ARTIGO

Functionality of the N400 component and its application in studies of figurative language processing: a systematic review

Funcionalidade do componente N400 e aplicação em estudos de processamento figurativo: uma revisão sistemática

Edgard Pereira Neves
pnevesedgard@gmail.com
Universidade Federal do ABC – UFABC

André Mascioli Cravo
andrecravo@gmail.com
Universidade Federal do ABC – UFABC

Maria Teresa Carthery-Goulart
teresa.carthery@gmail.com
Universidade Federal do ABC – UFABC

Linguagem em Foco
Revista do Programa de Pós-Graduação em Linguística Aplicada da UECE

FLUXO DA SUBMISSÃO
Submissão do trabalho: 31/10/2021
Aprovação do trabalho: 08/02/2022
Publicação do trabalho: 11/03/2022

doi 10.46230/2674-8266-13-7263

COMO CITAR
NEVES, Edgard Pereira; CRAVO, André Mascioli; CARTHERY-GOULART, Maria Teresa. Functionality of the N400 component and its application in studies of figurative language processing: a systematic review. **Revista Linguagem em Foco**, v.13, n.4, 2021. p. 204-229. Disponível em: https://revistas.uece.br/index.php/linguagememfoco/article/view/7263.

Abstract
Electrophysiological studies have increased in the field of Neuroscience of Language as they allow for detailed analyses of the temporal course of language processing. In figurative language research, the most employed technique is Event Related Potentials (ERPs). The N400 is the most studied ERP component and is usually described as a marker of difficulty or effort in semantic processing. However, many controversial results about N400 have been found. Two main reasons can be raised for these results: (1) methodological difficulties in controlling psycholinguistic variables that are known to affect language processing and (2) different interpretations regarding what the N400 indexes. In this systematic review our main goals are: (1) providing a general panorama of studies on the N400 component applied to metaphoric and idiomatic language and (2) discussing the different views adopted by researchers regarding functional N400 theories and how ERP studies of figurative processing fit into each of those theories. We found two main interpretations about the process which is indexed by the N400: N400 as a marker of lexical access or as a marker of semantic integration. We also found that this dichotomy has an impact on the choice of models to explain the results about figurative language processing.

Keywords
N400. Figurative processing. ERP. Metaphor. Idiom. Lexical access.

Resumo
Os estudos eletrofisiológicos permitem análises detalhadas sobre o

curso temporal do processamento da linguagem. Na pesquisa em linguagem figurativa, a técnica mais empregada são os Potenciais Relacionados a Eventos (ERPs). O N400 é o componente ERP mais estudado e geralmente é descrito como um marcador de dificuldade ou esforço no processamento semântico. No entanto, muitos resultados controversos sobre o N400 foram encontrados. Duas razões principais podem ser levantadas para esses resultados: (1) dificuldades metodológicas no controle de variáveis psicolinguísticas que sabidamente afetam o processamento da linguagem e (2) diferentes interpretações sobre o que o N400 indexa. Nesta revisão sistemática nossos principais objetivos são: (1) fornecer um panorama geral dos estudos sobre o componente N400 aplicado à linguagem metafórica e idiomática e (2) discutir as diferentes visões adotadas por pesquisadores em relação às teorias funcionais do N400 e como o ERP estuda o processamento figurativo se encaixam em cada uma dessas teorias. Encontramos duas interpretações principais sobre o processo que é indexado pelo N400: N400 como um marcador de acesso lexical ou como um marcador de integração semântica. Também descobrimos que essa dicotomia tem impacto na escolha de modelos para explicar os resultados sobre o processamento da linguagem figurativa.
Palavras-chave
N400. Processamento figurativo. ERP. Metáfora. Idiomático. Acesso lexical.

Introduction

Figurative language is defined as the use of a word in a phrase or expression in an alternative meaning or sense in relation of literal use, such as “*it’s raining cats and dogs”* or “*He broke my heart”*. This is an important topic in language research due to its common occurrence in everyday language as well as the fact that it persists as a universal manifestation across different cultures (LAKOFF, 1979; LAKOFF; JONHSON, 1980). These notorious aspects of figurative language lead to the development of important theories in Philosophy, Psycholinguistics, Cognitive Psychology and Neuroscience of Language on how humans process these types of stimuli.

Grice (1975) and Searle (1979) proposed that we need additional cognitive mechanisms to process figurative language compared to literal meanings. At the same period, Lakoff and Jonhson (1980) proposed that figurative language is more than a simple linguistic tool. In their view, the nature of mind is metaphorical, which implicates the existence of a general mechanism to access meaning in a figurative way.

The impact of these theories led to the emergence of research focusing on figurative processing in the brain. Since 1980, several models have been proposed to explain how the brain processes figurative sentences using different research methods. The main goal has been to explore differences between comprehension of literal and figurative stimuli in terms of stages and levels of processing and what the neural mechanisms underlying this ability are. In general, the main models proposed to explain figurative processing are the Hierarchical Hypothesis, the Direct View Hypothesis and the Graded Salience Hypothesis of figurative processing. In the first view, figurative language has higher processing costs than literal processing and involves more neural resources or pathways. In the Direct View Hypothesis , it is proposed that, with sufficient context, figurative language is processed similarly as literal language (GIBBS, 2001). Finally, the Graded Salience model proposes that the pattern of processing of figurative content is also influenced by other factors that facilitate the comprehension (GIORA, 1999, 2002).

Reaction time has been the main quantitative measurement applied and investigated in figurative language studies. Despite being very informative, this complex measure requires careful interpretation, given that it involves additional processes such as perceptual and motor components (FEDERMEIER, 2014). Particularly for figurative language, studies based on measuring reaction times may lead to incomplete conclusions since they are less sensitive to be mapped into a complex set of variables (FEDERMEIER, 2014).

To study brain responses more precisely and complement the reaction time data, there has been an increase of research on the past decades that combine electrophysiological measurements analyzed in combination with reaction times (THOMA; DAUM, 2006).

Despite the numerous methods in electrophysiology, event related potentials (ERP) has been the most employed procedure on language studies (SANEI; CHAMBERS, 2007). This technique exploits the optimal temporal resolution of EEG to provide an electrical pattern related to stimuli presentation (LUCK, 2014; SANEI; CHAMBERS, 2007). Generally, the ERP analysis considers the polarity and amplitude of electrical peaks following stimulus onset. Depending on the aim of the study, additional analyses of latency and topography can also be considered (LUCK, 2014).

Some peaks – the so-called ERP components – are repeatedly found for particular types of stimuli. The most well-known component in language literature, the N400, is described as a deflection with a peak close to 400 milliseconds classically related to anomalous association at the last word of a sentence (FEDERMEIER, 2014). The N400 was first described by Kutas e Hillyard (1980), starting a long series of articles reporting the N400 in linguistic stimuli related to dissociation in word pairs, anomalous sentence finalization, violation in high constraint contexts and unexpected words (FEDERMEIER, 2014). Beyond language, the N400 was also found in numerical cognition, face and object recognition paradigms tasks and semantic memory (FEDERMEIER, 2014).

The N400 is possibly the most well studied ERP in language and several studies have shown its relation to different aspects of language comprehension (for a review see FEDERMEIER, 2014; KUTAS; HILLAYARD, 1984; HAGOORT; LEVINSON, 2004; LAU; PHILLIPS; POEPPEL, 2008). Here, we will focus on how the N400 has been used to investigate figurative processing. Currently, there are two main views about what N400 indexes within this framework: (I) an integrative account, in which the N400 is depicted as representing a relatively late processing due to semantic integration of critical words with their respective previous context (FEDERMEIER, 2014; LAU et al., 2008); (II) a lexical account, in which the N400 is a marker of lexical retrieval from memory in addition to previous context (FEDERMEIER, 2014; LAU et al., 2008).

In order to clarify these issues, the current review approaches the different information conveyed by the N400 in recent figurative processing studies. Our main objectives are: (1) to show a general panorama of studies on the N400 component applied to metaphoric and idiomatic language, summarizing types of task regarding modality (visual or auditory), language demands of the task, form of analyses (spatial or temporal), conditions, sentence extension, context support and control of independent variables (psycholinguistic factors); (2) to discuss the different views adopted by researchers in the field regarding functional N400 theories and how ERP studies of figurative processing fit to each theory.

Method

We conducted a systematic review of ERP studies in figurative language processing. The search was performed on two databases in January 2018: PubMed and the National Library of Medicine of the United States of America. The search was guided by the following keywords, using “AND” only as a boolean operator: “ERP AND metaphor”; “N400 AND metaphor”; “ERP AND idiom”; “N400 AND idiom”; “ERP AND figurative language”; “N400 AND figurative language”.

No specific time-window was established for the database search. We excluded studies that did not use N400 data in debates on how the brain processes figurative stimuli, as well as studies using fMRI, PET, CT, NIRS experiments (except if concomitant with EEG) were. Apart from that, we excluded studies on irony, proverbs, metonymy and other types of figurative language that do not include metaphors and/or idiomatic expressions to restrict the research and discussion to relative similar figurative expression structures (idioms and metaphor expression can build differing only aspects as familiarity). Studies that used figurative language to analyze other aspects of language and cognition such as bilingualism, gesture processing, psychiatric syndromes, philosophy, and religion were also excluded. Finally, we excluded all works that did not show a clear ERP methodology or minimal aspects of the procedure.

Results

After organization and removal of all duplicates, 69 different articles were selected by title and abstract. We excluded 38 after reading the full text as they did not correspond to the criteria stated in methodology (Table 1). Another, 14 studies were excluded for not being related to figurative processing with ERPs; 11 articles were excluded due to irrelevant discussion about figurative language relative to our scope (i.e. religion, gesture processing and affective debate which do not fit in review scope); eight studies were excluded due to their focus on other types of figurative language (not regarding metaphors or idioms); three studies were excluded because they used TMS or fMRI instead of EEG. The PubMed database provided the majority of studies (for all tag searches) approached in this review, all results were consulted and accessed in January 2018.

Table 1 - Results of each tag search and number of selected articles

Descriptor	**PubMed Result**
ERP AND metaphor	43
N400 AND metaphor	31
ERP AND idiom	7
N400 AND idiom	8
ERP AND figurative language	20
N400 AND figurative language	19
ERP AND non-literal	26
N400 AND non-literal	18
Total of different articles	69
Total of selected articles	32

Table 1 Result of the search in PubMed database for each keyword and the total selected articles according to the criteria used.

Despite some methodological differences among the selected studies, the general experimental design shared similarities, as summarized in Table 2.

In almost all of these studies, visual stimuli were displayed on a monitor; only two studies were performed with auditory stimuli (LORUSSO et al., 2015; WEILAND; BAMBINI; SCHUMACHER, 2014), with one of them using masked prime word on monitor and sentences in earphones in a cross-modal manner (WEILAND et al., 2014).

In the majority experiments, stimuli consisted of sentences which were presented word-by-word with variable intervals between words specific to each experiment. This paradigm is named Rapid Serial Visual Presentation (RSVP), in which stimuli are presented serially and the volunteer makes the decision at the end of sequence. In some studies, however, stimuli were presented as entire sentences followed by a target sentence (LAURENT, 2006; RUTTER et al., 2012; SOTILLO et al., 2005), fragment by fragment (YANG; XUE, 2011), word pairs (ARZOUAN; GOLDSTEIN; FAUST, 2007a, 2007b; FORGÁCS; LUKÁCS; PLÉH, 2014; GOLDSTEIN; ARZOUAN; FAUST, 2012; STRANDBURG et al., 1997; WEILAND et al., 2014), or logogram pairs in a Chinese study (ZHAO et al., 2011). Also, there were cases in which stimuli consisted of pictorial cues that guided a semantic relation judgment with a target word (LU; ZHANG, 2012; MA et al., 2016). Furthermore, from the selected articles (32) experiments were applied to native speakers of English (11) and Chinese (6). A smaller number of studies were conducted in Hebrew (4), Italian (4), German (3), French (3), Dutch (1) and Spanish (1).

The stimuli (sentences) held different extensions and combinations of conditions across studies. The most frequent comparisons were between conventional and unconventional metaphors and between literal and anomalous metaphors. But there were also experiments that used idiomatic stimuli in comparison with literal and anomalous sentences (see Table 3).

The studies reported different sets of measurements of psycholinguistic features of stimuli and conditions (Table 3). The most common variable registered (due to its possible effect on N400 amplitudes) was the participants' familiarity with the stimuli. This feature was accessed using psycholinguistic scales of familiarity, conventionality, frequency, and cloze probability (obtained from Likert scales in each study or linguistic databases). The cloze probability is an important factor associated with N400 and is often mentioned in N400 linguistic experiments. This concept refers to the probability of a sentence or string of words to be completed or followed by a specific word. For example, in the sentence "The dentist recommends brushing your teeth twice a ____" the probability of the "day" is much higher than "year" to fill the targeted gap (BLOCK.; BALDWIN, 2010).

Also, several experiments controlled variables such as: word-length of critical words (number of characters or syllables), sentence length (number of words in a sentence) and grammatical- and orthographical category of the stimuli (variables were summarized in Table 3).

In general, the studies analyzed the differences between N400 amplitudes across conditions (summarized in Table 4). The N400 time window considered for analysis across all studies shows some variability; different time windows were found across studies, with a majority of nine studies working with a time-window of 300-500 milliseconds (COULSON; VAN PETTEN, 2002; FORGÁCS et al., 2015; LU; ZHANG, 2012; SCHNEIDER et al., 2014; TZUYIN LAI; CURRAN, 2013; VESPIGNANI et al., 2010; ZHAO et al., 2011; ZHOU; ZHOU; CHEN, 2004). In addition, various studies used other ERP components as further evidence to make inferences regarding figurative comprehension. The Late Positive Component (LPC) and the P600 were the components most commonly used (summarized in Table 4).

Almost all studies recorded behavioral measurements in their tasks, only a few did not register behavioral responses (LU; ZHANG, 2012; MA et al., 2016; PYNTE et al., 1996; SOTILLO et al., 2005; TARTTER et al., 2002; YANG; XUE, 2011).Most studies measured accuracy (ACC), reaction time (RT) or both (Table 4). Temporal analyses on the N400 were performed in all studies, generally investigating amplitudes to each type of stimulus.

In addition, some articles investigated the latency of the N400 or used time-frequency analysis (MA et al., 2016; ROMMERS; DIJKSTRA; BASTIAANSEN, 2013). Only a subset of articles included additional spatial analyses on N400, as an attempt to identify brain areas involved or evidence of hemisphere differences in figurative processing. This was generally done by hemifield stimuli presentation analysis (COULSON, 2007) and use of Low Resolution Brain Electromagnetic Tomography (LORETA) (ARZOUAN et al., 2007b; PROVERBIO et al., 2009; SOTILLO et al., 2005). Further analysis of N400 was based on the use of the Fuzzy C-Mean Algorithm (FCMA)(ZHOU et al., 2004) and Statistical Parametric Mapping (SPM) (ARZOUAN et al., 2007b; LORUSSO et al., 2015; ZHOU et al., 2004). A total of two studies used a mixed analysis with Near-Infrared spectroscopy (NIRS)(SCHNEIDER et al., 2014, 2015).

Although different studies focused on different time windows, almost all found that metaphors evoke a stronger N400, i.e a more negative amplitude) than literal and idiomatic sentences. The most common encountered result was graded N400 amplitudes across conditions, with a stronger N400 for anomalous sentences, an intermediate amplitude to metaphors and an even smaller or absent N400 to literal sentences (ARZOUAN et al., 2007a, 2007b; COULSON; VAN

PETTEN, 2002; DE GRAUWE et al., 2010; GOLDSTEIN et al., 2012; LORUSSO et al., 2015; PYNTE et al., 1996; RUTTER et al., 2012; SCHMIDT-SNOEK et al., 2015).

This pattern also occurred when the metaphor condition was divided into conventional and unconventional metaphors, this difference is linked with how frequent a given metaphor is used. Some studies, have shown differences between variations or categories of metaphors, e.g. conventional metaphors having a smaller N400 negativity than novel ones (ARZOUAN et al., 2007a, 2007b). Likewise, there are studies that did not find differences between novel, conventional and anomalous metaphors (LAI; CURRAN; MENN, 2009).

In relation to the N400 evoked by idiomatic expressions it was found that type of stimulus can evoke even less negativity than literal sentences (LAURENT et al., 2006; STRANDBURG et al., 1997). Only one experiment did not find this difference (PROVERBIO et al., 2009). Additionally, it was found the more lexicalized the idiom is, the less negativity it evokes.

Concerning these results about metaphors and idioms, it is important to remember that they are attached to specific models. The literature references used to support and explain their results were based on a variety of models. In general, indirect views of figurative processing were mainly considered in experiments. Basically, a direct view and indirect view of figurative language processing defends differences concerning to phases and time of processing. The direct view of figurative language makes an appoint that figurative language have no difference relative to literal one to be processed and indirect view argues that figurative processing has different phases do be processed in relation of literal sentence. In indirect view, that being said, mapping approaches were the most frequently mentioned models (ARZOUAN et al., 2007a, 2007b; COULSON, 2007; COULSON; VAN PETTEN, 2002; LAI et al., 2009; LU; ZHANG, 2012; RUTTER et al., 2012; SCHNEIDER et al., 2016; TZUYIN LAI; CURRAN, 2013; WEILAND et al., 2014; ZHAO et al., 2011).

There are, however, citations of other models such as Graded Salience Hypothesis (DE GRAUWE et al., 2010; GOLDSTEIN et al., 2012; LAURENT et al., 2006) or Configuration Hypothesis (VESPIGNANI et al., 2010) for idioms and also – but less frequently – authors that support a direct view about figurative processing. Lastly, it is important to mention that not all authors explicitly chose or mentioned a model; there are also studies that raise discussions with other aims, e.g. describing the profile of different clinical subgroups without any specific concern on cognitive models about figurative language. These models will be better discussed in next sections of this review.

Another important matter was how authors explained their results by adopting (explicitly or not), a particular view on the functionality of N400. Seven authors explicitly considered that N400 indexes post-lexical processes related to some level of semantic integration (COULSON; VAN PETTEN, 2002; GOLD; FAUST; GOLDSTEIN, 2010; LU; ZHANG, 2012; MA et al., 2016; PYNTE et al., 1996; SCHMIDT-SNOEK et al., 2015; YANG; XUE, 2011; ZHAO et al., 2011); 11 authors explicitly considered N400 as a marker of lexical retrieval from memory (ARZOUAN et al., 2007a, 2007b; BAMBINI et al., 2016; DE GRAUWE et al., 2010; GOLDSTEIN et al., 2012; IAKIMOVA et al., 2009; LAURENT et al., 2006; ROMMERS et al., 2013; RUTTER et al., 2012; VESPIGNANI et al., 2010; WEILAND et al., 2014; ZHOU et al., 2004). One of the authors held the view that N400 could be a response for an even more initial processing, close to something during lexical processing (TARTTER et al., 2002). However others have denied this view (LAI et al., 2009; TZUYIN LAI; CURRAN, 2013). There were other authors that did not explicitly choose any view on N400 indexing. Some studies displayed difficulties in interpreting what the N400 theory would specifically represent (COULSON, 2007; FORGÁCS et al., 2014; LORUSSO et al., 2015; PROVERBIO et al., 2009; SCHNEIDER et al., 2014, 2015; SOTILLO et al., 2005; STRANDBURG et al., 1997).

Table 2 - Results of characterization of experiment procedures in metaphor and idioms studies

Author	**modality of presentation**	**stimuli presentation**	**type of task**	**Groups**	Task language
ARZOUAN; GOLDSTEIN; FAUST, 2007.	Visual	word pair	semantic judgment	1 group	Hebrew
ARZOUAN; GOLDSTEIN; FAUST, 2007.	Visual	word pair	semantic judgment	1 group	Hebrew
BAMBINI, 2016.	Visual	word-by-word	semantic judgment	1 group	Italian
COULSON, 2007.	visual, hemifield presentation	word-by-word	semantic judgment (sentences question prove)	1 group	English
COULSON; VAN PETTEN, 2002.	Visual	word-by-word	semantic judgment	1 group	English
DE GRAUWE, 2010.	Visual	word-by-word	semantic judgment	1 group	English
FORGÁCS et al., 2015.	Visual	word pairs	semantic judgment	1 group	English
GOLD; FAUST; GOLDSTEIN, 2010.	Visual	word pairs	semantic judgment	2 groups (Asperger, controls)	Hebrew
GOLDSTEIN; ARZOUAN; FAUST, 2012.	Visual	word pairs	semantic judgment	1 group	Hebrew
IAKIMOVA et al., 2005.	Visual	word-by-word	semantic judgment	2 groups (Schizophrenic , controls)	French
KAZMERSKI; BLASKO; DESSALEGN, 2003.	Visual	word-by-word	semantic judgment	3 groups (low IQ, medium IQ, high IQ)	English
LAI; CURRAN; MENN, 2009.	Visual	word-by-word	semantic judgment (in 4 levels)	1 group	English
LAURENT, 2006.	Visual	(prime)+ sentence + last word	semantic judgment	1 group	French
LORUSSO et al., 2015.	Auditory	Speech sequence	semantic judgment	3 groups (Language impairment, Nonverbal impairment, Control)	Italian
LU; ZHANG, 2012.	Visual	picture cue and word	semantic judgment and lexical decision	1 group	Chinese
MA et al., 2016.	Visual	picture cue and word	appropriateness judgment	1 group	Chinese
PROVERBIO, 2009.	Visual	word-by-word	semantic judgment	1 group	Italian
PYNTE, et al. 1996.	Visual	word-by-word	semantic judgment	1 group	French
ROMMERS; DIJKSTRA; BASTIAANSEN, 2013.	Visual	word-by-word	lexical decision (target last word) and read for comprehension without task	1 group	Dutch
RUTTER, B. et al., 2012.	Visual	sentence +last word	usualness and appropriateness judgment	1 group	German
SCHMIDT-SNOEK et al., 2015.	Visual	sentence+ last word	sematic judgment	1 group	English
SCHNEIDER, 2014.	Visual	sentence+ last word	sematic judgment	1 group	German
SCHNEIDER, 2015.	Visual	word-by-word	semantic judgment	2 groups (Schizophrenic, controls)	German
SOTILLO, 2005.	Visual	complete sentence + Word target	sematic relation judgment	1 group	Spanish
STRANDBURG, 1997.	Visual	word pairs	semantic judgment	2 groups (Schizophrenic, controls)	English
TARTTER, et al., 2002.	Visual	word-by-word	semantic judgment	1 group	English
TZUYIN LAI; CURRAN, 2013.	Visual	prime + word-by-word	prime task semantic judgment	1 group	English
VESPIGNANI, 2010.	Visual	word-by-word (with two points of analyses)	semantic judgment	1 group	Italian
WEILAND; BAMBINI; SCHUMACHER, 2014.	visual and auditory (cross modal)	visual prime, auditory sentences word pairs	semantic judgment	1 group	German
YANG; XUE, 2011.	Visual	fragment by fragment	semantic judgment	1 group	Chinese
ZHAO, 2011.	Visual	word pairs (Logogram)	semantic judgment	1 group	Chinese
ZHOU; ZHOU; CHEN, 2004.	Visual	character-by-character (logogram)	semantic judgment	1 group	Chinese

Table 3 - Conditions and psycholinguist features

author	Conditions used (principal experiment)	Factors that modify conditions	extension – sentence e.g.	Controlled psycholinguist variables.
ARZOUAN; GOLDSTEIN; FAUST, 2007. a	literal; conventional metaphor; novel metaphor; unrelated	0	lucid mind	metaphoricity; familiarity; novelty; frequency; length; class; grammars.
ARZOUAN; GOLDSTEIN; FAUST, 2007. b	literal; conventional metaphor; unfamiliar metaphor; unrelated semantic	0	burning + fire	familiarity; frequency; concreteness; length; class;
BAMBINI, 2016.	literal; metaphor	minimal context; support context; familiar; unfamiliar	**minimal context**: It is a shark; **supportive context**: That fish is really aggressive. It is a shark	familiarity; cloze; orthography; length; frequency;
COULSON, 2007.	high-cloze literals; low-cloze literals; low-cloze metaphors	0	**Sentence** When he stepped into the elevator, he knew Jennifer had just been there by the smell of her perfume. **Question** He saw Jennifer in the elevator. (T/F)	cloze; frequency; length
COULSON; VAN PETTEN, 2002.	literal; metaphor; literal mapping	0	He + knows + that + power + is a + strong + intoxicant.	cloze; metaphoricity;
DE GRAUWE, 2010.	literal; metaphor; anomalous	minimal context; support context	Cholera is a plague; Cholera is a plague that affects many people.	cloze; familiarity; figurativeness; plausibility; syntactic structure
FORGÁCS et al., 2015.	metaphorical adjective; concrete adjective; abstract adjective	0	**magnetic** (metaphorical adjective) + **brain** (noun); **Slimy** (concrete adjective) + **brain**(noun)	concreteness; literalness; meaningless; Latent Semantic Analysis;
GOLD; FAUST; GOLDSTEIN, 2010.	literal; conventional metaphor; novel metaphor; unrelated	0	firm + words; blossoming + smile	familiarity; concreteness; frequency; length; grammatical category
GOLDSTEIN; ARZOUAN; FAUST, 2012.	literal; conventional metaphor; novel metaphor; unrelated words	conventionalization process	lucid + mind	familiarity; frequency; concreteness; length; class;
IAKIMOVA et al., 2005.	literal; metaphor; incongruous	0	This girl + is a + little + fairy	familiarity; frequency; cloze; length word; length sentences.
KAZMERSKI; BLASKO; DESSALEGN, 2003.	literal; metaphorical; scrambled	0	The + beaver + is a + lumberjack	familiarity; interpretability; cloze (posterior); syntactic structure.
LAI; CURRAN; MENN, 2009.	literal; conventional metaphor; novel metaphor; anomalous	0	The path turned in a new direction + ROAD (target or source)	familiarity; interpretability; cloze; length; syntactic structure;
LAURENT, 2006.	weakly idiom; strong idiom; literal; literal and idiom filler	target related; target unrelated	Rendre les + Armes (final word) + Abandonner (target)	cloze; familiarity; compositionality; literality; idiomaticity; frequency of last word; latent semantic analyses, length all words in sentence;
LORUSSO et al., 2015.	concrete; abstract; unrelated concrete; unrelated abstract	0	to prepare a coffee (auditory presentation)	concreteness; frequency; imageability (verb); age of acquisition syntactic structure.
LU; ZHANG, 2012.	literal-only; metaphor-only; met-lit; control	picture parts of body (ten types)	palm (figure)+ hand (critical word)	frequency; familiarity;

MA et al., 2016.	appropriate metaphor; unappropriate metaphor	sport car; minicar	appropriateness (e.g., sport utility vehicles + tiger) vs. inappropriateness (e.g., sport utility vehicles+cat)	appropriateness.
PROVERBIO, 2009.	idiomatic; literal	associated; not associated	To solve the problem I turned to some / experts (idiomatic key) / advice (Target)	comprehensibility; familiarity; cloze; frequency; concreteness; length o words, sentences and target; syntactic structure.
PYNTE, et al. 1996.	literal; unfamiliar metaphor; familiar metaphor	relevant context; irrelevant context	Those + babies + are + lambs	cloze; familiarity; context; length; syntactic structure;.
ROMMERS; DIJKSTRA; BASTIAANSEN, 2013.	Correct (COR); Related (REL); Unrelated (UNREL)	idiomatic; literal	After + many + transactions +the+ careless+ scammer +eventually +walked +against the + **lamp** + yesterday.	familiarity; cloze; transparency; plausibility; length; syntactic structure;
RUTTER, B. et al., 2012.	high unusual + high appropriate; low unusual + high appropriate; high unusual + low appropriate	0	Die Wolken haben über der Stadt + getanzt.	frequency; appropriateness; usually; length; syntactic structure;
SCHMIDT-SNOEK et al., 2015.	literal; metaphor; anomalous	motion; auditory	His smile + was a + charming **dodge.** (motion) His memoirs + were a + toilet **flush.** (auditory)	cloze; Familiarity; Imageability; Figurativeness; valence; characters; words; content words; frequency; frequency; Concreteness.
SCHNEIDER, 2014.	literal; metaphor; meaningless	0	The one hundred-year-old man is an *oak*	imagery; figurativeness; meaningfulness; familiarity; connotation; grammars, syntactic structure, tense.
SCHNEIDER, 2015.	literal; metaphor; meaningless	0	the old man is a doter	length; syntactic structure; tense; CW frequency.
SOTILLO, 2005.	related; unrelated (sentence + word)	0	Furious tube, that whistles and howls + Train	metaphoricity; familiarity; syllables.
STRANDBURG, 1997.	idiom; literal; nonsense	0	pot + LUCK,	frequency;
TARTTER, et al., 2002.	literal; metaphor; anomalous	0	The orchestra filled the concert hall with sunshine	meaningfulness; figurativeness; cloze;
TZUYIN LAI; CURRAN, 2013.	literal; conventional metaphor; novel metaphor; anomalous	prime; unprime	**target sentence:** Ideas +can +sometimes + be +bumpy; **sentence prime :** I can see the path of his ideas; **simile or literal comparison prime** : Ideas are like roads	familiarity; cloze; interpretability; length of primes and target sentences; syntactic structure;
VESPIGNANI, 2010.	idiomatic; substitution; violation	0	Giorgio + aveva + **un buco**[1] + **nello**[2] + stomaco + quella + mattina	familiarity; cloze; frequency; concreteness; number of characters; grammatical class; age of acquisition;
WEILAND; BAMBINI; SCHUMACHER, 2014.	metaphor; literal control	prime; unprime	These lobbyists are hyenas, if you the kindergarten teacher believe - furry (prime)	familiarity; sensicality; cloze; category cloze; coherence; syllables; word; frequency;
YANG; XUE, 2011.	horizontal metaphor; horizontal metaphor B; vertical metaphor	0	the upper week (in Chinese ideogram pair)	frequency, aspects of temporal; syntactic structure;
ZHAO, 2011.	Attributive metaphor; relational metaphor	0	argument war + life +journey	acceptability; frequency;
ZHOU; ZHOU; CHEN, 2004.	idiom; incongruous	0	shou-zhu-dai-tu	cloze; frequency, n° of ideograms (extension).

Table 4 - Results of mean analyses performed in each experiment

author	Behavioral	ERP Analysis focus:	Additional ERP Analysis	EEG channels	Other ERP components	N400 Window ms
ARZOUAN; GOLDSTEIN; FAUST, 2007.	acc; RT	Temporal approach	0	65	N2;P2; LPC	350-450
ARZOUAN; GOLDSTEIN; FAUST, 2007.	acc; RT	*Spatial and Temporal Approach*	LORETA; Statistical Parametric Mapping (SPM)	65	N400 (350–450 ms) and P600 (550–750 ms)	peak around 425
BAMBINI, 2016.	acc; RT	Temporal approach	topography	64	LPC (560-700)	320-440
COULSON, 2007.	acc; RT	*Spatial and Temporal Approach*	hemifield presentation	29	LPC (600-900)	350-500
COULSON; VAN PETTEN, 2002.	acc; RT[1]	Temporal approach	0	13	LPC	300-500
DE GRAUWE, 2010.	acc	Temporal approach	topography		early N400; late N400; early LPC; late LPC;	325-400-500
FORGÁCS et al., 2015.	acc	Temporal approach	0	26	0	300-500
GOLD; FAUST; GOLDSTEIN, 2010.	RT	Temporal approach applied to neuropsychology issues	topography	64	LPC	380-430
GOLDSTEIN; ARZOUAN; FAUST, 2012.	acc; RT	Temporal approach	0	65	LPC (550-800)	350-450
IAKIMOVA et al., 2005.	acc; RT	Temporal approach applied to neuropsychology issues	0	12	LPC	N400 (control: 200-400; squizo300-500;
KAZMERSKI; BLASKO; DESSALEGN, 2003.	acc; RT	Temporal approach applied to neuropsychology issues	regression IQ x ERP Metaphorical Interference Effect	16	200-400; 400-600; 600-800; 800-1000; 1000-1200; 1200-1400; 1400-1600	200-400; 400-600
LAI; CURRAN; MENN, 2009.	4 levels acc; RT	Temporal approach	0	128	lateN400	320-440
LAURENT, 2006.	acc; RT	Temporal approach	0	12	P600	320-500
LORUSSO et al., 2015.	acc;RT	Spatial and temporal approach applied to neuropsychology issues	Statistical Parametric Mapping (SPM)	19	0	350-500
LU; ZHANG, 2012.	?	Temporal approach	latency analysis	64	lateN400 (500-700)	300-500
MA et al., 2016.	0	*Temporal Approach*	topography; Time-frequency Analysis; (regression between ERP and TF)	64	N300 (280-360); P600 (550-750)	380–460
PROVERBIO, 2009.	RT	*Spatial and Temporal Approach*	LORETA	128	P300 (300-780)	360-590
PYNTE, et al. 1996.	0	Temporal approach	0	7	LPC	300-600
ROMMERS; DIJKSTRA; BASTIAANSEN, 2013.	acc; RT	Temporal approach	analysis of frequency; topography	61	early N400 (300-400); late N400 (400-500); LPC (500-800)	300-400-500
RUTTER, B. et al., 2012.	RT	Temporal approach	0	64	LC (500-900)	350-500
SCHMIDT-SNOEK et al., 2015.	acc; RT	*Temporal Approach*	topography	64	P600 (500-650)	350-500
SCHNEIDER, 2014.	acc; RT	*Spatial and Temporal Approach*	NIRs	21	P200	300-500
SCHNEIDER, 2015.	acc;RT	Spatial and temporal approach applied to neuropsychology issues	NIRs	26	early N400 (280-380) late N400 (380-500); P200	280-500
SOTILLO, 2005.	0	*Spatial Approach*	LORETA	24	0	peak around 500
STRANDBURG, 1997.	acc; RT	Temporal approach applied to neuropsychology issues	PCA to ERP data	17	CNV; P300;	320-580
TARTTER, et al., 2002.	0	Temporal approach	0	31	early N400 (200-300)	300-500
TZUYIN LAI; CURRAN, 2013.	RT	Temporal approach	0	128	0	300-500
VESPIGNANI, 2010.	Acc	Temporal approach	latency analysis	30	P300	300-500
WEILAND; BAMBINI; SCHUMACHER, 2014.	acc	Temporal approach	0	26	0	250-500
YANG; XUE, 2011.	?	Temporal approach	0	32	Various temporal windows	340-500
ZHAO, 2011.	acc; RT	Temporal approach	0	64	P300; LNC; P2; LPC	300-500
ZHOU; ZHOU; CHEN, 2004.	acc	*Spatial Approach*	Fuzzy C-Mean Algorithm Based (FCMA); Statistical Parametric Mapping (SPM)	19	Early Negativities (100-300; ELAN; ERHN); P600	300-500

1. Accuracy (acc) and Reaction Time (RT)

Discussion

We conducted a systematic review of the literature aimed at characterizing the N400 results for metaphor- and idiom processing in light of current functional theories of the N400. To achieve this, we summarized the experimental designs and methodology of those studies and examined their results involving N400 markers.

Our search included studies published up to January 2018, and as we have not determined a specific period for our review, it was possible to identify those studies using EEG to investigate metaphor and idiom processing have begun relatively recently (in the 1990's) and are still scarce (only 32 studies). In addition, most experiments were conducted in European or North American Universities in a reduced number of languages (8 idioms).

General characteristics of ERP figurative studies

The majority of experiments were performed in Germanic languages, less than one third were conducted in Roman languages. In almost one third of studies, the stimuli were presented in Hebrew and Chinese (see Table 2). Above all, it is worth mentioning that no studies were conducted in Portuguese, the second major Roman language. Nonetheless, in our review we found no specific N400 pattern that could be related to the language used in the studies.

Regarding stimuli presentation the majority of studies used visual tasks, such as word-pairs or sentences (presented word-by-word) and subjects were asked to make semantic judgments of the stimuli (i.e. semantic associations or sentence semantic plausibility). It is important to note that these two types of stimuli (word-pairs vs. sentences) differ not only in extension but also in the quantity of context offered for critical words, which is known to affect the comprehension process. For sentences, the cloze probability is currently considered the measure that best depicts the role of the context in predicting the critical word. On the other hand, word pair studies require measures of semantic association between the two items. Whereas most sentence studies informed the cloze probability of the critical word, only a few word-pair studies measured semantic associations (see Table 3). Moreover, word pairs can have different uses according to the language (e.g. in Hebrew some word-pairs can be considered a single expression – Arzouan and Faust, 2000).

We found two studies that used auditory stimuli. Lorusso et. al. (2015) performed a purely auditory task and Weiland and colleagues (2014) mixed visual and auditory presentation in their experiment. It is well known that the N400 is elicited by both modalities (visual and auditory) and that both forms elicit similar responses. Nevertheless, it is important to highlight that upon auditory presentations, the N400 can appear earlier and tends to be more frontal and less right biased (concerning the topography of activation) compared to the "classical" N400 found in visual tasks (FEDERMEIER, 2014; LAU et al., 2008).

Most studies compared three conditions in their experiments: "Literal x Metaphorical x Anomalous" (see Table 3). Some variations were found such as conventional *vs.* unconventional and familiar *vs.* unfamiliar metaphors. Those different designs allow analyses of specific linguistic aspects and their influence on N400 patterns. The use of conditions such as anomalous and literal sentences, conventional and unconventional metaphors identifies factors such as conventionality of a sentence and elucidates how this affects the N400 component and figurative processing in general (ARZOUAN et al., 2007a; GOLDSTEIN et al., 2012; LAI et al., 2009). The same holds for the cloze probability (LAURENT et al., 2006) and familiarity manipulations (DE GRAUWE et al., 2010). Also, cloze probability has been one of the main psycholinguist aspects researched in relation to N400, due to the strong negative correlation between the two measures (r=-0.9, for a review see FEDERMEIER, 2014), representing a measure from the degree of constraint of a word in a specific context (FEDERMEIER, 2014).

In our review, most studies used EEG to investigate temporal aspects of figurative processing and reported results of the N400 component. However, some studies have included localization analyses and other ERP components such as LPC and P600. The combined analysis of N400 and events of late windows is important to inform on different steps during processing. For instance, stronger N400 deflections combined with greater LPC is a biphasic pattern commonly used to assign indirect views of figurative processing (BAMBINI et al., 2016). The absence of this pattern gives other valuable insights that we will discuss in the following section (BAMBINI et al., 2016).

N400 as a marker in figurative models

Virtually all studies of this review have shown differences in the N400, with a stronger N400 for figurative processing than for literal sentences. This is commonly used as evidence to support indirect views on figurative processing, as these dissimilarities are understood as "proof" of different steps in the comprehension process or, at least, different processing patterns between the conditions.

A significant part of experiments reported results of a general pattern wherein the N400 component amplitudes display a graded form with largest amplitudes for anomalous stimuli, followed by metaphors and then literal statements. As mentioned in the results, these conditions are the most commonly reported across studies, with small variations.

This N400 difference between anomalous and metaphorical sentences is particularly important because it cannot be supported either by the Standard Pragmatic Model (whose assumptions have been weakly supported by electrophysiological studies) or by the Direct Model. The Standard Pragmatic Model proposes a step of negation of the literal sense and a later access to the figurative sense of a sentence. If this assumption is correct, metaphorical, and anomalous sentences should generate similar N400 patterns. Likewise, the difference between literal and metaphorical conditions does not grant that these sentences forms are accessed in the same way as proposed by the Direct Model. This gradual pattern of N400 amplitudes denotes that there are other factors influencing the comprehension process.

In our sample of studies, we found minimal variations across stimuli conditions. Most studies controlled sentences or target words on psycholinguistic patterns such as familiarity, cloze probability, conventionality, concreteness, and figurativeness. All of these factors are known to influence reaction times in comprehension tasks and also the N400 amplitudes, which results in different levels of this ERP component (DE GRAUWE et al., 2010; FEDERMEIER, 2014; LAU et al., 2008).

The most cited models in the investigated studies are mapping theories that encompass the graded pattern and these confounding factors (psycholinguistic variations of stimuli). These theories propose that a same system is used to understand figurative and literal concepts, but the comprehension process involves mapping the attributes from the target word to context information in order to establish similarities (COULSON; VAN PETTEN, 2002). In these explanations, N400 amplitudes are a proxy of the amount of activation involved in the process of comparison of two objects in a sentence (or word-pair), that is, the N400 varies according to the effort used to undertake comparisons that depend on how those two objects are distant in meaning one from another.

One of the first ERP studies that considered the Conceptual Mapping Model was conducted by Coulson and Van Petten (2002). They designed three conditions: literal, metaphoric and literal mapping. The literal mapping condition

(Eg."The ring was made of tin, with a pebble instead of a gem") was thought to be an expression of an intermediate mapping effort in which the critical word of a sentence would near certain attributes of context similarly to metaphorical sentences (Eg."After giving it some thought, I realized the new idea was a gem"). Their results confirmed that hypothesis with N400 amplitudes varying across conditions being largest to metaphoric, intermediate for literal mapping and lowest to literal sentences. These results suggest that the higher the distance of two concepts the larger the N400 amplitude.

Other prominent studies have reached the same conclusions but using other experimental settings. Arzouan, Goldstein and Faust (2007) used word-pairs and found that the smallest N400 amplitude occurred in the literal condition and then was progressively larger in the following conditions respectively: conventional metaphor< novel metaphors < unrelated pairs. The authors explained their results in terms of effort rates applied to establish mapping between concepts and support that this is the main aspect modulating the N400 amplitudes. They argued that it is important to maintain the literal meanings for concept mapping and in a later step accessing the metaphorical meaning.

The Graded Salience Hypothesis is another possible explanation for this graded pattern found across studies. According to this model, the initial comprehension process is not really influenced by figurative content but by salience characteristics of sentences and content words. In this view, factors such as familiarity, conventionality, frequency and cloze probability facilitate the retrieval of word knowledge and consequently, sentence comprehension (GIORA, 1999, 2002). In accordance with this model, the N400 is mostly linked with the expectancy offered by the previous context on the critical word, that is, an unexpected word finalization would produce larger N400 amplitudes in comparison to more familiar sentences (salient sentences in this model refer to more familiar sentences) (DE GRAUWE et al., 2010; LAURENT et al., 2006; STRANDBURG et al., 1997). Also, it is possible to see these same patterns considering idioms' studies. The results of Laurent and colleagues (2006) suggest that salience of a sentence is a strong factor modulating the N400. The authors demonstrate in a well-controlled study that compared two categories of idioms (strong and weak idiomaticity) to literal expressions that N400 results followed the graded pattern, in which larger amplitudes were found for weak salience conditions (DE GRAUWE et al., 2010; GOLDSTEIN et al., 2012; LAURENT et al., 2006)

In our review, the Indirect Views (Conceptual Mapping, Structure Mapping and Salience Hypothesis) predominated across studies. However, the Direct

Access View has also received some support. This hypothesis proposes that figurative and literal meanings present the same course of processing as long as they share the same quantity of supportive context. It received some support by the results of Iakimova et. al. (2005), a clinical study, in healthy participants did not find significant differences in ERPs between figurative and literal sentences. In that case, the major objective of this study was to analyses differences in processing between individuals with depression in comparison with healthy participants. In the present review it was very difficult to find studies about figurative processing that lent support to the direct view.

All models of figurative language have received some support of previous ERP studies, showing that ERP data have conflicting findings and that explanations are far from straightforward. It is possible that conflicting findings come not only from the control divergences of psycholinguistic variables and task type, but also from the interpretation that has been given to the N400 marker. For example, some N400 findings (gradual amplitudes across conditions and correlations with linguistic variables) may be equally explained by the Salience and Mapping models. On the other hand, functionality views of the N400 marker corroborates in a different way with each figurative language model. This means that according to the functionality of N400 that is assumed, different conclusions can be drawn.

There are two main theories about the types of processing indexed by the N400. In one view, the N400 is supposed to index a post-lexical process involved in semantic integration of the final word to the previous context. In another view, it is proposed that the N400 is a marker of lexical retrieval (from memory storage) (FEDERMEIER, 2014; KUTAS, 2006; LAI et al., 2009; LAU et al., 2008; NIEUWLAND; VAN BERKUM TZUYIN, 2006; LAI; CURRAN, 2013; WEILAND et al., 2014). Critically, it has also been proposed that both word meaning and world knowledge are recruited and integrated very rapidly by similar brain regions (HAGOORT; LEVINSON, 2004). These finding is particularly important as it suggests that what words and stimuli are used can possibly bias what affects the most the observed N400.

These two views seem similar or even the same phenomena, but they have particularities that may lead to different explanations for the process studied (FEDERMEIER, 2014; FEDERMEIER, WLOTKO; DE OCHOA-DEWALD; KUTAS, 2007; KUTAS, 2006).

Considering the mapping model, the effort for comprehension resides on the process of conceptual integration between items from a sentence. The assumption that the N400 is a marker of semantic integration would point that the attribution of literal or figurative senses is a context integration process that happens in around 400 milliseconds, that is, in a relative early temporal window, considering that semantic integration is a conscious process.

Now, considering the N400 as a marker of lexical retrieval from memory, the mapping theory would produce a different explanation, indicating that around 400 milliseconds an unconscious process would occur: the recovery of word information that matches (or not) with previous context (even when context is just one word). In this view, semantic integration is characterized as late process which might be indexed by other ERP components (as P600, LPC, LPP). In figurative language studies, this type of explanation was explicitly considered just by Lai and colleagues (2009, 2013) that discussed each possibility according with the functional role assumed for the N400.

Taking these functional views into consideration, the salience theory seems a better model for explaining the influence of linguistic factors on N400 amplitudes in the first window of processing. Salient meanings are accessed before than non-salient ones due to a memory facilitation process. Non-salient stimuli would need additional inferential processes best explained by later ERP components and in behavioral measures by larger RTs. In addition, considering the lexical-retrieval view of N400 there is a possibility of some overlap on the type of ERP pattern between the Graded Salience Hypothesis and Conceptual Mapping Views. It is very well established on the literature that lexical retrieval is largely influenced by factors such as cloze probability, familiarity, conventionality what fits with the idea that salience affects N400 amplitudes (FEDERMEIER, 2014; LAU et al., 2008).

GIORA (2002) considers that the salience model represents aspects of an initial step of sentence processing. If the semantic integration view of N400 is correct, this would imply that lexical retrieval occurs in even earlier windows and ERPs in later windows would index other types of event. For the N400 to be a marker of a conscious process it would be necessary to become clear that there is an integration aspect of context on final word responses in experiments. Pynte et al (1996) proposes that context have a central role on the N400. The authors compared familiar metaphors with previous irrelevant context information with unfamiliar metaphors with previous relevant context and found that the irrelevant context condition had more influence on N400 independent manipulations than familiarity. In this study, the N400 was affected mostly by the expectancy offered by context (BAMBINI et al., 2016; GIORA, 2002).

On a well-controlled study Bambini et al. (2016) manipulated the context and found that in conditions of supportive context (see their experiment 2) there are no N400 differences between literal and figurative conditions. In their study, the N400 was greatly suppressed by context and differences on late ERP windows were found. On the other hand, the lack of context (experiment 1) produced a biphasic pattern in which N400 negativity is found followed by a later positive component such as P600 or LPC for metaphorical conditions. That biphasic pattern is recurrent across studies that support indirect views of figurative language, as those patterns of ERPs matches with the idea that the N400 is an index of lexical retrieval and that in later windows other processes would take place (i.e analysis and reanalysis of semantic aspects of sentence) (BAMBINI et al., 2016).

In agreement with this, results of studies with idiomatic expressions show that strongly lexicalized idioms (items with high cloze probability) don't elicit the typical N400 deflections seen in other types of figurative sentences. It is frequently argued that this finding reflects the process in which the entire phrase structure is accessed from memory. Thus, conditions that do not present lexical retrieval would not generate N400 deflections. In weak idiomaticity (low cloze probability) or literal conditions the presence of the N400 deflection is thought to reflect of the role of expectancy of the final sentence word disambiguating the previous context (LAURENT et al., 2006; VESPIGNANI et al., 2010).

Combined, these results also satisfy the proposals of a probabilistic factor affecting word retrieval in written or auditory comprehension processes (eg. reading or listening sentences). Our brains automatically pre-activate upcoming words according to the constraints provided by the previous context in a mechanism that facilitates reading or listening processes (DELONG; URBACH; KUTAS, 2005; FEDERMEIER, 2007; TROYER; KUTAS, 2020). If expectation is broken by an unexpected word (low cloze probability condition) the result is a larger N400 deflection that denotes search mechanisms (from memory) to retrieve the semantic characteristics of this new lexical item in the sentence being processed (DELONG et al., 2005; FEDERMEIER et al., 2007). In addition, the N400 amplitude can be modulated by the semantic relationship between the expected and the unexpected word. The N400 amplitudes are smaller when those items share many semantic features (eg. hammer vs saw that belong to the same semantic category). Thus, when the unexpected word is semantically related to the expected word a reduced N400 effect is usually explained in terms of a facilitation of the lexical retrieval process (WLOTKO; FEDERMEIER, 2015).

The studies sample selected for this review seems to be more supportive of an early automatic processing view of the N400 evoked-response. This support could evidence that figurative processing, if it is indexed by N400, might not comprise conscious processes or integrative difficult of semantic in the previous context.

The assumption that the N400 appears for unexpected words (is a marker of expectations breaches) offers a better explanation for the results seen in figurative sentences in which the N400 varies with some psycholinguistic aspects linked with the frequency and familiarity of some word in certain context (i.e great correlation of N400 with Cloze Probability)(KUTAS; FEDERMEIER, 2014). This explanation is also suitable to explain findings with other grammatical classes (DELONG et al., 2005; WICHA; MORENO; KUTAS, 2004) wherein is thought that there is a general predictive mechanism in language processing that is indexed by the N400 (when the unexpected item appears). But not many studies have tried to replicate this finding for other grammatical classes. In fact, in a recent study NIEUWLAND et al. (2018) have not reproduced Delong et al. (2005) results, especially for articles (gramatical) N400 and refuse the entire probabilistic claim of N400. Nieuwland et al. (2018) discuss that the N400 cannot be explained only by phenomena linked to cloze probability (NIEUWLAND et al., 2018). On the other hand, Delong et al. (2017) responded to the study and raise questions of whether the cloze probability in the replication study was properly controlled (NIEUWLAND et al., 2018)

In our review, it seems that some studies have explicitly explained their results according to the functional view of the N400 functionality - nearly ten studies have interpreted their results considering the semantic integration view (COULSON; VAN PETTEN, 2002; Gold et al., 2010; KAZMERSKI; BLASKO; DESSALEGN, 2003; LU; ZHANG, 2012; MA et al., 2016; PYNTE et al., 1996; SCHMIDT-SNOEK et al., 2015; YANG; XUE, 2011; ZHAO et al., 2011) and about thirteen studies based their results on the lexical-retrieval view (ARZOUAN et al., 2007a, 2007b; BAMBINI et al., 2016; DE GRAUWE et al., 2010; GOLDSTEIN et al., 2012; IAKIMOVA; LAURENT; HARDY-BAYLE, 2005; LAURENT et al., 2006; ROMMERS et al., 2013; RUTTER et al., 2012; VESPIGNANI et al., 2010; WEILAND et al., 2014; ZHOU et al., 2004). Also from all these studies, just three discussed the implications of their findings according to each functionality view (BAMBINI et al., 2016; LAI et al., 2009; TZUYIN LAI; CURRAN, 2013).

It is also important to highlight that many studies don't make this functional approach very clear. Specifically, studies that tried to characterize clinical subgroups or focused on the localization of the N400 did not provide explanations for the cognitive mechanisms that may be implied. In fact, in those studies, no functional view is chosen to explain results (COULSON, 2007; FORGÁCS et al., 2015; LORUSSO et al., 2015; PROVERBIO et al., 2009; SCHNEIDER et al., 2014, 2015; SOTILLO et al., 2005; STRANDBURG et al., 1997). Although this is not a problem in studies not aimed at comprehension of the time-course of cognitive processing, they may lead to incorrect conclusions. For example, if it is assumed that the N400 is indexing lexical retrieval, the absence or different modulation of this component in Schizophrenia would mean that pathology affects processes of lexical retrieval and not processes of semantic integration of stimuli to previous context.

The debate about figurative models is far to be finished. This review emphasizes the need of deeper analyses and interpretations of the N400 functionality to provide more precise insights of electrophysiological applications in figurative language research. In addition, we believe that figurative language research may bring evidence for the functionality of N400.

This can build up certain robust views about functionality of N400 if authors appropriate the functionality debate on experimental design or discuss the N400 in effect studied. Here we show that it is not all authors who consider this issue for results discussion and how this implicates in model explanations (LAI et al., 2009). It is undeniable the importance of N400 theory aspect for support certain figurative processing model.

Final Considerations

Our review aimed to show a general panorama and discussion of the use of N400 on figurative language research. In sum, there are a variety of experimental set and uses of N400 in figurative research and different interpretation of what N400 can evidence or mark. Considering these different views about the N400 marker or even the absence of this discussion in some studies can influence the conclusions about figurative processing. Our results also show that the N400 can show different patterns according to the characteristics of sentences and words used in the stimuli set and the N400 function theory. Our review didn't consider the association of EEG and others methods such as fMRI, NIRS, MEG, which could add more data to the characterization of N400 functionality and elucidate part of this debate. Similarly, other EEG analysis methods, such as time-frequency and LORETA, can advance our understanding of the N400.

Acknowledgements

This research was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) – Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil and by Federal University of ABC (UFABC).

We thank our colleagues from the Human Cognition Laboratory (109 Lab) who provided insights and productive discussions that greatly assisted the research.

We thank Prof. Maria Alice Parente (UFABC) for encouraging the production of this literature review

We thank Prof. Claudinei Biazolli (UFABC) and Prof. Antônio Jaeger (UFMG) for comments that greatly improved the manuscript.

We are also grateful to Hannah Lewerentz and Camila Agostino for their comments on an earlier version of the manuscript.

References

ARZOUAN, Yossi; GOLDSTEIN, Abraham; FAUST, Miriam. Brainwaves are stethoscopes: ERP correlates of novel metaphor comprehension. **Brain research**, v. 1160, p. 69-81, 2007.

ARZOUAN, Yossi; GOLDSTEIN, Abraham; FAUST, Miriam. Dynamics of hemispheric activity during metaphor comprehension: Electrophysiological measures. **Neuroimage**, v. 36, n. 1, p. 222-231, 2007.

BAMBINI, Valentina et al. Disentangling metaphor from context: an ERP study. **Frontiers in psychology**, v. 7, p. 559, 2016.

BLOCK, Cady K.; BALDWIN, Carryl L. Cloze probability and completion norms for 498 sentences: Behavioral and neural validation using event-related potentials. **Behavior research methods**, v. 42, n. 3, p. 665-670, 2010.

CARROLL, DAVID W. Psychology of Language. 5ª ed. 2008.

COULSON, Seana; VAN PETTEN, Cyma. A special role for the right hemisphere in metaphor comprehension?: ERP evidence from hemifield presentation. **Brain research**, v. 1146, p. 128-145, 2007.

COULSON, Seana; VAN PETTEN, Cyma. Conceptual integration and metaphor: An event-related potential study. **Memory & cognition**, v. 30, n. 6, p. 958-968, 2002.

DE GRAUWE, Sophie et al. Electrophysiological insights into the processing of nominal metaphors. **Neuropsychologia**, v. 48, n. 7, p. 1965-1984, 2010.

DELONG, Katherine A.; URBACH, Thomas P.; KUTAS, Marta. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. **Nature neuroscience**, v. 8, n. 8, p. 1117-1121, 2005.

FORGÁCS, Bálint et al. Metaphors are physical and abstract: ERPs to metaphorically modified nouns resemble ERPs to abstract language. **Frontiers in Human Neuroscience**, v. 9, p. 28, 2015.

FORGÁCS, Bálint; LUKÁCS, Ágnes; PLÉH, Csaba. Lateralized processing of novel metaphors: disentangling figurativeness and novelty. **Neuropsychologia**, v. 56, p. 101-109, 2014.

GIBBS JR, Raymond W. Evaluating contemporary models of figurative language understanding. **Metaphor and symbol**, v. 16, n. 3-4, p. 317-333, 2001.

GIORA, Rachel. On the priority of salient meanings: Studies of literal and figurative language. **Journal of pragmatics**, v. 31, n. 7, p. 919-929, 1999.

GIORA, Rachel. Literal vs. figurative language: Different or equal?. **Journal of pragmatics**, v. 34, n. 4, p. 487-506, 2002.

GOLD, Rinat; FAUST, Miriam; GOLDSTEIN, Abraham. Semantic integration during metaphor comprehension in Asperger syndrome. **Brain and Language**, v. 113, n. 3, p. 124-134, 2010.

GOLDSTEIN, Abraham; ARZOUAN, Yossi; FAUST, Miriam. Killing a novel metaphor and reviving a dead one: ERP correlates of metaphor conventionalization. **Brain and language**, v. 123, n. 2, p. 137-142, 2012.

GRICE, Herbert P. Logic and conversation. In: **Speech acts**. Brill, 1975. p. 41-58.

HAGOORT, Peter; LEVINSON, Stephen C. Neuropragmatics. In: **The cognitive neurosciences**. MIT Press, 2014. p. 667-674.

IAKIMOVA, Galina et al. ERPs of metaphoric, literal, and incongruous semantic processing in schizophrenia. **Psychophysiology**, v. 42, n. 4, p. 380-390, 2005.

KEMMERER, David. Are the motor features of verb meanings represented in the precentral motor cortices? Yes, but within the context of a flexible, multilevel architecture for conceptual knowledge. **Psychonomic Bulletin & Review**, v. 22, n. 4, p. 1068-1075, 2015.

KUTAS, Marta; FEDERMEIER, Kara D. Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). **Annual review of psychology**, v. 62, p. 621-647, 2011.

KUTAS, Marta; HILLYARD, Steven A. Reading senseless sentences: Brain potentials reflect semantic incongruity. **Science**, v. 207, n. 4427, p. 203-205, 1980.

KUTAS, Marta; HILLYARD, Steven A. Brain potentials during reading reflect word expectancy and semantic association. **Nature**, v. 307, n. 5947, p. 161-163, 1984.

LAI, Vicky Tzuyin; CURRAN, Tim; MENN, Lise. Comprehending conventional and novel metaphors: An ERP study. **Brain research**, v. 1284, p. 145-155, 2009.

LAKOFF, G. Metaphor and Thought. Cambridge University Press. 1979

LAKOFF, George; JOHNSON, Mark. **Metaphors we live by**. University of Chicago press, 1ª ed. 1980.

LAU, Ellen F.; PHILLIPS, Colin; POEPPEL, David. A cortical network for semantics:(de) constructing the N400. **Nature reviews neuroscience**, v. 9, n. 12, p. 920-933, 2008.

LAURENT, Jean-Paul et al. On understanding idiomatic language: The salience hypothesis assessed by ERPs. **Brain Research**, v. 1068, n. 1, p. 151-160, 2006.

LORUSSO, Maria Luisa et al. Processing sentences with literal versus figurative use of verbs: an ERP study with children with language impairments, nonverbal impairments, and typical development. **Behavioural neurology**, v. 2015, 2015.

LU, Aitao; ZHANG, John X. Event-related potential evidence for the early activation of literal meaning during comprehension of conventional lexical metaphors. **Neuropsychologia**, v. 50, n. 8, p. 1730-1738, 2012.

LUCK, Steven J. **An introduction to the event-related potential technique**. MIT press, 2014.

MA, Qingguo et al. Neural correlates of multimodal metaphor comprehension: evidence from event-related potentials and time-frequency decompositions. **International Journal of Psychophysiology**, v. 109, p. 81-91, 2016.

NIEUWLAND, Mante S.; VAN BERKUM, Jos JA. When peanuts fall in love: N400 evidence for the power of discourse. **Journal of cognitive neuroscience**, v. 18, n. 7, p. 1098-1111, 2006.

PROVERBIO, Alice M. et al. The role of left and right hemispheres in the comprehension of idiomatic language: an electrical neuroimaging study. **BMC neuroscience**, v. 10, n. 1, p. 1-16, 2009.

PYNTE, Joel et al. The time-course of metaphor comprehension: An event-related potential study. **Brain and language**, v. 55, n. 3, p. 293-316, 1996.

ROMMERS, Joost; DIJKSTRA, Ton; BASTIAANSEN, Marcel. Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. **Journal of Cognitive Neuroscience**, v. 25, n. 5, p. 762-776, 2013.

RUTTER, Barbara et al. Can clouds dance? Part 2: An ERP investigation of passive conceptual expansion. **Brain and Cognition**, v. 80, n. 3, p. 301-310, 2012.

SANEI, Saeid; CHAMBERS, Jonathon A. **EEG signal processing**. John Wiley & Sons, 2007.

SCHMIDT-SNOEK, Gwenda L. et al. Auditory and motion metaphors have different scalp distributions: an ERP study. **Frontiers in human neuroscience**, v. 9, p. 126, 2015.

SCHNEIDER, Julie M. et al. Developmental differences in beta and theta power during sentence processing. **Developmental Cognitive Neuroscience**, v. 19, p. 19-30, 2016.

SCHNEIDER, Sabrina et al. Beyond the N400: Complementary access to early neural correlates of novel metaphor comprehension using combined electrophysiological and haemodynamic measurements. **Cortex**, v. 53, p. 45-59, 2014.

SOTILLO, María et al. Neural activity associated with metaphor comprehension: spatial analysis. **Neuroscience letters**, v. 373, n. 1, p. 5-9, 2004.

STRANDBURG, Robert J. et al. Event-related potential correlates of linguistic information processing in schizophrenics. **Biological Psychiatry**, v. 42, n. 7, p. 596-608, 1997.

TARTTER, Vivien C. et al. Novel metaphors appear anomalous at least momentarily: Evidence from N400. **Brain and Language**, v. 80, n. 3, p. 488-509, 2002.

THOMA, Patrizia; DAUM, Irene. Neurocognitive mechanisms of figurative language processing—evidence from clinical dysfunctions. **Neuroscience & Biobehavioral Reviews**, v. 30, n. 8, p. 1182-1205, 2006.

TROYER, Melissa; KUTAS, Marta. Harry Potter and the Chamber of What?: The impact of what individuals know on word processing during reading. **Language, cognition and neuroscience**, v. 35, n. 5, p. 641-657, 2020.

LAI, Vicky Tzuyin; CURRAN, Tim. ERP evidence for conceptual mappings and comparison processes during the comprehension of conventional and novel metaphors. **Brain and Language**,

v. 127, n. 3, p. 484-496, 2013.

VESPIGNANI, Francesco et al. Predictive mechanisms in idiom comprehension. **Journal of Cognitive Neuroscience**, v. 22, n. 8, p. 1682-1700, 2010.

WEILAND, Hanna; BAMBINI, Valentina; SCHUMACHER, Petra B. The role of literal meaning in figurative language comprehension: Evidence from masked priming ERP. **Frontiers in Human Neuroscience**, v. 8, p. 583, 2014.

YANG, Jie; XUE, Jin. Spatial metaphor processing during temporal sequencing comprehension. **Experimental brain research**, v. 213, n. 4, p. 475, 2011.

ZHAO, Ming et al. The neuromechanism underlying verbal analogical reasoning of metaphorical relations: An event-related potentials study. **Brain research**, v. 1425, p. 62-74, 2011.

ZHOU, S.; ZHOU, W.; CHEN, X. Spatiotemporal analysis of ERP during Chinese idiom comprehension. **Brain Topography**, v. 17, n. 1, p. 27-37, 2004.

About the authors

Edgard Pereira Neves - Center of Mathematics, Computing and Cognition, Federal University of ABC (CMCC-UFABC), São Bernardo do Campo (SP), Brazil.

André Mascioli Cravo - Center of Mathematics, Computing and Cognition, Federal University of ABC (CMCC-UFABC), São Bernardo do Campo (SP), Brazil.

Maria Teresa Carthery-Goulart - 1 - Center of Mathematics, Computing and Cognition, Federal University of ABC (CMCC-UFABC), São Bernardo do Campo (SP), Brazil; 2 - Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino (INCT-ECCE), São Carlos (SP), Brazil; 3 - Cognitive and Behavioral Neurology Research Group, Hospital das Clínicas, Department of Neurology, School of Medicine of the University of São Paulo (GNCC-HCFMUSP), São Paulo (SP), Brazil.