Decoding syntactic and prosodic structures from fMRI activity patterns

A.Bachrach[a,c,1], A.Gramfort[b], V.Michel[b], E.Cauvet[a], B.Thirion[b], C.Pallier[a,2]

[a]*INSERM, U992, Neuroimaging Unit, Neurospin center, Gif-sur Yvette, France*
[b]*INRIA Parietal, Neurospin center, Gif sur Yvette, France*
[c]*CNRS UMR 7023, Univ Paris 8,St Denis, France*

Abstract

One of the central open questions in the cognitive study of language regards the nature of the neural representation of abstract sentence structure (the syntax). To date it is still unclear whether syntactic structure by itself is coded in the brain (in a fashion accessible to neuroimaging techniques) or wether differences between different syntactic structures reflect other, related differences (e.g. semantics, working memory, prosody). In an fMRI experiment, we presented participants with different categories of utterances obtained by crossing three different types of syntactic configurations and two levels of prosodic embedding. In order to control for semantic, lexical and pragmatic factors, while retaining the ecological validity of the stimuli, we made use of a restricted fragment of French grammar, coordination structure. As a result, our conditions were optimally matched on all non-syntactic dimensions, permitting us to interpret eventual differences in brain activation in terms of differences in the syntactic structure per-se. Using a multivariate classification procedure, we show that it is possible to recover the syntactic construction, as well as the level of prosodic embedding, from the subjects' activation maps.

Is the syntactic structure of the linguistic input explicitly represented in the brain and if so how? This question, central to the cognitive neuroscience investigation of language, has important ramifications both for linguistic theory and to our understanding of neural computation more generally. Indeed, the neural representation of syntactic structure has been the object of numerous studies including both patient data (in particular Broca Aphasia, Thompson et al. 2013) and imaging studies (Friederici and Gierhan, 2013).

However, one main challenge remains largely unanswered in the current state of the art. Syntactic structure covaries with a number of other linguistic properties (semantics, working memory, frequency of use of the governing verb, etc.). For example, Stromswold et al. (1996) have famously contrasted, using PET, object and subject relatives (a syntactic distinction), however, as Caplan himself demonstrated in later work (Caplan et al., 2002), that original distinction was confounded with animacy distinctions, questioning the original interpretation of their 1996 results. Another pertinent example is the work by Makuuchi et al. (2009). In a previous paper (Bahlmann et al., 2008) the authors used an artificial grammar to distinguish between nested and un-nested dependencies, an important syntactic factor which has been a central topic in theoretical, empirical and computational linguistics. In the follow up paper the authors attempted to identify the brain correlates of the same distinction using this time a natural language (German). Nested dependencies were modeled using sentences containing complex subjects with doubly embedded relative clauses, while non-nested dependencies were modeled using simple declarative sentences with temporal and locative modifiers. Even if we adopt the syntactic assumptions of the authors, it is the case that these two sentence types differ also on multiple other dimensions beyond the purely syntactic configuration/processing (e.g.lexical content, semantics, pragmatics, working memory). These additional differences complicate the interpretation of any activation differences across the conditions as syntactically driven.

2

A particular dimension of syntactic structure that has been addressed from theoretical and computational linguistics (Kayne, 1994; Abney and Johnson, 1991) as well as psycholinguistics (Cheung and Kemper, 1992)) regards the configuration of embedding. It is common to distinguish purely right branching structures (structures where only right branches are complex) from structures with a complex left branch and structures with a complex internal branch (sometimes called center embedding). One important question (a specific version of the more general one we opened with) is whether this grammatical distinction per-se correlates with distinct neural patterns of activation and if so, whether current imaging tools can distinguish them. Two recent papers, using an habituation paradigm have failed to observed an habituation effect related purely to differences in the syntactic structure (Devauchelle et al., 2009; Santi and Grodzinsky, 2010).

To address this question we presented subjects three sets of linguistic phrases corresponding to the three structures mentioned above. We then used the imaging data to train a state of the art classification algorithm to identify the syntactic structure of a previously not seen input. We addressed the challenge of covariance discussed above by using a novel linguistic domain where it is possible to vary the syntactic structure of a linguistic expression while keeping most other properties identical. The material was chosen in an attempt to reduce to the minimum other, non-syntactic, differences between the three condition. In addition, we introduced some variability in the stimuli to assure that the successful classification would depend on the syntactic analysis rather then physical dimensions of the stimuli. Finally, we replaced the univariate subtraction logic with a multivariate classification procedure which is more sensitive in detecting 'brain-states' differences across the experimental conditions.

In order to reduce to a minimum the non syntactic differences across conditions we made use of a specific construction of French syntax, multi-membered coordination structure. Coordination is a universal and central feature of natural language (Haspelmath, 2004), pervasive both cross linguistically and across most if not all parts of speech (nouns, adjectives, adverbs, verbs and proposi-

	Syntactic Analysis	Prosodic Grouping
Complex right branch	[A [B [C D]]]	The dog — or the pig and the cow and the chicken
Complex left branch	[[A [B C]] D]	The dog — or the pig and the cow —— and the chicken
Complex center branch	[[A [[B C] D]]	The dog —— or the pig and the cow — and the chicken

Table 1: 3 different syntactic analyses of the coordinated phrase 'The dog or the pig and the cow and the chicken' and an example of an appropriate prosodic grouping

tions). In a number of languages coordination replaces clausal embedding found in languages such as English (Mithun, 1984). The advantage of the coordination fragment (a limited subset of syntax) is that syntactic structure can be manipulated with minimal effect on other linguistic dimensions (same lexical content, similar semantic content, no differences in long distance dependencies or other working memory parameters, arguably minimal frequency differences across structures).

Multiple-member Coordination structures allow for different realizations in term of perceptual grouping. 4 member structures can be realized in as many as 11 different ways. We based our paradigm on a fragment deriving all possible groupings in English coordination developed in Wagner (2010). That analysis, couched in the general framework of strictly binary branching phrase-structure grammar, derives the possible groupings via a mixed approach. According to this analysis certain boundaries are a reflex of syntactic constituency while other boundaries (sometimes referred to as optional) are purely prosodic,(that is, do not affect the syntactic parse). Based on a French adaptation of this fragment, we constructed exemplars of three syntactic configurations: complex right branch, complex middle branch (center embedding) and complex left branch, using the same lexical material (4 noun phrases). The resulting stimuli were used in a verification task that required correct parsing of the different abstract syntactic structures.

—————————-

insert Table 1 here

4

——————-

In table (1) we provide three different syntactic parsings for the same string of words. The string ' The dog or the pig and the cow and the chicken'' can be syntactically parsed in three different ways, depending on the perceived grouping. The length of the line separating the noun phrases stands for the relative strength of the grouping boundary (no line equals minimal boundary): If the three final elements are perceived to form a group, excluding the first member, the phrase is assigned an analysis with a complex right branch (first line of 1). A grouping of the three first elements with the exclusion of the 4th indicates a complex left constituent structure (second line of 1). If the two middle members are perceived to form a separate group from both the first and the last elements, the syntactic structure would contain a complex central constituent (last line of 1).

In French coordination (as well as in English and many other languages) , grouping is achieved via intonational modulation (e.g. length of pause, relative length of the final vowel in each group). In order to make sure that eventual brain differences across conditions reflect differences in the induced syntactic structure and not simply the intonational differences across conditions we introduced a second factor to our design: prosodic complexity. For each tree type we constructed a second realization in term of grouping. Both realizations reflected the same syntactic structure but the second realization induced the percept of an additional, embedded, non-syntactic, grouping. In table (2) we summarize the resulting 6 perceptual groupings formed by crossing syntax and prosodic complexity. Group 1 and Group 2, for example, both share the same syntactic structure (complex right branch) but differ with respect to prosodic embedding. Group 2 and group 4 represent different syntactic structures but both contain an embedded prosodic grouping : ——————-

Insert Table 2 here

——————-

In order to further insure that the eventual differences in brain activity pattern across conditions reflect the abstract syntactic and prosodic differences

Perceptual grouping	Condition	syntax	prosody
Group 1	syntax 1 prosody 2	[A [B [C D]]]	A —B C D
Group 2	syntax 1 prosody 2	[A [B [C D]]]	A —— B —C D
Group 3	syntax 2 prosody 1	[[A B C] D]]	A B C — D
Group 4	syntax 2 prosody 2	[[A B C] D]]	A — B C —— D
Group 5	syntax 3 prosody 1	[A [[B C] D]]	A — B C — D
Group 6	syntax 3 prosody 2	[A [[B C] D]]	A —— B C — D

Table 2: The six experimental conditions constructed by crossing the syntactic factor(1-3) and the prosodic factor (1-2). A-D stand for the noun phrases, the square brackets in the Syntax column represent constituency. The vertical lines in the prosody column represent perceptual groupings. An absence of a line indicates a minimal boundary, a single line indicates an intermediary boundary and a double line stands for a major boundary. Audio examples corresponding to the 6 different groupings can be found in the supplementary materials

rather then surface cues, we added two additional variables: A lexical variable (4 different sets of nouns) and logic (the use of one *or* and two *and*'s, or one *and* and two *or*'s), based on the finding that disjunction is more psychologically costly then coordination, (García-Madruga et al. 2001).

The use of a drastically reduced fragment of French grammar (which includes, in effect, only a single, recursive, rule) can be seen as a bridge between work making use of artificial grammars (Bahlmann et al., 2008) and work making use of the full power of natural language grammar. Artificial grammars are attractive in their simplicity and in the experimenter's ability to maximally control the different dimensions of the stimuli. However, as their names indicate, they are artificial and might or might not make use of the neural 'wetware' serving parallel natural language computations. The coordination fragment, while remaining within the bounds of natural language, shares with artificial grammars structural simplicity. As a consequence, we were able to manipulate the syntactic factor (branching structure) with minimal side effects.

Unlike most of the existing work on the neural representation of syntactic structure, the theoretical syntactic distinction we are addressing here is not

immediately associated with predictions regarding the relative amount of neural activity necessitated for the processing or encoding of each of the structures. While center embedding structures have been shown to be more costly to process than right branching structures (Stromswold et al., 1996), it is possible that this difficulty is due to the long distance relation formed between the verb and its argument (a factor eliminated in our design) rather than to the hierarchical structure per-se. Right and left branching structures are also distinguished in the computational processing literature, however the decision of which one is simpler to process depends on one's assumptions regarding the nature of the human parser mechanism (Abney and Johnson, 1991) which is still an open empirical question.

In effect, in the current experiment we posed a different question: can we detect differences in brain states associated with the processing of the different grouping structures. To test this hypothesis, we therefore chose to use a multivariate classification approach, that can exploit and combine statistical effects across brain regions. In such an approach (Cox and Savoy, 2003; Mitchell et al., 2004; Haynes and Rees, 2005; Kamitani and Tong, 2005; Norman et al., 2006), a classification algorithm (e.g. support vector machine, logistic regression, linear discriminant analysis) learns on a subset of the data (the training step) how to differentiate fMRI volumes corresponding to two different experimental conditions. In the test step, the classifier is used to predict the category, here the experimental condition, of unseen data. If the classifier does statistically better than chance, it is said to have successfully classified the two conditions given the data. The advantage of this technique is that it makes use of multivariate patterns in the data (across voxels) in order to be more sensitive than the subtraction approach (Davis and Poldrack, 2013).

Multivariate classification has been used to investigate neural correlates of sentence processing in a number of recent studies but primarily to distinguish intelligible from non intelligible speech (Evans et al., 2013) or grammatical from non-grammatical strings (Herrmann et al., 2012). To our knowledge only one other study (Allen et al., 2012) has used this method to distinguish the process-

ing of different syntactic structures. In that study the authors made use of the dative alternation in English (e.g. Jessica sold Mike a hot dog vs. Jessica sold a hot dog to Mike). The use of such alternation to distinguish different syntactic structures elegantly avoids differences in (open class) lexical items across conditions. However, the two variants in this alternation differ on a number of properties (apart from the syntactic difference) such as their frequency (with respect to each specific verb), sensitivity to animacy and detailed event structure (see (Bresnan et al., 2007) for a large scale corpus analysis). A second noteworthy aspect of the work here is that the classification was done at the group level (across subjects). In other words, the training is performed over a subset of the subjects and the testing on the remaining subjects.

1. Materials and method

1.1. Stimuli

Auditory stimuli:

We constructed a total of 48 French phrases of the form "A x B y C z D", where A-D stand for nouns and x, y, z stand for coordinators 'et' (*and*) or 'ou' (*or*), 8 per each of the six experimental conditions described in table (2). The six experimental conditions were created by crossing two factors, Syntactic structure (1) and depth of prosodic embedding (2):

(1) Branching structure:

- Complex right branch (syntax 1): [A [B [C D]]]

- Complex left branch (syntax 2): [[A [BC]] D]

- Complex middle branch (syntax 2I): [A [[BC] D]]

(2) Prosodic embedding: Each phrase contained either a single level of prosodic embedding (prosody 1) or two levels (prosody 2)

In order to introduce variability within each condition, four sets of noun phrases (furniture, animals, colors and food items) were used (12 phrases for
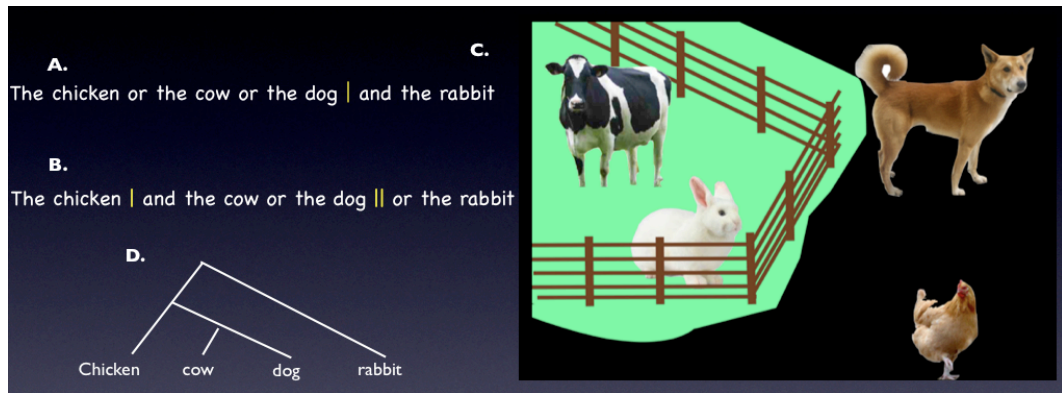
8

Figure 1: A schematic representation of a true (A) and a false (B) phrase given a particular scene (C). The target region is the green region of the image. The two phrases share the same right branching structure (D) but differ with respect to prosodic embedding. The phrase in A contains a single prosodic embedding, while the phrase in B contains two levels of embedding.

each set of nouns). In addition, half of the phrases in a given condition contained 2 instances of *and* and one instance of *or*, and the other half 2 instances of *or* and one instance of *and*. Noun phrases and connector distribution varied equally across the 6 conditions.

The phrases were produced by a trained native speaker (duration: mean=3.4s; std. dev.=0.52s). A second native speaker verified that the groupings were perceived as intended.

Visual stimuli:

On each trial, an auditory phrase was associated with a visually presented image. The image contained instances of the 4 nouns mentioned in the phrase. A subset of these appeared in a target location on the image. 96 images were constructed; two images per auditory phrase. Only for one of these two images the auditory phrase correctly designated the subset of items in the target location (i.e a true response to the question: 'which items are in the target location?', see Fig. 1).

————————-

Insert figure 1 here

9

—————————-

1.2. Participants:

28 native speakers of French (10 Females) between the ages of 18 and 36 participated in the experiment. Subjects were compensated for their participation.

1.3. Procedure:

In each trial, the image appeared at the onset of the auditory phrase and remained on the screen for 6 seconds. The subject was asked to judge whether the phrase constituted a true and complete response to the question 'Which of the 4 items is in the target location?'. The subject was instructed to press the right button for a true response and otherwise the left button ('false'). Reaction times were measured from the offset of the stimuli. The image/phrase pairs were constructed as to delay as much as possible the point in the phrase where the correct answer can be determined. During the experiment itself no feedback was provided. Before entering the scanner, participants went through a short training session were the task was explained and the 4 different scenes and target regions were introduced. The subjects then practiced the task itself first at a self-paced rhythm with feedback (and when needed clarifications by the experimenter) and then without feedback and in a rhythm similar to the one in the actual experiment (12 training trials).

We chose picture verification as the experimental task since, while being rather naturalistic and not meta-linguistic, it permitted us to verify that subjects have constructed the intended parse at each trial. The question ('Which of the 4 items are in the target location?') was introduced during the training and was not repeated at each trial. As discussed above, the auditorily presented response to the question consisted of (only) a complex noun phrase, or a fragment, rather then a full sentence (e.g. "in the target region there are X and B or C and D"). fragment answers (e.g. Q: "who will carry the sofa?" A: "John and Peter or Bill and Dave") are common in naturalistic conversations (Fernandez

and Ginzburg, 2002). While theoreticians have proposed competing analyses of fragment answers (Merchant, 2005), these differences are not pertinent to the manipulation used here.

1.4. fMRI design

Scanning was conducted in a Siemens 3T Trio scanner at our center in 2 runs. Each run lasted 6 minutes and included 48 trials with an average SOA of 7.5 seconds, jittered using Optseq (http://surfer.nmr.mgh.harvard.edu/optseq/, 1.5 seconds jitter). We acquired whole brain EPI images (TR=2s) at a resolution of 3*3*3 mm using online gradient correction for motion (PACE). A high resolution (1*1*1mm) MPRAGE T1 structural image was also acquired for each participant.

1.5. fMRI data analyses

1.5.1. Computation of individual subject maps

SPM8 (Penny et al., 2007) was used for all image preprocessing stages and first level (within subject) analysis. The EPI images of each subject were movement corrected and coregistered with the subject's structural image. The resulting images were then spatially normalized (using non-linear warps) with respect to the SPM avg152 template from the Montreal Neurological Institute. The images were then smoothed with an isotropic Gaussian kernel (FWHM=4mm).

Single subject data were analysed using a General Linear Model with 7 regressors for the 6 experimental conditions and a condition grouping all the incorrect trials). Each event was modeled as a box car lasting 2 seconds with its onset set to 2 seconds prior to response time. The choice to use reaction time rather than the onset of the stimulus as a temporal anchor was taken in order to take into account the variance due to this factor but also because we were primarily interested in brain activity related to the computation of the syntactic structure rather than perceptual process (Caplan et al., 2001). The events were further modeled using parametric modulators for response hand and duration of auditory input. For each subject we thus obtained 6 contrast

images corresponding to the 6 experimental conditions (i-vi). These images were then entered into the group analysis performed with a multivariate classification method.

In order to have the same number of voxels for all subjects, the 6 contrast images from each subject were masked by the intersection of the individual masks produced by the SPM first level analysis. Some regions, defined by AAL (Amygdala, Olfactory, Occipital, Cuneus) were removed from the analysis resulting in a mask formed of $54, 808$ voxels.

1.5.2. Multivariate decoding

The maps obtained from the single subjects were used as an input for several group-level tests using a multivariate decoder. fMRI decoding relies on a prediction function that is plagued by the curse of dimensionality; the number of features (voxels, regions) is much larger ($\sim 10^5$) than the numbers of samples (images) ($\sim 10^2$). Consequently, inference methods are prone to overfitting, thus leading to poor performance when trying to predict labels for new data.

To address this issue, a standard solution consists in reducing the number of features (or voxels) prior to classification (Haynes and Rees, 2006). This step is commonly known as dimensionality reduction. This can be done by selecting the most informative features in the training set, *i.e.* performing a *feature selection*. To date, the most widely used method for feature selection is voxel-based Anova (Analysis of Variance), that evaluates each brain voxel independently. The selected features can be redundant, and are not constrained by spatial information, thus can be spread in large regions within the whole brain without any emerging structure. More fundamentally such a dimensionality reduction assumes that information at the voxel level is relevant to achieve a prediction at the subject level. Unfortunately, due to anatomical variabilities between subjects as well as imperfect normalization procedures, there is no perfect voxel-by-voxel correspondence between 2 subjects.

In the context of inter-subjects decoding, a better way of reducing the number of features is based on *features agglomeration* rather than features selection.

12

It consists in replacing voxel-based signals by local averages (*a.k.a. parcels*) (Flandin et al., 2002; Thirion et al., 2006; Michel et al., 2010). These parcel-based averages of fMRI signals are then used to fit a predictive model from a reduced number of features. Starting from approximately $50,000$ voxels the number of parcels ranges from a few hundreds to a few thousands.

Features agglomeration is performed using Ward's algorithm (Ward, 1963), which is a model-free hierarchical agglomerative clustering method. This algorithm is based on a variance-minimizing criterion, in order to ensure that parcels averages provide an accurate representation of the signal. Connectivity constraints are added to this algorithm, so that it takes into account both spatial and signal information. Estimated parcels thus form interpretable clusters of voxels, while parcels averages provide an accurate representation of the data.

The mean signal in each parcel is then used as input for a sparse logistic regression classification (Hastie et al., 2003; Ryali et al., 2010) implemented as described in Fan et al. (2008). The procedure is evaluated with a leave-one-subject-out cross-validation. Each training set contains the data of 21 subjects and the test set consists of the left out subject.

The classifier is parametrized by the number of parcels for the clustering step and the regularization parameter in the logistic regression. Both parameters are fit on the training set using a 3 folds cross-validation procedure (7 subjects per fold). The number of clusters can vary between 400, 800, 1200 and 1600 while the regularization of the logistic regression is optimized on a logarithmic grid of 10 values between 1 and 10^7. The method is known as nested cross-validation.

Data from the 22 participants included according to their behavioral results were used for the classification analysis. Four classification tests were conducted: Prosody 1 vs. Prosody 2 (on 6*22=132 images), Syntax 1 vs. Syntax 2 (on 4*22=88 images), Syntax 1 vs. Syntax 3 (on 88 images) and Syntax 2 vs. Syntax 3 (on 88 images). For each test, the accuracy of the classifier was determined as the percentage of successful classifications in the test stage. With balanced classes as it is the case in this study the chance level is 50%.

2. Results

2.1. Behavioral data

After inspecting the distribution of hit rates across participants (see supplementary Figure S1), we decided to exclude the subjects who had a performance inferior to 70% to keep only those who we could be reasonably confident parsed the stimuli correctly (22 subjects out of 28).

Average hit rates and reaction times were computed for each participant in each of the six conditions obtained by crossing the factors Syntax (3 levels) and Prosody (2 levels). The groups means are displayed on Figure 2. Analyses of Variance with two within-subjects factors, Syntax and Prosody were performed on these data. An arcsin transform was applied to hit rates and Greenhouse-Geisser's correction for nonsphericity was applied to p-values. In the analysis of hit rates, both factors interacted significantly ($F(2,42)=3.9$; $p<.05$) and produced main effects (Syntax $F(2,42)=3.3$; $p<.05$; Prosody: $F(1,21)=30.1$ $p<.001$). This pattern is due to the fact that, overall, Prosody P2 was more difficult than Prosody P1, and that the effect of Prosody was less marked for the second syntactic configuration than for the others. In the analysis of reaction times, a main effect of Syntax was observed ($F(2,42)=7.5$; $p<.01$) but must be qualified by a marginal interaction between Syntax and Prosody ($F(2,42)=2.7$: $p=.08$) reflecting the fact that the effect of Syntax was mostly due to a trend to slow down for the first configuration (complex right branch) in Prosody 2 compared to Prosody 1 ($t(21)=2.0$; $p=0.6$) (see Fig. 2).

————————————

Insert figure 2 here

————————————-

2.2. fMRI data

The classification algorithm correctly classified images as belonging to either prosody 1 or prosody 2 on 68% of the cases (p-value=0.00002 according to an Exact Binomial Test). It was also able to distinguish between images corresponding to syntax 1 and syntax 2 (69%; p-value=0.0002), syntax 1 and syntax
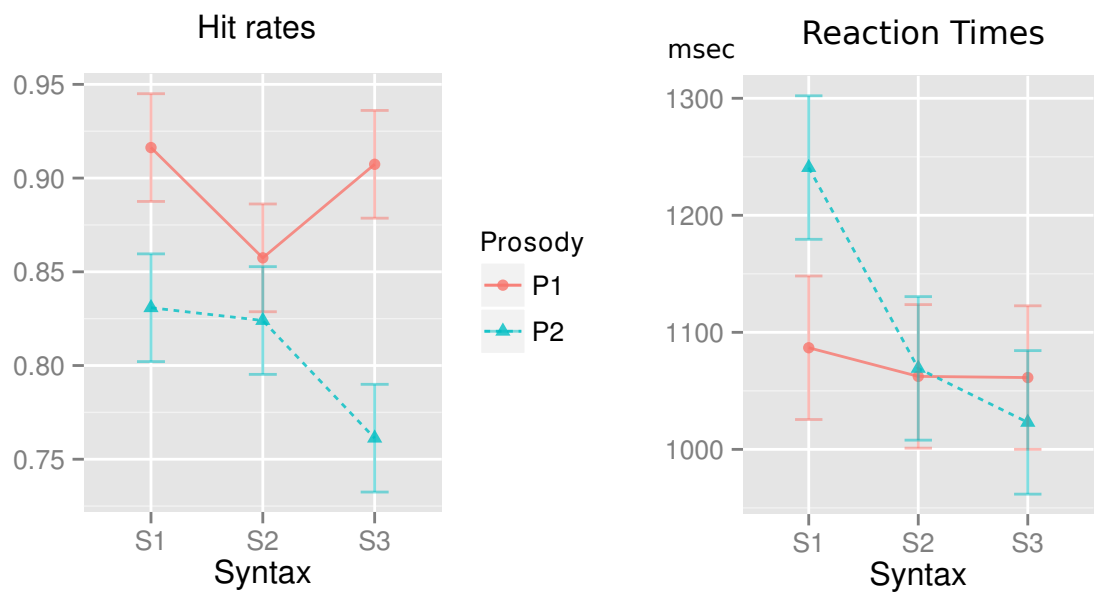
14

Figure 2: Average hit rates and reaction-times, as a function of Syntactic Configuration and Prosody (1 or 2 prosodic embeddings). Error bars represent Fisher Least Significant Difference computed from the ANOVAs with Syntax and Prosody as within-subjects factors.
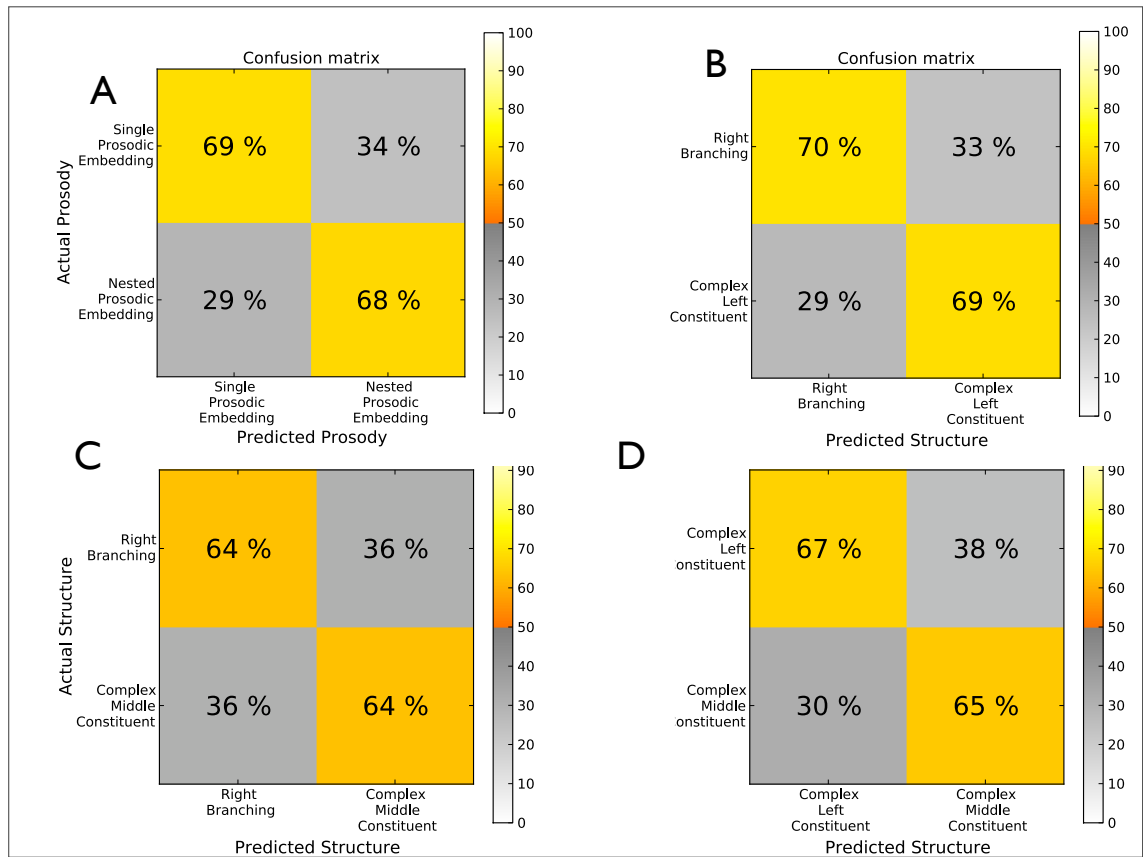
Figure 3: Confusion matrices for the prosody classification (A), and the three syntactic classifications (B-D).

3 (64%; p-value=0.007) and between syntax 2 and 3 (66%; p-value=0.002). The confusion matrices for the four classification tests are displayed in Fig.3.

————————

Insert figure 3 here

————————

Subjects were presented with 6 grouping types. Each of the 6 types varied internally with respect to the lexical items used and the combination of the conjunctures. Our theoretical assumptions predicted two possible classifications (three ways by syntax and two ways by prosody). In order to strengthen our

16

Contrast	classification score	p-value
syntax 1 vs. syntax 2 (1+2 vs 3+4)	0.693	0.0002
syntax 2 vs. syntax 3 (3+4 vs 5+6)	0.659	0.0018
syntax 1 vs. syntax 3 (1+2 vs 5+6)	0.647	0.003
1+6 vs. 2+5	0.636	0.006
1+6 vs. 2+3	0.590	0.054
1+4 vs 3+6	0.557	0.168
2+3 vs. 4+5	0.545	0.227
1+6 vs. 4+5	0.534	0.297
1+4 vs. 2+3	0.522	0.374
1+4 vs. 2+5	0.477	0.702

Table 3: Classification results for the 3 classes predicted by the theory and the 6 classes not predicted by the theory ordered by classification score.

theoretical interpretation of eventual successful classification of the 6 types into 3 syntactic classes we also tested our classifier on 6 classes (pairings of types) not predicted by our theoretical model. In table 3 we present the classification results for both theoretically motivated and unmotivated classes ordered by success score. As it is easy to see, classification scores for the 3 theoretically motivated classes were superior to all other tests. The probability that the predicted 3 classifications would produce better results then the non-predicted classifications by chance (that is, a false positive) is 0.011.

———————

Insert Table 3

———————

3. Discussion

The main finding of these experiment is that it is possible, from fMRI activation patterns, to classify syntactic configurations better than chance. We will first discuss behavioral data and then will come back to this result.

Concerning behavioral data, we had to discard one fifth of the subjects due to low hit rates. These low scores could be attributed to difficulties to perceive the auditory cues to the syntactic structure of the stimuli in the less than ideal MRI scanner environement. We also think that the session of familiarization with the task before entering the scanner may have been too short (12 trials). An indication that this may be the case is that performance improved significantly between the two within-scanner sessions (from 77% to 83%, t(27)=3.8; p¡.001).

Focusing now on the effects of the manipulated factors on behavior, the most remarkable result is that of the additional prosodic embedding on accuracy. This effect could reflect the fact that the inference of an embedded grouping requires a 3-way prosodic boundary strength distinction (and only a 2-way distinction for the non-embedded condition). Multiple studies have produced evidence for the sensitivity of hearers to the relative strength of a prosodic boundary during online parsing (Frazier et al., 2006). However, this is the first study we are aware of to demonstrate that the addition of a strength distinction can produce increased cognitive cost.

Interestingly, this effect of the additional prosodic boundary depended on the syntactic context. The effect was the larger for the complex middle branch (center embedding) conditions (syntax 3). Center embedding structures have been long considered particularly difficult to parse, compared to similar right branching structures (Miller and Isard, 1964; Stromswold et al., 1996):

(3)   The woman saw the boy that heard the man that left. (Right Branching)

(4)   The man the boy the woman saw heard left. (Center Embedding)

In their discussion of the contrast between (3) and (4), Weckerly and Elman (1992) list a number of the explanations given to this effect in the literature: violation of word order canonicity, difficulty in association between the matrix verb and the matrix subject and the distance between the verbs and their respective subjects.

None of these factors could explain the result here since there are no subject-verb relations or differences in word order. Instead, the finding here could

suggest a novel perspective on this issue, namely specific difficulty in forming non-edge embedded prosodic groupings. Such a cognitive factor would explain the interaction observed here and could be potentially extended to cases such as (4) where the grouping structure corresponding to the desired interpretation also exhibits a non-edge embedded prosodic embedding:

(5)   The man – the boy- the woman saw - heard – left.

The question why processing of such prosodic structures should be more difficult requires further research. However, this interpretation of the interaction effect suggests that the main prosodic effect is not due simply to the existence of additional boundary strength contrast but to the particular grouping pattern that that prosodic embedding, coupled with a complex center branch structure, brings about.

The analysis of reaction time produced a distinct picture. Prosody did not have any effect while syntactic structure did, with responses to the complex right branch slower then to the two other structures. The fact that hit rate for right branching structures did not induce any particular difficulty (numerically this condition had the highest hit rate) suggests that slower RT does not represent additional global cognitive load. A possible explanation of this effect could be in terms of parsing. As can be clearly observed from the difference in the number of right brackets at the right edge of the syntactic representations in table (1), in complex right branching structures, all the syntactic constituents are 'closed off' at the right edge of the parsed string. If closing off of constituents has an (additive) temporal cost this could explain the increased RT result.

Turning to the fMRI data, using a novel linguistic paradigm, we were able to train a classifier to distinguish between brain states corresponding to the processing of distinct syntactic structures and prosodic configurations. While classification has been used previously for decoding linguistic properties (Mitchell et al., 2008), this is the first demonstration of classification task applied to classes of complex abstract objects (syntactic and prosodic representations of phrases). The classification procedure used here, itself, was novel since it was performed

19

at the group level (rather than intra-subject) and used feature agglomeration (clustering) rather than feature reduction.

Santi and Grodzinsky (2010) used an fMRI habituation paradigm to explore the neural correlates of syntactic movement (by alternating between subject and object extraction) and branching structure (by alternating between complex left branch and complex right branch). In apparent contradiction to the findings here, the authors report no dishabituation due solely to alternation between right and left branching structures. This difference could be due to the fMRI paradigm itself (it is possible that the processing of syntactic structure per-se is not subject to habituation, (Devauchelle et al., 2009)), the materials used (these authors used copular constructions of the type 'John is the thief ' vs. 'the thief is John', which are arguably surface variants of the same deep structure copular construction, Heycock 1994) or the task used (unlike the task used here, the task used in the cited paper did not force subjects to fully parse the sentences).

While the topic of the neural representation of prosody has been investigated by numerous brain imaging studies, most have concentrated on the distinction between emotional and sentential prosody (Wildgruber et al., 2009). Of the few papers specifically interested in sentential (or linguistic) prosody a number have made use of altered speech (where prosodic information has been eliminated from the signal) to create subtractive designs (+ vs. -prosody, Humphries et al. 2005; Meyer et al. 2004). The inherent problem of such designs is the interpretation of brain response to such highly unnatural stimuli. There are only two published fMRI papers that have manipulated sentential prosody in a naturalistic context (Doherty et al., 2004; Ischebeck et al., 2008). The former has focused on information structure aspects (question versus assertion), a different domain from the one manipulated here. Ischebeck et al. (2008) have manipulated the presence of an embedded prosodic boundary and as such is the closest antecedent to the work reported here. There are two significant differences though. First, the manipulation there co-varied the syntactic structure and the prosodic structure which were orthogonalized here. The second difference is related to the abstraction level. The stimuli in that experiment varied along a

constant acoustic level property while the stimuli here varied along an abstract level representation of prosodic structure (while the specific acoustic level features were different). To our knowledge, this is the first non-confounded fMRI evidence for the brain correlates of an embedded prosodic boundary or grouping.

As in most neuroimaging studies on sentence processing or syntactic structure, we made use of theoretical assumptions or classifications established by theoretical linguists or grammarians (e.g. the dative alternation study by Allen et al. 2012). However, as argued for by Ferreira (2005) (and others) data from neuroimaging (and other neuro-cognitive methods) should in potential inform linguistic theory as well. Indeed, our results do provide preliminary evidence in favor of a specific linguistic theory of coordination (and syntactic dependencies more generally).

While our materials included six different grouping structures (out of the 11 possible ones), the grouping of these conditions into 3 syntactic types is a consequence of the specific assumptions of our analysis; namely the restriction to binary branching in Wagner 2010 and more generally in contemporary generative theory (Chomsky, 1994). An alternative theory that permits multiple branching would not necessarily predict the same clustering of the 6 grouping structures into 3 classes. For example, as can be seen in table (4), according to the binary branching theory assumed here, groupings 5 and 6 share the same complex middle branch syntactic representation, while groupings 1 and 2 share a complex right branch structure. However, the multiple branching representation of these 4 groupings does not offer a principled explanation for the successful classification (1&2 vs. 5&6) observed here.

―――――――――――-

Insert table 4 here

―――――――――――

The fact that our classifier produced better classification results for the 3 classes we have hypothesized, compared to other logically possible groupings (table 3) supports the theoretical approach to coordination structure developed by Wagner (2010) and adapted here. In other words, our classification results do

Perceptual grouping	prosody	binary branching	multiple branching
Group 1	A —B C D	[A [B [C D]]]	[A [B C D]]
Group 2	A —— B —C D	[A [B [C D]]]	[A [B [C D]]]
Group 5	A — B C — D	[A [[B C] D]]	[A [BC] D]
Group 6	A —— B C — D	[A [[B C] D]]	[A [[B C] D]]

Table 4: Hypothesized syntactic structures for groupings 1,2,5,6 under the binary branching and multiple branching theories of coordination. A-D stand for the noun phrases, the square brackets in the Syntax columns represent constituency. The vertical lines in the prosody column represent perceptual groupings. An absence of a line indicates a minimal boundary, a single line indicates an intermediary boundary and a double line stands for a major boundary.

not only distinguish between different syntactic parses of the same string, but also provide preliminary support for a particular theoretical analysis of these differences.

Table legends:

Table 1: 3 Different syntactic analyses of the coordinated phrase 'The dog or the pig and the cow and the chicken' and an example of an appropriate prosodic grouping

Table 2: The six experimental conditions constructed by crossing the syntactic factor(1-3) and the prosodic factor (1-2). A-D stand for the noun phrases, the square brackets in the Syntax column represent constituency. The vertical lines in the prosody column represent perceptual groupings. An absence of a line indicates a minimal boundary, a single line indicates an intermediary boundary and a double line stands for a major boundary.

Table 3: Classification results for the 3 classes predicted by the theory and the 6 classes not predicted by the theory ordered by classification score.

Table 4: Hypothesized syntactic structures for groupings 1,2,5,6 under the binary branching and multiple branching theories of coordination. A-D stand for the noun phrases, the square brackets in the Syntax columns represent constituency. The vertical lines in the prosody column represent perceptual groupings. An absence of a line indicates a minimal boundary, a single line indicates an intermediary boundary and a double line stands for a major boundary.

Figure legends:

Figure 1:A schematic representation of a true (A) and a false (B) phrase given a particular scene (C). The target region is the green region of the image. The two phrases share the same right branching structure (D) but differ with respect to prosodic embedding. The phrase in A contains a single prosodic embedding, while the phrase in B contains two levels of embedding.

Figure 2: Average hit rates and reaction-times, as a function of Syntactic Configuration and Prosody (1 or 2 prosodic embeddings). Error bars represent Fisher Least Significant Difference computed from the ANOVAs with Syntax and Prosody as within-subjects factors.

Figure 3: Confusion matrices for the prosody classification (A), and the three syntactic classifications (B-D).

Abney, S. P., Johnson, M., 1991. Memory requirements and local ambiguities of parsing strategies. Journal of Psycholinguistic Research 20, 233–250, 10.1007/BF01067217.

Allen, K., Pereira, F., Botvinick, M., Goldberg, A. E., Dec. 2012. Distinguishing grammatical constructions with fMRI pattern analysis. Brain and Language 123 (3), 174–182, 00009.
URL http://www.sciencedirect.com/science/article/pii/S0093934X12001599

Bahlmann, J., Schubotz, R. I., Friederici, A. D., 2008. Hierarchical artificial grammar processing engages broca's area. NeuroImage 42 (2), 525 – 534.

Bresnan, J., Cueni, A., Nikitina, T., Baayen, R. H., others, 2007. Predicting the dative alternation. Cognitive foundations of interpretation, 69–9400417.
URL http://www.ehu.es/seg/$_m$edia/morf/5/5/6/sak/predicting$_t$he$_d$ative$_a$lternation.pdf

Caplan, D., Vijayan, S., Kuperberg, G., West, C., Waters, G., Greve, D., Dale, A. M., 2001. Vascular responses to syntactic processing: Event-related fmri study of relative clauses. Human Brain Mapping 15 (1), 26–38.

Caplan, D., Vijayan, S., Kuperberg, G., West, C., Waters, G., Greve, D., Dale, A. M., 2002. Vascular responses to syntactic processing: Event-related fMRI study of relative clauses. Human Brain Mapping 15 (1), 26–38.

Cheung, H., Kemper, S., 1992. Competing complexity metrics and adults. Applied Psycholinguistics 13 (1), 53–76.

Chomsky, N., 1994. Bare phrase structure. Massachusetts Institute of Technology, Cambridge.

Cox, D. D., Savoy, R. L., 2003. Functional magnetic resonance imaging (fMRI) "brain reading": detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19, 261–270.

Davis, T., Poldrack, R. A., 2013. Measuring neural representations with fMRI: practices and pitfalls. Annals of the New York Academy of Sciences 1296 (1), 108–134.

Devauchelle, A.-D., Oppenheim, C., Rizzi, L., Dehaene, S., Pallier, C., 2009. Sentence syntax and content in the human temporal lobe: An fmri adaptation study in auditory and visual modalities. Journal of Cognitive Neuroscience 21 (5), 1000–1012.

Doherty, C. P., West, W. C., Dilley, L. C., Shattuck-Hufnagel, S., Caplan, D., 2004. Question/statement judgments: an fmri study of intonation processing. human brain mapping 23 (2), 85–98.

Evans, S., Kyong, J. S., Rosen, S., Golestani, N., Warren, J. E., McGettigan, C., Mourão-Miranda, J., Wise, R. J. S., Scott, S. K., Apr. 2013. The pathways for intelligible speech: Multivariate and univariate perspectives. Cerebral Cortex, bht08300004 PMID: 23585519.
URL http://cercor.oxfordjournals.org/content/early/2013/04/12/cercor.bht083

Fan, R.-E., Chang, K.-W., Hsieh, C.-J., Wang, X.-R., Lin, C.-J., 2008. LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research 9, 1871–1874.

Fernandez, R., Ginzburg, J., 2002. Non-sentential utterances: A corpus study. TAL. Traitement automatique des langues 43 (2), 13–42.

Ferreira, F., 2005. Psycholinguistics, formal grammars, and cognitive science. The Linguistic Review 22 (2-4), 365–380, 00060.
URL http://www.degruyter.com/view/j/tlir.2005.22.issue-2-4/tlir.2005.22.2-4.365/tlir.2005

Flandin, G., Kherif, F., Pennec, X., Malandain, G., Ayache, N., Poline, J.-B., 2002. Improved detection sensitivity in functional MRI data using a brain parcelling technique. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI'02). pp. 467–474.

Frazier, L., Carlson, K., Clifton Jr, C., 2006. Prosodic phrasing is central to language comprehension. Trends in Cognitive Sciences 10 (6), 244–249.

Friederici, A. D., Gierhan, S. M., Apr. 2013. The language network. Current Opinion in Neurobiology 23 (2), 250–254.

García-Madruga, J. A., Moreno, S., Carriedo, N., Gutiérrez, F., Johnson-Laird, P. N., 2001. Are conjunctive inferences easier than disjunctive inferences? a comparison of rules and models. The Quarterly Journal of Experimental Psychology Section A 54 (2), 613–632, PMID: 11394065.

Haspelmath, M. (Ed.), 2004. Coordinating constructions. No. 58 in Typological studies in language, ISSN 0167-7373. Amsterdam, Pays-Bas, Etats-Unis.

Hastie, T., Tibshirani, R., Friedman, J. H., 2003. The Elements of Statistical Learning. Springer.

Haynes, J.-D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. Nature Neuroscience 8 (5), 686–691.

Haynes, J.-D., Rees, G., 2006. Decoding mental states from brain activity in humans. Nature Reviews Neuroscience 7 (7), 523–534.

Herrmann, B., Obleser, J., Kalberlah, C., Haynes, J.-D., Friederici, A. D., 2012. Dissociable neural imprints of perception and grammar in auditory functional imaging. Human brain mapping 33 (3), 584–595, 00010.
URL http://onlinelibrary.wiley.com/doi/10.1002/hbm.21235/full

Heycock, C., 1994. Layers of Predication: the Non-Lexical Syntax of clauses. Garland.

Humphries, C., Love, T., Swinney, D., Hickok, G., 2005. Response of anterior temporal cortex to syntactic and prosodic manipulations during sentence processing. Human Brain Mapping 26 (2), 128–138.

Ischebeck, A. K., Friederici, A. D., Alter, K., 2008. Processing prosodic boundaries in natural and hummed speech: An fmri study. Cerebral Cortex 18, 541–552.

Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. Nature Neuroscience 8 (5), 679–685.

Kayne, R., 1994. The antisymmetry of syntax. MIT Press.

Makuuchi, M., Bahlmann, J., Anwander, A., Friederici, A. D., May 2009. Segregating the core computational faculty of human language from working memory. Proceedings of the National Academy of Sciences 106 (20), 8362–8367, PMID: 19416819.

Merchant, J., Jan. 2005. Fragments and ellipsis. Linguistics and Philosophy 27 (6), 661–738.

Meyer, M., Steinhauer, K., Alter, K., Friederici, A. D., von Cramon, D., May 2004. Brain activity varies with modulation of dynamic pitch variance in sentence melody. Brain and Language 89 (2), 277–289.

Michel, V., Eger, E., Keribin, C., Poline, J.-B., Thirion, B., 2010. A supervised clustering approach for extracting predictive information from brain activation images. MMBIA'10.

Miller, G. A., Isard, S., Sep. 1964. Free recall of self-embedded english sentences. Information and Control 7 (3), 292–303.

Mitchell, T. M., Hutchinson, R., Niculescu, R. S., Pereira, F., Wang, X., Just, M., Newman, S., 2004. Learning to decode cognitive states from brain images. Machine Learning V57 (1), 145–175.

Mitchell, T. M., Shinkareva, S., Carlson, A., Chang, K., Malave, V., Mason, R., Just, M., 2008. Predicting human brain activity associated with the meanings of nouns. Science 320, 1191–5.

Mithun, M., Sep. 1984. How to avoid subordination. Proceedings of the Annual Meeting of the Berkeley Linguistics Society 10 (0).

Norman, K. A., Polyn, S. M., Detre, G. J., Haxby, J. V., 2006. Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends Cogn Sci 10 (9), 424–430.

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., 2007. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Elsevier.

Ryali, S., Supekar, K., Abrams, D. A., Menon, V., 2010. Sparse logistic regression for whole-brain classification of fMRI data. NeuroImage 51 (2), 752 – 764.

Santi, A., Grodzinsky, Y., 2010. fmri adaptation dissociates syntactic complexity dimensions. NeuroImage 51 (4), 1285–1293.

Stromswold, K., Caplan, D., Alpert, N., Rauch, S., 1996. Localization of syntactic comprehension by positron emission tomography. Brain and Language 52 (3), 452–473.

Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B., 2006. Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27 (8), 678–693.

Thompson, C. K., Meltzer-Asscher, A., Cho, S., Lee, J., Wieneke, C., Weintraub, S., Mesulam, M.-M., Jan. 2013. Syntactic and morphosyntactic processing in stroke-induced and primary progressive aphasia. Behavioural Neurology 26 (1), 35–54.

Wagner, M., 2010. Prosody and recursion in coordinate structures and beyond. Natural Language & Linguistic Theory 28, 183–237, 10.1007/s11049-009-9086-0.

Ward, J. H., 1963. Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58 (301), 236–244.

Weckerly, J., Elman, J. L., 1992. A PDP approach to processing center-embedded sentences. In: Proceedings of the fourteenth annual conference of the Cognitive Science Society. Vol. 14. p. 414.

Wildgruber, D., Ethofer, T., Grandjean, D., Kreifelts, B., 2009. A cerebral network model of speech prosody comprehension. International Journal of Speech-Language Pathology 11 (4), 277–281.