博士論文

Computational principles of syntax in the language areas: Verification of the syntactic operations using fMRI

（言語野における文法の計算原理：**fMRI** による統辞操作の検証）

太田　真理

Shinri Ohta

Sakai Laboratory

Department of Life Sciences

Graduate School of Arts and Sciences

The University of Tokyo

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Meaning
Acc	accusative case
ACC	anterior cingulate cortex
AC–PC line	anterior to posterior commissure line
AF	arcuate fasciculus
AG	angular gyrus
ANCOVA	analysis of covariance
b0	image without diffusion-weighting
BA	Brodmann's area
BMS	Bayesian model selection
Comp	complementizer
Conjoined	conjoined sentence
Dat	dative case
DCM	dynamic causal modeling
d	dorsal
DoM	Degree of Merger
DTI	diffusion tensor imaging
EPI	echo-planar imaging
F3O	pars orbitalis of the inferior frontal gyrus
F3op	pars opercularis of the inferior frontal gyrus
F3t	pars triangularis of the inferior frontal gyrus
FA	flip angle
FDT	FMRIB's Diffusion Toolbox
FLIRT	FMRIB's Linear Image Registration Tool
fMRI	functional magnetic resonance imaging
FMRIB	Oxford Centre for Functional Magnetic Resonance Imaging of the Brain
FNIRT	FMRIB's Nonlinear Image Registration Tool
FOV	field of view
FSL	FMRIB's Software Library
FWE	family-wise error
FWHM	full-width at half maximum

Abbreviation	Meaning
Gen	genitive case
GLM	general linear model
L	left
LPMC	lateral premotor cortex
M	medial
MNI	Montreal Neurological Institute
n.s.	not significant
n/a	incalculable
N	(pseudo)noun phrase
N	node
Nested	nested sentence
Nom	nominative case
R	right
r^2	coefficient of determination
rANOVA	repeated-measures analysis of variance
Reverse	reverse-order string
ROI	region of interest
RSS	residual sum of squares
RT	reaction time
S, S’, S”	sentence
Same	same-order string
SEM	standard error of the mean
Simple	simple sentence
SLF	superior longitudinal fasciculus
SMG	supramarginal gyrus
SPM	statistical parametric mapping
t	trace
TE	echo time
Top	topic
TR	repetition time
V	(pseudo)verb phrase
VOI	volume of interest

Abstract

The nature of the computational principles of syntax remains to be elucidated. One promising approach to this problem is to construct formal and abstract linguistic models that parametrically predict the activation modulations in the language areas of the brain. In this dissertation, I verify computational principles of syntax in the language areas. First, I introduce the two fundamental linguistic operations: Merge (which combines two words or phrases to form a larger structure) and Search (which searches and establishes a syntactic relation of two words or phrases). I also illustrate certain universal properties of human language, and I propose Degree of Merger (DoM) as a key computational concept, which can be defined as the *maximum depth* of merged subtrees (called Mergers) within a given domain. I present hypotheses on how sentence structures are processed in the brain. Hypothesis I is that the DoM is a key computational concept to properly measure the complexity of tree structures. Hypothesis II is that the basic frame of the syntactic structure of a given linguistic expression is essentially determined by functional elements, which trigger Merge and Search. I then present our recent functional magnetic resonance imaging (fMRI) experiment, demonstrating that the DoM is indeed a key syntactic factor that accounts for syntax-selective activations in the left inferior frontal gyrus (L. F3op/F3t) and supramarginal gyrus (L. SMG). Using jabberwocky sentences with distinct constructions, I fitted various parametric models of syntactic, other linguistic, and nonlinguistic factors to activations measured with fMRI. I demonstrated that the models of the DoM and "DoM + number of Search" were optimum measures for explaining activations in the L. F3op/F3t and L. SMG, respectively. I further introduced letter strings, which had neither lexical associations nor grammatical particles, but retained both the matching and symbol orders of sentences. By directly contrasting jabberwocky sentences with letter strings, I found that localized activations in the L.

F3op/F3t and L. SMG were indeed independent of matching and symbol orders. Moreover, based on the dynamic causal modeling and diffusion tensor imaging results, I revealed the significance of the top-down connection from the L. F3op/F3t to L. SMG, suggesting that information about the DoM is transmitted through this specific dorsal pathway. These results indicate that the identified network of the L. F3op/F3t and L. SMG subserves the calculation of DoM in recursively merged sentences. Hypothesis III is that the DoM domain changes dynamically in accordance with iterative Merge applications, the Search distances, and/or task requirements. I confirm that the DoM accounts for activations in various sentence types. Hypothesis III successfully explains activation differences between object- and subject-relative clauses, as well as activations during explicit syntactic judgment tasks. Future research on the computational principles of syntax will further deepen our understanding of uniquely human mental faculties.

Chapter 1. General Introduction

Tree structures are among the most ubiquitous structures in nature, appearing in the branching of rivers, lightning, snowflakes, trees, blood vessels, nervous systems, etc., and can be simulated in part by fractal geometry (Mandelbrot, 1977). To properly quantify the complexity of such tree structures, various models have been proposed. The number of nodes would be one of the simplest models; this approach consists of simply counting the total number of nonterminal nodes (branching points) and terminal nodes of a tree structure (Figure 1A). This model obviously cannot capture hierarchical levels within the tree ("sister relations" in linguistic terms). To properly measure the hierarchical levels of a tree structure, I have proposed the Degree of Merger (DoM) as a key computational concept (Figure 1B) (Ohta, Fukui, & Sakai, 2013a, 2013b). The DoM is defined as the *maximum depth* of merged subtrees (called Mergers) within a given domain. With this model, the same numbers are assigned to the nodes with an identical hierarchical level. The DoM corresponds to the number of iterations for generating fractal figures, when the tree structures are self-similar.

In this dissertation, I examine whether parametric models based on modern linguistics can account for activations in the language areas of the brain measured with functional magnetic resonance imaging (fMRI). I first explain certain universal properties of human language discovered in modern linguistics, and I present hypotheses of how sentence structures are processed in the brain (Ohta, Fukui, & Sakai, 2013b). I then introduce our fMRI study, which demonstrates that the DoM is indeed a key syntactic factor that accounts for syntax-selective activations in the language areas (Ohta, Fukui, & Sakai, 2013a). I also show that the top-down connectivity from the left inferior frontal gyrus to the left supramarginal gyrus is critical for the syntactic processing. Next, I clarify that the DoM can account for activation modulations in the frontal region, depending on different sentence structures. Finally, I hypothesize that the DoM domain changes dynamically in accordance with iterative Merge applications, the distance required for Search operations (or simply the

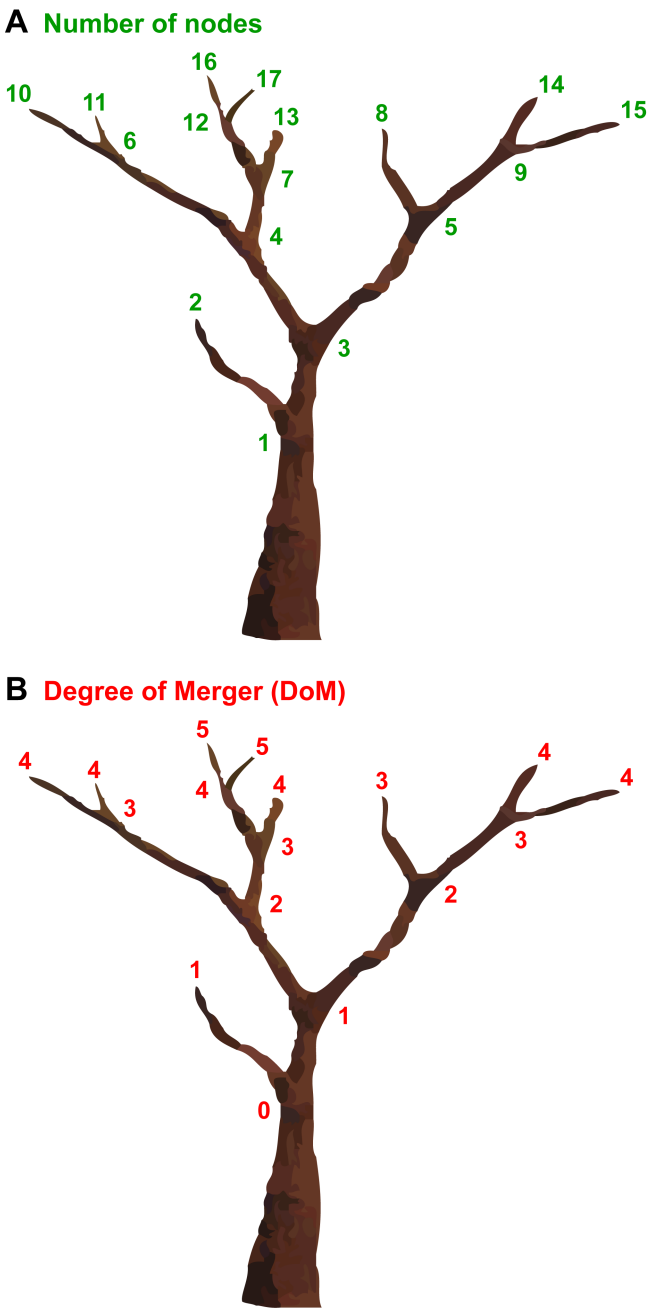


Figure 1. Two models for measuring the complexity of tree structures.
(**A**) The "number of nodes" counts the total number of nonterminal nodes (branching points) and terminal nodes of a tree structure. The number of nodes in the tree structure shown is 17. (**B**) The "Degree of Merger" (DoM) quantifies the maximum depth of merged subtrees, or the degree of branching. We increased the number one by one for each node, starting from the trunk (zero) to terminal nodes. The DoM of the tree structure shown is 5.

"Search distance"), and/or task requirements. This hypothesis accounts for activation differences between subject- and object-relative clauses, as well as activation during explicit syntactic judgment tasks.

1.1. Theoretical background

Modern linguistics has clarified universal properties of human language that, directly or indirectly, reflect the computational power, or engine, of the human language faculty. A sentence is not a mere string of words, but is made of phrase structures (called constituent structures). Moreover, a single phrase contains the key element (i.e., the "head") that determines the basic properties of the phrase. Furthermore, a sentence can be recursively embedded within other sentences, as in "*I think that John believes that Mary assumes that…*," and there is in principle no upper bound for the length of sentences. This property is the so-called *discrete infinity* made possible by the computational power, or engine, of the human language faculty. These universal properties can be adequately and minimally expressed by hierarchical tree structures with a set of relevant structural relations defined on such structures (Chomsky, 1957; Chomsky, 1965).

To construct hierarchical tree structures, modern linguistics has proposed the fundamental linguistic operation of *Merge* (capitalized in linguistics to indicate a formal operation). Merge is a structure-building operation that combines two syntactic objects (words or phrases) to form a larger structure (Chomsky, 1995). Merge would be theoretically "costless," requiring no driving force for its application (Saito & Fukui, 1998; Chomsky, 2004; Fukui, 2011). Besides Merge, Fukui and Sakai (2003) have proposed the *Search* operation, which applies to a syntactic object already constructed by Merge, to search for syntactic features. Search couples and connects two distinct parts of the same structure, thereby assigning relevant features from one to the other part. Various other "miscellaneous" operations that have been employed in the linguistics literature, such as Agree, Scope determination, Copy, etc., are in fact different manifestations of one and the same, i.e., more generalized, operation of Search (Fukui & Sakai, 2003). Thus, Agree, which has been assumed to be an operation mainly responsible for the agreement of grammatical features, is

actually not an operation specific to agreement, but rather is just an instance of the basic operation of Search, when it assigns specific features. Human language, therefore, should minimally contain two universal operations, Merge and Search. The total number of Merge and Search applications within an entire sentence are simply denoted here as "number of Merge" and "number of Search," respectively. The number of Merge applications in a sentence is always one less than the number of terminal nodes, *irrespective of sentence structures* (see Chapter 2, Materials and methods, Stimuli).

1.2. Symbol sequences and formal languages

Three specific types of languages have been discussed in the linguistics literature for formal symbol sequences beyond the bounds of finite state languages: (i) "counter language," (ii) "mirror-image language," and (iii) "copying language" (cf. Chomsky, 1957, p. 21).

> (i) *ab*, *aabb*, *aaabbb*, …, and in general, all sentences consisting of *n* occurrences of *a* followed by *n* occurrences of *b* and only these;
>
> (ii) *aa*, *bb*, *abba*, *baab*, *aaaa*, *bbbb*, *aabbaa*, *abbbba*, …, and in general, all sentences consisting of a string *X* followed by the 'mirror image' of *X* (i.e., *X* in reverse), and only these;
>
> (iii) *aa*, *bb*, *abab*, *baba*, *aaaa*, *bbbb*, *aabaab*, *abbabb*, …, and in general, all sentences consisting of a string *X* of *a*'s and *b*'s followed by the identical string *X*, and only these.

Counter languages can be handled by a counting mechanism to match the number of each symbol, while the mirror-image language contains a mirror-image dependency, requiring more than a simple counter. If the number of symbols is not fixed (i.e., infinite), both of these

languages are beyond the bounds of finite-state grammars and are to be generated by context-free (simple) phrase structure grammars. Furthermore, copying languages with a cross-serial dependency clearly go beyond the bounds of even context-free phrase structure grammars, requiring a more powerful device, viz., context-sensitive phrase structure grammars or transformational grammars (Chomsky, 1959; Hopcroft & Ullman, 1979).

It remains a central issue in cognitive sciences whether the faculty of language is also shared by animals. Animals have thus been tested with regular symbol sequences such as $A^n B^n$ ($n \geq 2$; i.e., AABB, AAABBB, …) and $(AB)^n$ ($n \geq 2$; i.e., ABAB, ABABAB, …), which differ in *symbol order*. In an animal study, songbirds were trained to discriminate patterns of $A^n B^n$ and $(AB)^n$ in more than ten thousand trials (Gentner et al., 2006). However, this learning can be achieved by tracking symbol repetition or counting strategy alone (Corballis, 2007). There is also a recent report that songbirds seemed to discriminate strings with or without nesting (Abe & Watanabe, 2011); however, this learning can be achieved by simply remembering partial strings (Beckers et al., 2012). Along the lines of contrasting $A^n B^n$ and $(AB)^n$, fMRI studies have tested participants with different symbol sequences, such as $A_2 A_1 B_1 B_2$ versus $A_1 B_1 A_2 B_2$ (each subscript denotes a matching order), which also differ in matching order (Bahlmann, Schubotz, & Friederici, 2008). The difference in activation patterns can be simply explained by differences in any factor associated with matching and symbol orders, i.e., temporal order-related factors. It is thus necessary to completely control these general factors when extracting any syntactic factor from among the cognitive factors involved in actual symbol processing.

Since the number of symbols is inevitably fixed (i.e., finite) in any actual experiment, it should be noted that any symbol sequence can be expressed by a regular (finite state) grammar, i.e., the least powerful grammar in the so-called Chomsky hierarchy. Therefore, one cannot, in principle, claim from experiments that individual grammars (e.g.,

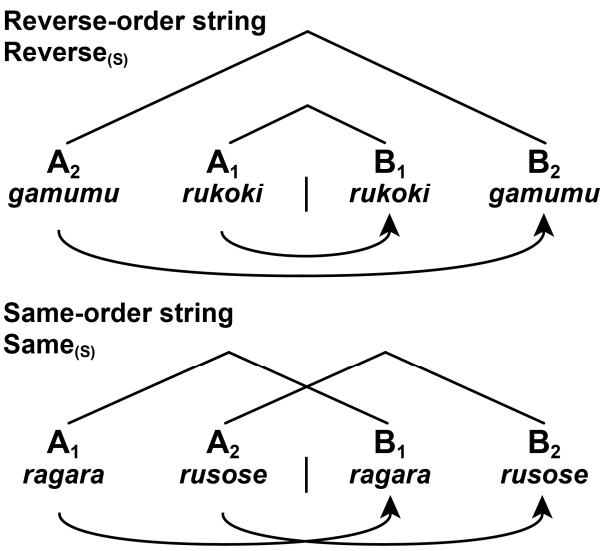


Figure 2. Two basic types of letter strings related to formal languages.
We tested two string conditions with short [(S) as a subscript] stimuli: Reverse$_{(S)}$ and Same$_{(S)}$. Each letter string was formed by jumbling letters of either a pseudonoun or pseudoverb (see Figure 4). We also tested the long stimuli with six items. Each curved arrow with an arrowhead denotes a Search operation, as in the following figures.
Symbols used: A, sample stimulus; B, comparison stimulus.

context-free phrase structure grammars vs. regular grammars) are represented differently in the brain. Thus, the neural representation of individual grammars was *not* within the scope of the present study. In addition to the various models examined, other nonstructural and nonsymbolic models with simple recurrent networks have been proposed to process even some context-free and context-sensitive phrase structure languages, generalizing these models to some degree to longer strings than the training set (Rodriguez, 2001). However, these models do not account for any parametric modulation of the activation reported in the present study, except for the length of sentences.

In the present experiment, I introduced letter strings with no lexical associations but having both symbol orders (e.g., AABB and ABAB) and matching orders (e.g., A_2 A_1 B_1 B_2). There were two basic types of strings: reverse-order strings (Reverse) and same-order strings (Same). In the Reverse strings, the first and second halves of the strings were presented in the reverse order, while in the Same strings, the halves were presented in the same order (Figure 2). Under these conditions, there was actually no path connecting the nonterminal nodes of

symbol pairs (e.g., A_1 B_1 and A_2 B_2), as there was *no* Merge application to connect the multiple pairs. Both the Reverse and Same strings took type (i) above (A^n B^n) in terms of symbol order. With regard to the matching orders, the Reverse strings took the type (ii) (A_2 A_1 B_1 B_2 or A_3 A_2 A_1 B_1 B_2 B_3), while the Same strings took the type (iii) (A_1 A_2 B_1 B_2 or A_1 A_2 A_3 B_1 B_2 B_3).

1.3. Hypothesis I

Given a tree structure with a formal property of Merge and *iterativity* (recursiveness) (Fukui, 2011), we propose the following hypothesis (Hypothesis I):

(1) The DoM, which can be defined as the *maximum depth* of merged subtrees within a given domain, is a key computational concept to properly measure the complexity of tree structures.

The DoM can be used to quantify and compare various syntactic phenomena, such as self-embedding, scrambling, *wh*-movement, etc. Furthermore, when Search applies to each syntactic object with its hierarchical structure, the calculation of the DoM plays a critical role. Indeed, from a nested sentence "[[*The boy*$_2$ [*we*$_3$ *like*$_3$]$_2$]$_1$ *sings*$_1$]$_0$" (subscripts denote the DoM for each node), two sentences "[*The boy* …]$_1$ *sings*$_1$" and "*we*$_3$ *like*$_3$" are obtained, where relevant features (numbers and persons here) are searched and matched between the nodes with the identical DoM. Since such analyses of hierarchical structures would produce specific loads in syntactic computation, we expect that the DoM and associated "number of Search" would affect performances and cortical activations.

Sentences with various constructions have been previously discussed in terms of sentence acceptability (cf. Chomsky, 1965, p. 12).

(i) nested constructions

(ii) self-embedded constructions

(iii) multiple-branching constructions

(iv) left-branching constructions

(v) right-branching constructions

Nested constructions are created by *centrally* embedding a phrase within another phrase (with some non-null element to its left and some non-null element to its right), and self-embedded constructions are a special case of nested construction in which nesting occurs within the *same* type of phrases (e.g., noun phrases). Multiple-branching constructions consist of conjoining phrases at the same hierarchical level, and left/right-branching constructions are yielded by merging a phrase in the left-most or right-most phrase. The degrees of nesting and self-embedding have already been proposed for modeling the understanding of sentences (Miller & Chomsky, 1963). By generalizing this attractive idea to include any construction with merged phrases, I introduced DoM as a key computational concept.

Based on nested (self-embedded), left/right-branching, and multiple-branching constructions, three basic types of sentences can be distinguished: the nested sentence (Nested), simple sentence (Simple), and conjoined sentence (Conjoined), respectively. Figure 3 shows some examples in Japanese. Given syntactic structures like the ones shown, the correspondence of each subject–verb pair becomes fixed. Here N and V denote a noun phrase and a verb phrase, respectively. For the sentence shown in Figure 3A, an entire sentence is constructed by nesting sentences in the form of $[N_2[N_1\ V_1]V_2]$, in which $[N_i\ V_i]$ represents a subject–verb pair of a sentence. Since Japanese is a head-last, and hence an SOV (verb-final) language, main verbs appear after subordinate clauses. Therefore, Japanese sentences

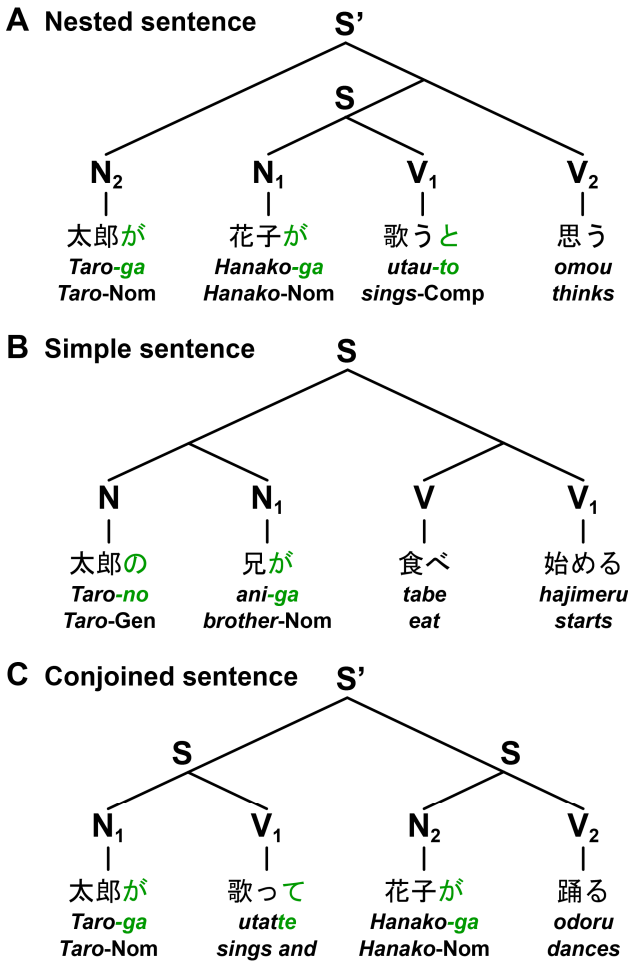


Figure 3. Japanese sentences with three major constructions.
The figure shows three basic types of sentences in Japanese: the nested sentence, simple sentence, and conjoined sentence. Based on contemporary linguistics (O'Grady et al., 2010), each diagram represents a unique tree structure of each sentence (S and S') constructed from Ns and Vs. Below each example, word-by-word translations in English are shown. (**A**) A sentence (S) at the lowest hierarchical level was nested into an entire sentence (S') ("*Taro-ga Hanako-ga utau-to omou*": "*Taro thinks that Hanako sings*"). (**B**) A simple sentence was constructed by adding the same number of left/right branches to both nouns and verbs ("*Taro-no ani-ga tabe hajimeru*": "*Taro's brother starts eating*"). (**C**) An entire sentence (S') was constructed by conjoining two sentences ("*Taro-ga utatte Hanako-ga odoru*": "*Taro sings, and Hanako dances*").
Symbols used: S and S', sentence; N, noun phrase; V, verb phrase; *-ga*, nominative case marker; *-no*, genitive case marker; *-to*, complementizer; *-te*, gerundive form; Nom, nominative case; Gen, genitive case; Comp, complementizer.

naturally yield nested structures without having to employ, as in English, object-relative clauses (e.g., "*The boy who*$_i$ *we like* t_i *sings*"), which require "movement" of an object (i.e., with more Merge applications) and thus leave behind a "trace" (t_i, subscripts denote the same entity). For the sentence shown in Figure 3B, a simple sentence is constructed by adding the same number of left/right branches to both Ns and Vs. The last noun (i.e., head) in the branches of Ns creates a subject–verb pair with the last verb (the head) of a compound verb.

Each simple sentence thus takes the form [(NN_1) (VV_1)]. For the sentence shown in Figure 3C, an entire sentence is constructed by conjoining sentences in the form of [N_1 V_1][N_2 V_2]. When considering longer sentences such as N_3 N_2 N_1 V_1 V_2 V_3, these constructions have distinct values for DoM.

1.4. Hypothesis II

In any sentence, functional elements, such as inflections, auxiliary verbs, and grammatical particles, serve an essentially grammatical function without descriptive content. We thus propose the following hypothesis (Hypothesis II) for the fundamental roles of these functional elements:

> (2) The basic frame of the syntactic structure of a given linguistic expression (e.g., sentence) is determined essentially by functional elements, which trigger Merge and Search operations.

In the nonsense poem "Jabberwocky" by Lewis Carroll, e.g., "*'Twas* ('*It was*') *brillig, and the slithy toves did* ...," the basic frames of syntactic structures are indeed determined by the functional elements "*'Twas*," "*and*," "*the*," "*-s*," and "*did*." In the Japanese language, grammatical particles and morphosyntactic inflections are functional elements. The sentences shown in Figure 3 actually contain only three kinds of grammatical particles, which represent *canonical* (i.e., in a prototypical use) case markings and syntactic information in Japanese: *-ga*, a nominative case marker; *-no*, a genitive case marker; and *-to*, a complementizer. It should be noted that both the nested and simple sentences have the same symbol order (N^n V^n), but they have different grammatical particles and syntactic structures. In contrast, both the simple and conjoined sentences have the same tree structures as a result, but different

symbol orders ($N^n V^n$ or $(NV)^n$ ($n \geq 2$)). It is the grammatical particles and morphosyntactic inflections, but not symbol orders or matching orders themselves, that determine the basic frame of syntactic structures of a sentence.

Following morphosyntactic and phonological features of Japanese verbs (Tsujimura, 2007), Vs take a nonpast-tense form (*-ru*), a past-tense form (*-ta*), or a gerundive form (*-te*); Vs ending with *-to* and *-te* introduce *that*-clauses and *and*-conjunctives, respectively. Gerundive forms can be used not only in *and*-conjunctives, but in compound verbs (e.g., "*tabete-simau*": "*to finish eating*"; actual Japanese words will be translated hereafter), much as gerunds can in English. The *-ga*, *-no*, *-to*, and *-te* endings (*green* letters in Figures 3–5), together with the first verb of a compound verb in an adverbial form (e.g., "*tabe*"), are associated with Merge applications to connect multiple nouns/verbs or sentences, amounting to "number of Merge." The Japanese language lacks "agreement features" (i.e., number, person, gender, etc.), but is nevertheless equipped with the general Search procedure employed in agreement phenomena in other languages. This Search mechanism is in fact attested for various other phenomena in Japanese (see Fukui and Sakai, 2003, for further discussion). For example, the Japanese language exhibits a phenomenon called "honorification," in which a noun phrase denoting an honored person and the form of honorific used on a verb match (Gunji, 1987; Ivana & Sakai, 2007).

In this chapter, I provided some theoretical discussions based on modern linguistics, focusing on two fundamental linguistic operations, Merge and Search. I hypothesized that the DoM is a key computational concept to properly quantify the complexity of tree structures, and that the basic frame of the syntactic structure of a given linguistic expression is determined essentially by grammatical particles and morphosyntactic inflections that trigger Merge and Search operations.

Chapter 2. Syntactic Computation in the Human Brain:

The DoM as a Key Syntactic Factor Elucidated by an fMRI Experiment

2.1. Introduction

It is widely accepted that in human language, a sentence can be expressed by a unique tree structure with recursive branches (Hopcroft & Ullman, 1979; O'Grady et al., 2010). One possible way to elucidate the neural basis of the computational properties of natural language is to examine how the brain responds to the modulation of specified syntactic factors. An early attempt with fMRI reported that activations in the language areas were modulated by noncanonical/canonical word orders and the presence/absence of lexical contents (Röder et al., 2002), in which multiple factors, including memory-related and semantic factors, could account for these activations. Therefore, we should not be content with such a general cognitive factor as so-called "syntactic complexity" or "syntactic working memory," which could involve both linguistic and nonlinguistic factors. We should instead identify minimal factors that *sufficiently* explain any activation change obtained. In addition, the size of linguistic constituents may also modulate cortical activations. A recent fMRI study reported that left frontal activations increased with the number of words or terminal nodes (symbols) in a phrase (Pallier, Devauchelle, & Dehaene, 2011), but, as rightly pointed out by the authors, precise phrase structures remained to be taken into account. We focused on different sentence constructions, and identified *minimal* syntactic factors associated with phrase structures, which parametrically modulated cortical responses measured with event-related fMRI (Ohta, Fukui, & Sakai, 2013a).

Modern linguistics has accumulated mounting evidence that the construction of any grammatical phrases or sentences can be adequately and *minimally* explained by hierarchical syntactic structures with a set of relevant structural relations defined on such structures (Chomsky, 1957; Chomsky, 1965). To properly measure the depth of a tree structure with a formal property of Merge and *iterativity* (recursiveness) (Fukui, 2011), we hypothesize that the DoM is a key computational concept, which can be defined as the *maximum depth* of

merged subtrees (i.e., Mergers) within a given domain. Moreover, the DoM can quantify and compare various syntactic phenomena, such as self-embedding, scrambling, *wh*-movement, etc. Furthermore, when Search applies to each syntactic object within a hierarchical structure, the calculation of the DoM plays a critical role. Indeed, from a nested sentence "[[*The boy*$_2$ [*we*$_3$ *like*$_3$]$_2$]$_1$ *sings*$_1$]$_0$" (subscripts denote the DoM for each node, see Figure 4A), two sentences "[*The boy* …]$_1$ *sings*$_1$" and "*we*$_3$ *like*$_3$" are obtained, where relevant features (numbers and persons here) are searched and checked between the nodes with the identical DoM. Because such analyses of hierarchical structures would produce specific loads in syntactic computation, we hypothesize that the DoM and associated "number of Search" modulate neural activations. As mentioned in Chapter 1, Merge would be theoretically "costless" (Saito & Fukui, 1998; Chomsky, 2004), and thus the "number of Merge" itself may not affect activations, which can be easily expected for *flat* structures (see Figure 4B).

In the present study, we used jabberwocky sentences, which consist of pseudonoun phrases (Ns) and pseudoverb phrases (Vs) that lack lexical associations but have grammatical particles and morphosyntactic inflections. According to Hypothesis II stated above, these jabberwocky sentences had the same syntactic structures as normal sentences (see Materials and methods, Stimuli). Based on the nested (self-embedded), left/right-branching, and multiple-branching constructions (see Chapter 1, Hypothesis I), we introduced three basic types of sentence constructions: nested sentences (Nested), simple sentences (Simple), and conjoined sentences (Conjoined) (Figure 4A). According to the second hypothesis stated above, the jabberwocky sentences had the same syntactic structures as normal sentences even without lexical meanings (Figures 3 and 4). When constructing syntactic structures like the ones shown in Figure 4A, the correspondence of each subject–verb pair is most crucial. To test whether participants actually paid attention to this correspondence, we used a matching task in which the vowel of a subject (N_i as a sample stimulus) was matched with the last

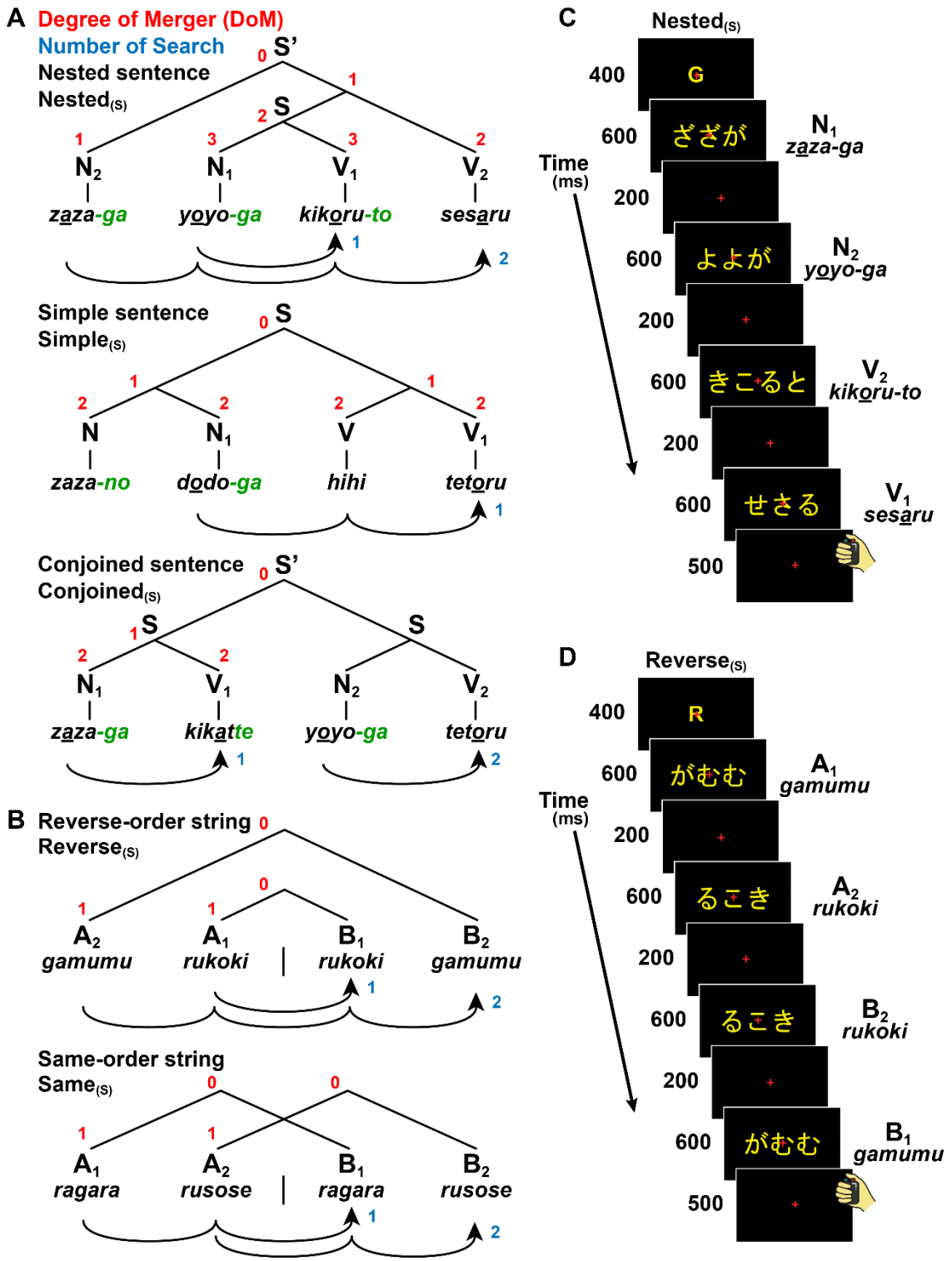


Figure 4. A paradigm for testing jabberwocky sentences and letter strings.
Examples of short [(S) as a subscript] matching stimuli are shown here with the Romanization system, but actual stimuli were presented in hiragana without hyphens (see C and D). Note the syntactic structures of these jabberwocky sentences are same as those of real sentences in Figure 3. We also tested the long stimuli with six words. (**A**) Three sentence conditions with short stimuli: Nested(S), Simple(S), and Conjoined(S). The digits shown in red and blue denote the DoM for each node and "number of Search," respectively (see Table 3). The curved arrows denote the matching of sequentially presented stimuli. (**B**) Two string conditions with short stimuli: Reverse(S) and Same(S). Each letter string was formed by jumbling letters of either N or V. (**C** and **D**) Examples of stimulus presentation. Here examples of matching stimuli are shown in hiragana for the Nested(S) and Reverse(S). Between the Nested(S) and Reverse(S), both of the symbol orders (the order of Ns, Vs, As, and Bs) and matching orders (denoted by subscripts) were identical.

vowel of a corresponding verb root (V_i as a comparison stimulus) (e.g., "*rara-ga tetaru*," underlined vowels within pseudowords), probing the goal with the same vowel as explained above. These features of vowels were only *experimentally* introduced, and this matching involved a factor of encoding (i.e., memorization of features necessary for matching). By generalizing the role of Search, we assumed that Search applied to a subject–verb pair (Ohta, Fukui, & Sakai, 2013a, 2013b). Because Vs lacked grammatical (agreement) features (e.g., number, person, gender, etc.), as do regular Japanese verbs, this matching property did not mimic agreement itself, but involved a formal association between sample and comparison stimuli. It follows that the same syntactic structures were constructed from matching and nonmatching stimuli (e.g., "*rara-ga teturu*," Tables 1 and 2), which were both well-formed, i.e., *grammatical*, in Japanese. A matching strategy (counting, for example, the first and the fourth stimuli for matching) was useful in solving the task, but performing the task was *not* a prerequisite for constructing syntactic structures. The matching task differed from the classification tasks for symbol orders (e.g., AABB vs. ABAB, where A and B are symbols representing certain sets of stimuli), which can be solved by counting the maximum number of consecutively repeated symbols. We further examined whether cortical activations were modulated by the length of sentences: short (S as a subscript, e.g., $\text{Conjoined}_{(S)}$; four-phrase sentences) and long (L as a subscript, e.g., $\text{Conjoined}_{(L)}$; six-phrase sentences) (Figure 5A), where the DoM domain spanned four and six relevant words, respectively.

We tested various parametric models of syntactic, other linguistic, and nonlinguistic factors (Table 3; see Materials and methods, Operational definitions of all factors examined), some of which were based on structure-based models (Figures 6–8). Given these factors with a limited number of experimental conditions, we wanted to narrow down the models as much as possible by adopting effective contrasts. For both short and long sentences, estimates for the "number of Merge," as well as those for the "number of case markers (*-ga*/*-no*)" and the

Table 1. Examples of short nonmatching stimuli.

Condition	Error type	Stimulus example
Nested sentence, short	$[N_2[N_1\ V_1]V_2{*}]$	*mumu-ga rara-ga tetaru-to hih<u>o</u>ru*
(Nested$_{(S)}$)	$[N_2[N_1\ V_1{*}]V_2{*}]$	*dodo-ga gugu-ga tet<u>o</u>ru-to kik<u>a</u>ru*
	$[N_2[N_1\ V_1{*}]V_2]$	*rara-ga mumu-ga ses<u>o</u>tta-to kikatta*
Simple sentence, short	$[(NN_1)\ (VV_1{*})]$	*rara-no gugu-ga tete ses<u>a</u>tta*
(Simple$_{(S)}$)		
Conjoined sentence, short	$[N_1\ V_1][N_2\ V_2{*}]$	*rara-ga hihatte gugu-ga tet<u>o</u>ru*
(Conjoined$_{(S)}$)	$[N_1\ V_1{*}][N_2\ V_2{*}]$	*yoyo-ga hih<u>u</u>tte rara-ga ses<u>o</u>tta*
	$[N_1\ V_1{*}][N_2\ V_2]$	*gugu-ga tet<u>o</u>tte zaza-ga hiharu*
Reverse-order string, short	$A_2\ A_1\ B_1\ B_2{*}$	*nododo rukiku rukiku <u>donodo</u>*
(Reverse$_{(S)}$)	$A_2\ A_1\ B_1{*}\ B_2{*}$	*gayoyo settasa <u>sasseta</u> <u>gadodo</u>*
	$A_2\ A_1\ B_1{*}\ B_2$	*mugamu sessota <u>kittako</u> mugamu*
Same-order string, short	$A_1\ A_2\ B_1\ B_2{*}$	*ruteta gugagu ruteta <u>yogayo</u>*
(Same$_{(S)}$)	$A_1\ A_2\ B_1{*}\ B_2{*}$	*yonoyo rusesu <u>donodo</u> <u>rususe</u>*
	$A_1\ A_2\ B_1{*}\ B_2$	*norara kuruki <u>noyoyo</u> kuruki*

For each condition, nonmatching stimuli with errors in different positions are listed in descending order of frequency. Under the sentence conditions, Vs with asterisks represent matching errors (underlined vowels in stimulus examples here; no underline in the real stimuli). Under the string conditions, Bs with asterisks represent matching errors (underlined letter strings in stimulus examples here; no underline in the real stimuli).

"depth of postponed symbols," were identical among the three sentence conditions. By taking one of sentence conditions as a reference, these three factors could be eliminated from the analyses. Moreover, a reference condition should be chosen separately for each of short and long sentences, as we tested the short and long stimuli on separate days. The Conjoined condition was actually *simplest* among the three sentence conditions and thus served as an appropriate reference, because the Conjoined condition had same or smaller estimates as Nested and Simple conditions for all factors except for numbers of Search and encoding.

Table 2. Examples of long nonmatching stimuli.

Condition	Error type	Stimulus example
Nested$_{(L)}$	$[N_3[N_2[N_1\ V_1]V_2]V_3{*}]$	*mumu-ga zaza-ga yoyo-ga kikotta-to hihatta-to tetatta*
	$[N_3[N_2[N_1\ V_1]V_2{*}]V_3]$	*dodo-ga rara-ga mumu-ga hihuru-to teturu-to kikoru*
	$[N_3[N_2[N_1\ V_1]V_2{*}]V_3{*}]$	*dodo-ga mumu-ga zaza-ga tetaru-to sesoru-to hihuru*
	$[N_3[N_2[N_1\ V_1{*}]V_2]V_3{*}]$	*rara-ga mumu-ga yoyo-ga hiharu-to teturu-to sesoru*
Simple$_{(L)}$	$[((NN)N_1)\ ((VV_1)V_1{*})]$	*rara-no gugu-no yoyo-ga kiki sesotte teturu*
	$[((NN)N_1)\ ((VV_1{*})V_1)]$	*gugu-no zaza-no dodo-ga kiki tetatte sesoru*
	$[((NN)N_1)\ ((VV_1{*})V_1{*})]$	*yoyo-no rara-no mumu-ga tete hihotte kikotta*
Conjoined$_{(L)}$	$[N_1\ V_1][N_2\ V_2][N_3\ V_3{*}]$	*dodo-ga tetotte mumu-ga sesutte zaza-ga hihoru*
	$[N_1\ V_1][N_2\ V_2{*}][N_3\ V_3]$	*gugu-ga kikutte zaza-ga tetotte dodo-ga sesoru*
	$[N_1\ V_1][N_2\ V_2{*}][N_3\ V_3{*}]$	*zaza-ga sesatte yoyo-ga kikutte gugu-ga tetotta*
	$[N_1\ V_1{*}][N_2\ V_2][N_3\ V_3{*}]$	*mumu-ga sesotte rara-ga kikatte dodo-ga hihutta*
Reverse$_{(L)}$	$A_3\ A_2\ A_1\ B_1\ B_2\ B_3{*}$	*gazaza rusose gunogu gunogu rusose nozaza*
	$A_3\ A_2\ A_1\ B_1\ B_2{*}\ B_3{*}$	*ragara hiruhu gayoyo gayoyo huruhi gazaza*
	$A_3\ A_2\ A_1\ B_1\ B_2{*}\ B_3$	*serusa gugagu hohiru hohiru nogugu serusa*
	$A_3\ A_2\ A_1\ B_1{*}\ B_2\ B_3{*}$	*gunogu ruteta nododo noyoyo ruteta gugagu*
Same$_{(L)}$	$A_1\ A_2\ A_3\ B_1\ B_2\ B_3{*}$	*hiruho gunogu haruhi hiruho gunogu hahiru*
	$A_1\ A_2\ A_3\ B_1\ B_2{*}\ B_3{*}$	*dogado rutetu zagaza dogado suruse zanoza*
	$A_1\ A_2\ A_3\ B_1\ B_2{*}\ B_3$	*kattaki yonoyo tutetta kattaki nododo tutetta*
	$A_1\ A_2\ A_3\ B_1{*}\ B_2\ B_3{*}$	*noyoyo tahihha munomu nododo tahihha mugamu*

For the Simple$_{(L)}$, there were other error types, i.e., $V_1{*}(VV_1{*})$, $V_1(VV_1{*})$, and $V_1{*}(VV_1)$. For the Nested$_{(L)}$ and Conjoined$_{(L)}$, we included nonmatching stimuli with a maximum variety of vowels in Vs, as shown here; for the Simple$_{(L)}$, we included nonmatching stimuli with a least variety of vowels in Vs. Therefore, the strategy of noting the variety of vowels was not effective. Under all conditions, there were more variations in error for the long than short stimuli.

We further introduced letter strings, which had neither lexical associations nor grammatical particles, but retained both the matching orders and symbol orders of sentences. There were two string conditions: reverse-order strings (Reverse) and same-order strings

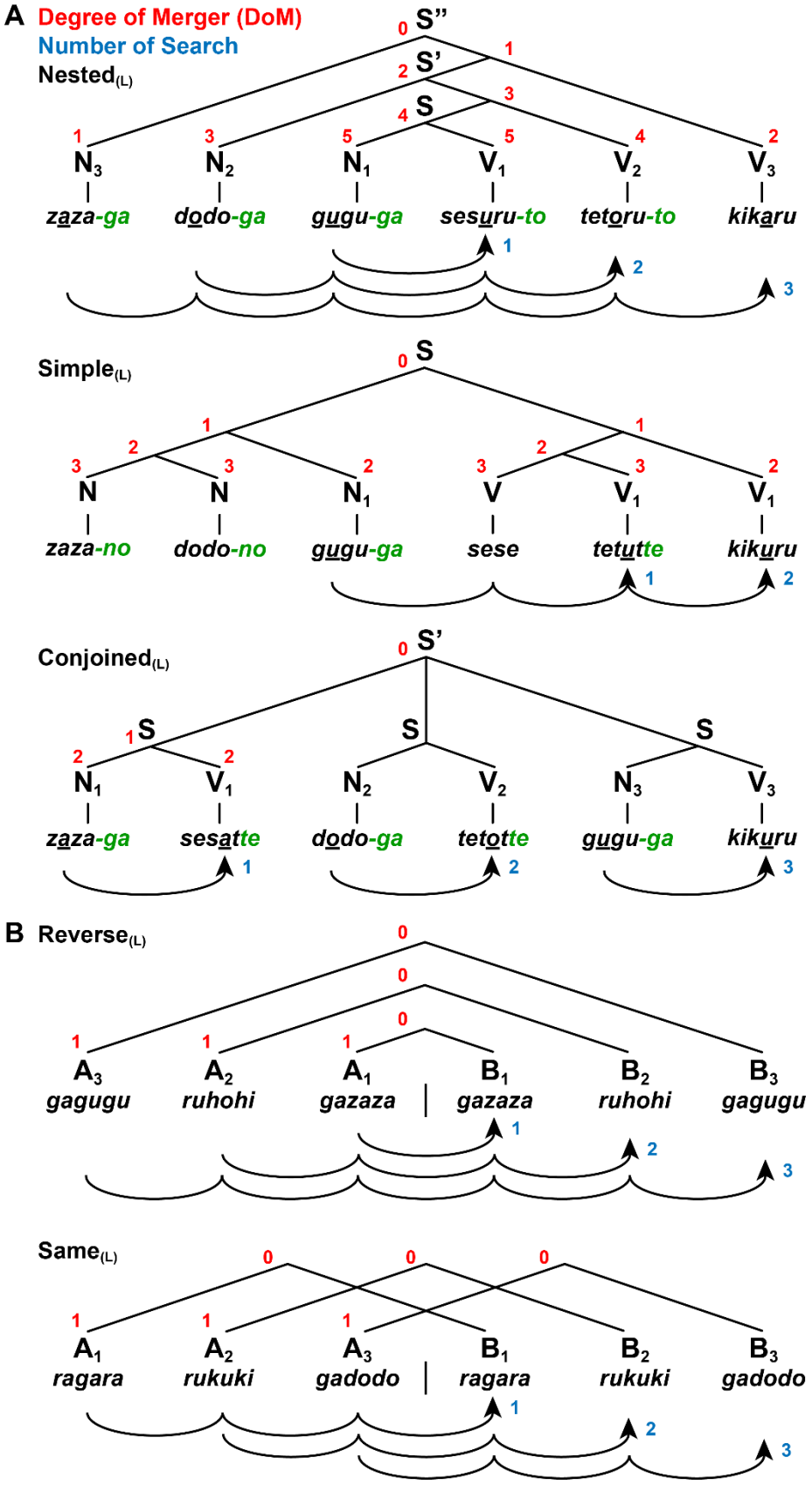


Figure 5. Examples of long matching stimuli.
(A) Three sentence conditions with long [(L) as a subscript] stimuli: Nested(L), Simple(L), and Conjoined(L). (B) Two string conditions with long stimuli: Reverse(L) and Same(L).

Table 3. Estimates of various factors to account for activations under the sentence conditions.

Syntactic factors	**Nested$_{(L)}$**	**Nested$_{(S)}$**	**Simple$_{(L)}$**	**Simple$_{(S)}$**	**Conjoined$_{(L)}$**	**Conjoined$_{(S)}$**
Degree of Merger (DoM)	5	3	3	2	2	2
No. of Search	3	2	2	1	3	2
No. of Merge	5	3	5	3	5	3

	Nested$_{(L)}$ − Conjoined$_{(L)}$	**Nested$_{(S)}$ − Conjoined$_{(S)}$**	**Simple$_{(L)}$ − Conjoined$_{(L)}$**	**Simple$_{(S)}$ − Conjoined$_{(S)}$**
DoM	3	1	1	0
DoM + No. of Search	3	1	0	−1
No. of Search	0	0	−1	−1
No. of Merge	0	0	0	0

Other linguistic factors	**Nested$_{(L)}$**	**Nested$_{(S)}$**	**Simple$_{(L)}$**	**Simple$_{(S)}$**	**Conjoined$_{(L)}$**	**Conjoined$_{(S)}$**
No. of case markers (*-ga*/*-no*)	3	2	3	2	3	2
No. of tense markers (*-ru*/*-ta*)	3	2	1	1	1	1
Degree of nesting	2	1	1	1	1	1
Degree of self-embedding	2	1	1	0	1	0
No. of nodes	11	7	11	7	10	7
Depth of postponed symbols	3	2	3	2	3	2
Integration costs	5	3	3	2	1	1
Storage costs	3	2	2	2	1	1
Syntactic interference	2	1	0	0	0	0
Positional similarity	3	2	2	0	0	0

	Nested$_{(L)}$ − Conjoined$_{(L)}$	**Nested$_{(S)}$ − Conjoined$_{(S)}$**	**Simple$_{(L)}$ − Conjoined$_{(L)}$**	**Simple$_{(S)}$ − Conjoined$_{(S)}$**
No. of case markers (*-ga*/*-no*)	0	0	0	0
No. of tense markers (*-ru*/*-ta*)	2	1	0	0
Degree of nesting	1	0	0	0
Degree of self-embedding	1	1	0	0
No. of nodes	1	0	1	0
Depth of postponed symbols	0	0	0	0
Integration costs	4	2	2	1
Storage costs	2	1	1	1
Syntactic interference	2	1	0	0
Positional similarity	3	2	2	0

Nonlinguistic factors	**Nested$_{(L)}$**	**Nested$_{(S)}$**	**Simple$_{(L)}$**	**Simple$_{(S)}$**	**Conjoined$_{(L)}$**	**Conjoined$_{(S)}$**
Memory span	4	2	2	1	0	0
Counting	2	1	2	1	0	0
No. of encoding	6	4	3	2	6	4

	Nested$_{(L)}$ − Conjoined$_{(L)}$	**Nested$_{(S)}$ − Conjoined$_{(S)}$**	**Simple$_{(L)}$ − Conjoined$_{(L)}$**	**Simple$_{(S)}$ − Conjoined$_{(S)}$**
Memory span	4	2	2	1
Counting	2	1	2	1
No. of encoding	0	0	−3	−2
Memory span + counting	6	3	4	2
Memory span + No. of encoding	4	2	−1	−1

We define the estimate of a factor as the largest value that the factor can variably take within an entire sentence. For each factor, its unit load should be invariable among all sentence conditions, making an independent subtraction between estimates of the *same* factor possible. Estimates under the Conjoined condition as a reference were subtracted from those under the other Nested and Simple conditions (e.g., DoM for [Nested$_{(L)}$ − Conjoined$_{(L)}$], 5 − 2 = 3), separately for long and short sentences. Excluding the "number of Merge," the estimates of which were null for all four contrasts, we regarded the "DoM + number of Search" (i.e., adding the estimates of two factors) as an additional factor. Among the nonlinguistic factors, "memory span + counting" and "memory span + number of encoding" were regarded as additional factors, because they were temporal order-related and memory-related factors, respectively.

(Same) (Figures 4B, 5B, and Table 4). As with the sentence conditions, we used the same matching task under both string conditions, in which the first half of a string (A_i as a sample stimulus) was matched with the corresponding second half (B_i as a comparison stimulus) in the reverse or same order. These string conditions also controlled for any involvement of matching strategy stated above. Between the Nested ($N_2 N_1 V_1 V_2$ or $N_3 N_2 N_1 V_1 V_2 V_3$, where each subscript denotes a matching order) and Reverse ($A_2 A_1 B_1 B_2$ or $A_3 A_2 A_1 B_1 B_2 B_3$) conditions, the curved arrows shown in Figures 4 and 5 represent the *same* matching order of sequentially presented stimuli (e.g., for $N_2 N_1 V_1 V_2$, the inner symbol pair of N and V is matched first, and then the outer symbol pair is matched). The symbol order was also identical among the Nested, Simple, Reverse, and Same conditions, and took the form $N^n V^n$ or $A^n B^n$. To control both matching and symbol orders, we directly compared the Nested with the Reverse, using the Simple and Same conditions as respective references (Table 4), i.e., (Nested − Simple) > (Reverse − Same), where we combined the short and long stimuli. Combining these multiple conditions, we were able to properly examine whether different structures were actually constructed for sentences and strings. The goal of these thorough controls was to demonstrate that the *purely* syntactic factors of the DoM and "number of Search" actually modulate neural activations.

It has been reported that more complex sentences elicit larger activations in the pars opercularis and pars triangularis of the left inferior frontal gyrus (L. F3op/F3t) (Stromswold et al., 1996; Dapretto & Bookheimer, 1999; Embick et al., 2000; Hashimoto & Sakai, 2002; Friederici et al., 2003; Suzuki & Sakai, 2003; Musso et al., 2003; Kinno et al., 2008), suggesting that the L. F3op/F3t is critical for syntactic processing as a grammar center (Sakai, 2005). On the other hand, the left angular gyrus and supramarginal gyrus (L. AG/SMG) have been suggested as important for vocabulary knowledge or lexical processing (Lee et al., 2007; Pattamadilok et al., 2010). To examine the functional specialization of any regions,

Table 4. Estimates of nonlinguistic and syntactic factors to account for activations.

Nonlinguistic factors	Nested$_{(L)}$	Nested$_{(S)}$	Simple$_{(L)}$	Simple$_{(S)}$	Reverse$_{(L)}$	Reverse$_{(S)}$	Same$_{(L)}$	Same$_{(S)}$
Memory span	4	2	2	1	4	2	2	1
Counting	2	1	2	1	2	1	2	1
No. of encoding	6	4	3	2	6	4	6	4

	Nested	Simple	Reverse	Same
Memory span	6	3	6	3
Counting	3	3	3	3
No. of encoding	10	5	10	10

	Nested − Simple	Reverse − Same
Memory span	3	3
Counting	0	0
No. of encoding	5	0

Syntactic factors	Nested − Simple	Reverse − Same
DoM	3	0
DoM + No. of Search	5	0

For Nested, Simple, Reverse, and Same conditions, estimates for short and long stimuli were added together because each factor's unit load would be invariable between short and long stimuli under each of the sentence and string conditions. Because the matching orders or symbol orders were identical between the Nested and Reverse conditions, the unit load of memory span or counting was invariable between the Nested and Reverse conditions, which was also invariable between the Reverse and Same conditions, thus invariable among the Nested, Simple, Reverse, and Same conditions. Note that the estimates of memory span in [Nested − Simple] and [Reverse − Same] also became identical, and that the Reverse − Same contrast makes the listed estimates null, except memory span. The last two syntactic factors, whose models were best in Tables 7 and 8, consistently accounted for the results of Figure 11F. All estimates of the other factors unlisted here were null in [Reverse − Same], which cannot account for the results of Figures 13C and 13D.

including the L. F3op/F3t and L. AG/SMG, in an unbiased manner, we adopted whole-brain analyses (Friston & Henson, 2006). The spatial and temporal resolution of fMRI, as well as its sensitivity, has been proven to be high enough to confirm various hypotheses, such as ours, about human cognitive functions. We also performed effective connectivity analyses by using dynamic causal modeling (DCM) (Friston, Harrison, & Penny, 2003) to examine the

functional integration of identified regions. To provide an empirical backup for the connection derived from DCM, we checked the anatomical plausibility of the network with diffusion tensor imaging (DTI). The findings of recent DTI studies have raised a debate about the functional roles of two different pathways in language processes: the dorsal tracts of the superior longitudinal fasciculus and arcuate fasciculus (SLF/AF), and the ventral tracts of the middle longitudinal fasciculus and extreme capsule (Saur et al., 2008; Wong et al., 2011; Wilson et al., 2011; Griffiths et al., 2013). Both pathways connect the inferior frontal and superior/middle temporal areas. The present study would elucidate the most crucial network and pathways for syntactic computation.

2.2. Materials and methods

2.2.1. Participants

Eighteen native Japanese speakers (all males, aged 19–25 years), who had not majored in linguistics, participated in an fMRI experiment. Fifteen additional participants (14 males, aged 19–40 years) were tested in a DTI experiment. All participants in the fMRI and DTI experiments were healthy and right-handed (laterality quotients: 11–100), according to the Edinburgh inventory (Oldfield, 1971). Prior to participation in the study, written informed consent was obtained from each participant after the nature and possible consequences of the studies were explained. Approval for the experiments was obtained from the institutional review board of the University of Tokyo, Komaba.

2.2.2. Stimuli

Each visual stimulus consisted of two to five yellow letters in hiragana (Figures 4C and 4D). The stimuli were visually presented against a dark background through an eyeglass-like MRI-compatible display (resolution, 800 × 600; VisuaStim XGA; Resonance Technology

Inc., Northridge, CA). The visual stimuli were always presented at the center of the monitor. At the start of every trial of the Nested, Simple, and Conjoined conditions, the cue "G" (for grammar conditions with all grammatical sentences) was shown for 400 ms. The cue "R" (for reverse orders) was shown for the Reverse, and "M" (for memorizing orders) for the Same. Four (short) or six (long) stimuli were each sequentially presented to the participants for 600 ms, with an interstimulus interval of 200 ms, leading to 4.5 s and 6 s trials for the short and long stimuli, respectively. For fixation, a red cross was always displayed at the center of the monitor. During fMRI experiments, stimulus presentation, as well as acquisition of responses and reaction times (RTs), was controlled using the LabVIEW software and interface (National Instruments, Austin, TX).

Under the sentence conditions of Nested, Simple, and Conjoined with the same structures shown in Figure 3, the jabberwocky sentences consisting of pseudonoun and pseudoverb phrases alone were visually presented to participants in a phrase-by-phrase manner. We made six pseudonouns by repeating the same syllables with voiced consonants and any one of /a/, /u/, or /o/: *rara*, *zaza*, *mumu*, *gugu*, *yoyo*, and *dodo*. We also made four pseudoverb roots by repeating the same syllables with voiceless consonants and either /i/ or /e/: *kiki*, *hihi*, *sese*, and *tete*. The transitions between consecutive phrases or sentences were thoroughly randomized. Nonmatching stimuli included at least one odd vowel of V_i as a matching error (Tables 1 and 2). All matching and nonmatching stimuli were phonotactically allowable, but lacked lexical associations in Japanese. There were 10 conditions (Figures 4 and 5); we prepared a set of 36 sentences for each of the sentence conditions, and a set of 36 letter strings for each of string conditions. Each set consisted of 18 matching and 18 nonmatching stimuli.

We used only three kinds of grammatical particles, which represented *canonical* (i.e., in a prototypical use) case markings and syntactic information in Japanese: -*ga*, a nominative

case marker; -*no*, a genitive case marker; and -*to*, a complementizer. In all jabberwocky sentences, the distinction between Ns and Vs was clear without the need to memorize pseudowords, because Ns, but not Vs, ended with either -*ga* or -*no*, i.e., case markers in Japanese such as -*ga* and -*no* can be generally attached only to nominal phrases (e.g., "*momo-ga minoru*" and "*momo-no iro*": "*the peach ripens*" and "*the peach's color*"). Moreover, Vs took a nonpast-tense form (-*ru*), past-tense form (-*ta*), or gerundive form (-*te*), consistent with the morphosyntactic and phonological features of regular Japanese verbs (Tsujimura, 2007); Vs ended with -*to* and -*te* introduced *that*-clauses and *and*-conjunctives, respectively (see examples in Figure 3 legend). Including the first verb of a compound verb in an adverbial form (e.g., "*hihi*" and "*sese*"), all Ns and Vs with -*ga*, -*no*, -*to*, and -*te* endings (green letters in Figures 3, 4A, and 5A) were associated with Merge applications to connect multiple nouns/verbs or sentences, amounting to the "number of Merge."

In the Japanese language, all regular verbs have either an *ichidan* (one-tier) or *godan* (five-tier) conjugation (Shibatani, 1990). In the present experiments, Vs took a nonpast-tense form (-*ru*), past-tense form (-*ta*), or gerundive form (-*te*) (e.g., "*teteru*," "*teteta*," and "*tetete*"), inflecting like normal *ichidan* verbs (e.g., "*tateru*," "*tateta*," and "*tatete*": "*build*," "*built*," and "*building*"). In subordinate clauses, Vs took tense markers (-*ru*/-*ta*) just before -*to*. The last V in an entire sentence always included a tense marker. As the tense of verbs can be independently marked in a Japanese nested sentence (e.g., "*Taro-ga Hanako-ga utau-to omotta*": "*Taro thought that Hanako would sing*"), we regarded the "number of tense markers (-*ru*/-*ta*)" as another linguistic factor (Table 3).

In Japanese, there are a number of morphologically/semantically related pairs of transitive and intransitive verbs with vowel changes (e.g., "*kakeru*" and "*kakaru*": "*hang up*" and "*hang down*") (Tsujimura, 2007). Not only vowel dissimilation, but vowel assimilation, i.e., *vowel harmony*, is commonly observed in natural languages (Nevins, 2010). Vowel

harmony is possible between adjacent or distant vowels within a word (Mailhot & Reiss, 2007). Indeed, the presence of vowel harmony has been indicated in the history of the Japanese language (Shibatani, 1990). In the present study, vowel harmony was adopted to change the last, i.e., the second, vowel of the verb root, so that this vowel harmonized with the vowel (i.e., /a/, /u/, or /o/) of the corresponding subject (e.g., "*rara-ga tetaru*" and "*rara-ga tetatta*" from "*teteru*" and "*teteta*," respectively, underlined vowels within pseudowords). This property of vowel harmony made each V_i inflect like a normal *godan* verb, in which the root always ends in a consonant, and this change made V_i distinct from the original form of *ichidan* verbs. It should be noted that V_i in nonmatching stimuli also inflected like a normal *godan* verb (Tables 1 and 2). When a *godan* verb root ends in *r*, *t*, or *w*, a euphonic change to a geminated consonant *t* occurs before *-ta* or *-te* (e.g., "*kakatta*" from "*kakari-ta*": "*hung down*"). Each V_i with *-ta* or *-te* (e.g., "*tetatta*") was consistent with its adverbial form *ri* ("*tetari-ta*"), *ti* ("*tetati-ta*"), or (*w*)*i* ("*tetai-ta*"). On the other hand, the first verb of a compound verb remained an *ichidan* verb in an adverbial form (e.g., "*tete*").

A few grammatical, but noncanonical (i.e., in a special use), usages of *-ga* or *-no* exist: parallel subjects marked with *-ga* (e.g., "*Taro-ga yuujin-ga sorezore utatta*": "*Taro and his friend each sang*"), an object marked with *-ga* (e.g., "*Taro-ga yuujin-ga suki-da*": "*Taro likes his friend*"), an external possessor marked with *-ga* (e.g., "*Taro-ga yuujin-ga sinsetu-da*": "*Taro's friend is kind*"), and a subject marked with *-no* (e.g., "*Taro-no suki-na yuujin*": "*the friend Taro likes*"). Considering such canonicity, we regarded "number of case markers (*-ga*/*-no*)" as another linguistic factor (Table 3). However, these noncanonical case markings are rare in both comprehension and production, as shown by previous behavioral experiments (Uehara & Bradley, 2002; Miyamoto, 2002). We assured participants that case markings for the stimuli were always canonical (see Task instructions). Actual usage of canonical case markings was fully guaranteed by the high accuracy under the sentence conditions, as the

matching task could not be performed correctly if such noncanonical case markings were employed.

Under the string conditions, for the first and second halves of the string, stimuli were presented in the reverse order for Reverse, whereas they were presented in the same order for Same (Figures 4B and 5B). Each letter string was formed by jumbling the letters of either the N or V, which had no lexical associations. For the Reverse and Same, there was actually no path connecting the nonterminal nodes of symbol pairs (e.g., A_1 B_1 and A_2 B_2) since there was *no* Merge application to connect the multiple pairs. The letter strings lacked *-ga*, *-no*, *-to*, or *-te* endings, and their flat constructions were determined by the cue of "R" or "M" alone. We estimated the syntactic factors for the letter strings, but all estimates of these factors were null in [Reverse − Same] (see Table 4).

We imposed the three following constraints on the letter strings. First, for the first half of a string (As), letter strings derived from Ns and Vs (denoted here as $\mathcal{N}$s and $\mathcal{V}$s, respectively) were in the order of $\mathcal{NV}$ or $\mathcal{VN}$ for the short stimuli, and $\mathcal{NVN}$ or $\mathcal{VNV}$ for the long stimuli. Secondly, we avoided endings with *-ga*/*-no* for $\mathcal{N}$s, but some $\mathcal{V}$s with *-ru*/*-ta* endings were used as stimuli. Lastly, neither Vs with *-to*/*-te* endings nor Vs in the adverbial form were used for making $\mathcal{V}$s.

Examples of long sentences are shown in Figure 5A. For the Nested$_{(L)}$ condition, a sentence at the lowest hierarchical level (S) was self-embedded twice into an entire sentence (S"). For the Simple$_{(L)}$ condition, we tested both stimuli of $(VV_1)V_1$ and $V_1(VV_1)$, where two verbs, i.e., VV_1 (a compound verb) and V_1, were conjoined. Branching constructions for Ns in the Simple$_{(L)}$ were ambiguous between $(NN)N_1$ and $N(NN_1)$, like "*Japanese history teacher*" ([[*Japanese history*] *teacher*] and [*Japanese* [*history teacher*]], respectively), both of which yielded the same DoM. For Japanese relative clauses, as well as noun phrases with a genitive case marker *-no*, left-branching constructions predominate (cf. Miller & Chomsky,

1963, pp. 471–472). For the Conjoined$_{(L)}$ condition, the DoM for a node was increased by one from the top nodes of entire conjoined sentences (S') to the same hierarchical level of conjoined sentences (Ss). Since all the conjoined sentences were equivalent with respect to their status in multiple branching, the two Merge applications involved here were assumed to follow, as a marked (i.e., "exceptional") option, the associative law (i.e., $[[a \# b] \# c] = [a \# [b \# c]]$, where # represents Merge). This type of associative Merge (which yields n-ary structures by applying $n - 1$ times) has been argued to be permissible for a certain class of marked constructions in human language (Fukui, 2011). In this way, the number of Merge operations in a sentence is always one less than the number of terminal nodes, *irrespective of sentence structures*.

2.2.3. Task

For each trial of a matching task under the sentence conditions or string conditions, participants judged whether all pairs of the sample stimulus (N or A) and comparison stimulus (V or B) matched, and responded by pressing one of two buttons (right for matching, and left for nonmatching) after the last stimulus appeared (Figures 4C and 4D). The accuracy and RTs were collected until 500 ms after the last stimulus disappeared. No feedback on any trial's performance was given to any participant.

For the Nested condition, an entire sentence was constructed by nesting sentences in the form of $[N_2[N_1\ V_1]V_2]$ or $[N_3[N_2[N_1\ V_1]V_2]V_3]$, where $[N_i\ V_i]$ represents a subject–verb pair of a sentence (Figures 4A and 5A). In head-last languages, the key element (the "head") that determines the properties of a phrase appears at the end of the phrase. Because Japanese is a head-last, and hence an SOV (verb-final) language, a main verb is placed after a subordinate clause. Therefore, Japanese sentences naturally yield nested structures of $N^n\ V^n$ without having to employ, as in English, object-relative clauses (e.g., "*The boy who$_i$ we like t_i*

sings"), which require "movement" of an object (i.e., with more Merge applications) leaving behind a "trace" (t_i). For the Simple condition, a simple sentence was constructed by adding the same number of left/right branches to both Ns and Vs. The last noun (i.e., head) in the branches of Ns made a subject–verb pair with the last verb (i.e., head) of a compound verb. Each simple sentence thus took the form of [(NN_1) (VV_1)], etc. For the Conjoined condition, an entire sentence was constructed by conjoining sentences in the form of [N_1 V_1][N_2 V_2] or [N_1 V_1][N_2 V_2][N_3 V_3].

In a single run of 60 trials for the short stimuli, there were 10 trials each for the sentence conditions ($\text{Nested}_{(S)}$, $\text{Simple}_{(S)}$, and $\text{Conjoined}_{(S)}$), and 15 trials each for the string conditions ($\text{Reverse}_{(S)}$ and $\text{Same}_{(S)}$). Trials alternated between sentence and string conditions. If the sentence and string sequences were separated, the order of the Nested, Simple, and Conjoined was pseudo-randomized without repetition, and the order of the Reverse and Same was counterbalanced as Same-Reverse-Reverse-Same-… or Reverse-Same-Same-Reverse-… In a single run of 50 trials for the long stimuli, there were 10 trials each for the sentence conditions ($\text{Nested}_{(L)}$, $\text{Simple}_{(L)}$, and $\text{Conjoined}_{(L)}$) and string conditions ($\text{Reverse}_{(L)}$ and $\text{Same}_{(L)}$), in the order of string-sentence-string-sentence-sentence-string-… With a maximum of nine runs, the same sentence stimulus appeared no more than three times for each participant.

2.2.4. Task instructions

Before the experiments, all participants were fully informed about the stimuli and the task. We instructed and trained them in the sentence conditions in the order of $\text{Simple}_{(S)}$, $\text{Conjoined}_{(S)}$, $\text{Nested}_{(S)}$, $\text{Simple}_{(L)}$, $\text{Conjoined}_{(L)}$, and $\text{Nested}_{(L)}$, as the number of vowel extractions increased in this order. The following is a translation of the Japanese-language task instructions.

[*Day 1 instructions*]
Words used in a task

Special pseudowords and letter strings will be used in a task. ***You don't have to remember the following words***.

Pseudowords used as nouns (six items) are "*rara*," "*zaza*," "*mumu*," "*gugu*," "*yoyo*," and "*dodo*."

Grammatical particles attached to nouns (two items) are "*-ga*," which marks a subject, and "*-no*," which marks a modifier (e.g., "*zaza-ga*," "*mumu-no*," etc.).

Pseudowords used as verbs (four items) are "*hihi*," "*kiki*," "*sese*," and "*tete*."

Verb endings (five kinds) and verb conjugation patterns are the following:

1. Verb endings: "*-ru*" (present tense), "*-tta*" (past tense), "*-tte*" (e.g., "*utatte*"), "*-ru-to*" (e.g., "*odoru-to*"), and "*-tta-to*" (e.g., "*odotta-to*"); and
2. Verb conjugation: ***When a subject corresponds to a verb within a sentence****, the latter vowel* (second syllable) *of the verb root will change to match the vowel of the corresponding subject*. You don't have to recall the root form of the verb.

Correct examples:

"*rara-ga hiharu*" (conjugation of "*hihi*")
"*mumu-ga kikutta*" (conjugation of "*kiki*")
"*dodo-ga sesoru*" (conjugation of "*sese*")
"*rara-ga tetatta*" (conjugation of "*tete*")

Wrong examples:

"*zaza-ga hihoru*"
"*gugu-ga kikotta*"
"*yoyo-ga sesuru*"
"*zaza-ga tetoru*"

Other letter strings include "*ragara*," "*nogugu*," "*huhhita*," "*kottaki*," etc.

Types of conditions

1) Grammar (G) conditions, 2) Memory (M) condition, and 3) Reverse (R) condition.

You will perform a task under the three types of conditions during scanning of your head images inside the MRI scanner. At the beginning of each trial, a cue ("G," "M," or "R") denoting a condition will appear. By noting these cues, please judge which condition will be presented. During the task, a small red cross will remain at the center of the monitor. Focus on this red cross as much as you can, but you may blink. *Please do not speak or read aloud during the task.*

During the task, please respond by pressing a button. Hold the switch box with both hands, with the red button to your right. When you press a button, use your right thumb without looking at the buttons. Please note the following crucial points:

- Press a button only once, when necessary.
- *Please press the button as fast as you can, while the stimuli are still being presented.*
- Every stimulus consists of a cue denoting a condition and of four or six phrases or strings.
- *Do not press a button before the last stimulus*.

1. Grammar conditions

There are six patterns under the Grammar conditions. The task is to judge whether the vowel of the subject matches the last vowel of the corresponding verb root. At the beginning of each trial, the cue "G" denoting a condition will appear. Do not overlook this cue, so that you can answer correctly. For every pattern shown below, "G" will appear.

1.1 Pattern 1 (Simple$_{(S)}$)

Pattern 1 is similar to "*Taro-no ani-ga sinobi aruita*" ("*Taro's brother sneaked around*"). The subject is "*ani*," and the verb is "*aruita*." Please judge whether the vowel of the *second* phrase (subject) matches that of the *fourth* phrase (corresponding verb). Please press *the rightmost button* if correct, and press *the second one from the right* if wrong. Press the button quickly with your right thumb, while the fourth stimulus is presented.

Correct examples:

"*zaza-no yoyo-ga hihi tetoru*"

"*zaza-no yoyo-ga hihi tetotta*"

Wrong examples:

"*zaza-no yoyo-ga tete sesuru*"

"*zaza-no yoyo-ga tete sesatta*"

1.2 Pattern 2 (Conjoined$_{(S)}$)

Pattern 2 is similar to that in "*Taro-ga utatte Hanako-ga odoru*" ("*Taro sings, and Hanako dances*"). Please judge whether the vowel of the *first* phrase matches that of the *second* phrase, and the vowel of the *third* phrase matches that of the *fourth* phrase.

Correct examples:

"*zaza-ga tetatte yoyo-ga kikoru*"

"*zaza-ga tetatte yoyo-ga kikotta*"

Wrong examples:

"*zaza-ga tetatte yoyo-ga kikuru*"

"*zaza-ga tetotte yoyo-ga kikatta*"

1.3 Pattern 3 (Nested$_{(S)}$)

Pattern 3 is similar to that in "*Taro-ga Hanako-ga utau-to omotta*" ("*Taro thought that Hanako would sing*"). Please judge whether the vowel of the *first* phrase matches that of the *fourth* phrase, and the vowel of the *second* phrase matches that of the *third* phrase.

Correct examples:

"*zaza-ga yoyo-ga tetoru-to kikaru*"

"*zaza-ga yoyo-ga tetotta-to kikatta*"

Wrong examples:

"*zaza-ga yoyo-ga tetoru-to kikoru*"

"*zaza-ga yoyo-ga tetatta-to kikutta*"

Tips

1. Please do not forget the first noun.
2. Some sentences have multiple errors; *please check them all the way through the last phrase*.
3. Please press the button *as fast as you can, while the last phrase is still being presented*.
4. *When the last phrase disappears, please stop pressing the button, and concentrate on the next trial*.

2. Memory condition (Same$_{(S)}$)

The task is to memorize the presented letter strings. At the beginning of each trial, the cue "M" denoting a condition will appear. Do not overlook this cue, so that you can answer correctly.

Four strings will appear one by one on the monitor. Please memorize these four strings, and judge whether the *first* and *third* strings, and the *second* and *fourth* ones, are exactly the same.

Correct examples:

"*yogayo tarute yogayo tarute*"
"*hiruhu garara hiruhu garara*"

Wrong examples:

"*yonoyo tettata yonoyo tatetta*"
"*hihhata nogugu settaso gunogu*"

3. Reverse condition (Reverse$_{(S)}$)

The task is to memorize the presented letter strings. At the beginning of each trial, the cue "R" denoting a condition will appear. Do not overlook this cue, so that you can answer correctly.

Four strings will appear one by one on the monitor. Please memorize these four strings, and judge whether the *first* and *fourth* strings, and the *second* and *third* ones, are exactly the same.

Correct examples:

"*yogayo tarute tarute yogayo*"
"*hiruhu garara garara hiruhu*"

Wrong examples:

"*yonoyo tettata tatetta yonoyo*"
"*hihhata nogugu nogugu settaso*"

[*Day 2 instructions*]

1. Grammar conditions

1.1 Pattern 4 (Simple$_{(L)}$)

Pattern 4 is similar to that in "*Taro-no ani-no yujin-ga hasiri mawatte utatta*" ("*Taro's brother's friend run around and sang*"). The subject is "*yujin*," and the verbs are "*mawatte*" and "*utatta*." Please judge whether the vowel of the *third* phrase matches that of the *fourth* or *fifth* phrase, and the vowel of the *third* phrase matches that of the *sixth* phrase. Please press *the rightmost button* if correct, and press *the second one from the right* if wrong. Press the button quickly with your right thumb, while the sixth stimulus is being presented.

Correct examples:

"*zaza-no yoyo-no mumu-ga tete kikutte hihuru*"

"*zaza-no yoyo-no mumu-ga tetutte kiki hihutta*"

Wrong examples:

"*zaza-no yoyo-no mumu-ga tetutte kiki hiharu*"

"*zaza-no yoyo-no mumu-ga tete kikatte hihotta*"

1.2 Pattern 5 (Conjoined$_{(L)}$)

Pattern 5 is similar to that of "*Taro-ga odotte Hanako-ga utatte Jiro-ga asobu*" ("*Taro dances, Hanako sings, and Jiro plays*"). Please judge whether the vowel of the *first* phrase matches that of the *second* phrase, the vowel of the *third* phrase matches that of the *fourth* phrase, and the vowel of the *fifth* phrase matches that of the *sixth* phrase.

Correct examples:

"*zaza-ga tetatte yoyo-ga kikotte mumu-ga hihuru*"

"*zaza-ga tetatte yoyo-ga kikotte mumu-ga hihutta*"

Wrong examples:

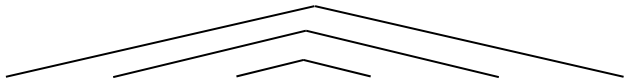
"*zaza-ga tetatte yoyo-ga kikotte mumu-ga hiharu*"

"*zaza-ga tetutte yoyo-ga kikotte mumu-ga hihatta*"

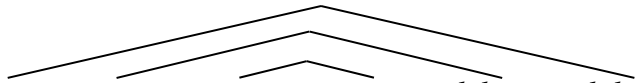
1.3 Pattern 6 (Nested$_{(L)}$)

Pattern 6 is similar to that of "*Taro-ga Hanako-ga Jiro-ga utau-to omou-to kangaeta*" ("*Taro supposed that Hanako would think that Jiro would sing*"). Please judge whether the vowel of the *first* phrase matches that of the *sixth* phrase, the vowel of the *second* phrase matches that of the *fifth* phrase, and the vowel of the *third* phrase matches that of the *fourth* phrase.

Correct examples:

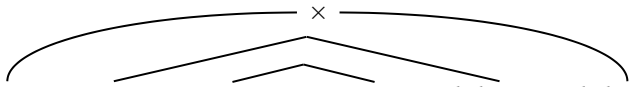


"*zaza-ga yoyo-ga mumu-ga teturu-to kikoru-to hiharu*"

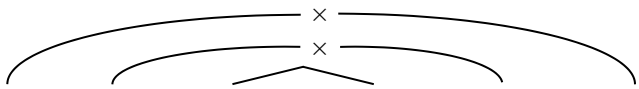


"*zaza-ga yoyo-ga mumu-ga tetutta-to kikotta-to hihatta*"

Wrong examples:



"*zaza-ga yoyo-ga mumu-ga teturu-to kikoru-to hihoru*"



"*zaza-ga yoyo-ga mumu-ga tetutta-to kikatta-to hihotta*"

2. Memory condition (Same$_{(L)}$)

Six strings will appear one by one on the monitor. Please memorize these six strings, and judge whether the *first* and *fourth* strings, the *second* and *fifth* ones, and the *third* and *sixth* ones, are exactly the same.

Correct examples:

"*yogayo teruta gagugu yogayo teruta gagugu*"
"*huruhi garara ruseso huruhi garara ruseso*"

Wrong examples:

"*noyoyo tattate gunogu noyoyo tattate munomu*"
"*huttahi ranora sottase hihhuta ranora tattate*"

3. Reverse condition (Reverse$_{(L)}$)

Six strings will appear one by one on the monitor. Please memorize these six strings, and judge whether the *first* and *sixth* strings, the *second* and *fifth* ones, and the *third* and *fourth* ones, are exactly the same.

Correct examples:

"*yogayo teruta gagugu gagugu teruta yogayo*"
"*huruhi garara ruseso ruseso garara huruhi*"

Wrong examples:

"*noyoyo tattate gunogu munomu tattate noyoyo*"
"*huttahi ranora sottase tattate ranora hihhuta*"

2.2.5. Training procedures

Experiments with short or long stimuli were performed on separate days with short stimuli presented before long. Before scanning, the participants were trained until they scored 80 % at each of the following stages. Each of the Simple, Conjoined, and Nested conditions were separately tested with self-paced reading in 10 to 30 trials with this order. Only one

participant completed 50 trials for the Nested$_{(L)}$. When these conditions were randomized, all participants took additional 20 or 40 trials. The Reverse and Same conditions were much easier to perform, and so required only 10 or 20 trials for a randomized sequence from the beginning. When all of these conditions were mixed, participants additionally took 20 or 40 trials. Finally, participants were tested on a sequence of all conditions at the regular rate of presentation (see Figures 4C and 4D) in 40 and 20 trials for the short and long stimuli, respectively. Only three participants required 60 or 80 trials for the short stimuli. Due to technical problems, four participants were retested with short or long stimuli on another day; two participants received 20 trials for the mixed conditions with self-paced reading, and all received 20–60 trials for the final stage.

2.2.6. Operational definitions of all factors examined

We operationally defined syntactic factors within an entire sentence (see Table 3) as follows. If a tree structure [a Phrase-marker (P-marker) associated with a linguistic expression] contains as its subtree a domain in which a node N immediately dominates n elements ($n > 1$), then we can say that the domain constitutes a *merged* structure. Note that under the binary Merge hypothesis, n equals 2, except for relatively rare "multiple branching" structures (see Stimuli). In the present study, we abstract away from the noun/verb vs. noun phrase/verb phrase distinction, as well as the sentence vs. complementizer phrase distinction. The operational definitions of syntactic factors examined here are as follows (see Figures 4A and 5A). "Number of Merge" is the total number of binary branches. "Number of Search" is the total number of correspondences between sample and comparison stimuli. The DoM is the largest integer m meeting the following condition: There is a continuous path passing through $m + 1$ nodes $N_0, \ldots, N_m$, where each N_i ($i \geq 1$) is *merged* in the subtree dominated by N_{i-1}.

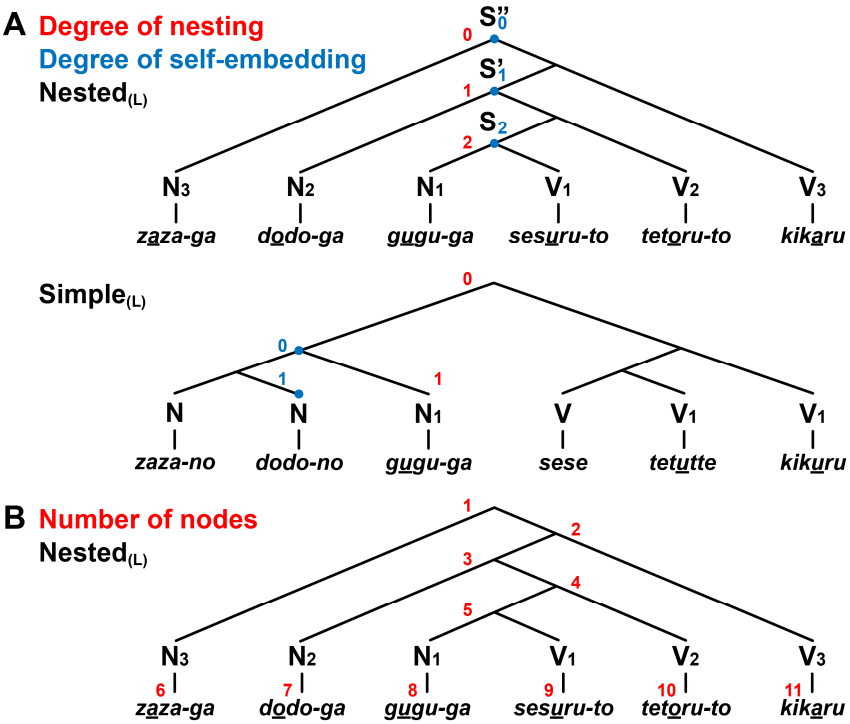


Figure 6. Application of other structure-based models to sentences with complex structures, I.
(**A**) The digits shown in red and blue denote "degree of nesting" and "degree of self-embedding," respectively. Nested and self-embedded constructions occur within sentences (Ss). Note that each shortest "zigzag path" counts one for the degree of nesting or self-embedding. For the Nested$_{(L)}$, S_1 dominates [N_2 S_2 V_2], and S_0 in turn dominates [N_3 S_1 V_3], i.e., [N_3[N_2 S_2 V_2]V_3]; the degree of nesting or self-embedding is thus two (the number of blue dots minus one). For the Simple$_{(L)}$, both of (NN)N_1 and N(NN_1) yield the same maximum degree of nesting or self-embedding for an entire sentence. (**B**) The digits shown in red denote the "number of nodes."

From both theoretical and experimental points of view, we also examined in detail the validity of other structure-based models, here categorized as "other linguistic factors." If a merged structure is surrounded by non-null elements on both sides, we get a "nested" structure. If a nesting structure occurs within the *same* type of elements, the structure is called a "self-embedded" structure. "Degree of nesting" (or "degree of self-embedding") is the largest integer m meeting the following condition (Figure 6A): There is a continuous path passing through $m + 1$ nodes $N_0, \ldots, N_m$, where each N_i ($i \geq 1$) is nested (or fully self-embedded) in the subtree dominated by N_{i-1} (Miller & Chomsky, 1963). "Number of nodes" is the total number of nonterminal nodes and terminal nodes (Figure 6B).

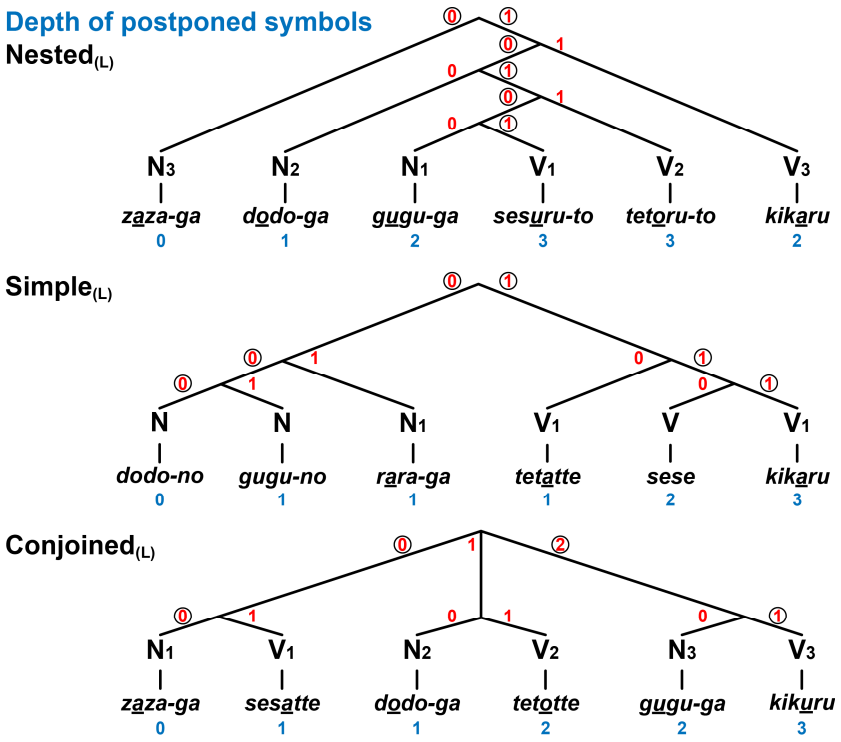


Figure 7. Application of other structure-based models to sentences with complex structures, II.
The digits shown in red and blue denote the number of branches from each node and "depth of postponed symbols," respectively (Yngve, 1960). The largest estimate can be obtained by adding together the digits shown in red with circles. For the Simple$_{(L)}$, the largest estimate of "depth of postponed symbols" is obtained, when Vs take a right-branching construction of $V_1(VV_1)$. For the Conjoined$_{(L)}$, the depth of postponed symbols is increased by two to reach the rightmost branches, when conjoining three sentences at a multiple-branching node.

As for "depth of postponed symbols," its original definition was for *producing* a given output sequence (Yngve, 1960). As we tested stimuli for *understanding* a given input sequence, we reversed the numbering to reflect the listener/reader's perspective as follows: "Depth of postponed symbols" is the amount of temporary storage needed to parse a given input sequence, which can be calculated in the following way: First, number the branches of each node from 0 to $n - 1$ (Figure 7, the digits shown in red), where n is the number of branches from that node. Start numbering from the left. Then, compute the depth d of each terminal node by adding together the numbers written along all branches leading to that terminal node, starting from the leftmost branch (i.e., the first input for the listener/reader).

According to the dependency locality theory (Gibson, 2000), two components of sentence parsing consume computational resources: "integration costs," which are the costs

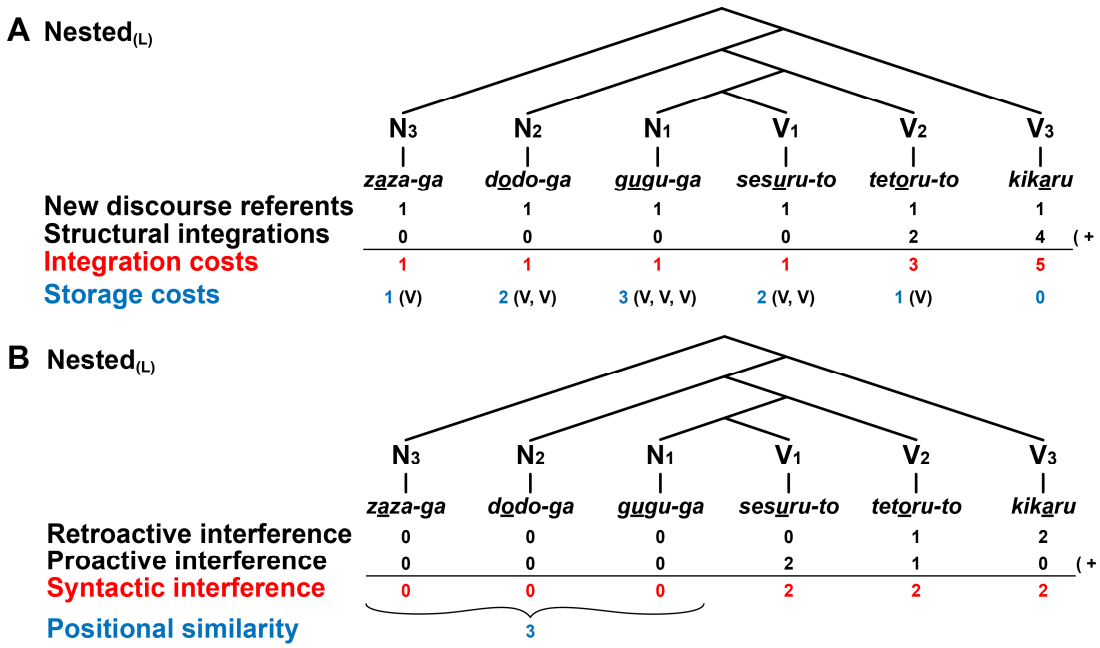


Figure 8. Application of other structure-based models to sentences with complex structures, III.
(**A**) The digits shown in red and blue denote "integration costs" and "storage costs," respectively (Gibson, 2000). Integration costs are estimated for every stimulus by adding together "new discourse referents" and "structural integrations." For example, at V_2 of the Nested$_{(L)}$, N_1 and V_1 intervene while making [N_2[N_1 V_1]V_2] (structural integrations = 2), and one verb completes the input with *-to* or *-te* (storage cost = 1). Note that the estimate of maximum structural integrations in a sentence matches with that of memory span in the present paradigm. (**B**) The digits shown in red and blue denote "syntactic interference" and "positional similarity," respectively (Lewis & Nakayama, 2002). Syntactic interference is estimated at every stimulus by adding together "retroactive interference" and "proactive interference." For example, at V_2 of the Nested$_{(L)}$, the attachment of V_2 to N_2 suffers from one unit of retroactive interference from N_1, and from one unit of proactive interference from N_3 (syntactic interference = 2). There are three adjacent nominative noun phrases in this sentence (positional similarity = 3).

of connecting words in the structure for the input thus far, and "storage costs," which are the minimum number of words required to complete the current input as a grammatical sentence (Figure 8A). Integration costs are the sum of "new discourse referents" (Ns and Vs in the present paradigm) and "structural integrations" (the number of discourse referents in the intervening region). According to similarity-based interference theory (Lewis & Nakayama, 2002), processing costs are the combined effects on syntactic attachments of "retroactive interference" (the number of nominative noun phrases between the subject–verb pair when a verb is processed) and "proactive interference" (the number of nominative noun phrases preceding the subject and still active in the parse) (Figure 8B). By adding both interference

effects together, “syntactic interference” is determined. Another source of interference is “positional similarity” which is the number of adjacent syntactically similar noun phrases (i.e., marked with similar case markers).

Nonlinguistic factors may also variably contribute to the processing load of sentences with different constructions. At least three basic nonlinguistic factors may be involved in the present experiment: memory span, counting, and “number of encoding.” The operational definitions of these nonlinguistic factors are as follows. Memory span is the maximum cost needed to maintain an item for matching against intervening or skipped stimuli (e.g., zero for N_1 V_1 in the Conjoined$_{(S)}$, and one for N_1 VV_1 in the Simple$_{(S)}$), and its operational definition is the maximum number of cusps in the curved arrows (Figures 4 and 5). Counting is an operation needed to track symbol repetition, and its operational definition is the maximum number of consecutively repeated symbols (e.g., zero for NVNV or NVNVNV, one for NNVV, and two for NNNVVV). Encoding is the memorization of features necessary for matching, and “number of encoding” is the total number of sample and comparison stimuli. Memory span and counting were considered temporal order-related factors in the present experiment; memory span was related to matching orders, while counting was related to symbol orders. On the other hand, memory span and “number of encoding” were memory-related factors.

2.2.7. MRI data acquisition

Depending on the time of the experiments, the fMRI scans were conducted on a 1.5 T scanner (Stratis II, Premium; Hitachi Medical Corporation, Tokyo, Japan) with a bird-cage head coil, and the DTI scans were conducted on a 3.0 T scanner (Signa HDxt; GE Healthcare, Milwaukee, WI) with an 8-channel phased-array head coil. For the fMRI, we scanned 26 axial slices that were 3-mm thick with a 1-mm gap, covering from $z = -40$ to 63 mm from the

anterior to posterior commissure (AC-PC) line, with a gradient-echo echo-planar imaging (EPI) sequence [repetition time (TR) = 3 s, echo time (TE) = 51 ms, flip angle (FA) = 90°, field of view (FOV) = 192 × 192 mm^2, resolution = 3 × 3 mm^2]. In a single scanning run, we obtained 92 volumes for the short stimuli and 101 volumes for the long stimuli following three dummy images, which allowed for the rise of the MR signals. For each participant, five to nine runs for each of the short and long stimuli were tested, and four to nine runs without head movement were used for analyses. After completion of the fMRI session, high-resolution T1-weighted images of the whole brain (145 axial slices, 1 × 1 × 1 mm^3) were acquired from all participants with a radio-frequency-spoiled steady-state acquisition with a rewound gradient echo sequence (TR = 30 ms, TE = 8 ms, FA = 60°, FOV = 256 × 256 mm^2).

For the DTI, we scanned 50 axial slices that were 3-mm thick without a gap, covering from $z = -60$ to 90 mm from the AC–PC line, with a diffusion-weighted spin-echo EPI sequence (b-value = 1,000 s/mm^2, TR = 15 s, TE = 87 ms, FOV = 256 × 256 mm^2, resolution = 2 × 2 mm^2, number of excitations = 2). A single image without diffusion-weighting (b0) was initially acquired, and then diffusion-weighting was isotropically distributed along 60 diffusion-encoding gradient directions. After completion of the DTI sessions, high-resolution T1-weighted images of the whole brain (192 axial slices, 1 × 1 × 1 mm^3) were acquired from all participants with a fast-spoiled gradient recalled acquisition in the steady state sequence (TR = 10 ms, TE = 4 ms, FA = 25°, FOV = 256 × 256 mm^2).

2.2.8. fMRI data analyses

Data analyses of fMRI were performed in a standard manner using SPM5 statistical parametric mapping software (Wellcome Trust Centre for Neuroimaging, London, UK; http://www.fil.ion.ucl.ac.uk/spm/) (Friston et al., 1995), implemented on MATLAB software (MathWorks, Natick, MA). The acquisition timing of each slice was corrected using the

middle slice (the thirteenth slice chronologically) as a reference for the EPI data. We realigned the EPI data to the first volume in each run, and removed runs that included data with a translation of >2 mm in any of the three directions and with a rotation of >1.4° around any of the three axes; these thresholds of head movement were empirically determined from the previous studies (Hashimoto & Sakai, 2002; Suzuki & Sakai, 2003; Kinno et al., 2008).

Each participant's T1-weighted structural image was co-registered to the mean functional image generated during realignment. The co-registered structural image was spatially normalized to the standard brain space as defined by the Montreal Neurological Institute (MNI) using the "unified segmentation" algorithm with medium regularization, which is a generative model that combines tissue segmentation, bias correction, and spatial normalization in the inversion of a single unified model (Ashburner & Friston, 2005). All of the normalized structural images were visually inspected and compared with the standard brain for the absence of any further deformation. After spatial normalization, the resultant deformation field was applied to the realigned functional imaging data in each run, which was resampled every 3 mm using seventh-degree B-spline interpolation. All normalized functional images were then smoothed by using an isotropic Gaussian kernel of 9 mm full-width at half maximum (FWHM). Low-frequency noise was removed by high-pass filtering at 1/128 Hz.

In a first-level analysis (i.e., fixed-effects analysis), each participant's hemodynamic responses induced by the trials were modeled with a boxcar function with a duration of 3.5 s (short stimuli) or 5.1 s (long stimuli), i.e., from the onset of the first stimulus (N or A) to 500 ms after the disappearance of the last stimulus (Figures 4C and 4D), and the boxcar function was convolved with a hemodynamic response function. These functions were used as run-specific covariates for matching or nonmatching stimuli for each condition in a general linear model (GLM). Only event-related responses to correct trials were analyzed.

For a second-level analysis (i.e., random effects analysis) using either an analysis of covariance (ANCOVA) with *t*-statistics or a one-sample *t*-test, contrast images were generated for each participant and used for intersubject comparisons. For all fMRI data analyses, the statistical threshold was set to $P < 0.05$ for the voxel level, corrected for multiple comparisons [family-wise error (FWE) correction] across the whole brain. To discount any general cognitive factors related to task difficulty from the evaluation of cortical activation, accuracy was used as a nuisance variable for each contrast of random effects analyses. Note that accuracy was more sensitive than RTs in the matching task (see Results, Condition and length effects on the accuracy/RTs). If a subtraction between conditions served as a reference, an exclusive mask (uncorrected $P < 0.01$) was applied to reduce the contribution of deactivation. For example, in (Nested − Conjoined) > (Simple − Conjoined), an exclusive mask of the [Conjoined − Simple] contrast was applied to reduce the contribution of deactivation in [Simple − Conjoined].

We used a factorial design for a two-way ANCOVA with condition [Nested − Conjoined, Simple − Conjoined] × length [Long, Short], where activations in [Nested − Conjoined] or [Simple − Conjoined] for long or short sentences were estimated for each participant, and then the main contrast (as denoted by a greater-than symbol) was calculated for intersubject comparisons. A one-way ANCOVA was used for a direct comparison of ($\text{Nested}_{(L)} - \text{Conjoined}_{(L)}$) > ($\text{Simple}_{(S)} - \text{Conjoined}_{(S)}$), and for (Nested − Simple) > (Reverse − Same) contrasting the sentence and string conditions. To ensure an independent statistical test in the latter contrast, the inclusive mask of (Nested − Conjoined) > (Simple − Conjoined) (corrected $P < 0.05$) was iteratively applied to the contrast image of each participant, thereby leaving out that participant from the calculation of the mask, i.e., the "leave-one-subject-out cross-validation" approach (Esterman et al., 2010). One-sample *t*-tests were used for

[Nonmatching − Matching] separately under the sentence and string conditions, and for [Reverse − Same].

For the anatomical identification of activated regions, we basically used the Anatomical Automatic Labeling method (Tzourio-Mazoyer et al., 2002). In region of interest (ROI) analyses, the percent signal changes averaged across the voxels in each activated region were extracted using the MarsBaR-toolbox (http://marsbar.sourceforge.net/). To fit a model of each factor to activations, we calculated a fitting scale and residual sum of squares (RSS) using MATLAB; we obtained the fitted values by multiplying the estimates by the fitting scale. For a no-intercept model, the coefficient of determination (r^2) should be calculated as $r^2 = 1 - \Sigma(y - \hat{y})^2/\Sigma y^2$, where $\hat{y}$ and y denote the fitted values and the signal changes for each contrast, respectively (Kvålseth, 1985). For this calculation, we used R software (http://www.r-project.org/). By using a restricted maximum-likelihood method, we further fitted "linear mixed-effects models" with individual activations as dependent variables, the estimates of each factor as a regressor, and the participants as random effects. For this calculation, we used an nlme (linear and nonlinear mixed-effects models) package (http://cran.r-project.org/web/packages/nlme/) on R software.

2.2.9. DCM data analyses

Data analyses of DCM were performed using DCM10 on SPM 8 (Friston, Harrison, & Penny, 2003). We concatenated the scans from the separate runs, and reanalyzed the preprocessed data with GLM, which contained regressors representing the Nested, Simple, Reverse, and Same conditions for making a meaningful contrast (correct trials alone; see Table 4), as well as a regressor representing all conditions (correct trials alone including those for the Conjoined condition) for driving inputs. The regressor representing the Nested

condition was also used for a modulatory effect. In addition, the effects of transition between runs were taken into account by adding regressors for each run.

We set the following functional and anatomical constraints for selecting ROIs of the L. F3op/F3t or L. SMG for each participant with individually preprocessed data (Stephan et al., 2007). The individual local maxima should be significant in (Nested − Simple) > (Reverse − Same) estimated for each participant (uncorrected $P < 0.05$); they should be nearest to the group local maxima of (Nested − Simple) > (Reverse − Same) within twice the FWHM of the smoothing kernel; and they should locate to the L. F3op/F3t or L. SMG. The averaged MNI coordinates of these individual local maxima were (−50, 26, 25) and (−39, −45, 43) for the L. F3op/F3t and L. SMG, respectively. With the volume-of-interest (VOI) tool in SPM8, the time series was extracted by taking the first eigenvariate across all suprathreshold voxels within 6 mm of the individual local maxima, defined as a ROI in this case.

We specified nine models with systematic variations in a modulatory effect and driving inputs (Figure 9). After estimating all models for each participant, we identified the most likely model by using random-effects Bayesian model selection (BMS) on DCM10. Inferences from BMS can be based on the expected probability, i.e., the expected likelihood of obtaining the model for any randomly selected participants, or on the exceedance probability, i.e., the probability that the model is a better fit to the data than any other models tested. We adopted the family inference method (Penny et al., 2010), in which three modulatory families were compared first, and then the input models within the winning family were further compared. After determining the best model, the parameter estimates of this particular model were evaluated by a one-sample *t*-test (Stephan et al., 2010).

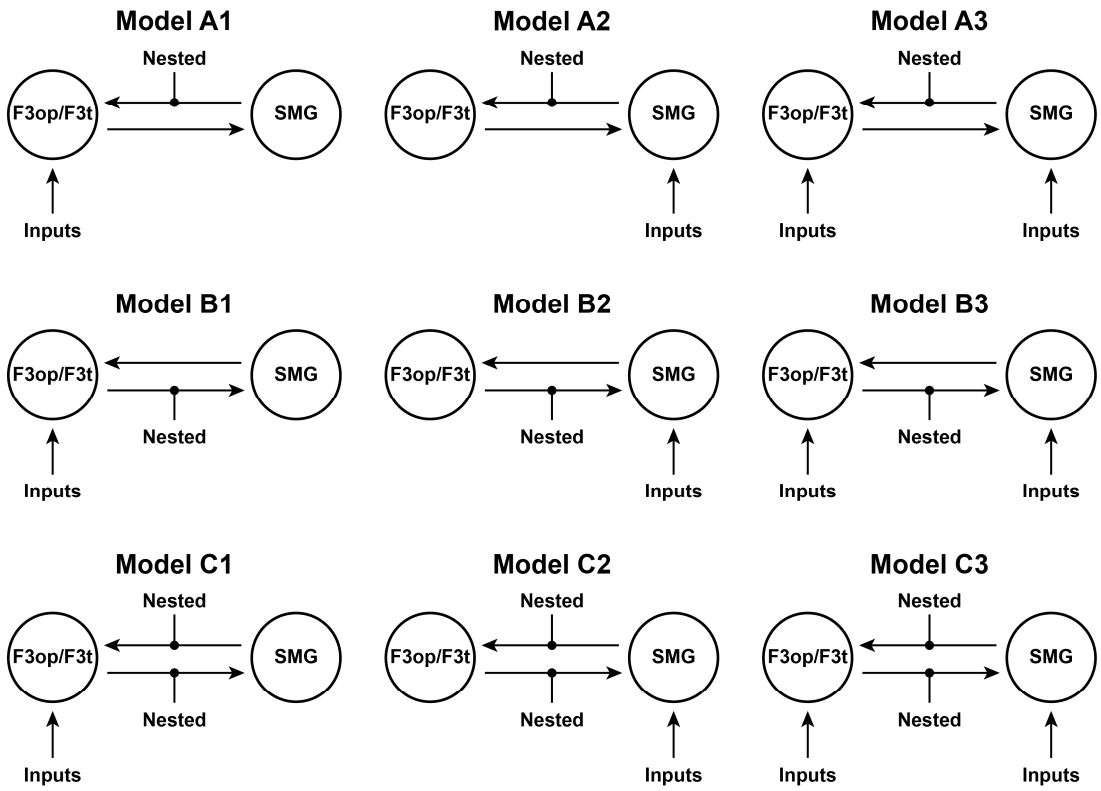


Figure 9. The DCM models tested.
We assumed bidirectional connectivity between the L. F3op/F3t and L. SMG. The models were grouped into three modulatory families based on the modulations of the connections under the Nested condition: Family *A* (*A1–A3*), in which the connection from the L. SMG to the L. F3op/F3t was modulated, Family *B* (*B1–B3*), in which the connection from the L. F3op/F3t to the L. SMG was modulated, and Family *C* (*C1–C3*), in which both connections were modulated. Each family was composed of three "input models" with regard to the regions receiving driving inputs.

2.2.10. DTI data analyses

Data analyses of DTI were performed using FSL [Oxford Centre for Functional MRI of the Brain's (FMRIB) Software Library 4.1.7; http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/] with FDT (FMRIB's Diffusion Toolbox 2.0) (Smith et al., 2004). Diffusion-weighted images were first resliced to isotropic voxels of 1 mm^3, and then eddy current distortions and motion artifacts were corrected using affine registration to the b0 image. We then extracted the brain shape from the b0 image and created the binary mask image (i.e., zero for the outside of the brain) for each participant. Markov Chain Monte Carlo sampling was performed to build up distributions on diffusion parameters at each voxel, which allowed for an estimation of the most probable pathway by Bayesian estimation (number of fibers modeled per voxel = 2)

(Behrens et al., 2007). The implicit modeling of noise in a probabilistic model made it possible to track the fibers near the gray matter.

By using FLIRT (FMRIB's Linear Image Registration Tool) on FSL, the b0 image was first coregistered to the individual T1-weighted image for each participant, and the T1-weighted image was spatially normalized to the MNI space by using both affine and nonlinear transformations with FLIRT and FNIRT (FMRIB's Nonlinear Image Registration Tool). With the transformation matrices and estimated deformation fields, the peak MNI coordinates of each region were transformed back to the individual b0 images, and a sphere of 6-mm radius centered at the transformed coordinates was defined as a seed mask for the probabilistic tractography. All fiber tracking was conducted in an individual DTI space. To find the connections between two ROIs, we set two seed masks and retained only those tracts that passed through both seed masks. Probabilistic fiber tracking was initiated from all voxels within the seed masks to generate 10,000 streamline samples, with a 0.5-mm step length, a maximum number of steps of 2,000, a curvature threshold of 0.2 (±78.5°), and a loopcheck option.

In the connectivity distributions obtained, each voxel value represented the total number of the streamline samples passing through that voxel. The connectivity probability maps were then created for each participant by dividing the connectivity distributions with a sum of the waytotal values, i.e., the total number of generated tracts from one seed mask that reached the other seed mask. This normalization approach allowed for a comparison of the connectivity probability values across participants; note that the pattern of connectivity did not change by this scaling.

To remove any spurious connections, the pathways in individual participants were thresholded to include only voxels that had at least 1% connectivity probability values (Flöel et al., 2009). The thresholded pathways in each participant were then spatially normalized as

above and binarized using “fslmaths” on FSL. The binarized pathways were overlaid across participants to produce a population probability map for each pathway, in which the voxel values represent the number of participants with a pathway through that voxel. The population probability map with thresholding (at least seven out of 15 participants) was smoothed and presented using MRIcroN software (http://www.mccauslandcenter.sc.edu/mricro/mricron/).

2.3. Results

2.3.1. Condition and length effects on the accuracy/RTs

Accuracy data, as well as RTs measured from the onset of the last stimulus, are shown in Figure 10. The high accuracy under both sentence and string conditions indicated the participants’ reliable and consistent judgments on the matching task. A two-way repeated-measures analysis of variance (rANOVA) with the condition [Nested, Simple, Conjoined, Reverse, Simple] × length [Long, Short] for the accuracy showed a significant main effect of condition [$F(4, 68) = 15$, $P < 0.0001$] and an interaction of condition by length [$F(4, 68) = 12$, $P < 0.0001$], but a main effect of length was not significant [$F(4, 68) = 3.8$, $P = 0.07$]. The RTs also showed a significant main effect of condition [$F(4, 68) = 43$, $P < 0.0001$] and an interaction of condition by length [$F(4, 68) = 13$, $P < 0.0001$], but a main effect of length was not significant [$F(4, 68) = 1.1$, $P = 0.30$]. Post-hoc paired *t*-tests among all conditions (significance level at $\alpha = 0.005$, Bonferroni corrected) showed that the accuracy for the Nested was significantly lower than that under the other conditions including the Reverse ($P < 0.0001$). This result indicates that the Nested was the most demanding condition, which cannot be explained by the *nonlinguistic* factors we examined (cf. the same estimates for the Nested and Reverse conditions in Table 4, as well as its notes). On the other hand, post-hoc paired *t*-tests showed that the RTs under each sentence condition were significantly longer

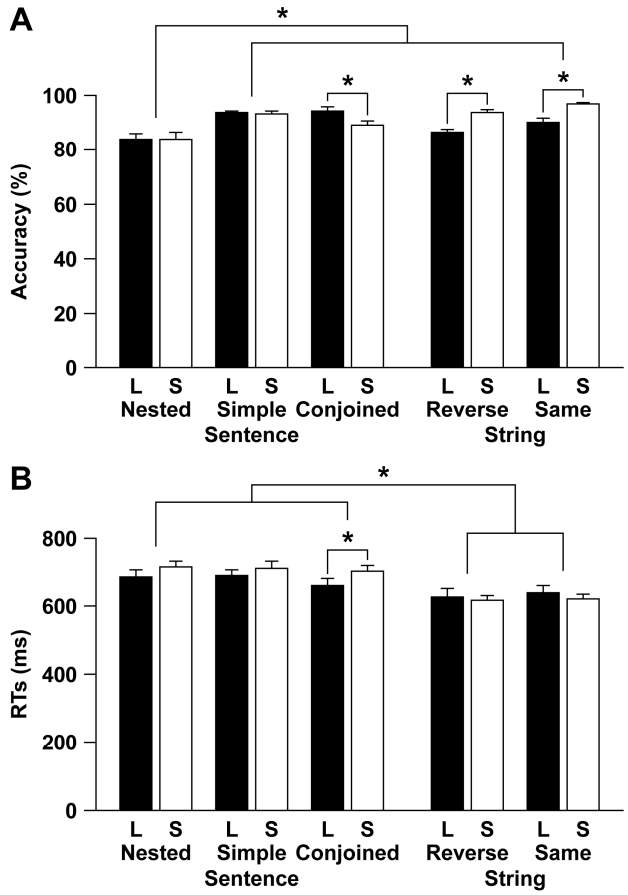


Figure 10. Condition and length effects on the accuracy/RTs.
(**A**) The accuracy (mean ± SEM) for long (L) and short (S) stimuli, denoted by filled and open bars, respectively. Asterisks indicate the significance level at corrected $P < 0.05$ (paired *t*-tests). (**B**) RTs from the onset of the last stimulus.

than those under each string condition ($P < 0.0001$). This difficulty was not in the task itself, but in vowel extraction; the sentence conditions, but not the string conditions, involved vowel extraction from the second syllable of V_i presented in hiragana, especially for the last V_i that were directly linked with RTs (Figures 4C and 4D). The load for vowel extraction would also be larger for the *short* stimuli, as we tested the short and long stimuli on separate days in the order short, then long. Indeed, the accuracy for the Conjoined$_{(S)}$ was significantly lower than that for the Conjoined$_{(L)}$ [$t(17) = 3.1$, $P = 0.006$] (significance level at $\alpha = 0.01$, Bonferroni corrected), and the RTs for the Conjoined$_{(S)}$ were significantly longer than those for the Conjoined$_{(L)}$ [$t(17) = 2.8$, $P = 0.01$], probably reflecting associated effects for novices. For the Conjoined, length effects were apparently absent, and the estimates of both memory span and

counting, which were associated with length effects, were indeed null for the Conjoined alone (Table 3). In the present study, we mainly analyzed activations that would show length effects (i.e., Long > Short), excluding the involvement of vowel extraction or effects for novices. Moreover, we used the Conjoined condition, which showed such effects most strongly, as a reference for both the Nested and Simple conditions. Therefore, we can safely conclude that any elicited effects did not directly relate to the task.

Under the string conditions, the accuracy for the long stimuli was significantly lower than that for the short stimuli ($P < 0.001$), indicating length effects. For the Nested and Simple conditions, in contrast, the effects for novices and length would have been canceled out, as neither the accuracy nor RTs differed significantly between the short and long stimuli ($P > 0.05$). Under the string conditions, the accuracy was more sensitive than the RTs.

2.3.2. Functional evidence of syntactic computation in the language areas

We examined brain activation under the sentence conditions, focusing in particular on selective activations for the most-demanding Nested condition. In a two-way analysis of covariance (ANCOVA) with the condition [Nested − Conjoined, Simple − Conjoined] × length [Long, Short], the main effect of condition, i.e., (Nested − Conjoined) > (Simple − Conjoined) while combining Long and Short, resulted in left-dominant activation, especially in the L. F3op/F3t, left lateral premotor cortex and F3op (L. LPMC/F3op), and L. SMG (Figure 11A and Table 5). Other significantly activated regions were the right (R) F3op/F3t, R. LPMC, anterior cingulate cortex (ACC), and R. SMG. The main effect of length, i.e., Long > Short while combining [Nested − Conjoined] and [Simple − Conjoined], also showed significant activations in the same regions, while there were more significant voxels in the right hemisphere (Figure 11B). Therefore, length effects alone could not account for the

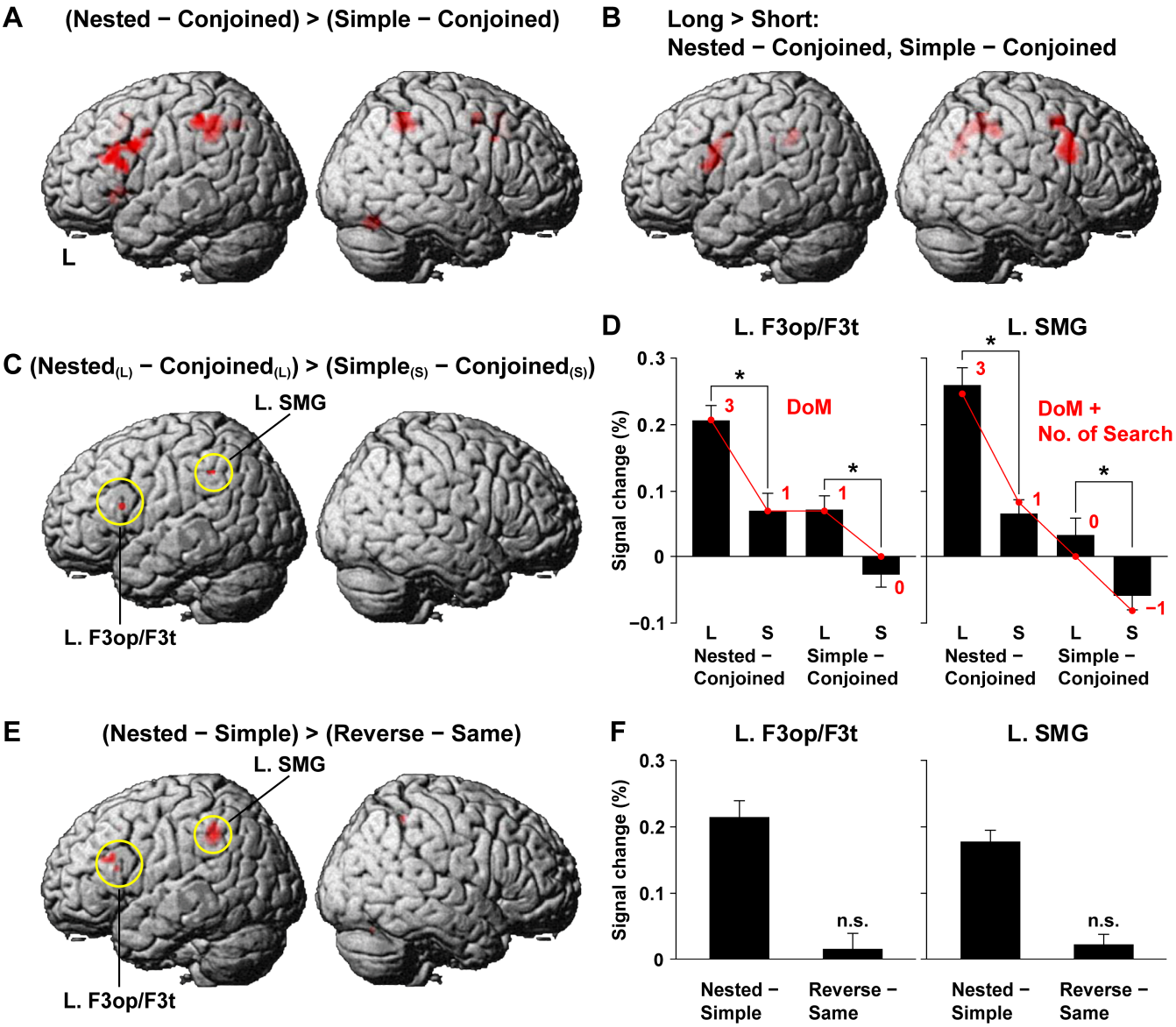


Figure 11. Functional evidence of syntactic computation in language areas.
For (**A**) and (**B**), we used a two-way ANCOVA with condition × length; for (**C**) and (**E**), a one-way ANCOVA was used. Activations were projected onto the left (L) and right lateral surfaces of a standard brain. See Tables 5 and 6 for their stereotactic coordinates. (**A**) Regions identified by the main effect of condition, i.e., (Nested − Conjoined) > (Simple − Conjoined). (**B**) Regions identified by the main effect of length, i.e., Long > Short while combining [Nested − Conjoined] and [Simple − Conjoined]. (**C**) Regions identified by (Nested$_{(L)}$ − Conjoined$_{(L)}$) > (Simple$_{(S)}$ − Conjoined$_{(S)}$), which reflected both main effects. (**D**) Percent signal changes for [Nested − Conjoined] and [Simple − Conjoined], averaged across the L. F3op/F3t and L. SMG in (C) (mean ± SEM). Overlaid red dots and lines denote the values fitted with the estimates (digits in red) for the best models: the DoM for the L. F3op/F3t and the "DoM + number of Search" for the L. SMG. (**E**) Regions identified by (Nested − Simple) > (Reverse − Same). (**F**) Percent signal changes for [Nested − Simple] and [Reverse − Same], averaged across the L. F3op/F3t and L. SMG in (E).
Symbols used: L, long sentences; S, short sentences; n.s., not significant.

consistent activation in these regions. An interaction of condition by length did not show any significant activation.

Table 5. Regions related to the sentence conditions.

Contrast	Brain region	BA	Side	*x*	*y*	*z*	*Z* Value	Voxels
Main effect of condition, (Nested − Conjoined) > (Simple − Conjoined)	F3op/F3t	44/45	L	−51	27	24	5.6	109
	LPMC/F3op	6/44	L	−48	9	30	5.4	*
	F3op/F3t	44/45	R	54	15	36	4.9	2
	LPMC	6	R	33	3	51	5.3	12
	Insula	—	L	−30	24	−3	5.7	20
	ACC	6/8/32	M	−3	18	48	6.1	45
	SMG	40	L	−54	−33	48	5.3	101
				−39	−42	39	5.9	*
			R	42	−48	54	5.3	64
	AG/SMG	39/40	L	−30	−60	48	4.9	11
	Cerebellum, lobule VI	—	R	27	−69	−21	5.6	26
Main effect of length, Long > Short: Nested − Conjoined, Simple − Conjoined	F3op/F3t	44/45	L	−48	12	18	5.9	63
	LPMC/F3op	6/44	L	−48	3	39	4.7	3
			R	48	6	30	6.0	129
	F3op/F3t	44/45	R	54	12	30	5.8	*
	LPMC	6	R	30	0	48	5.9	*
	ACC	6/8/32	M	0	27	39	4.9	9
	SMG	40	L	−57	−30	36	4.7	1
				−36	−45	39	5.3	26
			R	42	−42	42	5.5	116
	AG/SMG	39/40	R	33	−63	27	5.4	*

Stereotactic coordinates (*x*, *y*, *z*) in the Montreal Neurological Institute (MNI) space (mm) are shown for each activation peak of *Z* values (corrected $P < 0.05$). BA, Brodmann's area; L, left hemisphere; R, right hemisphere; M, medial; F3op/F3t, pars opercularis and pars triangularis of the inferior frontal gyrus; LPMC, lateral premotor cortex; ACC, anterior cingulate cortex; SMG, supramarginal gyrus; AG, angular gyrus. The region with an asterisk is included within the same cluster shown one row above.

To further narrow down candidate regions, we tested ($\text{Nested}_{(L)} - \text{Conjoined}_{(L)}$) > ($\text{Simple}_{(S)} - \text{Conjoined}_{(S)}$), which reflected both main effects, and found significant activation in the L. F3op/F3t and L. SMG (Figure 11C and Table 6). The data used for selecting these regions and those for subsequent analyses were not independent, which might cause a selection bias (Kriegeskorte et al., 2009). Among the four contrasts, however, [$\text{Nested}_{(L)} - \text{Conjoined}_{(L)}$] and [$\text{Simple}_{(S)} - \text{Conjoined}_{(S)}$] yielded two extremes of the estimates of most factors, without apparent bias among the factors (see Table 3). In addition to both main effects, the percent signal changes in the L. F3op/F3t and L. SMG (Figure 11D), averaged across significant voxels, showed a significant length effect within each of the [Nested − Conjoined] and [Simple − Conjoined] contrasts (paired *t*-test, $P < 0.02$; significance level at $\alpha = 0.025$, Bonferroni corrected). For these two critical regions, we examined the percent signal changes under the Nested and Simple conditions by subtracting those under the Conjoined condition, which had the simplest tree structures (Figure 4 and Table 3), separately for long and short sentences. Since we used appropriate references of the $\text{Conjoined}_{(L)}$ and $\text{Conjoined}_{(S)}$, we examined whether likewise *subtracted* estimates of each factor (e.g., DoM for [$\text{Nested}_{(L)} - \text{Conjoined}_{(L)}$]; see Table 3) directly explained the parametric modulation of activations in the four contrasting pairs [$\text{Nested}_{(L)} - \text{Conjoined}_{(L)}$], [$\text{Nested}_{(S)} - \text{Conjoined}_{(S)}$], [$\text{Simple}_{(L)} - \text{Conjoined}_{(L)}$], and [$\text{Simple}_{(S)} - \text{Conjoined}_{(S)}$]. The percent signal changes in the L. F3op/F3t and L. SMG indeed correlated exactly in a step-wise manner to the parametric models of the DoM [3, 1, 1, 0] and "DoM + number of Search" [3, 1, 0, −1], respectively.

We further examined 19 models proposed in theoretical linguistics, psycholinguistics, and natural language processing to verify that the models of the DoM and "DoM + number of Search" best explained the cortical activations in the L. F3op/F3t and L. SMG. All contrasts of [$\text{Nested}_{(L)} - \text{Conjoined}_{(L)}$], etc. predicted that the activations should be exactly zero when a factor produced no effect or load relative to the Conjoined. We thus

Table 6. Regions related to the sentence conditions and/or string conditions.

Contrast	Brain region	BA	Side	*x*	*y*	*z*	*Z* Value	Voxels
(Nested$_{(L)}$ − Conjoined$_{(L)}$) > (Simple$_{(S)}$ − Conjoined$_{(S)}$)	F3op/F3t	44/45	L	−45	18	18	4.8	1
	SMG	40	L	−42	−45	42	4.8	2
(Nested − Simple) > (Reverse − Same)	F3op/F3t	44/45	L	−51	24	24	5.8	5
	ACC	6/8/32	M	−3	18	45	5.2	1
	SMG	40	L	−39	−45	42	5.7	27
			R	39	−48	54	4.9	2
	Cerebellum, lobule VI	—	R	27	−69	−24	4.9	1
Nonmatching − Matching: Sentence	F3op/F3t	44/45	R	54	18	30	5.2	14
	LPMC/F3op	6/44	L	−45	9	30	4.8	1
	ACC	6/8/32	M	6	27	42	6.9	52
Nonmatching − Matching: String	F3op/F3t	44/45	R	54	18	30	5.3	21
			R	39	18	33	4.7	1
	SMG	40	R	42	−30	48	5.0	2
Reverse − Same	LPMC	6	R	27	−9	51	4.7	1

BA, Brodmann's area; L, left hemisphere; R, right hemisphere; M, medial; F3op/F3t, pars opercularis and pars triangularis of the inferior frontal gyrus; LPMC, lateral premotor cortex; ACC, anterior cingulate cortex; SMG, supramarginal gyrus; AG, angular gyrus.

adopted a no-intercept model, in which the percent signal changes in each region were fitted with a single (thus minimal) scale parameter to a model of each factor using its subtracted estimates. For the four contrasts, a least-squares method was used to minimize the RSS for the four fitted values (i.e., four estimates multiplied by a fitting scale) against the

corresponding signal changes averaged across participants (Tables 7 and 8). Among a number of parametric models tested, the model of the DoM for the L. F3op/F3t, as well as that of the "DoM + number of Search" for the L. SMG, indeed produced by far the least RSS (≤ 0.0020) and largest r^2 (≥ 0.97). Goodness of fit was further evaluated for each model by using a one-sample *t*-test (significance level at $\alpha = 0.0125$, Bonferroni corrected) between the fitted value for each contrast and individual activations. The model of the DoM for the L. F3op/F3t, as well as that of the "DoM + number of Search" for the L. SMG, produced no significant deviation for the four contrasts (one-sample *t*-test, $P \geq 0.17$). For the L. SMG, the second-best model was the DoM (RSS = 0.0063, $r^2 = 0.92$, and its smallest $P = 0.013$ was marginal). To further take account of interindividual variability, we fitted "linear mixed-effects models" with individual activations (Tables 7 and 8), and found that the models of the DoM and "DoM + number of Search" were by far more likely for the L. F3op/F3t and L. SMG, respectively.

Even if we took the Simple condition as a reference for subtracted estimates, we obtained the same results as the best models (RSS ≤ 0.0017, $r^2 \geq 0.96$). Moreover, these best models produced no significant deviation for the four contrasts (e.g., [$\text{Nested}_{(L)} - \text{Simple}_{(L)}$]) ($P \geq 0.17$). The linear mixed-effects models also showed that these models were more likely than other models tested (likelihood ratio ≤ 0.19). These results demonstrated that the DoM, as well as the number of Search, was indeed a key syntactic factor that accounted for the syntax-selective activation of the L. F3op/F3t and L. SMG.

Next, we examined whether the selective activation in these regions was replicated even after controlling both matching and symbol orders (e.g., $N_2\ N_1\ V_1\ V_2$ and $A_2\ A_1\ B_1\ B_2$) between the Nested and Reverse, i.e., in (Nested − Simple) > (Reverse − Same) combining the short and long stimuli, which further controlled various linguistic and nonlinguistic factors, such as the number of Merge, number of case markers, number of nodes, memory

Table 7. Fittings and likelihood of various models tested for the L. F3op/F3t.

Syntactic factors	RSS	r^2	*P* values for four contrasts	Log-likelihood	Likelihood ratio
*DoM	0.0007	0.99	0.17, 0.92, 0.97, 0.99	65.0	1.0
DoM + No. of Search	0.0065	0.88	0.0035, 0.064, 0.63, 0.88	59.2	3.1×10^{-3}
No. of Search	0.052	<0.1	<0.0001, 0.018, 0.019, 0.031	33.4	2.0×10^{-14}
No. of Merge	0.053	0	<0.0001, 0.0035, 0.018, 0.17	n/a	n/a
Other linguistic factors					
No. of case markers (*-ga*/*-no*)	0.053	0	<0.0001, 0.0035, 0.018, 0.17	n/a	n/a
No. of tense markers (*-ru*/*-ta*)	0.0067	0.87	0.0035, 0.17, 0.32, 0.56	59.7	4.8×10^{-3}
Degree of nesting	0.010	0.80	0.0035, 0.018, 0.17, >0.99	57.1	3.7×10^{-4}
Degree of self-embedding	0.015	0.71	0.0035, 0.0075, 0.019, 0.17	53.3	8.7×10^{-6}
No. of nodes	0.015	0.72	0.0050, 0.0082, 0.018, 0.17	53.7	1.2×10^{-5}
Depth of postponed symbols	0.053	0	<0.0001, 0.0035, 0.018, 0.17	n/a	n/a
Integration costs	0.0066	0.88	0.0017, 0,15, 0.48, 0.53	59.0	2.5×10^{-3}
Storage costs	0.014	0.74	<0.0001, 0.024, 0.83, 0.85	53.8	1.3×10^{-5}
Syntactic interference	0.0067	0.87	0.0035, 0.17, 0.32, 0.56	59.7	4.8×10^{-3}
Positional similarity	0.0055	0.90	0.051, 0.12, 0.17, 0.19	60.1	7.8×10^{-3}
Nonlinguistic factors					
Memory span	0.0066	0.88	0.0017, 0,15, 0.48, 0.53	59.0	2.5×10^{-3}
Counting	0.017	0.67	0.0003, 0.0013, 0.035, 0.72	50.8	7.0×10^{-7}
No. of encoding	0.051	<0.1	<0.0001, 0.014, 0.018, 0.12	32.9	1.2×10^{-14}
Memory span + counting	0.0099	0.81	0.0007, 0.035, 0.15, 0.76	55.5	7.9×10^{-5}
Memory span + No. of encoding	0.015	0.72	<0.0001, 0.10, 0.46, 0.59	52.5	3.6×10^{-6}

Percent signal changes in the L. F3op/F3t were fitted with a single scale parameter to a model of each factor using its subtracted estimates (Table 3) for the four contrasts of [$\text{Nested}_{(L)} - \text{Conjoined}_{(L)}$], [$\text{Nested}_{(S)} - \text{Conjoined}_{(S)}$], [$\text{Simple}_{(L)} - \text{Conjoined}_{(L)}$], and [$\text{Simple}_{(S)} - \text{Conjoined}_{(S)}$]. The *P* values for the *t*-tests are shown in ascending order. Note that the models of the DoM (with an asterisk) resulted in the best fit of 19 models tested for explaining activations in the L. F3op/F3t, i.e., with the least residual sum of squares (RSS), largest coefficient of determination (r^2), and larger *P* values. The likelihood of models with all null estimates was incalculable (n/a). The likelihood ratio was taken as the ratio of each model's likelihood to the best model's likelihood. The best models of the DoM for the L. F3op/F3t were by far more likely than the other models.

Table 8. Fittings and likelihood of various models tested for the L. SMG.

Syntactic factors	RSS	r^2	*P* values for four contrasts	Log-likelihood	Likelihood ratio
DoM	0.0063	0.92	0.013, 0.083, 0.44, 0.49	58.8	0.079
*DoM + No. of Search	0.0020	0.97	0.22, 0.30, 0.42, 0.62	61.4	1.0
No. of Search	0.075	<0.1	<0.0001, 0.0061, 0.045, 0.090	23.6	3.8×10^{-17}
No. of Merge	0.076	0	<0.0001, 0.0061, 0.013, 0.22	n/a	n/a
Other linguistic factors					
No. of case markers (*-ga*/*-no*)	0.076	0	<0.0001, 0.0061, 0.013, 0.22	n/a	n/a
No. of tense markers (*-ru*/*-ta*)	0.0079	0.90	0.013, 0.023, 0.22, 0.34	55.9	4.1×10^{-3}
Degree of nesting	0.0088	0.88	0.0061, 0.013, 0.22, >0.99	55.5	2.8×10^{-3}
Degree of self-embedding	0.023	0.69	0.0002, 0.0018, 0.013, 0.22	45.5	1.2×10^{-7}
No. of nodes	0.033	0.56	0.0004, 0.0005, 0.0061, 0.013	40.1	6.0×10^{-10}
Depth of postponed symbols	0.076	0	<0.0001, 0.0061, 0.013, 0.22	n/a	n/a
Integration costs	0.021	0.72	0.0001, 0.014, 0.028, 0.18	46.3	2.7×10^{-7}
Storage costs	0.032	0.58	<0.0001, 0.0014, 0.084, 0.49	40.3	7.1×10^{-10}
Syntactic interference	0.0079	0.90	0.013, 0.023, 0.22, 0.34	55.9	4.1×10^{-3}
Positional similarity	0.020	0.73	0.0039, 0.0052, 0.013, 0.029	47.6	1.0×10^{-6}
Nonlinguistic factors					
Memory span	0.021	0.72	0.0001, 0.014, 0.028, 0.18	46.3	2.7×10^{-7}
Counting	0.041	0.46	<0.0001, <0.0001, 0.0039, 0.77	35.6	6.2×10^{-12}
No. of encoding	0.076	<0.1	<0.0001, 0.0061, 0.017, 0.16	22.5	1.4×10^{-17}
Memory span + counting	0.028	0.63	<0.0001, 0.0018, 0.0086, 0.44	41.3	1.9×10^{-9}
Memory span + No. of encoding	0.011	0.85	0.0034, 0.051, 0.13, 0.81	52.1	9.7×10^{-5}

Percent signal changes in the L. SMG were fitted with a single scale parameter to a model of each factor using its subtracted estimates for the four contrasts. Note that the model of the “DoM + number of Search” (with an asterisk) resulted in the best fit of 19 models tested for explaining activations in the L. SMG. The likelihood of models with all null estimates was incalculable (n/a). The likelihood ratio was taken as the ratio of each model’s likelihood to the best model’s likelihood. The best model of the “DoM + number of Search” for the L. SMG was by far more likely than the other models.

span, and counting. This contrast indeed resulted in significant activation in the L. F3op/F3t and L. SMG (Figure 11E and Table 6). In both regions, the signal changes in [Reverse − Same] were not significantly different from 0 (one-sample *t*-test, $P > 0.1$) (Figure 11F). Moreover, the models of the DoM and "DoM + number of Search" were also consistent with the signal changes in both [Nested − Simple] and [Reverse − Same] (Table 4). Number of encoding might explain the results of Figure 11F, but its estimates cannot consistently explain the results of Figure 11D. These results directly support Hypotheses I and II, such that the basic frames of syntactic structures are determined essentially by functional elements, whereas the DoM, along with the number of Search, is a key factor to properly quantifying the complexity of the syntactic structures.

2.3.3. Significance of the connectivity between the L. F3op/F3t and L. SMG

To elucidate the relationships between the L. F3op/F3t and L. SMG, we modeled effective connectivity between the L. F3op/F3t and L. SMG with DCM. Our interest was in identifying the direction of the connectivity modulated by the Nested condition with the largest DoM, which has the largest DoM of all conditions. First, we assumed intrinsic, i.e., task-independent, bi-directional connections, and the models were grouped into three "modulatory families": families *A*, *B*, and *C*, corresponding to the modulation for the bottom-up connection from the L. SMG to the L. F3op/F3t, for the top-down connection from the L. F3op/F3t to the L. SMG, and for both connections, respectively. Each family was composed of three "input models" corresponding to the regions receiving driving inputs (see Figure 9 for all DCM models tested). Using a random-effects BMS, we found that family *A* was the most likely family (expected probability = 0.66, exceedance probability = 0.85) (Figures 12A and 12B). According to a second BMS for the input models within family *A*, model *A1*, in which the L. F3op/F3t received driving inputs, was the best and most probable model

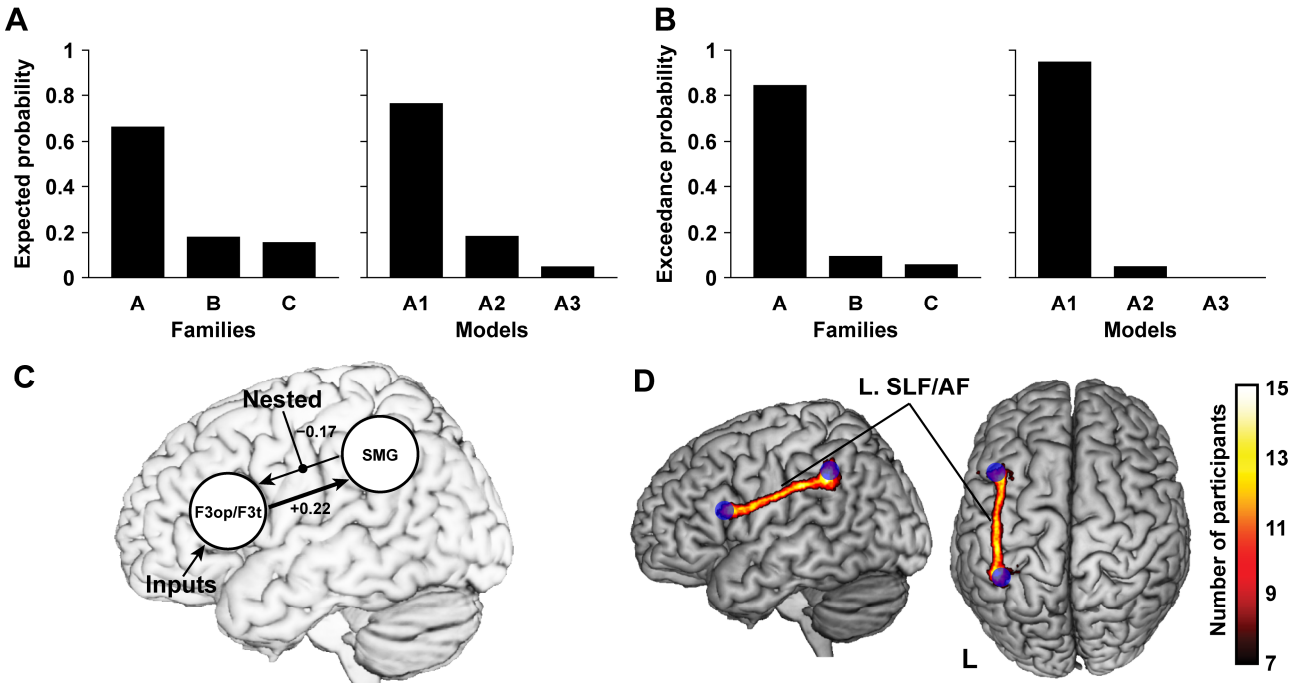


Figure 12. Effective and anatomical connectivity between the L. F3op/F3t and L. SMG.
(**A**–**C**) The results of DCM, testing effective connectivity between the L. F3op/F3t and L. SMG (see Figure 9). Bar graphs show expected probabilities (**A**) and exceedance probabilities (**B**) for each modulatory family and for the input models of the winning family *A*. The best model *A1* (**C**) included a significant top-down connection from the L. F3op/F3t to L. SMG (a thick line). (**D**) Anatomical connectivity between the L. F3op/F3t and L. SMG revealed by DTI. The population probability map is shown on the left lateral and dorsal surfaces of a standard brain with maximum intensity projection. Blue spheres represent seed regions of the L. F3op/F3t and L. SMG.

(expected probability = 0.77, exceedance probability = 0.95). For this particular model, we further tested whether the parameter estimates were significantly different from zero. The intrinsic connection from the L. F3op/F3t to the L. SMG was significantly positive [+0.22; one-sample *t*-test, $t(17) = 4.8$, $P < 0.0002$] (significance level at $\alpha = 0.025$, Bonferroni corrected within a parameter class of intrinsic connections) (Figure 12C), indicating that this top-down connection was consistent among the participants. The modulatory effect for the bottom-up connection was inhibitory [−0.17; $t(17) = 1.4$, $P = 0.17$], though it did not reach the significance level.

To further confirm the anatomical plausibility of the network between the L. F3op/F3t and L. SMG revealed by DCM, we used DTI with a probabilistic tractography. Seed masks were set in the pair of the L. F3op/F3t and L. SMG, both of which were

significantly activated in ($Nested_{(L)}$ − $Conjoined_{(L)}$) > ($Simple_{(S)}$ − $Conjoined_{(S)}$). We observed that a single continuous cluster of the left SLF/AF that connected these regions (cluster size, 3,189 mm^3), together with much smaller clusters or islands (Figure 12D). Moreover, the left SLF/AF was consistently observed in all participants (see Appendices, Figure S1).

2.3.4. Modulation of right frontal activation by nonlinguistic factors

We further examined the involvement of any error-related factors, which were residual factors that might induce cortical activation or deactivation. It should be noted that the factors listed in Tables 3 and 4 were equivalent between the matching and nonmatching stimuli. The [Nonmatching − Matching] contrast under either the sentence conditions (i.e., [Nested + Simple + Conjoined]) or the string conditions (i.e., [Reverse + Same]) consistently resulted in right-dominant activation, especially in the R. F3op/F3t (Figures 13A and 13B), which was in accordance with the same demand of the matching task (Figures 4C and 4D). Other significantly activated regions were the L. LPMC/F3op and ACC under the sentence conditions, as well as the R. SMG under the string conditions (Table 6). As for the [Matching − Nonmatching] contrast, no significant activation was seen under either the sentence or string conditions.

We also examined the activation in [Reverse − Same] for the effect of matching orders (e.g., $A_2\,A_1\,B_1\,B_2$ vs. $A_1\,A_2\,B_1\,B_2$; Figures 4B and 5B). The significant activation was observed only in the R. LPMC (Figure 13C and Table 6), which suggested that activations could indeed be estimated by one and only one non-null factor of memory span in [Reverse − Same] (Table 4). In [Nested − Simple], signal changes in the R. LPMC were also significant (one-sample *t*-test, $P < 0.05$), but were not significantly different between [Nested − Simple] and [Reverse − Same] (paired *t*-test, $P = 0.98$) (Figure 13D). This result was consistent with the equivalent estimates of memory span between [Nested − Simple] and [Reverse − Same].

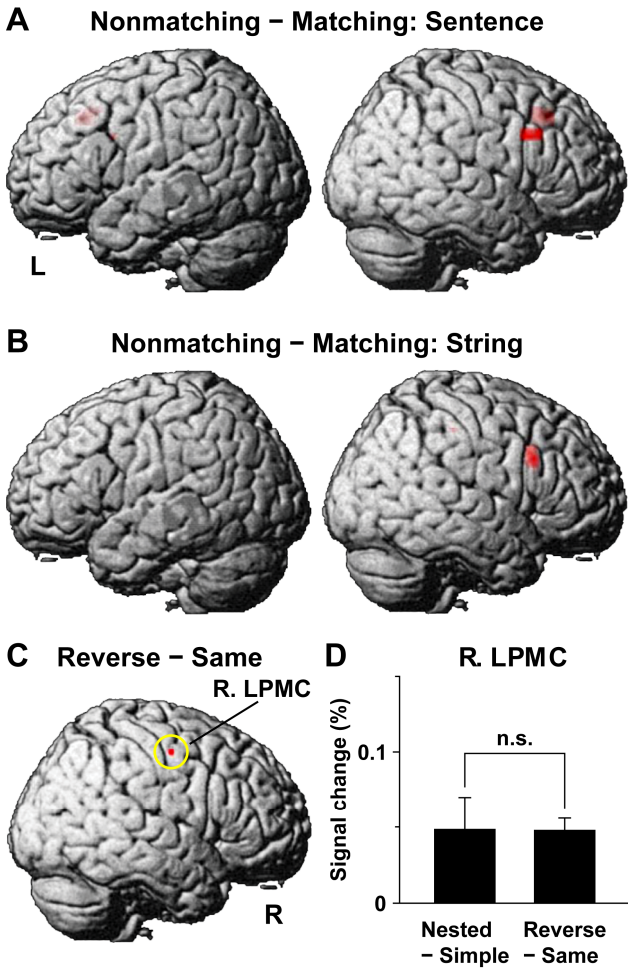


Figure 13. Modulation of the right frontal activations by nonlinguistic factors. One-sample *t*-tests were used for the contrasts indicated. (**A**) Regions identified by the Nonmatching − Matching contrast under the sentence conditions, related to error-related factors. Note the right-dominant activation, especially in the R. F3op/F3t. (**B**) Regions identified by the Nonmatching − Matching contrast under the string conditions. (**C**) Regions identified by the Reverse − Same contrast. This contrast revealed the difference in matching orders (e.g., A_2 A_1 B_1 B_2 vs. A_1 A_2 B_1 B_2). Note the significant activation in the R. LPMC. (**D**) The percent signal changes in the R. LPMC, which was consistent with the equivalent estimates of memory span (see Table 4).

It should be noted that R. LPMC activation was also observed for the main effects of condition and length (Figures 11A and 11B), which probably reflected the factor of memory span.

2.4. Discussion

By employing a novel paradigm to directly contrast jabberwocky sentences (Nested, Simple, and Conjoined) with letter strings (Reverse and Same) (Figures 4 and 5), we obtained four striking results. First, we found that the DoM was indeed a key syntactic factor that

could account for syntax-selective activations in the L. F3op/F3t and L. SMG, localized by the ($Nested_{(L)}$ − $Conjoined_{(L)}$) > ($Simple_{(S)}$ − $Conjoined_{(S)}$) contrast (Figures 11C and 11D). By constructing a model of each syntactic, other linguistic, or nonlinguistic factor using the estimates of each factor (Table 3), we demonstrated that the models of the DoM and "DoM + number of Search" were the models that best accounted for L. F3op/F3t and L. SMG activation, respectively (Tables 7 and 8). Second, by directly contrasting jabberwocky sentences with letter strings, i.e., (Nested − Simple) > (Reverse − Same), we showed that selective activation in the L. F3op/F3t and L. SMG, which was consistent with the involvement of the syntactic factors demonstrated above, was replicated irrespective of identical matching and symbol orders (e.g., N_2 N_1 V_1 V_2 and A_2 A_1 B_1 B_2 for the Nested and Reverse, respectively) (Figures 11E and 11F). This point is particularly important, because temporal order-related or memory-related factors have often been confused with differences in structure or grammar type. Our results strongly support syntactic structures being recursively constructed when well-formed sentences are given. Third, by using DCM, we found the best model to be that with an inhibitory modulatory effect for the bottom-up connectivity from the L. SMG to L. F3op/F3t, and with driving inputs to the L. F3op/F3t (Figures 12A and 12B). For this best model, the top-down connection from the L. F3op/F3t to L. SMG was significantly positive (Figure 12C). By using DTI, we also confirmed that the left dorsal pathway of the SLF/AF consistently connected these two regions (Figure 12D). These results suggest a transmission of information about the DoM through this specific dorsal pathway. Finally, we established that nonlinguistic order-related and error-related factors significantly activated mostly right frontal regions. The difference in memory span significantly modulated the R. LPMC activation in [Reverse − Same], suggesting that this region plays a major role in tracking matching orders (Figures 13C and 13D), while error-related factors in [Nonmatching − Matching] consistently modulated the R. F3op/F3t

activation under both sentence and string conditions (Figures 13A and 13B). In summary, these results indicate that the identified network of the L. F3op/F3t and L. SMG subserves the calculation of the DoM in recursively merged sentences, and that the R. LPMC monitors memory span to drive a memory-maintenance system. If multiple factors, such as the number of nodes, memory span, etc., are equally plausible for explaining activations, then a superordinate concept, such as "syntactic complexity," can be a more useful factor than individual factors. However, in the present experiment, the minimal factor of the DoM *sufficiently* explained the activation pattern observed, while other factors were by far less likely to (see Tables 7 and 8). Therefore, syntactic complexity was restricted and replaced by the DoM as a more fundamental concept, just like the historical development from "gene" to DNA.

Our finding that the L. F3op/F3t subserves the syntactic computation further extends the functional specialization of this region reported previously (Embick et al., 2000; Musso et al., 2003; Sakai, 2005; Kinno et al., 2008). Some previous fMRI studies have interpreted the L. F3op/F3t activation as reflecting temporal order-related or memory-related factors (Bahlmann et al., 2007; Santi & Grodzinsky, 2010). However, these previous studies contrasted hierarchically complex sentences with simpler sentences, while it is clear that syntactic factors, including the DoM, were also involved. Moreover, the previously reported modulation of L. F3op/F3t activation by scrambling word orders (Röder et al., 2002) can be consistently explained by the DoM, because scrambling requires "movements" of noun phrases to higher nodes by applying more Merge operations, thus increasing the DoM. The size of linguistic constituents also correlates with the DoM, especially when the number of left/right branches increases as in the case of Pallier, Devauchelle, and Dehaene (2011). In the present study, we characterized the neural substrates of syntactic computation by segregating a number of possible factors, and demonstrated that the exact activations in the L.

F3op/F3t can be used to calculate the DoM. Indeed, each structure of our jabberwocky sentences was uniquely represented by the DoM, together with the numbers of Merge and Search (see Table 3).

A previous fMRI study involving the implicit learning of an artificial regular grammar has reported that the "ungrammatical − grammatical" contrast for symbol sequences activated the L. F3op/F3t, suggesting that such activation was due to artificial syntactic violations among any error-related factors (Petersson, Forkstam, & Ingvar, 2004). However, this result may not depend on the presence of errors themselves, but on other rule-related processes associated with error correction, etc. In contrast, we have previously demonstrated that an explicit syntactic decision enhanced the L. F3op/F3t activation under *both* grammatical and ungrammatical conditions (Suzuki & Sakai, 2003). On the other hand, a recent fMRI study has compared nested and branching constructions, suggesting that activation in the bilateral posterior superior temporal cortex reflects an integration of lexico-semantic and syntactic information (Friederici, Makuuchi, & Bahlmann, 2009). However, this result also reflected the effects of semantic factors that were inevitably confounded with any structural processing, because real German sentences were used as stimuli in the study. Furthermore, according to our paradigm, the temporal cortex in neither of the hemispheres showed any significant activation for the Nested condition (Figure 11). It was thus quite important to verify that activation in the L. F3op/F3t, but not in the temporal cortex, is indeed crucial for syntactic processing.

In the present study, we found that the L. SMG activations were modulated by the "DoM + number of Search." Consistent with the suggested role of the L. AG/SMG for vocabulary knowledge or lexical processing (Lee et al., 2007; Pattamadilok et al., 2010), the number of Search is likely to induce such a modulation, in the sense that Search assigns a specific feature that can be linked with morphosyntactic changes. The Japanese language

happens to lack agreement of grammatical features, but it is nevertheless equipped with the general Search procedure attested for various phenomena in the language (Fukui & Sakai, 2003). Our results suggest that Search actually applied to a subject–verb pair of a jabberwocky sentence in the present paradigm, where the relevant features (in this case, vowels) are experimentally "inserted." It should also be noted in this connection that the Japanese language exhibits a phenomenon called "honorification," in which the case of an honored person and the form of honorifics on the verb optionally match (Gunji, 1987; Ivana & Sakai, 2007). Search assigns such features as honorifics. Our previous fMRI study using an honorification judgment task reported activation in the L. F3op/F3t and L. LPMC, as well as in the left inferior parietal gyrus and L. AG (Momo, Sakai, & Sakai, 2008), which is consistent with activation in the L. AG/SMG in the present study (Tables 5 and 6).

Our DCM and DTI results further indicate that L. SMG activation reflecting the DoM mirrored a top-down influence from the L. F3op/F3t through the left dorsal pathway of the SLF/AF. A recent DCM study with a picture–sentence matching task has suggested that the L. F3op/F3t received driving inputs (den Ouden et al., 2012), which was consistent with our DCM results. Moreover, our previous studies revealed that the functional connectivity between the L. F3t/F3O (pars orbitalis) and L. AG/SMG was selectively enhanced during sentence processing (Homae, Yahata, & Sakai, 2003), and that the L. AG/SMG was also activated during the identification of correct past-tense forms of verbs, probably reflecting an integration of syntactic and vocabulary knowledge (Tatsuno & Sakai, 2005). Considering the role of the L. AG/SMG in lexical processing, the Search operation based on the DoM would be essential in assigning relevant features to syntactic objects derived from lexical items.

In [Nonmatching − Matching], the R. F3op/F3t was consistently activated under both sentence and string conditions (Figures 13A and 13B), whereas the L. LPMC/F3op, ACC, or R. SMG were activated under either condition. These four regions were also

activated in (Nested$_{(L)}$ − Conjoined$_{(L)}$) > (Simple$_{(S)}$ − Conjoined$_{(S)}$), and in Long > Short while combining [Nested − Conjoined] and [Simple − Conjoined]; the ACC and R. SMG were activated in (Nested − Simple) > (Reverse − Same) as well. It appears likely that a part of the activation in these four regions reflects error-related factors including the detection and correction of errors, which would be more demanding with the Nested condition as well as in the Long > Short contrast. Because the L. LPMC has been known to selectively subserve syntactic processing (Indefrey et al., 2001; Hashimoto & Sakai, 2002; Kinno et al., 2008), a weak activation in the L. LPMC/F3op only under the sentence conditions may reflect the confirmation of sentence constructions when confronted with nonmatching stimuli. On the other hand, it has been suggested that the dorsal ACC plays a major role during conflict monitoring during a highly demanding task, e.g., a Stroop task (Botvinick, Cohen, & Carter, 2004). Our recent magnetoencephalography study also suggested that the anterior portion of the ACC is a candidate region for monitoring syntactically anomalous sentences (Iijima, Fukui, & Sakai, 2009). Moreover, previous studies on response inhibition, typically tested with a No-go task, suggested that the R. F3op/F3t, ACC, and R. SMG were also involved in monitoring anomalous stimuli (Chikazoe et al., 2007). In contrast to these factors that activated mostly the right and medial regions, it is noteworthy that syntactic factors clearly activated the left frontal and parietal regions.

Any factors associated with matching and symbol orders might influence activation in the language areas, but we clearly showed that the R. LPMC was activated in [Reverse − Same] (Figure 13C) for the effect of memory span related to matching orders. The study of real German sentences also reported activation in the right dorsal premotor area for the contrast between nested vs. branching constructions (Friederici, Makuuchi, & Bahlmann, 2009); however, the right dorsal premotor area was not the same region as the R. LPMC in the present study. In this German study, memory span was controlled by the insertion of

some words, while matching and symbol orders still differed, and thus factors other than memory span were inevitably introduced to interpret the right dorsal premotor activation. The identification of critical factors in language processing thus inevitably depends on an experimental design that involves an effective contrast of conditions. One promising direction for research is further clarifying activations modulated by other linguistic and nonlinguistic factors, which may eventually make possible the elucidation of all aspects of linguistic information in the human brain.

Chapter 3. Further Confirmation of Hypotheses I and II

3.1. A picture–sentence matching paradigm

We further examined whether our hypotheses held for various cases discussed in previous studies. In our fMRI study (Kinno et al., 2008), we used a picture–sentence matching task with three sentence types in Japanese: active, passive, and scrambled sentences (Figure 14A). In the picture–sentence matching task, participants read a sentence silently and judged whether the action depicted in a picture matched the meaning of the sentence. Each sentence had two *arguments*, noun phrases that each assume a different grammatical relation ("subject, direct object, or indirect object" in linguistic terms) in a sentence, as well as a different semantic role ("agent, experiencer, or patient" in linguistic terms, i.e., an agent performs the action, and an experiencer/patient is affected by it). These three conditions were thus called Two-argument conditions. More specifically, the active, passive, and scrambled sentences corresponded to "agent and patient" (subject and direct object), "experiencer and agent" (subject and indirect object), and "patient and agent" (direct object and subject) types, respectively. Pictures consisted of two stick figures, each of which was distinguished by a "head" symbol: a circle (○), square (□), or triangle (△). These sentences excluded the involvement of pragmatic information about word use (e.g., "*An officer chases a thief*" is more acceptable than "*A thief chases an officer*"). To minimize the effect of general memory demands, a whole sentence of a minimal length was visually presented for a longer time than was needed to respond.

In Japanese syntax, grammatical relations are first marked by grammatical particles (nominative, dative, or accusative), which in turn allow the assignment of semantic roles. In the active sentences we used, a noun phrase with the nominative case marker *-ga* (*green* letters in Figures 14) is associated with an agent, and the one with the accusative case marker *-o* is associated with a patient. For the passive sentences we used, however, a noun phrase with the nominative case marker *-ga* is associated with an experiencer (a person experiencing

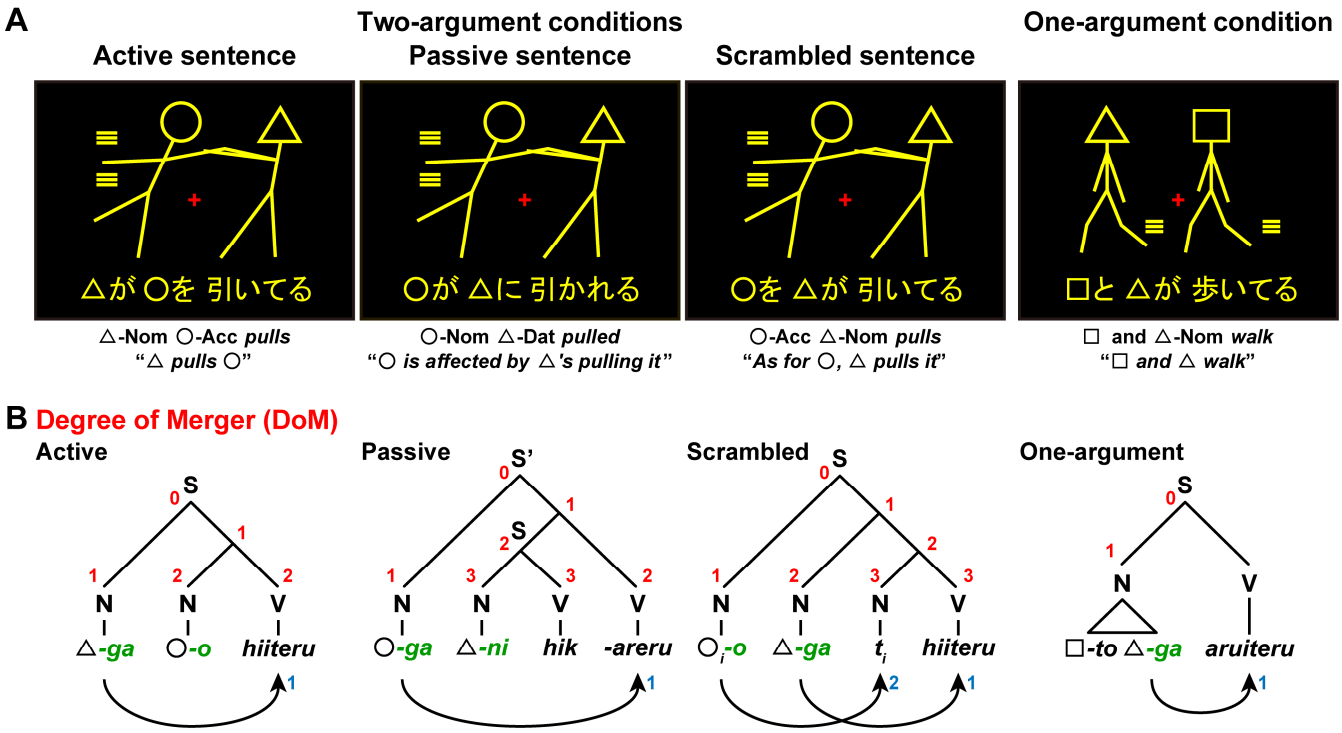


Figure 14. A picture–sentence matching paradigm in Kinno et al. (2008).
(**A**) A picture–sentence matching task under either Two-argument conditions or a One-argument condition. Each stimulus consisted of one picture (top) and one sentence (bottom). Below each example, word-by-word and full translations in English are shown. An identical picture set was used under the Two-argument conditions, where they tested three sentence types: active sentences ("△*-ga* ○*-o hiiteru*"), passive sentences ("○*-ga* △*-ni hik-areru*"), and scrambled sentences ("○*-o* △*-ga hiiteru*"). Under the One-argument condition, they presented syntactically simpler active sentences ("□*-to* △*-ga aruiteru*"). (**B**) The syntactic structures of three sentence types. The digits shown in red and blue denote the DoM for each node and "number of Search," respectively. Symbols used: S and S', sentence; N, noun phrase; V, verb phrase; Nom, nominative case; Acc, accusative case; Dat, dative case; *-ga*, nominative case marker; *-o*, accusative case marker; *-ni*, dative case marker; *-to*, coordinator; t_i, trace (subscripts denote the same entity).

a situation), whereas a passive bound verb "-(*r*)*areru*" marks passiveness, making a subject–verb pair with the experiencer. In contrast, a noun phrase with the dative marker *-ni* is associated with an agent, whereas an action verb (e.g., "*hik*(*u*)": "*pull*") makes a subject–verb pair with the agent, forming a subordinate clause within the main clause "○*-ga* … -(*r*)*areru*." Note that similar causative structures exist in both Japanese and English: "*Hanako-ga kare-ni hik-aseta*": "*Hanako made him pull*." Actually, there are two types of passivization in Japanese: *ni* passive (e.g., "*Hanako-ga Taro-ni hik-areru*": "*Hanako is affected by Taro's pulling her*") and *ni yotte* passive (e.g., "*Hanako-ga Taro-ni yotte hik-areru*": "*Hanako is pulled by Taro*"). According to Kuroda (1992), the *ni* passive involves no noun-phrase

movement, while the *ni yotte* passive involves a movement similar to that in English. For the scrambled sentences, an object moves from its canonical position to higher nodes by undergoing another Merge operation. This type of construction is perfectly normal, not only in Japanese but in German, Finnish, and other languages. We also tested the One-argument condition, under which each sentence was presented with an intransitive verb and double agents. This condition did not involve two-argument relationships, and was thus syntactically simpler than any of the Two-argument conditions.

3.2. Hypothesis III

Here we present the following hypothesis (Hypothesis III):

(3) The DoM domain changes dynamically in accordance with iterative Merge applications, the Search distances, and/or task requirements.

Since Merge combines two syntactic objects to form a larger structure, it always produces a one-level higher node. When Merge applies iteratively to an existing phrase or sentence, the DoM domain becomes thus larger in accordance with the number of Merge applications. The Search distance is the structural distance between two distinct parts to which the Search operation applies, regardless of the nodes that are irrelevant to the Search operation. As observed from Figure 4, the DoM domain changes in accordance with the Search distance. On the other hand, for every sentence stimulus in the present study, the construction of syntactic structures was ensured by task requirements: The three sentence types were completely mixed and had to be distinguished. Task requirements include not only certain constraints required by experimental tasks, but also the detailed parsing naturally required to

understand a part of phrases or sentences (e.g., subject–verb relationships and noun–pronoun (coreference) relationships).

In the above-mentioned paradigm (Kinno et al., 2008), the four task conditions (three sentence types under the Two-argument conditions, as well as one type under the One-argument condition) were completely mixed (see Figure 14A). With such task requirements, the DoM domain spanned three relevant words for all sentence types under the Two-argument conditions. Under the One-argument condition, the action of two stick figures was always identical, and thus a subject (a triangle just below N in Figure 14B) was regarded as a unit. Under these four task conditions, participants were required to check at least one of the argument–verb relationships, demanding Search at least once. For the scrambled sentences alone, an additional Search operation should match the identical indices of the moved object and its trace. For the active, passive, and scrambled sentences, the estimates of DoM were 2, 3, and 3, respectively, while those of the DoM was 1 under the One-argument condition.

3.3. Applying the DoM to various sentence types

In the study of Kinno et al. (2008), we directly contrasted passive and active sentence conditions to identify a cortical region activated by purely syntactic processes. This stringent contrast resulted in significant activation in the left dorsal F3t (L. dF3t) alone [(−48, 24, 21), $Z = 3.8$] (Figure 15A), which was very close to the L. F3op/F3t activation in our study (Ohta, Fukui, & Sakai, 2013a). The L. dF3t activation was significantly enhanced under both the passive and scrambled sentence conditions compared to that under the active sentence condition ($P \leq 0.033$) (Figure 15B), whereas there was no significant difference between the passive and scrambled sentence conditions ($P = 0.15$). Taking the One-argument condition as a reference for subtracted estimates, the signal changes in the L. dF3t were precisely correlated in a step-wise manner with the parametric model of the DoM [1, 2, 2],

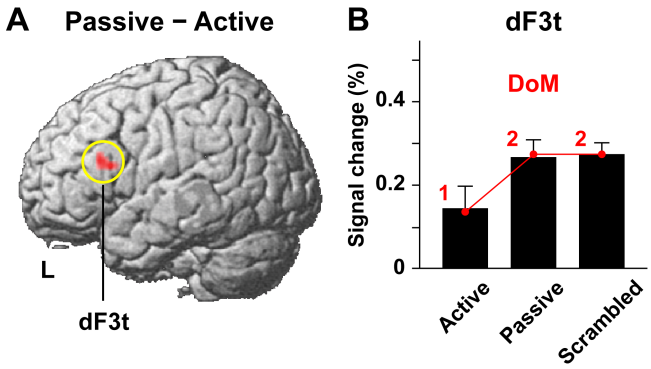


Figure 15. Activations in the L. dF3t modulated by the DoM.
(**A**) A region identified by the Passive − Active contrast (see Figure 14). Activations were projected onto the left (L) lateral surface of a standard brain. (**B**) Percent signal changes for the active, passive, and scrambled sentence conditions in the L. dF3t, taking the One-argument condition as a reference. Overlaid red dots and lines denote the values fitted with the estimates (digits in red) for the model of the DoM.

producing the RSS of 0.0001 and r^2 of 0.99, without significant deviation for the three contrasts ($P \geq 0.87$). The model of the DoM thus *sufficiently* explains the L. dF3t activations. It should be noted that the parametric model of the "number of nodes" [2, 4, 4] also yielded the same fitting results in this case. The design of experimental paradigms limited the separation of multiple factors.

In a recent fMRI study, only right-branching constructions were examined, and activations in the L. F3t were modulated by the size of constituents (i.e., number of terminal nodes) (Pallier, Devauchelle, & Dehaene, 2011). Since the estimates of the DoM were identical to those of the "number of Merge" or the "number of nonterminal nodes" in this case, it was not possible to separate these factors. Taking the simplest condition (lists of unrelated words) as an appropriate reference, the model of the DoM actually showed a comparable or better goodness of fit for activations in the L. F3t, when compared with their log-fitting models.

Chapter 4. General Discussion

4.1. The effect of the Search distance on the DoM

Neuroimaging and psycholinguistic studies have reported that English sentences with object-relative clauses have higher processing loads than those with subject-relative clauses (Stromswold et al., 1996; Just et al., 1996; Gibson, 2000). To properly parse relative clauses, the relative pronoun and its antecedent are coindexed; "*who*$_i$" and "*the boy*$_i$," respectively, in the example shown in Figure 16. In a subject-relative clause, a relative pronoun "*who*$_i$" was displaced from the *subject* position denoted by a trace t_i (originally, "*the boy*$_i$ *likes the girl*"), while in an object-relative clause, a relative pronoun was displaced from the *object* position (originally, "*the girl likes the boy*$_i$"). Following the proposal by Hawkins (1999), we assume that the relative pronoun searches the corresponding trace within tree structures of a sentence (see curved arrows in Figure 16). In a subject-relative clause, Search ends at the initiation of the verb phrase, whereas in an object-relative clause, Search ends *after* a verb appears within a subordinate clause. In accordance with the Search distances for these examples, the DoM would become one unit larger for the object-relative clause than the subject-relative one. Higher processing loads observed with object-relative clauses are consistent with this inference.

4.2. The effect of task requirements on the DoM

If Hypothesis III is correct, then L.F3op/F3t activation can differ by task requirements, even when the same sentences are presented. In our previous fMRI study, we compared three explicit linguistic tasks with the same set of normal two-word sentences: syntactic decision, semantic decision, and phonological decision tasks (Suzuki & Sakai, 2003). In the syntactic decision task, the participants judged whether the presented sentence was syntactically correct, and this judgment required syntactic knowledge about the distinction between transitive and intransitive verbs (e.g., normal sentence, "*yuki-ga tumoru*":

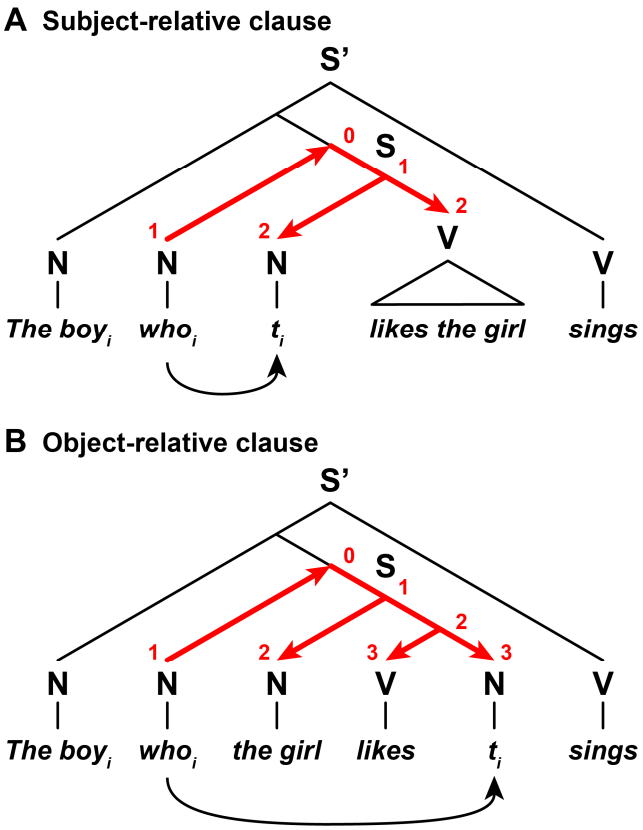


Figure 16. DoM domains varied with the Search distances.
(**A**) A sentence with a subject-relative clause. (**B**) A sentence with an object-relative clause. In these relative clauses, a relative pronoun *who*$_i$ is displaced from its subject or object position denoted by a trace t_i. A set of red straight arrows corresponds to the DoM domains. The digits shown in red denote the DoM for each node within the domain.
Symbols used: S and S', sentence; N, noun phrase; V, verb phrase; t_i, trace (subscripts denote the same entity).

"*snow lies* (*on the ground*)"; anomalous sentence, "*yuki-o tumoru*": "(*something*) *lies snow*"). In the semantic decision task, lexico-semantic knowledge about selectional restrictions was indispensable. In the phonological decision task, phonological knowledge about accent patterns was required. Neither the semantic decision task nor the phonological decision task, both with *implicit* syntactic processing, elicited significant activations in the L. F3op/F3t (−57, 9, 6), which was significantly activated during *explicit* syntactic processing, even by a direct comparison between the syntactic decision task and the other tasks. These results suggest the presence of the DoM domain in accordance with the task requirements of explicit syntactic processing.

4.3. The mixed effects of Search distances and task requirements on the DoM

In another fMRI study, we directly compared syntactic decision and short-term memory tasks (Hashimoto & Sakai, 2002). In this unique paradigm, we visually presented nested sentences that included two proper nouns, two verbs, and one pronoun, in which either a verb or pronoun was underlined. After presenting one complete sentence in a phrase-by-phrase manner, paired phrases, including an underlined phrase, were shown. In one syntactic decision task (SYN-1), participants were required to judge whether the subject of an underlined verb corresponded to the person in paired phrases (Figure 17A). In this case, the Search distance was the structural distance between the subject and verb of the same clause. In the other syntactic decision task (SYN-2), the participants were required to judge whether an underlined pronoun was able to refer to the person in paired phrases (Figure 17B). In this case, the Search distance was the structural distance between the coindexed noun and pronoun. In these syntactic decision tasks, the Search distance, and consequently the DoM domain, changed dynamically in accordance with the different task requirements, even when the same sentences were presented. The estimate of the resultant DoM was 2 for both cases. In a short-term memory task with a sentence, the participants memorized the linear order of the phrases, and judged whether the left-hand phrase preceded the right-hand one in the original sequence (Figure 17C). With such a task requirement, the factor of DoM would become less effective. Indeed, we found that activations in the L. F3op/F3t were equally enhanced in both syntactic decision tasks when compared with the short-term memory task.

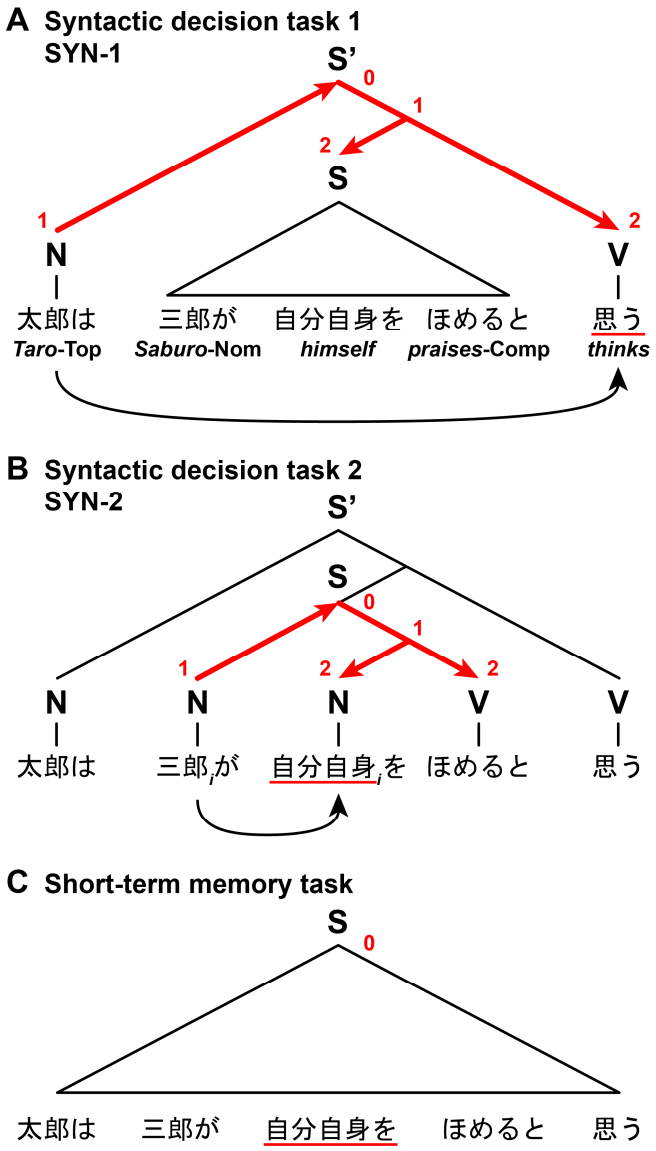


Figure 17. DoM domains varied with the Search distances and task requirements. In this task, participants read Japanese nested sentences ("*Taro-wa Saburo-ga jibunjishin-o homeru-to omou*": "*Taro thinks that Saburo praises himself*"), in which phrases were sequentially presented. (**A**) A syntactic decision task 1, in which participants judged subject–verb relationships. A set of red straight arrows corresponds to the DoM domains. The digits shown in red denote the DoM for each node within the domain. (**B**) A syntactic decision task 2, in which participants judged noun-pronoun (coreference) relationships (subscripts denote the same entity). (**C**) A short-term memory task with a sentence, in which participants judged the temporal order of the phrases. Symbols used: S and S', sentence; N, noun phrase; V, verb phrase; Top, topic; Nom, nominative case; Comp, complementizer.

Conclusions

In this dissertation, I tried to elucidate the computational principles of syntax in the language areas from the following perspectives. First, I provided theoretical discussions about the hierarchical tree structures of sentences, and introduced the two fundamental linguistic operations of Merge and Search. I also proposed the DoM as a key computational concept. I then presented our hypotheses that the DoM is a key computational concept to properly measure the complexity of tree structures (Hypothesis I), and that the basic frame of the syntactic structure of a given linguistic expression is determined essentially by functional elements, which trigger Merge and Search operations (Hypothesis II). Second, I presented our fMRI studies, which have demonstrated that the DoM, together with the number of Search, is indeed the key syntactic factor that accounts for syntax-selective activations in the L. F3op/F3t and L. SMG (Ohta, Fukui, & Sakai, 2013a). Moreover, based on the DCM and DTI results, I revealed the significance of the top-down connection from the L. F3op/F3t to L. SMG, suggesting that information about the DoM is transmitted through this specific dorsal pathway. Third, I further hypothesized that the DoM domain changes dynamically in accordance with iterative Merge applications, the Search distances, and/or task requirements (Hypothesis III). I showed that the DoM sufficiently explains activation modulations due to different structures reported in previous fMRI studies (Kinno et al., 2008; Pallier, Devauchelle, & Dehaene, 2011). Finally, I confirmed that Hypothesis III accounts for higher processing loads observed with object-relative clauses, as well as activations in the L. F3op/F3t during explicit syntactic decision tasks, reported in the previous neuroimaging and psycholinguistic studies (Stromswold et al., 1996; Just et al., 1996; Gibson, 2000; Hashimoto & Sakai, 2002; Suzuki & Sakai, 2003). It is likely that the DoM serves as a key computational principle for other human-specific cognitive capacities, such as mathematics

and music, both of which can be expressed by hierarchical tree structures. A future investigation into the computational principles of syntax will further deepen our understanding of uniquely human mental faculties.

Acknowledgments

First and foremost, I would like to thank my supervisor, Professor Kuniyoshi L. Sakai, for his constant encouragement and guidance throughout the years. His patience and kindness, as well as his academic experience, have been invaluable to me. I owe my deepest gratitude to Professor Naoki Fukui (Sophia University). Without his insightful comments on theoretical linguistics, this dissertation would not have materialized. I would like to express my gratitude to the members of my research committee, Professor Yoshihiko Tanno, Professor Kazuo Okanoya, Professor Shoichi Ishiura, and Professor Suguru Kawato, for their valuable advice and suggestions. I would like to offer my special thanks to Professor Noam Chomsky (Massachusetts Institute of Technology) for his valuable comments. I also thank the late Professor Shige-Yuki Kuroda (University of California, San Diego) for his helpful discussion.

I have greatly benefitted from Dr. Ryuta Kinno (Showa University), Dr. Xavier Perrot (Université de Lyon), Dr. Manabu Tamura (Tokyo Women's Medical University), Dr. Hiroyuki Miyashita, Mr. Kazuki Iijima, Mr. Tomoo Inubushi, Mr. Tomoya Nakai, Ms. Kayako Yamamoto, Mr. Yuichiro Shimizu, and Mr. Masatomi Iizawa for their invaluable comments on the earlier manuscript. I would also like to thank Ms. Naoko Komoro and Ms. Natsuki Saeki for their technical assistance, and Ms. Saori Matsukura and Ms. Hiromi Matsuda for their administrative assistance.

I would like to offer my gratitude to my family, Satoshi, Rie, and Manami, for their moral support and warm encouragement. I would like to thank Reverends Kazuhiro and Sonoe Hama, for their warm support. Last, but not least, I would like to express my sincere appreciation to Atsuyo Nishigaki for her support, love, and patience during my years in graduate school.

References

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience 14*, 1067-1074. doi: 10.1038/nn.2869

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage 26*, 839-851. doi: 10.1016/j.neuroimage.2005.02.018

Bahlmann, J., Rodriguez-Fornells, A., Rotte, M., & Münte, T. F. (2007). An fMRI study of canonical and noncanonical word order in German. *Human Brain Mapping 28*, 940-949. doi: 10.1002/hbm.20318

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage 42*, 525-534. doi: 10.1016/j.neuroimage.2008.04.249

Beckers, G. J. L., Bolhuis, J. J., Okanoya, K., & Berwick, R. C. (2012). Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *NeuroReport 23*, 139-145. doi: 10.1097/WNR.0b013e32834f1765

Behrens, T. E. J., Johansen-Berg, H., Jbabdi, S., Rushworth, M. F. S., & Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage 34*, 144-155. doi: 10.1016/j.neuroimage.2006.09.018

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences 8*, 539-546. doi: 10.1016/j.tics.2004.10.003

Chikazoe, J., Konishi, S., Asari, T., Jimura, K., & Miyashita, Y. (2007). Activation of right inferior frontal gyrus during response inhibition across response modalities. *Journal of Cognitive Neuroscience 19*, 69-80. doi: 10.1162/jocn.2007.19.1.69

Chomsky, N. (1957). *Syntactic Structures*. The Hague, The Netherlands: Mouton Publishers.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control 2*, 137-167. doi: 10.1016/S0019-9958(59)90362-6

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: The MIT Press.

Chomsky, N. (2004). Beyond Explanatory Adequacy. In A. Belletti (Ed.), *Structures and Beyond: The Cartography of Syntactic Structures, Volume 3* (pp. 104-131). Oxford, UK: Oxford University Press.

Corballis, M. C. (2007). Recursion, language, and starlings. *Cognitive Science 31*, 697-704. doi: 10.1080/15326900701399947

Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron 24*, 427-432. doi: 10.1016/S0896-6273(00)80855-7

den Ouden, D.-B., Saur, D., Mader, W., Schelter, B., Lukic, S., Wali, E., Timmer, J., & Thompson, C. K. (2012). Network modulation during complex syntactic processing. *NeuroImage 59*, 815-823. doi: 10.1016/j.neuroimage.2011.07.057

Embick, D., Marantz, A., Miyashita, Y., O'Neil, W., & Sakai, K. L. (2000). A syntactic specialization for Broca's area. *Proceedings of the National Academy of Sciences of the United States of America 97*, 6150-6154. doi: 10.1073/pnas.100098897

Esterman, M., Tamber-Rosenau, B. J., Chiu, Y.-C., & Yantis, S. (2010). Avoiding non-independence in fMRI data analysis: Leave one subject out. *NeuroImage 50*, 572-576. doi: 10.1016/j.neuroimage.2009.10.092

Flöel, A., de Vries, M. H., Scholz, J., Breitenstein, C., & Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca's area predicts grammar learning success. *NeuroImage 47*, 1974-1981. doi: 10.1016/j.neuroimage.2009.05.046

Friederici, A. D., Makuuchi, M., & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *NeuroReport 20*, 563-568. doi: 10.1097/WNR.0b013e3283297dee

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*, 170-177. doi: 10.1093/cercor/13.2.170

Friston, K. J., Harrison, L., & Penny, W. (2003). Dynamic causal modelling. *NeuroImage 19*, 1273-1302. doi: 10.1016/S1053-8119(03)00202-7

Friston, K. J., & Henson, R. N. (2006). Commentary on: Divide and conquer; A defence of functional localisers. *NeuroImage 30*, 1097-1099. doi: 10.1016/j.neuroimage.2006.02.007

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping 2*, 189-210. doi: 10.1002/hbm.460020402

Fukui, N. (2011). Merge and Bare Phrase Structure. In C. Boeckx (Ed.), *The Oxford Handbook of Linguistic Minimalism* (pp. 73-95). Oxford, UK: Oxford University Press.

Fukui, N., & Sakai, H. (2003). The visibility guideline for functional categories: Verb raising in Japanese and related issues. *Lingua 113*, 321-375. doi: 10.1016/S0024-3841(02)00080-3

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature 440*, 1204-1207. doi: 10.1038/nature04675

Gibson, E. (2000). The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language,*

Brain: Papers from the First Mind Articulation Project Symposium (pp. 95-126). Cambridge, MA: The MIT Press.

Griffiths, J. D., Marslen-Wilson, W. D., Stamatakis, E. A., & Tyler, L. K. (2013). Functional organization of the neural language system: Dorsal and ventral pathways are critical for syntax. *Cerebral Cortex 23*, 139-147. doi: 10.1093/cercor/bhr386

Gunji, T. (1987). *Japanese Phrase Structure Grammar: A Unification-Based Approach*. Dordrecht, The Netherlands: D. Reidel Publishing Company.

Hashimoto, R., & Sakai, K. L. (2002). Specialization in the left prefrontal cortex for sentence comprehension. *Neuron 35*, 589-597. doi: 10.1016/S0896-6273(02)00788-2

Hawkins, J. A. (1999). Processing complexity and filler-gap dependencies across grammars. *Language 75*, 244-285. doi: 10.2307/417261

Homae, F., Yahata, N., & Sakai, K. L. (2003). Selective enhancement of functional connectivity in the left prefrontal cortex during sentence processing. *NeuroImage 20*, 578-586. doi: 10.1016/S1053-8119(03)00272-6

Hopcroft, J. E., & Ullman, J. D. (1979). *Introduction to Automata Theory, Languages, and Computation*. Reading, MA: Addison-Wesley.

Iijima, K., Fukui, N., & Sakai, K. L. (2009). The cortical dynamics in building syntactic structures of sentences: An MEG study in a minimal-pair paradigm. *NeuroImage 44*, 1387-1396. doi: 10.1016/j.neuroimage.2008.10.041

Indefrey, P., Hagoort, P., Herzog, H., Seitz, R. J., & Brown, C. M. (2001). Syntactic processing in left prefrontal cortex is independent of lexical meaning. *NeuroImage 14*, 546-555. doi: 10.1006/nimg.2001.0867

Ivana, A., & Sakai, H. (2007). Honorification and light verbs in Japanese. *Journal of East Asian Linguistics 16*, 171-191. doi: 10.1007/s10831-007-9011-7

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science 274*, 114-116. doi: 10.1126/science.274.5284.114

Kinno, R., Kawamura, M., Shioda, S., & Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Human Brain Mapping 29*, 1015-1027. doi: 10.1002/hbm.20441

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience 12*, 535-540. doi: 10.1038/nn.2303

Kuroda, S.-Y. (1992). On Japanese passives. In S.-Y. Kuroda (Ed.), *Japanese Syntax and Semantics: Collected Papers* (pp. 183-221). Dordrecht, The Netherlands: Kluwer Academic.

Kvålseth, T. O. (1985). Cautionary note about R^2. *The American Statistician 39*, 279-285. doi: 10.1080/00031305.1985.10479448

Lee, H., Devlin, J. T., Shakeshaft, C., Stewart, L. H., Brennan, A., Glensman, J., Pitcher, K., Crinion, J., Mechelli, A., Frackowiak, R. S. J., Green, D. W., & Price, C. J. (2007). Anatomical traces of vocabulary acquisition in the adolescent brain. *Journal of Neuroscience 27*, 1184-1189. doi: 10.1523/JNEUROSCI.4442-06.2007

Lewis, R., & Nakayama, M. (2002). Syntactic and positional similarity effects in the processing of Japanese embeddings. In M. Nakayama (Ed.), *Sentence Processing in East Asian Languages* (pp. 85-111). Stanford, CA: CSLI Publications.

Mailhot, F., & Reiss, C. (2007). Computing long-distance dependencies in vowel harmony. *Biolinguistics 1*, 28-48. Retrieved from http://www.biolinguistics.eu/index.php/biolinguistics/article/view/14

Mandelbrot, B. B. (1977). *The Fractal Geometry of Nature*. New York, NY: W. H. Freeman and Company.

Miller, G. A., & Chomsky, N. (1963). Finitary Models of Language Users. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, Volume II* (pp. 419-491). New York, NY: John Wiley and Sons.

Miyamoto, E. T. (2002). Case markers as clause boundary inducers in Japanese. *Journal of Psycholinguistic Research 31*, 307-347. doi: 10.1023/A:1019540324040

Momo, K., Sakai, H., & Sakai, K. L. (2008). Syntax in a native language still continues to develop in adults: Honorification judgment in Japanese. *Brain and Language 107*, 81-89. doi: 10.1016/j.bandl.2007.12.003

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature Neuroscience 6*, 774-781. doi: 10.1038/nn1077

Nevins, A. (2010). *Locality in Vowel Harmony*. Cambridge, MA: The MIT Press.

O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2010). *Contemporary Linguistics: An Introduction, 6th Edition*. Boston, MA: Bedford/St. Martin's.

Ohta, S., Fukui, N., & Sakai, K. L. (2013a). Syntactic computation in the human brain: The degree of merger as a key factor. *PLOS ONE 8, e56230*, 1-16. doi: 10.1371/journal.pone.0056230

Ohta, S., Fukui, N., & Sakai, K. L. (2013b). Computational principles of syntax in the regions specialized for language: Integrating theoretical linguistics and functional neuroimaging. *Frontiers in Behavioral Neuroscience 7, 204*, 1-13. doi: 10.3389/fnbeh.2013.00204

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia 9*, 97-113. doi: 10.1016/0028-3932(71)90067-4

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America 108*, 2522-2527. doi: 10.1073/pnas.1018711108

Pattamadilok, C., Knierim, I. N., Duncan, K. J. K., & Devlin, J. T. (2010). How does learning to read affect speech perception? *Journal of Neuroscience 30*, 8435-8444. doi: 10.1523/JNEUROSCI.5791-09.2010

Penny, W. D., Stephan, K. E., Daunizeau, J., Rosa, M. J., Friston, K. J., Schofield, T. M., & Leff, A. P. (2010). Comparing families of dynamic causal models. *PLOS Computational Biology 6, e1000709*, 1-14. doi: 10.1371/journal.pcbi.1000709

Petersson, K. M., Forkstam, C., & Ingvar, M. (2004). Artificial syntactic violations activate Broca's region. *Cognitive Science 28*, 383-407. doi: 10.1207/s15516709cog2803_4

Röder, B., Stock, O., Neville, H., Bien, S., & Rösler, F. (2002). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance imaging study. *NeuroImage 15*, 1003-1014. doi: 10.1006/nimg.2001.1026

Rodriguez, P. (2001). Simple recurrent networks learn context-free and context-sensitive languages by counting. *Neural Computation 13*, 2093-2118. doi: 10.1162/089976601750399326

Saito, M., & Fukui, N. (1998). Order in phrase structure and movement. *Linguistic Inquiry 29*, 439-474. doi: 10.1162/002438998553815

Sakai, K. L. (2005). Language acquisition and brain development. *Science 310*, 815-819. doi: 10.1126/science.1113530

Santi, A., & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage 51*, 1285-1293. doi: 10.1016/j.neuroimage.2010.03.034

Saur, D., Kreher, B. W., Schnell, S., Kümmerer, D., Kellmeyer, P., Vry, M.-S., Umarova, R., Musso, M., Glauche, V., Abel, S., Huber, W., Rijntjes, M., Hennig, J., & Weiller, C. (2008). Ventral and dorsal pathways for language. *Proceedings of the National Academy of Sciences of the United States of America 105*, 18035-18040. doi: 10.1073/pnas.0805234105

Shibatani, M. (1990). *The Languages of Japan*. Cambridge, UK: Cambridge University Press.

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage 23*, S208-S219. doi: 10.1016/j.neuroimage.2004.07.051

Stephan, K. E., Marshall, J. C., Penny, W. D., Friston, K. J., & Fink, G. R. (2007). Interhemispheric integration of visual processing during task-driven lateralization. *Journal of Neuroscience 27*, 3512-3522. doi: 10.1523/JNEUROSCI.4766-06.2007

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage 49*, 3099-3109. doi: 10.1016/j.neuroimage.2009.11.015

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and Language 52*, 452-473. doi: 10.1006/brln.1996.0024

Suzuki, K., & Sakai, K. L. (2003). An event-related fMRI study of explicit syntactic processing of normal/anomalous sentences in contrast to implicit syntactic processing. *Cerebral Cortex 13*, 517-526. doi: 10.1093/cercor/13.5.517

Tatsuno, Y., & Sakai, K. L. (2005). Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands. *Journal of Neuroscience 25*, 1637-1644. doi: 10.1523/JNEUROSCI.3978-04.2005

Tsujimura, N. (2007). *An Introduction to Japanese Linguistics, 2nd Edition*. Malden, MA: Blackwell Publishing.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage 15*, 273-289. doi: 10.1006/nimg.2001.0978

Uehara, K., & Bradley, D. C. (2002). Center-embedding Problem and the Contribution of Nominative Case Repetition. In M. Nakayama (Ed.), *Sentence Processing in East Asian Languages* (pp. 257-287). Stanford, CA: CSLI Publications.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., Ogar, J. M., DeLeon, J., Miller, B. L., & Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron 72*, 397-403. doi: 10.1016/j.neuron.2011.09.014

Wong, F. C. K., Chandrasekaran, B., Garibaldi, K., & Wong, P. C. M. (2011). White matter anisotropy in the ventral language pathway predicts sound-to-word learning success. *Journal of Neuroscience 31*, 8780-8785. doi: 10.1523/JNEUROSCI.0999-11.2011

Yngve, V. H. (1960). A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society 104*, 444-466. Retrieved from http://www.jstor.org/stable/985230

Appendices

In the present study, we regarded the "DoM + number of Search" (i.e., adding the estimates of two factors) as an additional syntactic factor. Here we discuss the validity of adding these two factors. When Search applies to each syntactic object, the number of Search operations covaries with the DoM in most cases. For example, from a sentence "[[*The boy*$_1$ *sings*$_1$]$_0$" (subscripts denote the DoM for each node), the estimate of the DoM increases by one, while relevant features (numbers and persons here) are searched and matched between the nodes with the identical DoM. The estimates of the DoM and number of Search can be manipulated independently (see the Simple$_{(S)}$ and Conjoined$_{(S)}$ conditions in Figure 4A). Considering the possibility that the Merge and Search operations are additively applied, we added the metrics of the DoM and "number of Search" together. Among the nonlinguistic factors, we also regarded "memory span + counting" and "memory span + number of encoding" as additional factors, because they were temporal order-related and memory-related factors, respectively.

Formal linguistic theories have demonstrated that a pushdown automaton accepts and generates context-free languages, such as the counter language and mirror-image language (Chomsky, 1963). A "stack" (i.e., pushdown storage) is a critical component of the pushdown automaton (Searls, 2002), and we quantified the stack as counting in the present study.

In the present study, there was no region whose activations were modulated by the number of Search alone. As the Conjoined condition had equal or larger estimates of the number of Search than the other conditions, the number of Search may not be an explicit factor in the present paradigm. If another control condition is introduced, it would be possible to observe any effects of the number of Search separately from the DoM.

References

Chomsky, N. (1963). Formal Properties of Grammars. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, Volume II* (pp. 323-418). New York, NY: John Wiley and Sons.

Searls, D. B. (2002). The language of genes. *Nature 420*, 211-217. doi: 10.1038/nature01255

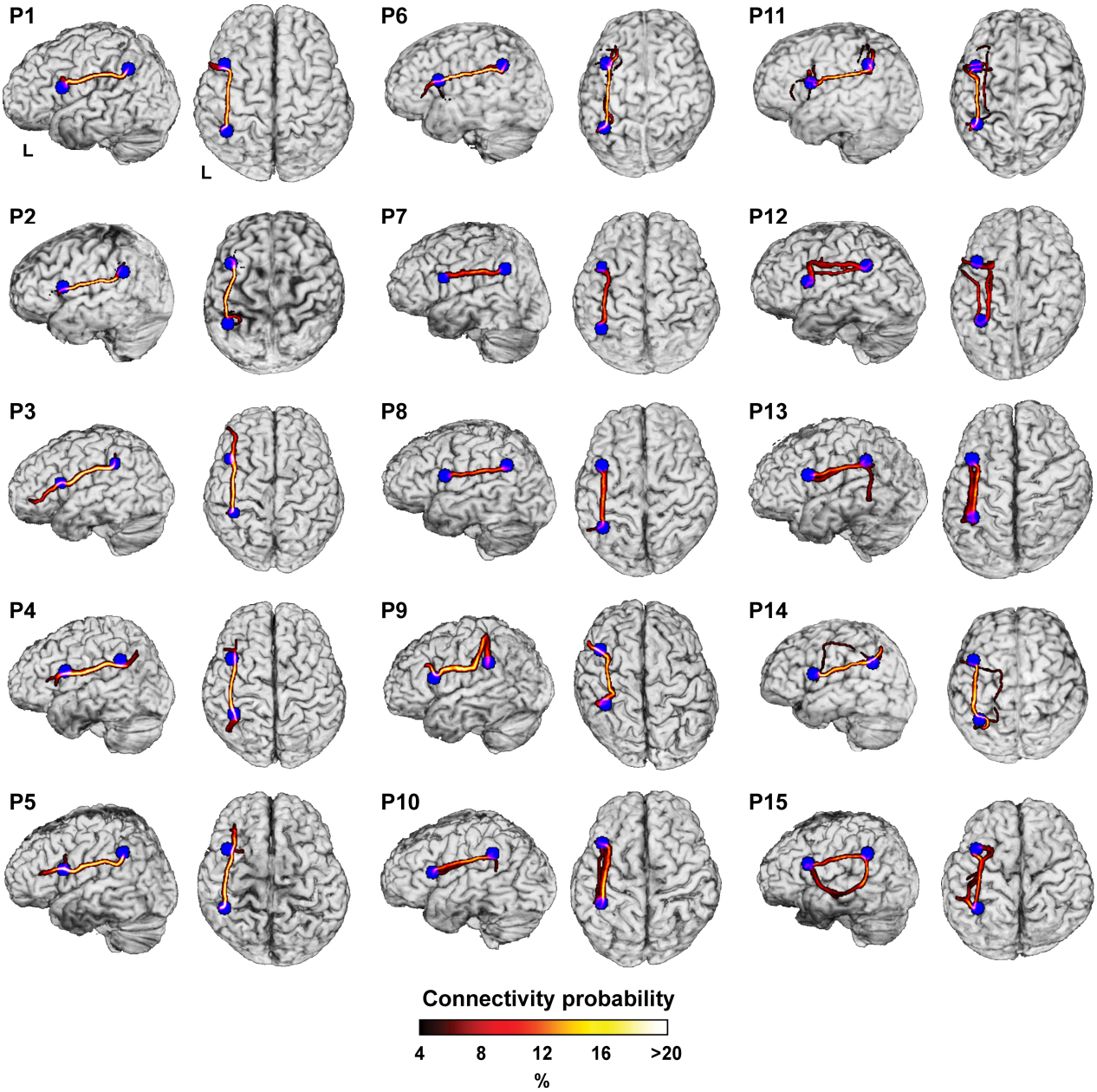


Figure S1. Anatomical connections between L. F3op/F3t and L. SMG for individual participants. Connectivity probability maps are shown on the left (L) lateral and dorsal surfaces of individual brains (P1–P15) with maximum intensity projection (cluster size = 629 ± 33 mm^3; cluster threshold, 30 mm^3). Blue spheres represent the seed regions of the L. F3op/F3t and L. SMG, the centers of which are activated foci (Figure 11C) transformed to individual brains. All participants showed the left dorsal pathway of the SLF/AF. For one participant (P15), the ventral pathway of the middle longitudinal fasciculus and extreme capsule was observed.