al., 1999) and activation increases as arithmetic complexity increases (Menon et al., 2000). Comparing the magnitude of two numbers, a task which requires some form of mental scale, also activates the HIPS (Chochon et al., 1999). In comparison and classification tasks, Thioux et al. (2005) found that animal names activated the left inferior temporal gyrus while numbers activated the intraparietal sulci, bilaterally, supporting the domain specificity of the HIPS. HIPS activation has also been shown to be parametrically modulated. That is, intraparietal activity increases and lasts longer when calculations involve larger numbers (Kiefer & Dehaene, 1997). Lastly, number processing is thought to take place in the HIPS even unconsciously. Dehaene, Naccache, et al. (1998) show that HIPS activity during a number comparison task is influenced by visual word primes which are too brief to be perceived.

The second parietal region is the (left-lateralized) angular gyrus (AG). It is believed that this region does not process quantities directly, but acts in collaboration with other (left-lateralized) perisylvian language areas to process heard numbers. In an attempt to map functional subdivisions of the parietal lobe, Simon et al. (2002) have subjects perform six tasks: grasping, pointing, saccades, attention, calculation, and phoneme detection. The calculation task activates two parietal regions: anterior IPS, bilaterally and a more posterior, left-hemispheric portion closer to the AG which overlaps with activity from the phoneme detection task. The AG also shows greater activation for operations that can be satisfied by accessing rote verbal memory facts, such as the multiplication of small numbers, than those that require more thoughtful quantity manipulation, such as subtraction of large numbers (Lee, 2000). This is understood to be due to the association between the language network and verbal memory.

The third parietal region is the posterior, superior parietal lobe (PSPL), bilaterally. While the PSPL has been shown to be activated in number comparison (Pesenti et al., 2000), it has also been shown to be involved in orienting attention to visual targets in space (Corbetta et al., 2000). The triple-code logic, then, is that the PSPL does not just orient the attention of the eyes to written numerals, but is also involved in orienting attention on the mental number line. In a transcranial magnetic stimulation study, Göbel et al. (2001) first identify bilateral poterior parietal sites where stimulation disrupts a visuospatial search task and then test whether stimulation at those sites disrupts performance during a two digit comparison task.

They find that rTMS does slow number comparison task response, lending support to the triple-code interpretation of the PSPL.

While the triple-code model (Dehaene, 1992; Dehaene et al., 2003) is dominant, work in this field progresses. Kadosh and Walsh (2009) challenge the consensus that numerical information is encoded abstractly, where "abstract" is defined (Dehaene, Kerszberg, et al., 1998, p. 356): "Adults can be said to rely on an abstract representation of number if their behavior depends only on the size of the numbers involved, not on the specific verbal or non-verbal means of denoting them." Some of their arguments are that similar behavioral effects could result from different regions, that the assumption that numbers are represented abstractly mainly comes from null findings (that is, no differences between notation or modality which could be due to a lack of paradigm sensitivity or statistical power), and that within a cortical region, similar activity could come from different neuronal populations within the same imaged voxel. Castelli et al. (2006) examine nonsymbolic stimuli seeking to disentangle discrete from continuous, analogue quantity processing. They use a clever experimental paradigm with two conditions in which participants judge whether there is more green or blue on the screen: a *discrete* condition in which discrete blue and green rectangles are presented and an *analogue* condition in which the green and blue elements of the display are continuous and transition smoothly into one another with no distinct boundary. The discrete condition is designed to target the question "How many?" while the analogue condition is designed to target "How much?" They find greater activation in IPS, bilaterally during the processing of discrete stimuli than analogue stimuli, suggesting two distinct processes. From the syntactic domain, Hung et al. (2015), investigate the neural correlates of complex number words in French and Chinese. Complex number words are of interest because they require the merging of multiple simple number words (e.g., "thirty" and "four" into "thirty four"). They find that the number of merges required for a complex number word correlates with activity in the left inferior frontal gyrus and in the left inferior parietal lobule.

§ 2:2 Grammatical Agreement

Agreement rules play an important part in natural language understanding, particularly in heavily inflected languages. The information that they provide aides in the construction of grammatical dependencies between different parts of a sentence during syntactic parsing. Agreement is also important for connecting references across long distances like sentences during discourse. Some agreement features include gender (semantic or grammatical), number, person, and case. In Spanish, for instance, common nouns have gender and grammatical number and a phrase or sentence in which an adjective, determiner, or pronoun does not agree in number and gender with the noun to which it refers is not grammatical.

Much research has been done to study the neural correlates of grammatical agreement from an electrophysiological perspective. Electroencephalography (EEG), due to its high temporal resolution, is an excellent tool for the study of online human sentence processing. While Kutas and Hillyard (1983) observe an effect for grammatical errors which is smaller and less consistent than the N400 that they find for semantically inappropriate stimuli, Hagoort et al. (1993) find the (now) expected P600/SPS (syntactic positive shift) for grammatical and syntactic violations. The P600 event-related potential (ERP) is associated with syntactic processing and with reanalysis when the human parser encounters data that cannot be integrated into the working parse. Friederici (1995) identifies an effect for syntactic anomalies similar to the effect identified in Kutas and Hillyard (1983): the left anterior negativity (LAN). Osterhout and Mobley (1995) find P600s for number agreement violations between a personal pronoun and its antecedent, a reflexive pronoun and its antecedent and a subject and verb. They also find P600s for gender agreement violations between pronouns and antecedents. In Spanish, for both gender disagreement word pairs and number disagreement word pairs, Barber and Carreiras (2005) find an N400-like effect for noun-adjective pairs when they are presented in isolation. For article-noun pairs presented in isolation, an additional LAN effect is found. When the word pairs are placed into sentences, LAN and P600 ERPs are found. Interestingly, when the violations occurr later in the sentence, gender violations have a stronger P600 effect than number violations. Because gender is attached to lexical representation rather than being a morphological feature like number, the interpretation is that reanalysis or repair after gender violation is more difficult than after number violation.

Some neuroimaging analyses of grammatical feature and agreement are also available. In separate sessions, Miceli et al. (2002) have subjects judge whether a written noun is masculine or feminine, animal or artifact, and if it contains a /tch/ or a /k/. For the grammatical feature task, they find increased activation in the inferior frontal gyrus (BA 45) compared to the baseline and the semantic tasks. They also find increased activation for the grammatical feature task in left middle/inferior temporal gyrus (BA 20/21), compared to the phonological task. Carreiras et al. (2010) present subjects with Spanish determiner-noun and noun-adjective word pairs of three conditions: gender and number agreement, gender violation, but number agreement, and gender agreement, but number violation. For both number and gender disagreement, they observe increased activation in left inferior frontal areas. Additionally, they find an increase in the right IPS and right superior parietal gyrus for number violation compared to baseline and gender violation. This effect, though, comes mainly from the determiner and noun word pairs and the authors speculate that that this activation is not regular during language processing, but that triple-code quantity comparison (Dehaene et al., 2003) results specifically from the number disagreement. For the identified parietal regions, they find no significant effects when contrasting plural > singular or singular > plural.

More neuroimaging studies are available analyzing syntax and semantic in a more general manner. In comparing the processing of sentences with syntactic and semantic violations, Friederici et al. (2003) find increased activation in the frontal operculum for syntactically, but not semantically anomalous sentences. In a study where linguistic constituent size is parametrically modulated, Pallier et al. (2011) identify inferior frontal and a posterior temporal region which show effects for constituent size regardless of whether the presented string consists of real words or pseudowords (jabberwocky stimuli). This suggests that these regions perform syntactic processing automatically, whether any meaningful content is present or not. In contrast, the temporal pole, anterior superior temporal sulcus, and temporo-parietal junction show effects for meaningful stimuli only.

§ 2:3 Number Semantics

A count noun (e.g., *cat*) is a noun which may be directly modified by cardinal numericals and a mass noun (e.g., *sand*) is a noun which cannot. While (3a) is perfectly acceptable, (3b) is not acceptable on the intended reading. There is a connection between the count/mass distinction and the counting/measuring distinction. While count nouns are counted (3a), mass nouns are measured (3c). It is not the case, however, that all languages make the count/mass distinction.

(3) a. two cats
    b. #two sands
    c. three buckets of sand

Chierchia (1998) proposes a three-way typological classification for languages based upon how they express counting. Chierchia's account is neo-Carlsonian, that is, it is based upon Carlson's (1977) investigation of bare plurals in English which proposes that nouns can either be predicates at type <*e*,*t*>, in which they denote a set of entities, or be arguments at type *e*, in which they denote kinds. The terms *predicate* and *argument*, here, are names for the semantic types <*e*,*t*> (functions from individuals to truth values) and *e* (entities of argumental type), respectively. *Kinds* are generally understood as regularities. For the property of being a cat, there is a corresponding kind: the cat-kind. In the other direction, a kind will have a property with which it corresponds: the property of belonging to the kind.

A noun (N) may fill an argument position if it is an argument, but if it is a predicate, it must combine with a determiner to reach the argument type. Chierchia's classification, then, is whether nouns in a language can occur as arguments, predicates, or both. From the features [+/-predicate] and [+/-argument], there are three possible language types: [+predicate, +argument], [-predicate, +argument], and [+predicate, -argument] ([-predicate, -argument] is not valid). English is [+predicate, +argument], Chinese is [-predicate, +argument], and French is [+predicate, -argument]. Chierchia argues that a language will have morphosyntactic properties based upon its features. The following section reviews these properties with data from Rothstein (2017, pp. 147–148).

With English being [+predicate, +argument], the nouns of English are either [+predicate] or [+argument]. Count nouns are predicates and mass nouns are arguments. Because they are predicates, singular count nouns must combine with a determiner to fill an argument position and it is predicted that bare singular count nouns are ungrammatical (4a). Plural count nouns can be shifted such that they yield a kind reading and and thus can occur as bare arguments. Mass nouns can occur as bare singulars in argument position (4b).

(4) a. I saw #(a) dog.

b. I bought wine.

Chinese, like other classifier languages, is [-predicate, +argument]. In these languages, bare singular nouns can occur as arguments (5). While nouns may occur as bare singulars, they may not be directly modified by cardinal numericals. Instead of directly taking bare nouns as complements, numericals take classifier + N sequences (6).

(5) a. wǒ kànjiàn gǒu le
I see dog PART
'I saw a dog/dogs, the dog(s).'

b. wǒ mǎi le jiǔ
I buy PFV wine
'I bought wine.'

(6) a. sān #(zhī) gǒu
three $\text{Cl}_{\text{small animal}}$ dog
'three dogs'

b. liǎng #(kē) shù
two Cl_{plant} tree
'two trees'

In an analysis of bare noun phrases in Chinese, Yang (2001) identifies the same readings identified by Carlson (1977) for English bare plurals: kind, generic, and narrowest-scope indefinite. While French and English necessarily mark definite NPs with determiners, Chinese has no determiners and bare NPs have definite readings that are not available in English. Since all nouns have the same properties, and no N can be directly modified by a numeral, there is no clear way to differentiate mass and count nouns grammatically. As compared to languages with mass/count distinction ([+predicate, +/- argument]), Chierchia's (1998) view is that in [-predicate, +argument] languages, every lexical noun is mass-like. Because the plural operator does not apply to kind or mass terms, classifier languages do not have nominal pluralization. Bare nouns in these languages have a number interpretation which is general and includes the plural (5a).

French, like other Romance languages, is [+predicate, -argument] and makes the count/mass distinction. Count nouns will be marked either singular or plural and all nouns (both count and mass) must occur with a determiner (7).

(7) a. J'ai vu #(un) chien.
I AUX saw a dog
'I saw a dog.'

b. J'ai acheté #(du) vin.
I AUX bought some wine
'I bought (some) wine.'

Interestingly, French is slightly more strict with its determiner requirement than Spanish and Italian which allow for bare plurals in well-governed conditions such as object (but not subject) position. The allowed bare plurals do not have kind or generic readings, though. Because English is [+predicate, +argument], its count nouns are similar to French nouns and its mass nouns are similar to Chinese nouns.

CHAPTER 3

DATA AND METHODS

§ 3.1 fMRI Datasets

The French dataset includes 30 healthy, right-handed adults (age range = 20-40). They self-identified as native French speakers and had no history of psychiatric, neurological, or other medical illness that could compromise cognitive functions. All participants gave written informed consent prior to participation, in accordance with the guidelines of the Regional Committee for the Protection of Persons involved in Biomedical Research.

The Chinese dataset includes 35 healthy, right-handed, young adults (15 female, mean age = 19.3, range = 18-25). They self-identified as native Chinese speakers and had no history of psychiatric, neurological, or other medical illness that could compromise cognitive functions. All participants were paid and gave written informed consent prior to participation, in accordance with the guidelines of the Ethics Committee at Jiangsu Normal University.

The French audio stimulus is an audiobook version of *The Little Prince* (*Le Petit Prince*) (de Saint-Exupéry, 1946), read by Nadine Eckert-Boulet. The Chinese audio stimulus is a Chinese translation of *The Little Prince*, read by a professional female Chinese broadcaster. The French audiobook lasts 98 minutes and the Chinese audiobook lasts 99 minutes. The presentation is divided into nine sections, each around ten minutes in length. Participants listened passively to the nine sections and completed quiz questions after each section. The questions were used to confirm participant comprehension.

A problem in fMRI data acquisition is that factors such as cardiac and respiratory activity and subject motion introduce noise which obscures the desired blood-oxygen-level dependent (BOLD) signal. See section 3.2 for a more detailed explanation of the BOLD signal. While registration (Jonsson et al., 1999) and filtering (Biswal et al., 1996) strategies have been developed for reducing noise for echo-planar imaging

(EPI), (Kundu et al., 2012) present a technique for multi-echo EPI which uses independent component analysis (ME-ICA) for removing noise and artifacts. Assuming mono-exponential decay, the signal across echo times is described by the function:

$$S(TE_n) = S_0 exp(\quad R_2 \quad TE_n)$$

where n is the echo number, R_2 is the relaxation rate, and S_0 ($TE = 0$) is initial signal intensity. In ME-ICA, ICA is used to decompose multi-echo data into components. The components are then scored based on degree to which their signal changes fit models for R_2 change and S_0 change. With high R_2 and low S_0 scores being BOLD-like and and low R_2 and high S_0 scores being noise-like, by identifying the noise and artifact components, the data can be denoised. ME-ICA was applied as part of the preprocessing pipeline for both of these datasets before we received them.

The French brain imaging data was collected with a Siemens Prisma fit 3T scanner. T1-weighted anatomical images were acquired with a 1 mm isotropic resolution. The EPI functional images were acquired with a resolution of 3:75 x 3:75 x 3:8 mm (34 axial slices with an interleaved acquisition scheme). The 3 echo times were 10 ms, 25 ms, and 38 ms. Preprocessing was performed with ME-ICA (Kundu et al., 2012) using the default parameters and spatial normalisation was done in the Montreal Neurological Institute (MNI) space. The final volumetric time series consists of 3:15 mm cubic voxels.

The Chinese brain imaging data were acquired with a 3T MRI GE Discovery MR750 scanner with a 32-channel head coil. Anatomical scans were acquired using a T1-weighted volumetric magnetization-prepared-rapid-gradient-echo pulse sequence. BOLD functional scans were acquired using a multi-echo EPI sequence with online reconstruction (TR = 2000 ms; TE's = 12:8, 27:5, 43 ms; FA = 77°; matrix size = 72 x 72; FOV = 240:0 mm x 240:0 mm; 2 x image acceleration; 33 axial slices; voxel size = 3:75 x 3:75 x 3:8 mm). The Chinese fMRI data were preprocessed using AFNI version 16 (Cox, 1996). The first 4 volumes in each run were excluded from analyses to allow for T1-equilibration effects. ME-ICA (Kundu et al., 2012) was used to denoise data for motion, physiology, and scanner artifacts. Images were then spatially normalized to the standard space of the MNI atlas, yielding a volumetric time series resampled at 2 mm cubic voxels.

§ 3:2 GLM Analysis of fMRI Data

Active neurons in the brain require oxygenated blood, with an increase in activation requiring an increase in oxygenated blood. The function which describes the increase in blood flow after brief neuronal activity is known as the *hemodynamic response function*, or HRF. It can be described as (Poldrack et al., 2011, p. 71): "the ideal, noiseless response to an infinitesimally brief stimulus." The HRF rises within 1-2 seconds after stimulus onset, reaching a peak within 4-6 seconds, and then returns to baseline 12-20 seconds after stimulus onset, typically with a brief undershoot of the baseline. Some researchers report an initial dip in the BOLD signal 1-2 seconds after stimulus onset which may be the result of quick, initial oxygen consumption (Buxton, 2012).

When neuronal population activity increases, the concentration of oxygenated hemoglobin increases, through factors such as blood flow, blood volume, and blood oxygenation. This increases the homogeneity of magnetic susceptibility, which in turn increases the T2*-weighted magnetic resonance signal. This signal is called the the *blood oxygenation level dependent* signal, or BOLD signal. In this way, by measuring BOLD signal, functional magnetic resonance imaging (fMRI) can be used to indirectly measure neuronal activity.

With respect to neural response, the BOLD signal is said to have *linear time-invariant* (LTI) properties (Boynton et al., 1996). This can be broken down into *linear* properties and *time-invariant* properties. By linearity, it is meant that if neural response is scaled by some factor x, then the BOLD signal is scaled by the same factor x. Linearity also applies to more than one event placed closely in time. The signal resulting from two close-together events is the sum of the two independent signals. By time-invariant, it is meant that if the stimulus is shifted s seconds, the BOLD signal will be shifted by s seconds.

In performing an fMRI analysis, we look at each voxel's BOLD signal through time to see if it changes in response to our stimuli. A voxel is a cubic parcellation of 3-dimensional space. In order to fit and identify this variation, we use a general linear model (GLM) where the observed BOLD time series is the dependent variable and the independent variables are the time series of our stimuli convolved with the HRF. As previously mentioned, while neuronal activity may be infinitesimally brief, the peak of the resulting BOLD signal will lag behind the stimulus approximately 5 seconds and not return to baseline for upwards of 20 seconds. In order to better model the BOLD signal, we convolve our stimuli time series with the HRF, that is, we blend the two functions together in an LTI manner. Analyses of BOLD data have characterized the HRF as a gamma function (Friston et al., 1994) and a single gamma function was commonly used as the canonical HRF until researchers began incorporating the post-timulus undershoot, which cannot be accounted for with a single gamma function. To incorporate the undershoot, a *double-gamma HRF* (Glover, 1999) can be used, where the first gamma function models the initial stimulus response and the second gamma function models the undershoot.

In the GLM, one or more predictors are related to a single, continuous response variable . The most basic form is simple linear regression which has only one predictor and takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

where Y is a vector of T values of the dependent variable, β_0 is the point at which the line described by the model crosses the y axis (intercept), β_1 is the slope of the line, X_1 is a vector of T values of the independent variable, and ϵ is a vector of T error terms. Simple linear regression can be extended from one predictor to p predictors $X_1, X_2, \ldots, X_p$ with *multiple linear regression* which takes the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p + \epsilon$$

where each parameter β_i is the effect of X_i, controlling for all other variables. It can be more concisely presented:

$$Y = X\beta + \epsilon$$

where X is a T x p matrix in which each column is an X_i and β is a column vector of length p + 1. Put another way, $\beta = [\beta_0, \beta_1, \ldots, \beta_p]'$.

By minimizing the squared distance between the data, Y, and estimates, $\hat{Y} = X\hat{\beta}$, known as the method of *least squares*, we can estimate the parameters β_i. The difference, $e = Y - \hat{Y}$, is known as the residual. We cannot simply solve the equation by multiplying each side by X^{-1} because X is not a square matrix and only square matrices have inverses. Rather, we first calculate the *normal equations* by multiplying each side by X':

$$X'Y = X'X\hat{\beta}$$

which can be solved by multiplying both sides by $(X'X)^{-1}$:

$$\hat{\beta} = (X'X)^{-1}X'Y$$

The parameter estimates $\hat{\beta}$ which satisfy the normal equations give the least squares solution by minimizing the sum of squares of the residual, $e'e$. This is assuming that the inveserse of $X'X$ exists, which requires that X have full column rank, that is, no columns which can be written as a linear combination of the other columns. The columns of X must be linearly independent. If the inverse does not exist, the normal equations can still be solved, but the solution may not be unique.

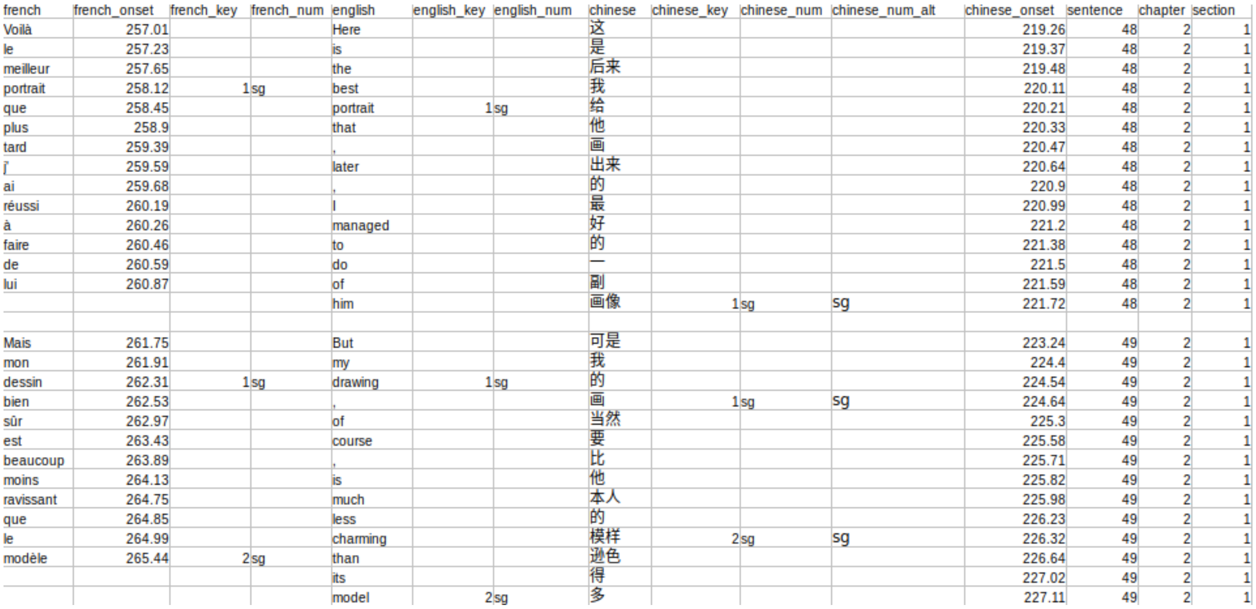
The GLM is used to analyze fMRI data at one of two levels. In a *first level analysis* (or subject level analysis), for each subject, for each voxel, a model is fit relating the convolved regressors to that voxel's activity time series. Each model will have a collection of β coefficients (one for each regressor) which reflect a regressor's effect on the signal when the others are held constant. One way to make use of these fitted models is to perform a *contrast*, that is, compare the activation associated with one regressor against another. A *subtraction* paradigm is common, in which the activity associated with one regressor (say, b) is subtracted from the activity associated with another (say, a). This results in a voxel-level *brain map* in which the values are either raw effect size ($\beta_a - \beta_b$), t-scores, or z-scores. In this way, we can identify voxels which more strongly respond to one regressor than another.

At the *second level* (or group level), the GLM relates subject groupings to the voxel-level estimates of the contrast in the first-level statistical maps ($a - b$). The *design matrix*, or regressor matrix, X, and contrast can be set up for a number of different analyses. A design matrix with a single column of 1s, one for each subject, and a contrast of $c = [1]$ is used to perform a one-sample t-test (H_0 : Overall mean $= 0$, equally $H_0 : c\beta = 0$). A design matrix with two columns, where a row value of $[1, 0]$ indicates belonging to group 1 and a row value of $[0, 1]$ indicates belonging to group 2 and a contrast of $c = [1, -1]$ is used to perform a two-sample t-test (H_0 : Mean of group 1 is different from group 2, equally $H_0 : \mu_{G1} - \mu_{G2} = 0$). Per its name, the GLM can be generalized to additional tests such as paired t-tests, ANOVAs, and ANCOVAs by changing the design matrix and contrast. The result is a second level statistial brain map.

Because the analysis described here models voxel-by-voxel BOLD signal, it is also referred to as a mass univariate analysis. This is in contrast to multivariate or multivoxel pattern analysis (MVPA) techniques such as decoding or searchlight (Kriegeskorte et al., 2006). It is possible that within a brain region, two predictors fail to contrast, but within the region, patterns of activation are still exhibited which would indicate that the region is involved. In this case, multivariate analysis is useful. For example, a full-brain decoding analysis could use the voxel-level activations as the input features to a classification algorithm for distinguishing experimental conditions. In a searchlight analysis, a "sphere" is moved throughout the brain, centering on each voxel, with the activations for the voxels within the sphere used as input features for a classification algorithm. In this way, regions whose patterns of activation have high classification accuracies for an experimental condition can be identified.

§ 3:3 Observations of Interest

In order to control for discourse factors which could modulate neural activity during naturalistic language processing, we align the storybook texts and select only parallel nouns for analysis, that is, nouns which occur in both stories and in the same context. The first step in this process is aligning sentences, which is done with the Hunalign bilingual sentence aligner (Varga et al., 2007). The Hunalign aligner makes use of both a bilingual dictionary (Chen, 1993) and sentence length. Gale and Church (1993) describe a bilingual sentence aligner based upon the principle that longer sentences in one language tend to be translated into longer sentences in another language and that shorter sentences in one language tend to be translated into shorter sentences in another language. In this way, for a proposed aligned sentence pair, a probabilistic score can be computed using the scaled difference in character lengths of the sentences. The probabilistic scores can then be used to find the maximum likelihood alignment of the sentences. If a bilingual dictionary is not available, Hunalign first aligns sentences based upon length information, creates a dictionary based upon this initial alignment, and then makes a second pass using the constructed dictionary. Hunalign was used to align the storybook texts for Chinese, French, and English and the alignments were checked and corrected by hand.

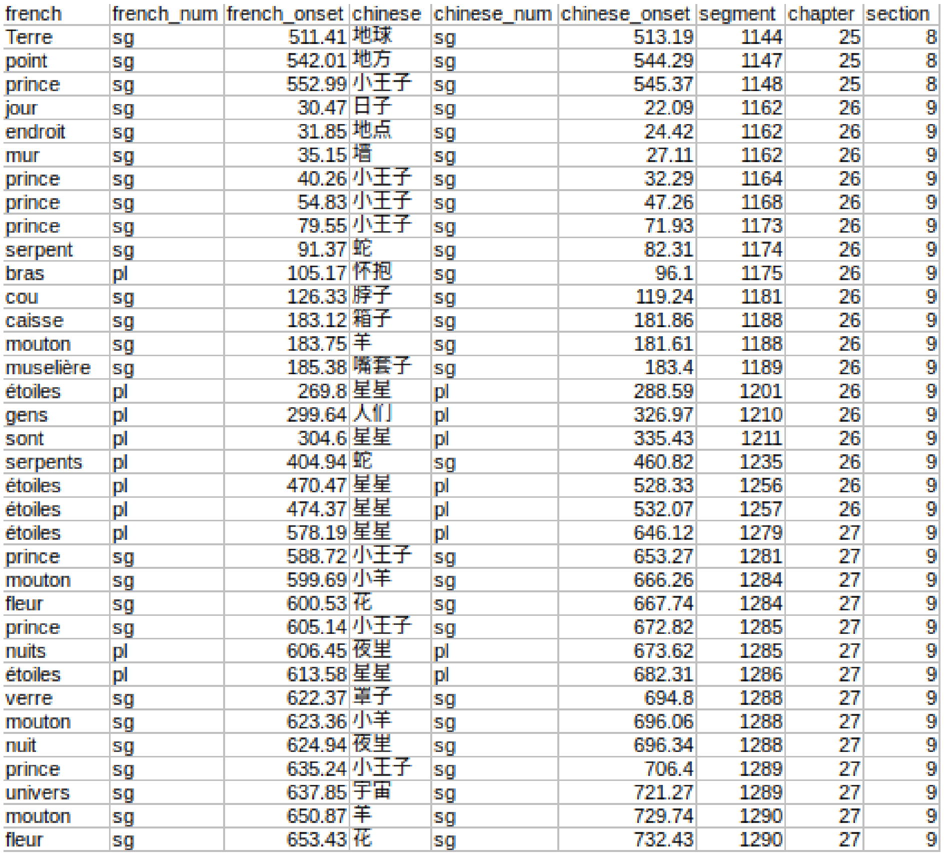


french	french_onset	french_key	french_num	english	english_key	english_num	chinese	chinese_key	chinese_num	chinese_num_alt	chinese_onset	sentence	chapter	section
Voilà	257.01			Here			这				219.26	48	2	1
le	257.23			is			是				219.37	48	2	1
meilleur	257.65			the			后来				219.48	48	2	1
portrait	258.12	1	sg	best			我				220.11	48	2	1
que	258.45			portrait	1	sg	给				220.21	48	2	1
plus	258.9			that			他				220.33	48	2	1
tard	259.39			,			画				220.47	48	2	1
j'	259.59			later			出来				220.64	48	2	1
ai	259.68			,			的				220.9	48	2	1
réussi	260.19			I			最				220.99	48	2	1
à	260.26			managed			好				221.2	48	2	1
faire	260.46			to			的				221.38	48	2	1
de	260.59			do			一				221.5	48	2	1
lui	260.87			of			副				221.59	48	2	1
				him			画像	1	sg	sg	221.72	48	2	1
Mais	261.75			But			可是				223.24	49	2	1
mon	261.91			my			我				224.4	49	2	1
dessin	262.31	1	sg	drawing	1	sg	的				224.54	49	2	1
bien	262.53			,			画	1	sg	sg	224.64	49	2	1
sûr	262.97			of			当然				225.3	49	2	1
est	263.43			course			要				225.58	49	2	1
beaucoup	263.89			,			比				225.71	49	2	1
moins	264.13			is			他				225.82	49	2	1
ravissant	264.75			much			本人				225.98	49	2	1
que	264.85			less			的				226.23	49	2	1
le	264.99			charming			模样	2	sg	sg	226.32	49	2	1
modèle	265.44	2	sg	than			逊色				226.64	49	2	1
				its			得				227.02	49	2	1
				model	2	sg	多				227.11	49	2	1

Figure 3.1: Sample French, English, and Chinese text, aligned at the sentence level with observations of interest marked with keys and for semantic number

Next, we identify the parallel nouns and filter the pairs with criteria which serve to maximize the typological distinction between French and Chinese. For the Chinese observations, we include only nouns which have no overt number marking, either morphological or through a number and classifier construction. For the French observations, we include only nouns indexed by the definite, common determiners: le, la, l', and les. While grammatical number annotation can be automated for the French nouns: le, la, and l' are singular and les is plural, annotation for the Chinese nouns is more challenging because number is not overtly marked and it is possible that different listeners will have different judgements. Because of this, we have two native Chinese speakers annotate the Chinese nouns and then calculate Cohen's kappa coefficient (Cohen, 1960), a measure of inter-rater reliability. The result is kappa = $0:96$, a high degree of inter-rater reliability. We do not use any nouns in the analysis for which the two annotators disagreed in their number judgements.

The time resolution of both of our fMRI data sets is $2:0$ seconds, much slower than a natural speech rate. Because of this, we remove observations where nouns of different number would occur together within the same volume. That is, if more than one singular noun occur in the same volume or if more than one plural noun occur in the same volume, they are retained. If a singular and plural noun occur in the same volume, however, they are not kept for analysis. After this, we end up with 288 parallel observations: 261 singular and 27 plural in the Chinese text and 263 singular and 25 plural in the French text.



french	french_num	french_onset	chinese	chinese_num	chinese_onset	segment	chapter	section
Terre	sg	511.41	地球	sg	513.19	1144	25	8
point	sg	542.01	地方	sg	544.29	1147	25	8
prince	sg	552.99	小王子	sg	545.37	1148	25	8
jour	sg	30.47	日子	sg	22.09	1162	26	9
endroit	sg	31.85	地点	sg	24.42	1162	26	9
mur	sg	35.15	墙	sg	27.11	1162	26	9
prince	sg	40.26	小王子	sg	32.29	1164	26	9
prince	sg	54.83	小王子	sg	47.26	1168	26	9
prince	sg	79.55	小王子	sg	71.93	1173	26	9
serpent	sg	91.37	蛇	sg	82.31	1174	26	9
bras	pl	105.17	怀抱	sg	96.1	1175	26	9
cou	sg	126.33	脖子	sg	119.24	1181	26	9
caisse	sg	183.12	箱子	sg	181.86	1188	26	9
mouton	sg	183.75	羊	sg	181.61	1188	26	9
muselière	sg	185.38	嘴套子	sg	183.4	1189	26	9
étoiles	pl	269.8	星星	pl	288.59	1201	26	9
gens	pl	299.64	人们	pl	326.97	1210	26	9
sont	pl	304.6	星星	pl	335.43	1211	26	9
serpents	pl	404.94	蛇	sg	460.82	1235	26	9
étoiles	pl	470.47	星星	pl	528.33	1256	26	9
étoiles	pl	474.37	星星	pl	532.07	1257	26	9
étoiles	pl	578.19	星星	pl	646.12	1279	27	9
prince	sg	588.72	小王子	sg	653.27	1281	27	9
mouton	sg	599.69	小羊	sg	666.26	1284	27	9
fleur	sg	600.53	花	sg	667.74	1284	27	9
prince	sg	605.14	小王子	sg	672.82	1285	27	9
nuits	pl	606.45	夜里	pl	673.62	1285	27	9
étoiles	pl	613.58	星星	pl	682.31	1286	27	9
verre	sg	622.37	罩子	sg	694.8	1288	27	9
mouton	sg	623.36	小羊	sg	696.06	1288	27	9
nuit	sg	624.94	夜里	sg	696.34	1288	27	9
prince	sg	635.24	小王子	sg	706.4	1289	27	9
univers	sg	637.85	宇宙	sg	721.27	1289	27	9
mouton	sg	650.87	羊	sg	729.74	1290	27	9
fleur	sg	653.43	花	sg	732.43	1290	27	9

Figure 3.2: Sample parallel French and Chinese nouns

§ 3:4 Statistical Analyses

We run separate French and Chinese GLM analyses using Nistats and Nilearn (Abraham et al., 2014; Pedregosa et al., 2011). At the first level, we include binary regressors for singular and plural nouns as well as coregressors for the root mean squared amplitude of the spoken narration (RMS) and spoken word rate. The singular and plural noun regressors are marked with a 1 at the onset of the nouns of interest, RMS is marked every 10 ms, and word rate is marked with a 1 at the onset of every word, except for the observations of interest. The coregressors are added to ensure that any results found are due to the differences between singular and plural nouns and not just effects of spoken language comprehension. The software adds thirteen drift and a constant coregressor. The first-level GLM is then:

$$BOLD \quad Plural + Singular + RMS + WordRate$$

It is difficult to form a markedness hypothesis for Chinese, with respect to the singular-plural contrast, as the bare nouns which we intentionally select for analysis are said to have "general number" (Rullmann & You, 2006, p. 1). On the other hand, in French, singular is taken to be the default form with plural as the non-default (de Swart & Farkas, 2010). In this vein, for both analyses, our first-level contrast subtracts singular from plural activation, producing t-value brain maps.

At the second level, we use the GLM to perform one-sample t-tests: "is the difference between plural and singular activation greater than 0 ?" We apply an 8 mm full width at half maximum Gaussian smoothing kernel to counteract inter-subject anatomical variation. Because we are analyzing tens of thousands of voxels, it is important to correct for multiple testing.

When performing more than one hypothesis test (a *family* of tests), the *familywise error rate* (FWER) is the probability of making one or more false discoveries. That is, the probability of making at least one type 1 error, the incorrect rejection of a true null hypothesis. It can be estimated with the formula:

$$FWER \quad 1 \quad (1 \quad _{it})^{c}$$

where $_{it}$ is the alpha level for an individual test (e.g., $0:05$) and c is the number of comparisons. Testing $50;000$ voxels at an alpha level of 5% would lead to a nearly 100% chance of making a false discovery. This can be counteracted, however, by dividing the alpha level by the number of tests performed, which brings FWER $\quad 0:05$. This is known as Bonferroni correction and in order for the result of a test to be significant, it must have a p-value less than or equal to the original alpha divided by the number of tests. While it is known to be conservative, FWER corrected results are customary in neurolinguistic fMRI analyses. To this end, we apply Bonferroni correction and the main-result brain maps are in terms of z-score with FWER < $.05$.

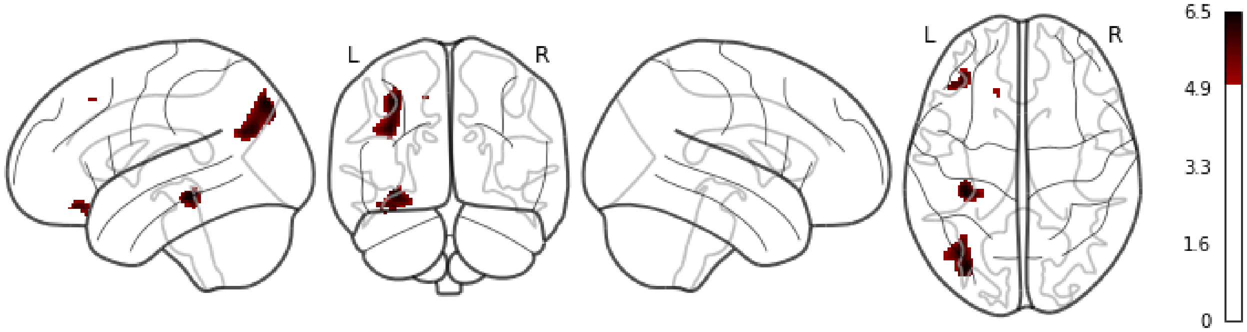
We want to identify any common regions of increased activation between the two languages. Because the Chinese data are higher resolution than the French data, so too are the results. Because of this, we downsample the Chinese results to the resolution of the French results. With the resulting statistical maps in the same space, we, voxel-by-voxel, take the intersection of the two. If a voxel's statistical result is significant in both language maps, it is recorded. This allows us to construct an intersection map identifying overlapping regions of increased activation. Lastly, while discourse-relevant semantic number is conveyed to the listener in both languages, it is overtly marked in the French, but not in the Chinese. Because of this, we run an additional second level analysis testing whether the increase in activation for plural over singular observations is greater in the French participants than the Chinese participants.

CHAPTER 4

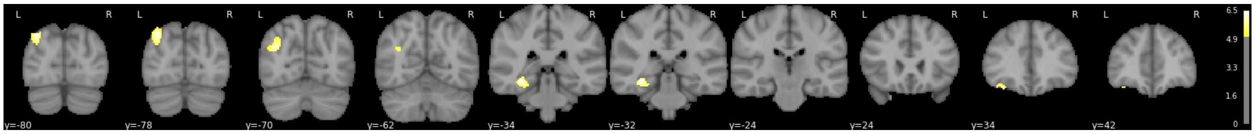
RESULTS

§ 4:1 Chinese Results

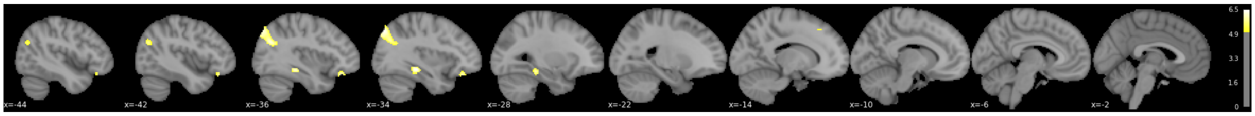
For the Chinese participants, we find 4 regions where activation for plural nouns is significantly greater than activation for singular nouns: the left parahippocampal gyrus (BA 36), the left angular gyrus (BA 39) extending into left visual association cortex (BA 19), the left pars orbitalis (BA 47), and a small cluster in left dorsomedial prefrontal cortex (BA 8).



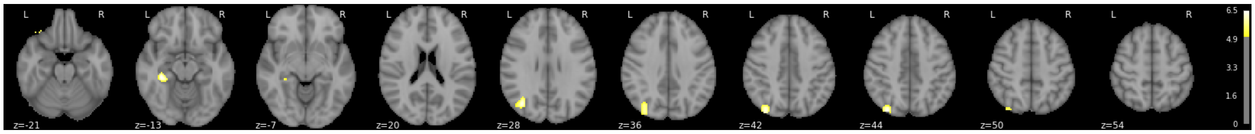
(a) z-score, glass brain



(b) z-score, coronal slices



(c) z-score, saggital slices



(d) z-score, axial slices

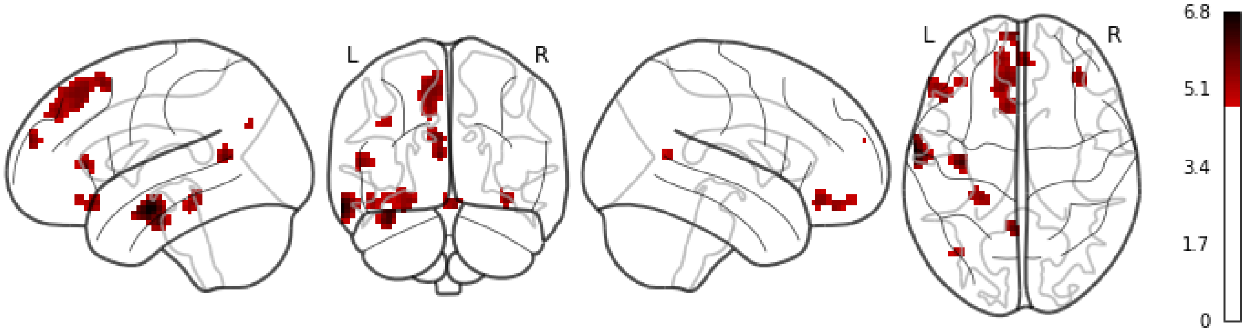
Figure 4.1: PLURAL > SINGULAR z-maps for Chinese, thresholded at FWER < 0:05

Table 4.1: PLURAL > SINGULAR clusters in Chinese, thresholded at FWER $< 0{:}05$. Region and Brodmann area labels come from the Yale BioImage Suite (Lacadie et al., 2008)

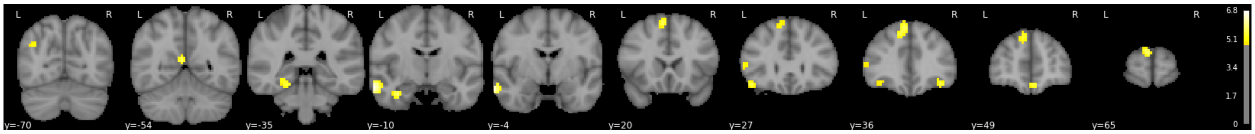
Region	Cluster size	MNI coordinates			z-score (peak level)
	(mm^3)	x	y	z	
L Parahippocampal Gyrus (BA 36)	992	-31.0	-36.0	-12.0	6.52
L Angular Gyrus (BA 39)	3224	-34.0	-80.0	44.0	6.42
L Visual Association Cortex (BA19)		-38.0	-70.0	28.0	6.33
L Pars Orbitalis (BA 47)	448	-34.0	38.0	-20.0	6.10
		-42.0	34.0	-20.0	5.88
L Dorsomedial Prefrontal Cortex (BA 8)	40	-16.0	28.0	46.0	5.17

§ 4:2 French Results

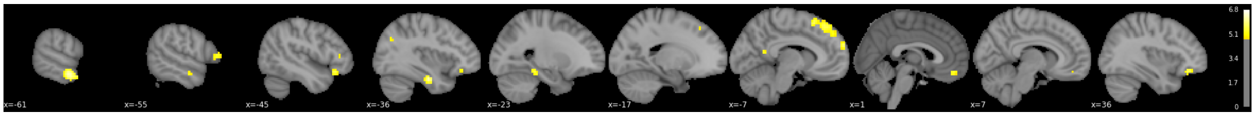
For the French participants, we find a number of regions where activation for plural nouns is significantly greater than singular nouns: the left middle temporal gyrus (BA 21), the left parahippocampal gyrus (BA 36), left posterior cingulate cortex (BA 23), left dorsomedial prefrontal cortex (BAs 8, 10), right orbitofrontal cortex (BA 11), the left and right pars orbitalis (BA 47), the left pars triangularis (BA 45), the left fusiform gyrus (BA 37), and the left angular gyrus (BA 39).



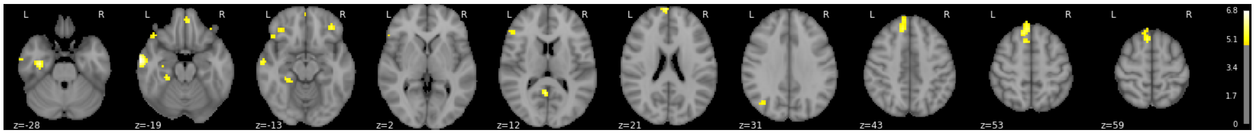
(a) z-score, glass brain



(b) z-score, coronal slices



(c) z-score, saggital slices



(d) z-score, axial slices

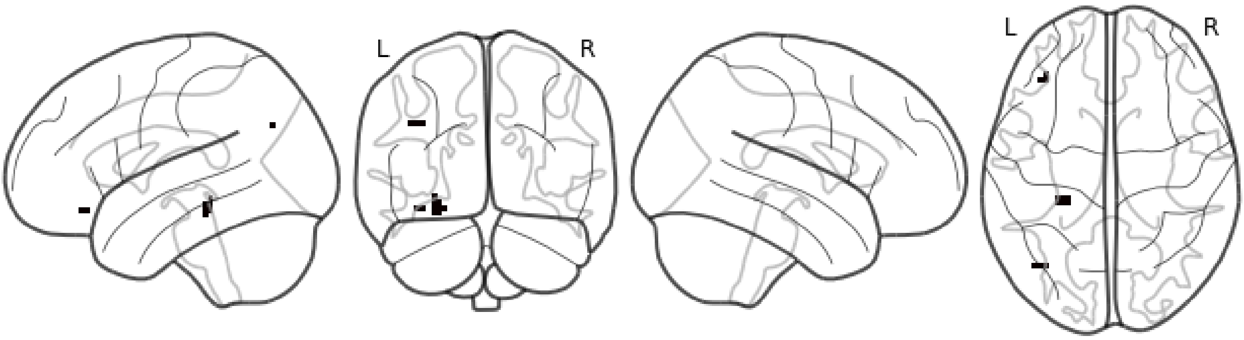
Figure 4.2: PLURAL > SINGULAR z-maps for French, thresholded at FWER < 0:05

Table 4.2: PLURAL > SINGULAR clusters in French, thresholded at FWER < $0{:}05$. Region and Brodmann area labels come from the Yale BioImage Suite (Lacadie et al., 2008)

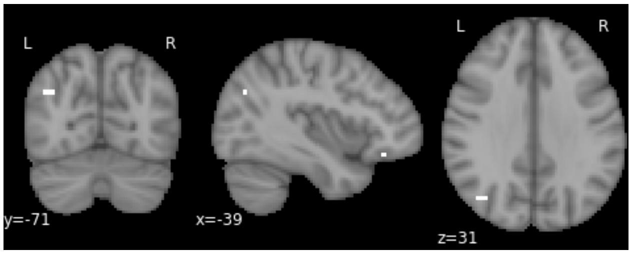
Region	Cluster size	MNI coordinates			z-score (peak level)
	(mm^3)	x	y	z	
L Middle Temporal Gyrus (BA 21)	2239	-65.0	-8.0	-19.0	6.79
L Parahippocampal Gyrus (BA 36)	1135	-37.0	-16.0	-26.0	6.22
L Posterior Cingulate Cortex (BA 23)	536	-5.0	-55.0	12.0	5.84
L Dorsomedial Prefrontal Cortex (BA 10)	662	-5.0	65.0	22.0	5.82
L Dorsomedial Prefrontal Cortex (BA 8)	4635	-11.0	30.0	53.0	5.68
		-8.0	21.0	56.0	5.66
		-11.0	49.0	41.0	5.51
R Pars Orbitalis (BA 47)	567	36.0	37.0	-16.0	5.60
L Pars Orbitalis (BA 47)	567	-46.0	27.0	-16.0	5.60
L Pars Triangularis (BA 45)	883	-52.0	30.0	9.0	5.52
L Fusiform Gyrus (BA 37)	851	-27.0	-3.0	-14.0	5.51
R Orbitofrontal Cortex (BA 11)	630	5.0	46.0	-19.0	5.49
L Pars Orbitalis (BA 47)	346	-36.0	37.0	-13.0	5.14
L Angular Gyrus (BA 39)	220	-39.0	-71.0	31.0	5.08

§ $4{:}3$ Intersection Main Results

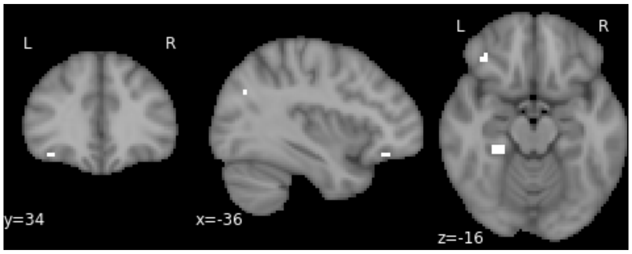
Taking the intersections of the Chinese and French statistical maps, we find three regions in common: the left angular gyrus (BA 39), the left pars orbitalis (BA 47), and the left parahippocampal gyrus (BA 36).



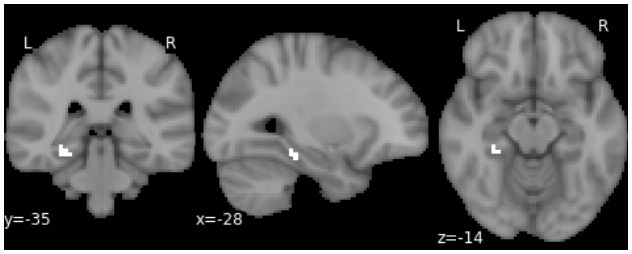
(a) Chinese and French intersection, glass brain



(b) Chinese and French intersection, angular gyrus



(c) Chinese and French intersection, pars orbitalis



(d) Chinese and French intersection, parahippocampal gyrus

Figure 4.3: Intersection of Chinese and French maps

Table 4.3: Intersection of Chinese and French. Region and Brodmann area labels come from the Yale BioImage Suite (Lacadie et al., 2008)

Region	Cluster size	MNI coordinates		
	(mm^3)	x	y	z
L Angular Gyrus (BA 39)	94	-39.0	-71.0	31.0
L Pars Orbitalis (BA 47)	94	-36.0	34.0	-16.0
L Parahippocampal Gyrus (BA 36)	378	-28.0	-35.0	-14.0

§ 4.4 Difference Between French and Chinese

We find no regions where the French contrast is significantly greater than the Chinese contrast.

CHAPTER 5

DISCUSSION

In contrasting neural activation between plural and singular nouns, we observe several common regions of increased activation between the two languages: the left pars orbitalis, the left angular gyrus, and the left parahippocampal gyrus, with additional regions occurring in the French results that do not occur in the Chinese results: the left middle temporal gyrus, the left pars triangularis, left posterior cingulate cortex, left dorsomedial prefrontal cortex, and the left fusiform gyrus.

Our findings do not align with what would be expected for triple-code number processing and quantity comparison (Dehaene et al., 2003). In the context of targeting conceptual knowledge storage, these regions have previously been implicated in a distributed network for semantic processing (Binder et al., 2009), but usually only one region or one region group at a time, depending on the specific stimulus. In addition, because our stimuli are not word-pair or concept-pair contrasts, but words embedded within naturalistic sentences where there is opportunity for continuous meaning composition, and because our results are consistent with previous fMRI results for semantic composition (Graessner et al., 2020; Graves et al., 2010; Husband et al., 2011), we posit a cross-linguistic role for semantic number in semantic composition, the process by which individual concepts are combined to form complex meaning. Lastly, we discuss the similarities and differences between the results and the role of linguistic typology in the neurobiology of language research domain.

§ 5:1 Quantity Comparison

The triple-code model (Dehaene, 1992) proposes a tripartite account for making sense of numbers and quantities with three portions of the parietal lobe facilitating this in different manners (Dehaene et al., 2003). The horizontal segment of the intraparietal sulcus (HIPS) serves as an internal number line which keeps track of size and distance between numbers and which is responsible for number representation, the left angular gyrus (AG) aids in processing heard numbers without processing quantities directly, and the posterior, superior parietal lobe (PSPL) orients attention both in space and on the internal number line. While the HIPS would be a plausible candidate for our plural > singular contrast, a significant difference in activation is not observed there. Even though Carreiras et al. (2010) observe an increase in activation for grammatical number disagreement in the right HIPS and right PSPL, they believe it unlikely that the parietal regions are specifically involved when processing language and it more likely that the activation is from quantity computation engaged by the grammatical judgement task. Indeed, they perform singular > plural and plural > singular contrasts for the identified parietal regions, but find no effect. While we do identify the left AG, we, importantly, do not identify the HIPS in our contrast of plural and singular nouns. If quantity comparison specific processing were happening, we would expect to identify the HIPS.

§ 5.2 Semantic Processing

Humans acquire and store information about about the world around them, learning the characteristic features (such as size, shape, color, and sound) of the things with which they interact. With the relationship between this collected knowledge and words known as *semantics*, the cognitive act of accessing the collected knowledge, or semantic memory, is known as *semantic processing*. In a meta-analysis of 120 semantic processing neuroimaging studies, Binder et al. (2009) identify 7 left-lateralized regions: the AG, the lateral temporal lobe, including the entire middle temporal gyrus (MTG) and the posterior inferior temporal gyrus, a ventromedial region of the temporal lobe centered around the mid-fusiform gyrus and adjacent parahippocampus, dorsomedial prefrontal cortex (DMPFC), the inferior frontal gyrus (IFG), particularly the pars orbitalis (POrb), ventromedial and orbital prefrontal cortex (VMPFC), and the posterior cingulate gyrus (PCC) and adjacent precuneus. They further categorize the AG, MTG, and fusiform gyrus into a posterior heteromodal association group, the DMPFC, VMPFC, and IFG into a heteromodal frontal group, and the parahippocampus and PCC into a medial paralimbic group with connection to the hippocampal formation.

While the studies reviewed by Binder et al. (2009) use some semantic contrast which compares, for example, high versus low meaningfulness or hypothetically distinct types of conceptual knowledge, there are other methods for investigating the semantic system. Huth et al. (2016), for instance, use voxel-wise modeling. They collect fMRI data while subjects listen to stories and construct a 958 dimension co-occurrence embedding for each word in the the stimulus with the logic that words with similar semantic values will occur in similar contexts and thus have high co-occurrence values. To estimate how the 958 semantic features influence BOLD response, they fit a cross-validated linear regression model with the embeddings (along with word rate and phonemic coregressors) to each voxel in each subject and find the best prediction performance in the medial, superior and inferior prefrontal cortex, lateral and ventral temporal cortex, and lateral and medial parietal cortex. Applying principal component analysis to the estimated models aggregated across subjects, they they retain 4 dimensions which explain a significant amount of variance. Projecting the word embeddings onto the 4 dimensions and applying k-means clustering, they identify 12 interpretable semantic categories: *tactile*, *visual*, *numeric*, *locational*, *abstract*, *temporal*, *professional*, *violent*, *communal*, *mental*, *emotional*, and *social*. In creating an atlas of semantic selectivity based upon their results, Huth et al. (2016) identify tiles responding to *social*, *numeric*, *visual*, or *tactile* concepts in lateral and medial parietal cortex and tiles responding to *social* concepts in medial superior frontal cortex. Other regions were less clear.

Binder et al. (2009) and Huth et al. (2016) demonstrate that the semantic network is both distributed and diverse. The regions which we find for our semantic number contrast are not out of place when compared to their results. However, their focus is on conceptual knowledge storage. In the case of Binder et al. (2009), the reviewed studies typically only identify one region or one of the region groups, depending on the semantic contrast. In the case of Huth et al. (2016), they find that while a concept like *self* may be selected for by portions of more than one region (e.g., PFC, AG, and MTG), those portions are small and interspersed among portions which select for other concepts in that region. In comparison, we identify a multitude of regions with our single contrast. In the case of the French results, all of the regions from Binder et al. (2009). Additionally, our stimuli are not word-pair or concept-pair contrasts, but words embedded within naturalistic sentences where there is opportunity for continuous meaning composition. Our interpretation of this is that we are observing an effect of semantic number during ongoing semantic composition.

A challenge for identifying the neural correlates of semantic composition is disentangling it from syntactic construction. One approach is to investigate sentences with richer semantics than syntax such as *coercion sentences*, which must undergo an additional compositional operation. An example of complement coercion would be: *The girl finished the book*, which is understood as though the girl had finished *reading* the book. While the verb *finished* should take some kind of event or activity complement, via semantic coercion, *the book*, which otherwise would be an incompatible entity, takes on an event meaning and semantic composition can succeed. Using magnetoencephalography (MEG), increased activity in the Anterior Midline Field in VMPFC has been found for complement coercion (Pylkkänen et al., 2009; Pylkkänen & McElree, 2007) and aspectual coercion (Brennan & Pylkkänen, 2008). In contrast, fMRI results from Husband et al. (2011) identify increased activation in left BA 45, the pars triangularis (PTri) of the IFG for complement coercion sentences.

Another approach is to try and identify minimal compositional mechanisms. In an MEG study of adjective-noun pairs (*red boat*), Bemis and Pylkkänen (2011) find, for written stimuli, increased activity in the left anterior temporal lobe (ATL) and VMPFC. Later, for similar, but spoken and written stimuli, Bemis and Pylkkänen (2013) identify the left ATL and left AG for the intersection of the two modalities. VMPFC does not reaching significance. In the fMRI domain, when Graessner et al. (2020) have participants listen to spoken meaningful (*fresh apple*) and pseudoword (*fresh gufel*) word pairs and provide a judgement for whether the phrase is meaningful or not, for the meaningful stimuli, they find an increase in activation in the left IFG, left DMPFC, bilateral AG, left pMTG, left ATL, left PCC, left precuneus, left hippocampus, bilateral ACC, bilateral pITG, bilateral insula, and right fusiform gyrus. Graves et al. (2010) investigate visually presented, highly meaningful noun-noun phrases (*lake house*) and their minimally meaningful reverses (*house lake*) in two experiments: a 1-back matching experiment for implicit processing and a meaningful/nonmeaningful judgement experiment for explicit processing. In the explicit task, for meaningful over reversed phrases, they find an increase in activation in the right AG, bilateral DMPFC, and bilateral PCC and precuneus.

In identifying an increase in activation in the left IFG (POrb, PTri), the left MTG bordering on the ATL, left DMPFC, left PCC, left parahippocampal gyrus, and left AG for plural over singular nouns (at least for the French results, only POrb, left parahippocampal gyrus, and left AG in the Chinese and French result intersection), our results seem to align with previous fMRI results for semantic composition (Graessner et al., 2020; Graves et al., 2010; Husband et al., 2011) and we believe that we identify an effect for semantic number in compositional semantic processing. That is, it is more difficult to integrate plural nouns into the current, working semantic representation than singular nouns. With semantic composition as the process by which individual concepts are combined to form complex meaning, it is understandable that whether there are *one* or *many* of someone or something would play a role in constructing meaning during language comprehension.

§ 5.3 Similarities and Differences Between the Results

The similarities that we see between the Chinese and French results are not unprecedented. In bilinguals, previous research has found overlap between the L1 and L2 regions which subserve lexicosemantic comparison (Crinion et al., 2006; Klein et al., 2006). Honey et al. (2012) expand this to narrative level stimuli, analyzing neural activity for monolingual English speakers and bilingual, Russian native, English L2 speakers in two conditions: listening to a Russian story and listening to an English translation of that story. When the participants listen to the story in their native language, they find a number of areas in common which reliably respond to the content of the narrative: the superior temporal sulcus, the AG, the supramarginal gyrus, the IFG, the precuneus, the middle frontal gyrus, and orbitofrontal cortex. These results, like ours, show that neural response patterns can be shared across groups despite differences in linguistic form.

With regard to the the differences between the Chinese and French results, although we observe no statistically significant differences in the increase in activation for plural over singular nouns in the French results than in the Chinese results, the Chinese results implicate only a subset of the French results: the left angular gyrus, the left parahippocampal gyrus, and the left pars orbitalis. One possible explanation is the salience of overt number marking. While the semantic number of our non-number marked Chinese observations is conveyed to the listener through discourse cues, for the French observations, grammatical number is always overtly marked on the determiner which occurs before the noun. Another potential factor is the differences between the two datasets. They were collected by different researchers in different facilities and the Chinese dataset has a higher resolution than the French dataset which leads to a more aggressive FWER correction.

§ 5.4 Linguistic Typology and the Neurobiology of Language

The neurobiology of language research domain (Bornkessel-Schlesewsky & Schlesewsky, 2016; Kemmerer, 2016; Poeppel et al., 2012) is interested in explaining how language is implemented in the human brain. One aspect that must be accounted for in any comprehensive model is the similarities and differences between languages described by research in linguistic typology. We advance this goal by investigating Chierchia's (1998) typological counting distinction in which languages differ with respect to whether their nouns can be predicates, arguments, or both. In French, nouns are predicates and in Chinese, nouns are arguments. To target this difference, we use the morphosyntactic properties which emerge. Our French nouns are marked for grammatical number by one of the common definite articles and our Chinese nouns have no overt number marking.

CHAPTER 6

CONCLUSION

In this project, we investigate the neural correlates of semantic number from a crosslinguistic perspective. We use parallel Chinese and French datasets, selecting these two languages because of the differences in their number semantics (Chierchia, 1998). While French nouns are explicitly marked for number and definiteness by the determiners which must occur before them, Chinese lacks nominal pluralization and bare nouns have a number interpretation which is general and includes the plural. Additionally, Chinese has no determiners and bare nouns have a definite reading. Despite the differences in their number semantics and morphosyntactic properties, we observe overlap in regions where neural activity is greater for plural than singular observations: the left pars orbitalis, the left angular gyrus, and the left parahippocampal gyrus, with additional regions occurring in the French that do not occur in the Chinese: left middle temporal gyrus, the left pars triangularis, left posterior cingulate cortex, left dorsomedial prefrontal cortex, and the left fusiform gyrus. These regions have previously been implicated in a distributed network for semantic processing (Binder et al., 2009).

In the context of targeting conceptual knowledge storage, these regions have previously been implicated in a distributed network for semantic processing (Binder et al., 2009), but usually only one region or one region group at a time, depending on the specific stimulus. In addition, because our stimuli are not word-pair or concept-pair contrasts, but words embedded within naturalistic sentences where there is opportunity for continuous meaning composition, and because our results are consistent with previous fMRI results for semantic composition (Graessner et al., 2020; Graves et al., 2010; Husband et al., 2011), we posit a cross-linguistic role for semantic number in semantic composition, the process by which individual concepts are combined to form complex meaning.

We base our analysis on Chierchia's (1998) typological counting distinction in which Chinese nouns are arguments and French nouns are predicates. This sort of typological research pushes forward the neurobiology of language research field which seeks a neurobiological model of language which is able to account for the similarities and differences between languages. One potential drawback to our study is that we approach the distinction from the morphosyntacic properties which come out of it: French nouns are overtly marked for number while Chinese nouns are not. The alternative would be do design a study which directly targets the argument/predicate linguistic semantic distinction with respect to nominals. Additionally, while we have investigated 2 out of 3 of the typological categories proposed by Chierchia (1998): [+predicate, -argument] (French) and [-predicate, +argument] (Chinese), there remains a third proposed category: [+predicate, +argument]. English is a member of this category and based upon the similarities of the Chinese an French results which we find here, we would expect to find similar results in English.

BIBLIOGRAPHY

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, *8*, 14.

Barber, H., & Carreiras, M. (2005). Grammatical gender and number agreement in spanish: An erp comparison. *Journal of cognitive neuroscience*, *17*(1), 137–153.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience*, *31*(8), 2801–2814.

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.

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.

Biswal, B., Deyoe, E. A., & Hyde, J. S. (1996). Reduction of physiological fluctuations in fmri using digital filters. *Magnetic resonance in medicine*, *35*(1), 107–113.

Blank, I. A., & Fedorenko, E. (2017). Domain-general brain regions do not track linguistic input as closely as language-selective regions. *Journal of Neuroscience*, *37*(41), 9999–10011.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2016). The importance of linguistic typology for the neurobiology of language. *Linguistic Typology*, *20*(3), 615–621.

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human v1. *Journal of Neuroscience*, *16*(13), 4207–4221.

Brennan, J., & Pylkkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and language*, *106*(2), 132–143.

Buxton, R. B. (2012). Dynamic models of bold contrast. *Neuroimage*, *62*(2), 953–961.

Carlson, G. N. (1977). Reference to kinds in english.

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.

Castelli, F., Glaser, D. E., & Butterworth, B. (2006). Discrete and analogue quantity processing in the parietal lobe: A functional mri study. *Proceedings of the National Academy of Sciences*, *103*(12), 4693–4698.

Chen, S. F. (1993). Aligning sentences in bilingual corpora using lexical information, In *31st annual meeting of the association for computational linguistics*.

Chierchia, G. (1998). Reference to kinds across language. *Natural language semantics*, *6*(4), 339–405.

Chochon, F., Cohen, L., Moortele, P. v. d., & Dehaene, S. (1999). Differential contributions of the left and right inferior parietal lobules to number processing. *Journal of cognitive neuroscience*, *11*(6), 617–630.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, *20*(1), 37–46.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature neuroscience*, *3*(3), 292–297.

Cox, R. W. (1996). Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, *29*(3), 162–173.

Crinion, J., Turner, R., Grogan, A., Hanakawa, T., Noppeney, U., Devlin, J. T., Aso, T., Urayama, S., Fukuyama, H., Stockton, K., Et al. (2006). Language control in the bilingual brain. *Science*, *312*(5779), 1537–1540.

de Saint-Exupéry, A. (1946). *Le petit prince [the little prince]*. Paris: Gallimard.

de Swart, H., & Farkas, D. (2010). The semantics and pragmatics of plurals. *Semantics and pragmatics*, *3*, 6–1.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*(1-2), 1–42.

Dehaene, S. (2001). Précis of the number sense. *Mind & language*, *16*(1), 16–36.

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the national Academy of Sciences*, *95*(24), 14529–14534.

Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G., van de Moortele, P.-F., & Le Bihan, D. (1998). Imaging unconscious semantic priming. *Nature*, *395*(6702), 597–600.

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

Duncan, J. (2010). The multiple-demand (md) system of the primate brain: Mental programs for intelligent behaviour. *Trends in cognitive sciences*, *14*(4), 172–179.

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences*, *108*(39), 16428–16433.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2013). Broad domain generality in focal regions of frontal and parietal cortex. *Proceedings of the National Academy of Sciences*, *110*(41), 16616–16621.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and language*, *50*(3), 259–281.

Friederici, A. D., Rüschemeyer, S.-A., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral cortex*, *13*(2), 170–177.

Friston, K. J., Jezzard, P., & Turner, R. (1994). Analysis of functional mri time-series. *Human brain mapping*, *1*(2), 153–171.

Gale, W. A., & Church, K. (1993). A program for aligning sentences in bilingual corpora. *Computational linguistics*, *19*(1), 75–102.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fmri study. *Brain and language*, *84*(3), 318–336.

Glover, G. H. (1999). Deconvolution of impulse response in event-related bold fmri1. *Neuroimage*, *9*(4), 416–429.

Göbel, S., Walsh, V., & Rushworth, M. F. (2001). The mental number line and the human angular gyrus. *Neuroimage*, *14*(6), 1278–1289.

Graessner, A., Zaccarella, E., & Hartwigsen, G. (2020). Differential contributions of left-hemispheric language regions to basic semantic composition. *bioRxiv*, 2019–12.

Graves, W. W., Binder, J. R., Desai, R. H., Conant, L. L., & Seidenberg, M. S. (2010). Neural correlates of implicit and explicit combinatorial semantic processing. *Neuroimage*, *53*(2), 638–646.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and cognitive processes*, *8*(4), 439–483.

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: Neural responses shared across languages. *Journal of Neuroscience*, *32*(44), 15277–15283.

Hung, Y.-H., Pallier, C., Dehaene, S., Lin, Y.-C., Chang, A., Tzeng, O. J.-L., & Wu, D. H. (2015). Neural correlates of merging number words. *NeuroImage*, *122*, 33–43.

Husband, E. M., Kelly, L. A., & Zhu, D. C. (2011). Using complement coercion to understand the neural basis of semantic composition: Evidence from an fmri study. *Journal of cognitive neuroscience*, *23*(11), 3254–3266.

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458.

Jonsson, T., Wennerberg, A. B., Forssberg, H., Glover, G. H., & Li, T.-Q. (1999). An image registration strategy for multi-echo fmri. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, *10*(2), 154–158.

Kadosh, R. C., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behav Brain Sci*, *32*(3-4), 313–328.

Kemmerer, D. (2016). Do language-specific word meanings shape sensory and motor brain systems? the relevance of semantic typology to cognitive neuroscience. *Linguistic Typology*, *20*(3), 623–634.

Kiefer, M., & Dehaene, S. (1997). The time course of parietal activation in single-digit multiplication: Evidence from event-related potentials. *Mathematical cognition*, *3*(1), 1–30.

Klein, D., Zatorre, R. J., Chen, J.-K., Milner, B., Crane, J., Belin, P., & Bouffard, M. (2006). Bilingual brain organization: A functional magnetic resonance adaptation study. *Neuroimage*, *31*(1), 366–375.

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, *103*(10), 3863–3868.

Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M., & Bandettini, P. A. (2012). Differentiating bold and non-bold signals in fmri time series using multi-echo epi. *Neuroimage*, *60*(3), 1759–1770.

Kutas, M., & Hillyard, S. A. (1983). Event-related brain potentials to grammatical errors and semantic anomalies. *Memory & cognition*, *11*(5), 539–550.

Lacadie, C., Fulbright, R., Arora, J., Constable, R., & Papademetris, X. (2008). Brodmann areas defined in mni space using a new tracing tool in bioimage suite, In *Proceedings of the 14th annual meeting of the organization for human brain mapping*.

Lee, K.-M. (2000). Cortical areas differentially involved in multiplication and subtraction: A functional magnetic resonance imaging study and correlation with a case of selective acalculia. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, *48*(4), 657–661.

Li, P., Jin, Z., & Tan, L. H. (2004). Neural representations of nouns and verbs in chinese: An fmri study. *Neuroimage*, *21*(4), 1533–1541.

Menon, V., Rivera, S., White, C., Glover, G., & Reiss, A. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage*, *12*(4), 357–365.

Miceli, G., Turriziani, P., Caltagirone, C., Capasso, R., Tomaiuolo, F., & Caramazza, A. (2002). The neural correlates of grammatical gender: An fmri investigation. *Journal of cognitive neuroscience*, *14*(4), 618–628.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520.

Obleser, J., Meyer, L., & Friederici, A. D. (2011). Dynamic assignment of neural resources in auditory comprehension of complex sentences. *Neuroimage*, *56*(4), 2310–2320.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and language*, *34*(6), 739–773.

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), 2522–2527.

Parkman, J. M. (1971). Temporal aspects of digit and letter inequality judgments. *Journal of experimental psychology*, *91*(2), 191.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M., Et al. (2000). A cultural effect on brain function. *Nature neuroscience*, *3*(1), 91–96.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Pesenti, M., Thioux, M., Seron, X., & Volder, A. D. (2000). Neuroanatomical substrates of arabic number processing, numerical comparison, and simple addition: A pet study. *Journal of cognitive neuroscience*, *12*(3), 461–479.

Poeppel, D., Emmorey, K., Hickok, G., & Pylkkänen, L. (2012). Towards a new neurobiology of language. *Journal of Neuroscience*, *32*(41), 14125–14131.

Poldrack, R. A., Mumford, J. A., & Nichols, T. E. (2011). *Handbook of functional mri data analysis*. Cambridge University Press.

Pylkkänen, L., Martin, A. E., McElree, B., & Smart, A. (2009). The anterior midline field: Coercion or decision making? *Brain and language*, *108*(3), 184–190.

Pylkkänen, L., & McElree, B. (2007). An meg study of silent meaning. *Journal of cognitive neuroscience*, *19*(11), 1905–1921.

Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2p1), 274.

Rothstein, S. (2017). *Semantics for counting and measuring*. Cambridge University Press.

Rullmann, H., & You, A. (2006). General number and the semantics and pragmatics of indefinite bare nouns in mandarin chinese. *Where semantics meets pragmatics*, 175–196.

Simon, O., Mangin, J.-F., Cohen, L., Le Bihan, D., & Dehaene, S. (2002). Topographical layout of hand, eye, calculation, and language-related areas in the human parietal lobe. *Neuron*, *33*(3), 475–487.

Thioux, M., Pesenti, M., Costes, N., De Volder, A., & Seron, X. (2005). Task-independent semantic activation for numbers and animals. *Cognitive Brain Research*, *24*(2), 284–290.

Varga, D., Halácsy, P., Kornai, A., Nagy, V., Németh, L., & Trón, V. (2007). Parallel corpora for medium density languages. *Amsterdam Studies In The Theory And History Of Linguistic Science Series 4*, *292*, 247.

Yang, R. (2001). *Common nouns, classifiers, and quantification in chinese* (Doctoral dissertation). Rutgers University Doctoral dissertation.