Functional MRI Investigation on Paradigmatic and Syntagmatic Lexical Semantic Processing

Songsheng Ying[1][2], Sabine Ploux[2]*

*Corresponding author: sabine.ploux@ehess.fr

[1]*SJTU-ParisTech Elite Institute of Technology, Shanghai Jiao Tong University*

[2]*Centre d'Analyse et de Mathématique Sociales CNRS-UMR 8557,*

Ecole des Hautes Etudes en Sciences Sociales, Paris, France

Abstract—The word embeddings related to paradigmatic and syntagmatic axes are applied in an fMRI encoding experiment to explore human brain's activity pattern during story listening. This study proposes the construction of paradigmatic and syntagmatic semantic embeddings respectively by transforming WordNet-alike knowledge bases and subtracting paradigmatic information from a statistical word embedding. It evaluates the semantic embeddings by leveraging word-pair proximity ranking tasks and contrasts voxel encoding models trained with the two types of semantic features to reveal the brain's spatial pattern for semantic processing. Results indicate that in listening comprehension, paradigmatic and syntagmatic semantic operations both recruit inferior (ITG) and middle temporal gyri (MTG), angular gyrus, superior parietal lobule (SPL), inferior frontal gyrus. A non-continuous voxel line is found in MTG with a predominance of paradigmatic processing. The ITG, middle occipital gyrus and the surrounding primary and associative visual areas are more engaged by syntagmatic processing. The comparison of two semantic axes' brain map does not suggest a neuroanatomical segregation for paradigmatic and syntagmatic processing. The complex yet regular contrast pattern starting from temporal pole, along MTG to SPL necessitates further investigation.

Index Terms—fMRI encoding, semantic embedding, paradigmatic, syntagmatic, regression

I. INTRODUCTION

Some functional magnetic resonance imaging (fMRI) investigations on word-meaning processing in the human brain have correlated semantic embedding vectors with neural activity recording. Especially it probes into the representational and/or operational functionalities of neural regions. Previous studies [1]–[6] revealed a distributed pattern of cortical areas' correlation with word semantic models. The relatively well-modeled voxels are typically located in the temporal, parietal and frontal lobes. Studies show that context length [5], semantic category [2], [7], conceptual concreteness [8] and perceptual properties [9], [10] are all related factors to differentially elicit semantic activities in different lobes. Yet, despite abundant literatures on theoretical semantic network architecture[11], no concluding evidence enlightens each cortical region's functional role in the processing.

Classical view [12] supported a temporal construction of semantic processing. To bridge the temporal-centric semantic processing theories and the found distributivity, a structure of *semantic hub* [13], [14], located in the bilateral anterior temporal lobe (ATL), acting as a centralized location that hosts and/or processes semantics, is by far consistent with human neural modeling studies [15], anatomical analysis [16] and pathological observations [17]. For example, in patients with semantic dementia (SD), the ATL atrophy and hypometabolism lead to a conceptual deficit. These patients develop a specific loss of capability to use distinctive features (such as stripes for zebras) necessary to identify and characterize a concept. So that, they will fail to name specific categories (eg. zebra) and would use similar or hypernymous categories (eg. horse) instead. The semantic hub's involvement in linguistic and non-linguistic semantic tasks indicates a generalized supra-modal or amodal concept network storage in ATL, beyond word meanings. It is said to accept inputs from a specific semantic neural component and activates reciprocally other components to evoke a complete mental representation of a word. Therefore, the hub must support a holistic structure of all words with at least minimal information. [18] suggests that the left ATL is a non-syntactic conceptual hub involved in semantic composition. The convergence of ATL's relatedness to semantic composition of conceptual attributes, to the preservation of hypernymous or similarity links (zebra–horse) and to the loss of the distinctive features in case of lesion to this region, needs to be further explored. As for the other components in the semantic processing system, they are either responsible for a modality specific semantic information stream or a semantic control/computation function.

An analogy in structural linguistics for this two-fold semantic processing is the paradigmatic and syntagmatic axes proposed by De Saussure, Jakobson and Halle. The two-axis proposition is best illustrated by an example in Table I. To render the paradigmatic *absentia* candidates, the responsible neural structure should hold at least a local lexical similarity comparison in the paradigmatic sense, and sort out the most similar words. The syntagmatic axis looks for semantic associations mainly by collocation. The functional association of syntagmatic semantics with the neural components found active in the aforementioned previous studies needs to be further explored. In addition, Saussure's theoretical proposal is completed by a psycholinguistic meta-analysis by [20] on two dissociated groups of aphasics each associated with one semantic axis, namely selection-deficits for paradigmatically

TABLE I
AN EXAMPLE OF SYNTAGMATIC AND PARADIGMATIC AXES

syntagmatic →

The ridiculous	girl	fell	into the	pond.
silly	person	jumped		river.
foolish	woman	tripped		lake.
funny	lady	plunged		sea.
crazy	princess	walked		ocean.
klutzy	child	ran		pool.

(paradigmatic, vertical axis on left)

This example is adapted from [21]. Gray-colored texts are in *absentia*, black-colored texts are in *presentia*.

On a phrasal level, the twofold proposition projects lexical entities in a phrasal linguistic unit onto two semantic axes. The syntagmatic axis operates in *presentia*, it combines words (present in the context) into larger linguistic units (phrases, sentences and paragraphs) using contexts and syntactic rules. In this case, *ridiculous* and *girl* is combined into a larger semantic unit. Each word in the larger unit is combinatorially bound to another, and the combination adds additional semantic value to the whole larger linguistic unit.

In contrast, the paradigmatic axis provides alternative words (in *absentia*) to clarify and complete the meaning of the current present word in the phrase. These unselected candidates proposed by the paradigmatic axis are near synonyms to the default word, and they do not change the meaning of the phrase.

impaired aphasics, and contexture-deficits for syntagmatically impaired ones, implying potentially different cortical loci for the processing of two axes.

This study leverages the contrast of paradigmatism and syntagmatism to investigate the nature of each cortical region's involvement in lexical semantic processing.

II. METHOD

This study hypothesizes on a representational and/or operational account that the paradigmatic and syntagmatic axes engage different yet not necessarily dissociated neural structures during semantic processing. Consistent with other works using fMRI encoding experiments [1], we propose an investigation by modeling neural activity with semantic embedding features, separately for the paradigmatic and syntagmatic axis, to potentially segregate the respective responsible neural structures.

The experiment is conducted with priorly collected data from the framework project "Neural Computational Models of Natural Language" (Principle Investigators: John Hale and Christophe Pallier). The stimuli materials preparation, participant recruitment, fMRI experiment and fMRI data preprocessing are carried out in [22].

Following the standard fMRI encoding procedure, the first step is to model the time-sequences of the administered stimuli received by experiment participants during the fMRI sessions. The external stimuli model will rely on paradigmatic and syntagmatic representation models (semantic embeddings), which themselves are abstract vectorial representations of semantic values in a high-dimensional space. The second step consists of voxel-level regression analyses which map the theoretical stimuli temporal signals onto actual fMRI measured brain activities. The brain activities are modeled as a temporo-spatial 4-dimension data. In MRI recordings, the brain volume is divided into small 3D volume units, namely *voxels*. Each voxel has a time series of spatially aggregated brain-activity measurement within its volume. A voxel-wise regression model is a mathematical function mapping the temporal semantic sequences onto the brain activity signal space.

The result analysis relies on each voxel model's predictive power provided with its training feature dataset. Stronger the model performance (or the model performance improvement compared to another model), the stronger the correlation between the neural activity in that voxel and the semantic and non-semantic feature contained in the training set.

A. Participants

20 French native speakers (11 females, average age of 24.5 years-old, range 18–39 years-old, right handed according Edinburgh's inventory [23] adapted for French, averaged score 0.903, range 0.375–1, without antecedent neurological or psychiatric disorders) were recruited from NeuroSpin's volunteer inventory. The recruited participants declare to have not been exposed to any material related to *The Little Prince* within last 5 years, including written books, audiobooks, films in any languages, and to be unable to clearly recall the story of the book.

B. Procedure

After being received by a researcher and informed with the experience protocol, the participants are first examined by the lab doctor, then invited to scan an anatomical MRI which lasts 8 minutes. In the scanner, the participants are able to see the instructions and visual aides displayed on the screen through a mirror. During the anatomical scan, two illustrations from the first two chapters of *the Little Prince* [24] are presented to prepare the participants. A sound test phase follows: the introductory paragraphs are administered to the participants under the noise of MRI scanner. The auditory stimuli is adapted from the audiobook [25]. Once the sound adjusted, the functional imaging sessions begin.

For the comfort of the participants and their concentration during listening comprehension, the audiobook is divided into 9 blocks, so that each block lasts at most 15 minutes. For each chapter, the reading of chapter title is removed from the audio, and 3 seconds of silence is added. Within each block, the participants listened passively with eyes closed, to introduce the minimum amount of perturbation to the BOLD signal. At the end of each block, they were tested with comprehension multi-choice questions displayed on the screen and responded orally. Additionally, they were required to answer "comprehension questions" necessitating finer and deeper reflexions or to retell the synopsis of the past block. 4 or 5 blocks are played in the morning, and the rest in the afternoon. A final block is added, where the audio stimuli is comprised of normal French sentences and acoustically deformed sentences to localize language-processing related areas.

C. MRI Acquisition and Preprocessing

Structural and functional data are acquired on a whole-brain 3-Tesla Siemens scanner at NeuroSpin. Structural images are collected in 154 axial slices with 1 mm isotropic voxels. Functional images are collected with multi-echo EPI sequence ($3.159^3 mm^3$, TR = 2000 ms). Multi-echo sequence is used to boost BOLD and non-BOLD component identification based on TE dependence, resulting in a higher signal-to-noise ratio in the final image rendering [26].

MRI data is analyzed with Multi-Echo Independent Component Analysis (ME-ICA) toolkit (https://github.com/ME-ICA/me-ica). Raw structural data and 4-D time-series data grouped by TE are supplied to render de-noised sMRI and fMRI, which are also spatially warped in Montreal Neurological Institute (MNI) template.

D. Paradigmatic Embedding

The paradigmatic embedding is built based on a French version of WordNet [27], [28], WOLF [29], a knowledge graph of words and abstract concepts. WordNet models synonym groups (*synset*) as vertices in a graph, and various semantic relations as edges connecting two vertices. The explicit semantic relations in WordNet are hand-coded by linguists. WOLF is a machine translated version of WordNet, manually verified and corrected by its authors.

With WordNet-alike knowledge bases, it is possible to compare pure taxonomical semantic similarity. Therefore, the knowledge base supports a modeling implementation of the paradigmatic axis. Such speculation is backed by a meta-analysis [30]. [30] collected a set of proposed distance metrics on WordNet then compared the word-pair 'proximities' against human similarity/relatedness judgement data (word-pair proximity ranking task). The results show that with WordNet it is possible to derive paradigmatic-specific semantic information. [31] uses WordNet retrofitting to enhance distributional embeddings. With the same ranking task and benchmarks, the enhanced versions perform better in paradigmatic tests and significantly worse in syntagmatic settings.

In WordNet-like knowledge bases, some of the semantic relations are paradigmatic: synonymy, hypernymy, hyponymy, the relation where an adjective is a participle of a verb (*exhausting–exhaust*), adjective having similar meaning (*exhausting–effortful*) and adverb deriving from adjective (*essentially–essential*). Other relations are syntagmatic: collocations, meronymy/holonymy (*ceil–house*), entailment/causality (*sunset–milky-way*). To build the paradigmatic embedding, We use [32]'s algorithm to transform the tree-like knowledge base consisting of *synsets* into a lexical semantic embedding, using a sub-graph of WordNet with only paradigmatic edges.

Consider vertices, which are more closely and densely connected by paradigmatic edges, more proximate in the paradigmatic sense. The algorithm implements first an infinite random walk to compute the graph distance between every word-pair with a discount factor for longer paths. In Eq. 1, M is the adjacency matrix of the knowledge graph, γ is a decay parameter weighting how longer paths are dominated

by shorter ones. M_G is the convergent limit of the iterative process, which is the final state of the random walk. Then a normalized positive point-wise mutual information transformation is applied to reduce noises induced by unbalanced word occurrence frequency. Finally a principle component analysis (PCA) is applied for dimensionality reduction to render the final embedding output. The number of PCs are determined in the later feature engineering stage.

$$M_G^{(n)} = I + \gamma M + \gamma^2 M^2 + \cdots + \gamma^n M^n$$

$$M_G = \sum_{e=0}^{\infty} (\gamma M)^e = (I - \gamma M)^{-1} \tag{1}$$

E. Syntagmatic Embedding

The syntagmatic embedding is based on DepGloVe [33], a GloVe-alike [34] distributional semantic models [35]–[38], constructed based on the *distributional hypothesis* [39] that words having similar context have as well similar meanings. However, this consideration also brings close words consistently co-occurring in same contexts, such as *teacher* and *student*. In several analyses [30], [40], a set of pre-trained corpus-based distributional word-embeddings including GloVe are shown to always mix paradigmatic and syntagmatic axes into the actual embedding.

Since GloVe-like co-occurrence-based statistical distributed representation models contain both paradigmatic and syntagmatic information, it remains as an ideal source for syntagmatic information extraction as there exists no, to our knowledge, a pure syntagmatic embedding.

Under a linear-additive approximation of the mixture of the two semantic axes, a syntagmatic embedding can be extracted from a GloVe-like embedding by subtracting detectable paradigmatic counterpart. Suppose the abstract paradigmatic semantic relations modeled by the mixed embedding is also accounted by the paradigmatic embedding constructed in the previous section, then we can use the projection of the paradigmatic embedding onto the mixed space to reveal the paradigmatic component and separate the two semantic sub-spaces.

Consider `MIX`, `PAR` and `SYN` in Eq. (2) as 2D matrix representations respectively for the co-occurrence-based Dep-GloVe, the paradigmatic embedding built with WOLF, and the syntagmatic embedding to be constructed. The lexicons of the three embeddings are aligned such that each row with the same index from three matrices represents one same lexicon unit. F is a transformation matrix learned with general linear model, of which the computational objective is to minimize the L-2 norm of `SYN`.

$$\texttt{MIX} = \texttt{PAR}.F + \texttt{SYN} \tag{2}$$

F. Feature Engineering

To construct stimuli-based temporal sequence features, the audiobook and the original text of *the Little Prince* serve as the

base. The onset and offset of each word in the audio document is aligned with the text. To aggregate the semantic value of various inflected French words, stemming is performed for source corpora and *the Little Prince* with `spaCy` and `FrenchLefffLemmatizer` [41]. The lemmatized text is manually verified.

Five feature time-sequence groups are created. The acoustic energy `RMS` is modeled by the average square root of the amplitude each ten millisecond, calculated with Octave (https://www.gnu.org/software/octave/). The word presence indicator `WRATE` is a binary feature. The feature's activation function is set to 1 from the onset to the offset of the word. Content word presence indicator `CWRATE` is similar to `WRATE`, and the feature returns 1 only if the part-of-speech (POS) tag of the word is among nouns, verbs, adjectives and adverbs. Two semantic feature groups `PAR` (for paradigmatic) `SYN` (for syntagmatic) takes the latent values extracted from the corresponding embedding for the duration of the word utterance.

G. Encoding Regressors

Voxels are considered as independent units as in other projects [4]. The BOLD temporal signal associated with an arbitrary voxel j is modeled as a linear combination of activities elicited by non-semantic and semantic features (3). The f_is are the previously presented feature temporal sequences. For each voxel, five classes of encoding models are trained: `RMS`, `RMS+WRATE`, `RMS+WRATE+CWRATE`, `RMS+WRATE+CWRATE+PAR`, `RMS+WRATE+CWRATE+SYN`. The order of the feature groups are consistent across different semantic modeling and the feature regressors are orthonormalized to remove potential confounds in later contrast analyses. Here hrf is the hemodynamic function used in Statistical Parametric Mapping software (https://www.fil.ion.ucl.ac.uk/spm/), provided by the Python library `nistats` (https://nistats.github.io/) [42]. This function mocks the BOLD response to neural activations. The convolution of hrf with feature activation functions models the hereby elicited BOLD signal, converts the activation function into *regressors* which are used in later regression/encoding. The coefficients $\beta_{i,j}$ are to be determined via regression models.

$$\text{BOLD}_j(t) = \sum_i \beta_{i,j} \times f_i(t) * \text{hrf}(t) + \text{Noise}(t), \quad (3)$$

One particularity for `PAR` regressors is the filtering by regressor variance as they are based on PCA factored components. Regressors are dropped if its variance over the time-sequence is smaller than 10^{-5}, which resulted 100 residual `PAR` regressors.

H. fMRI Encoding

The relatively small observation samples collected from fMRI sessions and abundant regression features motivate the regularization with Ridge, which is described by (4) to promote the generalizability of the learned transformation matrix.

The Ridge regression training algorithm uses a regularization coefficient α_j to balance the complexity of the model and the signal loss between the proposed model (the deterministic part as described by (2)) and the real signal $\text{BOLD}_{\text{real},j}$) measured for a given voxel j. For model classes containing `PAR` or `SYN` features, feed-forward feature selection is applied to maximally avoid potential overfitting by taking only the first $N_{j,\text{cls}}$ number of features into consideration in voxel modeling. The values of the regularization coefficient α_j and $N_{j,\text{cls}}$ are determined independently for each voxel via a grid search in a later stage of computation.

$$\min_{\beta_{i,j}} \sum_t | \sum_i^{N_{j,\text{cls}}} \beta_{i,j} \times f_i(t) * \text{hrf}(t) - \text{BOLD}_{\text{real},j}(t)|^2$$
$$+ \alpha_j \sum_i^{N_{j,\text{cls}}} \beta_{i,j}^2 \quad (4)$$

Models are trained and evaluated by means of a block-wise cross-validation where an fMRI record block is left to test the model. The coefficient of determination r2 is is a measure of the quality of the prediction. For each voxel j of a subject, the performance of the voxel model is averaged over nine cross-validations. Then, for each model class *cls*, the grid search retains only the combination of α_j and $N_{j,\text{cls}}$ yielding the highest r2. Group scores are obtained by averaging the same voxel model scores of all subjects.

Voxel models are contrasted across model-classes and 3D maps summarizing voxel-wise differences reveal each feature's spatial distribution in the human brain for language processing tasks. Five contrasts are calculated (see (5)) to show the evolution of lower- to higher-level language processing in the human brain. The first four contrasts are computed with nested voxel models. For the fifth, a non-nested model comparison framework [43] is experimented for regression simplicity. The contrast maps are visualized with the help of `nilearn`[42].

$$\text{Contrast}_{\text{WRATE}} = r^2_{\text{RMS+WRATE}} - r^2_{\text{RMS}}$$
$$\text{Contrast}_{\text{CWRATE}} = r^2_{\text{RMS+WRATE+CWRATE}} - r^2_{\text{RMS+WRATE}}$$
$$\text{Contrast}_{\text{PAR}} = r^2_{\text{RMS+WRATE+CWRATE+PAR}} - r^2_{\text{RMS+WRATE+CWRATE}}$$
$$\text{Contrast}_{\text{SYN}} = r^2_{\text{RMS+WRATE+CWRATE+SYN}} - r^2_{\text{RMS+WRATE+CWRATE}}$$
$$\text{Contrast}_{\text{PS}} = r^2_{\text{RMS+WRATE+CWRATE+PAR}}$$
$$- r^2_{\text{RMS+WRATE+CWRATE+SYN}}$$
$$(5)$$

III. RESULTS

A. Twofold Dissociation in Embeddings

The constructed paradigmatic and syntagmatic embeddings successfully dissociates two semantic principles as shown by the word-pair semantic proximity ranking task (Table II), measured with paradigmatic-/syntagmatic-specific benchmarks. The French benchmark data are translated [44] from

TABLE II
SEMANTIC SPACE SEMANTIC RANKING TASK RESULTS

Semantic Space	Vocabulary Size	Dimension	*Paradigmatic* SimLex-999	*Syntagmatic* WS353-ASN
French				
PAR	56665	634	.2812	**.0511**
	Out of Vocabulary		.048	.04
MIX			.1449	.2078
SYN	24519	200	.0771	.1566
	Out of Vocabulary		.0797	.0711
English				
PAR	15000	511	.4989	**.0193**
MIX	2.2M		.3752	.5709
SYN	8157	300	.2133	.5918
	Out of Vocabulary		.002	.024

The semantic ranking task compares the correlation between the model predicted word-pair similarity scores and the gold-standard. The null hypothesis is a non-existent linear correlation. Spearman's r is reported to measure the quality of an embedding in one semantic axis. Scores marked in bold correspond to a p-value larger than 0.05.

Both PAR embeddings attain high correlation scores in paradigmatic test and near-zero syntagmatic scores. Both MIX spaces perform well in both task-sets, with a slight preference for association, consistent with [40]'s conclusion. The extracted SYN have comparable scores in syntagmatic tests with MIX, but have a non-zero, yet much reduced score in paradigmatic tests.

two widely used English datasets [45], [46] [1]. By paradigmatic sub-graph curation, the French PAR contains only paradigmatic semantic information. The constructed French SYN significantly reduced paradigmatic information compared to the original embedding, while its syntagmatic counterpart is also slightly impacted, as indicated by the task results. To visually control the embedding quality, we sampled a few tokens from *the Little Prince* and examined the vectorial neighbours with TensorFlow (http://projector.tensorflow.org/) embedding projector from different embeddings and were reassured of the axis-wise purity of each embedding, thus the effect of axis dissociation. To further validate the methodology, the embedding construction and benchmarking are replicated in English using WordNet and a pre-trained GloVe embedding [2]. The replication achieves a better dissociation of two semantic axes, supporting our methodology, and suggesting that the French embeddings could be further tested with more appropriate behavioral benchmark data.

B. Primary Auditory Processing

With acoustic feature RMS, the r2 distribution map illustrated with obtained voxel model performances (Fig. 1) reveals

---

[1]The translations are available on https://github.com/nicolasying/Similarity-Association-Benchmarks, commit c97583f.
[2]The pre-trained model can be downloaded from https://nlp.stanford.edu/projects/glove.
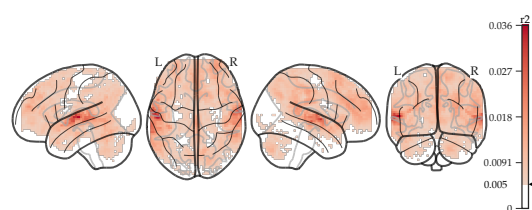

Fig. 1. Acoustic energy feature reveals bilateral primary auditory cortices with a slight left lateralization.
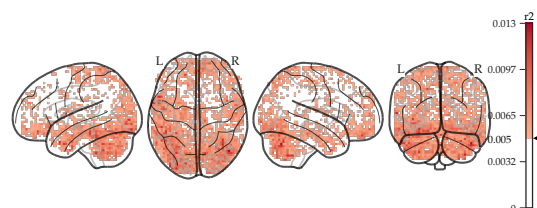

Fig. 2. Content word presence indicator feature improvements are mainly located in bilateral temporal pole, inferior temporal gyrus, frontopolar prefrontal.

several voxel-clusters with the most significant and largest activations from left posterior superior temporal gyrus (pSTG, Brodmann Area (BA) 41, peak at -60 -12 4, $r2_{max} = 0.0362$), followed by right pSTG (61 -13 2, $r2_{max} = 0.0328$). Other important yet much weaker activations are found in right middle cingulate cortex (MCC, BA23, 0 -24 29, $r2_{max} = 0.0189$) and right middle frontal gyrus (MFG, BA10, 29 56 20, $r2_{max} = 0.0180$).

C. Gross Semantic Processing

Possibly due to the collinearity between RMS and WRATE, the addition of word presence indicator to the feature set does not improve voxel-models. The results are omitted. CWRATE on the contrary, boosts a large portion of voxels' performance especially for those initially badly predicted with only RMS and WRATE features (Fig. 2), indicating that voxels processing semantic information has a limited usage of primary acoustic features.

The major improvement of CWRATE is remarked in left middle and inferior temporal pole (TP) (BA38, -53 11 -33), bilateral posteroinferior temporal gyrus (ITG, including fusiform gyrus FG, BA19/37, -47 -43 -24, 45 -69 -38), frontopolar prefrontal cortex (fpPFC, near rectus gyrus, -5 46 -26) and posterior cerebellum (Wilcoxon's W=136, Δr2>0.0067, p-value<$10^{-3.66}$ uncorrected).

D. Paradigmatic Processing

Fig. 3 illustrates the voxel performance improvements when paradigmatic semantic embedding features PAR are added to RMS+WRATE+CWRATE regressors. The most improved voxel clusters are located in bilateral middle temporal gyri (MTG, -51 -34 1, 57 -36 0, left cluster more significant), right angular gyrus (AG, 35 -65 44) and left superior parietal lobule
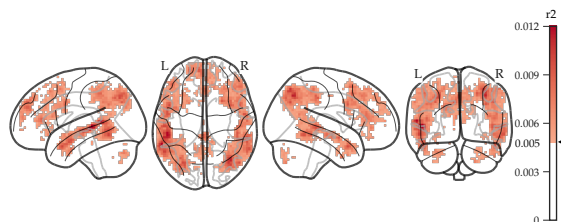
Fig. 3. Paradigmatic embedding improves voxel models located in bilateral middle temporal gyri, angular gyri and left superior parietal lobule.
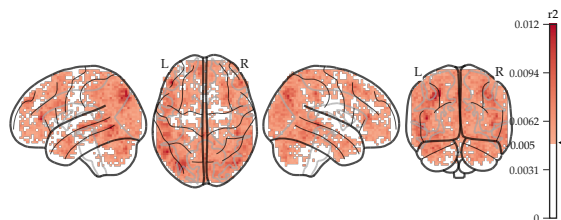

Fig. 4. Syntagmatic embedding improves voxels from the bilateral anterior temporal lobe, posterior middle temporal gyri, bilateral inferior frontal gyri, right angular gyrus and left middle occipital gyrus.

(SPL, -27 -72 44) (W=210, Δr2>0.0079, p-value<$10^{-4.35}$ uncorrected).

E. Syntagmatic Processing

Fig. 4 illustrates the voxel performance improvements when syntagmatic semantic embedding features are added to RMS+WRATE+CWRATE regressors. The most significant improvements are slightly different from paradigmatic neural responses. The most significantly improved cluster, which is located in left MTG, is more posterior than that found for PAR contrast (-53 -58 3). Other important clusters include bilateral inferior frontal gyrus pars triangularis (IFGtri, -46 37 11, 50 35 8), left middle occipital gyrus (MOG, -32 -78 41), right AG (43 -75 39) (W=190, Δr2>0.0065, p-value<$10^{-4.18}$ uncorrected).

F. Paradigmatic and Syntagmatic Particularity

Current data suggest that paradigmatic and syntagmatic semantic processing are not supported by macroscopically dissociated neural components: the voxels along left MTG, right anterior and posterior MTG, bilateral SPL and IFGtri are improved equally by PAR and SYN.

To reveal a differential map for two semantic axes' preference, the two classes of voxel models are contrasted with each other (Fig. 5). Most voxels are slightly better modeled by SYN models. Along the non-nested model comparison framework [43], the collinearity between paradigmatic and syntagmatic regressors are discovered. With PAR feature regressors, it is not possible to predict SYN features via a generalized linear model, whereas SYN regressors can be mapped to partially reconstruct first 5 PAR regressors.

The Wilcoxon test on voxel models of 20 subjects yields two small significant voxel clusters for PAR in left superior frontal cortex (SFC, BA10, -27 59 25) and left anterior cingulum
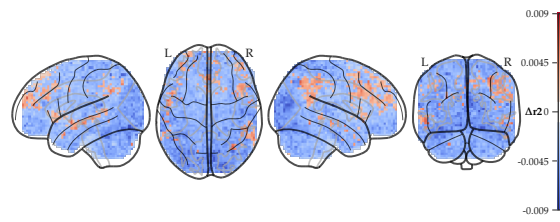

Fig. 5. Paradigmatic and syntagmatic voxel model contrast. Positive values correspond to preference for paradigmatic processing, negative values to syntagmatic.

cortex (ACC, -8 34 25) (Δr2>0.0069, p-value<0.05 voxel-wise multi-comparison corrected). Left aSTG, right pSTG, right posterior superior temporal sulcus (pSTS, BA22) and left aSTS/middle TP are also reported (p<0.001 uncorrected).

SYN found 17 small clusters in bilateral visual association areas (BA18), primary visual areas (BA17), ventrotemporal areas (ventral ITG, parahippocampal gyrus), left SPL, left thalamus and bilateral cerebellum (Δr2>0.0068, p-value<0.05 voxel-wise multi-comparison corrected).

IV. DISCUSSION

The acoustic signal regressor is able to provide a primary auditory processing locus located in bilateral BA41, consistent with existing literature [47]. The slight left lateralization of acoustic voxel-model performance is consistent with Tervaniemi and Hugdahl's [48] finding that left lateralization is linked to speech processing. The bilateral inferotemporal improvements, starting from fusiform gyri to anterior temporal regions, brought by content word presence indicator is associated with semantic memory processing [49], [50] and word-wise semantic priming effect in word meaning comprehension [51].

The spatial patterns of model improvements by paradigmatic and syntagmatic features in MTG interlace with each other, showing a gradient trend for both axes' processing in MTG. Both features improve bilateral posterior MTG voxel models. This region is associated with word-meaning access across modalities and categories of concept [50] and was also proposed as a semantic hub [52]. IFGtri, improved by both semantic axes as well, seems to be associated with general semantic processing [53]. ITG, predominantly improved by syntagmatic embedding, is called by [54] as a *unification space* enabling syntactico-semantic integration. It suggests that syntagmatic axis is used in this process. Further investigation on the feasibility of syntactic prediction via only syntagmatic and the sentential semantic integration process is needed.

On paradigmatic/syntagmatic preferential contrast, despite the small effect sizes, the trends are roughly symmetrical across two hemispheres and the distribution is spatially regular. However, no region-of-interest (ROI) level analysis has revealed regional preference for paradigmatic processing, indicating no macroscopic neural structures related to the specific semantic axis on a group level.

Another paradigmatic embedding, namely the projection of paradigmatic embedding onto the mixed space has also been contrasted with existing `SYN` models. The contrasts shown by the two paradigmatic embeddings against `SYN` are globally consistent, but are not stable on the voxel level. The intersection of two paradigmatic-predominant regions (obtained by the `PAR`/`SYN` contrasts) is restrained to two small voxel clusters located in left posterior to middle STG/MTG and right inferior parietal lobule.

In conclusion, the paradigmatic/syntagmatic separation in semantic fMRI encoding leads to minute regional contrasts for both semantic axes. Syntagmatic processing is dominant in ITG and MOG. The ITG aspect could be related to semantic unification process. MOG's linkage to imaginability[55] and its role in syntagmatic processing remain to be further investigated. The paradigmatic preference found in MTG is neighbored by syntagmatic counterpart, and in MTG both axes improve significantly model performance, indicating a gradient yet mixed nature of MTG semantic processing. The current results do not reveal a statistically established preference for the syntagmatic processing axis over the paradigmatic one and vice versa in the bilateral ATL, contrary to what might be expected following [18]. Future research could bridge syntactic processing, syntagmatic semantic processing and semantic composition. Distributional and structural neuroscience studies could also complete this current fMRI investigation by providing anatomical theoretical constructs.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENT

REFERENCES

[1] T. M. Mitchell, S. V. Shinkareva, A. Carlson, K.-M. Chang, V. L. Malave, R. A. Mason, and M. A. Just, "Predicting human brain activity associated with the meanings of nouns," *science*, vol. 320, no. 5880, pp. 1191–1195, 2008.

[2] A. G. Huth, S. Nishimoto, A. T. Vu, and J. L. Gallant, "A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain," *Neuron*, vol. 76, no. 6, pp. 1210–1224, Dec. 20, 2012, ISSN: 08966273. DOI: 10.1016/j.neuron.2012.10.014. pmid: 23259955. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0896627312009348.

[3] A. J. Anderson, J. R. Binder, L. Fernandino, C. J. Humphries, L. L. Conant, M. Aguilar, X. Wang, D. Doko, and R. D. S. Raizada, "Predicting Neural Activity Patterns Associated with Sentences Using a Neurobiologically Motivated Model of Semantic Representation," *Cerebral Cortex*, Aug. 12, 2016, ISSN: 1047-3211, 1460-2199. DOI: 10.1093/cercor/bhw240. [Online]. Available: http://cercor.oxfordjournals.org/cgi/doi/10.1093/cercor/bhw240.

[4] A. G. Huth, W. A. De Heer, T. L. Griffiths, F. E. Theunissen, and J. L. Gallant, "Natural speech reveals the semantic maps that tile human cerebral cortex," *Nature*, vol. 532, no. 7600, pp. 453–458, Apr. 27, 2016, ISSN: 14764687. DOI: 10.1038/nature17637. pmid: 27121839. [Online]. Available: http://www.nature.com/doifinder/10.1038/nature17637.

[5] S. Jain and A. Huth, "Incorporating Context into Language Encoding Models for fMRI," *bioRxiv*, p. 327601, May 21, 2018. DOI: 10.1101/327601. [Online]. Available: https://www.biorxiv.org/content/early/2018/05/21/327601.

[6] F. Pereira, B. Lou, B. Pritchett, S. Ritter, S. J. Gershman, N. Kanwisher, M. Botvinick, and E. Fedorenko, "Toward a universal decoder of linguistic meaning from brain activation," *Nature Communications*, vol. 9, no. 1, p. 963, Dec. 6, 2018, ISSN: 20411723. DOI: 10.1038/s41467-018-03068-4. pmid: 29511192. [Online]. Available: http://www.nature.com/articles/s41467-018-03068-4.

[7] Y. Shtyrov, O. Hauk, and F. Pulvermüller, "Distributed neuronal networks for encoding category-specific semantic information: The mismatch negativity to action words," *European Journal of Neuroscience*, vol. 19, no. 4, pp. 1083–1092, 2004, ISSN: 1460-9568. DOI: 10.1111/j.0953-816X.2004.03126.x. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.0953-816X.2004.03126.x.

[8] P. Hoffman, R. J. Binney, and M. A. Lambon Ralph, "Differing contributions of inferior prefrontal and anterior temporal cortex to concrete and abstract conceptual knowledge," *Cortex*, vol. 63, pp. 250–266, Feb. 2015, ISSN: 00109452. DOI: 10.1016/j.cortex.2014.09.001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0010945214002913.

[9] R. L. Moseley and F. Pulvermüller, "Nouns, verbs, objects, actions, and abstractions: Local fMRI activity indexes semantics, not lexical categories," *Brain and Language*, vol. 132, pp. 28–42, May 2014, ISSN: 0093934X. DOI: 10.1016/j.bandl.2014.03.001. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0093934X1400039X.

[10] V. Borghesani, F. Pedregosa, M. Buiatti, A. Amadon, E. Eger, and M. Piazza, "Word meaning in the ventral visual path: A perceptual to conceptual gradient of semantic coding," *NeuroImage*, vol. 143, pp. 128–140, Dec. 2016, ISSN: 10538119. DOI: 10.1016/j.neuroimage.2016.08.068. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1053811916304591.

[11] A. D. Friederici, "The Brain Basis of Language Processing: From Structure to Function," *Physiological Reviews*, vol. 91, no. 4, pp. 1357–1392, Oct. 2011, ISSN: 0031-9333. DOI: 10.1152/physrev.00006.2011. pmid: 22013214. [Online]. Available: http://physrev.physiology.org/cgi/doi/10.1152/physrev.00006.2011.

[12] D. Saumier and H. Chertkow, "Semantic memory," *Curr Neurol Neurosci Rep*, vol. 2, no. 6, pp. 516–522, Nov. 2002, ISSN: 1528-4042. pmid: 12359106.

[13] K. Patterson, P. J. Nestor, and T. T. Rogers, "Where do you know what you know? The representation of semantic knowledge in the human brain," *Nature Reviews Neuroscience*, vol. 8, no. 12, pp. 976–987, Dec. 2007, ISSN: 1471-0048. DOI: 10.1038/nrn2277. [Online]. Available: https://www.nature.com/articles/nrn2277.

[14] M. A. L. Lambon-Ralph, E. Jefferies, K. Patterson, and T. T. Rogers, "The neural and computational bases of semantic cognition," *Nature Reviews Neuroscience*, vol. 18, no. 1, pp. 42–55, Jan. 2017, ISSN: 1471-0048. DOI: 10.1038/nrn.2016.150. [Online]. Available: https://www.nature.com/articles/nrn.2016.150.

[15] G. Pobric, E. Jefferies, and M. A. Lambon Ralph, "Category-Specific versus Category-General Semantic Impairment Induced by Transcranial Magnetic Stimulation," *Current Biology*, vol. 20, no. 10, pp. 964–968, May 25, 2010, ISSN: 0960-9822. DOI: 10.1016/j.cub.2010.03.070. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0960982210004562.

[16] H. Damasio, D. Tranel, T. Grabowski, R. Adolphs, and A. Damasio, "Neural systems behind word and concept retrieval," *Cognition*, vol. 92, no. 1-2, pp. 179–229, 2004.

[17] P. J. Nestor, T. D. Fryer, and J. R. Hodges, "Declarative memory impairments in Alzheimer's disease and semantic dementia," *NeuroImage*, vol. 30, no. 3, pp. 1010–1020, Apr. 15, 2006, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2005.10.008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811905007895.

[18] L. Pylkkänen, "The neural basis of combinatory syntax and semantics," *Science*, vol. 366, no. 6461, pp. 62–66, Oct. 4, 2019, ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aax0050. [Online]. Available: http://www.sciencemag.org/lookup/doi/10.1126/science.aax0050 (visited on 12/06/2019).

[19] F. De Saussure, *Cours de Linguistique Generale: 3e Ed*. Payot, 1969.

[20] R. Jakobson and M. Halle, *Fundamentals of Language*. S-Gravenhage: Mouton & Co, 1963.

[21] *Example of paradigmatic and syntagmatic axes*. [Online]. Available: https://lmdresources.files.wordpress.com/2013/04/image002.jpg.

[22] S. Todorovic and C. Pallier, "Analyses IRMf lors de l'écoute de texte naturel," 2018, 98 pp.

[23] R. C. Oldfield, "The assessment and analysis of handedness: The Edinburgh inventory," *Neuropsychologia*, vol. 9, no. 1, pp. 97–113, Mar. 1, 1971, ISSN: 0028-3932. DOI: 10.1016/0028-3932(71)90067-4. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0028393271900674.

[24] A. de Saint-Exupéry, *Le Petit Prince*. 1943.

[25] A. de Saint-Exupéry and D. Wilkinson, *The little prince : a French/English bilingual book*. Rutland [England]: Omilia Languages Ltd, 2011, ISBN: 978-0-9567215-9-4.

[26] P. Kundu, S. J. Inati, J. W. Evans, W.-M. Luh, and P. A. Bandettini, "Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI," *Neuroimage*, vol. 60, no. 3, pp. 1759–1770, 2012.

[27] G. A. Miller, "WordNet: A lexical database for English," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, 1995.

[28] ——, *WordNet: An Electronic Lexical Database*. MIT press, 1998.

[29] B. Sagot and D. Fišer, "Building a free French wordnet from multilingual resources," in *OntoLex*, 2008.

[30] C. Lofi, "Measuring Semantic Similarity and Relatedness with Distributional and Knowledge-based Approaches," *Information and Media Technologies*, vol. 10, no. 3, pp. 493–501, 2015, ISSN: 09758887. DOI: 10.11185/imt.10.493. pmid: 9927097148801572. [Online]. Available: https://www.jstage.jst.go.jp/article/imt/10/3/10_493/_article/-char/ja/.

[31] M. Faruqui, J. Dodge, S. K. Jauhar, C. Dyer, E. Hovy, and N. A. Smith, "Retrofitting Word Vectors to Semantic Lexicons," in *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Denver, Colorado: Association for Computational Linguistics, 2015, pp. 1606–1615. DOI: 10.3115/v1/N15-1184. [Online]. Available: http://aclweb.org/anthology/N15-1184.

[32] C. Saedi, A. Branco, J. António Rodrigues, and J. Silva, "WordNet Embeddings," in *Proceedings of The Third Workshop on Representation Learning for NLP*, Melbourne, Australia: Association for Computational Linguistics, Jul. 2018, pp. 122–131. [Online]. Available: http://www.aclweb.org/anthology/W18-3016.

[33] E. de la Clergerie. (). DepGlove small server, [Online]. Available: http://alpage.inria.fr/depglove/process.

[34] J. Pennington, R. Socher, and C. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.

[35] C. Burgess and K. Lund, "Hyperspace analogue to language (HAL): A general model of semantic memory," in *Annual Meeting of the Psychonomic Society, Los Angeles*, 1995.

[36] O. Levy and Y. Goldberg, "Dependency-Based Word Embeddings," in *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Baltimore, Maryland: Association for Computational Linguistics, Jun. 2014, pp. 302–308. [Online]. Available: http://www.aclweb.org/anthology/P14-2050.

[37] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," Jan. 16, 2013. arXiv: 1301.3781 [cs]. [Online]. Available: http://arxiv.org/abs/1301.3781.

[38] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic Regularities in Continuous Space Word Representations," p. 6,

[39] Z. S. Harris, "Distributional Structure," WORD, vol. 10, no. 2-3, pp. 146–162, Aug. 1954, ISSN: 0043-7956, 2373-5112. DOI: 10.1080/00437956.1954.11659520. [Online]. Available: http://www.tandfonline.com/doi/full/10.1080/00437956.1954.11659520.

[40] G. Lapesa, S. Evert, and S. Schulte im Walde, "Contrasting Syntagmatic and Paradigmatic Relations: Insights from Distributional Semantic Models," in *Proceedings of the Third Joint Conference on Lexical and Computational Semantics (*SEM 2014)*, Dublin, Ireland: Association for Computational Linguistics and Dublin City University, 2014, pp. 160–170. DOI: 10.3115/v1/S14-1020. [Online]. Available: http://aclweb.org/anthology/S14-1020.

[41] B. Sagot, "The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French," in *7th International Conference on Language Resources and Evaluation (LREC 2010)*, 2010.

[42] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn," *Front. Neuroinform.*, vol. 8, 2014, ISSN: 1662-5196. DOI: 10.3389/fninf.2014.00014. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fninf.2014.00014/full.

[43] E. C. Merkle, D. You, and K. J. Preacher, "Testing nonnested structural equation models.," *Psychological Methods*, vol. 21, no. 2, pp. 151–163, 2016, ISSN: 1939-1463, 1082-989X. DOI: 10.1037/met0000038. [Online]. Available: http://doi.apa.org/getdoi.cfm?doi=10.1037/met0000038.

[44] A. Freitas, S. Barzegar, J. E. Sales, S. Handschuh, and B. Davis, "Semantic relatedness for all (languages): A comparative analysis of multilingual semantic relatedness using machine translation," in *European Knowledge Acquisition Workshop*, Springer, 2016, pp. 212–222.

[45] F. Hill, R. Reichart, and A. Korhonen, "SimLex-999: Evaluating Semantic Models With (Genuine) Similarity Estimation," *Computational Linguistics*, vol. 41, no. 4, pp. 665–695, Sep. 25, 2015, ISSN: 0891-2017. DOI: 10.1162/COLI_a_00237. [Online]. Available: https://doi.org/10.1162/COLI_a_00237.

[46] E. Agirre, E. Alfonseca, K. Hall, J. Kravalova, M. Paşca, and A. Soroa, "A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches," in *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, ser. NAACL '09, Stroudsburg, PA, USA: Association for Computational Linguistics, 2009, pp. 19–27, ISBN: 978-1-932432-41-1. [Online]. Available: http://dl.acm.org/citation.cfm?id=1620754.1620758.

[47] J. Rademacher, P. Morosan, T. Schormann, A. Schleicher, C. Werner, H. .-.-J. Freund, and K. Zilles, "Probabilistic Mapping and Volume Measurement of Human Primary Auditory Cortex," *NeuroImage*, vol. 13, no. 4, pp. 669–683, Apr. 1, 2001, ISSN: 1053-8119. DOI: 10.1006/nimg.2000.0714. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1053811900907146.

[48] M. Tervaniemi and K. Hugdahl, "Lateralization of auditory-cortex functions," *Brain Research Reviews*, vol. 43, no. 3, pp. 231–246, Dec. 1, 2003, ISSN: 0165-0173. DOI: 10.1016/j.brainresrev.2003.08.004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0165017303002236.

[49] M. Visser, K. V. Embleton, E. Jefferies, G. J. Parker, and M. A. L. Ralph, "The inferior, anterior temporal lobes and semantic memory clarified: Novel evidence from distortion-corrected fMRI," *Neuropsychologia*, vol. 48, no. 6, pp. 1689–1696, May 1, 2010, ISSN: 0028-3932. DOI: 10.1016/j.neuropsychologia.2010.02.016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0028393210000709.

[50] M. Visser, E. Jefferies, K. V. Embleton, and M. A. Lambon Ralph, "Both the Middle Temporal Gyrus and the Ventral Anterior Temporal Area Are Crucial for Multimodal Semantic Processing: Distortion-corrected fMRI Evidence for a Double Gradient of Information Convergence in the Temporal Lobes," *Journal of Cognitive Neuroscience*, vol. 24, no. 8, pp. 1766–1778, May 23, 2012, ISSN: 0898-929X. DOI: 10.1162/jocn_a_00244. [Online]. Available: https://doi.org/10.1162/jocn_a_00244.

[51] T. Wheatley, J. Weisberg, M. S. Beauchamp, and A. Martin, "Automatic Priming of Semantically Related Words Reduces Activity in the Fusiform Gyrus," *Journal of Cognitive Neuroscience*, vol. 17, no. 12, pp. 1871–1885, Dec. 1, 2005, ISSN: 0898-929X. DOI: 10.1162/089892905775008689. [Online]. Available: https://doi.org/10.1162/089892905775008689.

[52] Turken, U., and N. F. Dronkers, "The neural architecture of the language comprehension network: Converging evidence from lesion

and connectivity analyses," *Front Syst Neurosci*, vol. 5, p. 1, 2011, ISSN: 1662-5137. DOI: 10.3389/fnsys.2011.00001. pmid: 21347218.

[53] A. D. Friederici, "The cortical language circuit: From auditory perception to sentence comprehension," *Trends in Cognitive Sciences*, vol. 16, no. 5, pp. 262–268, May 1, 2012, ISSN: 13646613. DOI: 10.1016/j.tics.2012.04.001. pmid: 22516238. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661312000794.

[54] P. Hagoort, "On Broca, brain, and binding: A new framework," *Trends in Cognitive Sciences*, vol. 9, no. 9, pp. 416–423, Sep. 1, 2005, ISSN: 1364-6613. DOI: 10.1016/j.tics.2005.07.004. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364661305002093.

[55] G. Lewis and D. Poeppel, "The role of visual representations during the lexical access of spoken words," *Brain and Language*, vol. 134, pp. 1–10, Jul. 2014, ISSN: 0093934X. DOI: 10.1016/j.bandl.2014.03.008. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0093934X14000467 (visited on 12/14/2019).