Role of stress pattern in production and processing of compound words and phrases in Mandarin Chinese

Weilin Shen

DOCTORAL SCHOOL ED 261

«COGNITION, COMPORTEMENTS, CONDUITES HUMAINES »

Role of stress pattern in production and processing of compound words and phrases in Mandarin Chinese

by

Weilin Shen

PhD Thesis - Cognitive Psychology

Thesis Supervisor: Prof. Frédéric Isel

Committee:

Mireille Besson	Professor in Aix-Marseille University (Rapporteur)
Fanny Meunier	CNRS Research Director of University Lyon 1 (Rapporteur)
Anne Lacheret-Dujour	Professor of Paris Ouest Nanterre (Examiner)
Claire Saillard	Professor of Paris Diderot University (Examiner)
Fougeron Cécile	CNRS Research Director of New Sorbonne University (Examiner)
Frédéric Isel	Professor of Paris Ouest Nanterre (Thesis Supervisor)

This page is intentionally left blank.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Prof. Frédéric Isel, for the continuous support of my PhD study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Mireille Besson, Dr. Fanny Meunier, Prof. Anne Lacheret-Dujour, Prof. Claire Saillard and Dr. Fougeron Cécile for their helpful discussions and encouragement.

My sincere thanks also go to Prof. Jacqueline Vaissière who introduced me to a world of phonetic and for her insightful suggestions and important corrections on the chapter linguistic stress. I would also like to warmly acknowledge Dr. Xinyue Yu, Prof. Claire Saillard and Prof. Bernard Fradin for their valuable guidance and help in the selection of linguistic materials used in the thesis.

Last, I wish to thank my family for their endless love and support, without which no accomplishments can be appreciated. My mother and father: I have been so far away and too long from home, but you have always been with me. I also thank my beloved and invaluable colleagues and friends: Karin Heidlmayr, Mathilde Carminati and Xinyue Yu for their support and encouragement during the writing of this thesis.

This research was supported by the grant of the Doctoral School « Cognition, Comportements, Conduites Humaines » (ED 261).

Role of stress pattern in production and processing of compound words and phrases in Mandarin Chinese

By

Weilin Shen

ABSTRACT

The present doctoral thesis investigates the processing of stress pattern (i.e. word stress versus phrasal stress) and its implication in the auditory processing of disyllabic Mandarin Chinese ambiguous minimal pairs constituted of a compound word and a phrase. Two types of compound word/phrase minimal pairs were used (1) Compound words with a neutral tone on the final syllable (strong-weak) vs. Phrases with a full tone on the final syllable (strong-strong); (2) Verb-Noun (VN) compound words (strong-weak) and Verb-Object (VO) phrases (weak-strong) distinguished by the stress pattern. In a production study, phonetic analyses of the minimal pairs revealed a contrastive stress pattern indexed by longer duration, larger F0 variation and higher intensity in stressed syllables. At the neurophysiological level, the processing of the contrastive stress patterns of the two types of compound word/phrase minimal pairs was associated with two exogenous event-related brain potentials: An N200 thought to reflect an automatic language-specific processing of stress pattern on the one hand, and an P200 assumed to sign the processing of physical/acoustic cues, on the other hand. In addition to these two exogenous ERP components, we showed that one endogenous component, i.e. the N325, varied as a function of the stress pattern in the minimal pairs. Indeed, the N325 was larger for infrequent stress pattern, but this ERP effect was only observed in the stress discrimination task for non-delexicalized stimuli. We therefore argue that the N325 in response to the stress pattern in Mandarin Chinese may be a controlled language-specific ERP marker of stress pattern, possibly somehow related to allocation of attentional resources. Additionally, we found a late sustained MisMatch Negativity (MMN) in response to stress pattern congruency regardless of stress pattern and task. This late MMN suggests that stress pattern changes are processed relying on long-term representation of stress, not at single syllable level. Finally, our neurophysiological data showed that the neutral tone in the

minimal pairs was used by the processing system for assisting the analysis of the morpholexical structure of the ambiguous spoken sequences in Mandarin Chinese under investigation here. We demonstrated that once a neutral tone was detected, a right-to-left retroactive processing of the morphological structure was triggered in addition to the usual sequential left-to-right one. We assume that the right-to-left retroactive processing ensures a successful lexical access to the semantically opaque compound words in the minimal pairs. However, we cannot exclude that some acoustic information already carried in the first constituent may have informed the processing system that the sequence under processing is a compound word as our acoustic measures showed that the duration of the first constituent is shorter when it is followed by a neutral tone than by a full tone. Based on the present findings, we propose that neutral tone at the final syllable, probably combined with duration of the initial syllable, could be two *reliable* prosodic cues for assisting the analysis of the lexical morphology in Chinese. We concluded that, as already reported in other languages with various grammatical sequences, prosody might also have a function of prediction of ongoing grammatical information in Mandarin Chinese ambiguous minimal pairs. However, and most importantly, the present work highlights that in the Chinese minimal pairs we used, prosodic stress pattern delivered at the end of the sequences was employed for triggering a retroactive processing of the whole sequence. Taken together, the combined behavioral (production and perception studies) and neurophysiological processing studies presented in the current doctoral thesis allowed us to precise the functional and structural description of the Prosody-Assisted-Processing (PAP) model for Mandarin Chinese, especially by proposing an additional right-to-left retroactive processing route whose function is to morphologically reanalyze the Chinese sequences that carry a neutral tone on their final syllable.

Keywords: Compound words, Phrases, EEG, Prosody, Stress, Chinese, Neutral tone

Le rôle du patron accentuel dans la production et le traitement de mots composés et de syntagmes du Chinois Mandarin

par

Weilin Shen

RÉSUMÉ

La présente thèse étudie le traitement du patron accentuel (accent de mot vs. accent de syntagme) et son implication dans le traitement auditif des paires minimales du chinois mandarin constituées d'un mot composé ou d'un syntagme. Deux types de paires minimales ont été utilisés : 1) Mots composés avec un ton neutre sur la syllabe finale (fort-faible) vs. Syntagme avec un ton plein sur la syllabe finale (fort-fort) ; 2) Mots composés Verbe-Nom (VN, fort-faible) vs. Syntagmes Verbe-Objet (VO, faible-fort) se distinguant par le patron accentuel. Dans une étude de production, les analyses phonétiques des paires minimales ont révélé un patron accentuel contrastif indexé par une durée plus longue, une variation de F0 plus grande et une intensité plus élevée dans les syllabes accentuées. Au niveau neurophysiologique, le traitement du patron accentuel contrastif des deux types de paires minimales a été associé à deux potentiels cérébraux liés aux événements exogènes : une N200 supposée refléter un traitement langage-spécifique et automatique du patron accentuel, et une P200 supposée signer le traitement des signaux physiques/acoustiques. En plus de ces deux composantes ERP exogènes, nous avons montré qu'une composante endogène, à savoir la N325, varie en fonction du patron accentuel dans les paires minimales. En effet, la N325 était plus grande pour le patron accentuel rare, mais cet effet ERP n'a été observé que dans la tâche de discrimination de l'accent pour des stimuli non-délexicalisés. Nous argumentons donc que la N325 en réponse au patron accentuel du chinois mandarin peut être un marqueur ERP langage-spécifique et contrôlé du patron accentuel, éventuellement en quelque sorte lié à l'allocation des ressources attentionnelles. En outre, nous avons trouvé une négativité de discordance (MMN) tardive et soutenue en réponse au traitement de la congruence du patron accentuel indépendamment du patron accentuel et de la tâche. Cette MMN tardive suggère que les changements du patron accentuel sont traités sur la représentation à long terme de l'accent, pas au niveau de la syllabe. Enfin, nos données neurophysiologiques ont montré que le ton neutre dans les paires minimales a été utilisé par le système de traitement pour aider l'analyse de la structure morpho-lexicale des séquences parlées ambigües du chinois mandarin. Nous avons démontré que, une fois un ton neutre a été détecté, un traitement rétroactif de droite à gauche de la structure morphologique a été déclenché en plus de celui de gauche à droite séquentielle habituelle. Nous supposons que le traitement rétroactif de droite à gauche assure un accès lexical réussi pour les mots composés sémantiquement opaques dans les paires minimales. Cependant, nous ne pouvons pas exclure que certaines informations acoustiques déjà réalisées dans le premier constituant peuvent avoir informé le système de traitement que la séquence en cours de traitement est un mot composé comme nos mesures acoustiques ont montré que la durée du premier constituant est plus courte quand il est suivie d'un ton neutre que d'un ton plein. Basé sur les résultats actuels, nous proposons que le ton neutre sur la syllabe finale, probablement combiné avec la durée de la syllabe initiale, pourrait être deux indices prosodiques *fiables* pour aider l'analyse de la morphologie lexicale pour le chinois. Nous avons conclu que, comme déjà signalé dans d'autres langues avec diverses séquences grammaticales, la prosodie pourrait aussi avoir une fonction de prédiction de l'information grammaticale en cours dans les paires minimales ambigües du chinois mandarin. Cependant, et surtout, la présente étude souligne que dans les paires minimales chinoises que nous avons utilisées, le patron accentuel prononcée à la fin des séquences a été utilisée pour déclencher un traitement rétroactif de l'ensemble de la séquence. Pris ensemble, les études comportementales (études de production et de perception) et neurophysiologiques sur le traitement présentées dans cette thèse nous ont permit de préciser la description fonctionnelle et structurelle du modèle Prosody-Assisted-Processing (PAP) pour le chinois mandarin, en particulier en proposant une route de traitement supplémentaire de droite à gauche dont la fonction est de réanalyser morphologiquement les séquences chinoises qui portent un ton neutre sur la syllabe finale.

Mots-clés : Mots composés, Syntagmes, EEG, Prosodie, Accent, Chinois, Ton neutre

CONTENTS

1. Introduction

Spoken-word recognition relies upon different sources of information such as segmental (vowel, consonant) and suprasegmental phonology (e.g., stress, tone, duration) as well as morphological, syntactic, semantic and pragmatic information. Regarding the fact that the recognition can be achieved within several hundred milliseconds, the different grammatical information must be accessed and coordinated within milliseconds. The issue of how these sources of information are coordinated in real time (i.e. in a serial, parallel, or interactive fashion) is still a matter of considerable debate in the modeling of the language comprehension. The present work aimed to examine the temporal dynamics of the interaction between suprasegmental phonology, more precisely, stress pattern and morpholexical information in order to precise the role of prosody in language processing. We especially focused on morphologically complex structures, i.e., compound words and phrases. For this purpose, the electrophysiological technique with event-related brain potentials (ERPs) recordings was employed. The study of morphologically complex structures addresses one of the most important questions in linguistic morphology: how complex words are to be recognized. These questions are related to the issue of the organization of the mental lexicon in psycholinguistics (Collins & Quillian, 1969; Smith, Shoben & Rips, 1974; Collins & Loftus, 1975; Bock & Levelt, 1994). Two main approaches discuss the concept of mental lexicon. The first one considers that the mental lexicon contains only the information that is completely idiosyncratic, that means any property of a word that can be predicted by phonological, morphological, or syntactic rules will be excluded from the mental lexicon and the mental lexicon is thus simply a list of morphemes (Bloomfield, 1933). Another approach assumed a set of word formation rules for constructing words from morphemes (Pinker & Prince, 1988). In such an approach, all the mental lexicon needs to contain is a list of morphemes and necessary grammatical rules. A longstanding debate has been engendered on whether the representation of compound words in the mental lexicon is word-based (Marslen-Wilson & Welsh, 1978; Bybee, 1995) or organized according to morphological aspects, with monosyllabic morphemes rather than whole words being the primary unit of representation

(Taft & Forster, 1975, 1976; Taft, 2004).

The two different viewpoints on the representation of compound words in the mental lexicon led to three models on the processing of compound words. A *direct route model* postulates that compound words are acceded as its whole like monomorphemic words, since they consider that compound words are stored as full forms in the mental lexicon. According to *full-parsing model,* constituent morphemes of compound words are lexical storage in the mental lexicon and the processing of compound words involved a prelexical decomposition of compounds into their constituent morphemes. An intermediate model, so-called "*dual route*" has been proposed. The *dual route models* postulate two possible routes for processing compound words: (1) a decompositional route (or parsing route) in which the compound is decomposed into its morphological components and (2) a direct route in which the compound is accessed by a lexical entry based on stored full-form representation in the mental lexicon. It has been proposed that factors such as semantic transparency, lexical frequency and prosodic structure could influence which route would be chosen. For example, inspired by the works of Grosjean (1980) and Grosjean & Gee (1987) on the prosodic pattern of monomorphemic words, Isel, Gunter and Friederici (2003) showed for compound words that the length of the left syllable of German noun-noun compound words plays a critical role in the compound words processing. According to Isel et al. (2003), the length of the left syllable is analyzed in order to anticipate the morphological structure of German noun-noun compound words. In addition to syllable length, Koester, Gunter, Wagner and Friederici (2004) claimed that syllable F0 contour was involved in the spoken compound words processing. This was recently confirmed in production studies investigating acoustic-phonetic characteristics of VN compound words in Mandarin Chinese (Shen, Isel, & Vaissière, 2013) and in French (Rosenfeld, Fradin, & Isel, 2014). Because of the importance of the prosodic information in compound words processing, a Prosody-Assisted Processing (PAP; Isel et al., 2003) model was therefore proposed postulating an early role of prosodic information, i.e. length and F0 contour of initial constituents in compound words recognition.

Compound words and phrase are two structures sometime not distinct. For example, we have *blackbird* as a compound word and *black bird* as a phrase in English. The two construction types imply a structural ambiguity in the language such that identical concatenations can be either compound or phrase. It has been proposed that stress information involved in disambiguating the two structures. Chomsky & Halle (1968) proposed for English the Compound Stress Rule that stresses the left constituent of a compound and the Nuclear-Stress-Rule that places main stress on the rightmost constituent of a syntactic phrase. In the above example, the compound *BLACKbird*[1] has a stress on the left constituent *black*, however, the phrase *black BIRD* has its stress on the right constituent. The contrastive stress pattern in compounds and phrases were confirmed by acoustic analyses. For example, Farnetani, Torsello and Cosi (1988) showed that minimal pairs of compounds and phrases tend to differ in length, with the phrase being slightly longer (there are composed of two words, instead of one, with each characterized by word final lengthening, and there is an extra between-word boundary, as compared to the compound word.

This stress contrast creates interesting minimal pairs of segmentally identical but prosodically distinct compounds and phrases. There are morphological and semantic distinctions between compounds/phrases minimal pairs, such as *greenhouse* is a compound that refers to glasshouse for plants and *green house* is a phrase means a house that in green color. In addition, when we examine further the semantic of the constituents, sometimes we find that the first constituent of the compound becomes semantically opaque. In the above example, *green* in *greenhouse* represents no more a color. The phenomenon can be observed in more examples: *hotdog*/*hot dog*, *blackboard*/*black board*, etc. Psycholinguistic researches (Farnetani et al., 1988; Vogel & Raimy, 2002; McCauley, Hestvik, & Vogel, 2013) have showed that the processing of compound/phrasal stress is critical to parsing the ambiguous compounds/phrases minimal pairs. The compound/phrasal stress acquisition has thus an importance for foreign language learners to access the appropriate representation of the ambiguous compound/phrasal pairs. Another case of prosody-morphology ambiguity that would also be relevant for falsifying a model of word recognition that postulates a determining role of prosodic information is the case of phrases, which have a focus accent on the argument of the head like in the following example: “Is it a WHITE bird? No, it is a BLACK bird.”

[1] In the present work, we use the upper case to signifier stress.

Indeed, it would be first relevant to determine to which extent word stress and focus accent are acoustically different, and then to understand how are processed phrases with a focus accent.

Previous studies on the role of stress in compound words processing have been mostly carried out in stress languages such as English, Dutch, and German. The present study used Mandarin Chinese as target language to investigate the role of stress in Chinese compound words processing. Chinese has been described as a language of compound words. Compounding is a major and productive means of Chinese word formation (Ceccagno & Basciano, 2007). A corpus study by China Institute of Language Teaching and Research (1986) revealed that among a corpus of 1.3 million of words, 73.6% entries were disyllabic compound words. Moreover, in the corpus of neologisms proposed in The Contemporary Chinese Dictionary (2002) more than 90% of all new words are compounds (Ceccagno & Basciano, 2007). In addition to their large number, Chinese compound words are rich in their morphological types. Chinese compound words have three main types in compounding, i.e. subordinate, attributive and coordinate. Each type may be endocentric or exocentric. Many Chinese compound words are formed from syntactic phrases. Chinese verb-noun sequence often creates ambiguities in parsing, as it is the case in other languages. Indeed, Chinese Verb-Noun sequence can appear in at least two different relations, i.e. the verb-object phrasal relation and the modifier-head compound word relation. Thus, the Verb-Noun form is an especially interesting case in Chinese of apparent indeterminacy between morphology and syntax, because this form may be considered a Verb Object phrase or a compound. For instance, the verb-noun sequence *jian3zhi3*[2] 'cut - paper' may be understood as a verb-object phrase, i.e., 'cut paper' or a modifier-head compound word, i.e., 'paper-cut'.

With regard to the stress in Chinese, some linguists (Hyman, 1977; Selkirk & Shen, 1990) have considered Chinese as a language having no word stress because the suprasegmental information, i.e., the tone is used for distinguishing lexical meaning. However, linguists

[2] In the present work, Chinese is presented in *pinyin* that is the official phonetic system for transcribing the Mandarin pronunciations of Chinese characters into the Latin alphabet. The number refers to the tone: 1 = first tone: high-level; 2 = second tone: high-rising; 3 = third tone: falling rising (or dipping) and 4 = fourth tone: high-falling.

agreed that there are some “stress” phenomena in Chinese syllables (Lin, 2001; Duanmu 2000, 2007). The most acceptable stress phenomenon in Chinese is the ‘neutral tone’ (Chao, 1968). The neutral tone also called fifth tone or zeroth tone (Chinese: 轻声, literal meaning: ‘light tone’) in addition to the four full tones. It is claimed that in Standard Mandarin, about 15-20% of the syllables in written texts carried a neutral tone (Li, 1981). A corpus study (Chen, 2004) on the Modern Chinese Dictionary (1996 edition) with sixty thousand entries revealed that 2882 entries contain the neutral tone. The neutral tone is a neutralization of a corresponding full tone, and it is associated systematically with weak syllables. Neutral tones are reminiscent of reduced syllables in English. Syllables that have a neutral tone are pronounced in a short and light manner, thus sounds quite weak. Phonetic studies (e.g., Cao, 1992; Liu & Xu, 2005) associated the neutral tone with acoustic cues including shorter duration, smaller F0 variation and less intensity. The use of full tone and neutral tone leads to two degrees of “stress” (i.e. of strength) in Chinese, i.e. “stressed” (strong) that is loaded by a syllable with a full tone and “unstressed” (weak) loaded by a syllable with a neutral tone.

In Chinese, there are some disyllabic sequences differed from each other only in the stress pattern resulting two interpretations: 1) phrasal interpretation, with strong-strong stress pattern, in which the right syllable carried a full tone and the first morpheme was semantically transparent; or 2) compound word interpretation, with strong-weak stress pattern whose right syllable carried a neutral tone moreover the first constituent was semantically opaque. For example of the disyllabic sequence *dong - xi* ‘east - west’ consists of two morphemes, *dong* (means east) and *xi* (means west). When the second constituent carries a neutral tone (i.e., strong-weak), the sequence is a compound word, moreover the first constituent becomes semantically opaque (the meaning of the whole compound is ‘thing’, the first constituent does not contribute to the whole word meaning); when the second constituent carries a full tone (i.e., strong-strong), the sequence is a phrase (it means the east and the west), the first constituent is semantically transparent. These sequences are excellent materials for studying the interaction between stress information and morpholexical processing, because the semantic properties of the left morpheme of the sequence are assumed to be modulated by the stress information of the last syllable. Sequential continuous access theories claim that the ongoing acoustic-phonetic information is, left to right, continuously mapped onto representations stored in the mental lexicon regardless of the morphological structure of the word (Butterworth, 1983; Marslen-Wilson, 1987). Previous studies focused on the role of prosodic information of the initial constituent during lexical access of spoken compound words. In the Prosody-Assisted Processing (PAP; Isel et al., 2003) model, the length of the left morpheme of compound words is a determining factor for setting off/triggering a decompositional route. For these Chinese sequences, the prosodic information (stress) is available at the last syllable of the sequences and it modulates the semantic propriety of the first morpheme. These sequences were investigated in order to test the validity of the hypothesis of a right-to-left retroactive processing of prosodic information with a critical importance of the stress information of the last morpheme.

Chapter 2 will be devoted to morphology in word recognition. We will particularly insist on the compounding. We will firstly present the classification of compound words proposed in Linguistics, followed by the presentation of the morphology/syntax distinction for compound words and the morphological characteristics of Chinese compounds. A summary of the principal psycholinguistic literature on compound words recognition will be presented. In Chapter 3, we will introduce the description of stress in Linguistics with discussion of the phonetic correlates of stress production and stress perception. We will then introduce general prosodic characteristics and stress phenomenon in Mandarin Chinese. Moreover, particular focus will be given to the specification of compound vs. phrase stress. Finally, a series of psycholinguistic experiments that substantiate the role of the stress in the recognition of spoken words will be reviewed. In Chapter 4, we will present behavioural and neuro-cognitive models of compound words processing formulated by psycholinguists. Chapter 5 will introduce the methods applied in the current investigations. The technique of event-related brain potential (ERP) as well as some ERP components in language processing will be presented. Chapter 6 summarized our own work. Three ERP experiments that substantiate the role of stress in the processing of Chinese compound words as well as one production experiment will be presented and discussed. Chapter 7 will be devoted to the general discussion and further investigations. Chapter 8 is an additional chapter in which we introduce a method of the learning of Mandarin Chinese pronunciation. We developed a mobile software application for the learning of Mandarin Chinese pronunciation using pitch detection and automatic speech recognition techniques.

Chapter 2

Lexical morphology

2. Lexical morphology

2.1. Compounding

In linguistics, compounding is described as the morphological process of word formation that creates compound words containing two or more morphemes, such as the English word *teacup* composed of the nouns *tea* and *cup*. Most languages exhibit some form of compounding. In some languages, such as Chinese, Vietnamese, compounding is the only real evidence of morphological complexity.

The study of compound words addresses one of the most important questions in linguistic morphology: What is a word and how it is to be recognized? These questions are related to the issue of the organization of the mental lexicon[3] in psycholinguistics (Collins & Quillian, 1969; Smith, Shoben & Rips, 1974; Collins & Loftus, 1975; Bock & Levelt, 1994). In linguistics, two main approaches discuss the concept of mental lexicon. The first one considers that the mental lexicon contains only the information that is completely idiosyncratic, that means any property of a word, which can be predicted by phonological, morphological, or syntactic rules will be excluded from the lexicon and the lexicon is thus simply a list of morphemes (Bloomfield, 1933). Another approach assumed a set of word formation rules for constructing words from morphemes (Pinker & Prince, 1988). In such an approach, all the mental lexicon needs to contain is a list of morphemes and necessary grammatical rules. Consequently, the representation of compound words in mental lexicon led to a longstanding debate on whether the representation is word-based (Marslen-Wilson & Welsh, 1978; Bybee, 1995) or organized according to morphological aspects, with monosyllabic morphemes rather than whole words being the primary unit of representation (Taft & Forster, 1975, 1976; Taft, 2004). An hybrid approach postulating a dual-route processing system constituted of a direct route and a decompositional one for accessing the mental lexicon will discussed in details based on the data we collected in the present work using Mandarin Chinese compound words versus

[3] The mental lexicon is the component of a linguistic system which can be regarded as a list or network of words or lexical entries (also lexical items, lexemes) containing information regarding a word's meaning, pronunciation, and syntactic characteristics.

phrases.

In the present chapter, the first part will be devoted to the classification of compound words proposed in Linguistics, followed by the presentation of the morphology/syntax distinction for compound words and the morphological characteristics of Chinese compound words; in the second and last part, the psycholinguistic literature on compound words recognition will be reviewed, by focusing on the representation and processing of compound words.

2.1.1. Classification of compound words

2.1.1.1. Endocentric and exocentric compound words

A major division of the class of compounds is that between endocentric compounds and exocentric compounds. In endocentric compounds, one constituent functions as the semantic and syntactic head. The head represents the core meaning of the compound, and it is of the same word class as the compound as a whole. For instance, in *blackboard*, *board* is the head (a *blackboard* is a kind of *board*; *board* and *blackboard* are both nouns). Mathews (1974) used the following formula to explain endocentric compounds.

$[X]_{\alpha} + [Y]_{ß} \rightarrow [X+Y]_{ß}$

According to the formula, the word class ß of the compound [X+Y] is the same as the one of its head [Y]. In English and other Germanic languages, endocentric compounds tend to have heads on the right (i.e. right-headed compounds). For example, a *goldfish* is a particular type of *fish*, and not a particular kind of *gold*. French and Vietnamese compounds tend to have heads on the left (i.e. left-headed compounds). For instance, a French compound *chou-fleur* ‘cauliflower’ is a particular type of *chou* ‘cabbage’ (i.e. head).

In languages with grammatical gender, such as French or German, the gender of the compound is also determined by the gender of the head. For example, in French, the compound *timbre-poste* [stamp-post] ‘postage stamp’ is masculine where the head *timbre* is the masculine and the non-head *poste* is feminine, in German, for <u>*das*</u>$^{\text{neuter}}$ *Wein*$^{\text{masculine}}$<u>*glas*</u>$^{\text{neuter}}$, the gender of the head *glas* determine the syntactic gender of the compound.

In the case of exocentric compound words, there is no constituent that functions as the head. For example, in exocentric compounds that refer to people, we can isolate a predicate-type element and an argument-type element; however, neither element can be called the head of the construction. An example is the French verb-noun compound *garde-malade* [keep - patient] ‘nurse’, where the head is the agent who keeps patients, and neither keep nor patient is the head.

2.1.1.2. Subordinate, Attributive and Coordinate compound words

According to the different grammatical relation between the constituents of the compound words, Bisetto and Scalise (2005) proposed three types of compound words: subordinate, attributive and coordinate compound words. Each type may be endocentric or exocentric.

Subordinate compound words are those in which constituents have an argument-head (or head-argument) relation. This is clear in compounds with a deverbal head constituent, such as *taxi-driver* (argument-head), i.e. *somebody who drives a taxi*. A similar relation can be found in compounds that do not have a deverbal head, for example [N+N] compounds where the constituents are typically linked by an “of-relation”, as in *doorknob* (“knob of a door”).

Attributive compounds express a modifier relation between the head and the non-head, in which one element stands in a relation of attribute or modifier to another. The prototypical case involves attributive compounds in which the first constituent is an adjective, as in “*high school*”. However, we can find other structural types as well, for instance by two nouns, in which the non-head is very often used as a metaphoric attribute of the head, as in *swordfish* (“fish with a sword-like snout”) and not as a mere complement of it (*fish of a sword, fish for a sword).

There is a third kind of compounds, where both continents as equally sharing head-like characteristics. Coordinate compounds may have two heads. We can recognize them by possibility of adding "and" between the two heads. They can be a combination of parallel concepts, such as *producer-director*, which is the sum of *producer* and *director*. Coordinate compounds may also be a combination of antonyms where two heads describe contrary attributes of the compound. We can find many compounds of this type in Chinese, such as *mai3mai4* [buy - sell] 'business' is composed of 'buy' and 'sell' and *kai1guan1* [turn on - turn off] 'switch' is composed of 'turn on' and 'turn off'. Figure 2.1 summarizes the classification scheme for compounds based on Bisetto and Scalise (2005).

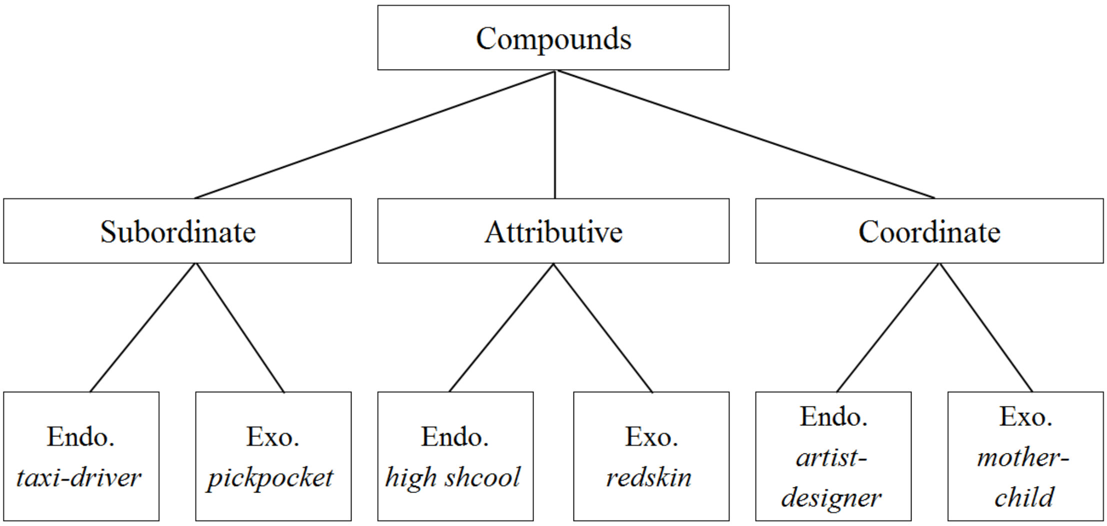


Figure 2.1. Classification of Compounds Bisetto and Scalise (2005)

2.1.2. Semantic transparency

Compounds vary in their semantic transparency, that is, the existence of the relationship between the meaning of the whole compound and the meaning of its constituents (Zwitserlood, 1994; Dirven & Verspoor, 1998). Compound words exhibit a continuum of transparency from fully transparent compound words to fully opaque compounds.

1. Both constituents of the compound and their semantic link are clearly analyzable and hence semantically transparent (i.e. TT), such as *snowball*.
2. The meaning of a compound is not clearly related to the meaning of its constituents and there is no semantic link between the constituents (i.e. OO), as in *honeymoon*, the compound is semantically opaque.

3. The meaning of a compound is only clearly related to one of its two constituents (i.e. OT or TO), such as *blackbird (OT)*, which does not denote a *black* type of *bird* but a *bird* species, the compound is partially transparent.

In psychological research, it has been shown that the degree of semantic transparency of the morphological complex words has an influence on the processing and representing of the words in mental lexicon (e.g., Sandra, 1990; Zwitserlood, 1994; Marslen-Wilson, Tyler, Waksler, & Older, 1994; Libben, 1998; Isel et al., 2003). These psychological researches will be discussed in detail in section 4.2.1.

2.1.3. Size of morphological family

The morphological family size of a word is the type count of all the compound words in which this word appears as a constituent (Baayen, Lieber, & Schreuder, 1997; Schreuder & Baayen, 1997). We can find in many languages that some words are more productive in compounding than others, and so they have a large morphological family size. It is the case in French for the word *porte* 'carry' that appears in several compound words such as *porte-parole* 'spokesman', *porte-monnaie* 'wallet', *porte-bébé* 'baby carrier', *porte-avion* 'aircraft carrier', *porte-malheur* 'jinx' and etc. In Chinese, the morpheme *da4* 'big' is for example very productive forming a large family of compound words including, *da4jia1* 'big - family: everybody', *da4cu2* 'big - cook: masterchef', *da4hai3* 'big - sea: ocean' and so on. Words with large morphological families can appear in initial or final position in compound words. We have a large *-man* in the last position of English compounds, like *policeman*, *superman*, *fireman*. The morphological family size seems to be an important factor in the processing of compound words (e.g., Berkowitz, 2009; Nicoladis & Krott 2007). Berkowitz (2009) found that participants responded faster to compounds from large as opposed to small morphological families in the naming and lexical decision tasks. Nicoladis and Krott (2007) asked French young children (from 3 years to 5 years) to explain the meaning of a compound word with constituents of varying family size. They showed that in order to describe the compound word, children used more frequently the constituent when it came from a large family than a small family. Taken together, these studies suggest that compound words with a large morphological family are more activated than compound words with a smaller one.

2.2. **Compound words in Mandarin Chinese**

Chinese has been described as a language of compound words. Unlike Indo-European languages, in Chinese derivational[4] (e.g., *-s*, *-ed*) and inflectional[5] (e.g., *-ly*, *-tion*) morphology plays a minor role in word formation (Li & Thompson, 1981). Compounding is the only major and productive means of Chinese word formation (Ceccagno & Basciano, 2007).

Chinese lexicon perfectly corresponds between the syllable and the written character (汉字 *hanzi*) which most times represent a morpheme. Corpus study (Yin, 1984; Yuan & Huang, 1998) showed that 93% of Chinese characters are a morpheme. A statistical analysis of the lexicon of common words in contemporary Chinese (Li, 2008) on a total of 56,008 words indicated that 6% are one-character words, 72% are two-character words and 22% are more than two characters words. The two-character words (i.e. disyllabic words) are therefore dominant in Chinese lexicon. Given the fact that, the syllable (i.e. character) in Chinese largely coincides with the morpheme, and therefore, almost all disyllables are made up of two lexical morphemes, it is not surprising that Chinese has been defined as a "language of compound words". A corpus study by China Institute of Language Teaching and Research (1986) recorded 73.6% disyllabic compound words among a corpus of 1.3 million of words.

The reason for richly developed compound form (i.e. disyllabic words) in modern Chinese is mainly explained in a history view of the evolution of Chinese lexicon. In old Chinese before 200 B.C., the dominant word form was monosyllabic, and disyllabic words represented only 20% of the lexicon. According to this functional explanation (Cheng, 1992; Packard,

[4] Morphological derivation is the process of forming new words on the basis of existing words. It often involves the addition of a morpheme in the form of an affix.
[5] Inflection is the modification of a word to express different grammatical categories such as tense, person, gender, number and case. The inflection of verbs is also called conjugation, and the inflection of nouns, adjectives and pronouns is also called declension.

2000), in the transition from a primitive society to a feudal one (between 1000 and 300 B.C.), there was a pragmatic need to create new words for new referents. Moreover, a simplification of the phonology of the language occurred in the period, in modern Chinese there are only 400 syllables with four distinct tones (Taylor & Taylor, 1995). The simplification of the phonology caused the loss of many distinctions, and many syllables that were once separated turned to homophonous. To avoid ambiguity, monosyllabic words increased syllable number and became disyllabic. Nowadays, morphemes keep on being combined to create new compound words (Zhou, 1994). In the corpus of neologisms proposed in The Contemporary Chinese Dictionary (2002) more than 90% of all new words are compounds (Ceccagno & Basciano, 2007).

2.2.1. Classification of Chinese compound words

In addition to their large number, Chinese compound words are rich in their types. Many approaches have been proposed for classifying Chinese compound words on different aspects: the morphological relationship of constituents (Xia, 1946), the semantics (Li & Thompson, 1981), the syntax (Chao, 1948; 1968), and the class description (Pacard, 2000). The classification of Ceccagno and Scalise (2006) may be one of the most suitable. This classification adopted the classification scheme put forth by Bisetto and Scalise (2005), quoted in the precedent section. We can classify Chinese compound words into three types in compounding, i.e. subordinate, attributive and coordinate. Each type may be endocentric or exocentric.

2.2.1.1. Subordinate compound words

Subordinate compound words express a complement relation between the head and the non-head. Examples are presented below:

(1) 待岗 *dai4gang3* $[V+N]_V$ "wait for + post" = "wait for a job"

(2) 干事 *gan4shi4* $[V+N]_N$ "do + matter/responsibility" = "secretary in charge of

something”

待岗 *dai4gang3* (1) is a verbal compound with a deverbal head *dai4* ‘wait’. The noun constituent acts as the internal argument of the verb. 干事 *gan4shi4* (2) is a compound with head *gan4* ‘do’ where the noun constituent acts as its argument. The compound is exocentric.

2.2.1.2. Attributive compound words

Attributive compounds are those in which one element stands in a relation of attribute or modifier to another. The prototypical case involves attributive compounds in which the first constituent is an adjective. However, we can find other structural types as well, for instance by two nouns. Many Chinese compounds are in this type, for example:

(1) 公牛 *gong1niu2* $[A+N]_N$ “male + cattle” = “bull”

(2) 足球 *zu2qiu2* $[N+N]_N$ “foot + ball” = “football”

(3) 婚介 *hun1jie4* $[N+V]_N$ “wedding + introduce” = “matchmaking”

公牛 *gong1niu2* (1) is a noun compound where the adjective constituent modifies the head. 足球 *zu2qiu2* (2) is a noun compound where the right noun constituent expresses the property of the head. 婚介 *hun1 jie4* (3) is a exocentric noun compound.

2.2.1.3. Coordinate compound words

Coordinate compounds may have two heads of parallel things or antonyms. We can find many compounds of this type in Chinese, such as *mai3mai4* [buy - sell] ‘business’ is composed of ‘buy’ and ‘sell’ and *kai1guan1* [turn on - turn off] ‘switch’ is composed of ‘turn on’ and ‘turn off’. For example:

(1) 体制 *ti3zhi4* $[N+N]_N$ ‘system + regime’ = ‘system of organization’

(2) 开关 *kai1guan1* $[V+V]_N$ ‘turn on + turn off’ = ‘switch’

(3) 东西 *dong1xi0*[6] $[N+N]_N$ ‘east + west’ = ‘thing’

[6] The tone number 0 refers to the neutral tone. The introduction of the neutral tone will be presented in the next chapter in section 3.2.

体制 *ti3zhi4* (1) is compound has two heads of parallel things. 开关 *kai1guan1* (2) is compound of combination of antonyms things. 东西 *dong1xi0* (3) is exocentric compound of combination of antonyms things.

Figure 2.2 shows the classification scheme illustrated with Chinese examples.

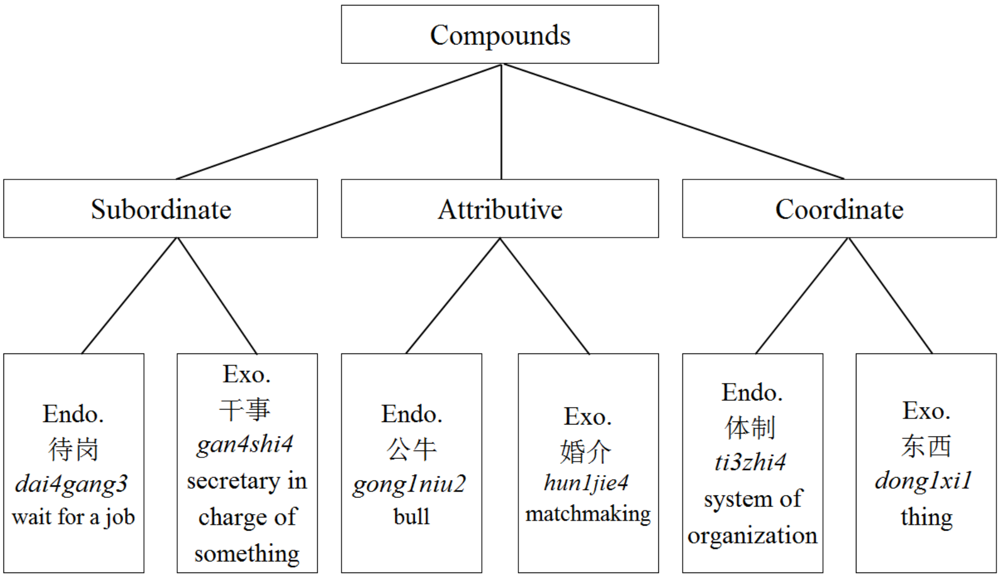


Figure 2.2. Classification of Compounds with Chinese examples

2.3. Compound and phrase distinction

In many languages, we are able to concatenate lexemes to form both complex lexemes such as compound words and phrases (Fabb, 1998). For example, in English, we have *blackbird* as a compound and *black bird* as a phrase. The two construction types are not distinct, therefore implies a structural ambiguity in the language such that identical concatenations can be either compound or phrase. We have often no satisfactory, univocal way of distinguishing between them.

Compounds have both sets of characteristic properties of syntax and morphology. The first set makes compounding resemble syntactic processes, the second set brings compounding closer to word formation. In French, the compounding system looks more like the lexicalization of syntax than a specific morphological compounding process. One main type of construction of French compound is formed from syntactic phrases (complete with function words and inflected lexical items) such as *le cessez-le-feu* 'ceasefire' ('cease-the-fire'), *les hors d'oeuvre* 'hors d'oeuvre'. A second type consists of a verb followed by its object: le *porte-feuille* 'wallet' ('carries- paper money), *le pince-nez* 'pince-nez' ('pinches-nose').

However, compounds also have a number of features that make them resemble words distinguished from a phrase:

1) Idiomaticity

Idiomaticity is the tendency of phrases to take on meanings that go beyond the compositional meanings of their parts. Compounds are often lexicalized and they may have idiomatic meanings, while the meaning of phrases tends to be compositional. For example, *blackboard* is not only a board in black. It could be also in other colors, like green and gray, etc. However, a phrase *black board* refers to the compositional meaning of the two constituents.

2) Generic/referential nouns

In [N+N] compounds, non-head nouns are always generic, i.e., refers to an entire class (non-referential). For example, neither *student* nor *film* in *student film society* serves to pick out any specific student or film. Nonetheless, in [N+N] phrases, non-head nouns can be generic or referential, but are usually referential.

3) Morphological cohesion

Compounds act as a *single unit* for the purposes of marking regular morphological processes. The non-head of compounds typically fails to derivational/inflectional morphological processes. Thus, **pick<u>ed</u>pocket* is not acceptable. We make the plural of a compound noun by adding -s to the head. Such as, toothbrush → toothbrushes not **t<u>ee</u>thbrushes*.

4) Syntactic cohesion

Compounds are syntactic cohesive in that they cannot generally be expanded by modifiers such as adjectives or adverbs (e.g. *crispbread* versus **very crispbread* ‘bread that is very crisp’).

5) Phonological cohesion

Compounds have stronger phonological cohesion and are distinguished from phrasal constructions, by a contrasting pattern of prosodic stress. A well-known example of this comes from English, where we have the Compound Stress Rule that stresses the left member of a compound and the Nuclear-Stress-Rule (Chomsky & Halle, 1968), which places main stress on the rightmost constituent of a syntactic phrase. For example, the compound *BLACKbird* has a stress on the left constituent *black,* however, the phrase *black BIRD* has its stress on the right constituent. We will discuss this concept in more detail in the following chapter.

2.3.1. Chinese Verb-Noun form

The verb-noun sequence in Chinese often creates ambiguities in parsing, as it is the case in other languages. Indeed, the verb-noun sequence can appear in at least two different relations, i.e. the verb-object relation and the modifier-head relation. Thus, the Verb-Noun form is an especially interesting case in Chinese of apparent indeterminacy between morphology and syntax, because this form may be considered a Verb Object phrase or a compound of different types (i.e. attributive compound or subordinate compound). Examples are presented below:

(1a) 剪纸 *jian3zhi3* $[V+N]_N$ ‘cut + paper’ = ‘paper-cut’

(1b) 剪纸 *jian3zhi3* V+O ‘cut + paper’ = ‘cut paper’

(2a) 干事 *gan4shi4* $[V+N]_N$ ‘do + matter/responsibility’ = ‘secretary in charge of something’

(2b) 干事 *gan4shi4* V+O ‘do + matter/responsibility’ = ‘do something’

剪纸 *jian3zhi3* (1a) is a lexicalized attributive compound in which the verb constituent modify the noun head. It has a strong syntactic cohesion. We cannot insert any modifiers such as adjectives. *剪红纸 *jian3hong2 zhi3* ‘*cut red paper’ is not acceptable. 剪纸 *jian3zhi3* (1b) is a Verb Object phrase whose syntactic cohesion is weak. The structure can be expanded, for example, 剪一张很大的纸 *jian3yi1zhang1hen3da4de0zhi3* ‘cut a very big paper’.

干事 *gan4shi4*(2a) represents a subordinate compound where the noun constituent acts as the argument. The compound is exocentric and lexicalized in the lexicon. 干事 *gan4shi4* (2b) is a Verb Object phrase. The extension is possible, for example, 干一件好事 *gan4yi1jian4hao3shi4* ‘do one good thing’.

The dual status of Verb-Noun form was explained from a historical view (Li & Thompson, 1981; Huang, 1984; Packard, 2000) that, since most Verb-Noun words come from syntactic phrases, they reflect a gradual process of phrases developing over time into completely fused words that are inseparable and completely idiomatic in meaning. Huang (1984) proposed that since such linguistic processes are gradual rather than abrupt, idiomaticity and separability are not discrete, ‘all-or-nothing properties’, but rather form a continuum, with idiomaticity and inseparability on one end, and compositionality and separability on the other one.

The dual status of Verb-Noun form is observed in many ambiguous compound/phrase forming what is called “minimal pairs” (For example, in French, “Le *garde-chasse* le sert chaque jour.” vs. “Le *garde chasse* le cerf chaque jour.”). There are neither morphological nor semantic or syntactic cues for making a distinction between them. The only available cues for distinguishing the two forms are given by the prosodic pattern. Therefore, prosodic information of the ambiguous compound/phrase minimal pairs may be used in their distinction. We have introduced above the Compound Stress Rule, which stresses the leftmost member of a compound, and the Nuclear-Stress-Rule that place main stress on the rightmost constituent of a syntactic phrase (Chomsky & Halle, 1968). It has been proposed that the Compound stress rule and the Nuclear-Stress-Rule are true for Mandarin Chinese, and they permit one to distinguish between compounds and phrases (Duanmu, 2002). An example in Duanmu (2002):

(1) 炒饭*CHAO3fan4* $[V+N]_N$ ‘fry rice’ = ‘fried rice’

(2) 炒饭*chao3FAN4* V+O ‘fry rice’ = ‘fry the rice’

炒饭*CHAO3fan4* (1) is a compound in which the verb modifies the noun head. According to Duanmu, this structure has the same stress pattern (i.e. stress on the left constituent) as like *BLACKbird* in English. With respect to the verb-object structure 炒饭*chao3FAN4* (2), the stress is placed on the right constituent such as *black BIRD* in English.

The question of the distinction of the Verb-Noun form is of central interest in the present work and is directly related to the role of prosodic structure in the distinction of compound words and phrases in Mandarin Chinese. It also addresses the issue of syntactic parsing directly related to the question of the prosodic boundaries. We will discuss in Chapter 3 the prosodic characteristics of such ambiguous forms.

2.4. Psycholinguistic research on morphology in word recognition

We will now review the psycholinguistic literature on the representation and processing of compound words. Behavioural and neurophysiological experimental evidence are presented to response to the main questions in morphological processing 1) how morphemes are represented in the mental lexicon 2) how morphological information is computed in the lexical processing and 3) what role does morphology structure play in morphological processing.

2.4.1. Morphological decomposition process

Morphological decomposition is the process of decomposing an incoming word into its constituent morphemes. Since the original work of Taft and Forster (1975; 1976), psycholinguists have collected abundant experimental evidence for or against morphological decomposition (see for a review, Isel, 2010).

2.4.1.1. Behavioural evidence

Using the lexical decision task, Taft and Forster (1975) firstly investigated the processing of affixed words. Taft and Forster (1976) extended this research to compound words, showing effects of morphological constituency in compound words. Participants were asked to do a lexical decision task on compound words. They found that the response latencies were modulated by the lexical status of the left constituent: the nonword **footmilge* (made up of a word and a nonword, respectively) took longer to classify as a nonword than did **trowbreak* (made up of a nonword and a word, respectively). The results can only be explained by a morphological decomposition processes during compound words recognition, that is to say, the **footmilge* is decomposed into *foot* and *milge*, the word status of *foot* slows the nonword judgment of the **footmilge*, while in **trowbreak* the first constituent *trow* is nonword thus no interference will take place. This finding lend support to a model of word recognition postulating that the left constituent is a sort of access code (i.e. a key) for accessing the mental lexicon.

Lima and Pollatsek (1983), in contrast, showed that nonword compounds in which both constituents formed real words (e.g., **turntribe*) took longer to reject in a visual lexical decision task than did those in which only the left constituent formed a word (the right constituent is nonword e.g., *hillsosk*). This result suggests that the lexical status of the right constituent also influences the processing of nonword compounds. However, the automatic decomposition processes did not go unchallenged and has been shown to be influenced by several factors, including semantic transparency, frequency and prosodic structure (e.g., Hyönä & Pollatsek, 1998; Libben, Gibson, Yoon & Sandra, 2003; Isel et al., 2003; Isel, 2010; Rosenfled, Fradin & Isel, 2014).

Libben (1994) and Libben, Derwing and de Almeida (1999) examined the decomposition in more detail using English ambiguous novel compounds. These words (e.g., *clamprod*) have two interpretable parses (e.g., *clam*^{palourde} + *prod*^{poussée} or *clamp*^{agrafe} + *rod*^{tringle}) thus offer an opportunity to investigate how the compound is to be decomposed. Using the semantic priming paradigm with a recall task, Libben et al. (1999) found that ambiguous compound words are able to prime semantic associates of the different constituents of the compound words, that is to say, in the above example, the compound *clamprod* can prime both the target word *sea*^{mer} that is semantically related to the constituent *clam*^{palourde} and the target word *hold*^{tenir} which is the semantic associate of the other possible constituent *clamp*^{agrafe}. The authors concluded that morphological parsing does not simply decompose a compound into its constituents, but rather generates all possible parses that are subsequently evaluated.

The role of semantic transparency on morphological decomposition has also been investigated. Sandra (1990), using semantic priming, showed no priming effects for semantically opaque compounds (e.g., *buttercup*) and pseudo-compounds (e.g., boycott). Constituents are accessed only for transparent compounds (e.g., *teaspoon*). A similar pattern of results was observed for German compound words (Isel et al., 2003). The authors investigated the auditory processing of 160 two-noun right-headed German compound words equivalently distributed in four different categories according to the semantic transparency of their constituents (40 Transparent-Transparent (*Bier*[*beer*]*glas*[*glass*])[*beer glass*], 40 Opaque-Opaque (*Schnee*[*snow*]*besen*[*broom*])[*whisk*], 40 Transparent-Opaque (*Knast*[*jail*]*bruder*[*brother*])[*jailbird*], and 40 Opaque-Transparent[head] (*Floh*[*flea*]*markt*[*market*])[*flea market*]) by monitoring semantic priming effects of the left constituent using a lexical decision task. Results revealed that German compound words with a transparent final head were decomposed, whereas compounds with an opaque head were not. Based on this finding, the authors suggested that the processing system might use two separate routes to access German compound words: A direct route[7] for compound words with an opaque head or a decompositional route for compound words with a transparent head.

Isel et al. (2003) also investigated the implication of the prosodic structure on the morphological processing by manipulating the acoustic structure of the left constituent of compounds. These authors highlighted the crucial role of length of the initial constituents of German compound words on the decomposition processes. Duration[8] of the left constituent is a determining phonetic cue, which is online analyzed for determining whether this segment of speech signal does or not constitutes the beginning of a longer word (a polymorphemic vs. a monomorphemic word). As most polysyllabic words in German are compound words, then Isel and colleagues proposed that duration would be a *valid prosodic cue* for predicting the morphological structure of word, at least in spoken German. The proposal that prosody might

---

[7] In the direct route, compound words are accessed via whole-form just like a monomorphemic word. In a decompositional route, compound words are decomposed into its morphological components. See Chapter 4 for more detail on the psycholinguistic models of compound words.

[8] The crucial role of duration in the process of lexical access for monomorphemic words has been reported in studies (Cutler, Dahan & van Donselaar 1997; Grosjean & Gee, 1987).

have a function of predictor of word's structure extended the original idea advanced for sentence processing by Grosjean (1983) that prosodic information is a good "predictor" of upcoming information during the auditory processing of sentences. Isel (2010) proposed that as soon as the processing system is able to decide based on the prosodic analysis of the onset of a speech segment under processing that it could be a morphologically complex word, then a decomposition mechanism would be triggered. Thus, duration of initial constituent of German compound words, in addition to its function of structure's predictor, might be able to *assist* the processing system in activating a decompositional route at the offset of these morphemes. Isel et al. (2003) and Isel (2010) formulated a prosody-assisted processing device for lexical access of compound words in German.

More recently, in three production studies Shen et al. (2013) in Mandarin Chinese, and Rosenfled et al., (2014) and Isel, Vaissière & Audibert (in preparation) in French showed that the duration of the rime of the first constituent as well as the size of the F0 difference between the rime of the first constituent and the first syllable of the following constituent are determining acoustic-phonetic cues of prosodic boundary. Critically, these authors were able to show by comparing acoustic-phonetic properties of compound words (l*e garde-chasse*) and phrases (*le garde chasse*) constituting minimal pairs that (1) the duration of the rime of the first constituent (e.g. *arde* in *garde* in the above-mentioned example) is shorter in compound words than in phrase[9], and (2) the F0 difference between the rime of the first constituent and the first syllable of the following constituent is smaller in compound words than in phrase. Thus, phrasal boundary, and thus complementarily, the absence of boundary in case of higher cohesion between the two constituents as in compound words, might be marked by specific variation of duration and F0. Furthermore, Rosenfeld et al. (2014) showed that high frequency French VN and low frequency French VN differ at the acoustic-phonetic level, lending support to a stronger phonological cohesion for high frequency compound words than for low frequency ones, the latter being more likely to be decomposed.

Another important factor that influences morphological decomposition of compound

[9] Similarly, using another material, Isel et al. (2003) showed that the word *Wein* (vine) is shorter when it is produced in a compound word (Weinberg/ than when it is produced in isolation.

words is the frequency. The frequency effects have been shown in eye-tracking data. The eye-tracking method was largely used in the morphological processing studies due to its high temporal sensitivity. Using eye tracking, researchers are thus able to investigate the role of morphological structure in the time course of lexical processing. By monitoring eye movements during the reading of Finnish two-noun compound words, Hyönä and Pollatsek (1998) showed frequency effects in the reading of Finnish compounds: the gaze duration (i.e., the total fixation time on an item before the reader moves on to the next item) on the compound word was 100 *ms* shorter when there was a frequent initial morpheme. This result suggests that left morpheme might play a role in the processing of the whole word access. The observation of a constituent frequency effect supports a morphological decomposition account. Andrews, Miller and Rayner (2004) replicated these results with English compound words. Pollatsek et al., (2000) found that the frequency of the whole compound had a similar influence on gaze duration and influenced eye movements at least as rapidly as did the frequency of the second constituent. These results suggested that the identification of these compound words involves parallel processing of both morphological constituents and whole-word representations.

Zhou and Marslen-Wilson (1994) investigated the nature of the mental representation of Chinese transparent disyllabic spoken compound words by varying the word frequency, morpheme frequency and syllable frequency with lexical decision tasks. The results showed that the frequency of the entire disyllabic compound is the dominant factor in determining response time in a lexical decision task, and that this effect does not interact with variations in either the morpheme or the syllable frequency of either of the two constituent morphemes, despite the salience and discriminability of morphemes and syllables in Chinese. The authors concluded that these findings rather favor a whole-word representation for Chinese disyllabic compound words.

2.4.1.2. Neurophysiological evidence

Event-related potentials (ERPs) because of their high temporal resolution (in the range of the millisecond) are suitable to investigate online the involvement of language processes known to be rapid as it is the case for prosodic and morphological processes implied in compound words recognition. As it will be presented in Chapter 5 on the neurophysiological marker (or biomarkers), at least two ERP components have been associated with the morphological decomposition process and the integration constituent processes, namely the Left Anterior Negativity (LAN[10]) and (e.g., the N400[11]) respectively.

For the processing of compound words in different languages, ERP data lend support to decompositional models (in French: Rosenfelt et al., 2014; in English: MacGregor & Shtyrov, 2013; in German: Koester, Holle, & Gunter, 2009; Koester et al., 2004; in Basque: Vergara-Martínez, Duñabeitia, Laka, & Carreiras, 2009; in Italian: El Yagoubi et al., 2008). Vergara-Martínez et al. (2009) reported ERP data on the effect of constituent's frequency in processing of Basque compound words that showed an early starting negativity modulated by the frequency of the first constituent with high-frequency first constituents eliciting larger negativities as well as an N400 effect modulated by the frequency of the second constituent with low-frequency second constituents eliciting larger N400 amplitudes than high-frequency second constituents. The authors interpreted the early starting negativity modulated by the frequency of the first constituent as an index of morphosyntactic decomposition and the N400 as an index of the cost of the selection and integration for the whole-word meaning in accessing the mental lexicon. Rosenfelt et al. (2014) investigated the decomposition of French Verb-Noun compounds as a function of the lexical frequency of the compound words (e.g. high frequency compound word like *brise-glace* (icebreaker); low frequency compound word like *brise-fer* (ironbreaker)) in a cross-modal repetition priming paradigm with a lexical decision task using EEG. The visual targets were the compound prime or its constituent (the first or second constituent). The authors observed a priming effect for the constituent presentation conditions indexing by a reduced N400. This result provides clear evidence of decomposition, as a priming effect on a constituent target can only be observed if the constituent was accessed during the processing of the compound prime. With respect to the

[10] LAN is a negative-going deflection that occurs approximately in the same time window as the semantic N400 effect, but generally has a more anterior scalp distribution and is sometimes left-lateralized. LAN is thought to be elicited by the initial morphosyntactic processing (Friederici, 1995) and has been found in all the studies that used morphosyntactic violations.

[11] N400 is a negative-going deflection that peaks around 400 ms post-stimulus onset. It is typically maximal over centro-parietal electrode sites. The N400 is assumed reflecting meaning processing (see Kutas & Federmeier, 2011 for review).

role of frequency in determining whether morphological decomposition occurs in French VN, the results did not provide a univocal interpretation. Indeed, the authors showed that while low frequency compound words undergo decomposition, results for high frequency are mixed as there is evidence of decomposition from first constituent. Based on these findings, Rosenfeld et al. (2014) suggested that maybe another factor is a better predictor of processing route, namely opacity may have played a role, such that all the compounds showed evidence of decomposition because all were transparent.

Koester et al. (2007) investigated the morphosyntactic decomposition of German auditory compound words. The authors constructed novel German compounds consisting of three constituents (nouns) for presentation together with a definite determiner (e.g., *der Stahlhakenpreis* the^{masc} steel^{masc}-hook^{masc}-price^{masc}). The gender agreement of the determiner with the first and the third constituent was manipulated independently resulting in a gender violation. A left anterior negativity (LAN) was observed in the ERP for gender incongruent of initial compound constituents providing online evidence for morphosyntactic decomposition. Contradictory findings challenging the view that an early morphosyntactic decomposition systematically occur during processing of spoken German compound words have been reported by Isel and Friederici (2006). The authors recorded ERP while native speaker of German listened noun phrases (e.g. *der Weinberg* (the vineyard)) composed of a definite article (e.g. *der* (the)) followed by a transparent compound word composed of two monosyllabic lexical morphemes (e.g. *Wein* (wine) and *Berg* (mountain)). The noun phrases were produced with a natural prosodic contour without to artificially introduce a salient word's boundary between the definite article and the first constituent. Isel and Friederici reported a biphasic N400/P600 ERP pattern in response to the processing of grammatical gender violation between the head of the compound words (i.e., the right-most constituent morpheme) and the preceding definite article. However, and critically, no ERP effect was found in response to the processing of similar anomalies between the first lexical morpheme, i.e. the non-head and the preceding definite article. Isel and Friederici (2006) claimed that these findings constitute an empirical argument against an early morphosyntactic decomposition, at least when the prosodic structure of the compound words is natural (there is no salient word's boundary between the definite article and the first constituent), that means its left constituent carries a prosodic pattern (duration and F0 contour) leading to a morphological interpretation in favor of a longer word.

In a recent study, MacGregor and Shtyrov (2013) investigated the neuro-cognitive processing of transparent vs. opaque compounds using oddball design. The authors presented monomorphemic standard stimuli, (e.g., *work*) as standard stimuli mixing with deviant stimulus including transparent (e.g., *homework*) or opaque (e.g., *framework*) compound word, or a meaningless pseudo-compound (e.g., **houndwork*, **grousework*). Whole-form frequency of the compounds was systematically varied over both opaque and transparent groups of stimuli. EEG data revealed that for opaque compounds, the MMN is larger for high frequency than low frequency compounds. This result is in line with previous findings of word frequency effects on the MMN amplitude for monomorphemic words, thus suggesting that opaque compounds is accessed in the manner of whole-form just like monomorphemic words. Moreover, the whole-form frequency effects are absent for transparent compounds. The lack of frequency effects for transparent compounds is clear evidence against a purely full-listing account of transparent compound.

2.4.1.3. Neuroimaging evidence

Functional neuroanatomical correlates of the processing of morphologically complex word came from studies of complex derived or inflected words (Bozic, Marslen-Wilson, Stamatakis, Davis, & Tyler, 2007; Devlin, Jamison, Matthews, & Gonnerman, 2004; Vannest, Polk, & Lewis, 2005; Vartiainen, Aggujaro, Lehtonen & Hultén, 2009; Davis, Meunier & Marslen-Wilson, 2004). Bozic et al. (2007) used event-related functional magnetic resonance imaging (efMRI) to investigate priming between pairs of words that potentially shared a stem, where this link was either semantically transparent (e.g., *bravely-brave*) or opaque (e.g., *archer-arch*). These morphologically related pairs were contrasted with identity priming (e.g., *mist-mist*) and priming for pairs of words that shared only form (e.g., *scandal-scan*) or meaning (e.g., *accuse-blame*). Morphologically related words produced significantly reduced activation in left frontal regions than control pairs (identity, form, or meaning), both for semantically transparent and opaque words. Morphological effects were observed separately from processing form and meaning. Based on the neuroimaging data, the authors proposed a decomposition of complex derived words triggered by their surface morphological complexity during complex words recognition. Moreover, results revealed an association between the left inferior frontal and structural processing of morphologically complex words. Another fMRI study (Devlin et al., 2004) using similar stimuli showed partially inconsistent results, that morphologically related pairs reduced blood oxygenation in the posterior angular gyrus bilaterally, left occipito-temporal cortex, and left middle temporal gyrus. Vartiainen et al. (2009) used magnetoencephalography (MEG) to study the time course of processing of morphological structure and frequency of written inflected monomorphemic Finnish words; furthermore, theses authors applied source localization method[12] to estimate the cortical regions responding to the MEG effects. Morphological complexity evoked a stronger and long-lasting activation of the left superior temporal cortex from 200 ms to 800 ms. Earlier effects of morphology before 200 ms were not found. These findings support the view that the well-established behavioral processing cost for inflected words stems from the semantic-syntactic level rather than from early decomposition. Moreover, Vartiainen et al. (2009) tested low-frequent and high-frequent Finnish morphologically complex words (inflected words) and showed both the morphology effect the two low- and high-frequency inflected words, the authors concluded that the majority of inflected Finnish words appear to be represented in a decomposed. However, morphological effects in functional imaging studies may be task- relevant. When using a task that does not require overtly morphological operations, such as a synonym-monitoring task (in which participants read single words presented on the screen and press a button if the current word was related in meaning to the immediately preceding word), Davis et al. (2004) did not report the morphological effects for four groups of English words with different morphological complexity (i.e., simple, monomorphemic words, complex derived or inflected words).

[12] Source localization methods devote to determine the brain regions where the EEG/MEG effects are generated. Due to the high temporal resolution of EEG/MEG measurement, performing the localization of EEG/MEG sources is thus of particular interest to better understand their generation and dynamic.

2.4.2. Morphological composition process

Morphological composition is the process of combining of morphemes' meaning/ concept after the morphological decomposition took place. Morphological composition is especially important for compound words processing. Given that, the meaning of a compound word is not the simple union of its constituents, a key question to be addressed concerns the mechanism that is used for creating the concept node for a compound word.

2.4.2.1. Behavioural evidence

There is emerging evidence that compound processing is affected by how compounds' constituents can be integrated. For example, by monitoring eye movements during the reading of German compounds, Inhof, Radach and Heller (2000) investigated the effect of inter-word spacing on German compound words. Complex compounds were shown with and without spaces between constituent words. They found shorter latencies in a naming task and shorter viewing durations in a reading task when interword spaces were inserted. Benefits derived from the availability of interword spaces occurred early in reading. In contrast, there was an inhibitory effect with longer final fixation times for compounds with inter-word spacing. The authors concluded that the interpretation of compounds involves two processes: One process involves accessing the constituent word forms and the other process involves integrating the constituents.

The compound words processing is not a simple combination of the meaning/concept of its constituents but also the morpho-semantic connection between them. The semantic composition involves *binding* the constituent concepts together in a particular way (Fiorentino & Poeppel, 2007; Taft, 2003). Studies using the relation priming on compound words lend support to these hypothesis. Gagné and Spalding (2004, 2009) investigated the relation-priming effect for English compounds. The reaction time results showed that snowball ("made-of" relationship) was processed more quickly following *snowfort* with a same relation than following *snowshovel* different relation (i.e. "for-modifier" relation). Jia, Wang, Zhang and Zhang (2013) replicated the same results for Chinese noun-noun compound words. Gagné and Spalding (2011) re-examined the facilitation effect on the relation-priming. The results indicated that the relation-priming effect is due to slower processing in the different-relation condition rather than to faster processing in the same-relation condition. The authors therefore concluded that the processing of compounds is affected by the competition between relational interpretations, such that activating an interpretation other than the established interpretation (as in the different-relation prime condition) increases the difficulty in selecting the established relational interpretation of the compound.

2.4.2.2 Neurophysiological evidence

Semantic composition of compound words using EEG is little investigated. Koester et al. (2007) proposed the N400 ERP component as electroencephalograhical marker of semantic integration process. The authors compared the N400 elicited by transparent German compound words, which require a semantic integration of constituents and by opaque compound words that did not require such an integration. Transparent compounds elicited a lager negativity that occurred during the presentation of the head constituent than opaque compound words; the authors proposed that the N400 might be related to the semantic composition of compound constituents.

Another EEG study from Koester et al. (2009) examined whether semantic integration is incremental or is delayed until the head (the last constituent in German) is available. All compounds consist of three nouns, and the semantic plausibility of the second and the third constituent was manipulated independently (high vs. low). Participants were asked to listen to the compounds and to evaluate them semantically. ERPs in response to the head constituents showed an increased N400 for less plausible head constituents, reflecting the lexical-semantic integration of all three compound constituents. Moreover, they observed as well an increased N400 in response to the second (less plausible) constituents. The occurrence of this N400 effect during the presentation of the second constituents suggests that the initial two non-head constituents are immediately integrated. The results support the view that lexical-semantic integration of compound constituents is an *incremental* process, at least in auditory compound processing.

Bai, Bornkessel-Schlesewsky, Wang, Hung, Schlesewsky, and Burkhardt (2008) examined the N400 for Chinese compounds by systematically manipulating the semantic similarity between the constituents. Two main categories compounds were used: 1) compounds whose constituents do not change semantically, such as 战争 *zhan4zheng1* 'fight + contest' = 'war'; 2) Compounds in which constituents have a semantic change (i.e., opposite meaning), like 开关 *kai1guan1* 'open + close' = 'switch'. The ERP data showed a significant influence of semantic information on the processing of compounds in Chinese indexing by a larger N400 for compounds that consist of semantically distinct meaning units. The authors concluded that the composition of such compounds consumes processing resources, which would be reflected by the modulation of the N400.

With regard to the interne morphological relation of constituents, Jia et al. (2013) investigated the relation-priming effect for Chinese noun-noun compounds using EEG. They observed a significant N400 amplitude reduction from 320 ms following word presentation when prime and target have same relation than when the relation is different. The findings show a composition processing for Chinese compound words, and that the relation information started to contribute to word meaning comprehension 320 ms after a word was presented, likely following meaning access to the first morpheme.

2.4.2.3. Neuroimaging evidence

Brooks and de Garcia (2015) used magnetoencephalography (MEG) with source localization method to investigate the neural bases of the composition process in English compound words recognition. MEG data were acquired while participants performed a word naming task in which three word types, transparent compounds (e.g., *roadside*), opaque compounds (e.g., *butterfly*), and morphologically simple words (e.g., *brothel*) were contrasted in a constituent priming paradigm where the target compound was primed by one of its constituent morphemes (e.g., *road* - *roadside* An analysis of the associated MEG activity revealed that only transparent compounds showed increased activity from 250 to 470 ms in a region of interest implicated in morphological composition, the Left Anterior Temporal Lobe

(LATL)[13]. However, this brain area was not sensitive to opaque compounds whose morphemes do not share a semantic relationship. Thus, these results suggest that semantics play a role in combining the meanings of morphemes.

[13] Neuroimaging studies on the binding mechanism in sentence building (Friederici et al., 2000) and on the composition of noun phrases (Bemis and Pylkkänen, 2011) reveal an implication of the left Anterior Temporal Lobe in the composition of words into phrases.

Chapter 3

Linguistic stress

3. Linguistic stress

Linguists have proposed a classification of languages based on rhythm (Abercrombie, 1967; Eling, Marshall, & van Galen, 1980; Ladefoged, 1975; Pike, 1945). It has been suggested that whereas many Germanic languages are stress-timed, many romance languages are syllable-timed, and languages like Japanese and Tamil are mora-timed. The rhythmic classes were first defined in terms of the timing units that define their rhythms. In stressed-timed languages, isochrony is supposed to be based on the regular occurrence of stressed syllables, whereas in syllable-timed and mora-timed languages, isochrony is supposed to rely on syllables and mora tending to have equal durations, respectively. Taken together, results of studies on the timing unit hypothesis failed to show differences in isochrony (see reviews by Lehiste, 1977, or Bertinetto, 1989). However, the classification of languages based on rhythm remains of central interest.

In languages where each content word has a primary stress like English and German, the syllables of the word are not uttered by the speaker with an equal articulatory effort. One syllable, that received primary (lexical) stress, is uttered with a longer duration and/or a stronger intensity and/or a particular F0 pattern and it is perceived stronger than other syllables; the other syllables may carry a secondary stress (in long words), or are unstressed or even reduced. *Word or lexical stress* corresponds to the accentuation of one syllable within a word which position is stored in the mental lexicon. When two or more content words form a phrase, one of the lexically stressed syllables of the words is more accented than the other lexically stressed syllables; it received a *phrase stress*. The phrasal stress is located by default in English on the lexically stressed syllable of the final word in the phrase. In the sentence, one of the stressed syllables is more accented: it corresponds to the *sentence stress*. The sentence stress is located in English by default on the stress of the final phrase, but in sentence with focus, it may be located on any (focused) word.

In long English words, there are four levels of stress: a primarily stressed syllable (that is the lexically stress syllable), an eventual secondarily stressed syllable, unstressed and reduced syllables. The primary stress is strongest degree of stress placed on a syllable, denoted by a superior vertical stroke (ˈ); Secondary stress is the weaker of two degrees of stress, denoted by a inferior vertical stroke (ˌ); Unstressed syllable receive no stress, no symbol of any type is placed on a nonstressed syllable. For example, the trisyllabic word *Chicago* /*ʃɪ ˈkaː,gəʊ*/ carries the primary stress on *ca*, the secondary stress on *go* and the first syllable *Chi* is unstressed. Some unstressed syllable undergo vowels reduction or neutralization by transcribing to the schwa /ə/, for example, in ***banana*** /*bəˈnanə*/, the first and the last vowel turn to /ə/.

Only one syllable that bears primary (lexical) stress (with few exceptions; see Hyman, 1977). The hierarchical relation of prominence between the different levels of stresses (lexical, phrasal and sentence) is described within the Metrical Phonology (Liberman, 1975; Liberman & Prince, 1977; Halle & Vernaud, 1978), which is a phonological theory describing the organization of syllables into rhythmic hierarchies. In Metrical Phonology, segments are organized into syllables, syllables into metrical feet[14], and feet into phonological words, these units are referred to as nodes with relational connections. The organization of the nodes is represented formally by the metrical tree (see Figure 3.1) and the metrical grid (see Figure 3.2), as displayed below.

In the metrical tree, the different degrees of prominence of a word are described by the relations between nodes in a branching tree. Prosodic patterning in the metrical tree is a relational property: a node is strong only by virtue of the fact that it is the sister of a weak node. The label *s* indicates relative stronger prominence, and the label *w* indicates weaker prominence. The Figure 3.1 shows an example of the metrical tree of the word *Alabama*:

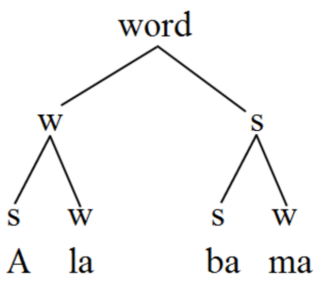


Figure 3.1. The metrical tree of the word *Alabama* (from bottom to top: the syllables, the feet, and the word).

[14] Feet (foot) are rhythmic units with one or more syllables, which make up words. Feet represent the rhythmic structure of the word and are the units that allow us to describe stress patterns.

The word *Alabama* contains four syllables. The primary word stress falls on the third syllable: *ba*. The other labels are defined relatively to that syllable, with the tendency to alternate strong and weak syllables, whenever it is possible (s, ws or sw, etc.); the word *Alabama* is therefore constituted by two feet.

Prosodic patterning can be also represented in a grid way, i.e., a metrical grid. The height of the grid columns is indicated by the number of stars (*), which represents the degree of prominence. Figure 3.2 displays an example of the metrical grid of the word *Alabama*. As we see, whereas the third syllable is the most prominent one (i.e. higher degree represented by three stars), the second and last syllables are the least prominent.

```
    *
*   *
* * * *
Alabama
```

Figure 3.2. The metrical grid of the word *Alabama.*

As rhythmic units of words, metrical feet allow us to describe stress patterns. There are two kinds of feet: left-dominant and right-dominant. If the left syllable node of a foot in the metrical tree, is strong, it is called that a left-dominant foot or trochaic foot, such as showed in Figure 3.3(a); in contrast, if the right syllable node of a foot is strong, that is a right-dominant foot or iambic foot, as in Figure 3.3(b).

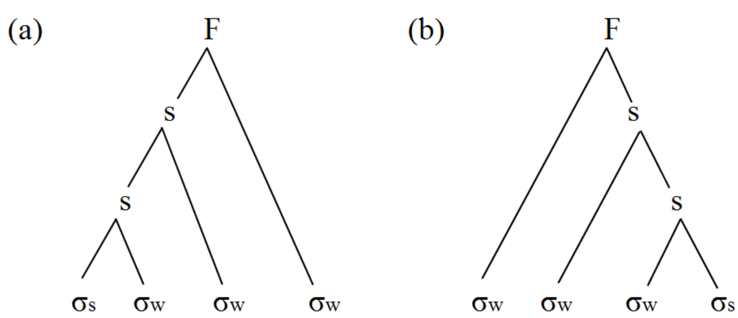


Figure 3.3. (a) A schema of left-dominant feet or trochaic feet; (b) a schema of right-dominant feet or iambic feet.

English is a left-dominant language. For example, *consultation* has two feet, /*kɔn.səl*/ and /*tæɪ.ʃən*/. In each of these feet, the left-most syllable is strong and the right-most is weak, that is, left-dominant. According to Duanmu (2004, 2007), Mandarin Chinese has left-dominant disyllabic feet, like in English.

Some languages have a so-called "fixed" word stress, meaning that the word stress always falls on a particular position within the word (e.g. on the first, peninitial (second), antepultimate (third-to-last), penultimate (second-to-last), final syllable). For example, in Finnish and in Czech, the lexically stressed syllable is usually the first syllable; in Polish, the penultimate syllable is lexically stressed; in French, the last syllable is lexically stressed. Other languages, like English, Russian, Germany, Italian, Portuguese and Spanish, the position of word stress in a word cannot be decided simply in relation to the position of the syllables of the word. In such languages with word stress with variable position within the word, the position of word stress may be used contrastively to distinguish meanings. For example, the German minimal pair *unterSTELLEN* 'to insinuate' and *UNTERstellen* 'to store' are distinguished only by the different position of primary word stress. In English, many two-syllable words whose meaning and word class change with the placement of the word stress. For example, the word *record* is a noun if the first syllable is lexically stressed (i.e. *REcord*) whereas it is a verb, when word stress falls the second syllable (i.e. *reCORD*). More examples: the words *insult, desert, import, export, contrast and protest* can all be nouns or verbs depending on whether the word stress is on the first or second syllable. These minimal pairs (e.g. *REcord* vs. *reCORD*) are especially interesting since they represent the interface between prosody and morpho-syntax *par excellence*. These minimal pairs were investigated in studies of prosodic information in psycholinguistics that address one of most important question in the recognition of spoken words: does and how prosodic information (e.g., word stress) plays a role in word recognition? Current psycholinguistic models of spoken-word recognition conceive of this role as activation and selection between potential candidate words (see Cutler, Dahan, & van Donselaar, 1997 for a review).

Word stress can be subdivided into simple word stress and complex word stress (Roach, 1998). Simple word denotes words that are not composed of more than one grammatical unit, so that, for example, *happy* is a simple word while *happiness* (derivation) and *happy hour* (compounding) is complex. The stress that placed on compound words is called the compound stress, which defines the prominence relation between the constituents of a compound. At higher levels than word, as mentioned before, the stress is referred to as phrasal stress (phrase or clause level) and sentence stress (sentence level). We will discuss in detail the characteristics of compound stress and phrase stress in section 3.3.

We presented above the basic description of stress in linguistics. In the next part of Chapter 3, we will firstly discuss the phonetic correlates of stress and stress perception; the general prosodic characteristics and stress phenomenon in Mandarin Chinese will also be presented; moreover we will introduce the specification of compound vs. phrase stress; finally, a series of psycholinguistic experiments that substantiate the role of the stress in the recognition of spoken words will be presented and discussed.

3.1. Acoustic-phonetic correlates of stress

According to Lehiste (1970), the production of increases or decreases in stress are ultimately due to increases or decreases in the speaker's physical effort. She states that "*A stressed syllable is one that the speaker consciously utters with greater effort than neighboring syllables - the listener hears this syllable as louder than unstressed syllables…*" (Lehiste, 1970).

A number of studies on stress perception have been carried out to determine what acoustic information makes a syllable perceived as prominent. From the pioneering works on the perception of stress by Fry (1955, 1958), researchers have agreed that stress correlates with a complex configuration of events of increased duration, heighten fundamental frequency (F0) and raised intensity, e.g. English (Fry, 1955, 1958; Lehiste, 1970), Spanish (Ortega-Llebaria, 2006), Dutch (Sluijter & van Heuven, 1996), Russian (Bondarko, Verbitskaya, & Tscherbakova, 1973), Polish (Jassem, Morton, & Steffen-Batog, 1968), Tongan (Garellek &

White, 2015), Chickasaw (Gordon, 2004), Indonesian (Adisasmito-Smith & Cohn, 1996) and Chinese (Liu & Xu, 2005; Cao, 1992; Shen et al., 2013); and that several cues may be functionally equivalent cross-linguistically (Vaissière, 2004: final lengthening in French as a phrase right boundary marker is equivalent to initial F0 resetting as a left boundary marker in Japanese). The syllable carrying sentence stress is produced with a supplement of air pressure, which produces higher intensity and higher fundamental frequency (Ladefoged, 2001).

Duration

Cross-linguistically, the length of the syllable is regarded as one of the most reliable cues for discriminating vowels with word stress from unstressed vowels. In the pioneering experiment (Fry, 1955), the implication of duration, F0 and intensity in the judgment of stress was investigated by using English minimal noun/verb pairs such as *OBject - obJECT*. Results showed that duration is a consistent correlate of stress at the word level in English and that it is a more effective cue than intensity and F0. Since then, researchers have started to give up the classical view that stress is equated to a higher degree of intensity and much research have showed the crucial role of duration in stress production and perception. A number of researches (Klatt, 1976; Crystal & House, 1988; Turk & Sawusch, 1996) reported that English vowel durations are influenced by stress and listeners attend more to duration as a most reliable cue in the perception of stress. Production and perception studies in other languages showed mostly an important role of the syllabic duration, and that the longer the length of a syllable the more prominent in stress. For example, Gordon (2004) investigated the word-level stress in Chickasaw. The production data showed significant overall effect of stress on duration and the same general pattern: Chickasaw primary stressed vowels were longest (87 ms averaged over all speakers), and unstressed vowels were shortest (58 ms), with secondary stressed vowels having intermediate duration values overall (76 ms). A recent study in Tongan (Garellek & White, 2015) showed that vowels with primary stress are significantly longer in duration (by about 30 ms) than unstressed vowels. Note however, that final unstressed syllable in a word is also lengthened due to the phenomenon of final lengthening: the role of duration is language dependent (Vaissière, 1982).

Data from Chinese full tone and neutral tone[15] syllables showed a similar pattern. Lin and Yan (1980, 1990) but also Cao (1992) showed in production studies that duration of the unstressed neutral tone syllable is systematically shorter (reduced by approximately 50%) than a syllable with full tone. Li, Gao, Jia, and Wang (2014) investigated the role of duration in different production contexts. Chinese target words were recorded (1) in isolation, (2) on-focus where target words were embedded in a carrier sentence in the focus position (e.g., in the sentence: Which word did Xiaozhao learn yesterday? Xiaozhao learned the word ***mo2ku1/mo2ku0***[16] yesterday.), and (3) post-focus where targets were embedded in a carrier sentence in the post-focus position (e.g., in the sentence: When did Xiaozhao learn the word *mo2ku1/mo2ku0*? It is **yesterday** that Xiaozhao learned the word *mo2ku1/mo2ku0*.). The production data showed that duration presents a rather constant impact on the focus perception across the three conditions. Perceptual tests were carried out in the study of Shen (1993) to examine the implication of phonetic cues in the perception on the relative syllable prominence in Mandarin Chinese. The results revealed that stress could be identified on the basis of duration and intensity alone. However, and similarly in other languages, duration is a more important cue than intensity.

<u>Fundamental frequency</u>

The physicist Joseph Fourier (1822) has discovered that a complex sound, and, therefore non-sinusoidal, could be decomposed into a succession of pure tones, called harmonics. The periodic signals are rarely purely sinusoidal and are in fact a mixture of sinusoidal signals whose respective frequencies are multiples of the fundamental frequency, the lower frequency. Thus, in acoustic, the fundamental frequency (F0) or pitch[17] is defined as the lowest frequency of a periodic waveform (first harmonic).

Lehiste (1970) states that "*the perception of stressedness appears to be based on a number*

---

[15] Neutral tone is sometimes thought of as a lack of tone. It is associated with weak syllables. For more information, see section 2 in the present chapter.

[16] The *mo2ku1* / *mo2ku0* are the target words where *mo2ku1* carries normal tone while *mo2ku0* carries a neutral tone on the last syllable. They are on-focus in this sentence as marked in bold.

[17] The fundamental frequency is referred to as the lowest frequency of a periodic waveform, e.g., speech. Pitch is the perceptual correlate of fundamental frequency. Pitch is closely related to the vibration frequency of the vocal folds during the sound productions.

of factors, the most influential of which is fundamental frequency…". F0 has been shown to be a major acoustic manifestation of suprasegmental structures in many studies. It is claimed by some researchers to be the strongest cue of word stress perception for stress languages (Cooper et al., 1985; Lieberman, 1960; Gussenhoven, Repp, Rietveld, Rump & Terken, 1997; Li et al., 2014). A simple increase in the F0 of a syllable may turn a lexically unstressed syllable into perceptively prominent syllable. Fry (1958) manipulated F0 of synthetic disyllabic words and showed that a fundamental frequency difference of 5 *Hz* is sufficient to influence which of the two syllables is perceived as stressed. In a word, the higher F0 in a syllable compared to the other syllables makes thus the syllable more prominent. To place some movement of pitch (e.g., rising or falling) on the syllable is even more effective cue of prominence than the level of the F0 (Roach, 2010). Ortega-Llebaria (2006) measured the pitch variation (calculated by the value of the standard deviation within syllable) in lexically stressed and unstressed Spanish syllables. Results demonstrated that the value of the standard deviation in the lexically stressed vowel increased dramatically suggesting that there was pitch variation on this vowel; while the unstressed vowel obtained a standard deviation close to zero showing no pitch movement. The pitch variation in stress perception is important for tonal languages. For example, in Mandarin Chinese, the pitch range[18] (or pitch variation) has been shown wider when syllables carry word stress (Shen, 1985; Moore, 1993; Liu & Xu, 2005; Shen et al., 2013). More specifically, when a Tone 3[19] carries word stress, it is dipped lower and, when a Tone 4[20] is lexically stressed, it starts higher and falls lower than the corresponding neutral tones, which are unstressed (Chao, 1968). Moreover, computational corpus studies (Kochanski, Shih, & Jing, 2003) have established quantitative F0 predictions in terms of the lexical tones and the prosodic prominence of each word. However, some investigations of English have shown that F0 is not a necessary cue, because stress can be identified on the basis of duration and intensity alone (Cutler & Darwin, 1981). In order to investigate whether F0 is necessary to perceive stress in Mandarin Chinese, Shen (1993) manipulated the recorded utterances by applying

---

[18] Difference between the max F0 and the min F0.
[19] Chinese Tone 3 is third tone having a falling-rising (or dipping) pitch contour.
[20] Chinese Tone 4 is fourth tone having a high-falling pitch contour. See section 2.1 for detailed information about Chinese tone system.

filters; the F0 was held constant at 135 Hz and duration as well intensity information was preserved in the filtered utterances. The results of perceptual tests showed that participants could identify relative syllable prominence in the filtered utterances suggesting that F0 is not a necessary cue to perceive stress in Mandarin Chinese and that duration and intensity alone is sufficient. Nevertheless, the author did not exclude F0 as a stress cue in Mandarin Chinese. He concluded that the non-prevalence of the F0 cue in stress judgments of Mandarin Chinese probably should be attributed to its contrastive lexical usage (the contour characteristics were likely to contribute mainly to tone discrimination) and the F0 variation in the acoustic signal is shared by tone and stress.

Intensity

Crystal (1969) defines stress as "*those variations in linguistically contrastive prominence primarily due to loudness*". Most people seem to feel that syllables perceived as stressed are louder than the ones perceived as unstressed; in other words, loudness is a component of prominence. In literature, the role of intensity for stress is not agreed upon. Fry (1955, 1958) showed that intensity was a less effective cue than duration on the perception of linguistic stress patterns. Nevertheless, some authors have argued that the strongest cue to prominence is intensity for English (e.g., Beckman, 1986; Turk & Sawusch, 1996). For Mandarin Chinese, the effect of the intensity is only secondary (Cao, 1986, 1992; Lin, 1985, 2006; Lin & Yan 1990). Studies on the neutral tone in Chinese showed that the intensity of the unstressed neutral tone is not necessary lower than the one with full tone (Cao, 1986, 1992). The unstressed neutral tone raises its intensity after Tone 3 (Lin, 2006; Cao, 1992). In a production study, Cao (1992) conducted acoustic analyses illustrating that the destressing of the neutral tone syllable is not related simply to its intensity. The intensity of a neutral tone syllable is lower than that of one with full tone in general (-6.5 dB), but the situation is reversed when it is preceded by a Tone 3 syllable (+0.7 dB).

3.2. The phonology of Standard Mandarin and the Neutral tone

Mandarin Chinese, one of the Chinese languages, which is spoken natively in China roughly in the north of the Yangtze River and in most of Southwest China. The phonology of Standard Mandarin[21], also known as Mandarin and Putonghua, is based on the Beijing Mandarin dialect. The phonological system of Standard Mandarin includes the segments (i.e., the vowels and consonants), and the tones that are applied to each syllable. There are 25 consonants in Standard Mandarin. Table 3.1 organizes the consonants in the way that Chinese linguists often do: labials, alveolars, alveolar sibilants, retroflexes, alveolo-palatals, and velars are presented in rows, while manners of articulation are presented in columns.

Table 3.1.

Mandarin consonant inventory

	Unaspirated	Aspirated	Nasals	Fricatives	Voiced continuants
Labials	p	p^h	m	f	
Alveolars	t	t^h	n		l
Alveolar sibilants	ts	ts^h		s	
Retroflexes	ʈʂ	$ʈʂ^h$		ʂ	ɻ
Alveolo-palatals	tɕ	$tɕ^h$		ɕ	j ɥ
Velars	k	k^h	ŋ	x	

There are five simple vowel phonemes in Mandarin: /*a, u, ə, i, u, y* / (Duanmu, 2007) as shown in Table 3.2.

[21] In this thesis, if no explanation is given, Standard Mandarin Chinese is referred to as Chinese.

Table 3.2.

Mandarin vowel phonemes

	Front		Central	Back
High	i	y		u
Middle			ə	
Low			a	

With regard to tones, Standard Chinese has four main tones that are high-level, rising, dipping, and high-falling, in addition to a neutral tone used on weak syllables. The most commonly adopted description of the main four lexical tones is Chao's five point scale (Chao 1956, 1968), in which a pitch range is divided into five levels, [5] being the highest and [1] being the lowest. The four tones of Standard Mandarin are marked as [55, 35, 214, 51] respectively (see Figure 3.4). In *pinyin*[22], the phonetic spelling system used in China, tone is indicated by the tone mark. The description of tones is shown in Table 3.3.

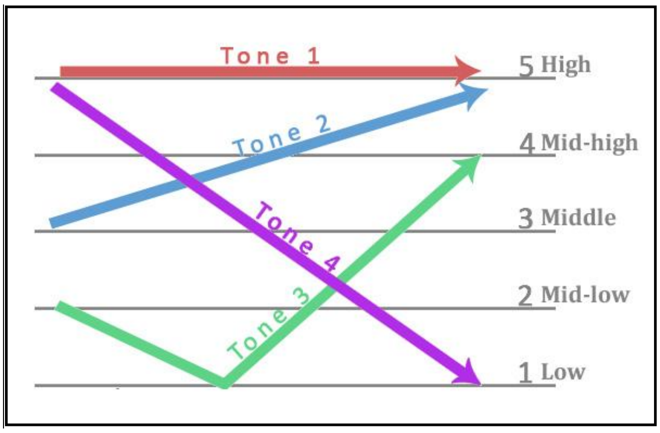


Figure 3.4. Chinese four tones at Chao's five-point scale.

[22] Pinyin (Chinese: 拼音) literally means 'spelled-out sounds', it is the official phonetic system for transcribing the Mandarin pronunciations of Chinese characters into the Latin alphabet.

Table 3.3.

Description of Mandarin tones.

Tone	Pitch description	Tone in *pinyin*	Chao's five-point scale	Notation in this study
Tone 1	High-level	ā	55	a1
Tone 2	High-rising	á	35	a2
Tone 3	Dipping	ǎ	214	a3
Tone 4	High-falling	à	51	a4
Neutral tone	Light	a	-	a0

As mentioned before, in addition to the four main tones, there is a neutral tone (Chao, 1968), also called fifth tone or zeroth tone (Chinese: 轻声, literal meaning: 'light tone'). Neutral tone is a neutralization of a corresponding full tone, and it is associated systematically with weak syllables. Syllables that have a neutral tone are pronounced in a short and light manner, thus sounds quite weak. What these weak syllables have in common is that 1) they must be attached to other syllable as a dependent unit in polysyllabic words; 2) they are usually the final syllable and never occur in the initial position in a word.

There are many cases where the presence of the neutral tone is morphologically predictable, including suffixes (e.g. the suffix *zi* in (1)), particles (e.g., the genitive/nominalizer marker *de* in (2)), reduplication (e.g., *kan* in (3)) and diminutive terms (e.g., *mei* in (4). However, the presence of the neutral tone in lexical items (e.g., the morpheme *xi* in (5)) is morphologically non-predictable.

(1) 桌子 *zhuo1**zi0*** 'table'

(2) 好的 *hao3**de0*** 'good'

(3) 看看 *kan4**kan0*** 'look'

(4) 妹妹 *mei4**mei0*** 'sister'

(5) 东西 *dong1**xi0*** 'thing'

A syllable with any of the four tones at a non-initial position can have a neutral tone when it does not carry word stress. The right-most syllable in certain suffixes and particles, such as the suffix *zi* in the above example (1) and the particle *de* in (2), has always a neutral tone and thus are unstressed. As mentioned before, the neutral tone syllables can be also found in the right-most position of some disyllabic compounds. However, they are not predictable and may be lexically determined in Standard Mandarin. However, they are not predictable and may be lexically determined in Standard Mandarin. It is claimed that in Standard Mandarin, about 15-20% of the syllables in written texts carried a neutral tone (Li, 1981). A corpus study (Chen, 2004) on the Modern Chinese Dictionary (1996 edition) with sixty thousand entries (word) revealed that 2882 entries contain the neutral tone, 1511 of them (52.4%) are morpho-syntactically based, like the above examples in (1), (2), (3) and (4) and the other 1191 (41.3%) are lexically based, such as in (5).

Chinese is considered as a language having no word stress at the word level because the tone is used for distinguishing lexical meaning (Hyman, 1977; Selkirk & Shen, 1990). Nevertheless, the clear difference between a syllable with a full tone and a weak (or reduced) syllable with a neutral tone makes people (e.g., Chao, 1968; Lin, 2001; Duanmu, 2000, 2007) to agree that there are two degrees of “stress” (i.e. of strength) in Chinese, i.e. “stressed” that is loaded by a syllable with a full tone and “unstressed” loaded by a syllable with a neutral tone. In the case where the two syllables of a word carry both a full tone, the relative syllable prominence is less evident and Chinese linguists do not have agreement on the pattern of the relative syllable prominence for normal syllables words. Chao (1968) proposed that a two-syllable word would have a slightly greater stress on the second syllable unless it is in the neutral tone. In contrast, Lin (2001) and Duanmu (2004, 2007) argued for the left prominence in a disyllabic Mandarin Chinese word.

A number of phonetic studies have been carried out to investigate the acoustic realization of the neutral tone. These quantitative data showed that the neutral tone has distinctive patterns compared to its corresponding full tone. Firstly, the neutral tone is consistently shorter in duration. Lin and Yan (1980, 1990) and Cao (1992) showed that duration of the neutral tone syllable is reduced by approximately 50% compared to a syllable with full tone. Another specific feature of neutral tone is that it loses of its original tone pattern and the shape of the F0 contour depends on the tone of the preceding syllable (Cao, 1992; Lin & Yan, 1980). For example, Cao (1992) found that the neutral tone rises slightly after a Tone 3 syllable, while it falls after all of the other tone's syllables; In addition, the F0 range of the neutral tone is flattened to practically zero (Shen, 1985; Liu & Xu, 2005; Shen et al., 2013). More specifically, a neutral Tone 3 almost not dip while the corresponding Tone 3 has a falling-rising contour and, for a neutral Tone 4, it starts lower and falls higher than its corresponding full tone (Chao, 1968). Moreover, the intensity of the neutral tone is generally lower than the corresponding full tone. However, it is not necessary lower than the one with full tone. The neutral tone may raise its intensity after Tone 3 (Lin, 2006; Cao 1992). Figure 3.5 presents the realization of a neutral tone and its corresponding full tone in Praat[23].

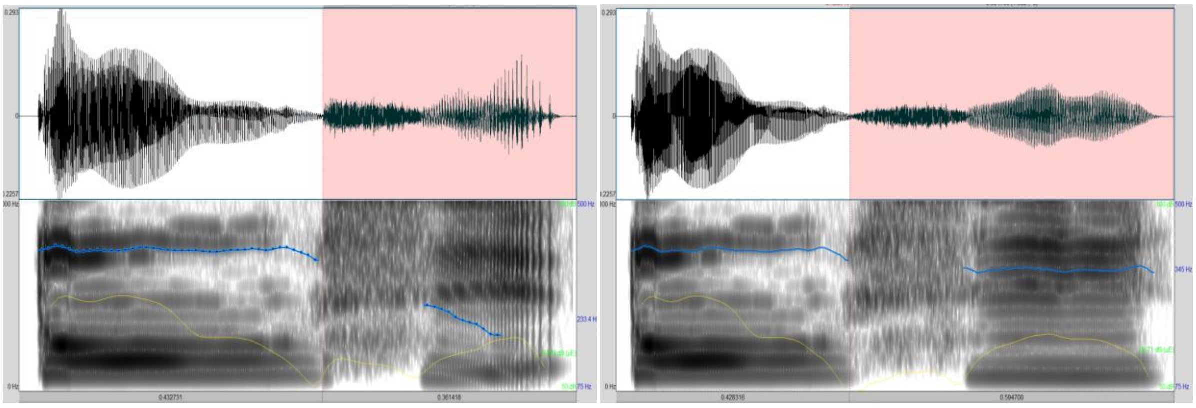


Figure 3.5. The left figure shows the realization of the compound *dong1xi0* 'thing' where the second syllable (colored) has a neutral tone (the corresponding full tone is Tone 1). The blue line represents the F0 contour; the yellow one represents the intensity. In the example, the mean F0 of the neutral tone is 233 Hz; the mean intensity is 60 dB; and the duration is 361 ms. The right figure displays the realization of the compound *dong1xi1* 'east and west' where the second syllable (colored) has a full tone. In the example, the mean F0 of the neutral tone is 345 Hz; the mean intensity is 61 dB; and the duration is 595 ms.

[23] Praat is a scientific computer software package for the analysis of speech in phonetics.

3.3. Compound/Phrasal Stress

As mentioned in the above chapter, in English, compounds and phrases constructions may be phonemically contrastive, as is evidenced by minimal pairs such as *blackbird* vs. *black bird.* In a compound, the first constituent receives the only primary lexical stressed syllable[24] and the phrasal stress; in phrases, both words bear primary lexical stress, with the second word receiving also the phrasal stress (Chomsky & Halle, 1968; Hayes, 1995; Gussenhoven, 2004; Plag, Kunter, Lappe, & Braun, 2008). As for the example of *blackbird* vs. *black bird,* the compound *BLACKbird* has a word stress on the left constituent *black* and the syllable *black* is the most prominent syllable in the compound; in the phrase *black BIRD*, *bird* receives a phrasal stress, and the second word *bird* is the most prominent syllable in the phrase. The contrastive stress pattern in compounds and phrases were confirmed by acoustic studies. Farnetani et al. (1988) showed that minimal pairs of compounds and phrases tend to differ in length, with the phrase being slightly longer (there are composed of two words, instead of one, with each characterized by word final lengthening, and there is an extra between-word boundary, as compared to the compound word. In addition, compounds and phrases differ also in their constituents. In a production experiment, Morrill (2012) examined the phonetic realization of stress pattern for English compounds and phrases of adjective-noun structures in a variety of intonational environments. Participants produced compounds and phrases such as *greenhouse* and *green house* in different sentence types and sentence positions to elicit various intonational environments (i.e., at position of the subject, clause-final, statement-final and question sentence). Phonetic analysis revealed clear acoustic patterns in intensity, duration and pitch of the constituents that distinguish compounds from phrases. Intensity is a correlate of stress that distinguishes between compounds and phrases in all environments except for question intonation. Duration differences are a reliable cue in question intonation environment, and also in clause-final intonation environment. The ratio of the duration between the first

[24] Despite the prevalence of the strong-weak pattern in compounds, other stress patterns also occur in English. Compounds such as "*silk SHIRT*" have the primary stress on the second constituent. Bell and Plag (2012; 2013) showed that the leftward or rightward stress assigned to compounds is influenced by the informativity of the constituents measured by multiples characteristics such as length, frequency, semantic transparency and family size, etc. Compounds with a relatively informative second constituent are more likely to receive a lexical stress on the second constituent than compounds with a less informative second constituent.

constituent and the second constituent in compounds is bigger than that in phrases. Distinctive pitch patterns were also observed across intonation environments.

This stress contrast creates interesting minimal pairs of segmentally identical but prosodically distinct compounds and phrases. There are morphological and semantic distinctions between compounds/phrases minimal pairs, such as *greenhouse* is a compound that refers to glasshouse for plants and *green house* is a phrase means a house that in green color. In addition, when we examine further the semantic of the constituents, sometimes we find that the first constituent of the compound becomes semantically opaque. In the above example, *green* in *greenhouse* represents no more a color. The phenomenon can be observed in more examples: *hotdog*/*hot dog*, *blackboard*/*black board*, etc. It is believed that this perceptual ability on stress contrast is critical to parsing the speech stream, in particular the ambiguous pairs. A number of psycholinguistic research have been carried out to investigate the role of prosody during language processing by using compounds/phrases minimal pairs (e.g., Farnetani et al., 1988; Vogel & Raimy, 2002; McCauley, Hestvik & Vogel, 2013). Research focused on the perception, processing and acquisition on the compound/phrasal stress, experimental evidence will be presented in section 3.4.

3.3.1. Specific minimal pair: Neutral tone compound vs. Full tone phrase

According to the literature, it is not very clear whether or not one syllable with a full tone received or not an underlying lexical (word) stress at the word level. F0 pattern is strongly constrained for marking the full tones, and therefore cannot be used for marking word stress like in English. Therefore, we will avoid using the expression word stress when considering Mandarin Chinese at least for word that all of its syllables carry full tones. However, not all syllables at the word level are uttered with equal strength. There are syllables with full tone and syllables with neutral tone.

The two syllables in a word in Mandarin Chinese can receive a full tone or one may receive a neutral tone. Neutral tones are reminiscent of reduced syllables in English. Many Chinese compounds (approximately 5%, i.e., 2882 included in the Modern Chinese Dictionary, 1996 edition) hold a neutral tone. Among them, we find certain can be phrase simultaneously, as shown in Table 3.4:

Table 3.4.

Examples of Chinese compounds/phrases minimal pairs

Example	Elements	Coumpound	Phrase
(1)	东西 dong1 + xi1	thing	the east and the west
(2)	马虎 ma3 + hu3	careless	horse and tiger
(3)	眉目 mei2 + mu4	prospect of a solution	eyebrow and eyes
(4)	兄弟 xiong1 + di4	brother	elder brother and younger brother

Note: the second syllable carries the neutral tone in compounds and the full tone in phrase.

There is contractive stress pattern in these minimal pairs, that is to say compound word (combined with normal and neutral tone) carries a strong-weak stress patterning and the corresponding phrase (combined with two full tones) presents strong-strong stress patterning.

In addition, some of them exhibit an alteration in the semantic manifestation of the first constituent, that is to say, when the compounds receive a neutral tone on the second syllable, the first constituent then alters to semantically opaque, thus, the first constituent such as the *east* in the above example (1) does not contribute to the meaning of the compound. The same pattern is also true for the examples (2), (3). This characteristic is interesting because the disambiguation prosodic cue (i.e., the neutral tone on the second constituent) could modulate the semantic representation of its preceding constituent. Indeed, it is assumed that the presence of a neutral tone on the second constituent does influence the semantic transparency of the first constituent, which in that case becomes semantically opaque. Therefore, a right-to-left retroactive processing of prosodic information would have to take place in Mandarin Chinese for reanalyzing the semantic proprieties access of the left constituent that will have impact of the recognition of the whole word. This retroactive process would work in addition to the usual sequential left-to-right spoken-word processing. To test the validity of the hypothesis of a right-to-left processing in Chinese compound words with a neutral tone on the second constituent, we conducted Experiment 2 using a semantic priming paradigm with a lexical decision task to investigate the possible right-to-left retroactive processing of prosodic information in Mandarin Chinese.

3.3.2. Specific minimal pair: VN/VO

We have discussed in section 2.3, the Verb-Noun form may be considered as a Verb-Object phrase or as a compound of different types (i.e. attributive compound or subordinate compound) and is an especially interesting case of apparent indeterminacy between morphology and syntax. The dual status of Verb-Noun form result in many ambiguous compounds/phrases minimal pairs. There are neither morphological nor semantic or syntactic cues for making a distinction between them. Nevertheless, prosodic information of the ambiguous compounds/phrases minimal pairs may be involved in their distinction. According to the Compound Stress Rule and the Nuclear-Stress-Rule (Chomsky & Halle, 1968), stress placement is different between compounds and the corresponding phrases, with left stress for compound and right phrasal stress for a phrase. Furthermore, each content word of a phrase keeps its own word stress. Taking the *black bird* pair as an example, when compound word is uttered in isolation: *black* carries word stress and phrasal stress while bird carries no stress, thus the stress pattern is 2:0, *black* carries the main stress (degree 2); when phrase is uttered in isolation, *black* receives the word stress and *bird* receive the word stress and phrasal stress, the stress pattern is 1:2, then *bird* carries the main stress (degree 2).

It has been proposed that the distinction of different degree of prominence between the syllables in compounds and phrases is true for Mandarin Chinese, and it permits one to distinguish between compounds and phrases (Duanmu, 2000). Duanmu (1990, 2000) proposed a general rule for assigning Mandarin Chinese compound and phrasal stress, the Non-head Stress Rule. This rule achieves similar effects as the Compound Stress Rule and the Nuclear Stress Rule combined and assumed that in the syntactic structure [X XP] (or [XP X]), where XP is the non-head and X the head, the non-head XP should be stressed. According to Non-head Stress rule, the non-head should have stress in the VN/VO minimal pairs, and because the non-head is on the left in compounds and on the right in most phrases, stress is assigned to the left in compounds and assigned to the right in phrases. An example in Duanmu

(2000):

(3) 炒饭 *CHAOfan* 'FRY rice' $[V+N]_N$ = 'fried rice'

(4) 炒饭 *chaoFAN* 'fry RICE' V+O = 'fry the rice'

炒饭 *CHAOfan* in (1) is a noun compound (a single word) in which the verb modify the noun head. The non-head *chao* holds the lexical compound and the phrasal stresses. With respect to the verb-object structure 炒饭 *chaoFAN* in (2), the phrasal stress is placed on the syntactic non-head *fan*.

However, there is little acoustic evidence for testing Duanmu's assumption on the distinction of stress pattern in Mandarin Chinese VN/VO minimal pairs. We therefore conducted a production study (see Experiment 3) to investigate the acoustic correlates of stress between VN/VO compounds and phrases in Mandarin Chinese. Another experiment (see Experiment 4) has been carried out to examine the perception of the stress pattern in VN/VO compounds and phrases.

3.4. Experimental evidence on the role of stress during language processing

3.4.1. Behavioural evidence

Spoken-word recognition involves the processing of the information about its constituting phonological segments, but also about a specific prosodic patterning, realized via suprasegmental parameters such as pitch contour, duration, amplitude, and consonant and vowel reduction/strengthening. Psycholinguistic studies revealed that the prosodic patterning information such as word stress is processed during spoken word recognition. Previous research in the area has examined:

a) Stress "deafness" in second language learning (Dupoux, Pallier, Sebastián-Gallés, & Mehler, 1997; Peperkamp & Dupoux, 2002);
b) The influence of correct word stress on the identification of phonemes (Connine, Clifton, & Cutler, 1987; Haggard, Ambler, & Callow, 1970);

c) The influence of incorrect word stress on the recognition of words (Slowiaczek, 1990; Small, Simon, & Goldberg, 1988; Koester & Cutler, 1997);

d) Identification of gated words using stress information (Cutler & Otake, 1999; Lindfield, Wingfield, & Goodglass, 1999; Wingfield, Lindfield, & Goodglass, 2000);

e) The use of word stress in a cross-modal fragment-priming task (Soto-Faraco, Sebastián-Gallés & Cutler, 2001; Tagliapietra & Tabossi, 2005; Cutler, 1986);

f) The role of stressed syllables in the segmentation of the speech signal (Cutler & Norris, 1988; van Heuven, 1985; Taft, 1984);

g) The role of prosodic information in compound words recognition (Isel et al., 2003; Koester et al., 2004).

a) Research on the perception of word stress suggests that speakers of languages with non-predictable (e.g., English, German and Spanish)[25] are more efficient than speakers of languages with fixed stress (e.g., French, Finnish and Polish) at distinguishing nonsense words contrasting in stress location. In contrast, speakers of languages with fixed stress exhibit difficulties to represent distinctive stress information in non-predictable stress languages because contrastive stress is not relevant for words of their native language. The phenomenon has been called stress 'deafness'. Dupoux et al. (1997) found that French listeners exhibit more difficulties than Spanish ones in discriminating non-words that differ only in the location of word stress. In French, unlike in English (e.g., *REcord* vs. *reCORD*), stress does not carry lexical information, but predictably falls on the word's final vowel. Speakers of French, then, do not need to process stress to identify lexical items; given its fixed position, stress may instead be used as a cue to word segmentation (Rietveld, 1980). Peperkamp and Dupoux (2002) studied more languages (i.e., Finnish, French, Hungarian, Polish) with non-contrastive stress, proposed a typology of stress deafness. They postulate that the size of the deafness effect corresponds to the degree of regularity of stress patterns in a language.

[25] Although English is not a fixed stress languages, some regularity of stress patterns emerge. The phenomenon has been used by Cutler & Norris (1988) for describing a model of word recognition in English based on the detection of strong syllable.

b) In order to examined the influence of correct word stress on the identification of phonemes, Connine et al. (1987) used two English words (i.e., *digress* and *tigress*) that differed phonemically only in the voicing feature of the initial stop consonant (/*d*/ vs. /*t*/) and in terms of their canonical stress pattern: *diGRESS* and *TIgress*. They synthesized two voicing continua, *diGRESS-tiGRESS* and *DIgress-TIgress*, that differed from one another only in their stress pattern. In the series, *diGRESS-tiGRESS*, the voiced endpoint (i.e., *d*) formed a real word; conversely, in *DIgress-TIgress*, the voiceless endpoint (i.e., *t*) formed a word. The initial segment in both series has an ambiguous status in the midrange of the continua. The authors found that subject's identification of the initial segment of these items was biased in the midrange of the continua in that they were more likely to report a segment that resulted in a real word than one that resulted in a nonword. The listeners clearly could use the stress information to resolve the phonetic ambiguity.

c) Slowiaczek (1990) carried out a series of experiments that systematically examines the influence of the word stress in the recognition of isolated spoken words. In an experiment, participants were asked to repeat an auditorily presented word aloud as quickly and accurately as possible. Results revealed that the response times were faster for correctly stressed word (e.g., *ANgry*) than for incorrectly stressed word (e.g., *anGRY*). In a lexical decision task, consistent results were repeated that correctly stressed words were classified faster than incorrectly stressed words. This work provides evidence across several experimental tasks for the important role of word stress in spoken word recognition. However, Small et al., (1988) failed to find the influence of the word stress in word recognition for homographs (e.g., *CONvert* vs. *conVERT*). The mis-stressing effects (slowing down word recognition) were only observed when listeners heard a misstressed word that resulted in a nonsense word; lexical access then appeared to be disrupted. In contrast, Dutch experiments on the importance of the word stress in word recognition showed consistent results. Koster & Cutler (1997) showed that for Dutch disyllabic words the effects of mis-stressing were as strong as the effects of segmental mispronunciation.

d) A number of studies (Lindfield et al., 1999; Wingfield et al., 2000; Cutler & Otake, 1999) have provided evidence that gated fragments[26] are identified faster when stress information is included in the signal than when it is not. Using the gating paradigm (Grosjean, 1980), Cutler & Otake (1999) addressed the question of whether pitch-accent information may be exploited in the process of recognizing spoken words in Tokyo Japanese. In the gating experiment, participants heard initial fragments of Japanese words then guessed what the words were, the confidence rating of the guess was also recorded. Their guesses overwhelmingly had the same initial accent structure as the gated word even when only the beginning syllable of the stimulus, (e.g., unstressed fragment *na* would be more guessed a word with same stress pattern like *naGASHI* rather than a word with different stress pattern such as *NAgasa*) was presented. In addition, participants were more confident in guesses with the same initial accent structure as the stimulus than in guesses with different accent.

e) The cross-modal priming paradigm has been largely used in psycholinguistic studies to understand the processing of word stress. In the cross-modal priming task, a participant is presented with an auditory stimulus (e.g., fragments of words, words or sentences); at some point during the stimulus, a visual target (a string of letters) appears on a screen, and the participant must decide whether it is a word. If the target has the same prosodic patterning as the prime, the response time on the target will be shorter, while if the prosodic patterning is mismatched, the response time will be longer. Using the cross-modal priming paradigm, Soto-Faraco et al. (2001) investigated the role of suprasegmental information in the activation of Spanish spoken words. In the experiment task, participants heard neutral sentences ending with word fragments (e.g., in the matching condition: *PRINci-* for the target *PRINcipe*; in the mismatching condition: *prinCI-* from the word prinCIpio, for the target *PRINcipe*; in the control condition: the fragment *mos* from the word *mosQUIto*, for the target *PRINcipe*) and

[26] The gated fragments refer to the fragments of a stimulus in a gating paradigm. The gating paradigm (Grosjean, 1980) involves the repeated presentation of a spoken stimulus (e.g., a word) such that its duration from onset is increased with each successive presentation. This is done utile the entire stimulus has been presented. After each presentation, subjects are asked to guess what are the stimulus is and to give the confidence rating of the guess

made lexical decisions on letter strings presented at fragment offset. Responses were compared for fragment primes that matched the spoken form (i.e., word stress pattern) of the initial portion of target words, versus control primes. Results showed facilitation effect for matching priming relative to the control condition suggesting the processing of the word stress in the Spanish spoken words recognition. Tagliapietra and Tabossi (2005) extended the results to Italian. Cutler (1986) presented in a cross-modal priming either of the two members of a minimal stress pair, for instance, either *FORbear* or *forBEAR*. The author found that whichever member of the stress pair had been heard, participant' responses to associates of both members of the pair were facilitated in comparison to control words. Cutler argued that the suprasegmental differences between, for instance, *FORbear* and *forBEAR* were ineffective in constraining lexical activation, so that for English listeners forbear was effectively a homophone.

f) Cutler and Norris (1988) proposed a model of speech segmentation, according to which the occurrence of a strong syllable triggers segmentation of the speech signal, whereas occurrence of a weak syllable does not trigger segmentation. They investigated the detection of a target word embedded in nonsense matrices with differing metrical structures. For example, the word *mint* was embedded either in *mintayf* /mintef/, in which the second syllable was strong, or *mintef* /mintəf/, in which the second syllable contained a schwa (i.e., was weak). Their experiments showed that listeners detected words more slowly when the disyllable had two strong syllables than when it had a strong and a weak syllable; The authors argued that this result is an effect of segmentation: When the second syllable is strong, it is segmented from the first syllable, and successful detection of the embedded word therefore requires assembly of speech material across a segmentation position. They proposed therefore a Metrical Segmentation Strategy (MSS) for speech segmentation in which the boundaries are postulated at the onset of strong syllables. The model can explain the findings that participants prefer identifying a stimulus as two words when it has a weak-strong stress pattern and as one word when it has a strong-weak stress pattern (Heuven, 1985; Taft, 1984).

g) With regard to the role of prosodic information in compound words processing, Isel et al. (2003) manipulated the length of the left syllable of German noun-noun compound words by splicing the two constituents that were spoken in isolation, yielding truncated compound words. Results revealed a 12 *ms* facilitatory priming effect for targets that are semantically related to the left constituents, indicating that left constituents representing the prosodic structure of individual words were activated. The authors concluded that the length of the left lexical morphemes of compound words is a determining factor for the activation of the morphological decomposition processing during processing of spoken German compounds. Syllable length is thus used by the language processing system in order to anticipate the morphological structure of the words. In addition to syllable length, Koester et al. (2004) claimed that syllable F0 contour was involved in the spoken compound words processing. As the importance of the prosodic information in compound words processing, a Prosody-Assisted Processing (PAP; Isel et al., 2003) model was therefore proposed postulating an early implication of prosodic information in compound words recognition. More detail about PAP model will be discussed in detail in section 4.2.3.

3.4.2. Neurophysiological evidence

As behavioural reaction times that constitute the by-product of different processes, it is difficult to differentiate these different processes in the perception of prosodic patterning. By contrary, Event-Related-Potentials (ERPs) offer the opportunity to track millisecond by millisecond the time course of language processes and then suitable (1) to investigate online the timing of processes that are intrinsically rapid as it is the case for the processing of stress, and (2) to disentangle the study of cognitive processes, which in some cases can occur in parallel.

Wang, Friedman, Ritter, and Bersick (2005) examined whether prosodically salient (i.e., stressed syllables) information serves as a cue to capture attention in speech sound analysis. In an *oddball* paradigm, two types of nonsense disyllables *BAga* and *baGA* (differing in stress pattern) were used as standard stimulus in two separate experiments session. For each standard stimulus, two types of deviants were created by replacing the voiced consonant by its corresponding unvoiced consonant (i.e., /*b*/ → /*p*/ or /*g*/ → /*k*/), thus resulting in the four following types of deviants: (1) the deviant occurring on the stressed initial syllable (i.e., ***P**Aga*), (2) the deviant on the unstressed initial syllable (i.e., ***p**aGA*), (3) the deviant on the stressed noninitial syllable (i.e., *ba**K**A*), and (4) the deviant on the unstressed noninitial syllable (i.e., *BA**k**a*). MMNs[27] were observed for all four types of deviants. In contrast, the P3a[28] was only seen when the deviance occurred on stressed syllables, but not when deviance occurred on an unstressed syllable, irrespective of its temporal position (i.e., at first or second syllable) in the disyllable. The findings suggested that salient syllabic prosody (i.e., stressed syllables) serves as a cue to capture involuntarily attention during speech sound perception.

Böcker, Bastiaansen, Vroomen, Brunia, and Gelder (1999) investigated the perception of metrical stress[29] in disyllabic Dutch words that start with either a weak or a strong syllable (i.e., weak-initial word or strong-initial word, respectively) and sought to associate metrical processing to ERP component. In their experiment, each trial consisted of a sequence of four disyllabic words auditory presented in succession. The first three words were either all strong-initial word (i.e., stressed on the first syllable) or all weak-initial word (i.e., stressed on the second syllable) and constituted the standard stimuli. The authors manipulated the metrical congruency by varying the stress pattern of the fourth word. In the metrically expected condition, the fourth word had a similar stress pattern as the previous three words. In the metrically unexpected condition, the fourth word had a different stress pattern. Participants were instructed to perform a stress discrimination task and a passive listing task. Böcker et al.

---

[27] The MMN is a frontocentral negative ERP component peaking at 100-250 ms invoked by infrequently, presented stimuli in an *oddball* paradigm (e.g., Näätänen, Gaillard, & Mäntysalo, 1978). The MMN has been argued to reflect a neural mismatch between the deviant stimulus and a representation of the standard stimulus held in memory (e.g., Näätänen, 1992). It is considered a neurophysiological marker of discrimination. In addition, because the MMN can be evoked in the absence of voluntarily directed attention to the eliciting event, mental processes manifested by the MMN are believed to be pre-attentive (e.g., Näätänen et al., 1978).

[28] The P3a is a frontocentrally distributed positive-going component whose peak occurs as early as 280 ms post-stimulus. It was first described by Squires, Squires, and Hillyard (1975), who showed it to be elicited by infrequent, unpredictable auditory stimuli regardless of whether the subject was paying attention to the auditory sequence or ignoring the tones while silently reading a book. The P3a has been thus characterized as a marker of involuntary attentional capture (e.g., Friedman, Cycowicz & Gaeta, 2001).

[29] Metrical stress is a conceptualization of stress based on the vowel quality in a syllable. Strong syllables are those that contain full vowels /e.g., ɪ, ʏ, ɛ, iː, uː, oː/. Weak syllables are those that contain reduced vowels, always a schwa /ə/ in Dutch.

(1999) found that concerning the process of lexical recognition, infrequent weak-initial disyllabic Dutch words (which constitute 12% of the Dutch lexicon) in comprising with strong-initial words elicited a larger left-frontal distributed negativity around 325 ms (named N325). Moreover, results showed that the N325 was influenced by task, as the amplitude de N325 was attenuated in passive listening task as compared to stress discrimination task, which suggested the endogenous nature of the effect. The authors proposed that the N325 reflects the extraction of metrical stress from the acoustic signal thus an ERP correlate of metrical stress perception. In addition to the N325, a larger P200, amplitude maximum at frontocentral region in the time window 171-271 ms, was evoked by infrequent weak-strong words than by strong-weak words in the passive listening task. The P200 component is an exogenous component assumed to reflect perceptual processing (Hillyard & Picton, 1987). The P200 has been found sensitive to acoustic properties (Marie, Magne, & Besson, 2011; Friedrich et al., 2001; Cunillera, Gomila, & Rodríguez-Fornells, 2008). Marie et al. (2011) manipulated the metric structure of French words by lengthening the penultimate syllable of trisyllabic words, and found metrically incongruous words elicited larger P200 components than metrically congruous words. Furthermore, the P200 effect was larger for musicians than for non-musicians. It is proposed that the amplitude enhancement of the P200 may reflect the automatic processing of the temporal attributes of words and, specifically, of the words' syllabic structure. The P200 in response to stress pattern information has been also observed with German words, Friedrich et al. (2001) found that German infrequent weak-strong words elicited a larger P200 than frequent strong-weak words. Magne, Gordon, and Midha (2010) applied the same paradigm used in Böcker et al. (1999) with English disyllabic words; moreover, the words were presented in visual modality. They found that metrically incongruent target words elicited larger negative components than metrically congruent target words in frontal and centro-frontal regions. The results suggest an online metrical stress processing in English even when reading. The finding reinforces the idea proposed by Fodor that implicit prosody is processed in silent reading, which means a default prosodic contour is projected onto the stimulus (Fodor, 2002). However, Friedrich, Alter and Kotz (2001) using the same paradigm but with German stimuli failed to replicate the N325 modulation for disyllabic German words neither with a weak-strong nor a strong-weak patterns. Their results showed that German infrequent weak-strong words elicited a larger P2 than frequent strong-weak words. The authors argued that unstressed pitch in the first syllable of weak-strong words is detected as a mismatch in relation to the more frequent initially stressed stress pattern in German. Friedrich, Kotz, Friederici, and Gunter (2004a, 2004b) showed in a cross-modal word fragment priming that targets with a stress pattern that matched the pitch contour of the primes (e.g., in the prime - target pair *RE* - *REgel*) evoked a reduced P350 (positive-going ERP deflection with a peak at 350 ms) than targets with a stress pattern that did not match the pitch of the primes (e.g., in the prime - target pair *re* - *REgel*). Reduced P350 can be interpreted as reflecting facilitated lexical identification. The authors conclude that the P350 effect indicates that pitch is used for lexical identification in spoken word recognition. Isel, Alter, and Friederici (2005) showed that mis-stressing of initially stressed German compounds (*weinBERG* instead of *WEINberg*) evoked an early sustained prosody-related bilateral anterior negativity starting approximately 100 ms after the onset of the initial constituent.

Investigating the role of metrical stress in syntactic processing, Schmidt-Kassow and Kotz (2009) examined the processing of German words during online auditory sentence comprehension. The metrical congruency and the syntactic congruency of the target word were systematically manipulated, resulting in four experimental conditions: metrically incorrect, syntactically incorrect, doubly incorrect and correct (control) condition. Participants were asked to judge syntactic correctness or metric homogeneity in two different sessions. Like Isel et al. (2005), results of the study of Schmidt-Kassow and Kotz (2009 showed a bilateral anterior negativity in response to metric violations in the time window 200-450 ms. Moreover, the authors also found the anterior negativity for the combined metric/syntactic violations. In both conditions (metric violations and metric-syntactic double violations) the anterior negativity deflected in the time window 200-450 ms earlier than the anterior negativity (i.e., LAN[30]) in the time window 500-650 ms elicited by syntactic violations alone,

[30] LAN is a negative going electrical potential presented between 300 and 500 ms post-stimulus onset. It is assumed reflecting morpho-syntactic processes (Friederici, Hahne, & Mecklinger, 1996).

and was all followed by a late posterior positivity, the P600, a late positivity peaking around 600 ms post-stimulus associated with violations of tense, agreement, and phrase structure (e.g., Hagoort, Brown, & Groothusen, 1993), as well as difficult syntactic integration (Kaan, Harris, Gibson, & Holcomb, 2000). Schmidt-Kassow and Kotz (2009) suggested that metric information is processed early before the syntactic processing; moreover, metric processing could be involved in a later integrational stage for the reanalysis (reflected by the P600) of a syntactically violated sentence.

McCauley, Hestvik, and Vogel (2013) investigated the perception and bias in the processing of compound versus phrasal stress (e.g., *GREENhouse* vs. *green HOUSE*) in online sentence processing. In a picture/word matching task, an image was presented (e.g., a green-colored house) establishing context and was followed by an utterance featuring the test item with either the congruent (phrase: *green HOUSE*) or incongruent (compound: *GREENhouse*) stress pattern. Participants were asked to judge whether the heard word described correctly or not the presented picture. Results showed that both compound and phrasal stress elicited a sustained left lateralized negativity (400-1000 *ms*) in incongruent condition. Moreover, incongruent compound stress elicited centro-parietal negativity (N400), while incongruent phrasal stress elicited a late posterior positivity (P600). The findings suggested the online use of stress information in lexical-semantic processing reflected by N400 and interactions between prosodic and syntactic information indexed by P600.

Isel and Friederici (2006) investigated constituents' prosodic role in the processing of German compounds. The authors recorded ERP while native speaker of German listened noun phrases (e.g. *der Weinberg* (the vineyard)) composed of a definite article (e.g. *der* (the)) followed by a transparent compound word composed of two monosyllabic lexical morphemes (e.g. *Wein* (wine) and *Berg* (mountain)). The noun phrases were produced with a natural prosodic contour without to artificially introduce a salient word's boundary between the definite article and the first constituent. Results revealed a biphasic N400/P600 ERP pattern in response to the processing of grammatical gender violation between the head of the compound words (i.e., the right-most constituent morpheme) and the preceding definite article. In contrast, no ERP effect was found in response to the processing of similar anomalies between the first lexical morpheme and the preceding definite article. These findings suggested that when the prosodic structure of the compound words is natural, early morphosyntactic decomposition does not occur.

Chapter 4

Psycholinguistic models of compound words recognition

4. Psycholinguistic models of compound words recognition

4.1. Full-parsing, Full-listing and Dual-route models

There is a longstanding debate how morphologically complex words such as compounds are processed and represented in the mental lexicon. Four decades ago, Taft and Forster (1976) used compound words in a lexical decision task to show the decomposition processes of compounds. The response latencies were influenced by the lexical status of constituents: The nonword **footmilge* (made up of a word and a nonword, respectively) took longer to classify as a nonword than did **trowbreak* (made up of a nonword and a word, respectively). Since then, numerous psycholinguistic researches have been carried out on compounds processing and three main models were proposed to account for the representation and processing of compound words. *Full-parsing or decompositional models* (e.g. Taft & Forster, 1976; Libben, Derwing, & de Almeida 1999; McKinnon, Allen, & Osterhout, 2003; Taft, 2004) and *full-listing or non-decompositional models* (e.g. Butterworth, 1983; Bybee, 1995) are two major competing models. The third class is constituted of hybrid models postulating both a full-parsing and full-listing depending of some linguistic factors like lexical frequency, semantic transparency, and prosodic pattern inherent to the compound words.

According to *full-parsing models,* constituent morphemes are lexical storage in the mental lexicon and the processing of compound words involved a prelexical decomposition of compounds into their constituent morphemes. If this were true, it would have advantages in terms of storage efficiency, as only mono-morphemes are stored in the mental lexicon no additional whole-form compound representation is needed. However, full-parsing models would create the need to decompose and recompose compound words each time they are perceived and produced; the decomposition and recomposition operations slow down the word recognition course. It would seem that such a system would have poor computational efficiency.

Full-listing models proposed a completely opposite hypothesis, claiming that all words are stored as full forms in the mental lexicon just like for monomorphemic words.

Consequently, these models exclude morphological decomposition thus morphological information is not utilized in compound words recognition. Although full-listing representation would have advantages in terms of computational efficiency, however a full-listing model alone has been proved to be very limited, because it cannot account for the productivity of compound words and the comprehension of novel compounds[31] that are not yet lexicalized in the mental lexicon. Moreover, such models would be highly consuming in terms of memory resources.

More recently, a third class of *hybrid* models, the so-called "*dual route*" models has been formulated, which is an intermediate theoretical approach between the other two accounts. These models postulate two possible routes for processing compound words: (1) a decompositional route (or parsing route) in which the compound is decomposed into its morphological components and (2) a direct route in which the compound is accessed by a lexical entry based on stored full-form representation in the mental lexicon. There are two most prominent examples of dual route models: the Augmented Addressed Morphology Model (AAM, e.g., Caramazza, Laudanna, & Romani, 1988; Chialant & Caramazza, 1995, among others), and the Morphological Race Model (MRM, e.g., Schreuder & Baayen, 1995; Baayen and Schreuder, 1999). According to AAM, the direct route is preferred for all known words, and familiar compound words are normally processed by the direct route, while unfamiliar compound words, such as novel or extremely rare compounds that are morphologically regular are decomposed and processed by the decompositional route. By contrast, the MRM posits that the direct route and the decompositional route operate in parallel. The outcome of the *race* between the two routes will be affected by a number of factors, including lexical frequency, semantic transparency and prosodic structure (Hyönä & Pollatsek, 1998; Libben et al., 2003; Isel et al., 2003).

[31] Novel compounds refer to compounds that are less familiar and their meanings are not yet established. To interpret a novel compound, one must therefore access the concepts denoted by its constituents and select a relation that link them together (Gagné, 2002).

4.2. Factors that influence compound words processing

4.2.1. The role of semantic transparency

One factor that may influence compound words representation and processing is *semantic transparency*. The meaning of transparent compounds can be understood from the combination of the meanings of their constituents, for example, a *snowball* is a ball made of snow. In contrast, the meaning of opaque compounds cannot be derived from their constituents, for example, *cocktail* is neither a cock nor a tail, and thus may require dedicated whole-form lexical storage in the mental lexicon.

Sandra (1990), using semantic priming, showed no priming effects for semantically opaque compounds (e.g. *buttercup*) and pseudo-compounds (e.g. *boycott*). Constituents are accessed only for transparent compounds (e.g. *teaspoon*). In a cross-modal semantic priming study (Isel et al., 2003), the authors showed that the first constituents of German spoken compounds primed visually presented targets only when the second constituent was transparent, but not when it was opaque, suggesting that activation of both constituents is dependent on the transparency of the second constituent (i.e., the head of compound word). The authors proposed a hierarchical processing of transparent head in German compound words.

The role of semantic transparency in compound words processing has been studied in real-time, using event-related potentials (ERPs). In a recent study, MacGregor and Shtyrov (2013) investigated the neuro-cognitive processing of transparent vs. opaque compounds using oddball design. The authors presented monomorphemic standard stimuli, (e.g., *work*) as standard stimuli mixing with deviant stimulus including transparent (e.g., *homework*) or opaque (e.g., *framework*) compound word, or a meaningless pseudo-compound (e.g., **houndwork*, **grousework*). Whole-form frequency of the compounds was systematically varied over both opaque and transparent groups of stimuli. EEG data revealed that for opaque compounds, the MMN is larger for high frequency than low frequency compounds. This result is in line with previous findings of word frequency effects on the MMN amplitude for monomorphemic words, thus suggesting that opaque compounds is accessed in the manner of whole-form just like monomorphemic words. Moreover, the frequency effects are absent for transparent compounds. The lack of frequency effects for transparent compounds is clear evidence against a purely full-listing account of transparent compound. A recent neuroimaging study (Brooks & de Garcia, 2015) used magnetoencephalography (MEG) to investigate the neural bases of the composition process in English compound word recognition. MEG data were acquired while participants performed a word naming task in which three word types, transparent compounds (e.g., *roadside*), opaque compounds (e.g., *butterfly*), and morphologically simple words (e.g., *brothel*) were contrasted in a constituent priming paradigm where the target compound was primed by one of its constituent morphemes (e.g., *road* - *roadside*). An analysis of the associated MEG activity revealed that only transparent compounds showed increased activity from 250 to 470 ms in a region of interest implicated in morphological composition, the Left Anterior Temporal Lobe (LATL)[32]. However, this brain area is not sensitive to opaque compounds whose morphemes do not share a semantic relationship. Thus, results suggest that semantics plays a role in combining the meanings of morphemes.

4.2.2. The role of frequency

A second potentially important factor in compound words processing is lexical frequency: compounds that are more frequent are more likely to benefit from readily available whole-form storage, whereas less frequently used compounds might have to be processed through a combinatorial mechanism. In addition to the whole word frequency, there is ample evidence that the frequency of compound constituents play a role during visual word recognition (e.g., Juhasz & Rayner, 2003; Hyönä & Pollatsek, 1998; Duñabeiti, Perea, & Carreiras, 2007, 2008; Bronk, Zwitserlood, & Bölte, 2013). Effects of constituent frequency are interpreted in favor of decomposition, or for the existence of two routes to compound word recognition. To address issues of full-listing and full-parsing in compounds processing, Bronk

[32] Neuroimaging studies on the binding mechanism in sentence building (Friederici et al., 2000) and on the composition of noun phrases (Bemis & Pylkkänen, 2011) reveal an implication of the left Anterior Temporal Lobe in the composition of words into phrases.

et al. (2013) manipulated the constituents frequency of German noun-noun compounds and maintained full-form frequency matched in visual lexical decision experiments. The idea is simple: if compounds are treated as morphologically simple words during word recognition, they should be recognized with a similar latency as monomorphemic words, when matched in overall frequency; if, however, morphological decomposition takes place upon lexical access, compounds should be accessed faster/slower, because of their more/less frequent constituents. Results revealed that reactions were faster to compounds with high-frequency constituents than to compounds with low-frequency constituents and the latter did not differ from the monomorphemic words. The difference in reaction time due to constituent frequency can only be explained by access to the constituents, and thus by decomposition.

Eye-tracking data also reveal the role of frequency in compound words processing. Hyönä and Pollatsek (1998) showed frequency effects in the reading of Finnish noun-noun compounds: the gaze duration on the compound word was 100 ms shorter when there was a frequent initial morpheme. The findings support a morphological decomposition account. Andrews et al., (2004) found that these results extend to English compounds. Pollatsek et al. (2000) found that the frequency of the whole compound had a similar influence on gaze duration and influenced eye movements at least as rapidly as did the frequency of the second constituent. These results suggested that the identification of these compound words involves parallel processing of both morphological constituents and whole-word representations.

4.2.3. The role of prosody and Prosody-Assisted Processing (PAP) model

The prosodic structure of compound constituents is another important factor in compound words processing. It has been showed that prosodic features such as duration (Isel et al., 2003) and fundamental frequency contour (Koester et al., 2004) of compound constituents bias either the direct or the decompositional route during compound word processing.

Isel et al. (2003) investigated German noun-noun compound words by monitoring semantic priming effects of the left constituent. In their study, the authors investigated the role of the prosodic structure in the morphological processing by manipulating the acoustic structure of the left constituent of compounds. Results revealed that the syllable length of initial morphemes is able to assist the processing system in activating a decompositional route at the offset of these morphemes. The authors therefore proposed a Prosody-Assisted Processing (PAP) model for compound words access. The PAP model is a dual-route models postulating that a decompositional route and a direct route work in parallel. The central assumption on which the model relies is that the length of the first lexical morphemes of compound words is a determining factor for setting off/triggering a decompositional route. At the onset of the speech input, a direct route is activated by default. Simultaneously, a prosodic analysis of the left-most morpheme is automatically carried out. The output of the prosodic analysis is crucial with respect to the activation of the decompositional route. If this output informs the processor that the left-most morpheme belongs to a compound word (i.e. short syllable), then the processing system activates a decompositional route. On the contrary, if this prosodic analysis indicates a monomorphemic word (i.e. long syllable), the processing system continues to match the acoustic-phonetic input with the word detectors at the word level, until the appropriate detector reaches the threshold for best match. In this case, the processing system has set a monomorphemic route configuration. Because of the sequential involvement of the two routes, the PAP model is considered as a cascading parallel model. Furthermore, the PAP model postulates that the syntactic elements of a compound word (i.e., head and modifier) are hierarchically activated, with primacy of the head. Therefore, the PAP model is a discontinuous hierarchical model that posits a discontinuous hierarchical lexical access of compound words, at least when the head noun is transparent. Moreover, the PAP model accounts for the constituent integration process by assuming a right-to-left (the direction of modification is leftwards, i.e. the right-most constituent is modified by the left-most one) semantic composition. The model assumes that the integration of the head and the modifier results from on a semantic binding.

The PAP model is based on studies on German compounds where the acoustic information for the extraction of the prosodic structure of a compound word is available as soon as its first constituent is presented. However, for some languages such as Chinese, the acoustic information may not be sufficient in the first constituent for determining the prosodic structure of a compound word. Indeed, in Chinese, in some compound words the crucial prosodic information is provided in the last constituent. As introduced in Chapter 3, Chinese neutral tone compound is a case. The prosodic structure of a neutral tone compound word is primarily determined by the right-most constituent. The current PAP model assumes a left-to-right analysis of prosodic information and then has its limit to account the processing of Chinese neutral tone compound whose major acoustic information is located at the right-most constituent. For the purpose, by using Chinese neutral tone compounds, we conducted the Experiment 1 & 2 to investigate the possible right-to-left retroactive processing of prosodic information leading to a complement of the sequential left-to-right spoken-word processing.

Chapter 5

Event-related brain potentials

5. Event-related brain potentials (ERPs)

Spoken-word recognition requires that different sources of information such as segmental and suprasegmental phonology as well as morphology, syntactic, semantic and pragmatic must be accessed and coordinated within milliseconds. Due to the precise temporal resolution of electrophysiological technique, the event-related potential (ERP) recordings offer the opportunity to track millisecond by millisecond the time course of language processes and then are suitable (1) to investigate the time course of the implication of the stress spoken-word recognition, and (2) to disentangle different cognitive processes, which in some cases can occur in parallel.

5.1. Methodological issue

In 1929, Hans Berger firstly revealed that one could measure the electrical activity of the human brain by placing an electrode on the scalp, amplifying the signal, and plotting the changes in voltage over time (Berger, 1929). This electrical activity is called the electroencephalogram (EEG). He found that the EEG could be influenced by external events. Since then, the EEG became to be a useful source in recording brain activity. However, it was very difficult to isolate individual neuro-cognitive processes by using raw EEG data because raw EEG reflects thousands of simultaneously ongoing brain processes as well as other bio-signals (e.g., electromyography) and noise (e.g., electromagnetic interference). This problem has been solved by applying a simple averaging procedure on the signal, which quickly became the primary tool of the cognitive neuroscientist (Davis, 1964; Cooper, Winter, Crow, & Walter, 1965; Sutton, Braren, Zubin, & John, 1965; Donchin & Cohen, 1967). The procedure consists in averaging together the waveforms, *time-locked* to an event. After that, non-event-related activity such as random brain activity and noise cancel out each other, however event-related signal is reinforced. The averaged EEG is called the event-related potential (ERP).

5.1.1. EEG recordings and electrodes positions

The EEG is measured as a difference potential between two electrodes. The ongoing difference potentials are first amplified and the continuous signal is then digitally sampled. High sampling rates make the recording data precise, but raise the size of the recording. A sampling rate at 1000 Hz (1000 samples per second) has been often used in cognitive studies, as it provides high precision in data, and a relative small data size in favor of data processing. One important point in EEG recording is the site of reference (or 'inactive') electrodes, relative to which the electrical activity in all other electrodes is measured. The choice of reference will determine the topography, polarity and amplitude of the recordings. The left/right earlobe or both of them are frequently used as reference electrode. Alternatively, the EEG may be recorded with any scalp electrode as a reference, and then the average reference is computed as a mean of all electrodes. It avoids all kind of asymmetry and makes the EEG recorded in various laboratories comparable. In the experiments of the present work, we used the FCz as the recording reference, and then the average reference was computed as a mean of all electrodes in the off-line analyses. Before the recording of the EEG data, the conductivity of electrodes connection should be tested by measuring the electrode impedance. The impedance between electrodes should be low, e.g., less than 5 kΩ, for a precise recording. High impedance causes a loss of band frequencies in the recording (Picton, Lins, & Scherg, 1995). In order to ensure comparability of electrode positions across measurements, guidelines for the positioning and nomenclature of electrodes have been developed. The international 10-20 system of Jasper (1958) defines 19 electrode positions that are situated in either 10% or 20% of the total front-back or right-left distance of the skull. Each site has a letter to identify the brain region and a number to identify the hemisphere location. The letters F, T, C, P and O stand for frontal, temporal, central, parietal, and occipital lobes, respectively. A 'z' (zero) is for electrode placed on the midline. Odd numbers (1, 3, 5, 7) refer to electrode positions on the left hemisphere, whereas even numbers (2, 4, 6, 8) refer to those on the right hemisphere. As many studies recorded the EEG from a greater number of electrode sites, the system was expanded in the guidelines of the *American Electroencephalographic Society*

(Sharbrough, Chatrian, Lesser, Lüders, Nuwer, & Picton, 1991). From the electrode positions specified in these guidelines, 64 positions were used in the experiments of the present work.

5.1.2. From the continuous signal to ERPs

The ERP at the scalp (5-10 µV) is substantially smaller in amplitude than the background EEG (50-100 µV). Due to the low signal-to-noise ratio, it is necessary to average EEG samples over many segments with 'similar' stimuli. The averaging procedure leads to an increase of the event related part of the signal and to a decrease of the part of the signal, which is due to random variation (noise or random brain activity). Figure 5.1 illustrates the averaging procedure on a repeated event (stimulus).

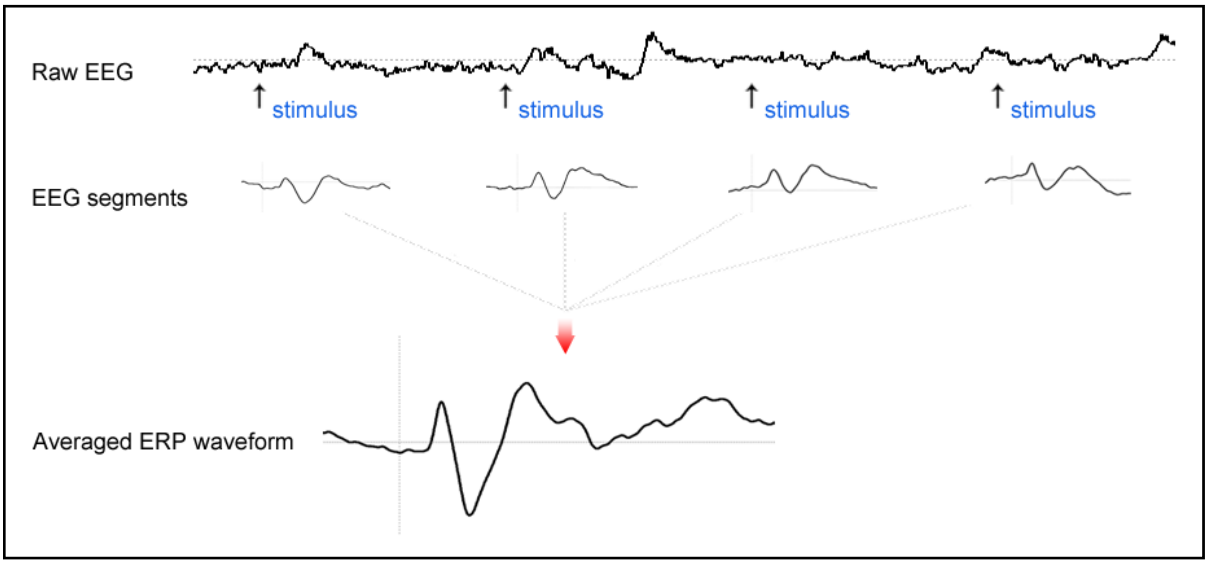


Figure 5.1. ERP averaging procedure.

The positive and negative deflections in the ERP vary in amplitude, topography, latency, and polarity. Characteristically, ERP amplitudes vary between 1 and 30 µV (Birmbaumer & Schmidt, 1990). In order to avoid pre-stimulus influence, the amplitude of ERPs is always measured in relation to a baseline time window (normally pre-stimulus, e.g., between -200 and 0 ms), in which the average amplitude is corrected to zero.

The identification of ERP component is not trivial and is usually done on the basis of a combination of properties related to the physiological source, such as polarity, latency and topography, and of functional characteristics, such as sensitivity to specific experimental manipulations (Coles & Rugg, 1995). Qualitative variations (i.e. variations in polarity and topography) in ERPs of different experimental conditions can stem either from different neural generators or from differences with regard to the relative contribution of the generators (Rugg, 1999). ERPs can therefore help dissociate cognitive operations involved in different conditions on the neural levels. However, it is important to keep in mind that similar ERP components do not allow for the conclusion that neural processes are identical. Similarities could also be found to the relative intensity of the EEG to a large part of brain activity. Several ERP components have been shown to be sensitive to different types of linguistic information. They provide information that can help disentangle different processes in language comprehension. In the following section, ERPs related to language processing will be introduced, with a specific focus on components related to prosodic, morpho-syntactic and semantic processes.

5.2. Some ERP components in language processing

5.2.1. N400

Kutas and Hillyard (1980) first reported the N400 in an experiment contrasting semantically congruent with semantically incongruent sentence. In a sentence comprehension task, semantically incongruent sentence-final word (e.g., ‘drink’ in ‘*The pizza was too hot to drink.’) elicited a larger negativity peaking at about 400 ms post-stimulus onset than congruent sentence-final word (e.g., ‘eat’ in ‘The pizza was too hot to eat.’). Since then, a number of studies replicated the N400 evoked by semantic anomalies in sentence context in different languages such as Dutch, French, German and Chinese as well as in different sensorial modalities (i.e., auditory and visually; for reviews see, Isel, Hahne, Maess, & Friederici, 2007 but also Kutas & Federmeier, 2009; Friederici, 2011). In addition to sentence context, the N400 is in response to words presented in lists or pairs. The following pattern of results has emerged: words that are unrepeated and semantically unrelated to previous words elicit larger N400 than repeated or semantically related ones (Radeau, Besson, Fonteneau, & Castro, 1998; Bentin, Kutas, & Hillyard, 1993). The typical paradigm is the semantic priming in which words are presented for identification in the context of a prime stimulus that either semantically or associatively related or unrelated to the target word. The robust behavioural finding is that target recognition is faster when the prime shares a semantic or associative relation with the target word than when it does not, namely semantic priming effect (see Neely, 1991 for a review). The semantic priming effect has been associated with the N400 (Bentin, McCarthy, & Wood, 1985; Kutas & Van Petten, 1994; Kutas & Federmeier, 2011). A number of studies have reported reduced N400 amplitudes to target words following semantically related as compared to unrelated primes. With respect to the cognitive processes the N400 reflects, a debate exists on whether it reflects an automatic process of lexical access, or a controlled process of lexical integration. Studies showed that the semantic priming N400 was not a purely automatic processes. For example, Brown and Hagoort (1993) failed to find the N400 priming effect using masking priming (the masking technique prevents a stimulus from reaching conscious perception, and as such it is claimed that masking by and large rules out controlled processing), while a reaction time priming effect was still observed. The N400 has been shown to be modulated by attentional instructions: Holcomb (1988) found the amplitude of the N400 was larger under instructions to attend to the prime than it does not. Although the N400 effect was enhanced by attention, Kutas and Hillyard (1989) showed that even in tasks in which it was not needed to process the meaning of words the N400 effect appeared.

5.2.2. P200, N325 and Closure Positive Shift (CPS)

An exogenous P200 component and an endogenous component N325 have been reported to be in response to stress pattern processing. Böcker et al. (1999) investigated the perception of metrical stress in disyllabic Dutch words that start with either a weak or a strong syllable (i.e., weak-initial word or strong-initial word, respectively) and sought to associate metrical processing to ERP component. They found that infrequent weak-initial disyllabic Dutch words in comparison with strong-initial words elicited a larger left-frontal distributed negativity around 325 ms. Moreover, results showed that the N325 was influenced by task, as the amplitude de N325 was attenuated in passive listening task as compared to stress discrimination task, which suggested the endogenous nature of the effect. The authors proposed that the N325 reflects the extraction of metrical stress from the acoustic signal thus an ERP correlate of metrical stress perception. In addition to the N325, a larger P200, amplitude maximum at frontocentral region in the time window 171-271 ms, was evoked by infrequent weak-strong words than by strong-weak words in the passive listening task. The P200 component is an exogenous component assumed to reflect perceptual processing (Hillyard & Picton, 1987). The P200 has been found sensitive to acoustic properties (Marie et al., 2011; Friedrich et al., 2001; Cunillera et al., 2008). Marie et al. (2011) manipulated the metric structure of French words by lengthening the penultimate syllable of trisyllabic words, and found metrically incongruous words elicited larger P200 components than metrically congruous words. Furthermore, the P200 effect was larger for musicians than for non-musicians. It is proposed that the amplitude enhancement of the P200 may reflect the automatic processing of the temporal attributes of words and, specifically, of the words' syllabic structure. The P200 in response to stress pattern information has been also observed with German words, Friedrich et al. (2001) found that German infrequent weak-strong words elicited a larger P200 than frequent strong-weak words. With respect to the processing of prosody information at sentence level, Steinhauer, Alter, and Friederici (1999) were the first to show that processing of major prosodic boundaries, i.e. intonational phrase boundaries (IPhs) is related to a specific electrical brain response that the authors called 'Closure positive Shift' (CPS). The CPS is centro-parietal positive-going waveform lasting about 500-1000 milliseconds. Critically, the CPS reflects lengthening of the syllable preceding the prosodic boundary as well as a rise of its F0.

5.2.3. MMN

The MMN is a frontocentral negative ERP component peaking at 100-250 ms invoked by infrequently, presented stimuli in an oddball paradigm (Näätänen et al., 1978). The MMN has been considered reflecting a neural mismatch between the deviant stimulus and a representation of the standard stimulus held in memory (e.g., Näätänen, 1992). Note that the MMN can be evoked in the absence of voluntarily directed attention to the eliciting event, mental processes manifested by the MMN are believed to be pre-attentive (e.g., Näätänen et al., 1978). The Mismatch Negativity (MMN) component has been shown to be particularly valuable to study the auditory language processing. As its exogenous nature, the MMN can be used to examine how the human brain processes linguistic information without requiring participants to make conscious decisions about the speech stimuli, and thus avoiding the interplay of extra-linguistic processes. In addition to changes in short-term representations of auditory stimuli, the MMN can also be elicited by changes in long-term representations of linguistic information (see Pulvermüller & Shtyrov, 2006 for a review). That is the case for stress pattern. In an oddball paradigm, Honbolygó and Csépe (2013) investigated the MMN in response to illegal stress pattern in Hungarian. The authors manipulated the position of the stress in disyllabic pseudowords in order to create illegal stress pattern, i.e., weak-strong, given that in Hungarian, stress in disyllabic words is always on the first syllable. Results revealed that illegal stress pattern evoked two consecutive MMNs only in the order in which the pseudoword with legal stress pattern was served as standard stimuli and the deviant stimuli was the pseudoword with illegal stress pattern. In contrast, in the reverse order where the deviant stimulus was the legal pseudoword, the MMN was not observed. The authors argued that illegal pseudoword mismatched both the short-term (acoustic parameter in single syllable) and long-term (stress pattern in syllables relation) thus evoked two consecutive MMNs; however, the legal pseudoword did not elicit the MMN because it did not mismatch the long-term stress representation. The findings suggest that the MMN could reflect changes in long-term representation of linguistic information such as stress pattern, and stress pattern changes are processed relying on long-term representation of word stress, not at single syllable level.

Chapter 6

Experimental section

6. Experimental section

6.1. Experiment 1: ERPs correlate of stress pattern of full-neutral/full-full tone pairs

Previous researches have suggested that sensitivity to rhythmic properties such as word stress (i.e. stress based on accentuation of syllables) or metrical (i.e. stress based on the vowel quality) stress is an important aspect of spoken word recognition in stress languages. With respect to word stress, Slowiaczek (1990) showed that English word stress plays an important role in the recognition of isolated spoken words. In an experiment, participants were asked to repeat an auditorily presented word aloud as quickly and accurately as possible. Results revealed that the response times were faster for correctly stressed word (e.g., *ANgry*) than for incorrectly stressed word (e.g., *anGRY*). In another lexical decision task, Slowiaczek (1990) found consistent results with correctly stressed words classified faster than incorrectly ones.

With regard to metrical stress, it has been shown that stress-timed languages such as English and Dutch use metrical information to segment the continuous speech stream into isolate word (Cutler & Norris, 1988; Vroomen & de Gelder, 1995). Cutler and Norris (1988) investigated the detection of a target word embedded in nonsense matrices with differing metrical structures. For example, the word *mint* (strong syllable) was embedded either in *mintayf* /mintef/, in which the second syllable, containing a full vowel, was strong, or *mintef* /mintəf/, in which the second syllable contained a schwa (i.e. weak syllable). Participants had to detect the word *mint*. Results showed that listeners detected words more slowly when the disyllable had two strong syllables /mintef/ than when it had a strong and a weak syllable /mintəf/. The authors argued that this result is an effect of segmentation: When the second syllable is strong, it is segmented from the first syllable, and successful detection of the embedded word therefore requires assembly of speech material across a segmentation position as the word *mint* belongs partially to both syllables and therefore was interfered. Cutler and Norris (1988) proposed a model of speech segmentation, i.e.the Metrical Segmentation Strategy (MSS) according to which the occurrence of a strong syllable triggers segmentation of the speech signal, whereas occurrence of a weak syllable does not trigger segmentation. Therefore, at least in English or more generally in stress-time languages, metrical stress may constitute a reliable boundary marker in the segmentation mechanism.

More recently, event related potentials (ERPs) have been used to investigate the neural correlates of word and metrical information in online language processing (Böcker et al., 1999; Isel & Friederici, 2005; Magne et al., 2010; Marie et al., 2011). Böcker et al. (1999) using a modified oddball paradigm to examine the ERP correlate of metrical stress using disyllabic Dutch words, which started with either a weak (e.g., *fregat* 'frigate') or a strong syllable (e.g., *franje* 'fringe'). In the modified oddball paradigm, sequences of four words were presented auditory in which the first three words had the same stress pattern (strong-strong or strong-weak; standard stimuli), while the fourth word had either the same or a different (deviant stimulus) stress pattern as the previous words of the sequence. Participants were asked to perform two tasks in two different blocks (1) one explicit task of stress discrimination (that means, whether the fourth word had same or different stress pattern with the first three words) and one implicit task of passive listening. The authors found that infrequent weak-initial Dutch words (12% of the Dutch lexicon) compared to strong-initial words evoke a larger positivity between 171 ms and 271 ms (P200) at frontocentral electrodes in the passive listening condition, and a larger negativity peaking at 325 ms post-stimulus at frontocentral electrodes in both tasks. Interestingly, the N325 was larger in the stress discrimination task than in the passive listening. These findings suggest that the early P200, i.e. an exogenous component, may reflect the processing of physical/acoustic information whereas the N325, i.e. an endogenous component (processing of cognitive information), is thought to sign the extraction of metrical stress from the acoustic signal. Marie et al. (2011) manipulated the metric structure of French words by lengthening the penultimate syllable of trisyllabic words, and found that metrically incongruous words elicited larger P200 components than metrically congruous ones. Furthermore, the P200 effect was larger for musicians than for non-musicians. It was proposed that the amplitude enhancement of the P200 may reflect the automatic processing of the temporal attributes of words and, specifically, of the words' syllabic structure. Magne et al. (2010) used the same paradigm used in the study of Böcker et al. (1999) with visually presented English disyllabic words (e.g., strong-weak: *table* weak-strong: *woman*) found that metrically incongruent target words elicited a larger negativity between 250 ms and 400 ms than metrically congruent target words in frontal and centrofrontal regions, that the authors called N400. However, unlike Böcker et al. (1999), they failed to show a P200. The results are partially in line with the finding of Böcker et al. (1999) and reinforce the idea that the frontocentral negativity could be a marker of extraction of metrical stress. Furthermore, the P200 may be modality-dependent as it was only observed in the auditory modality. Furthermore, the observation of ERP variations (here, the N325) in response to the visual processing of incongruent stimuli with respect to their stress patterning suggests that even in the visual modality an online metrical stress extraction can take place. This finding is consistent with the claim of Fodor (2002) that implicit prosody is processed in silent reading. That means, the readers may project prosodic information on the visual linguistic materials.

Chinese has been considered previously as a non-stress language because the tone is already used for distinguishing lexical meaning and therefore cannot be used for the realization of stress (Hyman, 1977; Selkirk & Shen, 1990). Although linguists agreed that, there are some stress phenomena in Chinese syllables (Lin, 2001; Duanmu 2000, 2007), the role of rhythmic properties of Chinese at word level has been little discussed. In linguistics, the most discussed stress phenomenon in Chinese is the 'neutral tone' (Chao, 1968). The neutral tone is also called fifth tone or zeroth tone (Chinese: 轻声, literal meaning: 'light tone') in addition to the four full tones. The neutral tone is a neutralization of a corresponding full tone, and it is associated systematically with weak syllables. Neutral tones are reminiscent of reduced syllables in English. Syllables that have a neutral tone are pronounced in a short and light manner, thus sounds quite weak. Acoustic studies (e.g., Cao, 1992; Liu & Xu, 2005) on neutral tone associated neutral tone with acoustical cues including shorter duration, smaller F0 variation and lower intensity. Compounds with neutral tone constitute 5% of the Chinese lexicon (Modern Chinese Dictionary, 1996 edition). The use of full tone and neutral tone leads to two degrees of "stress" (i.e. of strength) in Chinese, i.e. "stressed" (strong) that is loaded by a syllable with a full tone and "unstressed" (weak) loaded by a syllable with a neutral tone. What these weak syllables have in common is that 1) they must be attached to other syllable as a dependent unit in polysyllabic words; 2) they are usually the final syllable and never occur in the initial position in a word.

As we already mentioned, the role of rhythmic properties of Chinese at word level has been little discussed. To our best knowledge, no study has investigated the neurophysiological correlates of the processing of neutral tones in Chinese. Moreover, beyond the question of the perception of neutral tone in Chinese, more importantly, the study of the processing of neutral tones in final syllables of compound words is of central interest regarding the functional architecture of left-to-right sequential models postulating an early interaction between prosodic cues and morphological information. Indeed, in the case when the neutral tone is carried out by the final syllable of a compound word in Chinese, this prosodic information is not immediately available at the beginning of the word. Therefore, a left-to-right sequential processing model of compound word recognition that assumes an early involvement of prosody for assisting morphological analysis at the lexical level would not be designed to process grammatical sequences delivering critical prosodic information at the end of the word. This could have a dramatic consequence on lexical access and more generally on language comprehension in Chinese, particularly when meaning of the compound word has to be reconsidered once a neutral tone was detected on the last syllable of the compound word. This issue will be addressed in Experiment 2 of the present doctoral thesis.

Before to examine the lexical consequences of the processing of neutral tones in Chinese compound words, we wanted firstly to ensure that neutral tone is perceived by native speakers of Mandarin Chinese. For this purpose, we conducted a first electrophysiological experiment (Experiment 1), which aimed to study the ERPs correlate of the neutral tone using a modified oddball paradigm inspired by the study of Böcker et al. (1999). Participants performed two tasks, namely an explicit stress discrimination task and a task of passive listening (implicit task with respect to prosodic processing investigated here). Strong-final and weak-final Chinese words were auditory presented. In a sequence of four words, the first three words had the same stress pattern (strong-strong or strong-weak[33]), while the fourth word had either the same or different stress pattern as the previous words of the sequence. Congruent and incongruent stress pattern conditions were thus created. If stress information in strong-strong/strong-weak

[33] The weak syllable carries a neutral tone.

words was processed, it will be reflected in ERP components (P200, N325). In the explicit task, i.e. the stress discrimination, native Chinese listened to sequences of four disyllabic words in which the fourth word either had the same or a different stress pattern as the first three words, the task of participant was discriminated between sequences with same and different stress patterns. This task required that stress information be processed explicitly; in the second task, i.e. the passive listening, participants were asked to listen carefully to the words presented to them in the task, subjects the processing of stress pattern was implicit. Moreover, to ensure that data we collected in Experiment 1 does reflect processing of information and were not influenced by top-down lexical information, we created an additional experimental condition in which the compound words used in the congruent and incongruent stress patterning conditions was filtered so that the lexical information was removed and only the prosodic stress pattern information remained (delexicalized condition).

We expected to observe two types of ERP components to be related to differences in stress pattern processing. First, an early exogenous component (i.e., P200) was expected to occur to infrequent stress pattern (i.e., strong-weak). Secondly, a late endogenous component (i.e., N325) reflecting the extraction of metrical stress from the acoustic signal. Moreover, due to oddball-like design, we expected the mismatch negativity (MMN; Näätänen, Gaillard, & Mäntysalo, 1978) for targets in incongruent trials regardless of stress pattern. Moreover, we predicted that if stress pattern information relies on lexical information, that is to say, the processing of stress pattern is not only of the physical/acoustic stimulus parameters, stress processing related ERPs (i.e. the P200 and the N325) should only occur in the non-delexicalized condition.

6.1.1. Method

6.1.1.1. Participants

Eighteen native Mandarin Chinese speakers (10 females, 26.33 ±3.56 years, ranging from 21 to 33 years) were recruited in Paris. None of them has lived in France more than five years

(3.28 ± 1.36 years, ranging from 1 to 5 years) and all of them used Chinese frequently in daily life. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). They have normal or corrected to-normal vision and no auditory impairment. They have no history of current or past neurological or psychiatric illnesses. They were paid 10€ per hour for their time. All participants gave informed consent before participation and completed an information sheet.

6.1.1.2. Stimuli

A set of 60 pairs of Chinese compound words (see Appendix A) was selected from the Contemporary Chinese Dictionary (2005). They were disyllabic compounds in which the first syllable carries a full tone. The second syllable in the two compounds in each pair was the same, but different in stress status. One carried a neutral tone (i.e., strong-weak stress pattern), and the other carried a full tone (i.e., strong-strong stress pattern). Neutral tone compounds and full tone compounds were matched on the frequency[34] (40.6 occurrences per million; t (118) = 0.29, $p > 0.05$) and on the of strike number[35] [t (118) = 1.13, $p > 0.05$]. All words were recorded in an acoustic chamber by a native Mandarin Chinese speaker (female, 28 years old, from Nanjing). In order to verify the contrastive stress pattern in the neutral tone syllable and the full tone syllable, we conducted a phonetic analysis on the final syllable of the words. Results indicated that 1) syllable duration was shorter [t (118) = 19.37, $p < 0.001$] for neutral tone syllables (256.01 ms ± 56.83 ms) than for full tone syllables (518.08 *ms* ± 88.06 *ms*); 2) average intensity was lower [t (118) = 8.99, $p < 0.001$] for neutral tone syllables (46.92 dB ± 4.69 dB) than for full tone syllables (53.89 dB ± 3.74 dB); 3) average F0 was lower [t (118) = 3.68, $p < 0.001$] for neutral tone syllables (186.59 Hz ± 49.7 Hz) than for full tone syllables (214.62 Hz ± 31.73 Hz) and 4) F0 variation was smaller [t (118) = 7.46, $p < 0.001$] for neutral

[34] Frequency data is based on the Frequency Dictionary on the Modern Chinese Corpus with 200 million characters, approximately 100 million words (China State Language Commission, www.cncorpus.org).
[35] The strike number and the length of a Chinese word are two measures of the complexity in writing of the word. In the study, all words have the same length, i.e., 2 characters. The strike number has been shown as an important factor in the recognition of Chinese characters, a character with fewer strokes is accessed faster than a character with more strokes (Peng & Wang, 1997).

tone syllables (65.0 Hz ± 48.6 Hz) than for full tone syllables (147.6 Hz ± 70.7 Hz). Results on the first syllable only showed that the duration was shorter [t (118) = 3.24, $p < 0.001$] when the syllable is followed by a neutral tone (372.75 ms ± 57.10 ms) than by a full tone (415.04 ms ± 83.27 ms). For preparing the delexicalized stimuli, we applied a low-pass filter at 400 Hz to all stimuli in order to obtain stimuli where all lexical information was removed. According to two native speakers of Chinese, these delexicalized words were lexically unintelligible, however most stress pattern information[36] (e.g., duration, F0) was preserved.

6.1.1.3. Design and Procedure

Each trial started with a visual fixation at the center of a computer monitor for 1000 ms, this was followed by a sequence of four auditorily presented words. The stimulus onset asynchrony for subsequent words was 1500 ms. The target word (the fourth one) was presented for 3000 ms, and then was interrupted as soon as participants initiated their response. The intertrial interval (ITI) was 1500 ms. Stimulus delivery and behavioral response recording were controlled by E-prime (Psychology Software Tools, Inc., Pittsburgh, PA; http://www.pstnet.com/eprime). In each sequence, four spoken disyllabic compound words were presented. The first three words all had the same stress pattern, that is, either strong-weak (i.e., neutral tone on the second syllable) or strong-strong (i.e., full tone on the second syllable). Prosodic congruency was manipulated by varying the stress pattern of the fourth word. In the prosodically congruent condition, the fourth word had the same stress pattern as the previous three words. In the prosodically incongruent condition, the fourth word had a different stress pattern. The same condition was presented with delexicalized stimuli. Examples of the four possible stimulus sequences are presented in Table 6.1.1.

[36] The duration and the F0 are not affected by the low-pass filter at 400 Hz. The intensity is proportionally attenuated due to the filter.

Table 6.1.1.

Examples of the four experimental conditions

Condition	Word 1	Word 2	Word 3	Word 4	Congruency
1	SS *shi2fen1*	SS *lao3hu3*	SS *yang2mao2*	**SS** *tian1qi4*	Congruent
2	SW *bu4fen0*	SW *ma3hu0*	SW *mei2mao0*	**SW** *ke4qi0*	Congruent
3	SW *jia1huo0*	SW *shu1fu0*	SW *bu4fen0*	**SS** *tian1qi4*	Incongruent
4	SS *tong2huo3*	SS *zhi4fu4*	SS *shi2fen1*	**SW** *ke4qi0*	Incongruent

Note: SW = Strong-Weak stress pattern (with a neutral tone); SS = Strong-Strong pattern (with a full tone).

The participants passed the two types of stimuli, i.e., the non-delexicalized compounds and the delexicalized compounds. Moreover, two tasks were employed. The first task was a stress discrimination task in which the participants were instructed to discriminate between sequences of congruent and incongruent stress patterns by pressing the one button (*S* or *K* on keyboard) when the stress pattern of the target was congruent with that of the first three and the other button when it was incongruent. The second task was a passive listening task where the participants were asked to listen carefully to the words presented to them. The crossing of Lexicalization (two types of stimuli: non-delexicalized stimuli, delexicalized stimuli) × Task (two tasks: explicit, implicit) created four experimental blocks, i.e., (1) non-delexicalized and discrimination, (2) non-delexicalized and passive listening, (3) delexicalized and discrimination, (4) delexicalized and passive listening. Block order was counterbalanced across the participants. Each block contained 120 trials (i.e., 30 trials per condition) presented in a pseudo-randomized order. The following constraints were used for creating the pseudo-randomization order. First, target word was not repeated immediately. Secondly, no more than three sequences of the same experimental condition were presented in succession. The pseudo-randomized order of trials was created using the program Conan (Nowagk, 1998).

6.1.1.4. ERP Recordings and Analysis

The EEG was recorded from 64-channel electrodes using Brain Vision Recorder (Brain Products GmbH, Gilching, Germany). Data was sampled at 1000 Hz with an online 0.1-100 Hz frequency bandpass filter. All electrode impedances were maintained below 5 kΩ. Further processing of ERPs was done off-line. First, EEG data were re-referenced to an average reference, and then were filtered with a 0.1-12 Hz bandpass filter. Next, the continuous EEG was segmented into epochs from 200 ms pre-stimulus until 1500 ms post-stimulus onset, with 200 ms of each epoch corresponding to a pre-stimulus baseline. Then, a semi-automatic artifact decontamination procedure was applied. Artifacts were marked when 1) maximal amplitude was higher than 100 μV or minimal amplitude was lower than -100 μV; 2) absolute difference was more than 50 μV or less than 0.5 μV in an interval of 200 ms. Only trials with correct responses and with no artifacts were kept for further data analysis. In each experimental condition, the segmented ERP data was then averaged over trials and over participants (i.e., grand average). Finally, the ERP data were baseline-corrected to the mean amplitude of the 200 ms pre-stimulus interval.

Statistical analyses were conducted for three ERP signatures for which the time windows were selected based on previous studies adapted by visual inspection of the grand averages: N200 and P200 (180-300 ms), N325 (300-600 ms), and a late sustained negative-going potential (300-1000 ms). Repeated measures ANOVA were conduct for the Stress pattern effect (strong-weak vs. strong-strong) and for the Congruency effect of stress pattern (congruent vs. incongruent) for each four experimental block, i.e., 1) Block 1: stress discrimination task and non-delexicalized stimuli, 2) Block 2: passive listening task and non-delexicalized stimuli, 3) Block 3: stress discrimination task and delexicalized stimuli 4) Block 4: passive listening task and delexicalized stimuli. To permit examination of hemispheric differences, the data recorded at the lateral recording sites were treated separately from the data recorded at the midline electrode sites. All lateral electrodes were classified by two topographical variables: Region (anterior, posterior) and Hemisphere (left, right). Four regions of interest (ROIs) resulting from a complete crossing of the Region and Hemisphere variables were defined: left anterior (F7, F3, FT7, FC3), right anterior (F8, F4, FT8, FC4), left posterior (P3, P7, O1), and right posterior (P4, P8, O2). For each time window, a six-way repeated measures ANOVA with the factors Lexicalization (non-delexicalized, delexicalized), Congruency (congruent, incongruent), Stress pattern of the target (strong-strong, strong-weak), Task (stress discrimination, passive listening) and the topographical variables Hemisphere (left, right) and Region (anterior, posterior) was conducted. For the midline electrodes, a five-way repeated measures ANOVA including factors Lexicalization (non-delexicalized, delexicalized), Congruency (congruent, incongruent), Stress pattern of the target (strong-strong, strong-weak), Task (stress discrimination, passive listening) and Electrode (Fz, Cz, Pz) was run for each of the three time windows of interest. The Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied when evaluating effects with more than one degree of freedom in the numerator. Post-hoc pairwise comparisons at single electrode sites were made using a modified Bonferroni procedure (Keppel, 1991).

6.1.2. Results

6.1.2.1. Behavioral Results

Behavioral data[37], in the form of error rate and response times, were available for the discrimination task only. Table 5.1.2 shows error rate and response times in the four experimental conditions (congruent, incongruent, strong-strong, strong-weak) for both non-delexicalized stimuli and for delexicalized stimuli. For error rate, a three-way repeated measures ANOVA with the factors Lexicalization (non-delexicalized, delexicalized), Congruency (congruent, incongruent), and Stress pattern of the target (strong-strong, strong-weak) revealed a main effect of Lexicalization [F (1, 16) = 23.48, MSE = 0.017, $p <$ 0.001, $\eta 2_p = 0.60$], reflecting that error rate was higher for delexicalized stimuli (20.4 ± 15.9%) compared to non-delexicalized stimuli (9.5 ± 9.7%) and a main effect of Stress pattern [F (1, 16) = 7.92, MSE =.009, $p < .05$, $\eta 2_p = 0.33$], reflecting that error rate was higher for strong-weak target word (17.3 ± 14.8%) compared to strong-strong target word (12.6 ± 10.8%).

[37] Data of one participant was excluded from analyses, because of his high error rate (48.8% errors).

There was a significant interaction of Lexicalization × Stress pattern [F (1, 16) = 5.90, MSE = 0.002, p < 0.05, $\eta 2_p$ = 0.27]. The interaction reflected Stress pattern effect for non-delexicalized stimuli [F (1, 16) = 14.13, MSE = 0.005, p < 0.01, $\eta 2_p$ = 0.47], but not for delexicalized stimuli [F (1, 16) = 1.90, MSE = 0.006, p > 0.05]. Furthermore, the error rate was higher in incongruent condition (12.1 ± 11.7%) than in congruent condition (7.0 ± 7.6%) but only for non-delexicalized stimuli [F (1, 16) = 8.24, MSE = 0.005, p < 0.05, $\eta 2_p$ = 0.34]. For response times, a three-way repeated measures ANOVA revealed a main effect of Lexicalization [F (1, 16) = 12.70, MSE = 19843.772, p < 0.01, $\eta 2_p$ = 0.44], reflecting that response times were slower for delexicalized stimuli (790.4 ± 239.4 ms) compared to non-delexicalized stimuli (704.3 ± 256.0 ms) and a main effect of Stress pattern [F (1, 16) = 36.86, MSE = 9249.261, p < 0.001, $\eta 2_p$ = 0.70], reflecting that response times were slower for strong-strong target word (797.4 ± 247.0 ms) compared to strong-weak target word (697.3 ± 248.4 ms). No interaction was found.

Table 6.1.2.

Behavioral data for four experimental conditions

	Non-delexicalized				Delexicalized			
	Error [%]		RT [ms]		Error [%]		RT [ms]	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Congruent	7.0	7.6	703.1	243.4	18.7	15.8	780.8	247.3
Incongruent	12.1	11.7	705.4	268.6	22.2	16.0	800.0	231.5
Strong-strong	6.2	7.5	756.8	252.8	19.1	14.2	838.0	241.1
Strong-weak	12.8	11.9	651.8	259.2	21.8	17.6	742.8	237.7

Note: Strong-strong is the stress pattern of targets with full tones and strong-weak is the stress pattern of targets with a neutral tone.

6.1.2.2. Electrophysiological Results

Statistical analysis of the EEG data was conduct in each four experimental block (see Table 6.1.3), i.e., 1) Block 1: stress discrimination task and non-delexicalized stimuli, 2) Block 2: passive listening task and non-delexicalized stimuli, 3) Block 3: stress discrimination task and delexicalized stimuli 4) Block 4: passive listening task and delexicalized stimuli. Stress pattern effect was analyzed in the time window 180-300 ms and 300-600 ms and Congruency effect was analyzed in the time window 300-1000 ms.

Table 6.1.3.

Task and type of stimuli in the four experimental blocks

Block	Task	Stimuli
Block 1	Stress discrimination	Non-delexicalized
Block 2	Passive listening	Non-delexicalized
Block 3	Stress discrimination	Delexicalized
Block 4	Passive listening	Delexicalized

6.1.2.2.1. Stress pattern effect

Time window 180-300 ms (N200 & P200)

Block 1 (Task: stress discrimination; Stimuli: non-delexicalized)

Figures 6.1.1 and 6.1.2 display the ERP data. A three-way repeated measures ANOVA on lateral electrodes with the factors Stress pattern of the target, Hemisphere and Region did not reveal any main effect or interaction ($Fs < 1$). A two-way repeated measures ANOVA on the three midline electrodes with the factors Stress pattern of the target and Electrode showed a main effect of Stress [F (1, 16) = 12.71, MSE = 0.606, $p < 0.01$, $\eta 2_p = 0.44$], indicating that strong-weak pattern elicited a larger negativity than strong-strong one. Moreover, a significant Stress × Electrode interaction was found [F (2, 32) = 4.18, MSE = 0.798, $p < 0.05$, $\eta 2_p = 0.21$]. Post hoc analyses revealed that strong-weak target words elicited a larger negativity than strong-strong words (i.e., the N200) on the electrodes Fz [F (1, 16) = 15.33, MSE = 0.573, $p < 0.001$, $\eta 2_p = 0.49$] and Cz [F (1, 16) = 6.27, MSE = 0.601, $p < 0.05$, $\eta 2_p = 0.28$]. There was no significant N200 effect on the electrode Pz [$F < 1$].

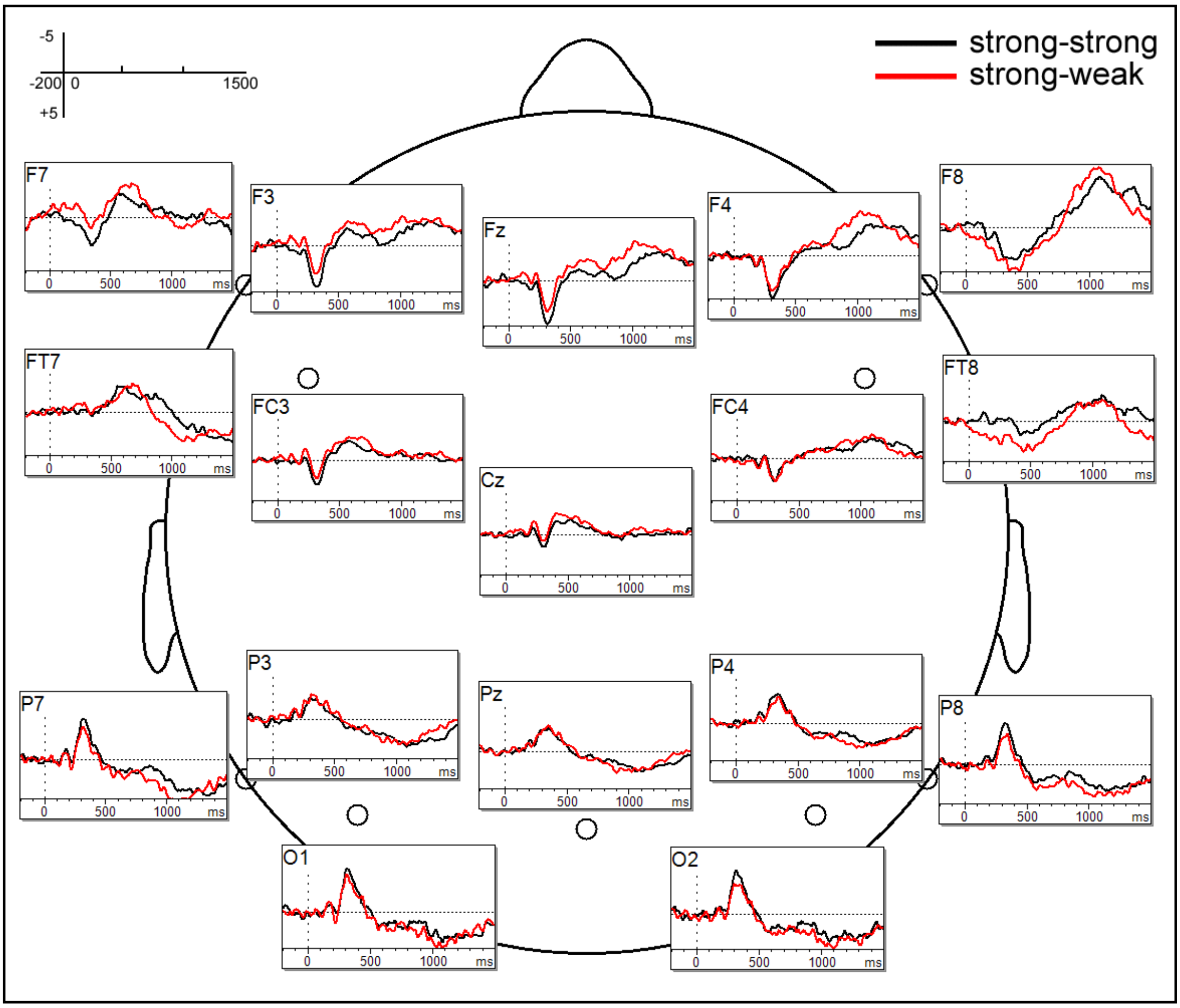


Figure 6.1.1. Block 1: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the 17 electrodes (F7, F3, FT7, FC3, F8, F4, FT8, FC4, P3, P7, O1, P4, P8, O2, Fz, Cz, Pz).

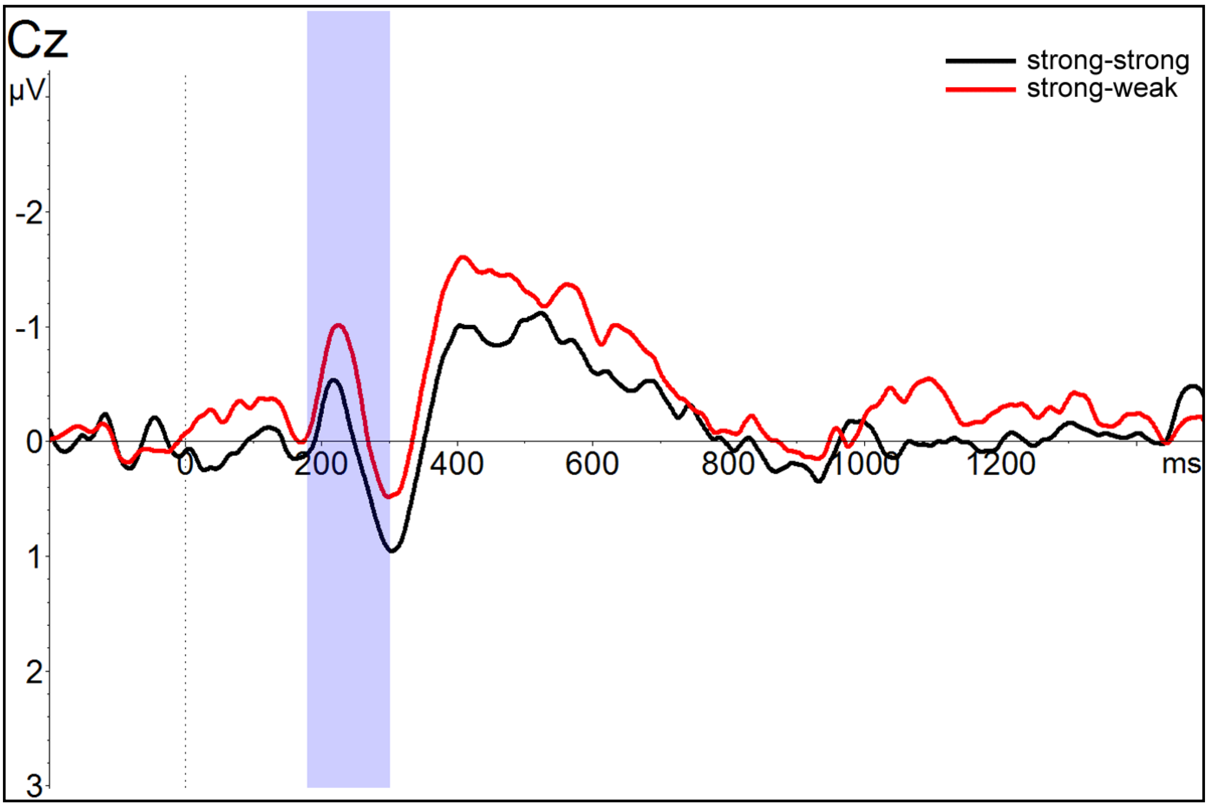


Figure 6.1.2. Block 1: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode Cz.

Block 2 (Task: passive listening; Stimuli: non-delexicalized)

For lateral electrodes, ANOVA results did not show any main effect or interaction (*Fs* < 1). For midline electrodes, a significant main effect of Stress was found [F (1, 16) = 6.27, MSE = 1.050, $p < 0.05$, $\eta 2_p = 0.28$], indicating a larger negativity in the 180-300 ms time window for the strong-weak pattern than for the strong-strong one. ERP waveforms at the Cz are plotted in Figure 6.1.3.

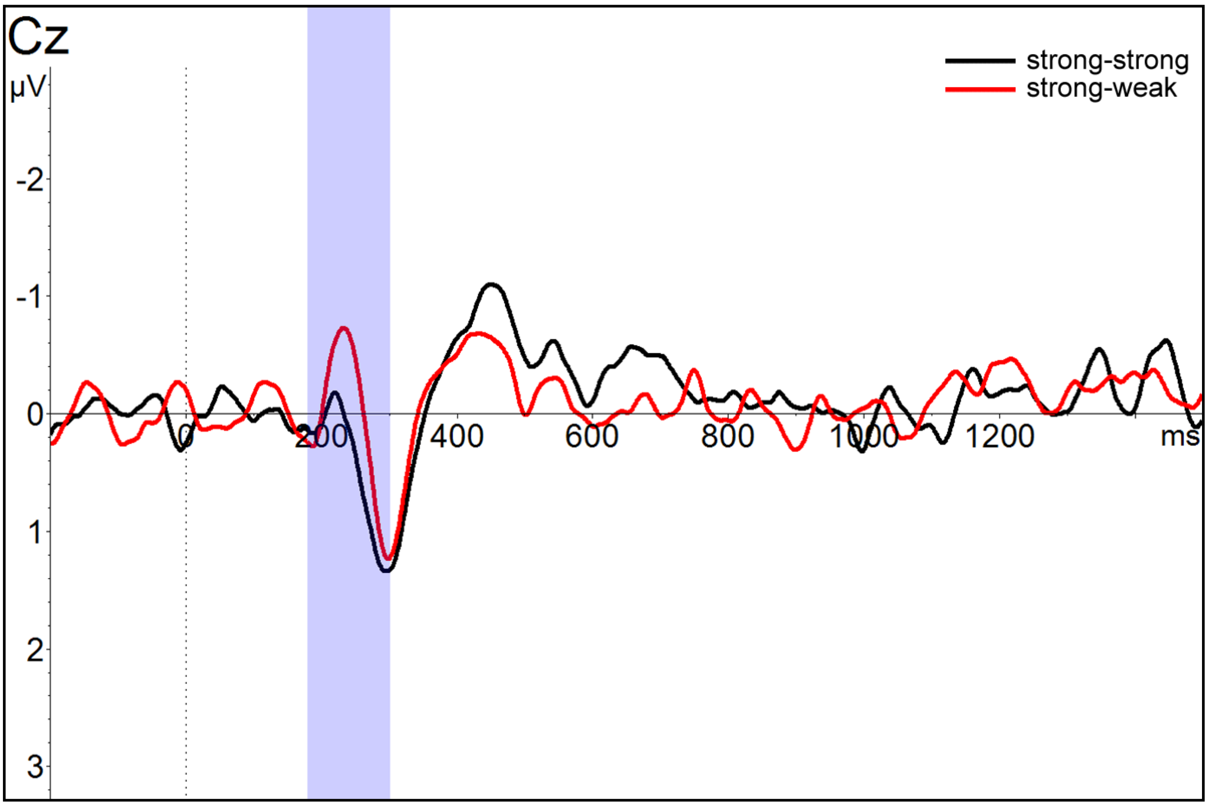


Figure 6.1.3. Block 2: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode Cz.

Block 3 (Task: stress discrimination; Stimuli: delexicalized)

For lateral electrodes, results showed a main effect of Stress [F (1, 16) = 4.55, MSE = 0.109, $p < 0.05$, $\eta 2_p = 0.22$], reflecting that strong-weak target words elicited a larger positivity in the 180-300 ms time window than the Strong-Strong words. There was no significant interaction. For the midline electrodes, results revealed a significant interaction between Stress and Electrode [F (2, 32) = 4.26, MSE = 0.433, $p < 0.05$, $\eta 2_p = 0.21$]. Post hoc analyses revealed a larger positivity in the strong-weak than in the strong-strong pattern but only at the electrodes Fz [F (1, 16) = 12.63, MSE = 0.482, $p < 0.01$, $\eta 2_p = 0.44$] (see Figure 6.1.4). Based on the latency and the topography of surface, one can reasonably assume that the observed effect may be a P200 effect.

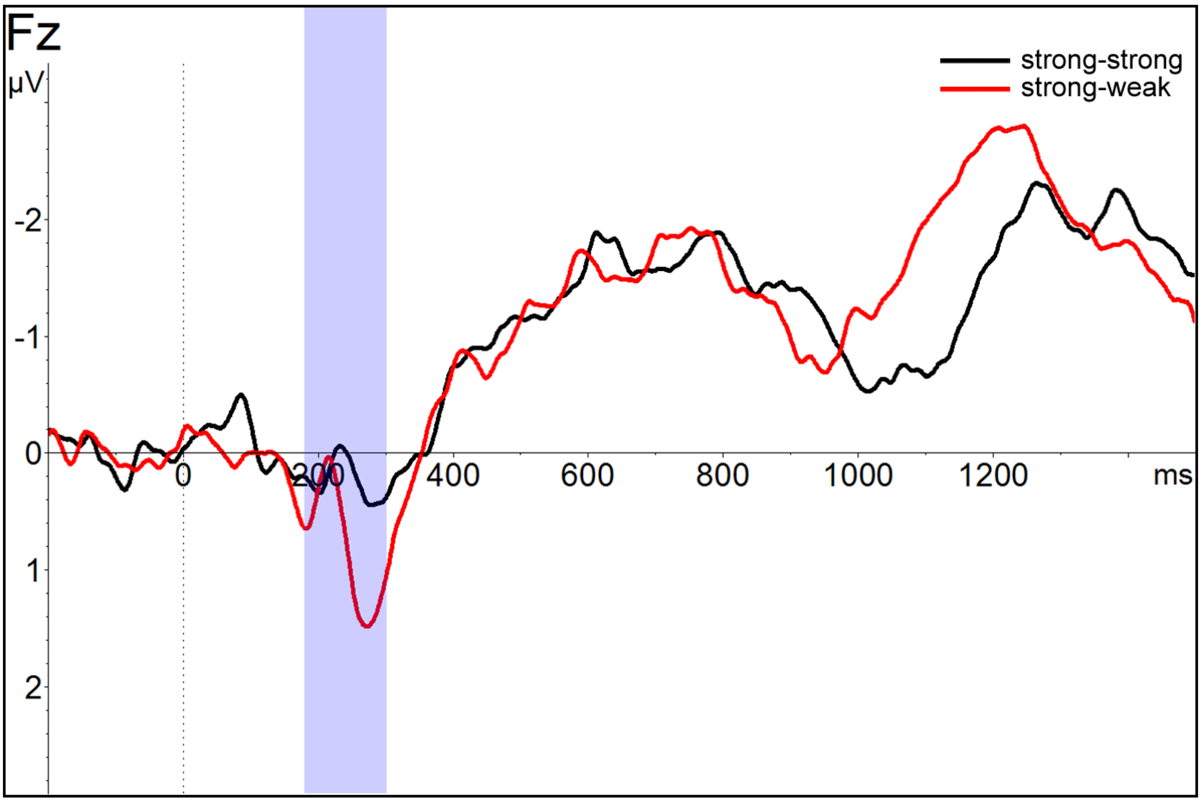


Figure 6.1.4. Block 3: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode Fz.

Block 4 (Task: passive listening; Stimuli: delexicalized)

A three-way repeated measures ANOVA on lateral electrodes with the factors Stress pattern of the target, Hemisphere and Region showed a main effect of Stress [F (1, 16) = 6.42, MSE = 0.195, $p < 0.05$, $\eta 2_p = 0.29$], indicating a larger positivity in the strong-weak condition than in the strong-strong one. Moreover, a significant Stress × Region interaction was found [F (1, 16) = 6.24, MSE = 1.289, $p < 0.05$, $\eta 2_p = 0.28$]. This interaction revealed a significant negativity effect at posterior electrodes [F (1, 16) = 6.94, MSE = 1.127, $p < 0.05$, $\eta 2_p = 0.30$], but only a marginally significant P200 effect ($p = 0.059$) at anterior electrodes. For the midline electrodes, the three-way repeated measures ANOVA revealed a significant Stress × Electrode interaction [F (2, 32) = 4.86, MSE = 5.622, $p < 0.05$, $\eta 2_p = 0.23$]. Post hoc analyses revealed a significant P200 effect only on the electrodes Fz [F (1, 16) = 6.46, MSE = 0.601, $p < 0.05$, $\eta 2_p = 0.29$], indicating that the amplitude of strong-weak target words (0.77 ± 1.6 μV) was more positive than the one of Strong-Strong words (0.29 ± 1.1 μV). ERP data at the Fz is shown in

Figure 6.1.5.

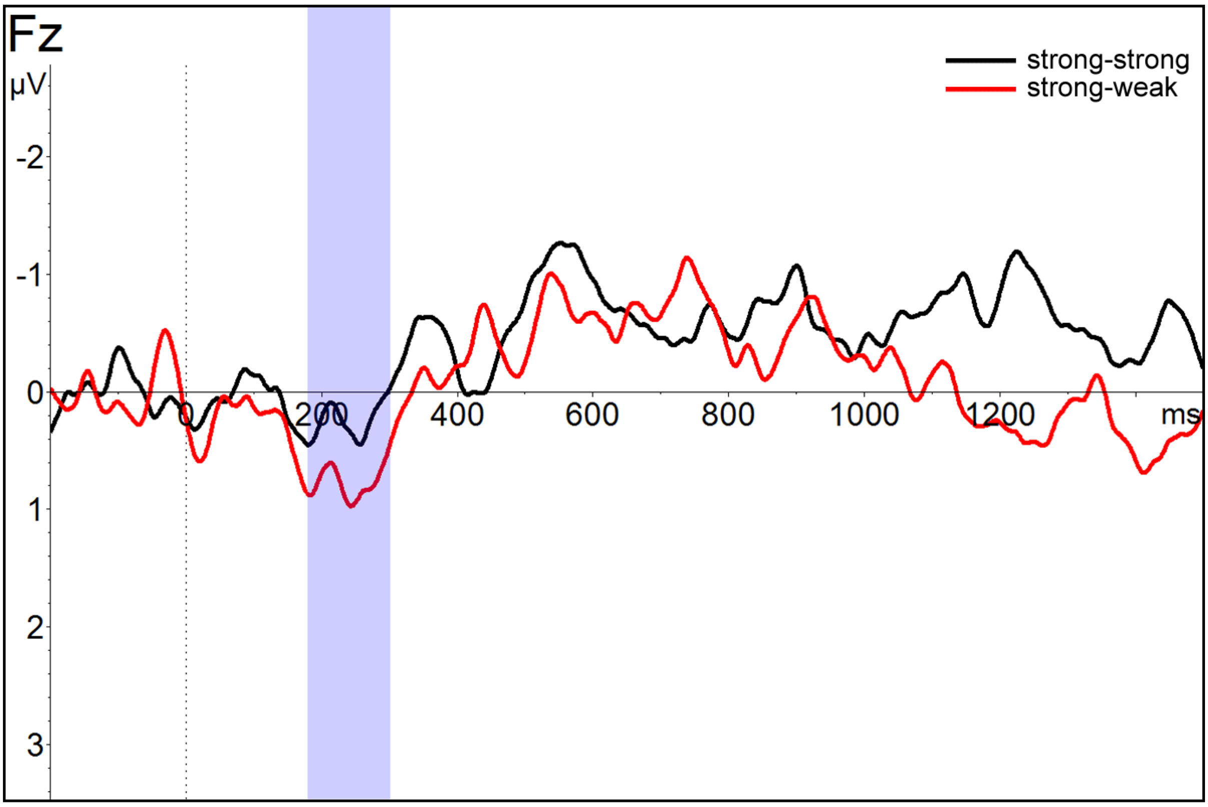


Figure 6.1.5. Block 4: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode Fz for (Task: passive listening; Stimuli: delexicalized).

Time window 300-600 ms (N325)

Block 1 (Task: stress discrimination; Stimuli: non-delexicalized)

A three-way repeated measures ANOVA on lateral electrodes with the factors Stress pattern of the target, Hemisphere and Region did not reveal a main effect of Stress. In contrast, a significant interaction of Stress × Hemisphere [F (1, 16) = 4.60, MSE = 0.996, $p < 0.05$, $\eta 2_p$ = 0.22], indicating that the N325 effect at the left hemisphere sites [F (1, 16) = 16.04, MSE = 1.749, $p < 0.001$, $\eta 2_p$ = 0.50], however, a larger positivity at the right hemisphere sites [F (1, 16) = 19.05, MSE = 18.37, $p < 0.001$, $\eta 2_p$ = 0.54]. Moreover, a marginally significant three-way Stress × Region × Hemisphere interaction was found [F (1, 16) = 3.12, MSE = 0.657, p = 0.097, $\eta 2_p$ = 0.16]. Post-hoc analyses showed a larger N325 effect at left anterior electrodes quadrant [F (1, 16) = 9.26, MSE = 0.414, $p < 0.01$, $\eta 2_p = 0.37$]. A two-way repeated measures ANOVA conducted on the midline electrodes with the factors Stress pattern of the target and Electrode showed a main effect of Stress [F (1, 16) = 13.54, MSE = 0.806, $p < 0.01$, $\eta 2_p = 0.46$]. The amplitude of the N325 was larger for strong-weak target words (-0.75 ± 1.4 μV) than strong-strong words (-0.28 ± 1.5 μV). Figure 6.1.6 displays the ERP data at the Cz. In contrast, the N325 was not significant at lateral or midline electrodes in Block 2, Block 3 or Block 4.

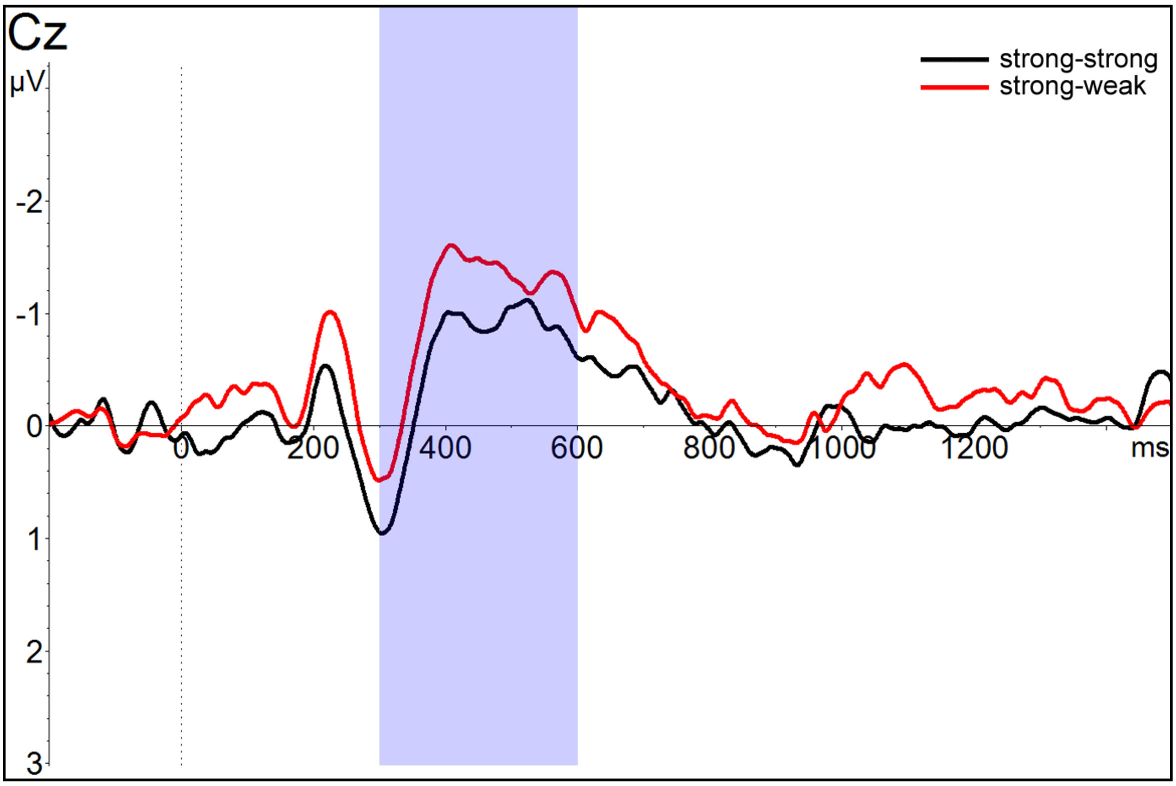


Figure 6.1.6. Block 1: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode Cz.

6.1.2.2.2. Congruency effect

Time window 300-1000 ms (late sustained negative)

Block 1 (Task: stress discrimination; Stimuli: non-delexicalized)

Figures 6.1.7 and 6.1.8 show the ERP data for Block 1. A three-way repeated measures

ANOVA on lateral electrodes with the factors Congruency on the stress pattern of the target, Hemisphere and Region showed a significant interaction of Congruency × Hemisphere [F (1, 16) = 15.34, MSE = 2.015, $p < 0.001$, $\eta 2_p = 0.49$] and a significant interaction of Congruency × region × Hemisphere [F (1, 16) = 14.46, MSE = 0.770, $p < 0.01$, $\eta 2_p = 0.48$]. Post-hoc analyses revealed a larger sustained negativity in the left anterior Region of Interest [ROI; F (1, 16) = 17.40, MSE = 1.149, $p < 0.001$, $\eta 2_p = 0.52$], but a larger positivity at right anterior electrodes [F (1, 16) = 11.89, MSE = 1.646, $p < 0.01$, $\eta 2_p = 0.42$] (see Figure 6.1.7). A two-way repeated measures ANOVA on midline electrodes with the factors Stress pattern of the target and Electrode showed neither a main effect of Congruency [F (1, 16) = 2.91, MSE = 0.849, $p > 0.05$] nor a significant interaction of Congruency × Electrode [$F < 1$].

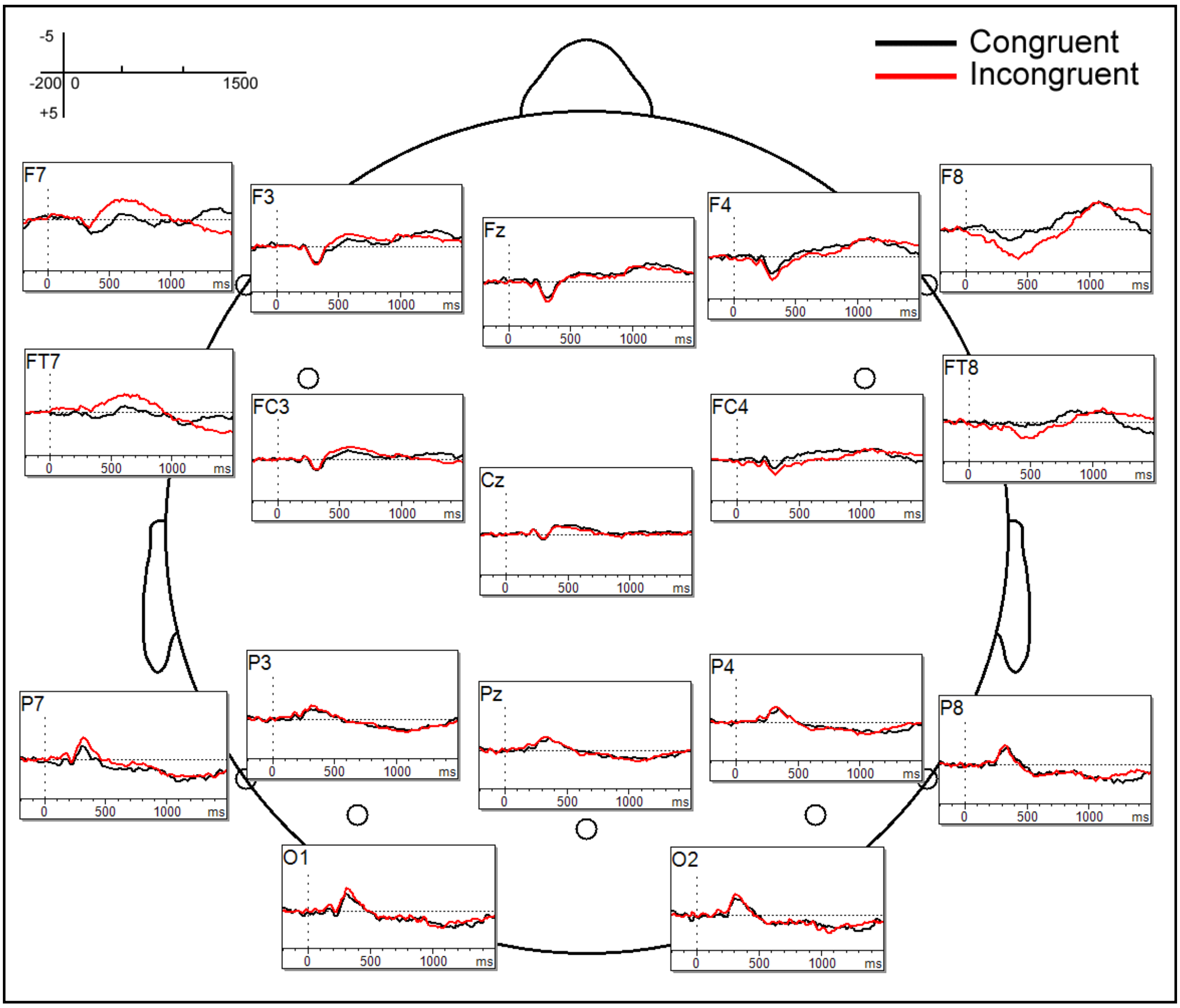


Figure 6.1.7. Block 1: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the 19 electrodes (F7, F3, FT7, FC3, F8, F4, FT8, FC4, P3, P7, O1, P4, P8, O2, Fz, Cz, Pz).

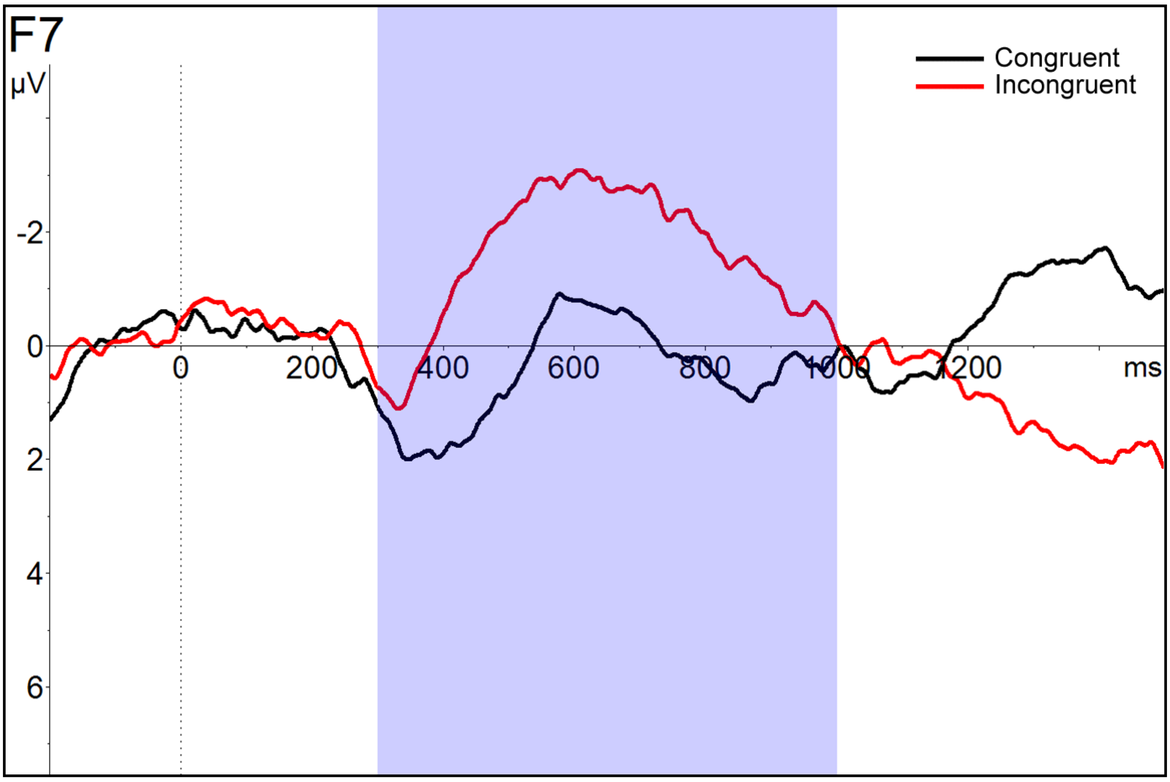


Figure 6.1.8. Block 1: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode F7.

Block 2 (Task: passive listening; Stimuli: non-delexicalized)

For the lateral electrodes, a significant interaction of Congruency × Region [F (1, 16) = 4.65, MSE = 2.960, $p < 0.05$, $\eta 2_p = 0.23$] was found. The interaction was due to a larger sustained negativity at the anterior electrodes [F (1, 16) = 3.92, MSE = 0.463, $p = 0.65$ marginally significant, $\eta 2_p = 0.20$] and a larger positivity at the posterior electrodes [F (1, 16) = 4.40, MSE = 3.451, $p = 0.52$ marginally significant, $\eta 2_p = 0.22$]. Figure 6.1.9 displays the ERP results at the F4. For the midline electrodes, results did not reveal neither a main effect of Congruency [$F < 1$] nor a significant interaction between Congruency and Electrode [F (2, 32) = 1.72, MSE = 2.866, $p > 0.05$].

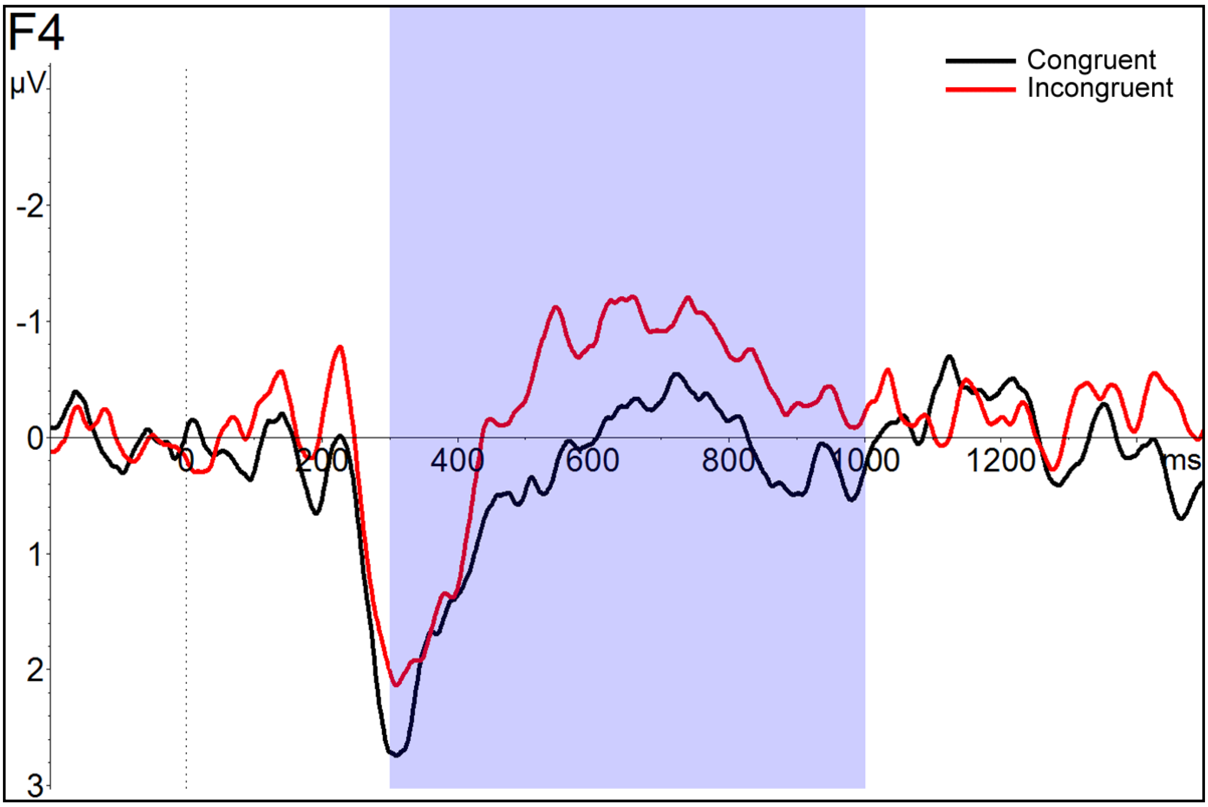


Figure 6.1.9. Block 2: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode F4.

Block 3 (Task: stress discrimination; Stimuli: delexicalized)

ERP waveforms are plotted in Figures 6.1.10 and 6.1.11. For the lateral electrodes, the ANOVA showed a significant Congruency × Hemisphere interaction [F (1, 16) = 20.74, MSE = 1.040, $p < 0.001$, $\eta 2_p = 0.57$]. Post-hoc comparisons showed that this interaction was due to a reversal polarity of the effects found in the two hemispheres, with a sustained negativity effect in the left hemisphere and a positivity effect in the right hemisphere. Moreover, a significant interaction between Congruency, Region and Hemisphere was observed [F (1, 16) = 6.51, MSE = 0.494, $p < 0.01$, $\eta 2_p = 0.45$]. Post-hoc analyses revealed a larger sustained negativity at the left anterior electrodes site [F (1, 16) = 24.01, MSE = 0.645, $p < 0.001$, $\eta 2_p = 0.60$] and a larger positivity effect at the right anterior electrodes site [F (1, 16) = 13.16, MSE = 0.808, $p < 0.01$, $\eta 2_p = 0.45$] and at right posterior electrodes [F (1, 16) = 8.58, MSE = 0.808, $p < 0.01$, $\eta 2_p = 0.35$]. For central electrodes, there was neither a main effect of Congruency nor a significant Congruency × Electrode interaction [$Fs < 1$].

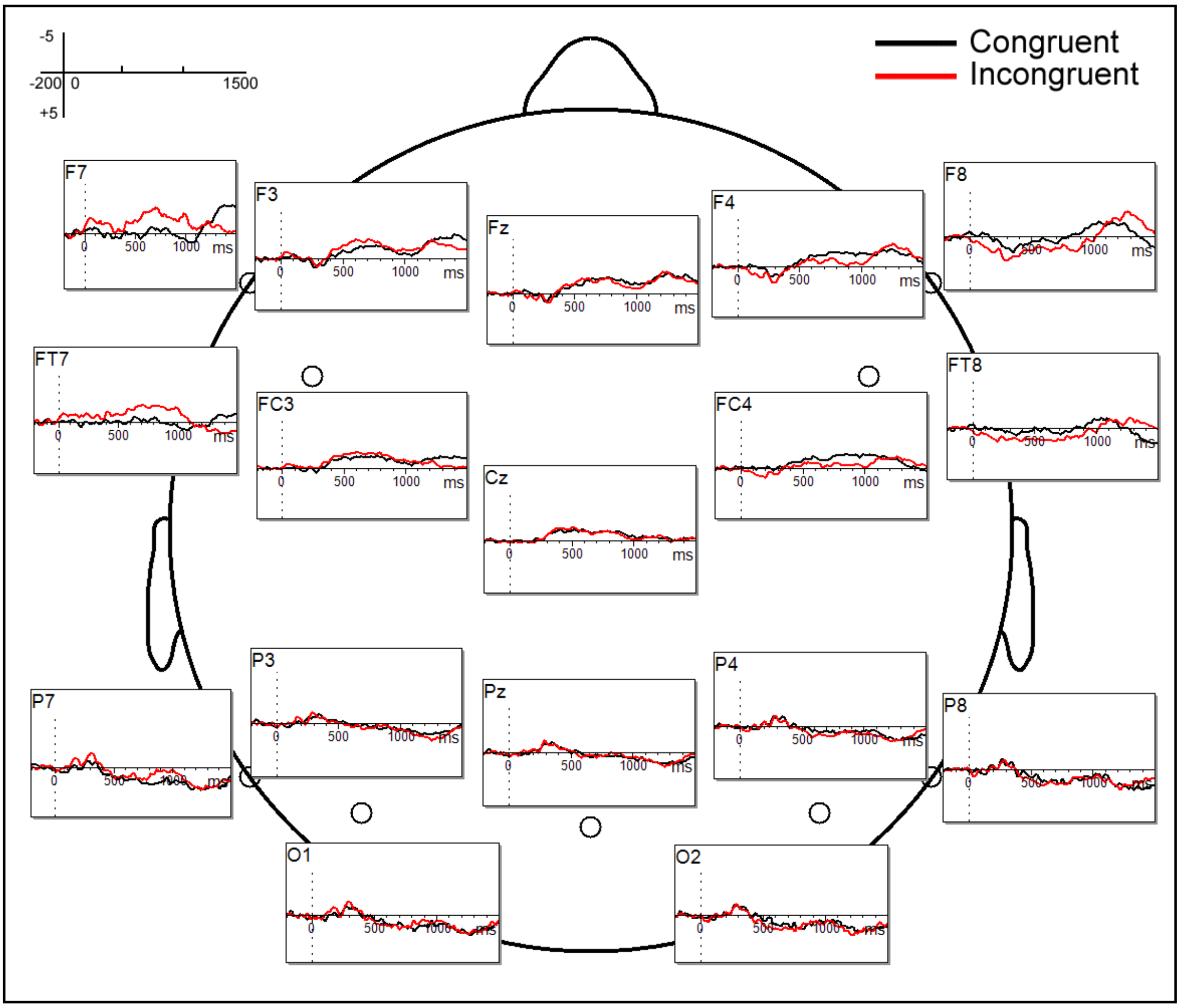


Figure 6.1.10. Block 3: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the 19 electrodes (F7, F3, FT7, FC3, F8, F4, FT8, FC4, P3, P7, O1, P4, P8, O2, Fz, Cz, Pz).

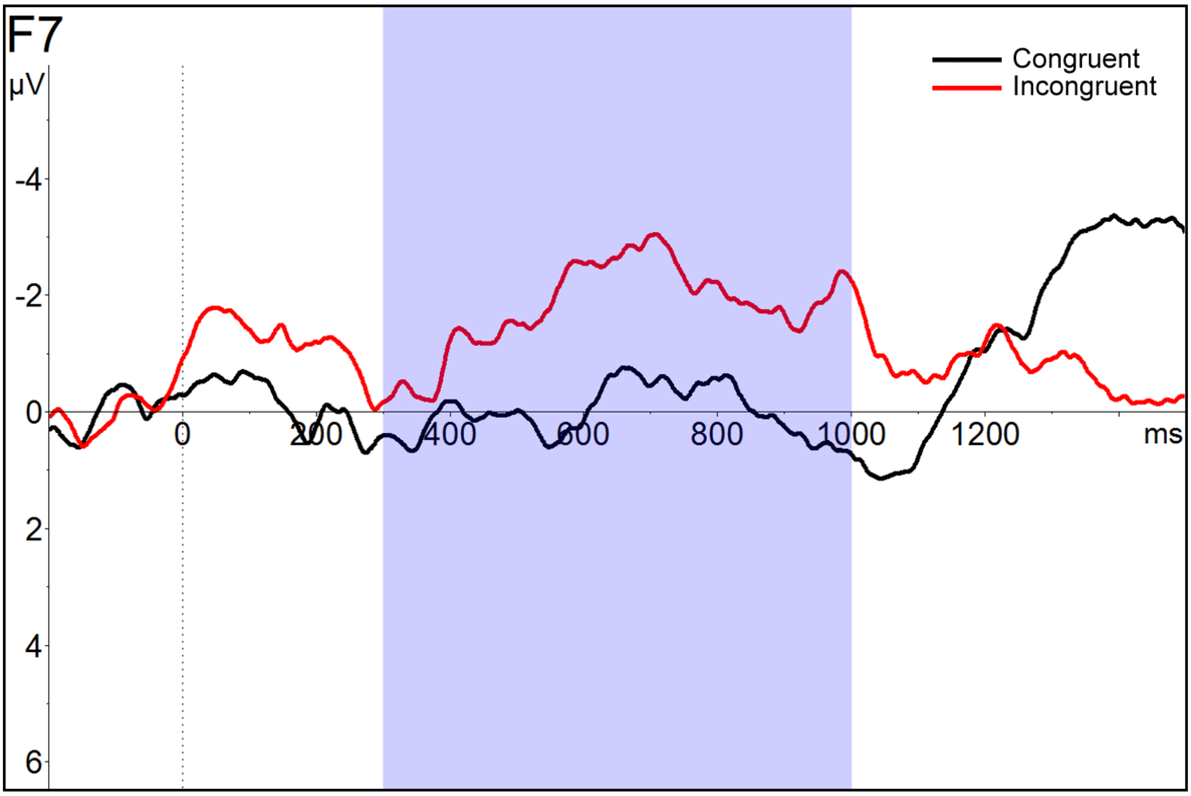


Figure 6.1.11. Block 3: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode F7.

Block 4 (Task: passive listening; Stimuli: delexicalized)

For the lateral electrodes, neither a significant main effect of Congruency nor significant interactions were found [*Fs* < 1]. For the midline electrodes, results revealed a significant main effect of Congruency [F (1, 16) = 13.071, MSE = 0.439, $p < 0.01$, $\eta 2_p = 0.45$], indicating that a larger negativity in incongruent condition than in congruent condition. Furthermore, a significant Congruency × Electrode interaction was found [F (2, 32) = 3.40, MSE = 0.538, $p < 0.05$, $\eta 2_p = 0.18$]. Post hoc analyses revealed a significant sustained negativity effect at the electrode Cz [F (1, 16) = 20.22, MSE = 0.238, $p < 0.001$, $\eta 2_p = 0.56$] and at the electrode Pz [F (1, 16) = 9.63, MSE = 0.349, $p < 0.01$, $\eta 2_p = 0.38$]. Figure 6.1.12 shows the ERP data at the Cz.

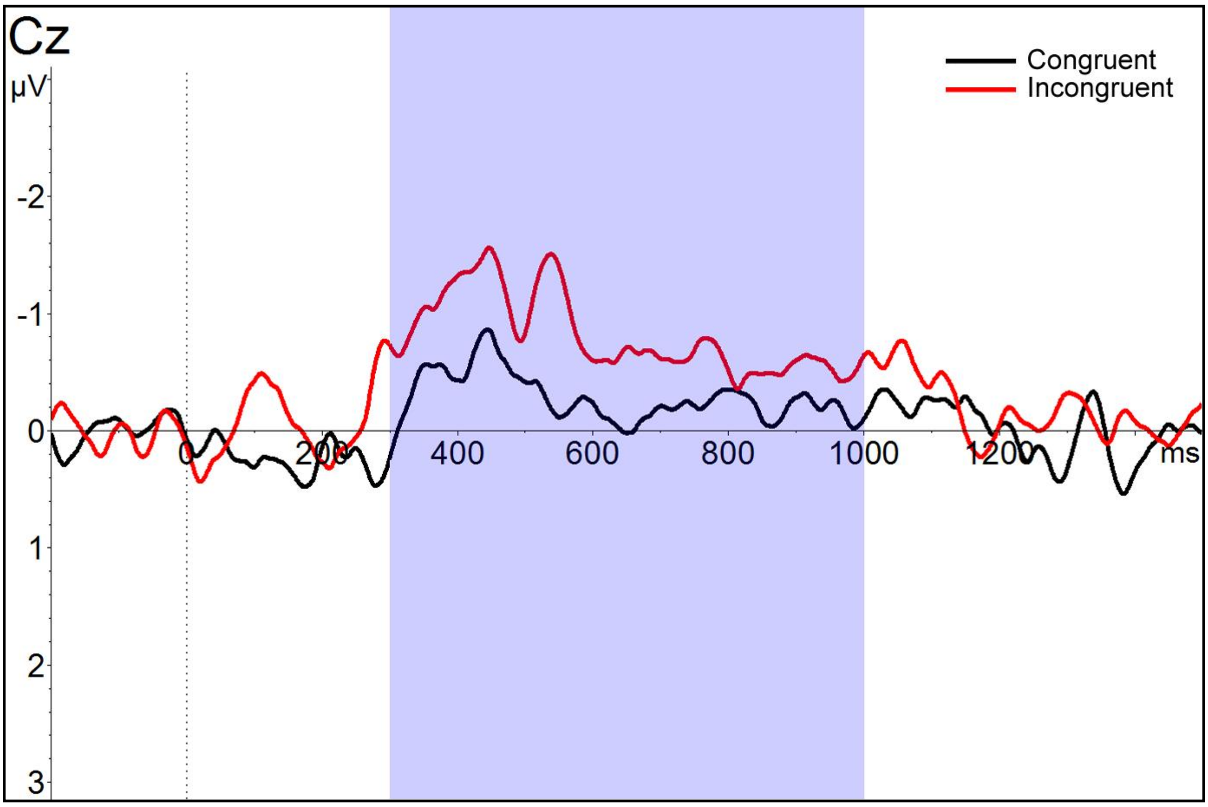


Figure 6.1.12. Block 4: The ERP in the two conditions (i.e., strong-weak, strong-strong) at the electrode Cz.

6.1.3. Discussion

The present experiment investigated ERPs correlate of the rhythmic property of Mandarin Chinese neutral vs. full tone. A set of 60 word pairs of strong-weak (with a neutral tone at the final syllable) vs. strong-strong stress pattern were used in an oddball-like paradigm. Participants performed an explicit task (stress discrimination) and an implicit task (passive listening). For the purpose to investigate whether the processing of stress pattern relies on lexical information, we created also a set of delexicalized stimuli by filtering lexical information of stimuli.

Behavioral data showed shorter response times for infrequent strong-weak words compared to frequent strong-strong words in the discrimination task. One possible interpretation for accounting for this effect is that strong-weak words (628.8 ms averaged all stimuli) is shorter in duration than strong-strong words (933.1 ms averaged all stimuli);

moreover, the compound word with a strong-weak stress pattern, which is rare in Chinese in comparison with the strong-strong stress pattern may have better captured the attention of the participants. Interestingly, strong-weak words led to a higher error rate than strong-strong words. Taken together, the RT data indicated a trade-off effect between rapidity and performance: For the rare Strong-Weak stress pattern, participants were faster to respond but did more errors. The ERP data revealed two exogenous components (N200, P200) and an endogenous component (N350) related to Mandarin Chinese stress pattern processing. Moreover, the MMN was found to be associated to stress pattern congruency.

Exogenous stress effect: N200

The N200 effect was larger for strong-weak words than for strong-strong ones in both stress discrimination and passive listening tasks with no size difference between the explicit and the implicit task. We considered the N200 to be a task-independent exogenous stress effect. Generally, the N200 was linked to unattended deviant auditory stimuli in oddball paradigms. Our observed N200 had a frontocentral topography, the same as the one in oddball paradigms (Näätänen et al., 1978; Näätänen & Gaillard, 1983). We recorded as well the ERP for the delexicalized stimuli, in which lexical information was filtered with the conservation of stress pattern information. Results on the delexicalized stimuli did not show the N200. We argued that the N200 observed on non-delexicalized stimuli was not just a response to unattended deviant auditory stimuli, thus it was language specific in response to the processing of stress pattern. Cutler and Norris (1988) proposed a Metrical Segmentation Strategy (MSS) for speech segmentation postulating an important role of the stress information. The model suggested boundaries at the onset of strong syllables. The segmentation process was presumed to occur at prelexical stage in word recognition. The model can explain the findings that participants prefer identifying a stimulus as two words when it has a weak-strong stress pattern and as one word when it has a strong-weak stress pattern (van Heuven, 1985; Taft, 1984). The N200 is probably an ERP correlate to the prelexical segmentation process. We suggest that prelexical and language-specific representations of stress pattern information reflected by the N200 are may be stored as ‘stress pattern templates’ in mental lexicon. These templates help to categorize rapidly and accurately speech information into different stress patterns, which is an important process in accessing the mental lexicon.

The N200 amplitude on strong-weak words was larger than on strong-strong words in both stress discrimination and passive listening tasks with no amplitude difference between the explicit and the implicit task. We considered the N200 to be a task-independent exogenous stress effect. Generally, the N200 was linked to unattended deviant auditory stimuli in oddball paradigms. Our observed N200 had a frontocentral topography, the same as the one in oddball paradigms (Näätänen et al., 1978; Näätänen & Gaillard, 1983). In order to clarify the nature of the N200, we recorded the ERP for the delexicalized stimuli, in which lexical information was filtered with the conservation of stress pattern information (e.g., duration, F0). Results on the delexicalized stimuli did not show the N200. We argued that the N200 observed on non-delexicalized stimuli was not just a response to unattended deviant auditory stimuli, thus it was language specific in response to the processing of stress pattern. Cutler and Norris (1988) proposed a Metrical Segmentation Strategy (MSS) for speech segmentation postulating an important role of the stress pattern information. The model suggested boundaries at the onset of strong syllables. The segmentation process was presumed to occur at prelexical stage in word recognition. The model can explain the findings that participants prefer identifying a stimulus as two words when it has a weak-strong stress pattern and as one word when it has a strong-weak stress pattern (van Heuven, 1985; Taft, 1984). The N200 is probably an ERP correlate to the prelexical segmentation process. We suggest that prelexical and language-specific representations of stress pattern information reflected by the N200 are stored as ‘stress pattern templates’ in mental lexicon. These templates help to categorize rapidly and accurately speech information into different stress patterns, which is an important process in accessing the mental lexicon.

Effect of the processing of the physical/acoustic stimulus parameters: P200

We found a frontocentral P200 effect peaking at 272 ms post-stimulus in response to the stress pattern in delexicalized stimuli in both tasks. The amplitude of the P200 was larger for infrequent strong-weak than frequent strong-strong stress pattern. The P200 component is an exogenous component assumed to reflect perceptual processing (Hillyard & Picton, 1987). The P200 has been found sensitive to acoustic properties (Marie et al., 2011; Friedrich et al., 2001; Böcker et al., 1999). Marie et al. (2011) manipulated the metric structure of French words by lengthening the penultimate syllable of trisyllabic words, and found metrically incongruous words elicited larger P200 components than metrically congruous words. Furthermore, the P200 effect was larger for musicians than for non-musicians. It is proposed that the amplitude enhancement of the P200 reflected the automatic processing of the temporal attributes of words and, specifically, of the words' syllabic structure. The P200 in response to stress pattern information has been also observed with German words, Friedrich et al. (2001) found that German infrequent weak-strong words elicited a larger P200 than frequent strong-weak words. Although the P200 component is of exogenous, it has been shown to be sensible to task, Böcker et al. (1999) revealed the P200 effect only in the passive listening task. Our P200 in Experiment 1 was found in both explicit and implicit tasks but only for delexicalized stimuli. Böcker et al. (1999) proposed that the P200 was an exogenous component that reflects the acoustic stimulus parameters. Our data supported this viewpoint of the P200 with evidence of delexicalized stimuli.

Endogenous stress effect: N325

We replicated the endogenous stress effect N325 in the study of Böcker et al. (1999), with the same frontocentral topography of surface. However, in our Experiment, the peak latency occurred later (i.e. at 405 ms post-stimulus onset) than in the Böcker et al. study. This might be due to a longer duration of the first syllable in our stimuli (393.90 ± 74.20 ms) than in the stimuli used by Böcker et al. (1999) (292 ± 57 ms). We had thus about 100 ms more duration in the first syllable of our stimuli probably resulting in the delay of the N325. Böcker et al. (1999) showed that the N325 would be influenced by task, the amplitude de N325 was attenuated in passive listening task, which suggested the endogenous nature of the effect. In our experiment with Chinese stimuli, we observed the N325 only presented in stress discrimination task for non-delexicalized stimuli. It was not found for the implicit passive listening task, or for delexicalized stimuli. The results suggested that the N325 was language-specific (non-language stimuli would not evoke the N325) and it need listeners to pay attention to the stress pattern for enabling the emergence of the N325, at least with the stress patterning of Chinese language. Unlike the exogenous N200 and P200 stress effect reflecting automatic stress pattern processing (stress pattern templates matching in mental lexicon or processing of the physical/acoustic stimulus parameters), the N325 is of controlled process reflecting the perception of the stress pattern following by of the automatic stress pattern processing indexed by the N200 and P200. More precisely, when attention was paid to stress pattern, the automatic stress pattern processing would be used to produce a perception of stress pattern reflected by the N325; when attention was not paid, the automatic stress pattern processing would be rapidly eliminated, a perception of stress pattern would then not be drawn, thus no N325 in an implicit task.

Congruency effect: MMN

The mismatch negativity (MMN) was expected for targets in incongruent trials regardless of stress pattern and task. In all four blocks, with each of the two tasks and each of two types of stimuli, a late sustained negative-going potential between 300 and 1000 ms was consistently found. We argued that this negativity was of the MMN. Firstly, the negativity represented automatic process because of the observation for implicit task. Secondly, it was not sensible to the nature of stimuli, language and lexicalized non-language stimuli both evoked this effect. Finally, this negativity peaked around 600 ms after the onset of the stimulus, too late for a MMN, however, if we subtracted the duration of the first syllable (393.9 ms averaged of all stimuli), the latency of the negativity matched the typical latency of MMN (150-250 ms). Listeners received critical information from the onset of the second syllable to determinate the stress pattern. It is thus reasonable to subtracted the duration of the first syllable from the latency of the negativity. The late MMN reflected indeed the stress pattern congruency effect in the *oddball*-like paradigm. This finding suggests that stress pattern changes are processed relying on long-term representation of stress, not at single syllable level.

6.1.4. Conclusion

The present experiment investigated ERPs correlate of the rhythmic property of Mandarin Chinese. The ERP data revealed two exogenous (N200 & P200) and an endogenous (N325) components related to Chinese stress pattern processing. Moreover, an MMN was found associated to stress pattern congruency. The N200 was larger for strong-weak stress pattern with a neutral tone (which constitute only 5% of the Mandarin Chinese lexicon) than for frequent strong-strong pattern. It was proposed to reflect automatic stress pattern processing. The P200 was larger for infrequent strong-weak than frequent strong-strong stress pattern reflecting the acoustic stimulus parameters. The followed endogenous N325 was presented if listeners paid attention to stress pattern processing, resulting in the perception of the stress pattern as we only reported an N325 variation in the stress discrimination task with non-delexicalized stimuli. Moreover, the incongruency in the stress pattern elicited the MMN in both explicit and implicit tasks and for both non-delexicalized and delexicalized stimuli. Taken together, findings of Experiment 1 allowed us to provide ERP evidence of the processing of stress information in Mandarin Chinese.

The next experiment (Experiment 2) examines in detail the influence of stress information in compound words recognition, more particularly to which extends the stress information interacts with the lexical processing, such as semantic access of morphemes in compound word. For this purpose, we conducted an ERP study using the semantic priming paradigm with a lexical decision task to investigate how stress information modulates semantic priming effect.

6.2. Experiment 2: Semantic/associative priming on neutral tone words

Spoken-word recognition involves the processing of the information about its constituting phonological segments, but also about a specific prosodic patterning, realized via suprasegmental parameters such as pitch contour, duration, amplitude, and consonant and vowel reduction/strengthening. Previous studies have shown important role of stress information in different aspects during spoken word recognition: 1) correct word stress facilitates the identification of phonemes (Connine et al., 1987; Haggard et al., 1970); 2) mis-stressing syllables disturb the recognition of words (Slowiaczek, 1990; Small et al., 1988; Koester & Cutler, 1997); 3) prosodic information assists compound words recognition (Isel et al., 2003; Koester et al., 2004); 4) stressed syllables play a critical role in the segmentation of the speech signal (Cutler & Norris, 1988; van Heuven, 1985; Taft, 1984).

In the present doctoral thesis, Experiment 1 has revealed an online processing of prosodic patterning information for Chinese strong-weak (full + neutral tone) and strong-strong (full + full tone) words. The ERP data indicated an early processing of stress pattern at prelexical stage indexed by the N200 and P200 and then the perception of the stress pattern reflected by the N325. Experiment 2 investigated to which extend the stress information interacts with the morpholexical information in spoken Chinese compound word.

The study of the processing of neutral tones in final syllables of compound words is of central interest regarding the functional architecture of left-to-right sequential models postulating an early interaction between prosodic cues and morphological information. Indeed, in the case where the neutral tone is carried by the final syllable of a compound word in Chinese, this prosodic information is not immediately available at the beginning of the word. Therefore, a left-to-right sequential processing model of compound word recognition that assumes an early involvement of prosody for assisting morphological analysis at the lexical level would not be suitable to process grammatical sequences that deliver critical prosodic information at the end of the word. This could have a dramatic consequence on lexical access and more generally on language comprehension in Chinese, particularly when meaning of the compound word has to be reanalyzed once a neutral tone was detected on the last syllable of the compound word. Experiment 2 addresses this issue. Specifically, we aimed to investigate to which extend processing of a neutral tone on a last constituent will modulate the analysis of the semantic information carried by the first constituent. Indeed, in some special Chinese disyllabic sequences the neutral tone and the semantic transparency of the first constituent have a particular relation. These sequences differ from each other only in the stress pattern resulting in two morpholexical interpretations: 1) phrasal interpretation, with strong-strong stress pattern, in which the right syllable carried a full tone and the first morpheme is semantically transparent; or 2) compound word interpretation, with strong-weak stress pattern whose right syllable carried a neutral tone and the first constituent was semantically opaque. For example, the disyllabic sequence *dong-xi* 'east-west' consists of two morphemes, *dong* (means east) and *xi* (means west). When the second constituent carries a neutral tone (i.e., strong-weak), the sequence is a compound word, moreover the first constituent becomes semantically opaque (the meaning of the whole compound is 'thing', the first constituent does not contribute to the whole word meaning); when the second constituent carries a full tone (i.e., strong-strong), the sequence is a phrase (it means the east and the west), the first constituent is semantically transparent (it means the east).

These sequences are excellent stimuli for studying the interaction between stress information and morpho-lexical processing, because the semantic propriety of the left morpheme of the sequence should be modulated by the stress information of the last syllable. Some theories on the lexical access to compound word such as sequential continuous access theories claim that the ongoing acoustic-phonetic information is, left to right, continuously mapped onto representations stored in the mental lexicon regardless of the morphological structure of the word (Butterworth, 1983; Marslen-Wilson, 1987). The prosodic information of English or German is available as soon as the words are produced. Previous studies focus on the prosodic role of the initial constituent during the auditory lexical access of compound words. Isel et al. (2003) proposed a Prosody-Assisted Processing (PAP; Isel et al., 2003) model for compound auditory words processing. The central assumption on which the model relies is that the length of the left morpheme of compound words is a determining factor for setting off/triggering a decompositional route. For Chinese, it is also true that the prosodic information delivered at the beginning of the words also plays important role in compound words processing. However, in Experiment 2, the stress information for our selected sequences was available at the last syllable of the sequences and the stress information modulated the semantic propriety of the first morpheme. It involves thus a right-to-left retroactive processing of prosodic information with a critical importance of the stress information of the last morpheme.

We used cross-modal semantic/associative priming paradigm in which, a participant was presented with an *auditory* stimulus (i.e., the selected sequences) as the prime; at some point during the auditory presentation of the prime, a *visual* target that was either semantically related or unrelated to the left morpheme appears on a screen, and the participant performed a lexical decision task on the target, that means they had to decide whether the target was a word or not. We manipulated systematically the temporal position of the onset of the target. Two positions were used: 1) initial position, i.e., at the onset of the auditory prime where the critical stress information in the right syllable of the prime was not yet available during the processing of the target; 2) middle position, i.e., at the onset of the second syllable of the prime where the stress information in the second syllable became available during the processing of the target. We hypothesized that the availability of the stress information would modulate the semantic/associative priming effect. Particularly, for strong-weak compound (the first morpheme was semantically opaque) the semantic/associative priming effect would occur when the target word was presented at the initial position, however, when it was presented at the middle position of the prime the semantic/associative priming effect would disappear, as the stress information deactivated the semantic access of the first morpheme which became thus semantically opaque; for strong-strong phrase (the first morpheme was semantically transparent), the priming effect would emerge for the two presentation positions.

The semantic/associative priming effect has been linked to the N400 peaking around 400 milliseconds post-stimulus onset, typically maximal over centro-parietal electrode sites (Bentin et al., 1985; Kutas & Van Petten, 1994; Kutas & Federmeier, 2011). A number of studies have reported reduced N400 amplitudes to target words following semantically related as compared to unrelated primes. We expected thus to find the N400 to be related to semantic priming effect that semantically unrelated target words would evoke larger amplitude of the N400 than related ones, the larger N400 reflected difficulty in the lexical access of the target word.

6.2.1. Method

6.2.1.1. Participants

Twenty native Mandarin Chinese participants (12 females, 24.32 $\pm$ 4.03 years, ranging from 19 to 31 years) in Paris were recruited. None of them has lived in France more than five years (2.2 $\pm$ 1.61 years, ranging from 0.2 to 5 years). They reported to use frequently the Chinese language in their daily life. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). They have normal or corrected to-normal vision and no auditory impairment. They have no history of current or past neurological or psychiatric illnesses. They were paid 10€ per hour for their time. All participants gave informed consent before participation and completed an information sheet.

6.2.1.2. Stimuli

A set of 20 Chinese disyllabic sequences (see Appendix B) was selected as auditory primes. Each sequence can be understood in two ways: 1) compound word, whose second syllable carried a neutral tone and the first constituent was semantically opaque; 2) phrase, in which the second part carried a full tone and the first part was semantically transparent. An example of a disyllabic sequence *dong-xi* is presented in Table 6.2.1.

Table 6.2.1.

The two interpretations of the disyllabic sequence dong-xi

Type	Stress pattern	Meaning	Tone of C2	Semantic of C1
Compound	dong1xi0 = strong-weak	Thing	Neutral tone	Opaque
Phrase	dong1xi1 = strong-strong	East and west	Full tone	Transparent

The sequence *dong xi* consists of two morphemes, C1 = *dong* (means *east*) and C2 = *xi* (means *west*). When the second constituent carries a neutral tone (i.e., strong-weak), the sequence is a compound word, moreover the first constituent becomes semantically opaque (the meaning of the whole compound is 'thing', the first constituent, *dong* (means *east*), does not contribute to the whole word meaning); when the second constituent carries a full tone (i.e., strong-strong), the sequence is a phrase (means 'east and west'), the first constituent, *dong* (means *east*) is semantically transparent as it does contribute to the meaning of the phrase.

All primes sequences were recorded in an acoustic chamber by a native Mandarin Chinese speaker (female, 22 years old, from Beijing). Phonetic analysis on the final syllable of the prime sequences were conducted to examine the contrastive stress pattern in the neutral tone syllable and the full tone syllable. Paired t-tests indicated that 1) syllable duration was shorter [t (19) = 9.16, $p < 0.001$] for neutral tone syllables (293.09 ms ± 66.39 ms) than for full tone syllables (502.99 *ms* ± 92.01 *ms*); 2) average intensity was lower [t (19) = 2.93, $p < 0.01$] for neutral tone syllables (60.55 dB ± 3.14 dB) than for full tone syllables (63.38 dB ± 5.24 dB); and 3) average F0 was lower [t (19) = 2.52, $p < 0.05$] for neutral tone syllables (258.46 Hz ± 42.34 Hz) than for full tone syllables (297.58 Hz ± 48.68 Hz). Results on the first syllable only showed that the duration was shorter [t (19) = 2.42, $p < 0.05$] when the syllable was followed by a neutral tone (497.65 ms ± 116.72 ms) than by a full tone (542.88 ms ± 79.42 ms). The results replicated the acoustic analysis on the materials used in Experiment 1 produced by another speaker.

Twenty words that presented an associative relation to the C1 and twenty control words that did not present this relation were selected as visual targets. To determine the semantic associates of the first constituents, we conducted a pretest of semantic association with 12 participants. Participants read each of the twenty C1 and were instructed to give as rapidly as possible the first word that came to mind (the word should start with the given C1). The most frequent word produced by the twelve participants for each C1 was selected as the associative word of the C1. Moreover, the strike number and the frequency for targets were controlled. Associative targets and non-associative targets were matched on the of strike number [16.1 ± 5.1 strikes in overall average; t (19) = 0.63, $p > 0.05$] and on the frequency (95.2 occurrences per million; t (19) = 0.10, $p > 0.05$). For constructing the lexical decision task design, an hundred and twenty pseudowords were randomly composited from two Chinese characters. All pseudowords were classified as meaningless by two native Chinese.

6.2.1.3. Design and Procedure

Each trial started with a visual fixation at the center of a computer monitor for 500 ms, this was followed by an auditory prime. The onset of the visual presentation of the target was placed at two positions in the prime: 1) at the onset of the prime (see Figure 6.2.1a), or 2) at the middle of the prime, i.e., onset of the C2 (see Figure 6.2.1b).

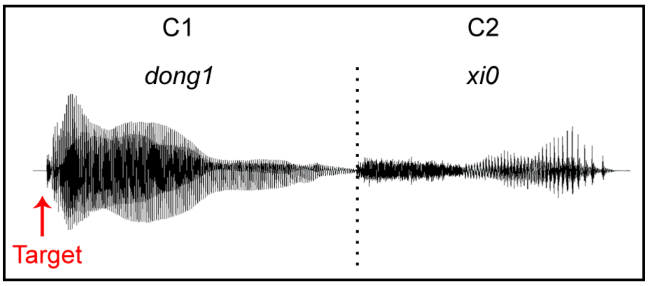


Figure 6.2.1a. Presentation of the visual target at the onset of the onset of the prime.
Prime: *dong1xi0* ‘thing’, C1: *dong1*, C2: *xi0*.

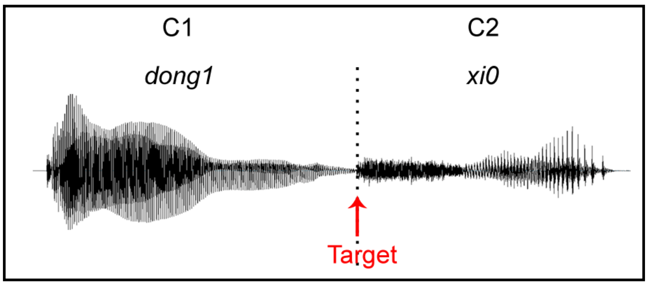


Figure 6.2.1b. Presentation of the visual target at the onset of C2.
Prime: *dong1xi0* ‘thing’, C1: *dong1*, C2: *xi0*.

The target word was presented for 300 ms. Participants were instructed to indicate as accurately and quickly as possible whether the visual target was a word or a pseudoword. The next trial began 2000 ms after the participant gave his response. Twelve practice trials were presented prior to the experimental trials. Each experimental session lasted about 90 minutes. Stimulus delivery and behavioral response recording were controlled by E-prime (Psychology Software Tools, Inc., Pittsburgh, PA; http://www.pstnet.com/eprime). Factors Type of stimuli (compound, phrase), Position of target (onset of C1, onset of C2) and Semantic association (related, unrelated and pseudoword) created twelve experimental conditions (see Table 6.2.2).

Table 6.2.2.

Experimental conditions

Condition	Type of stimuli	Target position	Semantic association
1	Compound (SW)	C1 onset	Related
2	Compound (SW)	C1 onset	Non-related
3	Compound (SW)	C1 onset	Pseudoword (filler)
4	Compound (SW)	C2 onset	Related
5	Compound (SW)	C2 onset	Non-related
6	Compound (SW)	C2 onset	Pseudoword (filler)
7	Phrase (SS)	C1 onset	Related
8	Phrase (SS)	C1 onset	Non-related
9	Phrase (SS)	C1 onset	Pseudoword (filler)
10	Phrase (SS)	C2 onset	Related
11	Phrase (SS)	C2 onset	Non-related
12	Phrase (SS)	C2 onset	Pseudoword (filler)

Note: Compounds carry SW (strong-weak) pattern, phrases carry SS (strong-strong) pattern.

Each condition contained 30 trials presented in a pseudo-randomized order in mixed blocks. A number of constraints were used for creating the pseudo-randomization. First, prime was not repeated immediately. Secondly, no more than three targets of the same condition were presented in succession. Lastly, no more than three same responses were allowed. The pseudo-randomized order of trials was created using the program Conan (Nowagk, 1998).

6.2.1.4. ERP Recordings and Analysis

The EEG was recorded from 64-channel electrodes using Brain Vision Recorder (Brain Products GmbH, Gilching, Germany). Data was sampled at 1000 Hz with an online 0.1-100 Hz frequency bandpass filter. All electrode impedances were maintained below 5 kΩ. Further processing of ERPs was done off-line. First, EEG data were re-referenced to an average reference, and then were filtered with a 0.1-12 Hz bandpass filter. Next, the continuous EEG was segmented into epochs from 200 ms pre-stimulus until 1500 ms post-stimulus onset, with 200 ms of each epoch corresponding to a pre-stimulus baseline. Then, a semi-automatic artifact detection procedure was applied. Artifacts were marked when 1) maximal amplitude was higher than 100 μV or minimal amplitude was lower than -100 μV; 2) absolute difference was more than 50 μV or less than 0.5 μV in an interval of 200 ms. Only trials with correct responses in the behavioral lexical decision task and with no artifacts were kept for further data analysis. In each experimental condition, the segmented ERP data was then averaged over trials and over participants (i.e., grand average). Finally, the ERP data were baseline-corrected to the mean amplitude of the 200 ms pre-stimulus interval.

Statistical analyses were conducted for the N400 for which the time windows were selected based on previous studies adapted by visual inspection of the grand averages: N400 (250-500). Based on previous the visual inspection of the grand averages and previous studies on the semantic N400, we focused the analyses of the N400 on central electrodes. For each four experimental condition, i.e., (1) compound prime and target at C1 onset, (2) compound prime and target at C2 onset, (3) phrase prime and target at C1 onset, and (1) phrase prime and target at C2 onset. A two-way repeated measures ANOVA with the factors Semantic association (related, non-related, pseudoword[38]) and Electrode (Fz, Cz, Pz) was conducted. The Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied when evaluating effects with more than one degree of freedom in the numerator. Post-hoc pairwise comparisons at single electrode sites were made using a modified Bonferroni procedure (Keppel, 1991).

6.2.2. Results

6.2.2.1. Behavioral Results

For error rate, a three-way repeated measures ANOVA with factors of Type of prime (compound, phrase), Position of target (onset of the prime, middle of the prime) and Semantic association (related, non-related, pseudoword) revealed a main effect of Semantic association [F (2, 38) = 4.23, MSE = 0.003, $p < 0.05$, $\eta 2_p = 0.18$]. Post-hoc analyses showed that error rate was higher in Pseudoword condition (3.3 ± 4.0%) than in Related condition (1.2 ± 2.2%) or in Unrelated condition (1.8 ± 2.6%). Moreover, the error rate was higher in Unrelated condition (1.8 ± 2.6%) than in Related condition (1.2 ± 2.2%). For response times, ANOVA revealed a main effect of Semantic association [F (2, 38) = 8.67, MSE = 38720.191, $p < 0.01$, $\eta 2_p = 0.31$]. Post hoc analyses indicated that the response times were slower for unrelated targets (598.0 ± 161.3 ms) than related targets (589.5 ± 159.7 ms) and the pseudoword showed a slower response times (673.9 ± 169.9 ms) than related or unrelated targets. Moreover, a significant Semantic association × Type interaction [F (2, 38) = 4.65, MSE = 2105.656, $p < 0.05$, $\eta 2_p$ = 0.20] was observed, reflecting that for Phrase prime, unrelated targets had a slower response times (599.1 ± 165.4 ms) than related targets (574.3 ±163.5 ms). However, for Compound prime, there was no significant difference between the response times for unrelated targets (596.9 ± 157.3 ms) and related targets (604.6 ±155.9 ms). Moreover, a significant interaction between Semantic association and Position [F (2, 38) = 9.23, MSE = 1552.588, $p < 0.001$, $\eta 2_p$

[38] Pseudoword served as fillers in the present study. However, we put the pseudoword condition into the analyses as a reference of the N400.

= 0.33]. Post hoc analyses revealed that for the target position at the onset of prime condition, unrelated targets had a slower response times (622.59 ± 175.9 ms, p < 0.001) than related targets (589.9 ±166.3 ms). However, for the target position at the middle of prime condition, unrelated targets had a faster (marginally significant, p = 0.057) response times (573.4 ± 1146.8 ms) than related targets (589.0 ±153.1 ms). Finally, a marginally significant interaction Semantic association × Type × Position [F (2, 38) = 2.94, MSE = 1070.456, p = 0.073, $\eta 2_p$ = 0.13]. Post hoc analyses revealed that 1) when the target was presented at the onset of the prime, a priming effect for Compound prime condition [27.21 ±7.23 ms, $p < 0.01$] as well as for Phrase prime condition [38.18 ±10.37 ms, $p < 0.01$] (see Figure 6.2.2); 2) when the target was presented at the middle of the prime, an inverse priming effect for Compound prime condition [42.66 ±7.61 ms, $p < 0.001$] was observed (see Figure 6.2.3).

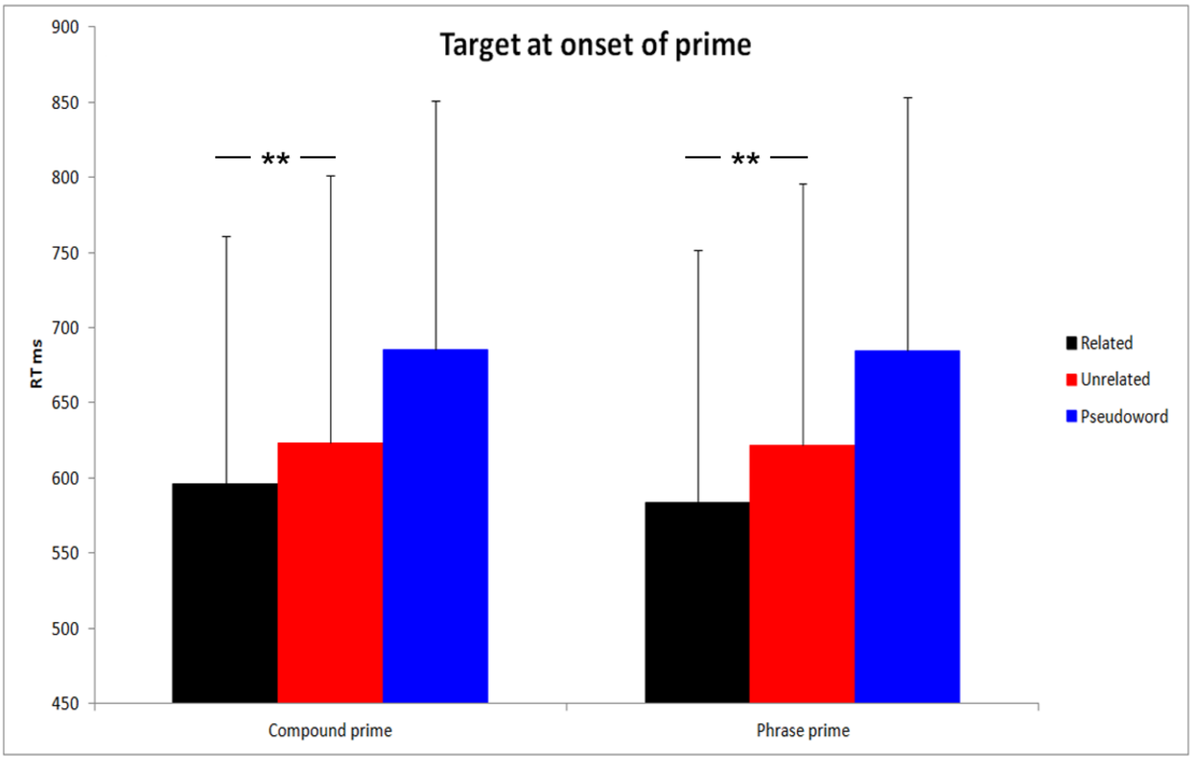


Figure 6.2.2. Mean response times in Related, Unrelated and Pseudoword conditions for Compound prime and for Phrase prime and the target at the onset of prime.
Note: ** $p < 0.01$.

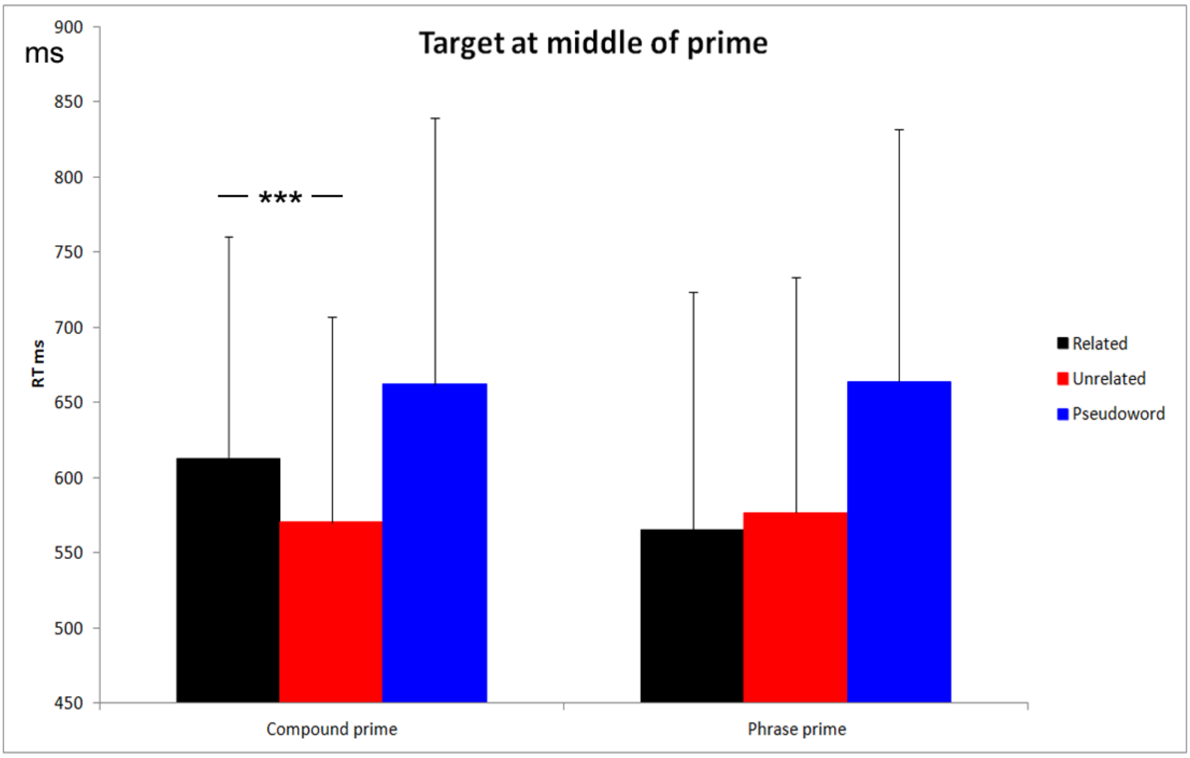


Figure 6.2.3. Mean response times in Related, Unrelated and Pseudoword conditions for Compound prime and for Phrase prime and the target at the middle of prime.
Note: *** p < 0.001.

6.2.2.2. Electrophysiological Results

Time window 250-500 ms (N400)

Statistical analysis of the EEG data was conduct in each four experimental conditions (see Table 6.2.3) in time window 250-500 ms.

Table 6.2.3.

Type of prime and Position of target in the four experimental conditions

Condition	Type of prime	Position of target
1	Compound $C1_{full}$ - $C2_{neutral}$	Onset of prime
2	Compound $C1_{full}$ - $C2_{neutral}$	Middle of prime
3	Phrase $C1_{full}$ - $C2_{full}$	Onset of prime
4	Phrase $C1_{full}$ - $C2_{full}$	Middle of prime

Condition 1: Prime $C1_{full}$ - $C2_{neural}$ (Compound) and Target at the onset of C1

Figure 6.2.4 displays the ERP data. A two-way repeated measures ANOVA on midline electrodes with the factors Semantic association (related, non-related, pseudoword) and Electrode (Fz, Cz, Pz) showed a main effect of Semantic [F (2, 38) = 19.21, MSE = 5.944, $p <$ 0.001, $\eta 2_p$ = 0.50]. Post hoc analyses revealed a larger N400 for unrelated targets (-1.18 ± 1.36 μV, $p < 0.05$) than for related targets (-0.63 ± 1.43 μV) at the Cz, as well as at the Pz with larger N400 for unrelated targets (0.78 ± 1.27 μV) than related targets (1.35 ± 1.23 μV), however, the main effect of Semantic at the Fz was due to a larger N400 between pseudowords (-1.79 ± 2.30 μV, $p = 0.053$) and related targets (-0.77 ± 1.38 μV). There was no significant interaction Semantic × Electrode [$F < 1$].

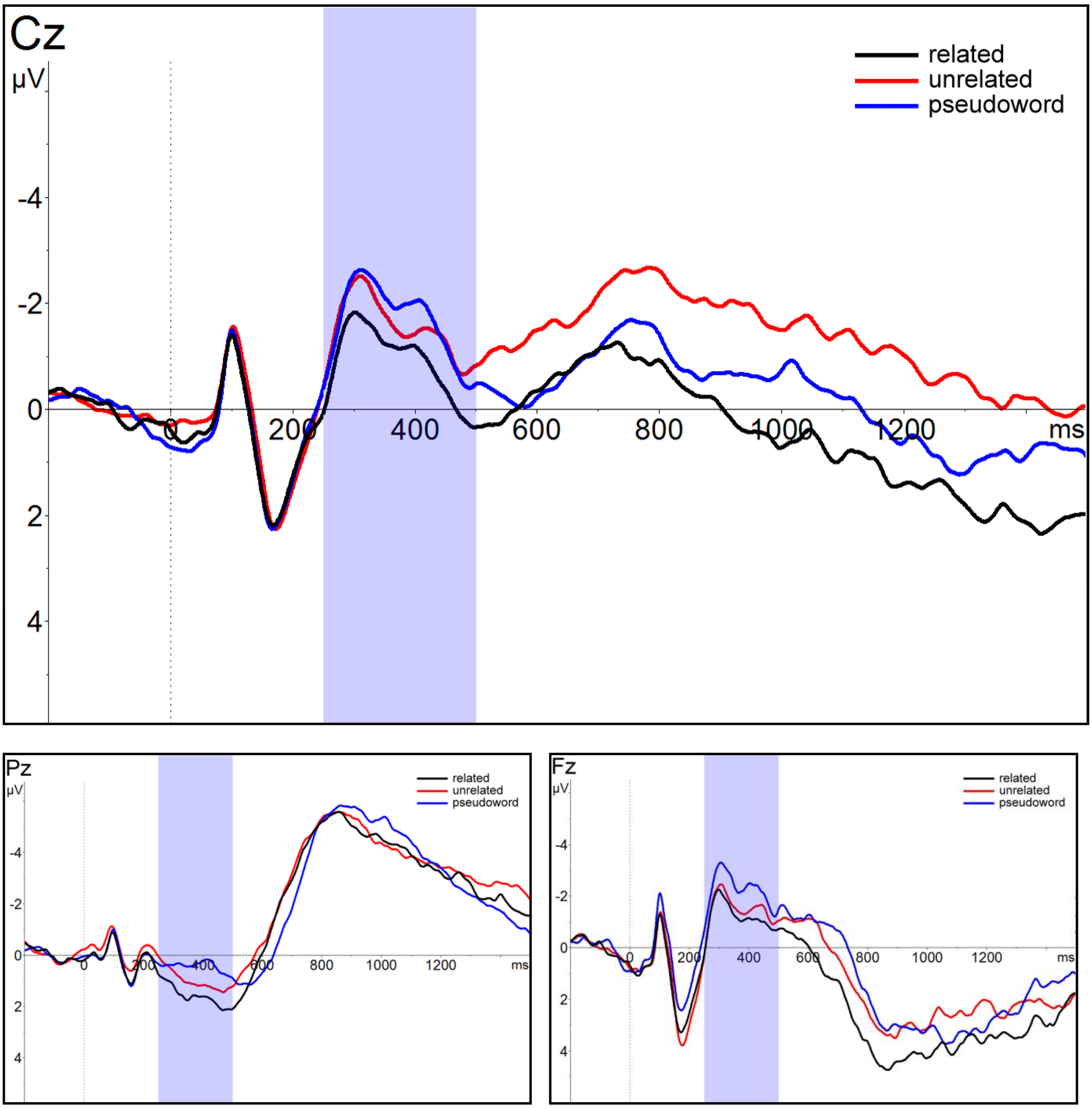
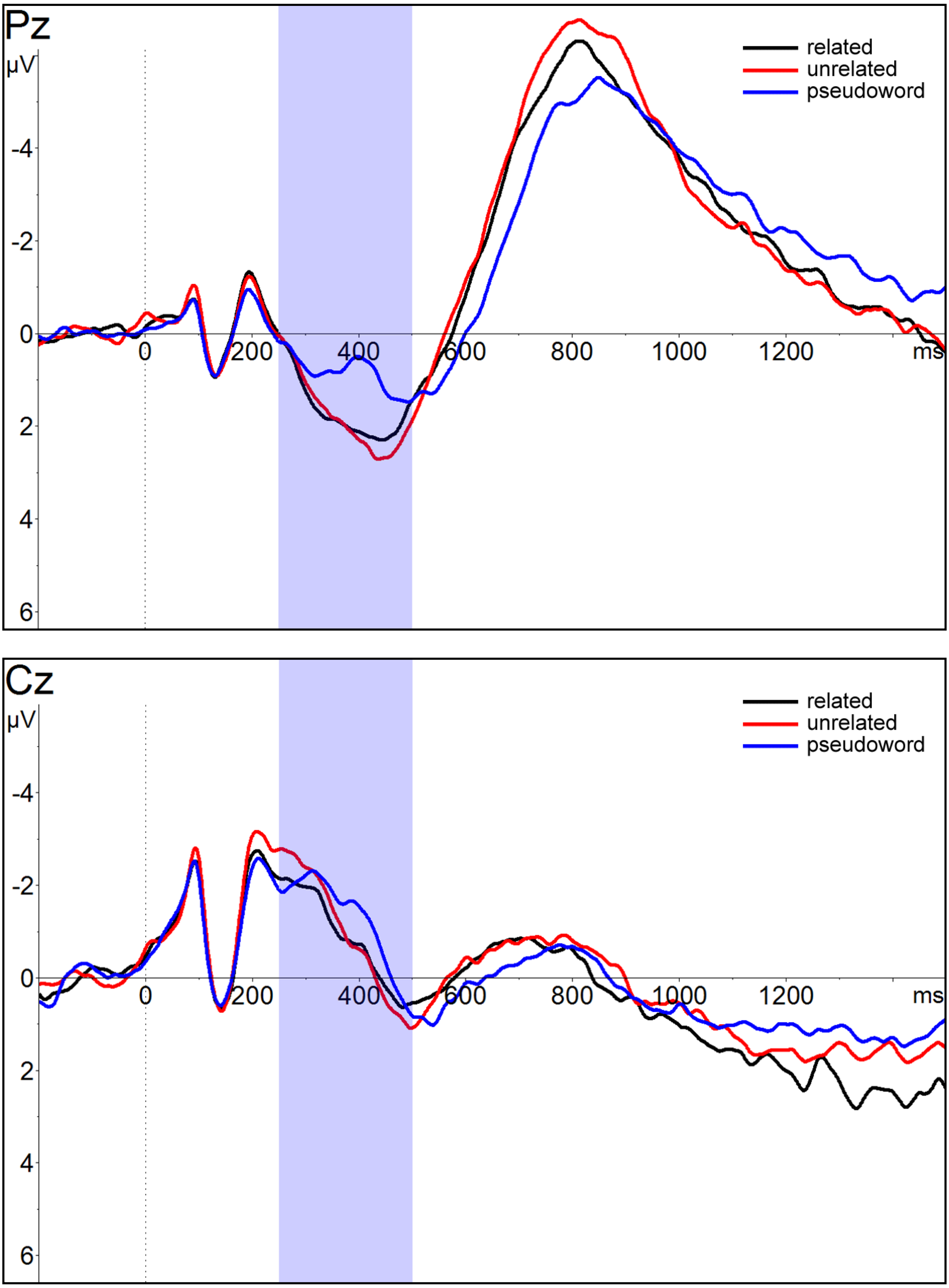


Figure 6.2.4. Condition 1: The ERP in the three conditions (i.e., related, unrelated, pseudoword) at the electrode Fz, Cz, Pz.

Condition 2: Prime C1$_{\mathbf{full}}$ - C2$_{\mathbf{neural}}$ (Compound) and Target at the onset of C2

Figure 6.2.5 displays the ERP data. ANOVA revealed a main effect of Stress [F (2, 38) = 17.21, MSE = 13.071, $p < 0.01$, $\eta 2_p = 0.48$]. Post hoc analyses indicated a larger negativity for pseudoword (-0.86 ±1.71 µV) than for related targets (-0.61 ±2.24 µV, $p < 0.001$) and than for unrelated targets (-0.71 ±1.95 µV, $p < 0.001$), however, the N400 effect did not reach significant between related and unrelated targets. Moreover, there was no significant interaction of Semantic × electrode [F (4, 76) = 2.39, MSE = 1.753, $p > 0.05$].


Pz
μV
-4
-2
0
2
4
6
0
200
400
600
800
1000
1200
ms
related
unrelated
pseudoword
Cz
μV
-4
-2
0
2
4
6
0
200
400
600
800
1000
1200
ms
related
unrelated
pseudoword

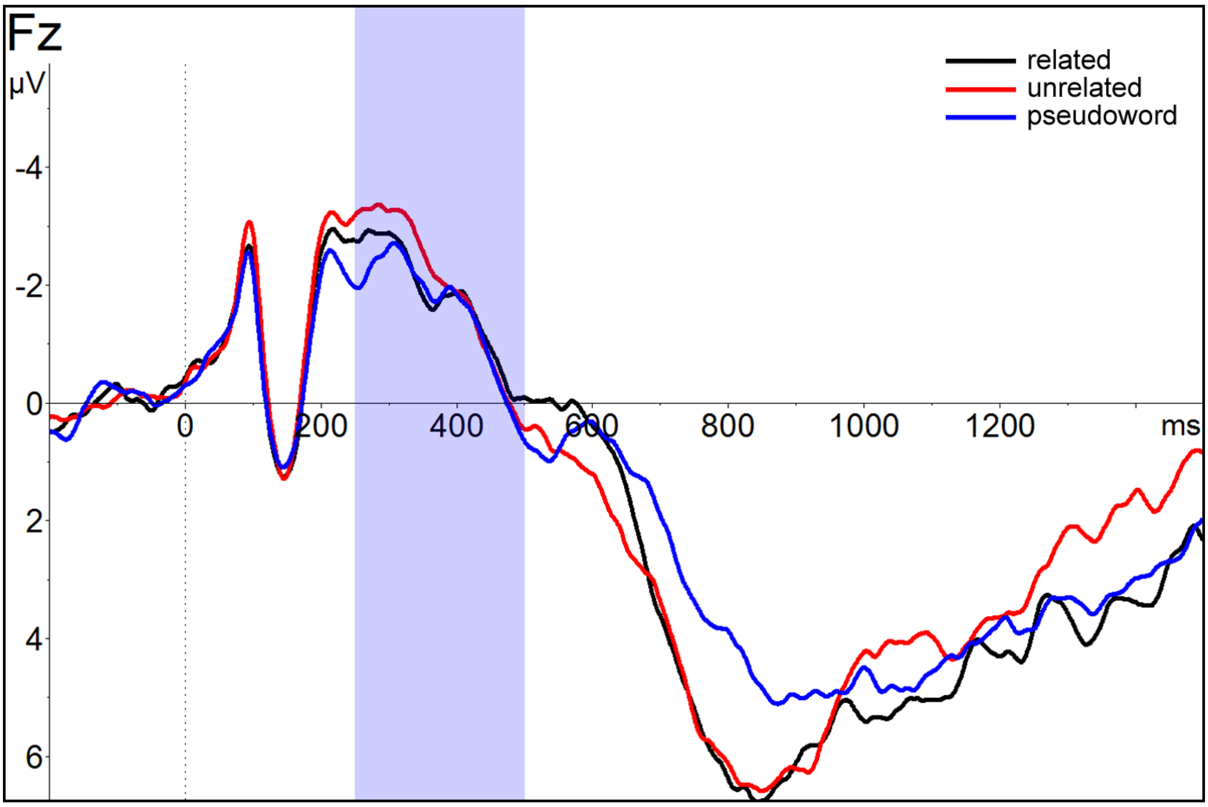
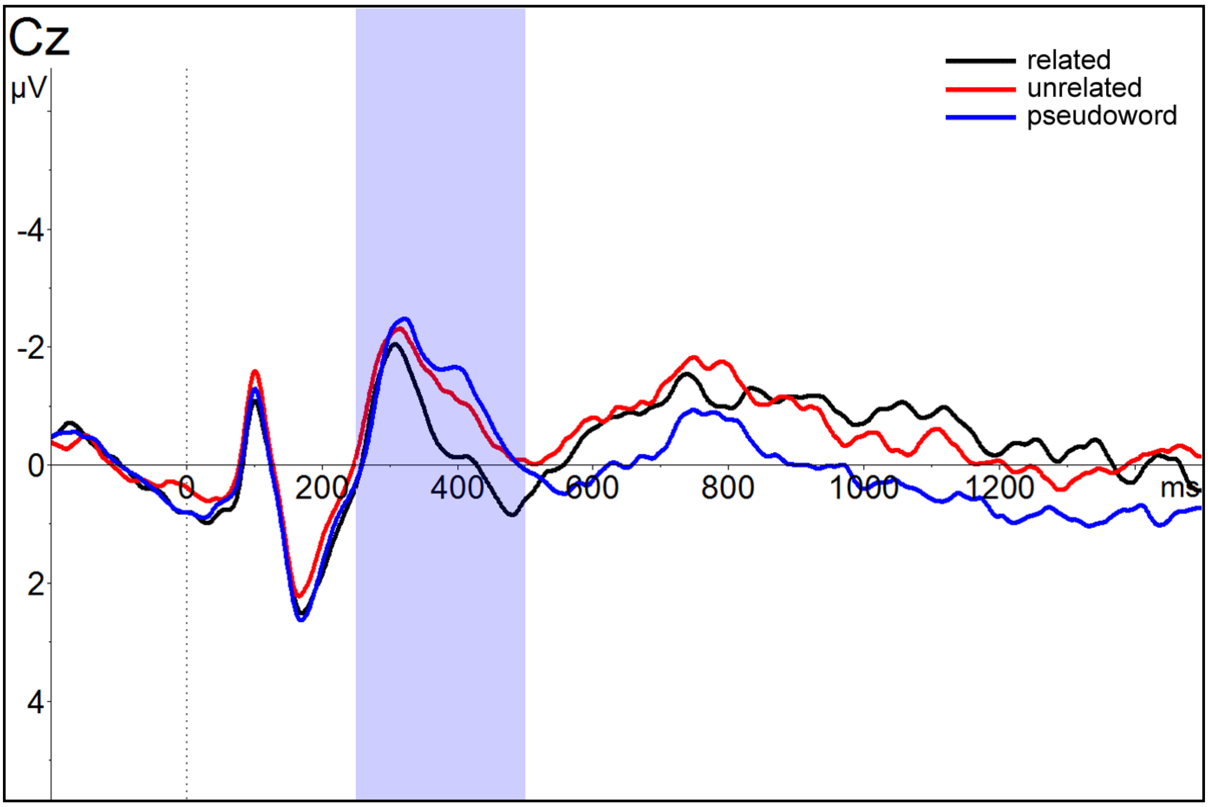
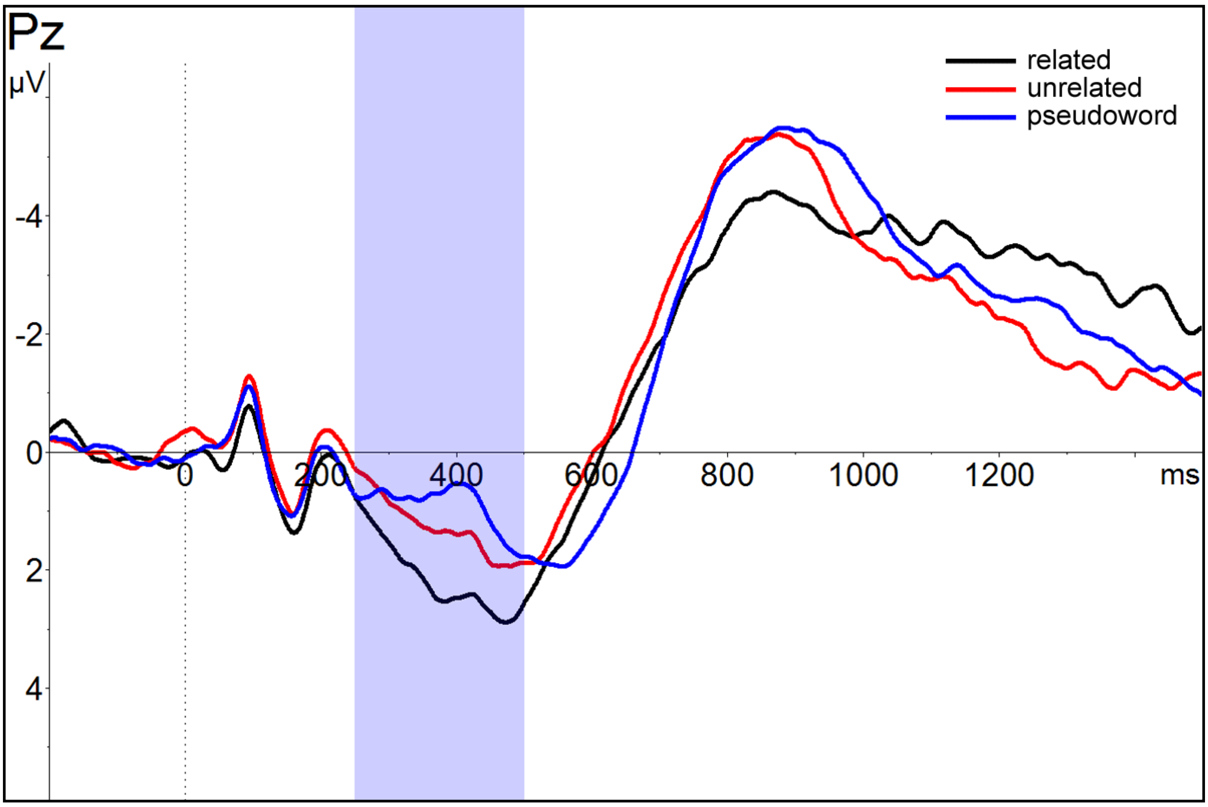


Figure 6.2.5. Condition 2: The ERP in the three conditions (i.e., related, unrelated, pseudoword) at the electrode Pz, Cz, Fz.

Condition 3: Prime C1$_{full}$ - C2$_{full}$ (Phrase) and Target at the onset of prime

Figure 6.2.6 presents the ERP data. A main effect of Semantic association was found [*F* (2, 38) = 12.96, MSE = 0.943, $p < 0.001$, $\eta 2_p = 0.41$]. Post hoc analyses revealed a larger N400 for unrelated targets than for related targets (-0.67 ± 0.12 µV, $p < 0.001$; N400 effect). No significant interaction Semantic × Electrode was found [*F* (4, 76) = 1.39, MSE = 2.489, $p >$ 0.05].


Cz
µV
-4
-2
0
2
4
0
200
400
600
800
1000
1200
ms
related
unrelated
pseudoword


Pz
µV
-4
-2
0
2
4
0
200
400
600
800
1000
1200
ms
related
unrelated
pseudoword

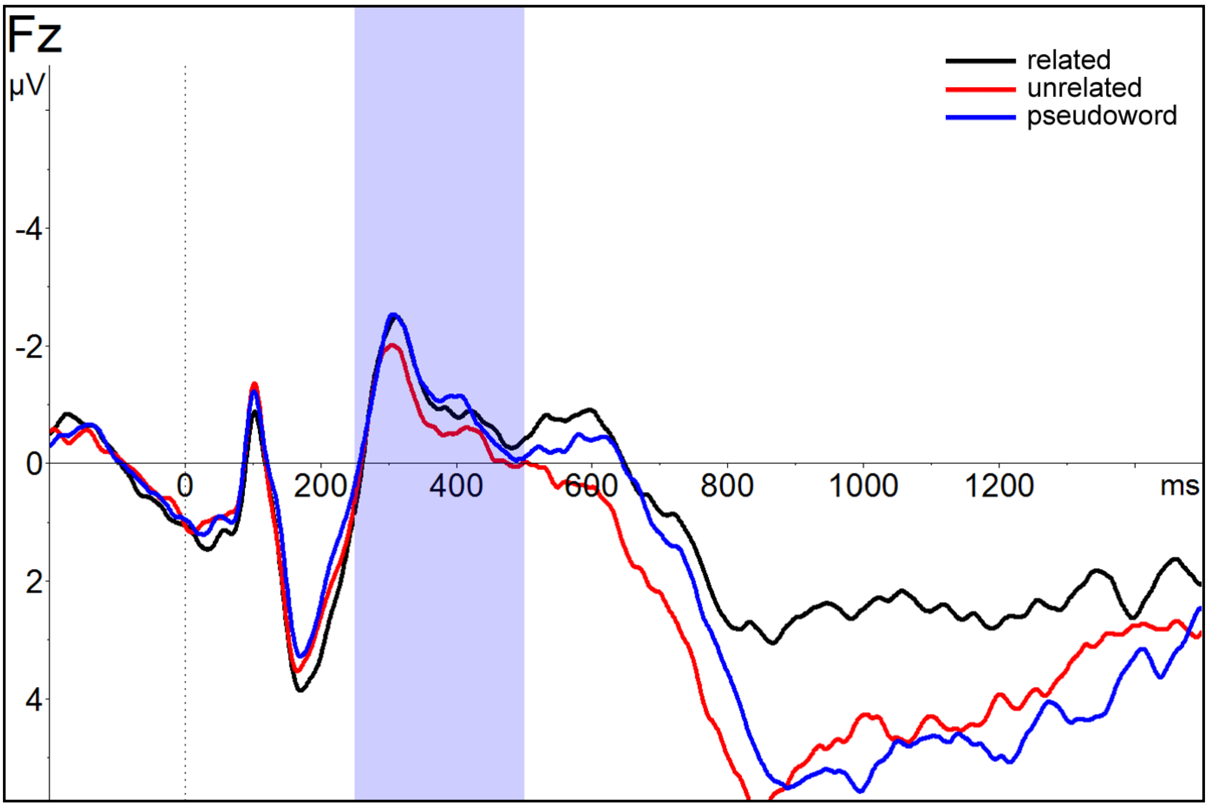


Figure 6.2.6. Condition 3: The ERP in the three conditions (i.e., related, unrelated, pseudoword) at the electrode Cz, Pz, Fz.

Condition 4: Prime $C1_{full}$ - $C2_{full}$ (Phrase) and Target at the onset of C2

Figure 6.2.7 displays the ERP data for condition 4. The main effect of Semantic association was not significant [F (2, 38) = 1.92, MSE = 1.047, $p > 0.05$]. In contrast, a significant Semantic association × Electrode interaction was observed [F (4, 76) = 6.35, MSE = 1.180, $p < 0.01$, $\eta 2_p = 0.25$]. Post hoc analyses revealed a larger N400 response for the unrelated targets than for the related ones (-0.27 ± 0.10 µV, $p < 0.05$; N400 effect), as well as a larger N400 for the pseudowords than for the related targets (-0.47 ± 0.17 µV, $p < 0.05$) at the Cz.

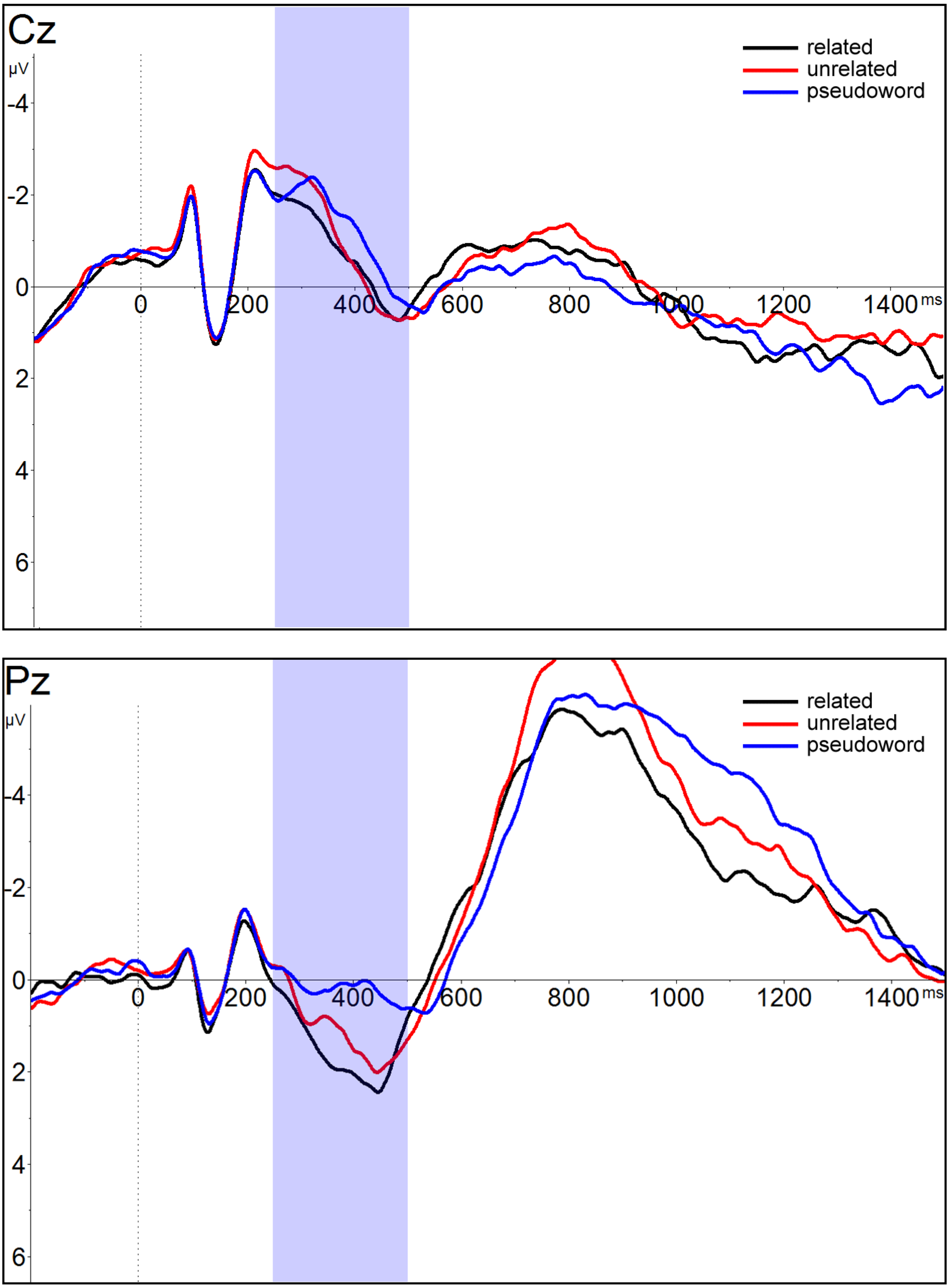


Figure 6.2.7. The ERP in the three conditions (i.e., related, unrelated, pseudoword) at the electrode Cz for Condition 4: Prime $C1_{full}$ - $C2_{full}$ & Target at the middle of prime.

6.2.3. Discussion

The present experiment investigated the influence of stress information in Chinese compound words recognition, more particularly to which extends the stress information interacts with morpholexical processing. Twenty minimal pairs of strong-weak (with a neutral tone on the final syllable) vs. strong-strong stress pattern were used. We recorded the ERP using a cross-modal auditory-visual associative priming paradigm with a lexical decision task to investigate how stress information of the last constituent modulates semantic priming effect on the first constituent whose semantic transparency depends on the prosodic characteristics of the last constituent.

As expected, behavioral data showed shorter response times for related target word as compared to unrelated ones for both types of prime (i.e., full-full tone phrase and full-neutral tone compound words) when the target was presented at the onset of the C1. However, results revealed a reverse priming effect for full-neutral tone compound words when the target was presented at the onset of the second constituent, i.e. C2 (middle of the prime). The lengthening of the RT in the related condition in comparison with the unrelated one could be view as a negative priming effect. Negative priming effect is usually assumed to result from active suppression (inhibition) of irrelevant information (Neill, 1977; Tipper & Cranston, 1985). This pattern of results suggested that the stress information of the last syllable of the prime may modulated the semantic propriety of the first morpheme of the prime. We propose that as soon as the stress information has been processed, the semantic representation of the first morpheme of the prime has to be deactivated (became semantically opaque), resulting in more difficulty in the lexical access of a semantically related target word (longer response times).

ERP data revealed the N400 related to associative priming effect. When the target word was presented at the onset of C1, consistent N400 effects were found for strong-weak compound and for strong-strong phrase (see Figures 6.2.4 and 6.2.6). The related targets evoked reduced N400 amplitudes than the unrelated ones (i.e. N400 associative priming effect). In the context of lexical access, the N400 amplitude is assumed to reflect the degree of difficulty in accessing the lexical-semantic representation, this access being facilitated for the related targets compared to the unrelated ones (i.e. facilitation of the access to the semantic representation of the targets)[39]. When the target was presented at the middle of the prime, i.e. at the onset of the C2, where stress information became available during the processing of the target, in the full-full tone phrase prime condition, the N400 effect was preserved (see Figure 6.2.7), suggesting that the semantic representation of the first morpheme of the prime remained activated when the second morpheme of the prime was processed. However, the N400 effect was not found in full-neutral tone compound prime condition suggesting that the semantic representation of the first constituent was not accessed. These findings suggested that the stress information in the last syllable of the ambiguous sequence served as an *access code* for selecting the appropriate processing routes to allow a successful access to the compound word. When the second morpheme (i.e., carrying the stress information) was processed, a decomposition route would be turn off, a direct route was reinforced, consequently the first constituent was not accessed thus it would not prime the semantically related target.

The Prosody-Assisted-Processing (PAP) model for compound words processing assumes that the length of the first lexical morphemes of compound words is a determining factor for setting off/triggering a decompositional route. At the onset of the speech input, a direct route is activated by default. Simultaneously, a prosodic analysis of the left-most morpheme is automatically carried out. The output of the prosodic analysis is crucial with respect to the activation of the decompositional route. If this output informs the processor that the left-most morpheme belongs to a compound word (i.e. short syllable), then the processing system activates a decompositional route. On the contrary, if this prosodic analysis indicates a monomorphemic word (i.e. long syllable), the processing system continues to match the acoustic-phonetic input with the word detectors at the word level, until the appropriate detector reaches the threshold for best match. In this case, the processing system has set a monomorphemic route configuration. Because of the sequential involvement of the two routes, the PAP model is considered as a cascading parallel model.

Regarding to our results in Mandarin Chinese, we observed a picture showing that at

[39] The pseudoword showed the largest N400 amplitude which reflects largest difficulty as lexical access never succeed.

the onset of the speech input, a decompositional route is activated by default. Constituent morphemes are lexically processed. Simultaneously, a prosodic analysis of the stress pattern is automatically carried out. When the prosodic analysis reaches to the last syllable of the input, the stress pattern could then be recognized. The output of the prosodic analysis is crucial with respect to the activation of the direct route. If this output informs the processor that the last morpheme carried a neutral tone, then the processing system activates a direct route and deactivates the decompositional route. On the contrary, if this prosodic analysis does not detect a neutral tone, the processing system continues the constituent morpheme processing in addition to the direct route. Our findings extend thus the PAP model by using compound words and phrase, more importantly by adding a right-to-left retroactive processing of prosodic information in Mandarin Chinese.

We assume that the right-to-left retroactive processing ensures a successful lexical access to the semantically opaque compound words in the minimal pairs. However, we cannot exclude that some acoustic information already carried in the first constituent may have informed the processing system that the sequence under processing is a compound word as our acoustic measures showed that the duration of the first constituent is shorter when it is followed by a neutral tone than by a full tone. Based on the present findings, we propose that neutral tone at the final syllable, probably combined with duration of the initial syllable, could be two *reliable* prosodic cues for assisting the analysis of the lexical morphology in Chinese. We concluded that, as already reported in other languages with various grammatical sequences, prosody might also have a function of prediction of ongoing grammatical information in Mandarin Chinese ambiguous minimal pairs.

6.2.4. Conclusion

The present experiment investigated to what extent the stress pattern information interacts with the morpholexical processing, such as semantic access of morphemes in Chinese compound word. Results showed semantic priming effects indexed by the N400 for strong-weak compound and for strong-strong phrase when the second syllable that carried the stress pattern information was not processed. Interestingly, when the second syllable was processed (i.e., stress pattern information was analyzed), there was no N400 effect suggesting that semantic representation of the first constituent was not accessed. In contrast, the N400 effect was preserved for strong-strong phrase. These findings suggested that the stress information in the last syllable of the ambiguous sequence served as an access code for the choice of the processing routes to allow a successful access to the compound word. Our findings extend the PAP model by using compound words and phrase, more importantly by adding a right-to-left retroactive processing of prosodic information in Mandarin Chinese.

6.3. Experiment 3: Acoustic analysis on VN versus/VO minimal pairs

Shen, W., Vaissière, J., & Isel, F. (2013). Acoustic Correlates of Contrastive Stress in Compound Words versus Verbal Phrase in Mandarin Chinese. *Computational linguistics and Chinese language processing, 18(3)*, 45-58.

As discussed in section 2.3.1 in Chapter 2, Chinese Verb-Noun form may be considered as a Verb-Object phrase or as a compound of different types (i.e. attributive compound or subordinate compound) and is an especially interesting case of apparent indeterminacy between morphology and syntax. The dual status of Verb-Noun form results in many ambiguous compounds/phrases minimal pairs. There are neither morphological nor semantic or syntactic cues for making a distinction between them. Nevertheless, prosodic information of the ambiguous compounds/phrases minimal pairs may be involved in their distinction. According to the Compound Stress Rule and the Nuclear-Stress-Rule (Chomsky & Halle, 1968), stress placement is different between compounds and the corresponding phrases, with left stress for compound and right phrasal stress for a phrase. Furthermore, each content word of a phrase keeps its own word stress. Taking the compound word *blackbird* as an example, when compound word is uttered in isolation: *black* carries word stress and phrasal stress while bird carries no stress, thus the stress pattern is 2:0, *black* carries the main stress (degree 2); when phrase is uttered in isolation, *black* receives the word stress and *bird* receive the word stress and phrasal stress, the stress pattern is 1:2, then *bird* carries the main stress (degree 2).

It has been proposed that the distinction of different degree of prominence between the syllables in compounds and phrases is true for Mandarin Chinese, and it permits one to distinguish between compounds and phrases (Duanmu, 2000). Duanmu (1990, 2000) proposed a general rule for assigning Mandarin Chinese compound and phrasal stress, the Non-head Stress Rule. This rule achieves similar effects as the Compound Stress Rule and the Nuclear Stress Rule combined and assumed that in the syntactic structure [X XP] (or [XP X]), where XP is the non-head and X the head, the non-head XP should be stressed. According to Non-head Stress rule, the non-head should have stress in the VN/VO minimal pairs, and because the non-head is on the left in compounds and on the right in most phrases, stress is assigned to the left in compounds and assigned to the right in phrases. An example in Duanmu (2000):

(1) 炒饭 *CHAOfan* [V+N]$_N$ 'fry rice' = 'fried rice'

(2) 炒饭 *chaoFAN* V+O 'fry rice' = 'fry the rice'

炒饭 *CHAOfan* in (1) is a noun compound (a single word) in which the verb modifies the noun head. The non-head *chao* holds the lexical compound and the phrasal stresses. With respect to the verb-object structure 炒饭 *chaoFAN* in (2), the phrasal stress is placed on the syntactic non-head *fan*.

However, there is little acoustic evidence for testing Duanmu's assumption on the distinction of stress pattern in Mandarin Chinese VN/VO minimal pairs. We therefore conducted a production study to investigate the acoustic correlates of stress between VN/VO compounds and phrases in Mandarin Chinese. Our hypotheses were that 1) VN modifier-head compound and VO phrases differ phonetically with left stress in VN modifier-head compounds and right stress in VO phrases and that 2) a different prosodic pattern is reflected in acoustic features in F0, duration and intensity.

6.3.1. Method

6.3.1.1. Materials

One hundred thirty-five minimal pairs (see Appendix C) presenting a morpholexical ambiguity (i.e. VN modifier-head compound vs. VO phrases) were selected from the Contemporary Chinese Dictionary 5th edition (Lu & Ding, 2008). Each pair had the same segmental characteristics and was assumed to differ from each other only in the stress pattern. The target words were not recorded in embedded utterance fragment:

1) 我说的不是偏正名词‘*编号*’而是动宾短语‘*编号*’.

[I did not say attributive noun compound ‘*bian-hao’* but said verb-object phrase ‘*bian-hao*’.]

The critical words in each pair change their position in the utterance fragment, giving

2) 我说的不是动宾短语‘*编号*’而是偏正名词‘*编号*’.

[I did not say verb-object phrase ‘*bian-hao’* but said attributive noun compound ‘*bian-hao’*.]

6.3.1.2. Recording Procedure

The recording was carried out in the laboratory of Phonetics and Phonology of University Sorbonne Nouvelle Paris 3. Speakers were recorded individually in an acoustic chamber, using an attached microphone, placed at a distance of about five centimeters from the speaker’s mouth. Speech samples were recorded digitally at 44100 Hz, 16-bit mono.

6.3.1.3. Participants

Six native Mandarin Chinese speakers (four females) participated in the experiment. All speakers had Mandarin both as mother tongue and as language of schooling.

6.3.1.4. Acoustic Measurements

The first syllable and the second syllable for each critical word were manually marked in Praat. A Praat script extracted the mean duration and the mean intensity value of each segment. The F0 was measured on the vowel of syllables. The average F0 was calculated, moreover, the F0 at onset/offset of the vowel, the $F0_{min}$ and the $F0_{max}$ were measured. In order to reconstruct the contour of the F0, we divided the vowel into ten segments normalized in time, we used then these ten F0 for representing the F0 contour.

6.3.2. Results

Two-way repeated measures ANOVA were performed for each acoustic feature (i.e., duration, F0, and intensity) with factors of Type of word (VN, VO) and Position of syllable (left syllable, right syllable).

6.3.2.1. Duration

The two-way ANOVA showed a main effect of Type [F (1, 269) = 5.86, MSE = 6091.635, $p < 0.05$, $\eta 2_p = 0.2$] indicating a longer duration for VO (593.90 ± 123.99 ms) than for VN (587.29 ± 123.98 ms). Moreover, a significant Type × Position interaction was observed [F (1, 269) = 418.61, MSE = 2021.034, $p < 0.001$, $\eta 2_p = 0.61$]. Post-hoc analyses showed that the duration of left syllable of VN (315.88 ± 63.69 ms) was longer than the one of VO (263.21 ± 47.50 ms) [F (1, 269) = 310.018, MSE = 1208.125, $p < 0.001$, $\eta 2_p = 0.54$] and the duration of right syllable of VO (330.69 ± 76.49 ms) was longer than the one of VN (271.41 ± 60.29 ms) [F (1, 269) = 198.95, MSE = 2384.585, $p < 0.001$, $\eta 2_p = 0.43$] (see Figure 6.3.1).

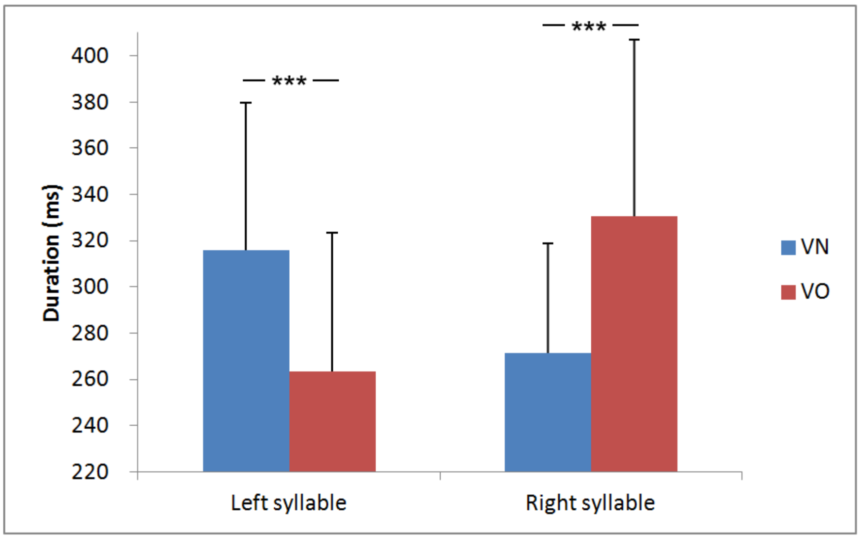


Figure 6.3.1. Mean syllable durations in ms for the left and the right syllable in VN and VO. *** $p < 0.001$

6.3.2.2. F0

The F0 was analyzed separately for each four tones. With regard to different contour characteristics of the four tones (i.e., Tone 1: high flat, Tone 2: rising, Tone 3: falling-rising, Tone: falling), analyses were conducted on the average F0 for Tone 1, on the onset/offset of F0 for Tone 2 and Tone 4 or on the variation of F0 for Tone 3.

Tone 1

A two-way ANOVA with factors of Type and Position was conducted on the average of the F0 for Tone 1. Results showed a significant interaction between Type and Position [F (1, 45) = 48.96, MSE = 184.65, $p < 0.001$, $\eta 2_p = 0.52$]. Post-hoc analyses showed that the average F0 of left syllable of VN (219.45 ±16.52 Hz) was a higher than the one of VO (205.41 ±13.53 Hz) [F (1, 61) = 38.22, MSE = 159.853, $p < 0.001$, $\eta 2_p = 0.39$] and the average F0 of right syllable of VO (215.15 ±18.92 Hz) was a higher than the one of VN (199.91 ±22.67 Hz) [F (1, 45) = 15.14, MSE = 352.423, $p < 0.001$, $\eta 2_p = 0.25$]. The contour of F0 for Tone 1 is shown in Figure 6.3.2.

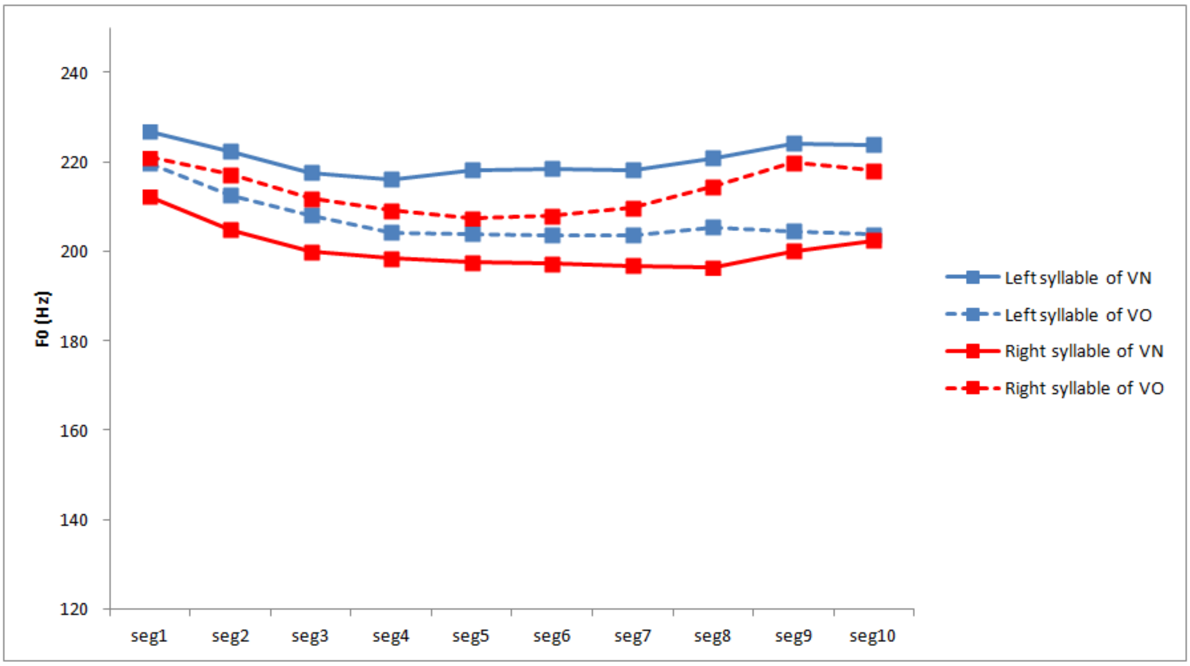


Figure 6.3.2. F0 contours of Tone 1 for the left and the right syllable in VN and VO.

Tone 2

As the contour of the F0 of Tone 2 is rising, we conducted analyses on the onset and on the offset of the F0. For onset F0, the ANOVA with factors of Type and Position failed to reveal either a significant main effect of Type or a significant interaction of Type and Position [$Fs < 1$]. For offset F0, the ANOVA showed a main effect of Type [F (1, 51) = 6.37, MSE = 321.576, $p < 0.05$, $\eta 2_p = 0.11$], indicating that..Moreover, a significant Type by Position interaction was found [F (1, 51) = 72.02, MSE = 317.616, $p < 0.001$, $\eta 2_p = 0.59$]. Post-hoc analyses revealed that the offset F0 of left syllable of VN (207.92 ± 15.96 Hz) was higher than the one of VO (193.07 ± 22.16 Hz) [F (1, 57) = 39.75, MSE = 160.939, $p < 0.001$, $\eta 2_p = 0.41$] and that the offset F0 of right syllable of VO (209.82 ± 10.60 Hz) was higher than the one of VN (182.57 ± 32.55 Hz) [F (1, 51) = 41.49, MSE = 465.287, $p < 0.001$, $\eta 2_p = 0.45$]. The contour of F0 for Tone 2 is displayed in Figure 6.3.3.

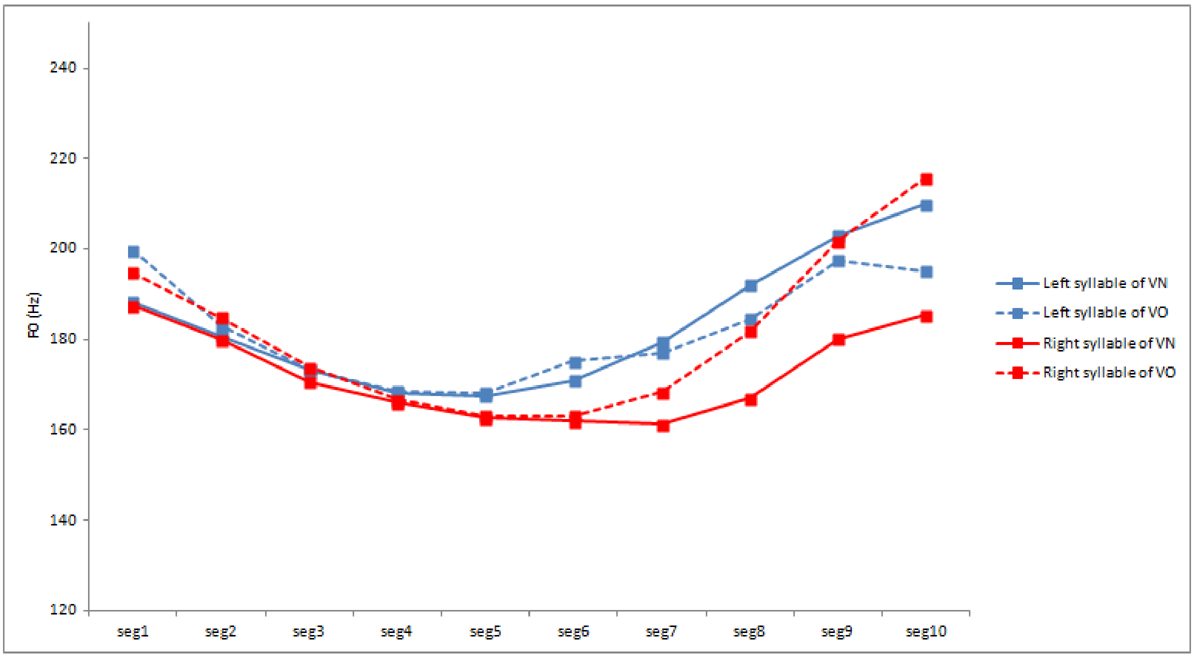


Figure 6.3.3. F0 contours of Tone 2 for the left and the right syllable in VN and VO.

Tone 3

As the contour of the F0 of Tone 3 is falling-rising, we conducted analyses on the variation of the F0 (i.e., the difference between $F0_{max}$ and $F0_{min}$). A two-way ANOVA with factors of Type and Position revealed a significant interaction between Type and Position [F (1, 53) =

23.95, MSE = 3858.002, $p < 0.001$, $\eta 2_p = 0.31$]. Post-hoc analyses showed that the variation of F0 of left syllable of VN (73.07 ±78.83 Hz) was higher than the one of VO (47.16 ±29.59 Hz) [F (1, 57) = 4.39, MSE = 4432.861, $p < 0.05$, $\eta 2_p = 0.07$], and conversely, the variation of F0 of the right syllable of VO (101.77 ±60.51 Hz) was higher than the one of VN (44.80 ±66.49 Hz) [F (1, 53) = 21.01, MSE = 4170.340, $p < 0.001$, $\eta 2_p = 0.28$]. In order to not confound tone sandhi influence for these analyses, two items with a Tone 3 - Tone 3 combination were excluded from the analyses. The contour of F0 for Tone 3 is presented in Figure 6.3.4.

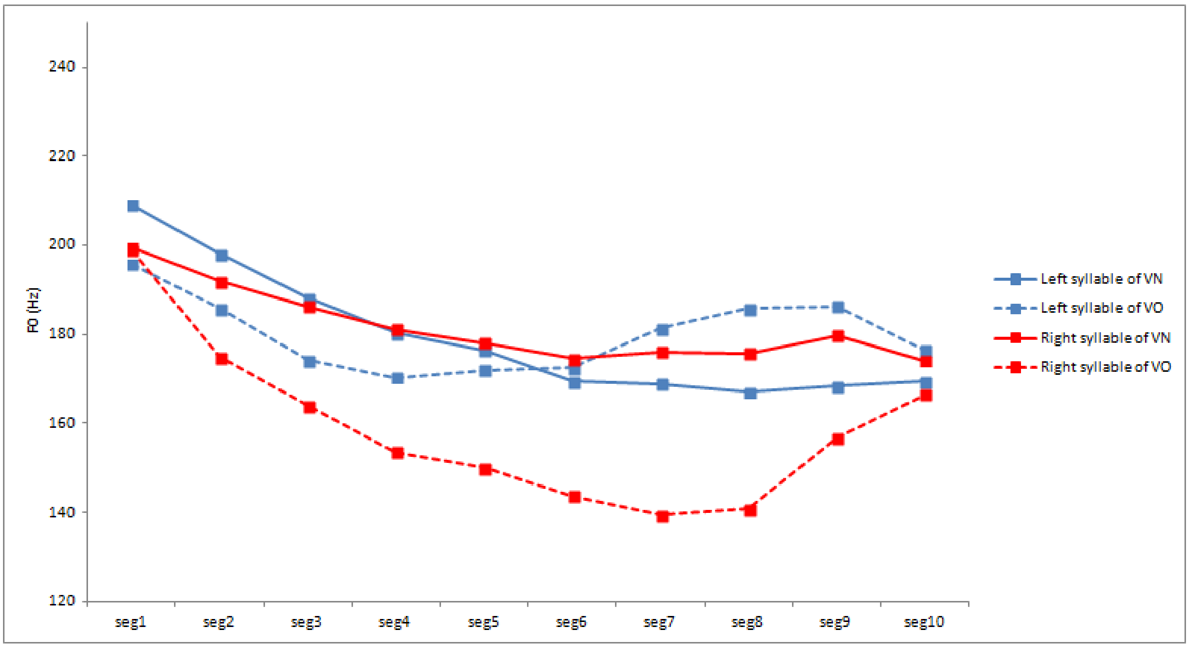


Figure 6.3.4. F0 contours of Tone 3 for the left and the right syllable in VN and VO.

<u>Tone 4</u>

As the contour of the F0 of Tone 4 is falling, we conducted analyses on the onset and on the offset of the F0. For onset F0, the ANOVA with factors of Type and Position showed a significant interaction between Type and Position [F (1, 91) = 17.62, MSE = 2142.689, $p < 0.001$, $\eta 2_p = 0.16$]. Post-hoc analyses revealed that the onset F0 of left syllable of VN (257.09 ± 48.38 Hz) was higher than the one of VO (233.29 ± 25.91 Hz) [F (1, 91) = 20.93, MSE = 1244.112, $p < 0.001$, $\eta 2_p = 0.19$] whereas the offset F0 of the right syllable of VO (227.01 ± 58.82 Hz) was higher than the one of VN (207.18 ± 51.20 Hz) [F (1, 116) = 8.71, MSE = 2640.542, $p < 0.01$, $\eta 2_p = 0.07$]. The contour of F0 for Tone 4 is shown in Figure 6.3.5.

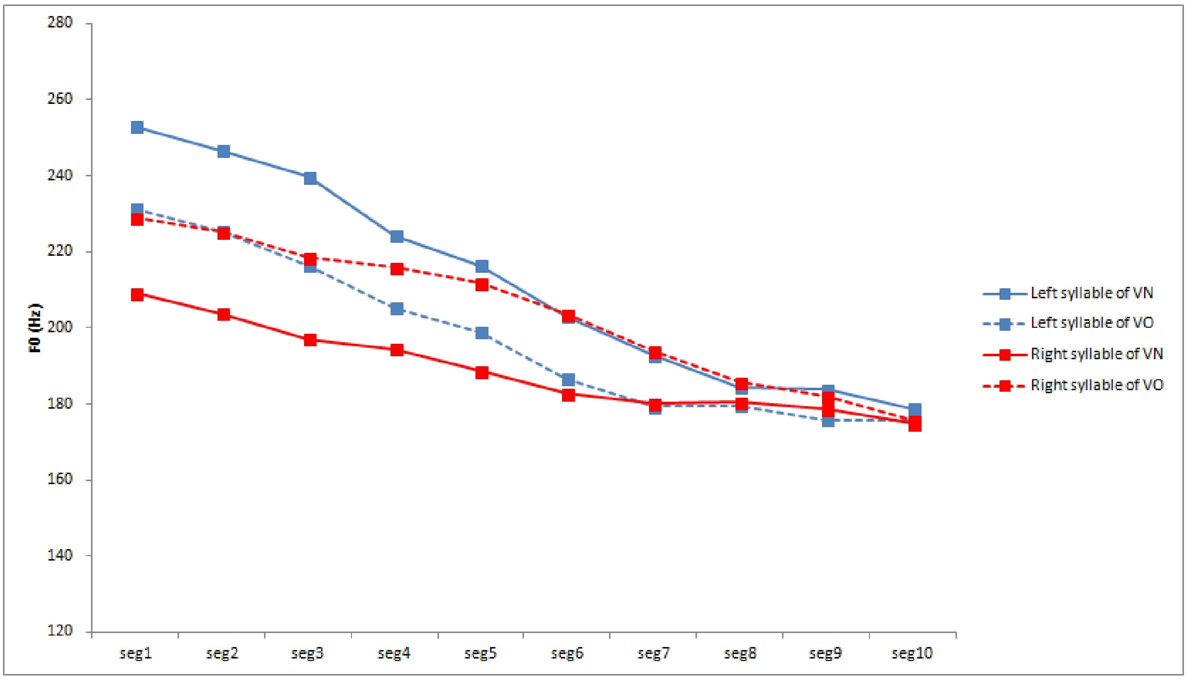


Figure 6.3.5. F0 contours of Tone 4 for the left and the right syllable in VN and VO.

6.3.2.3. Intensity

The two-way ANOVA showed a main effect of Type [F (1, 269) = 184.64, MSE = 3.674, p < 0.001, $\eta 2_p$ = 0.41] indicating a larger intensity for VO (60.13 ±3.2 dB) than for VN (58.54 ±3.58 dB). Moreover, a significant interaction between Type and Position was revealed [F (1, 269) = 396.67, MSE = 2.598, p < 0.001, $\eta 2_p$ = 0.60]. Post-hoc analyses showed that the intensity of left syllable of VN (60.65 ±3.70 dB) was larger than the one of VO (60.29 ±3.70 dB) [F (1, 269) = 7.26, MSE = 2.526, p < 0.01, $\eta 2_p$ = 0.26] and that the intensity of right syllable of VO (59.96 ±3.47 ms) was a larger than the one of VN (56.43 ±3.47 dB) [F (1, 269) = 451.26, MSE = 3.746, p < 0.001, $\eta 2_p$ = 0.63] (see Figure 6.3.6).

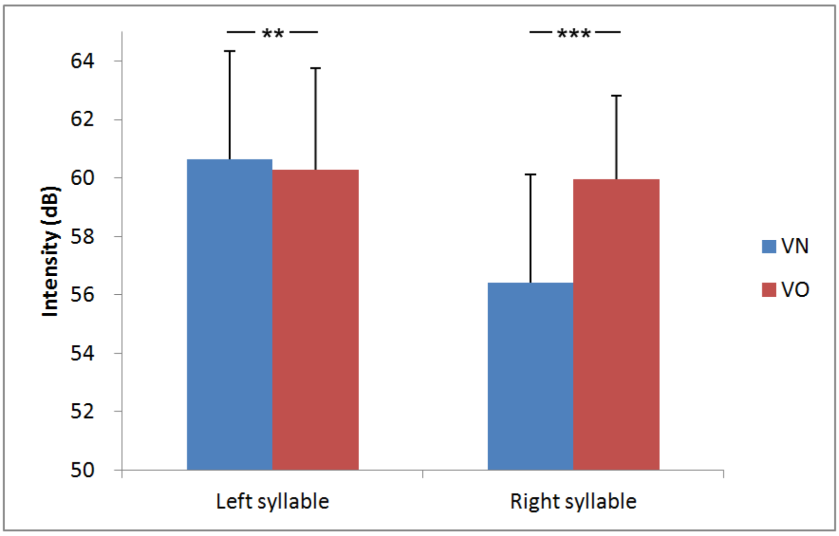


Figure 6.3.6. Mean intensity (dB) for the left and the right syllable in [VN] and [VO]. ** $p <$ 0.01, *** $p < 0.001$

6.3.3. Discussion

This production study investigated the acoustic correlates of stress pattern on the ambiguous structure Verb-Noun (i.e. VN vs. VO) in Mandarin Chinese. Results showed the implication of duration, F0, and intensity in the production of compound and phrasal stress in Mandarin. In particular, the duration was longer for the left syllable in VN than the one in VO, and the duration was longer for the right syllable in VO than the one in VN. Despite the fact that, in tone languages, F0 information should be attributed to its lexical usage, our results showed that F0 would be a reliable cue for the stress pattern in VN and VO. The average F0 was higher for the stressed left syllable in VN than the unstressed left one in VO. The average F0 was also higher the stressed right syllable in VO than the unstressed right one in VN. The variation of F0 was shown to link to the stress production for Tone 2, Tone 3 and Tone 4, which was in line with the predictions (Chao, 1968) that pitch variation is wider for stressed syllables, specifically, when a Tone 2 is stressed, it ends higher, when a Tone 3 is stressed, it dips lower, and, when a Tone 4 is stressed, it starts higher. The intensity was larger for the stressed left syllable in VN than for the unstressed left syllable in VO. Moreover, the amplitude of the stressed right syllable in VO was larger than the unstressed right syllable in VN. Our results showed that Mandarin Chinese follows the Compound Stress Rule and the Nuclear-Stress-Rule (Chomsky and Halle, 1968), stress placement is different between compounds and the corresponding phrases, with left stress for compound and right phrasal stress for a phrase. These findings confirmed a stress pattern strong-weak for VN and a stress pattern weak-strong for VO.

6.3.4. Conclusion

Our results demonstrated that (1) the left syllable in VN was longer than the one in VO, and the right syllable was more lengthened in VO than in VN; (2) the F0 range was larger in the right stressed syllable in VO than the one in VN; The intensity was larger in the left syllable in VN than the one in VO and larger in the right syllable in VO than the one in VN. Our results confirmed the left stressed pattern in VN and the right stressed pattern in VO. Taken together, the present acoustic study lends support to the hypothesis that principles of stress upward of word level are universal through different languages.

We have shown that VN compound words and VO phrases present different acoustic patterns with respect to the position of stress. The next step would be to verify whether this stress pattern is used by the listeners to differentiate the two forms in cases of segmental ambiguities. For this purpose, we conducted an ERP experiment to investigate the processing of stress pattern in the VN/VO pairs.

6.4. Experiment 4: ERPs correlate of stress pattern of VN/VO minimal pairs

The same paradigm as in Experiment 1 was applied to VN/VO pairs. Participants performed an explicit task: stress discrimination, and an implicit task: passive listening. Only non-delexicalized stimuli were used. We expected to replicate the results of Experiment 1. Two early exogenous components i.e., the N200 and P200 were expected to occur to infrequent stress pattern (i.e., strong-weak). In addition, the N325 component reflecting the perception of stress pattern was expected. Moreover, due to *oddball*-like design, we expected the mismatch negativity (MMN) for targets in incongruent trials regardless of stress pattern.

6.4.1. Method

6.4.1.1. Participants

Eighteen native Mandarin Chinese participants (9 females, 25.89 ± 3.29 years, ranging from 21 to 32 years) in Paris were recruited. None of them has lived in France more than five years (2.72 ± 1.23 years, ranging from 1 to 5 years) and all of them used Chinese frequently in daily life. All participants were right-handed according to the Edinburgh Handedness Inventory (Oldfield, 1971). They have normal or corrected to-normal vision and no auditory impairment. They have no history of current or past neurological or psychiatric illnesses. They were paid 10€ per hour for their time. All participants gave informed consent before participation and completed an information sheet.

6.4.1.2. Stimuli

A set of 20 Chinese disyllabic sequences (see Appendix D) was selected as auditory primes. Each sequence can be understood in two ways: 1) VN compound word; 2) VO phrase. An example of a disyllabic sequence *chao-fan* is presented in Table 6.4.1.

Table 6.4.1.

VN and VO of the disyllabic sequence chao-fan

Type	**Sequence**	**Stress pattern**	**Meaning**
VN	*chao3fan4*	strong-weak	fried rice
VO	*chao3fan4*	weak-strong	fry the rice

Note: The sequence *chao-fan* consists of two morphemes, the first morpheme = *chao* (verb, means 'fry') and the second morpheme = *fan* (noun, means 'rice').

VN carried strong-weak stress pattern and VO had weak-strong pattern. All sequences were recorded in an acoustic chamber by a native Mandarin Chinese speaker (female, 22 years old, from Xi'an). Phonetic analysis on these sequences were conducted to examine the contrastive stress pattern in VN and VO. Results indicated that 1) for duration, the ratio between first syllable and second syllable (i.e., S1 : S2) was larger [t (18) = 5.98, $p < 0.001$] for VN (1.36 ±0.56) than for VO (0.64 ±0.19); 2) for intensity, the ratio between first syllable and second syllable was larger [t (18) = 16.17, $p < 0.001$] for VN (1.32 ±0.07) than for VO (0.99 ± 0.07).

6.4.1.3. Design and Procedure

Experimental design was similar to the one in Experiment 1. Each trial started with a visual fixation at the center of a computer monitor for 1000 ms, this was followed by a sequence of four words presented auditory. The stimulus onset asynchrony for subsequent words was 1500 ms. The target word (the fourth one) was presented for 3000 ms, and then could be terminated by making a response. The intertrial interval was 1500 ms. Stimulus delivery and behavioral response recording were controlled by E-prime (Psychology Software Tools, Inc., Pittsburgh, PA; http://www.pstnet.com/eprime). A sequence consist four disyllabic items presented in succession auditory. The first three items all had the same stress pattern, that is, either strong-weak (i.e., VN) or weak-strong (i.e., VO). Prosodic congruency was manipulated by varying the stress pattern of the fourth item. In the prosodically congruent condition, the fourth item had the same stress pattern as the previous three items. In the prosodically incongruent condition, the fourth item had a different stress pattern. Examples of the four possible stimulus sequences are presented in Table 6.4.2.

Table 6.4.2.

Examples of the four experimental conditions

Condition	Item 1	Item 2	Item 3	Item 4	Congruency
1	VN *(SW)*	VN *(SW)*	VN *(SW)*	**VN** (*SW*)	Congruent
2	VO (*WS*)	VO (*WS*)	VO (*WS*)	**VO** (*WS*)	Congruent
3	VO (*WS*)	VO (*WS*)	VO (*WS*)	*VN* *(SW)*	Incongruent
4	*VN* *(SW)*	*VN* *(SW)*	*VN* *(SW)*	**VO** *(WS)*	Incongruent

Note: VN carry SW (strong-weak) pattern, VO carry WS (weak-strong) pattern.

Participants were asked to perform two tasks. The first task was a stress discrimination task in which the participants were instructed to discriminate between sequences of congruent and incongruent stress patterns by pressing the one button (*S* or *K* on keyboard) when the stress pattern of the final item was congruent with that of the first three and the other button when it was incongruent. The second task was a passive listening task where the participants were asked to listen carefully to the items presented to them. Trials of the two tasks were presented separately in two blocks. Block order was counterbalanced across the participants. Each block contained 120 trials (i.e., 30 trials per condition) presented in a pseudo-randomized order. A number of constraints were used for the pseudo-randomization. First, target item was not repeated immediately. Secondly, no more than three sequences of the same experimental condition were presented in succession. The pseudo-randomized order of trials was created using the program Conan (Nowagk, 1998).

6.4.1.4. ERP Recordings and Analysis

The EEG was recorded from 64-channel electrodes using Brain Vision Recorder (Brain Products GmbH, Gilching, Germany). Data was sampled at 1000 Hz with an online 0.1-100 Hz frequency bandpass filter. All electrode impedances were maintained below 5 kΩ. Further processing of ERPs was done off-line. First, EEG data were re-referenced to an average reference, and then were filtered with a 0.1-12 Hz bandpass filter. Next, the continuous EEG was segmented into epochs from 200 ms pre-stimulus until 1500 ms post-stimulus onset, with 200 ms of each epoch corresponding to a pre-stimulus baseline. Then, a semi-automatic artifact rejection procedure was applied. Artifacts were marked when 1) maximal amplitude was higher than 200 μV or minimal amplitude was lower than -200 μV; 2) absolute difference was more than 50 μV or less than 0.5 μV in an interval of 200 ms. Only trials with correct responses and with no artifacts were kept for further data analysis. In each experimental condition, the segmented ERP data was then averaged over trials and over participants (i.e., grand average). Finally, the ERP data were baseline-corrected to the mean amplitude of the 200 ms pre-stimulus interval.

Statistical analyses were conducted for three ERP signatures for which the time windows were selected based on previous studies and adapted by visual inspection of the grand averages: N200 & P200 (170-250 ms & 250-400 ms), and a late sustained negative-going potential (600-1000 ms). All analyses were quantified using a multivariate repeated measures approach and followed a hierarchical analysis schema. To permit examination of hemispheric differences, the data recorded at the lateral recording sites were treated separately from the data recorded at the midline electrode sites. All lateral electrodes were classified by two topographical variables: Region (anterior, posterior) and Hemisphere (left, right). Four regions of interest (ROIs) resulting from a complete crossing of the Region and Hemisphere variables were defined: left anterior (F7, F3, TT7, TC3), right anterior (F8, F4, TT8, TC4), left posterior (P3, P7, O1), and right posterior (P4, P8, O2). For each time window, a six-way repeated measures ANOVA with the factors of Stress pattern of the target (strong-weak, weak-strong), Congruency (congruent, incongruent), Task (stress discrimination, passive listening) and the topographical variables Hemisphere (left, right) and Region (anterior, posterior) was conducted. For the midline electrodes, a five-way repeated measures ANOVA including factors of Stress pattern of the target (strong-weak, weak-strong), Congruency (congruent, incongruent), Task (stress discrimination, passive listening) and Electrode (Fz, Cz, Pz) was run for each of the three time windows of interest. The Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied when evaluating effects with more than one degree of freedom in the numerator. Post-hoc pairwise comparisons at single electrode sites were made using a modified Bonferroni procedure (Keppel, 1991).

6.4.2. Results

6.4.2.1. Behavioral Results

Behavioral data[40], in the form of error rate and response times, are summarized in Table 6.4.3 for the four experimental conditions (congruent, incongruent, strong-weak, weak-strong). For error rate, a two-way repeated measures ANOVA with the factors Congruency (congruent, incongruent), and Stress pattern of the target (strong-weak, weak-strong) did not reveal a main effect of Congruency [$F < 1$] or a main effect of Stress [F (1, 16) = 3.85, MSE = 0.006, $p > 0.05$]. Moreover, there was no significant interaction between Congruency and Stress [$F < 1$]. For response times, a two-way repeated measures ANOVA did not show a main effect of Congruency [$F < 1$] or a main effect of Stress [F (1, 16) = 2.00, MSE = 11785.791, $p > 0.05$] or a significant interaction between Congruency and Stress [$F < 1$].

Table 6.4.3.

Behavioral data for four experimental conditions

	Error [%]		RT [ms]	
	Mean	SD	Mean	SD
Congruent	13.2	13.3	641.7	231.6
Incongruent	12.9	15.4	643.7	214.8
Strong-weak	11.1	11.5	624.1	215.8
Weak-strong	15.0	17.3	633.3	202.3

Note: Strong-weak is the stress pattern of compound target and weak-strong is the stress

[40] Data of one participant was not taken into the analyses due to damage in the data.

pattern of phrase target.

6.4.2.2. Electrophysiological Results

Statistical analysis of the EEG data was conduct in each two experimental block, i.e., 1) Block 1: stress discrimination task, 2) Block 2: passive listening task. Stress pattern effect was analyzed in time windows 170-250 ms and 250-400 ms and Congruency effect was analyzed in time windows 170-250 ms, 250-400 ms and 600-1000 ms. Stress pattern effect and Congruency effect were presented separately.

6.4.2.2.1. Stress pattern effect

Time window 170-250 ms (N200)

Block 1: stress discrimination task

The ERP data is shown in Figures 6.4.1 and 6.4.2. A three-way repeated measures ANOVA on lateral electrodes with the factors Stress pattern of the target, Hemisphere and Region showed a main effect of Stress [F (1, 16) = 7.52, MSE = 0.148, $p < 0.05$, $\eta 2_p = 0.32$] indicating a larger positivity for strong-weak targets (0.15 ± 1.51 μV) than for weak-strong targets (0.02 ± 1.53 μV) and a significant interaction of Stress × Region [F (1, 16) = 5.14, MSE = 2.477, $p < 0.05$, $\eta 2_p = 0.24$]. Post-hoc analyses revealed that the interaction was due to a larger positivity at posterior electrodes [F (1, 16) = 6.87, MSE = 1.557, $p < 0.05$, $\eta 2_p = 0.30$]. A two-way repeated measures ANOVA on central electrodes with the factors Stress pattern of the target and Electrode showed a main effect of Stress [F (2, 32) = 10.10, MSE = 0.421, $p < 0.01$, $\eta 2_p = 0.39$] indicating a larger negativity for weak-strong targets (-0.62 ± 1.21 μV) than for strong-weak targets (-0.33 ± 1.17 μV).

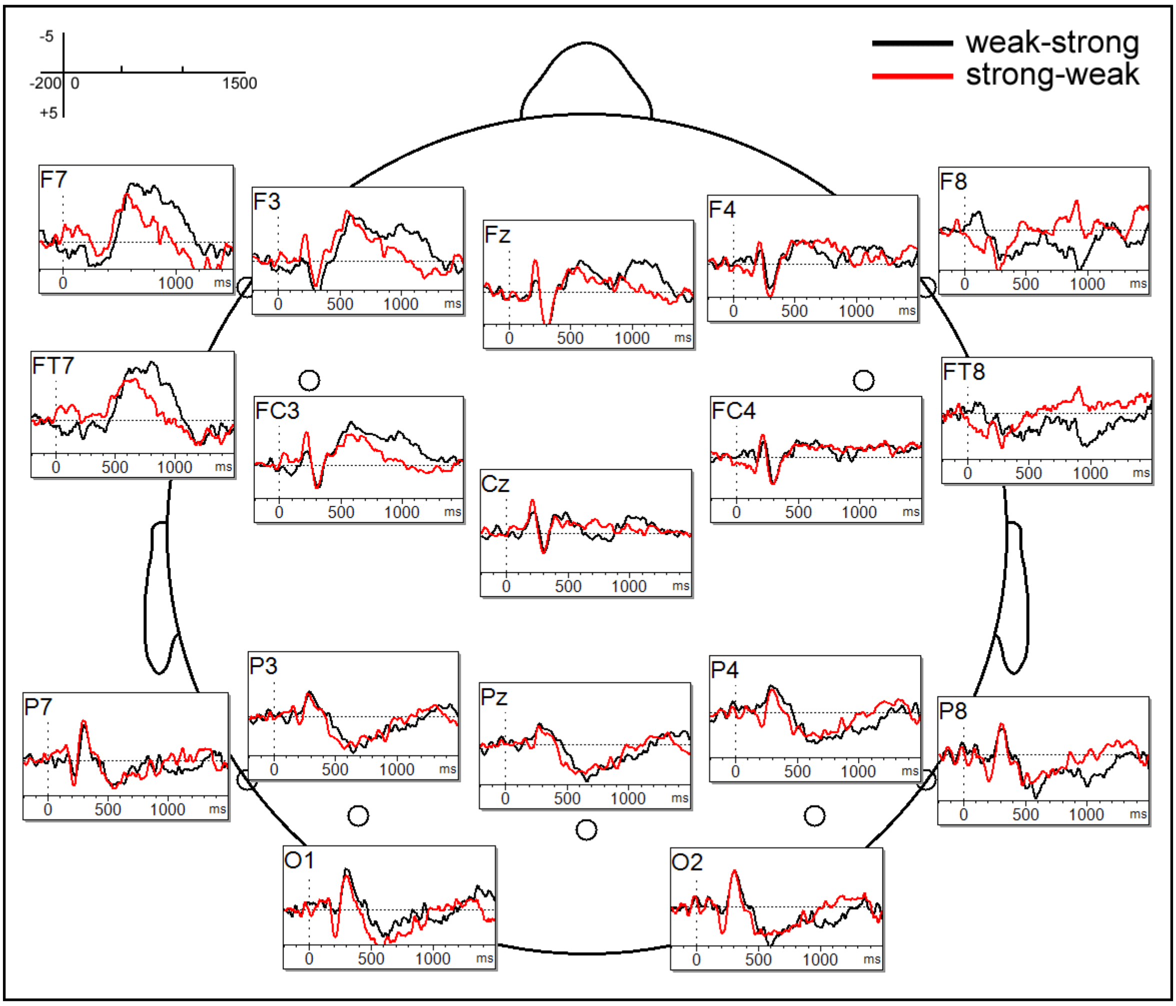


Figure 6.4.1. The ERP in the two conditions (i.e., strong-weak, weak-strong) at the 17 electrodes (F7, F3, FT7, FC3, F8, F4, FT8, FC4, P3, P7, O1, P4, P8, O2, Fz, Cz, Pz) for Block 1: stress discrimination task.

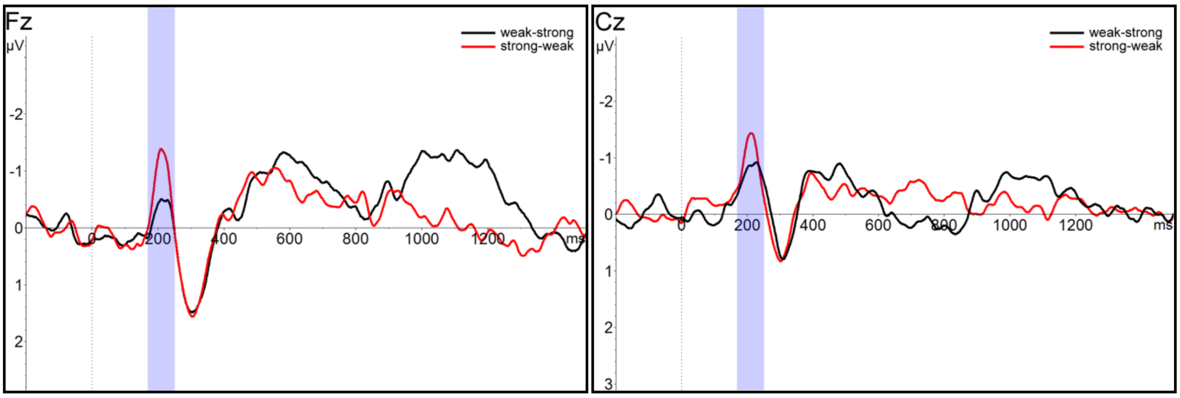


Figure 6.4.2. The ERP in the two conditions (i.e., strong-weak, weak-strong) at the electrode Fz and the Cz for Block 1: stress discrimination task.

Block 2: passive listening task

Figures 6.4.3. and 6.4.4. show the ERP data. For lateral electrodes, ANOVA results did not show any main effect or interaction (*Fs* < 1). For central electrodes, neither a significant main effect [F (1, 16) = 2.53, MSE = 1.035, $p > 0.05$] nor a significant Stress × Electrode interaction was found [F (2, 32) = 2.39, MSE = 2.302, $p > 0.05$].

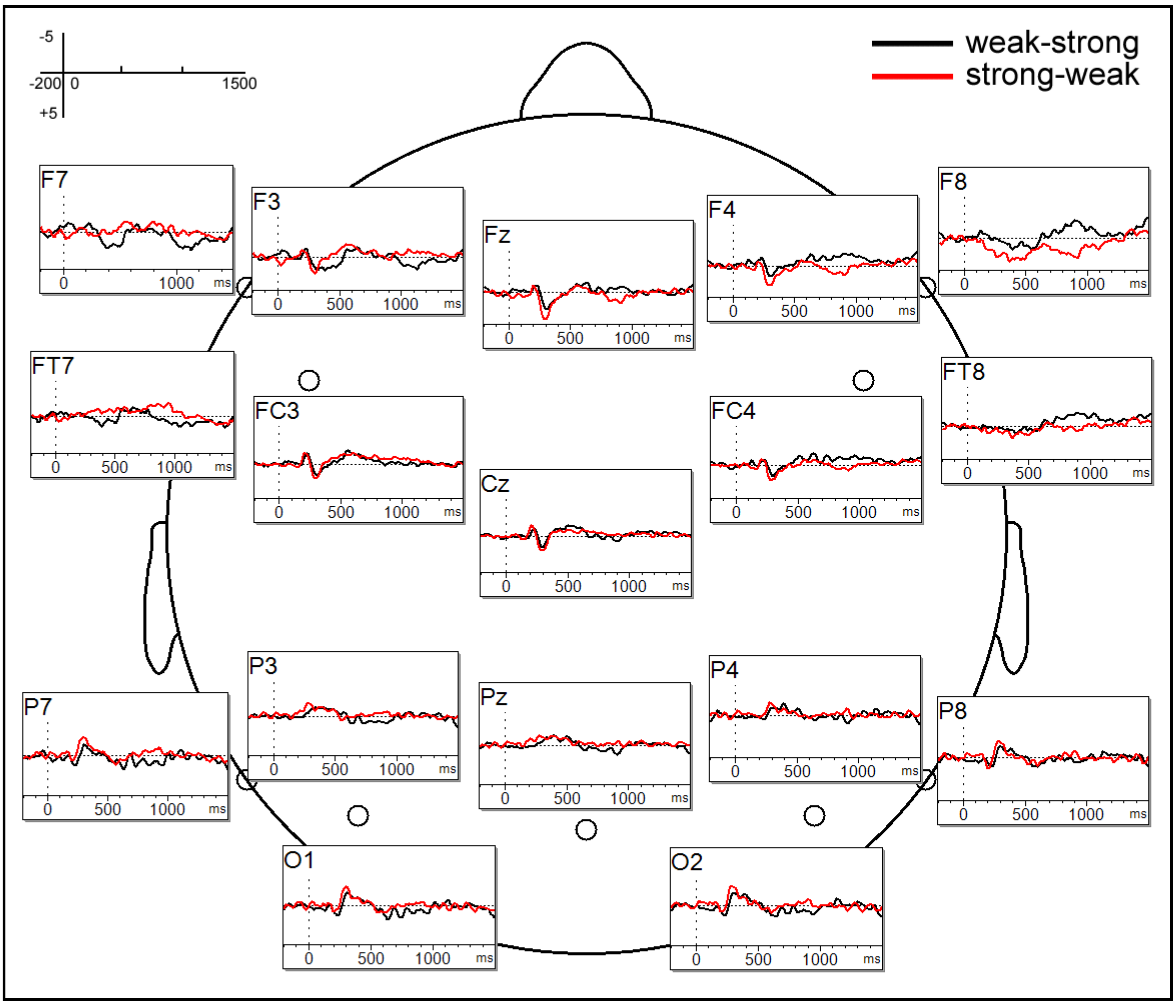


Figure 6.4.3. The ERP in the two conditions (i.e., strong-weak, weak-strong) at the 17 electrodes (F7, F3, FT7, FC3, F8, F4, FT8, FC4, P3, P7, O1, P4, P8, O2, Fz, Cz, Pz) for Block 2: passive listening task.

Time window 250-400 ms (P200)

Block 1: stress discrimination task

For lateral electrodes, there was neither main effect nor interaction (*Fs* < 1) which reached the significance level. For central electrodes, neither a significant main effect nor a significant Stress × Electrode interaction was found was found [*Fs* < 1].

Block 2: passive listening task

For lateral electrodes, ANOVA results showed neither significant main effect of Stress [F (1, 16) = 2.05, MSE = 0.219, $p > 0.05$] nor significant interactions. For the midline electrodes, a significant interaction between Stress and Electrode was found [F (2, 32) = 3.84, MSE = 0.743, $p < 0.05$, $\eta 2_p = 0.19$]. Post-hoc analyses revealed a larger positivity for the strong-weak pattern than for the weak-strong one, but only at the Fz [F (1, 16) = 5.33, MSE = 0.810, $p < 0.05$, $\eta 2_p = 0.25$] (see Figure 6.4.4).

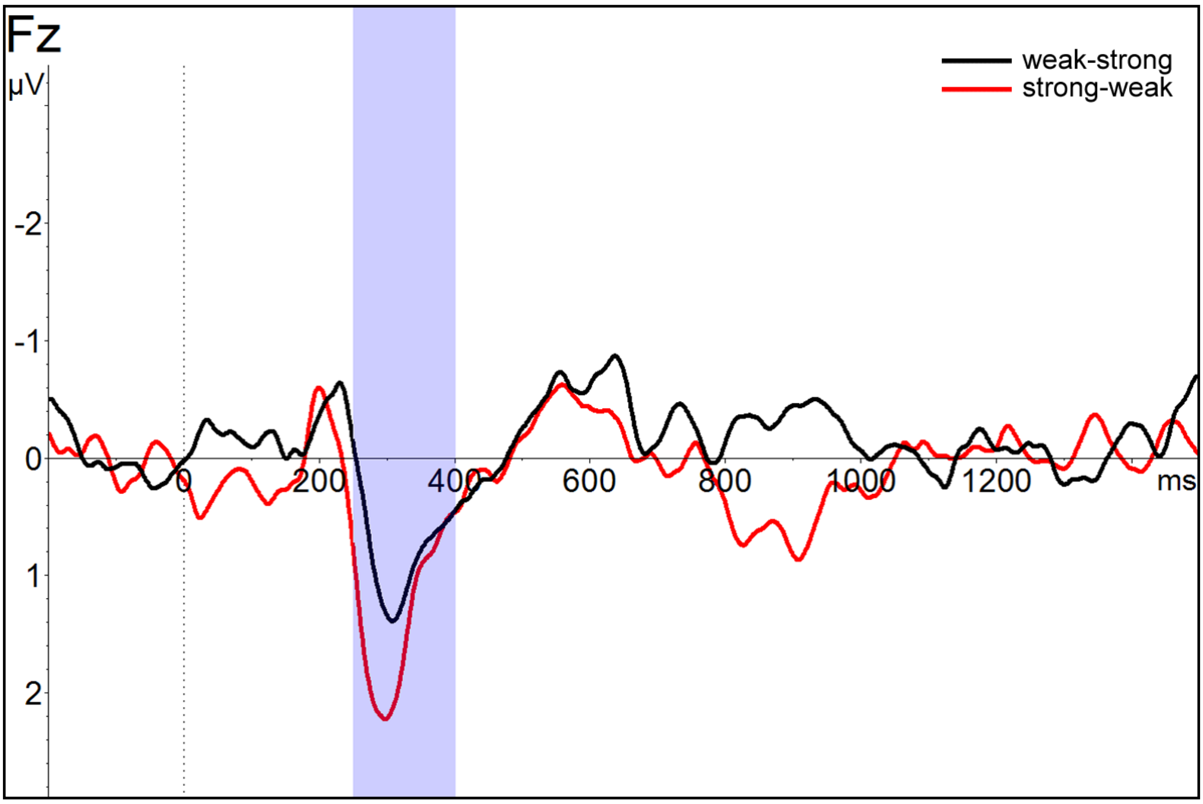


Figure 6.4.4. The ERP in the two conditions (i.e., strong-weak, weak-strong) at the electrode Fz for Block 2: passive listening task.

6.4.2.2.2. Congruency effect

Time window 600-1000 ms

Block 1: stress discrimination task

For lateral electrodes, neither main effect of Congruency [F (1, 16) = 1.80, MSE = 0.383, $p > 0.05$] nor interactions were found. For the midline electrodes, there was a significant main effect of Congruency [F (1, 16) = 7.00, MSE = 0.528, $p < 0.05$, $\eta 2_p = 0.30$], indicating a larger negativity for incongruent trials (-0.02 ± 1.76 μV) than for congruent trials (0.25 ± 1.63 μV). Furthermore, a marginally significant Congruency × Electrode interaction [F (2, 32) = 2.69, MSE = 1.279, $p = 0.083$, $\eta 2_p = 0.14$]. Post-hoc analyses revealed a larger negativity for incongruent trials (0.64 ± 1.62 μV) than for congruent ones (1.33 ± 1.90 μV) at the Pz electrode [F (1, 16) = 6.65, MSE = 1.230, $p < 0.05$, $\eta 2_p = 0.29$] (see Figure 6.4.5).

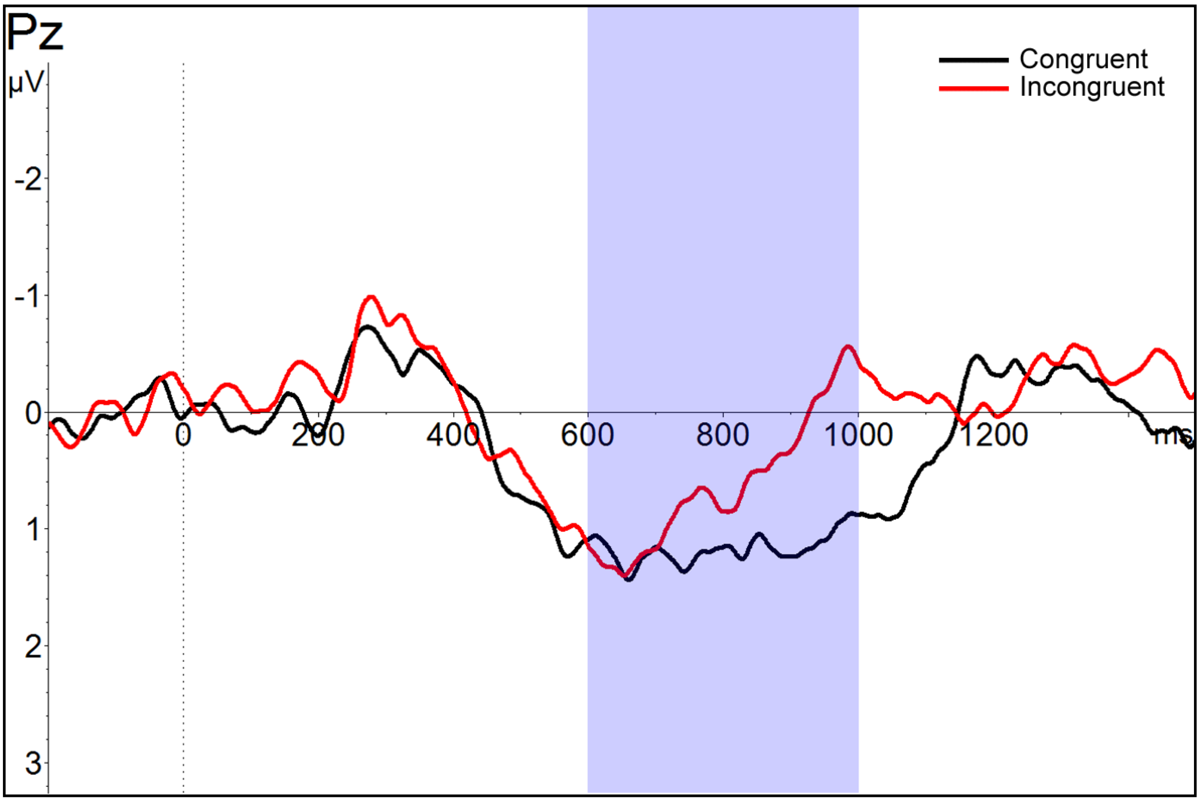


Figure 6.4.5. The ERP in the two conditions (i.e., strong-weak, weak-strong) at the electrode Pz for Block 1: stress discrimination task.

Block 2: passive listening task

For lateral electrodes, there was no main effect of Congruency [$F < 1$] nor interactions. For the midline electrodes, a significant interaction between Congruency and Electrode was found [F (2, 32) = 8.297, MSE = 0.844, $p < 0.01$, $\eta 2_p = 0.34$]. Post-hoc analyses revealed a larger negativity for incongruent trials (-0.32 ± 1.22 µV) than for congruent ones (0.52 ± 1.50 µV), but only at the Fz electrode [F (1, 16) = 9.81, MSE = 1.235, $p < 0.01$, $\eta 2_p = 0.38$] (see Figure 6.4.6).

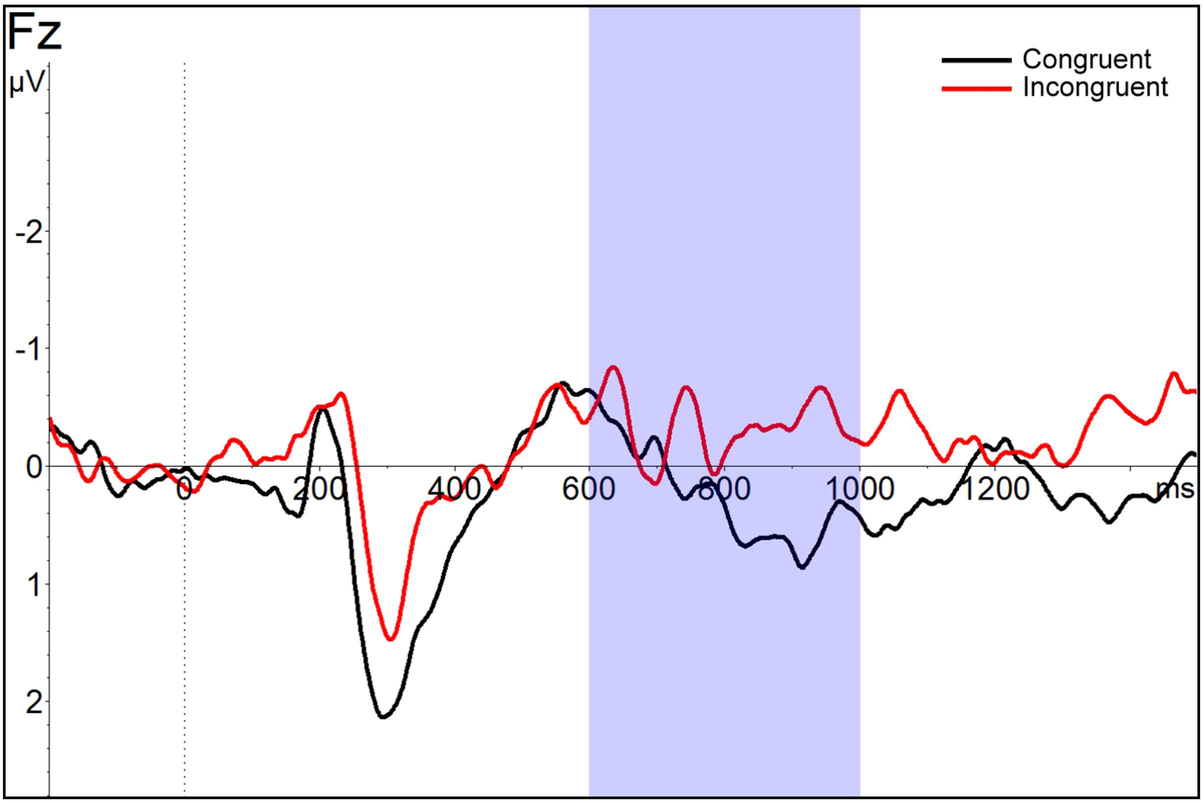


Figure 6.4.6. The ERP in the two conditions (i.e., strong-weak, weak-strong) at the electrode Fz for Block 2: passive listening task.

6.4.3. Discussion

The present experiment investigated ERPs correlate of the rhythmic property of Mandarin Chinese VN/VO minimal pairs. These VN/VO pairs have been shown in the previous production study (Experiment 3) to differ phonetically, with left stress in VN and right stress in VO phrases and that a different stress pattern (strong-weak, weak-strong) was reflected in acoustic features in F0, duration and intensity. Twenty pairs were auditory presented in an oddball-like paradigm as the one used in Experiment 1. Participants performed an explicit task (stress discrimination) and an implicit task (passive listening). The ERP data revealed two components related to Mandarin Chinese stress pattern processing of VN/VO pairs (N200 & P200). Moreover, the MMN was found associated to stress pattern congruency.

In Experiment 1, we found a larger P200 elicited by infrequent stress pattern compared to frequent pattern for neutral/full tone minimal pairs. In the present experiment, using VN/VO minimal pairs we replicated the P200. Moreover, the P200 effect was only observed in the passive listening task. The finding was consistent with the previous study (Böcker et al., 1999) that the P200 effect was presented only in the passive listening task. We proposed that the P200 was an exogenous component that reflects the physical/acoustic stimulus parameters.

The N200 amplitude on strong-weak words was larger than on weak-strong words in the stress discrimination but not for passive listening task. In Experiment 1, we found N200 in response to stress pattern of neutral vs. full tone pairs in both stress discrimination and passive listening task. We suggested that N200 indexes the automatic 'stress pattern templates' matching process during accessing the mental lexicon. We argue that the absence of the N200 in passive listening task was due to the less importance of the stress pattern distinction in VN/VO than the distinction of neutral/full tone in Chinese. A misstressed neutral/full tone may result in interpretation error. For example, if the compound word *dong1xi0* 'thing' is misstressed with a full tone at the final syllable instead of the neutral tone, listeners would misunderstand the meaning of the sequence because the meaning of the sequence with a full tone at the final syllable (i.e., *dong1xi1*) is 'the east and the west; however, the one with neutral tone at the final syllable (i.e., *dong1xi0*) represents completely different meaning, that is a 'thing". In contrast, the misstressing in VN/VO would not cause critical semantic change (no lexical value). For instance, the verb-object sequence *CHAO3fan4* with a left stress means 'fried rice' and *chao3FAN4* with right stress means 'fry the rice'. Therefore, when the task did not require to process stress information, stress pattern search and matching in the mental lexicon would not occur (absence of N200 in passive listening task), only if attention was paid to stress information, stress pattern search and matching processes would take place (observation of N200 in stress discrimination task).

Furthermore, the present experiment did not reveal the endogenous stress effect N325 in both tasks. According to Experiment 1, the N325 is of controlled process reflecting the perception of the stress pattern following by of the automatic stress pattern processing indexed by the N200 and P200. The absence the N325 even for the explicit task could be due to task strategy. We found that response times were fast (628.1 ms averaged of all trials), even earlier than the offset of the stimuli (724.7 ms averaged of all stimuli) which implied that participants might develop the strategy based on the acoustic information of the first syllable, that is when a more prominence was detected in the first syllable it was a strong-weak sequence, and when the first syllable was less prominent it was then of weak-strong. Consequently, it did not involve a stress pattern process but an acoustical process on the first syllable. Therefore, the N325, which associated to stress pattern perception, would be not observed.

With regard to the congruency effect, we observed a late sustained negative-going potential between 300 and 1000 ms in both tasks. The findings were in line with the Experiment 1. As proposed in Experiment 1, we considered this late task-independent negativity as the MMN evoked by incongruent stress pattern reflecting the stress pattern congruency effect in the *oddball*-like paradigm.

6.4.4. Conclusion

The present experiment investigated ERPs correlate of the rhythmic property of Mandarin Chinese VN/VO pairs. The ERP data revealed two exogenous (N200 & P200) related to stress pattern processing. The N200 was larger for strong-weak stress pattern in VN compound than for frequent weak-strong pattern in VO phrase in stress discrimination task and the P200 was larger for VN compound than for VO phrase in passive listening task. In contrast, the endogenous component (N325) was not found as short response times suggested that the distinction of stress pattern was based on task strategy. In addition, the MMN was found in response to stress pattern congruency.

Chapter 7

General discussion and Conclusion

7. General discussion and Conclusion

The present thesis investigated the role of stress pattern in auditory processing of compound words and phrases in Mandarin Chinese. The goals were to determine whether the stress pattern information was processed in compound words recognition and then to which extent the stress pattern information is implicated in lexical processing. Two types of compound/phrase minimal pairs were used: (1) Compound words with a neutral tone (strong-weak) vs. phrases with a full tone (strong-strong) distinguished by the final syllable tone realization, (2) VN compound words (strong-weak) and VO phrases (weak-strong) distinguished by the position of the stress.

In order to confirm the contrastive stress pattern in these two types of minimal pairs, phonetic analyses were firstly conducted. For the contrastive stress in the neutral tone and the full tone, we analyzed a set of 60 pairs of Chinese disyllabic words in which the first syllable carried a full tone, and the second carried either a neutral tone (i.e., strong-weak) in one item of the pair, or a full tone (i.e., strong-strong) in another item of the pair. The results replicated the findings of previous studies on the Mandarin neutral tone that the neutral tone has shorter duration, smaller F0 variation and less intensity as compared to the corresponding full tone. The neutral tone resulted in significant reduction of the duration (reduced by 51.6%), the F0 variation (reduced by 56.0%), and the intensity (reduced by 12.9%) as compared to the corresponding full tone. Thus, we confirmed the contrastive stress pattern in the neutral vs. full tone minimal pairs (see the section Stimuli in Experiment 1 & 2). For the contrastive stress in the minimal pairs VN vs. VO, we carried out a production experiment (Experiment 3). Duanmu (1990, 2000) proposed that Chinese VN and VO were distinguished by their stress pattern; moreover, the stress pattern follows the same Compound Stress Rule and the Nuclear-Stress-Rule (Chomsky & Halle, 1968) as stress languages like English. According to the Compound Stress Rule and the Nuclear-Stress-Rule, stress placement is different between compounds and the corresponding phrases, with left stress for compound and right phrasal stress for the phrase. However, there is little acoustic evidence for testing Duanmu's assumption on the distinction of stress pattern in Mandarin Chinese VN/VO minimal pairs. We are the first in the present doctoral thesis to report the acoustic results in Chinese VN/VO, which demonstrated that (1) the left syllable in VN was longer than the one in VO, and that the right syllable was more lengthened in VO than in VN; (2) the F0 range was larger in the right stressed syllable in VO than the one in VN; The intensity was higher in the left syllable in VN than the one in VO and higher in the right syllable in VO than the one in VN. These results confirmed the left stressed pattern in VN (strong-weak) and the right stressed pattern in VO (weak-strong). The acoustic study also lends support to the hypothesis that principles of stress upward of word level are universal through different languages.

Three ERP experiments have been then carried out to examine whether the stress pattern information represented by acoustic cues in these minimal pairs was processed during compound words recognition and then to which extent the stress pattern information plays a role in lexical processing in morphologically complex sequences in Chinese. All three experiments revealed an important role of the stress pattern in Mandarin Chinese processing.

With regard to the stress pattern perception, Experiment 1 showed two exogenous N200/P200 components and an endogenous component N325 in response to stress pattern processing of neutral (strong-weak) vs. full tone (strong-strong) minimal pairs. Using the same paradigm with VN/VO minimal pairs, Experiment 4 revealed the exogenous N200/P200 components but no endogenous component N325 in response to stress pattern in VN/VO. Both the two experiments showed the MMN effect in response to a deviant (incongruent) stress pattern compared to the stress pattern in three preceding sequences.

Exogenous stress effect: N200/P200

In Experiment 1, the N200 amplitude on infrequent stress pattern was larger than on infrequent stress pattern in both stress discrimination and passive listening tasks with no amplitude difference between the explicit and the implicit task. We considered the N200 to be a task-independent exogenous stress effect. Generally, the N200 was linked to unattended deviant auditory stimuli in oddball paradigms. Our observed N200 had a frontocentral topography, the same as the one in oddball paradigms (Näätänen et al., 1978; Näätänen &

Gaillard, 1983). In order to clarify the nature of the N200, in Experiment 1 (Block 3 & 4) we recorded the ERP for the delexicalized stimuli, in which lexical information was filtered with the conservation of stress pattern information (e.g., duration, F0). Results on the delexicalized stimuli did not show the N200. We therefore argued that the N200 observed on non-delexicalized stimuli was not just a response to unattended deviant auditory stimuli, that is to say it was language specific in response to the processing of stress pattern. We suggest that N200 may index an automatic 'stress pattern templates' matching process during accessing the mental lexicon. Stress pattern templates may be stored in the mental lexicon and may help to assign rapidly and accurately auditory speech input into the relevant stress patterns categories. The recognized stress pattern may then participate in different aspects of lexical processing, including morphological processing (e.g., determine the ambiguous morphological status of verb-object sequence) and semantic processing (e.g., determine the semantic access of constituents). A correct stress pattern thus facilitates the lexical processing as found in many studies (Slowiaczek, 1990; Small et., 1988; Koester & Cutler, 1997). The findings suggest that the mental lexicon probably store suprasegmental information, i.e., stress pattern in addition to segmental information of words. Previous neuroimaging studies showed that in the human brain whereas the right temporal region is thought support the identification of prosodic parameters, the right frontal cortex is involved in the processing of sentence melody (Friederici, 2002) but also for a neurophysiological dynamic dual pathway model discussing the lateralization of auditory language functions, see Friederici & Alter, 2004)

The P200 was observed for delexicalized stimuli in both tasks in Experiment 1 and for non-delexicalized stimuli only in the passive listening task in Experiment 4. In both experiments, the amplitude of the P200 was larger for infrequent stress pattern than for frequent one. The P200 component is an exogenous component assumed to reflect perceptual processing (Hillyard & Picton, 1987). The P200 has been found to be sensitive to acoustic properties (Marie et al., 2011; Friedrich et al., 2001; Böcker et al., 1999). Marie et al. (2011) manipulated the metric structure of French words by lengthening the penultimate syllable of trisyllabic words, and found that metrically incongruous words elicited larger P200 components than metrically congruous ones. Furthermore, the P200 effect was larger for musicians than for non-musicians. It was proposed that the amplitude enhancement of the P200 may reflect the automatic processing of the temporal attributes of words and, specifically, of the words' syllabic structure. The P200 in response to stress pattern information has been also observed with German words. Friedrich et al. (2001) found that German infrequent weak-strong words elicited a larger P200 than frequent strong-weak words. Although the P200 component is of exogenous, it has been shown to be sensible to task, Böcker et al. (1999) revealed the P200 effect only in the passive listening task. Our P200 in Experiment 1 was found in both explicit and implicit tasks but only for delexicalized stimuli. In Experiment 2, the P200 was observed only in the passive listening task. Taken together, our data suggest that the P200 reflects a non-language specific processing of acoustic stimulus parameters. When stress pattern processing was enhanced by using either explicit task (stress discrimination) or by using language stimuli (non-delexicalized), the P200 would be neutralized by a larger negativity (i.e., N200) in response to stress pattern processing, as the two components share similar time windows in the present study.

Endogenous stress effect: N325

We replicate in Experiment 1 the endogenous stress effect N325 previously reported in the study of Böcker et al. (1999), with the same frontocentral topography. However, in our study the peaking latency was later (at 405 ms post-stimulus). This might be due to a longer duration (393.90 ± 74.20 ms) of the first syllable in our stimuli compared to the duration (292 ± 57 ms) of the stimuli in the study of Böcker et al. (1999). In average, our the first syllable of our stimuli were about 100 ms longer, which can explains the delayed peak latency of the N325 we recorded. Böcker et al. (1999) showed that the N325 would be influenced by task, the amplitude de N325 was attenuated in passive listening task, which suggested the endogenous nature of the effect. In our experiment with Chinese stimuli, we observed the N325 only presented in stress discrimination task for non-delexicalized stimuli. It was not found for the implicit passive listening task, or for delexicalized stimuli. The results suggest that the N325 was task-dependant (it needs listeners to pay attention to the stress pattern) and language-specific (non-language stimuli would not evoke the N325). Unlike the exogenous N200/P200 stress effects reflecting automatic stress pattern processing the N325 is of controlled process and may reflect an output of the automatic stress pattern processing. More precisely, when attention was paid to stress pattern, the output of the automatic stress pattern processing was kept to produce a perception of stress pattern; when attention was not paid, the output of the automatic stress pattern processing would not be used to draw a perception of stress pattern. However, our ERP data on VN/VO minimal pairs in Experiment 4 did not reveal an endogenous stress effect N325 in both tasks. This was due to the less importance of the stress pattern distinction in VN/VO than the distinction of neutral/full tone in Chinese. A mis-stressed neutral/full tone may result in recognition error. For example, if the compound word *dong1xi0* 'thing' is mis-stressed with a full tone at the final syllable instead of the neutral tone, listeners would misunderstand the meaning of the sequence because the meaning of the sequence with a full tone at the final syllable (i.e., *dong1xi1*) is 'the east and the west; however, the one with neutral tone at the final syllable (i.e., *dong1xi0*) represents completely different meaning, that is 'a thing/an object''. In contrast, the mis-stressing in VN/VO would not occasion critical semantic change. For instance, the verb-object sequence *CHAO3fan4* with a left stress means 'fried rice' and *chao3FAN4* with right stress means 'fry the rice'. We argue that the good accuracy and short response times in the stress discrimination task was not due to the recognition of the stress pattern in VN/VO, and that it was due to some strategy processes, such as processing of the physical/acoustic stimulus parameters of the first syllable in VN/VO. As our stimuli had either strong-weak or weak-strong stress pattern, participants might have developed a strategy, that is when a more prominence was detected in the first syllable it was a strong-weak sequence, and when the first syllable was less prominent it was then of weak-strong. The strategy based on the first syllable was implied by the short response times (628.1 ms averaged of all trials). It was even earlier than the offset of the stimuli (724.7 ms averaged of all stimuli).

Congruency effect: MMN

The mismatch negativity (MMN) was expected for targets in incongruent trials regardless of stress pattern. In Experiment 1, a late sustained negative-going potential between 300 and 1000 ms was consistently found in all four blocks, with explicit or implicit task and non-delexicalized or delexicalized stimuli. We argued that this negativity is an MMN. Firstly, the negativity mirrors automatic processes because of its observation for implicit task. Secondly, it was not sensible to the nature of stimuli, language and lexicalized non-language stimuli both evoked this effect. Finally, this negativity peaked around 600 ms after the onset of the stimulus, too late for a MMN, however, ff we subtracted the duration of the first syllable (393.9 ms averaged of all stimuli), the latency of the negativity matched the typical latency of MMN (150-250 ms). Listeners received critical information from the onset of the second syllable to determine the stress pattern. It is thus reasonable to subtract the duration of the first syllable from the latency of the negativity. The late MMN reflected indeed the congruency effect in the *oddball*-like paradigm. These findings were in line with what has been proposed concerning the MMN, that is, in addition to changes in short-term representations of auditory stimuli, the MMN can also be elicited by changes in long-term representations of linguistic information (see Pulvermüller & Shtyrov, 2006 for a review). That is the case for stress pattern. In Experiment 4, we observed as well a late sustained negative-going potential between 300 and 1000 ms in both tasks. We considered this late task-independent negativity as the MMN evoked by incongruent stress pattern. Taken together, these finding suggest that stress pattern changes are processed relying on long-term representation of stress, not at single syllable level.

With regard to the implication of stress pattern in lexical processing, Experiment 2 showed both behavioral evidence and neurophysiological evidence. Behavioral data revealed shorter response times for related target word as compared to unrelated ones for both types of prime (i.e., full-full tone phrase and full-neutral tone compound) when the target was presented at the onset of the prime (onset of C1). However, results revealed an reverse priming effect for full-neutral tone compound when the target was presented at the onset of C2, i.e. the middle of the prime. This pattern suggested that the stress information of the last syllable of the prime modulated the semantic propriety of the first morpheme of the prime, when the stress information was processed, the semantic representation of the first morpheme of the prime was eliminated (became semantically opaque) resulting in more difficulty in the lexical access of a semantically related target word (longer response times).

ERP data mirrored the behavioural data. We found consistent N400 associative priming effects were for both strong-weak compound words and strong-strong phrase when the second syllable that carried the stress pattern information was not processed. Interestingly, when the second syllable was processed (i.e., stress pattern information was analyzed), the N400 effect inversed its polarity for strong-weak compound words reflecting that there was no more activation of the semantic representation of the first morpheme of the prime. In contrast, the N400 effect was preserved for strong-strong phrase. This pattern of results was expected and it may reflect a modulation of the processing of the first constituent morpheme by the stress information in the second syllable. These findings suggested that the stress information in the last syllable of the ambiguous sequence served as an access code for the selection of the processing routes to allow a successful access to the compound word. When the second morpheme (i.e., carrying the stress information) was processed, a decomposition route would be turn off and inhibited, a direct route would be reinforced, consequently the first constituent was not activated thus it would not prime the semantically related target.

The Prosody-Assisted-Processing (PAP) model for compound words processing assumes that the length of the first lexical morphemes of compound words is a determining factor for setting off/triggering a decompositional route. At the onset of the speech input, a direct route is activated by default. Simultaneously, a prosodic analysis of the left-most morpheme is automatically carried out. The output of the prosodic analysis is crucial with respect to the activation of the decompositional route. If this output informs the processor that the left-most morpheme belongs to a compound word (i.e. short syllable), then the processing system activates a decompositional route. On the contrary, if this prosodic analysis indicates a monomorphemic word (i.e. long syllable), the processing system continues to match the acoustic-phonetic input with the word detectors at the word level, until the appropriate detector reaches the threshold for best match. In this case, the processing system has set a monomorphemic route configuration. Because of the sequential involvement of the two routes, the PAP model is considered as a cascading parallel model.

Regarding our results in Mandarin Chinese, we observed a picture showing that at the onset of the speech input, a decompositional route is activated by default. Constituent morphemes are lexically processed. Simultaneously, a prosodic analysis of the stress pattern is automatically carried out. When the prosodic analysis reaches to the last syllable of the input, the stress pattern could then be recognized. The output of the prosodic analysis is crucial with respect to the activation of the direct route. If this output informs the processor that the last morpheme carried a neutral tone, then the processing system activates a direct route and deactivates the decompositional route. On the contrary, if this prosodic analysis does not detect a neutral tone, the processing system continues the constituent morpheme processing in addition to the activated direct route. Our findings extend thus the PAP model by using compound words and phrase, more importantly by adding a right-to-left retroactive processing of prosodic information in Mandarin Chinese.

In conclusion, the present doctoral thesis investigates the processing of stress pattern (i.e. word stress versus phrasal stress) and its implication on the auditory processing of disyllabic Mandarin Chinese ambiguous minimal pairs constituted of a compound word and a phrase. Two types of compound word/phrase minimal pairs were used (1) Compound words with a neutral tone on the final syllable (strong-weak) vs. Phrases with a full tone on the final syllable (strong-strong); (2) Verb-Noun (VN) compound words (strong-weak) and Verb-Object (VO) phrases (weak-strong) distinguished by the stress pattern. In a production study, phonetic analyses of the minimal pairs revealed a contrastive stress pattern indexed by longer duration, larger F0 variation and higher intensity in stressed syllables. At the neurophysiological level, the processing of the contrastive stress patterns of the two types of compound word/phrase minimal pairs was associated with two exogenous event-related brain potentials: An N200 thought to reflect an automatic language-specific marker of stress pattern on the one hand, and an P200 assumed to sign the processing of physical/acoustic cues, on the other hand. In addition to these two exogenous ERP components, we showed that one endogenous component, i.e. the N325, varied as a function of the stress pattern in the minimal pairs. Indeed, the N325 was larger for infrequent stress pattern, but this ERP effect was only observed in the stress discrimination task for non-delexicalized stimuli. We therefore argue that the N325 effect in response to the processing of stress pattern in Mandarin Chinese may be a controlled language-specific ERP marker of stress pattern, possibly somehow related to allocation of attentional resources. Additionally, we found a late sustained MisMatch Negativity (MMN) in response to stress pattern congruency regardless of stress pattern and task. This late MMN suggests that stress pattern changes are processed relying on long-term representation of stress, not at single syllable level. Finally, our neurophysiological data showed that the neutral tone in the minimal pairs was used by the processing system for assisting the analysis of the morpholexical structure of the ambiguous spoken sequences in Mandarin Chinese under investigation here. We demonstrated that once a neutral tone was detected, a right-to-left retroactive processing of the morphological structure was triggered in addition to the usual sequential left-to-right one. We assume that the right-to-left retroactive processing ensures a successful lexical access to the semantically opaque compound words in the minimal pairs. However, we cannot exclude that some acoustic information already carried in the first constituent may have informed the processing system that the sequence under processing is a compound word as our acoustic measures showed that the duration of the first constituent is shorter when it is followed by a neutral tone than by a full tone. Based on the present findings, we propose that neutral tone at the final syllable, probably combined with duration of the initial syllable, could be two *reliable* prosodic cues for assisting the analysis of the lexical morphology in Chinese. We concluded that, as already reported in other languages with various grammatical sequences, prosody might also have a function of prediction of ongoing grammatical information in Mandarin Chinese ambiguous minimal pairs. However, and most importantly, the present work highlights that in the Chinese minimal pairs we used, prosodic stress pattern delivered at the end of the sequences was employed for triggering a retroactive processing of the whole sequence. Taken together, the combined behavioral (production and perception studies) and neurophysiological processing studies presented in the current doctoral thesis allowed us to precise the functional and structural description of the Prosody-Assisted-Processing (PAP) model for Mandarin Chinese, especially by proposing an additional right-to-left retroactive processing route whose function is to morphologically reanalyze the Chinese sequences that carry a neutral tone on their final syllable.

Chapter 8

Applications in learning Mandarin Chinese pronunciation

8. Applications in learning Mandarin Chinese pronunciation

8.1. Pronunciation learning in Mandarin Chinese

Pronunciation is often regarded as one of the most challenging parts of Chinese learning for non-tonal languages speakers. As a typical tonal language, the same Chinese syllable can be pronounced with four tones, which link to different lexical meanings. Chinese pronunciation learning was thus mainly based on the training of the pronunciation of the four tones. It is a fact that Chinese stress phenomenon, such as the neutral tone is usually neglected in Chinese pronunciation learning. As introduced in Chapter 3, the neutral tone represents an important part in the Chinese lexicon. It is claimed that in Standard Mandarin, about 15-20% of the syllables in written texts carried a neutral tone (Li, 1981). The negligence of the neutral tone in Chinese pronunciation learning is because a misstressing of a neutral tone causes usually just a non-native accent but no problem in understanding. However, the neutral tone may play an important morpho-lexical role for some disyllabic Chinese sequences. For example in the minimal pair *dong1xi0* (compound word, means 'thing') vs. *dong1xi1* (phrase means 'east and west'), the neutral tone in *xi* determines the morphological structure and the meaning of the sequences. Therefore, the training on the neutral tone pronunciation/perception will interest learners to improve their pronunciation and comprehension in Chinese.

Learning pronunciation in general relies much on feedback from a teacher. It is very difficult to learn proper pronunciation independently. Although learners can get far by mimicking and listening to their recordings, but receiving feedback is essential. The problem is that most learners do not have native speakers around to ask all the time. A number of computer-based learning methods (e.g. software and application) have been developed to meet this requirement. For example, using pitch detection techniques[41] the software application *WaiChinese*[42] functions a Chinese tone visualizer that displays the shape of a spoken tone in

[41] The pitch detection techniques permit to estimate the tone contour or pitch of speech. See section 8.2.1 for an introduction.

[42] *WaiChinese* is available at www.waichinese.com

real-time. The display of tone contour serves as a visual feedback of learner's pronunciation quality with which learners can correct their pronunciation by mimicking the correct tone contour. Inspired by *WaiChinese,* we decided to develop a new application with more functions that helps learners pronounce better Chinese. This application would:

1) allow learners to perform pronunciation exercises on Chinese tones and get immediate feedbacks of the tone contour by using pitch detection techniques;
2) we would use the automatic speech recognition techniques which gives a great advantage for learning pronunciation because it can be served as a feedback on whether learner's pronunciation on a word is correct or not, that is to say, if the pronunciation is correct, it will lead to a correct recognition by the machine, if the pronunciation is bad, it will not be recognized;
3) we would also create a perception test on tone judgment including the neutral tone;
4) finally we would create a Chinese pronunciation dictionary to facilitate learners to get standard pronunciation;

The development of the application has been realized. Some screenshots of the application on the iPhone are shown below. We will present in the next section, the development/implementation of the application on the iOS platform[43].

---

[43] The iOS is a mobile operating system created and developed by Apple Inc. and distributed exclusively for Apple hardware including the iPhone, iPad, and iPod touch.

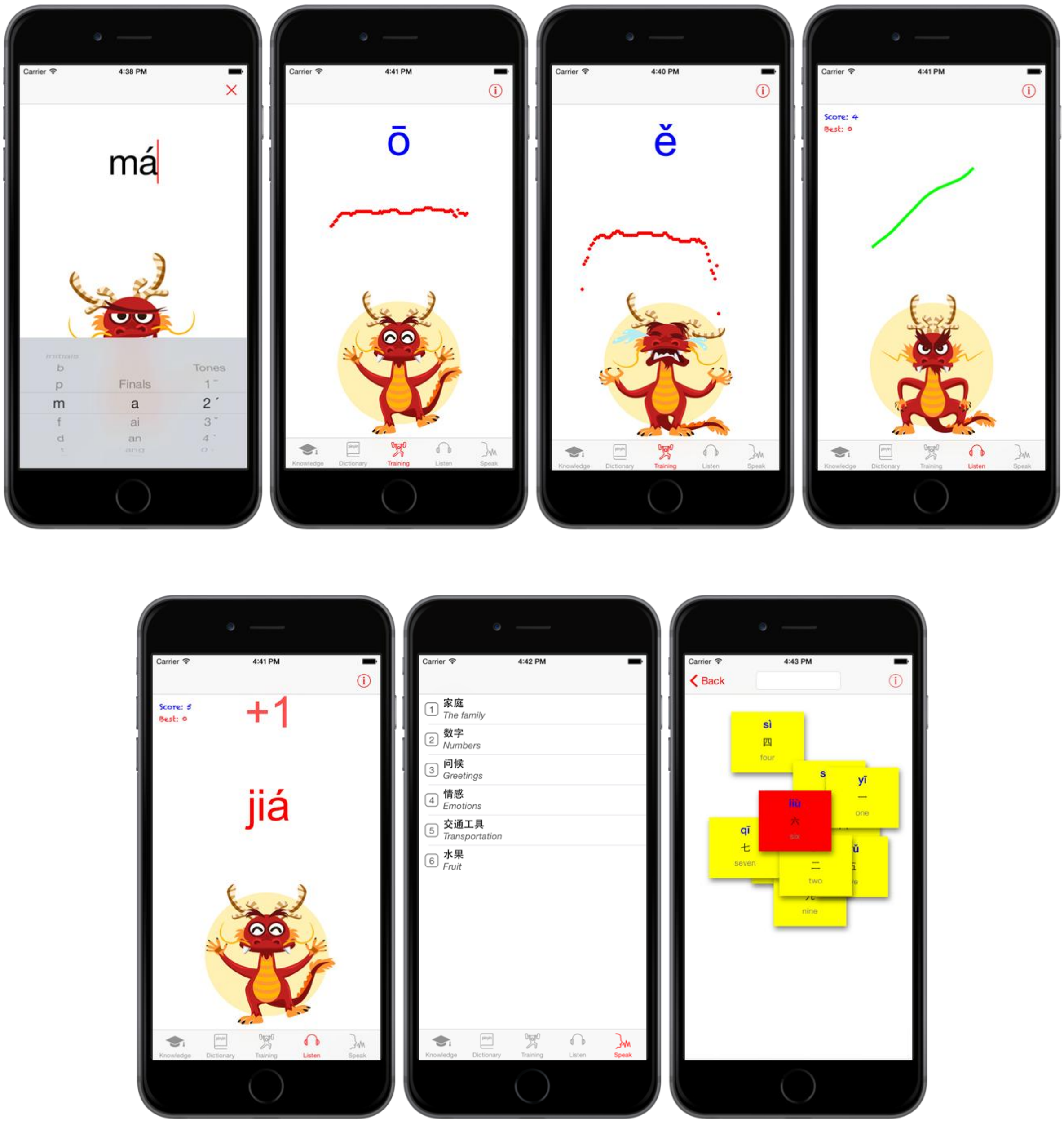


Figure 8.1. Screenshots of the application on the iPhone.

The first screenshot shows the interface of the pronunciation dictionary. The second and the third ones show the training of tone pronunciation with the red curves displaying the tone contour of learner's pronunciation (in this example, good pronunciation in the second screenshot and bad pronunciation in the third one). The fourth screenshot shows learner's response in the tone judgment test and the fifth shows the feedback. The last two screenshots show the pronunciation test using the automatic speech recognition where the first one displays all available categories (e.g., digits, family members, fruits etc.) and the second one shows the automatic speech recognition interface.

8.2. Implementation of the application

The present iOS-based application runs on the popular Apple mobile devices: iPhone, iPad, and iPod Touch. The application was developed in the language Objective-C[44] under Xcode[45]. The main algorithms for the pitch detection and the automatic speech recognition will be discussed in detail. The user interface design and the database design will not be mentioned.

8.2.1. Pitch Detection Algorithm (PDA)

The display of tone contour serves as a visual feedback of learner's pronunciation quality with which learner can correct their pronunciation by mimicking the correct tone contour. The tone contour or pitch can be estimated by using some methods of signal processing, so-called Pitch Detection Algorithm (PDA) algorithm. A number of methods have been proposed for pitch detection. The time-domain-based AutoCorrelation Function methods (ACF; Rabiner, 1977) are ones of the most robust and reliable. The ACF is calculated directly on the waveform in time domain and is a straightforward computation. The present application used the ACF for realizing the pitch detection.

8.2.1.1. Autocorrelation Function (ASF)

Autocorrelation function is the cross-correlation of a signal with itself at different points in time. Given a discrete time signal *x(n)*, the autocorrelation function is generally defined in: (where *m* represents lag)

$$D_{ACF}(m) = \frac{1}{N}\sum_{n=0}^{N-1} x(n) \times x(n+m)$$

The autocorrelation function is basically a transformation of the signal that is useful for

---

[44] Objective-C is the primary programming language for developing software for OS X and iOS. It is a superset of the C programming language and provides object-oriented capabilities and a dynamic runtime.

[45] Xcode is an integrated development environment (IDE) containing a suite of software development tools developed by Apple Inc. for developing software for OS X and iOS. Xcode supports Objective-C, Swift, and others languages like C, C++, Java, Python etc. We used the version 6.4.

displaying the periodic characteristic in the waveform. If we assume $x(n)$ is exactly periodic with period p, i.e., $x(n) = x(n + p)$, then it is easily shown that:

$$Dx(m) = Dx(m + p)$$

therefore, the autocorrelation function has also the same period p. Conversely, periodicity in the autocorrelation function indicates periodicity in the signal.

The autocorrelation function involves a series of speech processing. The implanted speech processing in the present application are 1) Pre-emphasis, 2) Framing, 3) Windowing, 4) Autocorrelation, 5) Smoothing. Figure 8.2 shows a block diagram of autocorrelation function.

Figure 8.2. Block diagram of autocorrelation function.

(1) Pre-emphasis

The first step emphasizes higher frequencies by passing of signal through a filter. The process will improve the overall signal-to-noise ratio[46]. The used filter can be defined in:

$$x'_{(n)} = x_{(n)} - 0.95 * x_{(n-1)}$$

This filter increase higher frequencies by 20 dB gain.

(2) Center clipping

The nonlinear processing is usually used in pitch detection to reduce the effects of the formant structure on the detailed shape of the autocorrelation function. We used the center-clipping method (Sondhi, 1968), which could be described by following relation:

$$y(n) = \begin{cases} x(n) - C_L, & x(n) > C_L \\ 0, & |x(n)| \le C_L \\ x(n) + C_L, & x(n) < -C_L \end{cases}$$

The clipping threshold (C_L) was chosen to 30% of the max amplitude.

[46] Signal-to-noise ratio is a measure that compares the level of a desired signal to the level of background noise.

(3) Framing

The speech signal should be segmented into small duration blocks of 20-40 ms known as frames to perform the short-term analysis. The short-term analysis is required as speech is a time varying signal but when it is examined on short time scales, its properties are stationary. In our system, we defined 1024 samples per frame (i.e. 23 ms at the sampling of 44.1k Hz), with adjacent frames separated by 512 samples (i.e. 50% size of the frame).

(4) Windowing

To reduce discontinuity at both ends of each frame, the original signal was multiplied by a Hamming window. The Hamming window is optimized to minimize the maximum (nearest) side lobe and the window function is given by

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{L-1}\right) \qquad 0 \geq n \geq L-1$$

(5) Autocorrelation Function

The autocorrelation function can be defined in:

$$D_{ACF}(m) = \frac{1}{N}\sum_{n=0}^{N-1} x(n) \times x(n+m)$$

N = 1024; m represents lag; for each frame the autocorrelation was calculated from lag = 0 to 1023.

(6) Smoothing

When we calculate the F0, there are often some outliers. To minimize these outliers, we have chosen to use a median smoothing. A window size of five was used, that is to say, for a F0 value at the time of t, i.e. F0(t), the five F0 values at t-2, t-1, t, t+1 and t+2 were involved in the smoothing. The function takes the median of the five values as the smoothed F0 at the time t.

After performing the above six steps, we obtain the F0 contour of a pronunciation of tones.

An algorithm was thus designed to classify various contours into the four tones. For each tone, the algorithm is described as follow:

(1) Tone 1: the tone 1 is plat and relatively high. To evaluate the flatness of the F0 contour, we calculate the standard deviation[47] on values of F0 contour. If only the standard deviation is less than a threshold, the F0 contour is classified into the tone 1. The threshold is defined at 5 Hz.
(2) Tone 2 & 4: the tone 2 and tone 4 has ascendant or descendant contour. The slope of the F0 contour is estimated by calculating the mean increment in each F0 value. A threshold +0.8 Hz is used for tone 2 and a threshold -0.8 Hz for tone 4.
(3) Tone 3: the tone 3 has a descendant then ascendant contour. In order to classify the complex contour, we firstly try to find the minimum F0 in the contour, and then estimate the slopes of the first and the second parts of the F0 contour divided by the minimum point. The same criteria used in the classification for tone 2 & 4 are applied. The tone 3 should show a descendant then ascendant contour.

The following Figure 8.3 illustrates two examples of the pitch contour classification.

[47] In statistics, the standard deviation is used to quantify the amount of variation or dispersion of a set of data values.

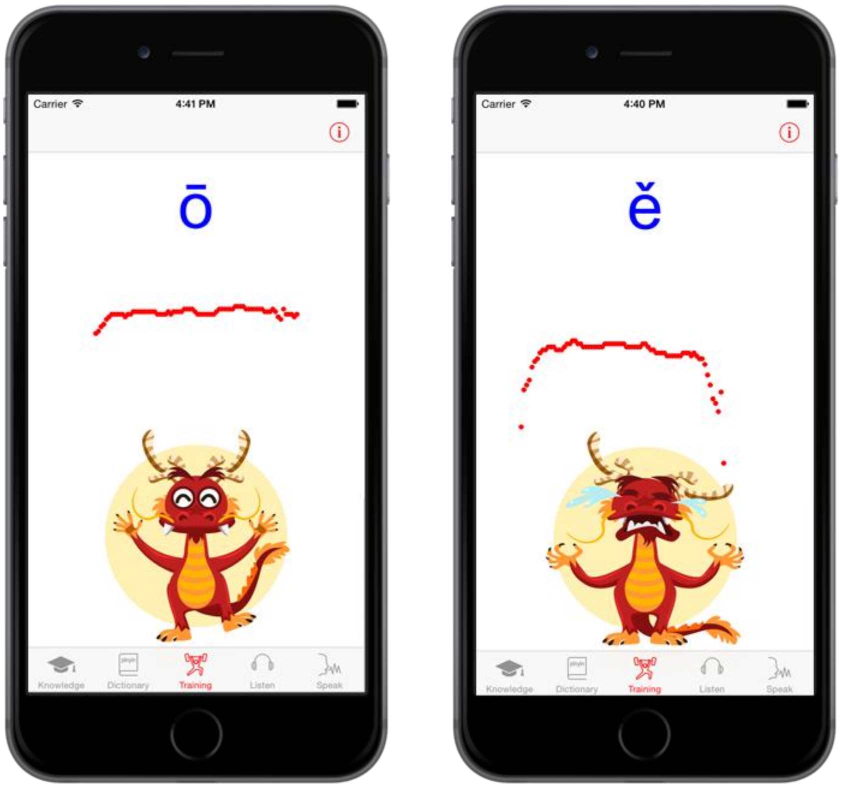


Figure 8.3. Screenshots of the pitch contour classification.
The red curves represent the pitch contour that the learner pronounced. In the first screenshot, the learner correctly pronounced the Tone 1 (/ *ō* /). However, as we see in the second screenshot, the pitch contour illustrated that the pronunciation of Tone 3 (/ *ě* /) was not correct.

8.2.1.2. Experimental results

We recorded the pronunciation of six basic vowels (i.e., *a, o, e, i, u, ü*) from a male and a female Chinese native speaker. Speakers repeated all vowels five times with each four tone. It produced sixty samples for each tone. These samples were used to evaluate the pitch detection system. The good classification rate for each tone is presented in Table 8.1.

Table 8.1.

Recognition rate for each tone

Tone	Classification rate
Tone 1	93%
Tone 2	82%
Tone 3	68%
Tone 4	78%

8.2.2. Automatic Speech Recognition (ASR)

We decided to use automatic speech recognition techniques in the application to serve as feedbacks of learner's pronunciation. If the learner's pronunciation is correct, it will lead to a correct recognition by the automatic speech recognition, if the pronunciation is bad, it will not be recognized. There is some commercial or free services dedicated to the automatic speech recognition, but they have some restrictions for the usage and cannot meet all our demand. For example, the famous speech recognition company Nuance Communications, Inc. provides free service[48] of the automatic speech recognition, however the service is based on web services, that is to say, an internet connection is obligatory for the use of the service, more importantly, it is not open source, developers cannot access to the source for adapting their requirements in the source. Another framework of the automatic speech recognition is OpenEars™. It is an open source framework for voice recognition and speech synthesis, however the voice recognition works only in English and in Spanish, and Chinese is not yet available. Therefore, instead of using existing services, we decided to develop our own automatic speech recognition function for the present application.

Automatic Speech Recognition (ASR) is technology that allows a computer to identify the words that a person speaks into a microphone and convert it to written text (Stuckless, 1994). It has a long history with several waves of major innovations. Three main approaches to

[48] The service is free with a limit of number of uses.

automatic speech recognition were sequentially proposed (see Anusuya & Katti, 2009 for a review): 1) Acoustic phonetic approach, 2) Pattern-matching approach and 3) Artificial intelligence approach. The pattern-matching approach has become the predominant method for the automatic speech recognition in the last six decades of the 20th century (Tran, 2000). It involves two essential steps namely, pattern representation and pattern comparison. A speech pattern representation can be in the form of a template or a statistical model (e.g., a HIDDEN MARKOV MODEL). The underlying idea of pattern-matching approach using template is simple. The system stores a set of prototypical speech patterns as reference templates representing the dictionary of candidate's words. Recognition is then carried out by matching an unknown spoken utterance with each of these reference templates and selecting the best matching one. Figure 8.4 shows the block diagram of the basic template-based system.

Figure 8.4. Basic model of template-based system

One key challenge in template-based approach in speech recognition is to derive typical acoustic features for a pattern (template). Another challenge is to temporarily align patterns to account for differences in speaking rates across speakers. A family of techniques has been proposed to overcome the two challenges, such as Mel Frequency Cepstral Coefficients (MFCC; Davis & Mermelstein, 1980) for acoustic feature extraction and dynamic time warping (DTW; Sakoe & Chiba, 1978; Rabiner & Juang, 1993) for dynamic alignment. The template-based approach is simple in the implementation compared to statistical approaches, and therefore has been widely used for speech recognition especially in small vocabulary system (Higgins & Wohlford, 1985; Khalid, Darabkh, Khalifeh, Baraa, Bathech & Sabah, 2013).

The present application used the template-based approach for the automatic speech recognition. The system consists of speech processing inclusive of 1) feature extraction using MFCC, and 2) recognition using DTW.

8.2.2.1. Feature extraction: Mel Frequency Cepstral Coefficients (MFCC)

Feature extraction is the parameterization of the speech signal. This is intended to retain useful information of the signal while to discard redundant and unwanted information (Kesarkar, 2003). In consideration of the large variability in the speech signal, the selection of acoustic features becomes a fundamental problem of speech recognition. The Mel-Frequency Cepstral Coefficient (MFCC) firstly introduced by Davis and Mermelstein (1980) is most widely used. The MFCC is designed using the knowledge of human auditory system. Human perception of frequency (i.e. subjective pitch) does not follow a linear scale in speech signal: humans are much better at discerning small changes in pitch at low frequencies than they are at high frequencies (Stevens, Volkmann, & Newman, 1937). The purpose of the MFCC processor is to mimic the behaviour of the human ears to makes the acoustic features match more closely that humans perceive. Main steps involved in MFCC are 1) Pre-emphasis, 2) Framing, 3) Windowing, 4) Fast Fourier Transform, 5) Mel filter bank, and 6) Cepstrum. Figure 8.5 shows a block diagram of the MFCC function.

Figure 8.5. Block diagram of the MFCC function.

(1) Pre-emphasis

This step emphasizes higher frequencies by passing of signal through a filter. The process will improve the overall signal-to-noise ratio. The used filter can be defined in:

$x'_{(n)}=x_{(n)} - 0.95*x_{(n-1)}$

This filter increases higher frequencies by 20 dB gain.

(2) Frame blocking.

The pre-emphasized speech signal should be segmented into small duration blocks of 20-40 ms known as frames to perform the short-term analysis. The short-term analysis is required as speech is a time varying signal but when it is examined on short time scales, its properties are stationary. In the system, we defined 512 samples per frame (i.e. 32 ms at the sampling of 16k Hz), with adjacent frames separated by 256 samples (i.e. 50% size of the frame).

(3) Windowing

To reduce discontinuity at both ends of each frame, the original signal was multiplied by a Hamming window. The Hamming window is optimized to minimize the maximum (nearest) side lobe and the window function is given by

$$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{L-1}\right) \qquad 0 \geq n \geq L-1$$

(4) Fast Fourier Transform

A Fast Fourier Transform (FFT)[49] was applied to convert each frame from the time domain into the frequency domain. After the FFT, we obtained the spectrum of speech to determine MFCC.

(5) Mel filter bank

As mentioned above, psychophysical studies have shown that human perception of the frequency is in a non-linear way. It is less sensitive at higher frequencies, roughly greater than 1000 Hz. The Mel scale filter bank is a series of triangular band pass filters that have been designed to simulate the band pass filtering believed to occur in the human's auditory system. In our system, we created a bank of 26 Mel triangular filters. Frequencies were converted to Mel scale using the following conversion formula:

$$f_{mel}(f) = 2595 \cdot log\left(1 + \frac{f}{700\,Hz}\right)$$

By applying the Mel filter bank, we obtained 26 Mel spectrums. They were then performed by a logarithmic transformation for the next step.

(6) Cepstrum

[49] A Fast Fourier Transform is an algorithm to decompose a function of time (e.g., a signal) into the frequencies that make it up.

In this step, we converted the log Mel spectrum back to time domain (i.e. cepstrum) using the Discrete Cosine Transform (DCT)[50]. The resulting features of DCT are called MFCC coefficients. For speech recognition, only the 12 lower ones of the 26 coefficients are kept.

As the MFCC coefficients do not capture frame energy that represents important feature of speech, we added the logarithmic signal energy to the 12 MFCC coefficients thus giving rise to 13 MFCC coefficients. The signal energy was calculated by the following formula:

$$E_j = \log \sum_{p=0}^{P-1} x_2^2(p; j)$$

(7) Deltas and Delta-Deltas

These 13 MFCC coefficients describe only the power spectral envelope of a single frame (i.e., in a static way), but it seems like that speech would also have information in its dynamic. It turns out that calculating the MFCC trajectories and appending them to the original MFCC increases recognition performance by quite a bit. We acquired 13 new delta-MFCC coefficients by calculating the change between the original MFCC coefficients of each frame. In same fashion, we calculated the change between the 13 delta-MFCC coefficients to obtain 13 delta-delta-MFCC coefficients. Finally, we got a total of 39 MFCC coefficients for every frame representing the acoustic features of speech in its static and dynamic.

8.2.2.2. Recognition: Dynamic Time Warping (DTW)

After the feature extraction from speech signal of a word, we converted a given word into a series of 39 dimensional feature vectors (i.e. MFCC coefficients). We applied the feature extraction function for every word in the vocabulary (e.g., a small vocabulary of digits from 0

---

[50] A discrete cosine transform (DCT) expresses a signal in terms of a sum of cosine functions oscillating at different frequencies.

to 9), and stored the calculated 39 dimensional feature vectors of each word as a template into a database. The simplest way to recognize an input word is then to compare its features against the stored templates and determine the best match. However, a direct comparison is difficult, because different recordings of a word never match in the precise timing and the durations of each sub-word (e.g., phoneme, syllable) within the word are not the same. As a result, efforts to recognize words by matching them to templates will give inaccurate results if there is no temporal alignment. The Dynamic Time Warping (DTW) is one of the pioneer approaches to solve the alignment problem (Sakoe & Chiba, 1978; Rabiner & Juang, 1993). The DTW is an algorithm for measuring similarity between two sequences (e.g., signal), which may vary in time or speed. It provides a procedure to align in the test and reference patterns to give the average distance associated with the optimal warping path. The DTW compares an incoming speech signal with each template in the database and then takes the closest match that has a minimal pattern distance. Figure 8.6 shows the graph of an example on Dynamic Time Warping.

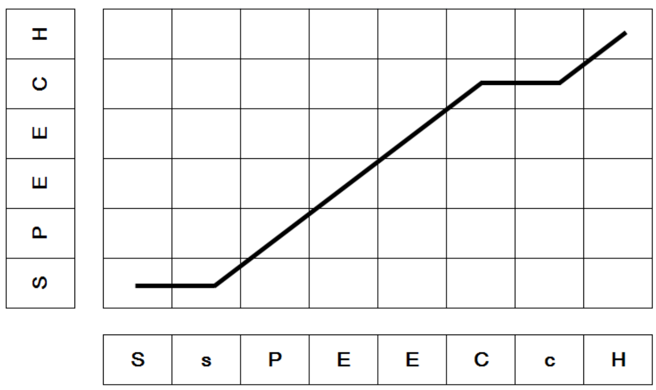


Figure 8.6. The warp path between "*SPEECH*" and "*SsPEECcH*".
The horizontal axis represents the time sequence of the input "*SsPEECcH*", and the vertical axis represents the time sequence of the template "*SPEECH*". The path shown results in the minimum distance between the input and template streams.

The final interface of the pronunciation test using the automatic speech recognition is illustrated with two screenshots of the application in Figure 8.7.

Figure 8.7. Screenshots of the interface of the pronunciation test using the automatic speech recognition. The two screenshots show the pronunciation test using the automatic speech recognition where the first one displays all available categories (e.g., digits, family members, fruits etc.) and the second one shows the automatic speech recognition interface.

8.2.2.3. Experimental results

This section provides the experimental results in recognizing the isolated words of Chinese digits from zero to nine. The templates of the ten digits were obtained by passing the feature extraction function using the utterances of a single native Chinese male speaker. The test samples were recorded from a male and a female Chinese native speaker. Every speaker repeated each digit five times, giving rise to ten samples for each digit. The recognition rate for each digit is presented in Table 8.2.

Table 8.2.

Recognition rate for ten digits

Digit	Recognition rate
ling1 ‘zero’	70%
yi1 ‘one’	80%
er2 ‘two’	70%
san1 ‘three’	70%
si4 ‘four’	50%
wu3 ‘five’	60%
liu4 ‘six’	70%
qi1 ‘seven’	60%
ba1 ‘eight’	70%
jiu3 ‘nine’	60%

8.2.3. A pronunciation dictionary

In order to facilitate learners to get the standard Chinese pronunciation, we created a pronunciation dictionary of syllables in the application. A male native Chinese speaker recorded all possible Chinese syllables with tones (i.e. in total 1257 recordings). In addition to the four full tones, syllables with neutral tones were also recorded. In the pronunciation dictionary, syllables are indexed by initials, finals and tones. When learner lookups a syllable, he or she chooses firstly the initial and final, then select its tone. To listen to the standard pronunciation, simply tap on the found syllable. Figure 8.8 shows a screenshot of the pronunciation dictionary.

Figure 8.8. A screenshot of the pronunciation dictionary.

8.2.4. A perception test on tone judgment

A perception test on tone judgment including the neutral tone was created in the application. In the perception test, the system plays randomly a syllable with one of the four tones or a neutral tone from the above database of the pronunciation of syllables; learners decide which tone is contained in the heard syllable by swiping the tone contour on the iPhone's screen. The use of finger swipes as responses interface instead of tapping on response buttons may enhance the association between the tone pronunciation and the tone contour (i.e., the shape that learners swipe on the screen[51]). The system displays after learner's response the presented syllable and the accuracy of the tone judgment as feedback. Figure 8.9 shows screenshots of the tone judgment test.

[51] As neutral tones have no stable contour, learner was asked to swipe a 'zero', i.e., '0' on the screen to indicate the neutral tone.

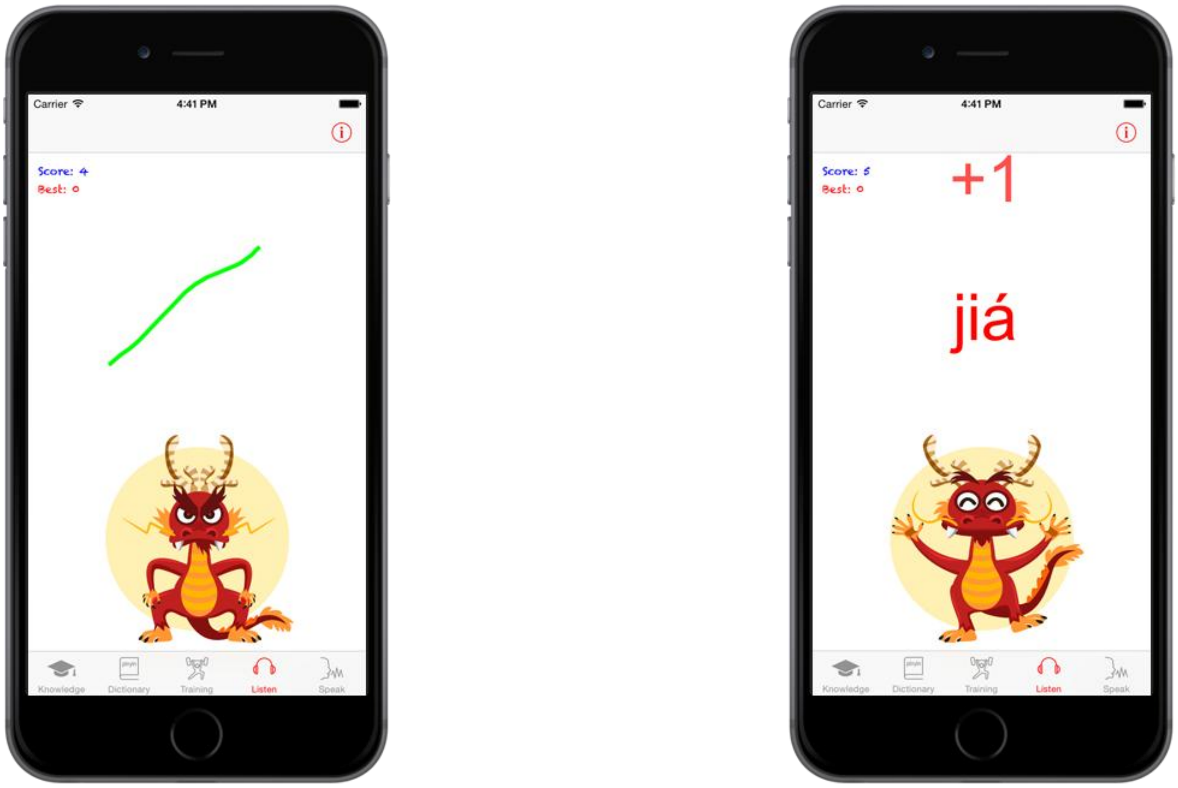


Figure 8.9. Screenshots of the tone judgment test. The left one shows the learner's response on an auditorily presented vowel with Tone 2. The green curve represents the learner's response, i.e., swipe on the screen. The response was good as displayed in the right screenshot.

8.2.4. Conclusion

We developed an iOS-based Chinese pronunciation learning application. The implemented pitch detection function displays in real-time the tone contour of learner's pronunciations of vowels. Results of the experiment showed a general classification rate on the four tones of 80.0%; in regards to speech recognition, the experiment data on the recognition of Chinese digits 0-9 from two independent speakers showed a recognition rate of 66.0%; moreover, a perception test on tone judgment including the neutral tone and a pronunciation dictionary were installed. Some improvements are planned for the future update, including more precise algorithms for pitch detection and automatic speech recognition.

Appendices

Appendix A

	Neutral tone	frequency	strike	Full tone	frequency	strike
1	巴掌 ba1zhang0	9	16	手掌 shou3zhang3	14	16
2	本事 ben3shi0	17	13	刑事 xin2shi4	15	14
3	部分 bu4fen0	312	14	十分 shi2fen1	335	6
4	畜生 chu4sheng0	5	15	孪生 ruan2sheng1	5	14
5	窗户 chuang1hu0	24	16	用户 yong4hu4	24	9
6	凑合 cou4he0	5	17	愈合 yu4he2	5	19
7	大意 da4yi0	5	16	大意 da4yi4	5	16
8	地道 di4dao0	7	18	地道 di4dao4	7	18
9	地方 di4fang0	532	10	北方 bei3fang1	65	9
10	东西 dong1xi0	437	11	东西 dong1xi1	437	11
11	动静 dong4jing0	12	20	安静 an1jing4	29	22
12	耳朵 er3duo0	53	12	花朵 hua1duo3	17	13
13	功夫 gong1fu0	28	9	车夫 che1fu1	10	8
14	姑娘 gu1niang0	183	18	大娘 da4niang2	36	13
15	故事 gu4shi0	125	17	军事 jun1shi4	138	14
16	寡妇 gua3fu0	16	20	产妇 chan3fu4	5	12
17	怪物 guai4wu0	12	16	事物 shi4wu4	224	16
18	闺女 gui1nü0	13	12	宫女 gong1nü3	5	12
19	含糊 han2hu0	9	22	浆糊 jiang4hu2	5	25
20	行当 hang2dang4	7	12	恰当 qia4dang4	27	15
21	合计 he2ji0	5	10	合计 he2ji4	5	10
22	和尚 he2shang0	25	16	时尚 shi2shang4	6	15
23	红火 hong2huo0	5	10	着火 zhao2huo3	7	15
24	厚道 hou4dao0	5	21	大道 da4dao4	11	15
25	记性 ji4xing0	5	13	天性 tian1xing4	7	12
26	家伙 jia1huo0	42	16	同伙 tong2huo3	5	12
27	将就 jiang1jiu0	5	11	成就 cheng2jiu4	77	18
28	客气 ke4qi0	16	13	天气 tian1qi4	80	8
29	口袋 kou3dai0	30	14	麻袋 ma2dai4	5	22
30	口气 kou3qi0	5	7	口气 kou3qi4	32	7
31	老子 lao3zi0	5	9	老子 lao3zi1	21	9
32	厉害 li4hai0	43	15	危害 wei1hai4	44	16
33	利落 li4luo0	5	19	掉落 diao4luo4	5	23
34	力气 li4qi0	28	6	氧气 yang3qi4	60	14
35	利索 li4suo0	5	17	铁索 tie3suo3	5	20
36	凉快 liang2kuai0	5	17	飞快 fei1kuai4	13	10
37	粮食 liang2shi0	113	22	零食 liang2shi2	5	22
38	龙头 long2tou0	5	10	龙头 long2tou2	7	10
39	马虎 ma3hu0	6	11	老虎 lao3hu3	22	14
40	买卖 mai3mai0	5	14	买卖 mai3mai4	39	14
41	眉毛 mei2mao0	17	13	羊毛 yang2mao2	13	10
42	门道 men2dai0	5	15	航道 hang2dao4	5	22
43	名字 ming2zi0	80	12	文字 wen2zi4	99	10
44	木匠 mu4jiang0	7	10	工匠 gong1jiang4	5	9
45	暖和 nuan3he0	7	21	求和 qiu2he2	5	15
46	盘算 pan2suan0	6	25	估算 gu1suan4	7	21
47	欺负 qi1fu0	9	28	正负 zheng4fu4	5	11
48	热闹 re4nao0	42	18	吵闹 cao3nao4	6	15
49	舒服 shu1fu0	31	20	制服 zhi4fu2	9	16
50	舒坦 shu1tan0	5	20	平坦 ping2tan3	13	13
51	岁数 sui4shu0	8	19	小数 xiao3shu4	5	16
52	下场 xia4chang0	5	9	下场 xia4chang3	11	9
53	下面 xia4mian0	5	12	下面 xia4mian4	146	12
54	相声 xiang4sheng0	6	16	歌声 ge1sheng1	25	21
55	消息 xiao1xi0	95	20	气息 qi4xi1	17	14
56	笑话 xiao4hua0	21	18	对话 dui4hua4	14	13
57	妖精 yao1jing0	5	21	酒精 jiu3jing1	23	24
58	月饼 yue4bing0	5	13	大饼 da4bing3	5	12
59	运气 yun4qi0	5	11	运气 yun4qi4	7	11
60	招呼 zhao1hu0	39	16	欢呼 hunan1hu1	12	14

Frequency: per million

Appendix B

	Prime	Meaning of the first morpheme	Related target		Frequency	Strike	Unrelated target		Frequency	Strike
1	东西 dong1xi0	east	南方	south	49	13	书法	handwriting	9	12
2	马虎 ma3hu0	horse	动物	animal	222	24	墙壁	wall	25	30
3	地道 di4dao0	ground	土地	land	193	9	价值	value	232	16
4	大意 da4yi0	big	微小	tiny	19	16	彩色	colorful	25	17
5	眉目 mei2mu0	eyebrow	眼睛	eyes	305	26	农业	agriculture	303	11
6	龙头 long2tou0	dragon	凤凰	phoenix	6	15	答复	response	17	21
7	红火 hong2huo0	red	绿色	green	56	17	膝盖	knee	9	26
8	反正 fan3zheng0	reversed	相反	in contrast	108	13	栽培	cultivate	31	22
9	合计 he2ji0	put together	结合	combine	314	15	岗位	post	37	14
10	枪手 qiang1shou0	gun	子弹	bullet	23	14	表格	table	8	18
11	门道 men4dao0	door	窗户	window	24	16	危险	danger	82	15
12	大方 da4fang0	big	尺寸	size	10	7	速度	speed	238	19
13	风光 feng1guang0	wind	下雨	rain	18	11	特征	characteristic	209	18
14	白话 bai4hua0	white	黑色	black	42	18	果实	fruit	36	16
15	方丈 fang1zhang0	square	圆圈	circle	9	21	考试	examination	38	14
16	点心 dian3xi0	point	面积	area	126	19	中央	center	150	9
17	厚道 hou4dao0	thick	扁平	flat	7	14	进化	evolution	45	11
18	风水 feng1shui0	wind	寒冷	frigid	25	19	知识	knowledge	321	15
19	下场 xia4chang0	down	上去	go up	88	8	赤道	equator	24	19
20	官司 guan1si0	officer	政府	government	293	17	会计	accountant	33	10

Frequency: per million

Appendix C

		verb-object	VN compound
1	帮工	help with farm work	helper
2	报价	to quote	a quote
3	报料	reveal a news	news
4	保险	to assure	insurance
5	备份	to backup	backups
6	备料	to stock	stock preparation
7	编号	to number	a number
8	编剧	w rite a play	playwright
9	编码	encoded	code
10	标价	mark the price	price
11	表情	express one's feelings	expression
12	比价	compare the price	price parity
13	拨款	allocate funds	Appropriation
14	补液	supply w ater	fluid supplementation
15	补益	to benefit	benefits
16	藏书	collect books	a collection of books
17	插话	interpose	episode
18	超人	exceed	superman
19	成人	grow up	adult
20	成文	become official dispatch	existing writings
21	吃水	absorb water	drinking water
22	创意	create	originality
23	创作	create	creation
24	传人	teach	successor
25	传闻	to rumour	a rumour
26	传言	pass on a message	hearsay
27	出口	to export	exports
28	存货	to stock	a stock
29	存粮	store up grain	grain in stock
30	出身	come from	class origin
31	出账	charge off	item of expenditure
32	存款	to deposit	a deposit
33	代表	to delegate	a delegate
34	答卷	answer questions in an exam	answer sheet
35	倒账	repudiate a debt	bad debts
36	得分	to score	a score
37	定案	decide on a verdict	verdict
38	订单	to order	an order
39	顶风	against the w ind	head w ind
40	定稿	finalize a manuscript	final version
41	定价	make a price	fixed price
42	定量	to quantify	norm
43	定时	definea time	definite time
44	定义	to define	definition
45	定员	designate members	fixed number of passengers
46	定址	select a venue	permanent venue
47	断层	fault	faultage
48	对话	have a dialogue	dialogue
49	罚金	to fine	a fine
50	罚款	impose a fine	a fine
51	发面	leaven dough	leavened dough
52	返程	to return	return
53	返利	to rebate	a rebate
54	发言	speak	speech
55	封口	to seal	seal
56	分界	have as the boundary	boundary
57	耕地	to plough	cultivated land
58	管家	manage	manager
59	鼓包	to lump	a lump
60	雇工	hire labour	hired labourer
61	耗材	consume	consumptive material
62	合力	join forces	resultant
63	护法	defend	custodian
64	回话	to reply	a reply
65	汇款	make a remittance	remittance
66	回礼	present a gift in return	a gift in return
67	回味	recollect pleasant flavour of ...	aftertaste
68	寄语	send word	message
69	剪报	cut new spaper	cuttings
70	兼差	work part-time	part-time jobs
71	讲话	speak	speech
72	剪影	to sketch	a sketch
73	兼职	hold two posts concurrently	concurrent post
74	剪纸	cut paper	paper-cut
75	结晶	crystallize	crystal
76	借款	to loan	a loan
77	结尾	to end	ending
78	进口	to import	importation
79	进账	register an income	income
80	纪实	to record actual events	record of actual events
81	纪事	to record	a record
82	捐款	contribute money	donation
83	决策	make a strategic decision	a strategic decision
84	开局	to open	opening
85	开头	make a start	beginning period
86	考绩	check achievement	merit of professional performance
87	理事	to manage	member of a council
88	留言	leave a message	a message
89	留影	take a photo as a memento	photograph
90	满堂	fill the hall	whole hall
91	满月	complet the first month of life	full moon
92	逆流	against the flow	backset
93	配方	make up a prescription	a prescription
94	平价	stabilize the price	fair price
95	评价	evaluate	evaluation
96	品味	to taste	taste
97	欠款	owe	debt
98	欠债	run into debt	amount due
99	欠账	to debit	debit
100	签证	to visa	a visa
101	签字	to sign	signature
102	起价	make a price	starting price
103	起头	to start	beginning
104	融资	to finance	a finance
105	说理	argue	argument
106	谈话	to talk	a talk
107	提包	carry a bag	handbag
108	题词	write an inscription	an inscription
109	贴息	pay interest	interest so deducted
110	替工	work as a substitute	temporary worker
111	投资	invest	investment
112	限价	limit the price	limited price
113	限量	limit the quantity of	limited quantity
114	限令	order sb. to do sth. in a certain time	restrictions
115	限期	set a time limit	deadline
116	显效	take effect	effect
117	选项	delect an option	option
118	续约	renew a contract	renewal of contract
119	押款	borrow money on security	a loan on security
120	演义	explain some reason or fact	historical novel
121	议价	negotiate a price	negotiated price
122	引例	to cite	citation
123	用语	choose words	choice of words
124	约期	fix a date	date of appointment
125	渔利	reap unfair gains	easy gains
126	造型	to model	modelling
127	摘要	make a summary	summary
128	掌舵	steer a boat	the man who steers a boat
129	征文	solicit articles	essay writing
130	转机	transfer	a favourable turn
131	转年	pass to the coming year	the coming year
132	铸币	to coin	coined money
133	驻军	to garrison	garrison
134	作文	write a composition	composition
135	作业	to operate	a job

Appendix D

	Sequence		VN	VO
1	炒饭	chao3fan4	fried rice	fry the rice
2	烤肉	kao3rou4	barbecue	grill the meat
3	驻军	zhu4jun1	garrison (n.)	garrison (v.)
4	转机	zhuan3ji1	a favourable turn	transfer
5	签字	qian1zi4	signature	sign
6	留言	liu2yan2	message	leave a message
7	捐款	juan1kuan3	donation	donate
8	耕地	geng1di4	cultivated land	cultivate
9	封口	feng1kou3	seal	shut up
10	罚款	fa2kuan3	penalty	impose a fine
11	得分	de2fen1	score	get a point
12	答卷	da2juan4	answer sheet	answer questions in an exam
13	存款	cun2kuan3	deposit	deposit money
14	拨款	bo1kuan3	funds	allocate funds
15	标价	biao1jia4	marked price	mark the price
16	编号	bian1hao4	number (n.)	number (v.)
17	备份	bei4fen4	backups	make a backup
18	报价	bao4jia4	quoted price	quote
19	开局	kai1ju2	opening	start a round
20	定时	ding4shi2	fixed time	make a date

Bibliography

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh University Press, Edinburgh.

Adisasmito-Smith, N., & Cohn, A. C. (1996). Phonetic correlates of primary and secondary stress in Indonesian: A preliminary study. *Working papers of the Cornell phonetics laboratory*, *11*, 1-16.

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology, 16*(1-2), 285-311.

Anusuya, M. A., & Katti, S. K. (2009). Speech recognition by machine, a review. *(IJCSIS) International Journal of Computer Science and Information Security, Vol.6,No.3.*

Baayen, H., & Schreuder, R. (1999). War and peace: Morphemes and full forms in a noninteractive activation parallel dual-route model. *Brain and language, 68*(1), 27-32.

Baayen, R. H., Lieber, R., & Schreuder, R. (1997). The morphological complexity of simplex nouns. *Linguistics, 35*, 861-877.

Bai, C., Bornkessel-Schlesewsky, I., Wang, L., Hung, Y. C., Schlesewsky, M., & Burkhardt, P. (2008). Semantic composition engenders an N400: Evidence from Chinese compounds. *Neuroreport*, *19*(6), 695-699.

Beckman, M. E. (1986). *Stress and non-stress accent* (Vol. 7). Walter de Gruyter.

Bell, M. J., & Plag, I. (2012). Informativeness is a determinant of compound stress in English. *Journal of Linguistics, 48*(03), 485-520.

Bell, M., & Plag, I. (2013). Compound stress, informativity and analogy. *Word Structure, 6*(2), 129-155.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *The Journal of Neuroscience*, *31*(8), 2801-2814.

Bemis, D. K., and Pylkkänen, L. (2011). Simple composition: a magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *J. Neurosci. 31*, 2801-2814.

Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology, 30*(2), 161-169.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and clinical Neurophysiology, 60*(4), 343-355.

Berger, H. (1929). Über das elektrenkephalogramm des menschen. *European Archives of Psychiatry and Clinical Neuroscience, 87*(1), 527-570.

Berkowitz, R. N. (2009). Recognizing English Compound Words: The Role of Morphological Family Size (Doctoral dissertation, Wesleyan University).

Bertinetto, P. M. (1989). Reflections on the dichotomy ‘stress’ vs.‘syllable-timing’. *Revue de phonétique appliquée, 91*(93), 99-130.

Birmbaumer , N. & Schmidt, R.F.(1990). *Biologische Psychologie*. Berlin: Springer.

Bisetto, A., and Scalise, S. (2005). Classification of Compounds. *Lingue e Linguaggio, 4*(2), 319-332.

Bloomfield, A. L., & Polland, W. S. (1933). *Gastric anacidity; its relation to disease*. Macmillan.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. (pp. 945-984). San Diego, CA US: Academic Press.

Böcker, K. B., Bastiaansen, M., Vroomen, J., Brunia, C. H., & Gelder, B. (1999). An ERP correlate of metrical stress in spoken word recognition. *Psychophysiology, 36*(6), 706-720.

Bondarko, L. V., Verbitskaya, L. P., & Tscherbakova, L. P. (1973). Ob opredelenii mesta udareniya v slove. *Izvestiya AN SSSR, Seriya OLYA, 2*.

Bozic, M., Marslen-Wilson, W. D., Stamatakis, E. A., Davis, M. H., & Tyler, L. K. (2007). Differentiating morphology, form, and meaning: Neural correlates of morphological complexity. *Journal of Cognitive Neuroscience, 19*(9), 1464-1475.

Bronk, M., Zwitserlood, P., & Bölte, J. (2013). Manipulations of word frequency reveal differences in the processing of morphologically complex and simple words in German. *Frontiers in psychology, 4*.

Brooks, T. L., & de Garcia, D. C. (2015). Evidence for morphological composition in compound words using MEG. *Frontiers in human neuroscience, 9*.

Brown, C., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*(1), 34-44.

Butterworth, B. (1983). Lexical representation. In B. Butterworth (Ed.), *Development, writing and other language processes*, Vol. 2, London: Academic Press.

Bybee, J. L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes, 10*,425-455.

Cao, J. F. (1986). Putonghua qingsheng yinjie fenxi [Acoustic analysis of the neutral tone syllables in Mandarin]. *Yingyong Shengxue [Applied Acoustics], 1986*(4):1-6. [Chinese Academy of Social Science].

Cao, J. f. (1992). On Neutral-Tone Syllables in Mandarin Chinese. *Canadian Acoustics, 20*(3).

Caramazza, A., Laudanna, A., & Romani, C. (1988). Lexical access and inflectional morphology. *Cognition, 28*(3), 297-332.

Ceccagno, A., & Basciano, B. (2007). Compound headedness in Chinese: an analysis of neologisms. *Morphology, 17*(2), 207-231.

Ceccagno, A., & Scalise, S. (2006). Classification, structure and headedness of Chinese compounds. *Lingue e linguaggio,* (2), 233-260.

Chao, Y. R. (1948). Mandarin Primer: an Intensive Course in Spoken Chinese. Cambridge, Mass: Harvard University Press.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley and Los Angeles: University of California Press.

Chao, Yuen-Ren (1956). *Tone, intonation, singsong, chanting, recitative tonal composition and atonal composition in Chinese*. In Moris Halle (ed.), For Roman Jakobson 52-59. The Hague: Mouton.

Chen, X.Y. (2004). On the Necessity and Consistency Principles of Qingsheng Words:Metric Research on Qingsheng Words in Modern Chinese Dictionary. Yuyan Wenzi Yingyong 1, 112-119.

Cheng, X.Q. (1992). Researches on pre-Qin disyllabic words. In X. Cheng, ed. *Researches on pre-Qin Chinese*. Jinan: Shandong Jiaoyu Chubanshe. 44-112.

Chialant, D., & Caramazza, A. (1995). Where is morphology and how is it processed? The case of written word recognition. *Morphological aspects of language processing*, 55-76.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English*. New York: Harper & Row.

Coles, M.G.H. & Rugg, M.D. (1995). Event-related brain potentials: An introduction. In M.G.H. Rugg M.D. and Coles (Ed.). *Electrophysiology of mind: Event-related brain potentials and cognition* (p. 1-23). New York: Oxford University Press.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407-428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning & Verbal Behavior, 8(*2), 240-247.

Connine, C. M., Clifton Jr, C., & Cutler, A. (1987). Effects of lexical stress on phonetic categorization. *Phonetica, 44*(3), 133-146.

Cooper, R., Winter, A. L., Crow, H. J., & Walter, W. G. (1965). Comparison of subcortical, cortical and scalp activity using chronically indwelling electrodes in man. *Electroencephalography and clinical neurophysiology, 18*(3), 217-228.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question–answer contexts. *The Journal of the Acoustical Society of America, 77*(6), 2142-2156.

Crystal, D. (1969). *What is linguistics?* Hodder Arnold.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected ‐ speech signals: Current results. *The journal of the acoustical society of America, 83*(4), 1553-1573.

Cunillera, T., Gomila, A., & Rodrǵuez-Fornells, A. (2008). Beneficial effects of word final stress in segmenting a new language: evidence from ERPs. *BMC neuroscience, 9*(1), 23.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*(3), 201-220.

Cutler, A., & Darwin, C. J. (1981). Phoneme-monitoring reaction time and preceding prosody: Effects of stop closure duration and of fundamental frequency. *Perception & Psychophysics, 29*(3), 217-224.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human perception and performance, 14*(1), 113.

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *The Journal of the Acoustical Society of America, 105*(3), 1877-1888.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and speech, 40*(2), 141-201.

Darabkh, K. A., Khalifeh, A. F., Bathech, B. A., & Sabah, S. W. (2013). Efficient DTW-Based Speech Recognition System for Isolated Words of Arabic Language. In

Proceedings of International Conference on Electrical and Computer Systems Engineering (ICECSE 2013) (pp. 699-702).

Darabkh, K. A., Khalifeh, A. F., Bathech, B. A., & Sabah, S. W. (2013, May). Efficient DTW-Based Speech Recognition System for Isolated Words of Arabic Language. *In Proceedings of International Conference on Electrical and Computer Systems Engineering (ICECSE 2013)* (pp. 699-702).

Davis, M. H., Meunier, F., & Marslen-Wilson, W. D. (2004). Neural responses to morphological, syntactic, and semantic properties of single words: an fMRI study. *Brain and language, 89*(3), 439-449.

Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *28*(4), 357-366.

Devlin, J. T., Jamison, H. L., Matthews, P. M., & Gonnerman, L. M. (2004). Morphology and the internal structure of words. *Proceedings of the National Academy of Sciences of the United States of America, 101*(41), 14984-14988.

Dirven, R. & Verspoor, M. (1998). *Cognitive Exploration of Language and Linguistics*. Amsterdam: Benjamins. Chapter 1: The cognitive basis of language: language and thought. pp. 1-24

Donchin, E., & Cohen, L. (1967). Averaged evoked potentials and intramodality selective attention. *Electroencephalography and Clinical Neurophysiology, 22*(6), 537-546.

Duanmu, S. (1990). Tonal patterns in Chinese regulated verses: phonological rules or phonological constraints. *MIT Working Papers in Linguistics, 12*, 177-189.

Duanmu, S. (2000). Stress in Chinese. In Xu, D.B. (Eds.), *Chinese Phonology in Generative Grammar*. New York: Academic Press, pp. 117-138.

Duanmu, S. (2002). *The phonology of standard Chinese*. Oxford University Press.

Duanmu, S. (2004). Left-headed feet and phrasal stress in Chinese. *Cahiers de linguistique-Asie orientale, 33*(1), 65-103.

Duanmu, S. (2007). *The phonology of standard Chinese*, 2nd Edition. Oxford University Press.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). Do transposed-letter similarity effects occur at a morpheme level? Evidence for morpho-orthographic decomposition. *Cognition, 105*(3), 691-703.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing" deafness" in French?. *Journal of Memory and Language, 36*, 406-421.

El Yagoubi, R., Chiarelli, V., Mondini, S., Perrone, G., Danieli, M., & Semenza, C. (2008). Neural correlates of Italian nominal compounds and potential impact of headedness effect: an ERP study. *Cognitive Neuropsychology, 25*(4), 559-581.

Eling, P. A., Marshall, J. C., & Van Galen, G. P. (1980). Perceptual centres for Dutch digits. *Acta psychologica, 46*(2), 95-102.

Fabb, N. (1998). Compounding. In A. Spencer, & A. M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 66-83). Oxford: Blackwell.

Farnetani, E., Torsello, C. T., & Cosi, P. (1988). English compound versus non-compound noun phrases in discourse: An acoustic and perceptual study. *Language and speech, 31*(2), 157-180.

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and language, 115*(3), 149-161.

Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge, Massachusetts: MIT Press.

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive processes, 22*(7), 953-1000.

Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. *In Proceedings of the SPEECH PROSODY 2002 Conference*, Aix-en-Provence, France, April 2002.

Fourier, J. (1822). *Theorie analytique de la chaleur, par M. Fourier.* Chez Firmin Didot, père et fils.

Friederici, A. D. (1995). The time course of syntactic activation during language processing: A model based on neuropsychological and neurophysiological data. *Brain and language, 50*(3), 259-281.

Friederici, A. D. (2001). Event-related brain potentials and aphasia. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (2nd ed., Vol. 3, pp.353–373). Amsterdam: Elsevier Science.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in cognitive sciences, 6*(2), 78-84.

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: a dynamic dual pathway model. *Brain and language, 89*(2), 267-276.

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(5), 1219.

Friederici, A. D., Wang, Y., Herrmann, C. S., Maess, B., and Oertel, U. (2000). Localization of early syntactic processes in frontal and temporal cortical areas: a magnetoencephalographic study. *Hum. Brain Mapp. 11*, 1-11.

Friedman, D., Cycowicz, Y. M., & Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. *Neuroscience & Biobehavioral Reviews, 25*(4), 355-373.

Friedrich, C. K., Alter, K., & Kotz, S. A. (2001). An electrophysiological response to different pitch contours in words. *NeuroReport, 12*(15), 3189-3191.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Alter, K. (2004a). Pitch modulates lexical identification in spoken word recognition: ERP and behavioral evidence. *Journal of Cognitive Brain Research, 20*, 300-308.

Friedrich, C. K., Kotz, S. A., Friederici, A. D., & Gunter, T. C. (2004b). ERPs reflect lexical identification in word fragment priming. *Journal of Cognitive Neuroscience, 16*(4), 541-552.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. The *Journal of the Acoustical Society of America, 27*(4), 765-768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and speech, 1*(2), 126-152.

Gagné, C. L. (2002). Lexical and relational influences on the processing of novel compounds. *Brain and Language, 81*(1), 723-735.

Gagné, C. L., & Spalding, T. L. (2004). Effect of relation availability on the interpretation and access of familiar noun-noun compounds. *Brain and Language, 90*(1), 478-486.
Gagné, L.C. & Spalding, T.L. (2009). Constituent integration during the processing of compound words: Does it involve the use of relational structure. *Journal of Memory and Language, 60*, 20-35.
Garellek, M., & White, J. (2015). Phonetics of Tongan stress. *Journal of the International Phonetic Association, 45*(01), 13-34.
Gordon, M. (2004). A Phonological and Phonetic Study of Word-Level Stress in Chickasaw 1. *International Journal of American Linguistics, 70*(1), 1-32.
Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*(2), 95-112.
Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267-283.
Grosjean, F. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics*, 21, 501-529.
Grosjean, F., & Gee, J. P. (1987). Prosodic structure and spoken word recognition. *Cognition, 25*(1), 135-155.
Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge University Press.
Gussenhoven, C., Repp, B. H., Rietveld, A., Rump, H. H., & Terken, J. (1997). The perceptual prominence of fundamental frequency peaks. *Journal of the Acoustical Society of America, 102*(5), 3009-3022.
Haggard, M., Ambler, S., & Callow, M. (1970). Pitch as a voicing cue. *The Journal of the Acoustical Society of America, 47*(2B), 613-617.
Haggard, M.; Ambler, S.; Callow, M.: Pitch as a voicing cue. *J. acoust. Soc. Am.* 47: 613-617 (1970).
Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and cognitive processes, 8*(4), 439-483.
Halle, M. & Vergnaud, J. (1978). *Metrical structures in phonology.* Unpublished manuscript, MIT.
Hayes, B. (1995). *Metrical stress theory: Principles and case studies*. University of Chicago Press.
Higgins, A. L., & Wohlford, R. E. (1985). Keyword recognition using template concatenation. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'85.* (Vol. 10, pp. 1233-1236). IEEE.
Hillyard, S. A., & Picton, T. W. (1987). Electrophysiology of cognition. In F. Plum (Ed.), Handbook of physiology: Section 1. The nervous system. Volume V. Higher functions of the brain, part 2 (pp. 519-584). Bethesda, MD: American Physiological Society.
Holcomb, P. J. (1988). Automatic and attentional processing: An event-related brain potential analysis of semantic priming. *Brain and language, 35*(1), 66-85.
Honbolygó, F., & Csépe, V. (2013). Saliency or template? ERP evidence for long-term representation of word stress. *International Journal of Psychophysiology, 87*(2), 165-172.
Huang, C. T. J. (1984). Phrase structure, lexical integrity, and Chinese compounds. *Journal*

of the Chinese Language Teachers Association, *19*(2), 53-78.

Hyman, L. (1977). On the nature of linguistic stress. Studies in stress and accent, 4.

Hyönä, J., & Pollatsek, A. (1998). Reading Finnish compound words: Eye fixations are affected by component morphemes. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(6), 1612.

Inhoff, A. W., Radach, R., & Heller, D. (2000). Complex compounds in German: Interword spaces facilitate segmentation but hinder assignment of meaning. *Journal of Memory and Language*, *42*(1), 23-50.

Institute of Language Teaching and Research. (1986). *A frequency dictionary of Modern Chinese*. Beijing: Beijing Language Institute Press.

Isel F. (2010) The role of linguistic prosody in the processing of morphologically complex words in phrase and sentence contexts: Evidence from ERP studies. HDR memory.

Isel, F., & Friederici, A. D. (2006). Does prosody guide lexical access? ERP evidence from spoken German. *The 12th annual Meeting of the organization for Human Brain Mapping,* Florence, Italy.

Isel, F., Alter, K., & Friederici, A. D. (2005). Influence of prosodic information on the processing of split particles: ERP evidence from spoken German. *Journal of Cognitive Neuroscience*, *17*(1), 154-167.

Isel, F., Gunter, T. C., & Friederici, A. D. (2003). Prosody-assisted head-driven access to spoken German compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 277.

Isel, F., Hahne, A., Maess, B., & Friederici, A.D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology*, 74, 337-346.

Jasper, H. (1958). The ten twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371-375.

Jassem, W., Morton, J., & Steffen-Batog, M. (1968). The perception of stress in synthetic speech-like stimuli by Polish listeners. *Speech analysis and synthesis*, 1, 289-308.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*(4), 406-430.

Ji, H., Gagné, C. L., & Spalding, T. L. (2011). Benefits and costs of lexical decomposition and semantic integration during the processing of transparent and opaque English compounds. *Journal of Memory and Language*, *65*(4), 406-430.

Jia, X., Wang, S., Zhang, B., & Zhang, J. X. (2013). Electrophysiological evidence for relation information activation in Chinese compound word comprehension. *Neuropsychologia*, *51*(7), 1296-1301.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(6), 1312.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and cognitive processes*, *15*(2), 159-201.

Keppel, G. (1991). *Design and analysis: A researcher's handbook*. Prentice-Hall, Inc.

Kesarkar, M. P. (2003, November). Feature extraction for speech recognition. In *M. Tech. Credit Seminar Report, Electronic Systems Group, EE. Dept, IIT Bombay* (pp. 1-11).

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America, 59*(5), 1208-1221.

Kochanski, G., Shih, C., & Jing, H. (2003). Hierarchical structure and word strength prediction of Mandarin prosody. *International Journal of Speech Technology, 6*(1), 33-43.

Koester, D., Gunter, T. C., Wagner, S., & Friederici, A. D. (2004). Morphosyntax, prosody, and linking elements: The auditory processing of German nominal compounds. *Journal of Cognitive Neuroscience, 16*(9), 1647-1668.

Koester, D., Holle, H., & Gunter, T. C. (2009). Electrophysiological evidence for incremental lexical-semantic integration in auditory compound comprehension. *Neuropsychologia, 47*(8), 1854-1864.

Koester, M., & Cutler, A. (1997). Segmental and suprasegmental contributions to spoken-word recognition in dutch. In *EUROSPEECH*.

Kutas, M., & Federmeier, K. D. (2009). N400. *Scholarpedia, 4*(10), 7790.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual review of psychology*, 62, 621.

Kutas, M., & Hillyard, S. (1989). An electrophysiological probe of incidental semantic association. *Cognitive Neuroscience, Journal of, 1*(1), 38-49.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. Biological psychology, 11(2), 99-116.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified. *Handbook of psycholinguistics*, 83-143.

Ladefoged, P. (1975). *A Course in Phonetics*. New York: Harcourt, Brace, Jovanovich.

Ladefoged, P. (2001). Vowels and consonants. *Phonetica, 58*(3), 211-212.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, Mass.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of phonetics*.

Li, A., Gao, J., Jia, Y., & Wang, Y. (2014). Pitch and duration as cues in perception of neutral tone under different Contexts in Standard Chinese. In Asia-Pacific Signal and Information Processing Association, *2014 Annual Summit and Conference (APSIPA)* (pp. 1-6). IEEE.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar.* Univ of California Press.

Li, C., & Thompson, S. (1981). A functional reference grammar of Mandarin Chinese.

Li, W. (1981). A Study of the neutral tone and the normal stress in Mandarin. *Chinese Language and Writing* 1, 35-40.

Li, X. (2008). *Lexicon of Common Words in Contemporary Chinese*. The Commercial Press, Beijing, China.

Libben, G. (1994). How is morphological decomposition achieved?. *Language and Cognitive Processes, 9*(3), 369-391.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language, 61*(1), 30-44.

Libben, G., Derwing, B. L., & de Almeida, R. G. (1999). Ambiguous novel compounds and models of morphological parsing. *Brain and Language, 68*(1), 378-386.

Libben, G., Derwing, B., & de Almeida, R. G. (1999) Ambiguous Novel Compounds and Models of Morphological Parsing. *Brain and Language*, 68, 378-386.

Libben, G., Gibson, M., Yoon, Y. B., & Sandra, D. (2003). Compound fracture: The role of semantic transparency and morphological headedness. *Brain and Language, 84*(1), 50-64.

Liberman, M. Y. (1975). *The intonational system of English* (Doctoral dissertation, Massachusetts Institute of Technology).

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic inquiry,* 249-336.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America, 32*(4), 451-454.

Lima, S. D., & Pollatsek, A. (1983). Lexical access via an orthographiccode? The basic orthographic syllable (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior,* 22, 310-322.

Lin, Hua (2006.) Mandarin Neutral Tone as a Phonologically Low Tone. *Journal of Chinese Language and Computing, 16*(2): 121-134.

Lin, Maocan & Jingzhu Yan. (1980). Beijinghua qingsheng de shengxue xingzhi [The acoustic nature of the Mandarin Neutral Tone]. Fanyan [Dialects], 1980(3):166-178.

Lin, Maocan & Jingzhu Yan. (1990). Putonghua qingsheng yu qingzhong yin [The neutral tone and stress in Mandarin.] *Yuyan Jiaoxue yu Yanjiu [Language Teaching and Linguistic Studies], 1990*(2):88-104.

Lin, T. H. (2001). *Light Verb Syntax and the Theory of Phase Structure*. Irvine: University of California dissertation.

Lin, Tao (1985). Tantao Beijinghua qingyin xingzhi de chubu shiyan (On neutral tone in Beijing Mandarin). In Shuangbao Hu (ed.), *Beijing yuyin shiyanlu (Working papers in experimental phonetics)* 1-26. Beijing: Peking University Press.

Lindfield, K. C., Wingfield, A., & Goodglass, H. (1999). The role of prosody in the mental lexicon. *Brain and Language, 68*(1), 312-317.

Liu, F., & Xu, Y. (2005). Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation. *Phonetica*, 62, 70-87.

Lu S. & Ding. S. (2008). *The Contemporary Chinese Dictionary,* 5th Edition (Xiandai Hanyu Cidian). Beijing: Commercial Press.

MacGregor, L. J., & Shtyrov, Y. (2013). Multiple routes for compound word processing in the brain: Evidence from EEG. *Brain and language, 126*(2), 217-229.

Magne, C., Gordon, R. L., & Midha, S. (2010). Influence of metrical expectancy on reading words: An ERP study. *Proceedings of Speech Prosody,* 2010, Chicago.

Marie, C., Magne, C., & Besson, M. (2011). Musicians and the metric structure of words. *Journal of Cognitive Neuroscience, 23*(2), 294-305.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102.

Marslen-Wilson, W., & Welsh, A. (1978). Processing inter-actions and lexical access during word recognition in continuous speech. *Cognitive Psychology,*10,29-63.

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological review, 101*(1), 3.

Mathews, P. H. (1974). *Morphology: an introduction to the theory of word-structure.* Cambridge UniversityPress, London.

Matthews, P.H. (1974). *Morphology*. Cambridge: Cambridge University Press.

McCauley, S. M., Hestvik, A., & Vogel, I. (2013). Perception and bias in the processing of compound versus phrasal stress: Evidence from event-related brain potentials. *Language and speech, 56*(1), 23-44.

McKinnon, R., Allen, M., & Osterhout, L. (2003). Morphological decomposition involving non-productive morphemes: ERP evidence. *Neuroreport, 14*(6), 883-886.

Moore. (1993). Phonetic observations on stress and tones in Mandarin Chinese. *Working papers of the Cornell Phonetics laboratory*. vol. 8, pp. 87-117.

Morrill, T. (2012). Acoustic correlates of stress in English adjective-noun compounds. *Language and speech, 55*(2), 167-201.

Näätänen, R. (1992). *Attention and brain function*. Psychology Press.

Näätänen, R., & Gaillard, A. W. K. (1983). 5 The orienting reflex and the N2 deflection of the event-related potential (ERP). *Advances in psychology*, 10, 119-141.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta psychologica, 42*(4), 313-329.

Neely, J. H. (1991). Semantic priming effects in visual word recognition: A selective review of current findings and theories. *Basic processes in reading: Visual word recognition,* 11, 264-336.

Neill, W. T. (1977). Inhibitory and facilitatory processes in selective attention. *Journal of Experimental Psychology: Human Perception and Performance, 3(*3), 444.

Nicoladis, E., & Krott, A. (2007). Word Family Size and French‐Speaking Children's Segmentation of Existing Compounds. *Language Learning, 57*(2), 201-228.

Nowagk, R. (1998). Conan: A Barbarian Tool for Constrained Randomization (Version 1.9). Leipzig: Max-Planck-Institute of Cognitive Neuroscience Leipzig.

Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Ortega-Llebaria, M. (2006). Phonetic cues to stress and accent in Spanish. *In Selected proceedings of the 2nd conference on Laboratory Approaches to Spanish Phonetics and Phonology* (pp. 104-118).

Packard, J. L. (2000). *The morphology of Chinese: A linguistic and cognitive approach*. Cambridge University Press.

Peng, D., & Wang, C. (1997). Basic processing unit of Chinese character recognition: Evidence from stroke number effect and radical number effect. *Acta Psychologica Sinica*.

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. *Laboratory phonology*, 7, 203-240.

Picton, T. W., Lins, O. G., & Scherg, M. (1995). The recording and analysis of event-related potentials. *Handbook of neuropsychology*, 10, 3-3.

Pike, K.L. (1945). *The Intonation of Amercian English*. Ann Arbor: University of Michigan Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28*(1-2), 73-193.

Plag, I., Kunter, G., Lappe, S., & Braun, M. (2008). The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language, 84*(4), 760-794.

Pollatsek, A., Hyönä, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance, 26*(2), 820.

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: the mismatch negativity as a tool for studying higher cognitive processes. *Progress in neurobiology, 79(*1), 49-71.

Rabiner, L. R. (1977). On the use of autocorrelation analysis for pitch detection. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, *25*(1), 24-33.

Rabiner, L., & Juang, B. H. (1993). *Fundamentals of speech recognition*. Upper Saddle River, New Jersey: Prentice Hall, USA.

Radeau, M., Besson, M., Fonteneau, E., & Castro, S. L. (1998). Semantic, repetition and rime priming between spoken words: behavioral and electrophysiological evidence. *Biological psychology, 48*(2), 183-204.

Rietveld, A. C. M. (1980). Word boundaries in the French language. *Language and Speech, 23*(3), 289-296.

Roach, P. (1982). On the distinction between 'stress-timed'and 'syllable-timed'languages. *Linguistic controversies,* 73-79.

Roach, P. (2010). *English Phonetics and Phonology Fourth Edition: A Practical Course*. Ernst Klett Sprachen.

Rosenfeld, L., Isel, F., & Fradin, B. (2014). Representation and processing of French verb-noun compounds. *In 9th International Conference on the Mental Lexicon* (pp. 21-22).

Rugg, M.D. (1999). Functional neuroimaging in cognitive neuroscience. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (p. 15-30). New York: Oxford University Press.

Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *Acoustics, Speech and Signal Processing, IEEE Transactions on, 26*(1), 43-49.

Sandra, D. (1990). On the representation and processing of compound words: Automatic access to constituent morphemes does not occur. *The Quarterly Journal of Experimental Psychology, 42*(3), 529-567.

Schmidt-Kassow, M., & Kotz, S. A. (2009). Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *Journal of Cognitive Neuroscience, 21*(9), 1693-1708.

Schreuder, R., & Baayen, R. (1997). How complex simplex words can be. *Journal of Memory and Language, 37*(1), 118-139.

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 131-154). Hillsdale, NJ: Erlbaum

Selkirk, E., & Shen, T. (1990). Prosodic domains in Shanghai Chinese. *The* *phonology-syntax connection*, 313, 37.

Sharbrough, F., Chatrian, G., Lesser, R., Lüders, H., Nuwer, M., & Picton, T. (1991). American electroencephalographic society guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, 8, 200-202.

Shen, J. (1985). Beijinghua shengdiaode yinyu he yudiao. [Pitch range of tone and intonation in Beijing dialect]. In: *Beijing Yuyin Shiyan Lu [Working Papers in Experimental Phonetics]*. Beijing: Beijing Daxue Chubanshe. 73-130.

Shen, W., Vaissière, J., & Isel, F. (2013). Acoustic Correlates of Contrastive Stress in Compound Words versus Verbal Phrase in Mandarin Chinese. *Computational linguistics and Chinese language processing, 18*(3), 45-58.

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech, 33*(1), 47-68.

Sluijter, A., & Van Heuven, V. J. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. In Spoken Language, 1996. ICSLP 96. *Proceedings., Fourth International Conference on (Vol. 2, pp. 630-633)*. IEEE.

Small, L. H., Simon, S. D., & Goldberg, J. S. (1988). Lexical stress and lexical access: Homographs versus nonhomographs. *Perception & Psychophysics, 44*(3), 272-280.

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review, 81*(3), 214-241.

Sondhi, M. M. (1968). New methods of pitch extraction. *Audio and Electroacoustics, IEEE Transactions on*, *16*(2), 262-266.

Song H.Q. (2009). *Dictionary of Mandarin Word Stress Pattern.* Shanghai Music Publishing House.

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*(3), 412-432.

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and clinical neurophysiology, 38*(4), 387-401.

Steinhauer, K., Alter, K., & Friederici, A. D. (1999). Brain potentials indicate immediate use of prosodic cues in natural speech processing. *Nature neuroscience, 2*(2), 191-196.

Stevens, S. S., Volkmann, J., & Newman, E. B. (1937). A scale for the measurement of the psychological magnitude pitch. *The Journal of the Acoustical Society of America, 8*(3), 185-190.

Stuckless, R. (1994). Developments in real-time speech-to-text communication for people with impaired hearing. *Communication access for people with hearing loss*, 197-226.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187-1188.

Taft, L. A. (1984). *Prosodic constraints and lexical parsing strategies*. University of Massachusetts. Graduate Linguistics Student Association.

Taft, M. (2003). Morphological representation as a correlation between form and meaning. In *Reading complex words* (pp. 113-137). Springer US.

Taft, M. (2004). Morphological decomposition and the reverse base frequency effect. *Quarterly Journal of Experimental Psychology Section A, 57*(4), 745-765.

Taft, M., & Forster, K. I. (1975). Lexical storage and retrieval of prefixed words. *Journal of*

Verbal Learning and Verbal Behavior, 14, 638-647.

Taft, M., & Forster, K. I. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of Verbal Learning & Verbal Behavior, 15*(6), 607-620.

Tagliapietra, L., & Tabossi, P. (2005). Lexical stress effects in Italian spoken word recognition. In *The XXVII Annual Conference of the Cognitive Science Society* (pp. 2140-2144).

Taylor, I., & Taylor, M. M. (1995). *Writing and literacy in Chinese, Korean, and Japanese.* Philadelphia: John Benjamins.

Tipper, S. P., & Cranston, M. (1985). Selective attention and priming: Inhibitory and facilitatory effects of ignored primes. *The Quarterly Journal of Experimental Psychology, 37*(4), 591-611.

Tran, D. T. (2000). *Fuzzy Approaches to Speech and Speaker Recognition* (Doctoral dissertation, university of Canberra).

Turk, A. E., & Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *The Journal of the Acoustical Society of America, 99*(6), 3782-3790.

Vaissière, J. (2004). *Perception of intonation. Handbook of Speech Perception*. D. B. Pisoni and R. E. Remez. Oxford, Blackwell.

Vaissière, J., (1982). Utilisation des paramètres suprasegmentaux en reconnaissance automatique de la parole comme aide à la segmentation. In A. D. Cristo, J.-P. Haton, M. Rossi, & J. Vaissière (Eds.), *Prosodie et reconnaissance automatique de la parole* (123-139). Groupement des Acousticiens de Langue Française.

van Heuven, V. J. (1985). Perception of stress pattern and word recognition. *The Journal of the Acoustical Society of America, 78*(S1), S21-S21.

Vannest, J., Polk, T. A., & Lewis, R. L. (2005). Dual-route processing of complex words: new fMRI evidence from derivational suffixation. *Cognitive, Affective, & Behavioral Neuroscience, 5*(1), 67-76.

Vartiainen, J., Aggujaro, S., Lehtonen, M., Hultén, A., Laine, M., & Salmelin, R. (2009). Neural dynamics of reading morphologically complex words. *Neuroimage, 47*(4), 2064-2072.

Vergara-Martínez, M., Duñabeitia, J. A., Laka, I., & Carreiras, M. (2009). ERP correlates of inhibitory and facilitative effects of constituent frequency in compound word reading. *Brain Research*, 1257, 53-64.

Vogel, I., & Raimy, E. (2002). The acquisition of compound vs. phrasal stress: the role of prosodic constituents. *Journal of Child Language, 29*(02), 225-250.

Vroomen, J., & De Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human perception and performance, 21*(1), 98.

Wang, J., Friedman, D., Ritter, W., & Bersick, M. (2005). ERP correlates of involuntary attention capture by prosodic salience in speech. *Psychophysiology, 42*(1), 43-55.

Wingfield, A., Lindfield, K. C., & Goodglass, H. (2000). Effects of age and hearing sensitivity on the use of prosodic information in spoken word recognition. *Journal of Speech, Language, and Hearing Research, 43*(4), 915-925.

Xia, M. (1946). Methods of composing two-character words [Cited in Pan, Yip and Han (1993: 37), no further citation given].

Yen-Hwei Lin (2001). San Duanmu (2000). The phonology of Standard Chinese. Oxford: Oxford University Press. Pp. xv+300. *Phonology*, 18, pp 458-463.

Yin, Fuyong (1984). Quantitative research on Chinese morphemes. *Zhongguo Yuwen* 5:338-347.

Yuan, C., Huang, C. (1998). Mandarin morpheme and morphology study based on morpheme database. *Chinese Teaching in the World*, 2, 83-88.

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language and Cognitive Processes, 9*(3), 393-422.

Zwitserlood, P. (1994). The role of semantic transparency in the processing and representation of Dutch compounds. *Language and cognitive processes, 9*(3), 341-368.