<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>An fMRI study of grammatical morpheme processing associated with nouns and verbs in Chinese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Yu, X; Bi, Y; Han, Z; Law, SP</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>PLoS One, 2013, v. 8 n. 10, article no. e74952</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2013</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/200906">http://hdl.handle.net/10722/200906</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Introduction

Languages vary widely in the complexity of their morphosyntactic system. For instance, on one end of the spectrum, the Chinese language is well-known for its impoverished inflectional morphology [1, 2]. The morphological and phonological structures of Chinese words stay the same during sentence construction. There is only aspectual marking for verbs, and no inflectional marker for nouns in the traditional sense. On the other end of the continuum, there are languages with rich inflectional morphology such as Italian, Polish, Hungarian, where the verb may change its form on the basis of tense, aspect, person, finiteness, negation, or modality, and a noun may be marked for gender, number, or case. Moreover, nouns and verbs in such languages may have different declensional and conjugational patterns, respectively, depending on their classification. Understanding whether and how representation of morphosyntactic processing at the brain level may differ as a function of complexity will significantly inform us about neurobiology of language [3], and possibly contribute to cross-linguistic observations of neural substrates underlying processing of grammatical morphemes from an analytic and a classifier language, and thereby deepen our understanding of neurobiology of human language.

Neural Substrates of Morphosyntactic Processing in European Languages

Early evidence for neural representation of a morphosyntactic component in the language system comes from behavioral observations of individuals with brain injury. A case study of an Italian-speaking individual with aphasia described a pattern of selective impairment to inflectional morphemes in spontaneous production of sentences and repetition of single words, including nouns, verbs, and adjectives, and relatively preserved production of derivational morphemes [6]. Specific disruption to inflectional morphology was similarly reported of an English aphasic speaker in a reading aloud task [7]. Subsequent case studies show that impaired production of inflectional morphology may occur to specific grammatical class, for instance, to verbs in English [8] and Greek [9] or nouns [10]. That homonyms (to watch, a watch) and pseudowords were used as stimuli for the noun and verb conditions in [8, 10] reduced the possibility that the dissociation patterns were confounded with psycholinguistic factors, and thereby demonstrated that inflectional morphology is specified for grammatical word class in the brain. These findings also suggest distinctive neural correlates of verbal and nominal inflectional morphology.

In the past decade, a number of functional imaging studies employing tasks that explicitly involved operations of inflectional morphology associated with nouns and verbs were conducted to identify brain areas of noun-specific and verb-specific morphosyntax (see [3] for a review of studies showing a left-lateralized fronto-temporal network supporting the processing of inflected spoken words). In most of these investigations, the participants were asked to provide singular-plural alternations for nouns and
present-past tense alternations or person agreement for verbs in phrasal or sentential contexts. In a series of studies [11,12,13], real words including abstract and concrete nouns and verbs with regular and irregular inflections, as well as pseudowords served as stimuli. Pseudowords were used in the attempt to eliminate semantic confounds between nouns and verbs. Greater activation was found in the left middle frontal gyrus (LMidFG) and bilateral superior posterior parietal regions for verb production and the left middle fusiform gyrus for noun production across conditions of lexicality, concreteness, and regularity of inflection in English [11]. Different cortical regions for production of nominal and verbal inflections were also reported in German, albeit in somewhat different areas – the left superior frontal gyrus extending anteriorly for verbs and the right superior temporal gyrus and the left fusiform gyrus for nouns – in a PET study [12]. The role of LMidFG, particularly its anterior portion (LaMidFG), in the processing of verbal inflection in English was highlighted in [13] using transcranial magnetic stimulation (TMS). Application of repetitive TMS (rTMS) to LaMidFG significantly slowed down morphological operation of verbs but not nouns. This observation was replicated in [14]. Moreover, the study found that production of both regularly and irregularly inflected verbs was inhibited, and the suppression resulting from application of rTMS was restricted to LaMidFG since no word class specific interference was noted in the neighboring Broca’s area or the posterior MidFG, or the right homologue of aMidFG. More recently, a linear decrease in activation was reported in the junction of the left inferior frontal gyrus (LIFG) and LMidFG to English inflected verb production over the course of a morphological transformation task [15]. The functional magnetic resonance (fMR) adaptation of the area was argued to be associated with repeated processing of verbal inflections. The LMidFG (Brodmann area (BA)9 and BA10) and LIFG (BA44 and BA45), in addition to the left inferior precentral gyrus (BA6), have similarly been reported to selectively activate for inflected verbs in Italian compared with repetition of verbs as a baseline [16]. In short, there is evidence suggesting that LMidFG underlies morphosyntactic processing of verbs.

However, the claim that an area in the left frontal region can be identified for processing of inflectional morphology specific to verbs has been challenged. More specifically, the left BA44/45 and BA47 were found to be more strongly activated for inflected nouns than verbs in English [17]. The researchers attributed the observation to greater processing difficulty of the noun stimuli, due to lower frequency and irregular inflections, compared to the verb stimuli. It has also been explicitly argued that the activity of the left frontal region is modulated by processing demands instead of reflecting operations of any particular grammatical class [18]. Differing from the studies reviewed thus far, Italian-speaking participants in [18] named pictures depicting events using infinitive verbs, inflected verbs, and action nouns. Since action nouns are not the preferred responses to pictured actions, learned relatedness judgment task and a semantic associate production task recently been reported. Converging evidence from a semantic relatedness judgment task and a semantic associate production task revealed a task-independent region in the left posterior superior and middle temporal cortices (LpSTG&MTG) that activated more strongly for verbs than nouns [25,26]. Therefore, the proposition that impoverished inflectional morphology would lead to a lack of word class distinction in a language needs to be reconsidered.

While Chinese lacks inflectional morphology, there are grammatical morphemes that take part in the syntax of the language typical of an analytic language, including those relevant to nouns and verbs. They appear before or after the content word without changing its form. There are five aspect markers in Mandarin Chinese: the perfective le5, experiential guo4, and continuous zhe3, which may be attached to the end of a verb, the progressive zai4, which occurs before a verb, and the delimitative yi1, which appears between a verb and its reduplicated form, i.e. V-yi1-V.

As for nouns, there is a class of morphemes called classifiers that must appear when a noun is preceded by a numeral and/or a demonstrative, such as yi1 ben3 shu4 ‘one + classifier + book’. In other words, Chinese is also a classifier language. Classifier languages are spoken by a large portion of the world’s population, including speakers of East and Southeast Asian languages, some Australian aboriginal languages, and some native American languages [28]. Two major types of classifiers can be distinguished in Chinese, sortal (or count-classifiers) and mensural (or mass-classifiers). The former are closed-class morphemes and often related to the noun, especially when it denotes an object, in terms of shape, animacy, function or social status [28,29,30], while the latter are open-class morphemes that quantify the amount of an object or objects (e.g. a group of students, a glass of water, a month of work). Estimates of the number of classifiers vary widely across sources, ranging from two dozen to several hundreds depending on whether mensural classifiers are also included. In Mandarin, it has been estimated that there are over 60 classifiers [1], but only...
about two dozen are “core classifiers” for most classifier use. This is consistent with the description in [28], which listed 126 “classifiers” but indicated that only 19 functioned solely as classifiers.

Given the relationship between nouns and sortal classifiers, it is not surprising that sortal classifiers have been studied extensively in child language development (e.g. [31,32]) and on the relationship between language and cognition (e.g. [28,33,34]). Nonetheless, it should be emphasized that there is “a non-trivial degree of arbitrariness” (p. 1127 in [28], and [34]) in the choice of a sortal classifier from the meaning of the noun (e.g. the classifier pu3 for both ‘horse’ and ‘bolts of cloth’). Moreover, not only concrete nouns or objects (which have been the focus of most previous work on classifiers) but also abstract nouns require a classifier (e.g. ‘news’, ‘hope’, ‘resentment’), and it is hard to discern any physical or functional relationship between the abstract noun and its classifier. It is also worth mentioning that in casual speech, most Mandarin Chinese speakers would use the general classifier 者 and in place of the proper classifier. Given all these characteristics, it has been proposed that classifiers can be seen as the Chinese counterparts of noun inflection [27,35].

The existence of sortal classifiers and aspect markers associated with Chinese nouns and verbs constitutes a very special case from the perspective of cross-linguistic study. The relative simplicity of the Chinese system raises the question of whether the degree of complexity or richness of a grammatical component would affect its representation in the brain, analogous to previous reports of null findings for representation of lexical (or derivational) morphology in English (e.g. [36,37] but see [38]), but positive findings in Hebrew (e.g. [39]), German (e.g. [40]), and Italian (e.g. [16,41]). In addition, contrary to most European languages in which the verbal paradigm is more complex than the nominal counterpart, the contrast between the nominal classifier and verbal aspect marker inventories in Chinese presents the opposite pattern. This difference renders Chinese a highly interesting testing ground for assessing the view that neural correlates of morphosyntactic processes, particularly in LIFG, specific to a grammatical class is driven by computational demands. If correct, one would expect to find areas in the left prefrontal cortex more strongly activated for nominal than verbal morphological operations in Chinese, and none for the reverse comparison. On the other hand, if neural representation of grammatical morphemes does not simply reflect processing demand but in fact is form class specific, it is possible to find separate neural correlates for classifiers and aspect markers.

We carried out two experiments, a production task (sentence completion) in which the participants supplied either a sortal classifier or an aspect marker to complete a sentence, and a grammaticality judgment task. In the latter, grammatical violation arose from inappropriate pairing between a noun and a classifier, or from the incongruity between an aspect marker and the lexical aspect (or semantic structure) of a verb. For instance, an atelic verb (i.e. a verb without an inherent end point, such as stative or psych verbs) followed by a perfective marker, e.g. *她珍视了它 she-treasure-PERF-it, or a telic verb coupled with a continuous marker, e.g. *它倒塌着 it-collapse-DUR, would result in ungrammaticality. Note that while this task could also be considered semantic judgment task as in [42,43,44], we prefer the term “grammaticality” since judgments in our experiment were not solely semantically based. Different from the exclusive use of object nouns in [42,44], our stimuli included both abstract and concrete nouns in the classifier condition. Hence, congruity judgment between an abstract noun and a classifier and that between the semantic structure of a verb and the grammatical meaning of an aspect marker were not driven only by semantic features in the typical sense.

Similar to [25,26], conjunction analyses were conducted in the production experiment to identify brain areas that were more activated for classifiers than aspect markers as well as those that were more active for verbal than nominal grammatical morphemes across concreteness conditions. The use of both concrete and abstract items and conjunction analyses across concreteness levels is an important aspect of the current design. It allowed us to identify regions that cannot be said to be mainly responsive to semantic features such as shape, function, and animacy in the case of classifiers. The classifier-specific and aspect marker-specific regions then served as regions-of-interest (ROI) to detect differential activation in grammaticality judgment to the classifier vs. aspect marker conditions.

**Methods**

**Ethics Statement**

Informed written consent was obtained from all participants before the study began. The experiments were performed in accordance with the Declaration of Helsinki with ethical approval from the Institutional Review Board of the State Key Laboratory of Cognitive Neuroscience and Learning in Beijing Normal University as well as the University of Hong Kong Human Research Ethics Committee for Non-Clinical Faculties.

**Participants**

Sixty-six native Mandarin speakers were recruited from Beijing Normal University to participate in the current study. Among them, 19 participants (11 female, Age mean = 23.9, SD = 4.40) took part in a pilot experiment. The remaining 47 participants carried out one of the imaging experiments, with 27 (16 females, Mean age = 20.8, SD = 2.14) in the grammaticality judgment experiment, and 20 (10 females, Mean age = 21.3, SD = 3.00) in the sentence completion experiment. All subjects who took part in the fMRI experiments were further required to be right-handed (assessed by the Edinburgh inventory, [45]), have normal or corrected to normal visual acuity, as well as no history of psychiatric or neurological disorders.

**Grammaticality Judgment Experiment**

**Materials and stimuli.** Two aspect markers (ASPs) — 着 (zhe5, continuous ASP) and 了 (le5, perfective ASP) were selected for verbs of both concrete and abstract concepts, while four sortal classifiers (CLs) were chosen with two — 个 (ge4) and 具 (ju1) for concrete nouns and two — 间 (jian4) and 条 (tiao2) for abstract items. Sixty unambiguous nouns and 60 unambiguous verbs that satisfied the following criteria were selected: 1) for each word, the frequency as the target grammatical class is at least 10 times larger than that of the second most-frequently used word class; 2) for each word class, half of the items were concrete, and half were abstract; 3) nouns and verbs were balanced in frequency; 4) half of the nouns or verbs at each concreteness level were congruent with the nouns or verbs at the same concreteness level. Note that although imageability was not balanced in the concrete level...
Table 1. Properties of materials in grammaticality judgment experiment with error rate (%) and response latency (ms) from pilot study.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Log frequency (SD)</th>
<th>Imagability (SD)</th>
<th>Example of grammatical sentence</th>
<th>Example of ungrammatical sentence</th>
<th>Error (%) (SD)</th>
<th>RT (ms) (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCL</td>
<td>1.14 (0.54)</td>
<td>6.6 (0.2)</td>
<td>&quot;This is six CL receipts&quot;</td>
<td>&quot;Those are six CL receipts&quot;</td>
<td>3.4 (8.2)</td>
<td>1272 (268)</td>
</tr>
<tr>
<td>AASP</td>
<td>1.20 (0.52)</td>
<td>3.5 (0.3)</td>
<td>&quot;I gasp he sings&quot;</td>
<td>&quot;He has invented ASP it&quot;</td>
<td>4.2 (8.5)</td>
<td>1366 (173)</td>
</tr>
<tr>
<td>CL</td>
<td>1.54 (0.50)</td>
<td>3.5 (0.3)</td>
<td>&quot;Those are two CL projects&quot;</td>
<td>&quot;Those are six CL receptos&quot;</td>
<td>6.0 (8.5)</td>
<td>1349 (340)</td>
</tr>
<tr>
<td>ASP</td>
<td>1.54 (0.50)</td>
<td>3.5 (0.3)</td>
<td>&quot;He has invented ASP it&quot;</td>
<td>&quot;He has invented ASP it&quot;</td>
<td>6.0 (8.5)</td>
<td>1425 (1182)</td>
</tr>
<tr>
<td>AASP</td>
<td>1.54 (0.50)</td>
<td>3.5 (0.3)</td>
<td>&quot;He has invented ASP it&quot;</td>
<td>&quot;He has invented ASP it&quot;</td>
<td>6.0 (8.5)</td>
<td>1425 (1182)</td>
</tr>
</tbody>
</table>

Note. CCL = concrete classifier; CAP = concrete aspect marker; ACL = abstract classifier; AASP = abstract aspect marker; RT = response latency.

A pilot study was conducted with 19 participants, in order to evaluate the acceptability of the stimuli as well as the processing demand for each condition as reflected in response latency (RT). All of the grammatical sentences were accepted by more than half (i.e. 10) of the participants (see error rates in Table 1); however, some of the ungrammatical sentences were not rejected by more than 10 subjects. Therefore, only 20 out of the 30 ungrammatical sentences were selected from each condition. For the abstract CL condition, 15 ungrammatical sentences took tiao2 and five had xiang4 as the incongruent CL, whereas for the other three conditions, the selected CLs or ASPs were equally represented. Based on the pilot results, both the grammatical and ungrammatical sentences could be balanced on processing demand in terms of error rates and RTs between CL and ASP conditions of both concreteness levels, except that ungrammatical CL sentences with concrete nouns showed a lower error rate than concrete ASP sentences (Table 1). One point worthy of mention is that, regardless of concreteness level or sentence structure, CL sentences with six characters were significantly longer than ASP sentences, which include four to six characters. Possible effects of sentence length were addressed in data analysis.

Four additional words (two nouns and two verbs), other than experimental materials, were selected and used to construct two grammatical and two ungrammatical sentences, functioning as lead-in trials during scanning.

Design and procedure. An event-related design was adopted. Sentences from each condition were combined and divided into four blocks of 50 items, with grammatical and ungrammatical sentences associated with the same CL or ASP condition distributed equally across blocks. Care was further taken to ensure that grammatical and ungrammatical trials containing the same nouns or verbs were not assigned to the same block. Items in each block were arranged according to optimal event type scheduling computed by Optseq software (http://surfer.nmr.mgh.harvard.edu/optseq/), forming one experimental run with one lead-in trial added at the beginning. The order of the four experimental runs was counterbalanced across participants.

The experiment was conducted in E-prime 1.2. During each run, a blank screen would first appear and last for 10 s, allowing the participants to adapt to the scanning environment. The lead-in
trial then appeared, followed by one experimental block. In each trial, a sentence stimulus would be presented in the center of the screen (visual angle: 5° ~ 7° depending on the sentence length) for 4 s. Participants were instructed to judge whether the sentence was grammatical or not by pressing the corresponding buttons with their left hand, as accurately and quickly as possible. Response accuracies and latencies were recorded. The stimulus was replaced by a blank screen, the duration of which was determined by the Optseq software (min = 2 s, max = 10 s, mean = 4 s), in order to optimize the event scheduling for better partition and estimation of each event type. The next trial began after the jittered ISI (inter-stimulus interval). Throughout the run, a red dot remained in the center of the screen as the fixation point.

Each run lasted about 7 minutes, and there was a self-paced break between runs (around 2 minutes). The entire experiment, including the practice and preparation, took approximately 50 minutes.

**MRI data acquisition, preprocessing and first-level analysis.** Functional MRI scans were collected on a 3.0 Tesla Siemens scanner using a 12-channel transmit/receive gradient head coil (Beijing Normal University, China). A T2*-weighted gradient-echo planar imaging (EPI) sequence was applied to acquire the blood oxygen level-dependent (BOLD) signals (flip angle 90°, TE = 30 ms, TR = 2000 ms, in-plane resolution = 3.125 * 3.125, slice thickness = 4 mm, slice gap = 0.8 mm).

Data preprocessing and analysis were performed using SPM5 (http://www.fil.ion.ucl.ac.uk/spm/software/spm5/). The first 9 TRs containing blank screen and lead-in trials were deleted from each run, before functional images were slice-time and head motion corrected for each run per subject. Subsequently, data were normalized to a standard template in Montreal Neurological Institute (MNI) space and then smoothed with an isotropic 8-mm full-width-half-maximal Gaussian kernel.

For the first-level analysis, due to excessive head movement (>2 mm or 2°) within at least one experimental run, four participants (three females) were excluded from further analysis. The images of the other 23 participants were entered into two models that were set up for different purposes.

A. Conventional model: this model was built to investigate the processing of different grammatical morphemes. Thus, eight event types, as decided by experimental manipulations (2 grammatical morpheme types * 2 concreteness levels * 2 grammaticality levels) were considered and modeled with the canonical hemodynamic response functions for estimation. The high-pass filter was set at 273 s, calculated based on the longest time interval between trials from the same condition. After model estimation, subjectwise contrast maps were computed for each of the following four conditions against fixation, which were grammatical and ungrammatical CL sentences (GCL, UCL), as well as grammatical and ungrammatical ASP sentences (GASP, UASP).

B. Length model: The second model was constructed to evaluate the effect of sentence length in terms of number of characters as it was not balanced across experimental conditions, which might have confounded with effects of contrasts between nominal and verbal grammatical morphemes. ASP sentences with concrete verbs were combined across grammaticality and re-divided based on verb transitivity into sentences with transitive verbs (Verb\textsubscript{transitive} 19 trials) and those with intransitive verbs (Verb\textsubscript{intransitive} 31 trials). Sentences in the former condition contained five or six characters, which were significantly longer than those in the latter condition (four or five characters, Mean length: Verb\textsubscript{transitive} = 5.2, Verb\textsubscript{intransitive} = 4.3; t(48) = 6.9, p<0.001). During model specification, Verb\textsubscript{transitive} and Verb\textsubscript{intransitive} as well as the other six event types as in the conventional model, were fed into the GLM. The high-pass filter was adjusted to 206 s. Subjectwise contrast maps for Verb\textsubscript{transitive} versus fixation, as well as Verb\textsubscript{intransitive} versus fixation were produced for subsequent analysis.

**Chronometric data analysis.** Subjectwise accuracies were first computed by averaging accuracies across all items. One participant with accuracy lower than 80% was excluded from further analysis. Thus, due to excessive head movement and/or poor performance, five participants in total were removed, leaving data from 22 subjects (13 females) for further behavioral and imaging data analyses.

For behavioral analyses, RT data were trimmed if responses were incorrect, absent, or 3 SDs away from the individual mean. Error rates and RTs were then entered into three-way ANOVA tests with item and subject as random factors, respectively, to evaluate the main effects of concreteness, word class and grammaticality, as well as their interactions. Results were considered reliable only if both by-item and by-participants analyses were significant.

**Sentence Completion Experiment**

**Materials and stimuli.** The 120 grammatical sentences in the grammaticality judgment task served as materials in this experiment (see Appendix for the entire list of sentences). The stimuli were created by masking the CL or ASP in each sentence. Four sentence stimuli (two CL and two ASP sentences) with nouns and verbs other than the experimental materials were further created, as lead-in trials in the imaging experiment.

Moreover, 50 words randomly selected from the experimental materials and 39 novel words other than the stimuli in the experiment were employed in a post-scanning memory probe test, in order to evaluate participants’ attentiveness during scanning.

**Design and procedure.** An event-related design was adopted as in the grammaticality judgment task. Items from each condition were mixed and divided into two blocks of 60 trials. In each block, sentences types were balanced across the four conditions, and further matched on the number of trials between sentences with either CL or ASP within each condition. Similar to the judgment experiment, item sequence in each block was computed by Optseq, forming one experimental run with two lead-in trials added at the beginning. To minimize the order effect, two lists composed of two blocks with different stimulus sequences were generated and randomly assigned to the 20 participants, with 10 for each list. The run order was counterbalanced across the 10 subjects for the same list.

E-prime 1.2 was used to run the experiment. At the beginning of each run, a blank screen was first presented for 10 s, followed by lead-in and experimental trials. During each trial, an incomplete sentence would be shown in the center of the screen (visual angle: 5° ~ 7°) for 4 s, during which participants were required to produce one CL (except the general CL δις or ASP appropriate for the stimulus sentence covertly, in order to minimize head movement. The stimulus was then replaced by a blank screen with a jittered ISI (computed by Optseq, min = 2 s, max = 12, mean = 4 s) before the next trial began. Throughout the run, a red dot remained in the center as the fixation point. Each of the two runs lasted for 8.4 minutes, and there was a 2-minute break in between. Immediately after the experiment, the participants were asked to attend a memory probe test. Each subject had to indicate if a stimulus word had been seen in the scanner. The test was self-paced and took approximately three minutes to complete.

The entire experiment, including the practice, preparation, and probe test, took approximately 35 minutes. The participants were required to return the next day to repeat the same experiment.
outside the scanner with overt responses, in order to collect their responses and response latencies.

MRI data acquisition, preprocessing and first-level analysis. The parameter setting for scanning and the procedure of preprocessing were identical to those in the grammaticality judgment experiment, except that the initial 13 TRs containing blank screens and lead-in trials were removed from each run. Due to excessive head movement (> 2 mm or 2° in at least one run), data of three participants (one female) were discarded.

For subject-level analysis, two models with the same purposes as those in the grammaticality judgment experiment were built. In the conventional model, regressors for the four event types (CCL, ACL, CASP, AASP) were included, and contrast maps for each condition versus fixation were computed after estimation. The high-pass filter was set at 191 s. For the length model, trials with concrete verbs were divided into Verbtransitive (18 trials, mean length = 4.3) and Verbtransitive (12 trials, mean length = 5.2), which significantly differed in sentence length, t(26) = 5.5, p < 0.001. Five event types (Verbtransitive, Verbtransitive, ASP, CCL, ACL) were entered into modeling, and contrast maps were computed accordingly. The high-pass filter was adjusted to 206 s for the second model.

Chronometric data analysis. The accuracy rate for the memory probe test was first calculated for each participant. Cut-off score was set at 65.2%, which was significantly above chance level. Two subjects (one of whom also had excessive head movement) with accuracies lower than the criterion were excluded. In the end, data of four participants were discarded due to excessive head movement or poor performance in the memory probe test. Data of the remaining 16 participants (nine females) underwent further statistical analyses.

The appropriateness of responses collected outside the scanner was judged by two raters, who were naive to the design and aims of the current experiment. CLs or ASPs that were rated as ungrammatical by one of the two participants were regarded as errors. Response times were trimmed if a) an erroneous response was given, b) the voice key was triggered by noise, such as cough, or c) the value was 3 SDs away from the subjectwise mean or less than 200 ms. Both the error and RT results were analyzed with two-way ANOVAs with items and subjects as random effects, respectively, to calculate the main effects of grammatical morpheme and concreteness, as well as their interaction. Similar to the grammaticality judgment experiment, results were considered reliable only if both by-item and by-participants analyses were significant.

Imaging Data Analysis Involving Sentence Completion and Grammaticality Judgment Tasks

Group-level analyses of imaging data were conducted across the two experiments to reveal task-independent effects. Regions more strongly activated for the CL or ASP condition were first obtained from the sentence completion experiment by a whole-brain analysis. Contrast maps of each condition versus fixation that were significantly more activated for the CL or ASP condition were first obtained. A three-way ANOVA with concreteness as one of the factors was in fact carried out, and none of the ROIs showed significant interaction between grammatical morpheme and concreteness. Hence, results of the two-way ANOVA are reported. Regions replicating the grammatical morpheme effects in the sentence completion task in terms of main effect or interaction were regarded as task-independent regions for further consideration.

In addition, possible confounding effects of sentence length in both experiments were estimated in CL-specific ROIs, in order to evaluate whether the observed finding was the result of higher visual processing load due to longer sentences in the CL conditions. Based on the length effect model in each experiment, percentage signal change in the conditions of intransitive verbs and transitive verbs were extracted and averaged, respectively, for each CL-specific region. T-tests were applied to compare the activation levels between the two conditions representing different lengths. Regions showing significant length effects were regarded as neural areas sensitive to visual processing demand, which might have confounded with the effects of stronger activation apparently induced by classifiers.

Results

Behavioral Results

For the sentence completion experiment, the two-way ANOVA analysis of error rates did not find any significant main effect or interaction effect. However, for the RT, main effects of both grammatical morpheme and concreteness were significant (grammatical morpheme: F1(1,15) = 11.2, p < 0.01; F1(1,116) = 15.7, p < 0.001; concreteness: F1(1,15) = 30.5, p < 0.001; F1(1,116) = 9.8, p < 0.01), with items of concrete concepts and ASP sentences responded to more quickly (see descriptive results in Table 2). The interaction between grammatical morpheme and concreteness was insignificant.

For the grammaticality judgment experiment, the three-way ANOVA revealed a significant main effect of concreteness on both error rates and RTs (error rates: F2(1,21) = 34.3, p < 0.001; F2(1,192) = 18.3, p < 0.001; RT: F2(1,21) = 47.5, p < 0.001; F2(1,192) = 17.4, p < 0.001) with concrete items easier and quicker to respond to, while main effects of grammatical morpheme and grammaticality were not significant in either error rates or RT (Table 2). The interaction between concreteness and grammaticality for error rates (F2(1,21) = 9.3, p < 0.01; F2(1,192) = 4.6, p < 0.01) was significant, with higher error rates for ungrammatical trials at the concrete level but a reversed tendency for trials with abstract concepts. However, post-hoc analyses did not reveal any significant simple effect (all p > 0.1). The other two-way interactions and the three-way interaction were not statistically reliable.
Imaging Results

Conjunction analyses of CL vs. ASP contrasts between the two concreteness levels in the sentence completion task revealed that the left posterior middle temporal gyrus (adjacent to the superior temporal gyrus) was activated more strongly for the ASP sentences than the CL sentences for both concrete and abstract levels, whereas regions showing greater activation for the CL conditions of both concreteness levels included bilateral calcarine and lingual gyr (area in the left hemisphere extended into posterior fusiform gyrus), bilateral orbital inferior frontal gyri and insula cortex (BA47, right BA47 (rBA47)), as well as the left supplementary motor area and superior medial frontal gyrus (LSMA&SMedFG).

In addition, the dorsal aspect of left triangular and opercular inferior frontal gyri (LIFG, BA44), with a smaller cluster size (k = 56, corresponding to cluster-level p = 0.06) was also activated more strongly for CL trials and therefore included for further consideration (see detailed information on each cluster in Table 3).

ROI analyses of the length effect revealed two CL specific clusters – (i) left calcarine, lingual and posterior fusiform gyri, as well as (ii) right calcarine and lingual gyri – which showed significantly greater activation for sentences with more characters (Verbtransitive) than those with fewer characters (Verbtransitive) in both experiments (see Table 4). Since the confounding effects of sentence length could not be separated from those of grammatical morpheme contrasts in the current study, these two regions would not be considered further for the sake of parsimony.

For the remaining four CL specific regions and one ASP specific area, results of the ROI analyses using two-way ANOVAs (grammatical morpheme x grammaticality) of thegrammaticality judgment task revealed a significant main effect of grammatical morpheme only in left posterior middle temporal gyrus, with larger signal changes induced by the ASP condition. This pattern was consistent with the results in the sentence completion task. The main effect of grammaticality was significant in bilateral BA47, LSMA&SMedFG, as well as left posterior middle temporal gyrus with ungrammatical sentences inducing stronger responses (Table 4).

A significant interaction effect between grammatical morpheme and grammaticality was also observed in BA47, rBA47, and LSMA&SMedFG. Among them, interaction effects in two regions – BA47 and LSMA&SMedFG – were caused by higher activation for CL sentences than ASP sentences in grammatical trials only (Figure 1). This pattern was consistent with the results of the sentence completion task. Post-hoc analyses contrasting grammatical CL with ASP sentences found significant differences in both regions (BA47: t(21) = 2.42, p < 0.05; LSMA&SMedFG: t(21) = 2.50, p < 0.05). For rBA47, the interaction exhibited a pattern of lower activation for ungrammatical CL sentences than ungrammatical ASP sentences with a reversed effect between CL.

Table 2. Behavioral results in error rate (%) and response latency (ms) of grammaticality judgment and sentence completion tasks.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Condition</th>
<th>Mean (%)</th>
<th>SD</th>
<th>Mean (ms)</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammaticity judgment</td>
<td>Grammatical CCL</td>
<td>4.2</td>
<td>6.3</td>
<td>1535</td>
<td>188</td>
</tr>
<tr>
<td></td>
<td>Grammatical CASP</td>
<td>5.3</td>
<td>7.7</td>
<td>1456</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>Grammatical ACL</td>
<td>15.2</td>
<td>12.6</td>
<td>1654</td>
<td>191</td>
</tr>
<tr>
<td></td>
<td>Grammatical AASP</td>
<td>9.4</td>
<td>8.2</td>
<td>1627</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical CCL</td>
<td>4.8</td>
<td>4.5</td>
<td>1546</td>
<td>155</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical CASP</td>
<td>7.5</td>
<td>5.8</td>
<td>1566</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical ACL</td>
<td>9.1</td>
<td>6.3</td>
<td>1664</td>
<td>131</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical AASP</td>
<td>8.2</td>
<td>8.7</td>
<td>1589</td>
<td>142</td>
</tr>
<tr>
<td>Sentence completion</td>
<td>CCL</td>
<td>2.7</td>
<td>6.5</td>
<td>1222</td>
<td>212</td>
</tr>
<tr>
<td></td>
<td>CASP</td>
<td>3.5</td>
<td>5.8</td>
<td>1117</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td>ACL</td>
<td>7.7</td>
<td>14.2</td>
<td>1339</td>
<td>187</td>
</tr>
<tr>
<td></td>
<td>AASP</td>
<td>2.5</td>
<td>5.6</td>
<td>1196</td>
<td>149</td>
</tr>
</tbody>
</table>

Note. Mean and SD were calculated across item-wise values within each condition. CCL = concrete classifier; CASP = concrete aspect marker; ACL = abstract classifier; AASP = abstract aspect marker.

doi:10.1371/journal.pone.0074952.t002

doi:10.1371/journal.pone.0074952.t003

Table 3. Results of whole-brain analysis from sentence completion experiment.

<table>
<thead>
<tr>
<th>Contrasts for conjunction</th>
<th>Activated region</th>
<th>Cluster size</th>
<th>x</th>
<th>y</th>
<th>z</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CCL-CASP) (\cap) (ACL-AASP)</td>
<td>Left calcarine, lingual and posterior fusiform gyri</td>
<td>412***</td>
<td>−15</td>
<td>−93</td>
<td>−9</td>
<td>5.72</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Right calcarine and lingual gyri</td>
<td>186***</td>
<td>9</td>
<td>−87</td>
<td>0</td>
<td>5.99</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Left triangular and opercular inferior frontal gyr (dorsal part, BA44)</td>
<td>56b</td>
<td>−39</td>
<td>12</td>
<td>30</td>
<td>4.92</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Left orbital inferior frontal gyrus and insula (BA47)</td>
<td>225***</td>
<td>−33</td>
<td>30</td>
<td>−15</td>
<td>8.6</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Right orbital inferior frontal gyrus and insula (rBA47)</td>
<td>82*</td>
<td>27</td>
<td>27</td>
<td>−9</td>
<td>5.15</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Left supplementary motor area and superior medial frontal gyrus</td>
<td>213***</td>
<td>−6</td>
<td>27</td>
<td>45</td>
<td>4.43</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>(CASP-CCL) (\cap) (AASP-ACL)</td>
<td>Left posterior middle temporal (adjacent to the superior gyrus)</td>
<td>60*</td>
<td>−57</td>
<td>−48</td>
<td>12</td>
<td>4.53</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Note. For the whole brain analyses, unless specified otherwise, significant threshold was held at \(p_{\text{cor}}<0.001\), \(k>60\), corresponding to corrected cluster-level \(p<0.05\). CCL = concrete classifier; CASP = concrete aspect marker; ACL = abstract classifier; AASP = abstract aspect marker.

*Peak coordinates are reported in the MNI system.

*Due to a relatively smaller cluster size, BA44 only showed a marginally significant effect of grammatical morpheme (\(p_{\text{cor}} = 0.06\)).

*p\(<0.05\), **p\(<0.01\), ***p\(<0.001\).

doi:10.1371/journal.pone.0074952.t000
and ASP trials for the grammatical condition. However, post-hoc analyses did not find significant simple effects between CL and ASP sentences in either grammaticality condition (grammatical trials: $p > 0.1$; ungrammatical trials: $p \geq 0.07$).

In summary, with respect to processing nominal classifiers and verbal aspect markers, convergence analyses have shown task-independent regions for greater response to classifiers in BA47 and LSMA&SMedFG, and to ASP stimuli in the left posterior middle temporal gyrus, as illustrated in Figure 2.

### Discussion

The neural bases underlying processing of Chinese classifiers and aspect markers were investigated through one expressive task—sentence completion, and one receptive task—grammaticality judgment, to look for converging evidence for the processing of Chinese grammatical morphemes associated with nouns and verbs. From the production task, we identified a number of regions that were more strongly activated during classifier selection and those that were more responsive to aspect marker selection, regardless of the concreteness of the relevant noun and verb. Signal changes in these regions from the judgment task were then

---

**Table 4.** Results of two-way ANOVAs from grammaticality judgment and of sentence length.

<table>
<thead>
<tr>
<th>Regions from conjunction analyses of sentence completion</th>
<th>Grammatical morpheme effect</th>
<th>Grammaticality effect (Ungrammatical &gt; Grammatical)</th>
<th>Interaction</th>
<th>Length effect: Judgment task (Long &gt; Short)</th>
<th>Length effect: Completion task (Long &gt; Short)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CL-specific regions</strong> (CL &gt; ASP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left calcarine, lingual and posterior fusiform gyri</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>Right calcarine and lingual gyri</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
<td>***</td>
</tr>
<tr>
<td>BA44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BA47</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
<td>***</td>
</tr>
<tr>
<td>rBA47</td>
<td>***</td>
<td></td>
<td></td>
<td>***</td>
<td></td>
</tr>
<tr>
<td>Left supplementary motor area and superior medial frontal gyrus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ASP-specific regions</strong> (ASP &gt; CL)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left posterior middle temporal</td>
<td>***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. CL = classifier; ASP = aspect marker. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

doi:10.1371/journal.pone.0074952.t004

---

Figure 1. Interaction effects between grammatical morpheme type and grammaticality in left BA47 and SMA&SMedFG. Note. GCL = grammatical classifier; GASP = grammatical aspect marker; UCL = ungrammatical classifier; UASP = ungrammatical aspect marker; BA47 = left Brodmann area 47; LSMA&SMFG = left supplementary motor area and superior medial frontal gyrus. *$p < 0.05$.

doi:10.1371/journal.pone.0074952.g001
extracted to assess effects of grammatical morphemes and grammaticality, and their interaction. Task-independent regions showing greater activation associated with classifier processing included left BA47 and SMA&SMedFG, whereas stronger response to aspect marker processing was found in the left posterior middle temporal gyrus. We consider below each of these regions in terms of its possible role in processing nominal and verbal grammatical morphemes in Chinese given previous claims that have been made about its function(s).

As discussed in the Introduction, two major functional roles relevant to the present investigation have been associated with the left ventral prefrontal cortex – domain general executive functions and language-specific functions. The domain general account contends that the prefrontal cortex supervises or coordinates with other parts of the brain to perform various cognitive activities. It functions as a control hub in extracting/gathering information from a large variety of sources and projecting back to these systems to guide some voluntary action/decision [47], or choosing an appropriate response according to the task at hand from amidst possible candidates [21]. It has also been linked to working memory, temporarily holding non-integrated pieces of information before processing [48]. As such, the activation in the prefrontal cortex is supposed to be sensitive to task difficulty, as higher task demand corresponds to heavier load exerting on computational and/or working memory resources in any task. Hence, its involvement in grammatical morpheme processing might simply reflect a difference in paradigmatic complexity between morphosyntactic operations of nouns and verbs [13,49,50]. Researchers have further explained the engagement of the prefrontal cortex in terms of functional or structural connections to other more domain specific regions [48,51], e.g. LIFG – left temporal pole network for semantic processing, LIFG-left posterior MTG for syntactic processing [52,53].

Alternatively, it has been argued that specific region(s) in the left prefrontal cortex are dedicated to linguistic processing. Moreover, functional fractionation within this region has been proposed – BA47 for semantic processing (e.g., [54], for Chinese [55], Broca’s area for syntactic processing (e.g., [56], for Chinese [57]), semantic processing (e.g. [42,44,57] for Chinese), and morphosyntactic processing (e.g., [58]). Some studies tried to demonstrate the specific role of these areas in linguistic processing regardless of influence of working memory or selection demand by balancing computational loads (e.g., [59] for semantic processing, and [11] for morphosyntactic processing), or adding orthogonal/independent manipulation of task difficulty (e.g., [60] for semantic processing, and [61]; [62] for syntactic processing, but see [63] for alternative interpretations of [62]).

Two left prefrontal areas, BA47 and LSMA&SMedFG, were found to respond more strongly to classifier than aspect marker processing in this study. While one may interpret these areas as underlying morphological processes of nouns, it is also reasonable to propose that participants may need more computational resources, perhaps in terms of working memory or selection demand, to process CL sentences as there are more activated morphemes in the case of CL than ASP operations. This interpretation is consistent with the signficantly longer RT in the classifier than ASP condition in the sentence completion task (Table 2). We are aware of previous claims from a series of Chinese studies that activation in the left inferior frontal cortex basically supports semantic processing, as the language lacks inflectional morphology, morphosyntactic processes, and purely syntactic violations [42]. However, it is important to note that all the evidence comes from tasks involving semantic judgment at the sentence level [42,44,57] or the lexical level [55,64,65,66,67,68,69]. While making judgment about semantic acceptability or relatedness clearly involves semantic processes, processing semantic incongruency or making relatedness decision about two items may also be more resource demanding, which may or may not be reflected in response latency. Our previous finding from a semantic relatedness judgment task also identified a marginally significant cluster in BA44 related to word class effects [25], but the region was no longer significant when we looked for convergence between the judgment task and a semantic associate production task [26]. Most relevant to our consideration of the functional role of BA47 is a study that employed reversible two-character Chinese words, e.g. lǐn ´NÔ ‘to lead’ and bēi ´N ‘a necktie’, and concluded that executive control processes of semantic retrieval modulated activities in that area [55]. Interestingly, that study also found the Broca’s area more responsive to the ‘low conflict’ compared with neutral condition. Given these findings in Chinese, it would be more parsimonious to attribute the stronger response for CL stimuli in BA47 and LSMA&SMedFG to higher processing demand, resulting from activation of morphemes of a larger CL inventory. One noteworthy finding in the current study was the significant interaction between grammaticality and grammatical morpheme type but without significant simple effects [25], but the region was no longer significant when we looked for convergence between the judgment task and a semantic associate production task [26]. Of most relevance to our consideration of the functional role of BA47 is a study that employed reversible two-character Chinese words, e.g. lǐn ´NÔ ‘to lead’ and bēi ´N ‘a necktie’, and concluded that executive control processes of semantic retrieval modulated activities in that area [55]. Interestingly, that study also found the Broca’s area more responsive to the ‘low conflict’ compared with neutral condition. Given these findings in Chinese, it would be more parsimonious to attribute the stronger response for CL stimuli in BA47 and LSMA&SMedFG to higher processing demand, resulting from activation of morphemes of a larger CL inventory. One noteworthy finding in the current study was the significant interaction between grammaticality and grammatical morpheme type but without significant simple effects [25].

Figure 2. Task-independent regions of grammatical morpheme processing associated with Chinese nouns and verbs. Note. Classifier-specific regions are drawn in yellow and aspect marker-specific region in blue.

doi:10.1371/journal.pone.0074952.g002
Finally, although the main findings of this study have come from brain imaging data, one aspect of the behavioral results is worth mentioning for future studies involving processing of nominal classifiers in Chinese and using abstract and concrete nouns. We reported earlier a significant two-way interaction in error rate, albeit insignificant simple effects, between grammatical morpheme and concreteness in the grammaticality judgment task with higher error rates for ungrammatical trials at the concrete level but a reversed pattern for trials with abstract content words. An examination of Table 2 suggests that the higher error rate of abstract grammatical trials than ungrammatical trials might be driven by the particularly high error rate of grammatical sentences containing abstract nouns. We propose that this may be related to the greater flexibility of use of classifier for abstract concepts. That is, compared with most concrete nouns with specific classifiers, there is relatively lower agreement on the most appropriate classifier for an abstract noun.

In conclusion, through contrasting the processing of classifiers and aspect markers representing, respectively, nominal and verbal morphological operations in receptive and expressive tasks, we have found converging evidence for brain regions differentially responsive to one type of stimuli over the other, and vice versa. We have attributed the activation in the left prefrontal cortex to greater paradigmatic complexity of classifiers than aspect markers, which may reflect domain general computational loads, consistent with views from studies of European languages [18,19,20,21,70], and the left posterior temporal gyrus to more demanding verb semantic processing stemming from incongruency between aspect markers and semantic structure of verbs. Our results have contributed for the first time to cross-linguistic study of neural representation of grammatical morpheme processing from an analytic and a classifier language.

**Supporting Information**

**Appendix S1** Sentence stimuli containing nominal classifiers or verbal aspect markers in sentence completion and grammaticality judgment tasks.

**Author Contributions**

Conceived and designed the experiments: YB ZH SPL. Performed the experiments: XY. Analyzed the data: XY. Wrote the paper: XY SPL.


