THE STRUCTURE AND PROCESSING OF PERSIAN COMPOUND WORDS IN THE MENTAL LEXICON

THE STRUCTURE AND PROCESSING OF PERSIAN COMPOUND WORDS IN THE MENTAL LEXICON: EXPERIMENTAL STUDIES ON COMPOUND WORD PROCESSING IN PERSIAN

BY

NARCISSE TORSHIZI, M.A., B.A.

A THESIS

SUBMITTED TO THE DEPARTMENT OF LINGUISTICS AND LANGUAGES

AND THE SCHOOL OF GRADUATE STUDIES

OF MCMASTER UNIVERSITY

IN PARTIAL FULFILMENT OF THE REQUIREMENT

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Doctor of Philosophy (2020) McMaster University

Linguistics & Languages Hamilton, Ontario, Canada

TITLE: The Structure and Processing of Persian Compound Words in the Mental Lexicon: Experimental Studies on Compound Word Processing in Persian

AUTHOR: Narcisse Torshizi

M.A., B.A. Linguistics: Translation Studies

SUPERVISOR: Dr. Elisabet Service

NUMBER OF PAGES: xv, 124

This dissertation is dedicated

to Dr. Nima Farzadnia

for

teaching me the true meaning of life

and to my parents

for

their endless love

Acknowledgements

I am writing this section with my heart filled with gratitude for people who have made the journey of these past six years achievable and enjoyable for me. Finishing this thesis marks the end of a long road full of ups and downs of learning, discovery, and much self-doubt. I would like to take a moment to express my deepest gratitude to the following people.

First, I want to thank my supervisor Dr. Elisabet (Lili) Service. Lili agreed to take over my supervision in the second year of my PhD program, when I decided to switch to psycholinguistics. I feel honored and incredibly grateful that she had faith in me and accepted me as her PhD student. I appreciate Lili's great knowledge in the field of psycholinguistics and experimental work, as well as her patience, time, ideas, and funding. Over the past few years, we have spent hours, days, and weeks in the lab, designing my experiments and interpreting data. It was through these many hours of discussion that I could expand my knowledge of psycholinguistics and cognitive science. I have learned so much from Lili, including how to be patient when experiments go wrong, how to find joy in research, and how to handle many different research projects at the same time. Lili has been a mentor, a friend, and my role model both in academia and in life. I feel profoundly grateful to have been Lili's PhD student and to have worked under her supervision for the past 5 years. Thank you, Lili – without your unconditional care and support, I would not even be close to where I am standing today.

I also want to say a special thank you to my first supervisor, Dr. Magda Stroińska. I met Magda at the program's open house before applying to McMaster, and she was one of the main reasons I decided on McMaster. Her warm smile, unique personality, and kind nature motivated me to be the strongest version of myself, even from my earliest months at McMaster. Magda supervised me like a mother, in the academic world and through very difficult moments of my personal life in the past six years. She made me feel like a family member and has done everything in her power to prevent me from falling behind, even in my most doubtful times. Magda, I am grateful to have you as a supervisor, a friend, and a mentor in my life; for all the priceless moments that we spent discussing linguistic and life issues; for making me feel like McMaster and your office are home when I didn’t have my family in Canada; and for changing my life academically and as an immigrant in Canada.

I would also like to thank Dr. Arsalan Kahnemuyipour for agreeing to be on my thesis committee and for his continuous support, guidance, and valuable feedback. Over the past few years, Arsalan has spent so many hours of his busy schedule discussing the details of my research with me. I am profoundly grateful to Arsalan for his great knowledge of syntax and Persian linguistics, as well as his valuable support and feedback on the theoretical part of my thesis. Arsalan has been very patient and supportive, has answered many big questions I have had, has challenged me to think outside of my comfort zone, and has been a friend while I was going through the discovery path of my research. I have enjoyed our thought-provoking conversations very much. Arsalan’s comments on my thesis drafts have added so much more insight and value to my thesis. It is fair to say that my PhD experiments and this dissertation would have never been settled without Arsalan’s guidance, feedback, and support.

I am also profoundly grateful to Dr. Gary Libben for being a part of my thesis committee and for his valuable feedback in my committee meetings. Gary’s insights have helped me see the bigger picture of my experiments and to realize how they can fill a gap in complex word processing. It was through Gary’s feedback on my first proposal presentation that I started thinking about the processing of compound words in Persian. I appreciate Gary’s unique knowledge and expertise in the field of psycholinguistics, his continuous support, and his valuable questions. He never hesitated to offer me help and support throughout the past 5 years. I feel honored to have Gary as a member of my PhD thesis committee.

I also wish to thank Dr. John Connolly and the Language, Memory and Brain Lab (LMBLab). John agreed to be on my committee for two years. While I did not end up having an EEG experiment as a part of my PhD thesis, I learned so much from John through taking his courses and working in his lab. Thank you, John, for allowing me to lead experiments in the EEG lab, for your continuous support, and for the interesting and insightful conversations we had. It was a pleasure and honor for me to be involved with the EEG lab for two years.

I would also like to thank the external examiner of this thesis, Professor Daphnée Simard, for her great comments and editions, which greatly added to the quality of this thesis.

Thank you to Dr. Daniel Pape for helping me record my experiments' auditory stimuli in the Phonetics Lab. I was also able to expand my knowledge of phonetics and phonology through his advanced graduate course in phonetics and phonology. I am grateful for Daniel's help and support over the past few years.

I must also acknowledge other faculty members in the department of Linguistics and Languages who have helped me strengthen my knowledge of linguistics and neurolinguistics: Dr. Catherine Anderson , Dr. John Colarusso, Dr. Wendy D'Angelo, Dr. Ivona Kučerová, Dr. Victor Kuperman, Dr. Anna Moro, Dr. Nikolai Penner, and Dr. George Thomas.

Special thanks go to Dr. Daniel Schmidtke and Dr. Aki Kyröläinen, for helping me with my statistical analysis. Your assistance, ideas, and questions have added so much more value to the data analysis of my thesis.

I am very grateful to Chia-Yu Lin, ARiEAL's lab manager, for her amazing support throughout the past three years. Also, thank you to Bianca James and Nanci Cole, for the great administrative support and help that they have provided to me over the past six years. I appreciate it very much.

I would like to specifically thank my best friends who have made my PhD journey easier and more enjoyable. Thank you, Dr. Rober Boshra, for always being there for me and for listening to me when I needed it the most. I really appreciate having you in my life, and I have enjoyed every part of our unique friendship. Thank you, Zoë Wälchli, for the many beautiful memories and for your true care and support. From our trail hikes to our academic challenges, we have been through many beautiful experiences together. Your friendship is very valuable to me. Thank you, Leah Richardson, for being a sister and a true friend in all ups and downs of the past eight years. I feel blessed to have you in my life. Thank you, Laura Beaudin for the time, love, and care that you have offered to me. Thank you for all those happy moments and memories that have helped me to be stronger in the valleys of my academic path. Thank you, Berrin Karabeyaz for being an amazing friend, and a huge source of love and support thorough all these years. Without the presence of all of you, this journey would not be as enjoyable.

As a member of the LMBLab, I am incredibly grateful for all of the ideas and helpful advice I have received from fellow lab members: Angie Lopez, Erin DeBorba, and Bre-Anna King. Special thanks go to my very good friends and colleagues, Chelsea Whitwell and Fareeha Rana: thank you for your kind help with my experiments, for your great ideas, and your emotional support over the past few years. I would also like to thank the undergraduate volunteers who have helped me in different phases of building my experiments: thank you Niloufar Farjam, Ariel Takach, Meliha Horzum, and Nasra Nadeem.

I would also like to thank Mary Coccia from McMaster's OSAP office who has been a huge help. She has been extremely kind and patient with me, and has gone above and beyond to find any possible sources of financial support for me. Thank you, Mary! I appreciate your kind help and warm personality very much.

My fellow graduate friends and colleagues in the Department of Linguistics and Languages and the ARiEAL Centre at Mac have had a huge role in helping me reach my goals. Here, I would like to specially thank Dr. Cassandra Chapman, Dr. Jitka Bartošová, Dr. Richard Mah, Natalia Malley, Dr. Kyle Ruiter, Dr. Connie Imbault, Dr. Mackenzie Salt, Dr. Ala Al-Kajela, and Dr. Edalat Shekari. Also, thanks to Amanda Ho, Sheryl Sawyer, Diane Doran, Gina Henrry, Heather Stevens, Mike Greencorn, Gaisha Oralova, Bryor Snefiella, Kelly Nisbet, Olivier Mercier, Melda Koscun, Nadia Kryvobok, Frederico Prado, and Nathalee Ewers.

I am so grateful to my family for their unconditional love and support. My parents have been a huge source of love and support throughout this journey and I feel blessed to have them beside me. Dad: I have always looked up to you – the dream of earning a PhD started from the time you received yours and you became my role model. Mom: your love, care, and beautiful reassuring smile have made it possible for me to be the strongest version of myself. I would also like to thank my brothers who have been there for me, not only as siblings but also friends.

Finally, I dedicate this thesis to Nima, my best friend and the biggest source of support in my life for the past 18 years. We have been through three degrees and so many chapters of life together. We have changed countries, titles, and pages of life, but our friendship, support, and care have remained the same. Thank you for your unconditional love, amazing support, for listening to me patiently, for not letting me quit my dreams, for showing me how to look for the bigger pictures in life, and for giving me wings to fly. I would not be able to do this without you.

Abstract

This thesis examines the processing of N-N compound words in Persian. The main research question is whether and how the order of their functional components, modifier and head, affects their processing. Persian is an Indo-European language with subject object verb (SOV) word order. Persian allows for different head-modifier orders in compounds and, therefore, opens an opportunity for researchers to investigate if the processing effort is different for head-final (right-headed or RH) versus head-initial (left-headed or LH) compounds. In a series of three experiments, this thesis (i) presents a novel database of familiarity, imageability, and age of acquisition (AOA) ratings for Persian compounds, (ii) investigates effects of constituent order and syntactic structure of compound words on their recall from working memory, and (iii) explores the effectivity of different constituents of compounds to activate the whole-word representation for naming.

Chapter 2 introduces a database of AOA, familiarity, and imageability ratings as lexical variables that affect compound word processing. The database also includes information on the right and left constituents of these compounds, their headedness, meaning in English, length in letters and in phonemes, and Google frequencies. Correlational analysis indicated that familiarity, imageability, and AOA are highly inter-correlated in both right-headed and left-headed compounds.

Chapter 3 presents data on compound word maintenance in short-term memory. The results revealed better recall of commonly occurring RH compounds compared to common LH ones. This supports the view that right headedness is the default form of compound words in Persian. The results also indicated that processing of an irregular or marked structure could negatively affect short term memory span in Persian, and confirmed a syntax-based decomposition approach for the mental lexicon in Persian.

Chapter 4 reports the results of a cross-modal primed naming time experiment. The results demonstrated that RH compounds have shorter reaction times than LH compounds. Importantly, there was an interaction between headedness (RH vs. LH) and whether the first or the second constituent was used as prime, suggesting that priming effects are different for the two types of compounds. The results of this chapter provide additional evidence on the role of certain syntactic structures of compounds in the processing and organization of these words in the mental lexicon.

Overall, the experimental data suggest that the head position of compound words can affect their processing in the mental lexicon. The three experiments are the first to provide such evidence for Persian. The findings also provide more proof for the decompositional routes of processing in the mental lexicon. The findings are in line with data from other languages, pointing to negative effects of marked structures on complex word access and maintenance. Another major contribution of this thesis is to provide evidence on the importance of the Ezafe construction on the processing of complex words in Persian.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AOA - Age of acquisition

Cons - Constituent

CPrs - Complex predicates

DP - Determiner phrase

EZ - Ezafe structure

Freq - Frequency

LH - Left-headed

NN - noun-noun

RH - Right-headed

RTs - Reaction times

SD - Standard deviation

SOV - Subject object verb

STM - Short Term Memory

SVO - Subject verb object

WM - Working Memory

1 Introduction

Among the unique characteristics of human languages is their ability to combine lexical units for conveying complex meanings. If we consider a single word as a lexical unit, combining two lexical units or items results in a compound word. Combining words to make new units reveals this essential ability as part of our cognitive system which directs language production and interpretation. Compound words are the result of combining two existing words to make a new lexical unit. Compounds reflect one of the core strategies for information packaging in the mental lexicon, i.e., how words are stored and processed in the brain. Jackendoff (2002) underlines the importance of compound words, indicating that they could be protolinguistic fossils from which more complex linguistic structures have been developed. Libben (2014) states that although the concept of compound production appears to be simple, understanding the neurological connections underlying compound word comprehension is central for identifying cognitive mechanism for morphological and lexical processing in general and cognitive representation of human language as a whole. The processing and representation of compound words in the brain has been a topic of debate for the past four decades. This discussion is mainly rooted in the fact that compounds can be characterized as units that are in between words and phrases. They retain the word characteristics of whole word representation and the syntactic structure akin to that of syntactic phrases and sentences. Therefore, studying compounds in general, and particularly in languages with internal structures different from English, opens up new windows on the fundamental processes of multimorphemic word production and comprehension.

This thesis studies Persian noun-noun (NN) compound words and their processing in the brain. Persian is an Indo-European language with subject object verb (SOV) word order, and it allows for variable head positions in compound words. The head of a compound is the constituent that carries the whole compound's grammatical category along with its syntactic and semantic properties (Semenza et al., 2011). For example, the Persian right-headed compound *mosaafer* ***xaane,*** literally *'traveller house'* meaning *"inn"* has ***xaane*** meaning *"house"* as its head and *mosaafer* meaning *"traveller"* as a modifier for house. An example of a left-headed compound in Persian is ***kuku*** *sabzi* literally *'frittata vegetable'* meaning *"vegetable frittata"*, which has frittata as its head in the left position. This work is mainly focused on examining the role of headedness on the processing of Persian NN compounds in the brain.

1.1 Background

1.1.1 Compounding

Compounding has relatively recently become the focus of interesting research in morphology (see for example Lieber and Štekauer, 2009; Scalise and Vogel, 2010). The approaches to compounds range from statements such as "compounds don't exist as a separate sort of word formation" (Lieber and Štekauer, 2009, p.2) to assertions that there are "compound-like and less compound-like complexes" (Lieber and Štekauer, 2009, p.10) and suggestions that "all NN constructions can be considered to be compounds" (Olsen, 2001). Although different theories have addressed structural and semantic features of compounding (see ten Hacken, 2016 for a review on these frameworks), there is still an ongoing debate on how to model different kinds of compounds. According to Marchand (1965), "[w]hen two or more words are combined into a morphological unit, we speak of a compound" (p.11).

A more comprehensive and crosslinguistic definition of compounding can be the one presented by Bauer (2006), in which compounding is referred to as a word formation process through which two or more lexemes join to form a new lexeme. Word formation is a process through which new words are made from the existing words (e.g., backformation: donation-donate) and, in the majority of compounding literature, it refers to derivation (e.g., happy-happiness), compounding (e.g., finger + print = fingerprint), and conversion/zero derivation (e.g., butterN- butterV) (see for example Štekauer & Lieber, 2006). Marchand (1969), on the other hand, believes that compounding is not a word formation process of its own. He believes that, linguistically, derivation and expansion are the only word formation processes. Under this view, compound words are constructed employing either one of these categories depending on whether their head is an independent morpheme or not. If the head of a compound word is an independent morpheme, it is an expansion (e.g., replay[1] and snow man), whereas a bound morpheme in the head position results in the compound being classified as a derivation (e.g., dissent). Therefore, in Marchand's view, compounds and prefixed words are categorized under a different category of word formation from suffixed words. Since the elements that form compounds in different languages might be different, i.e., can be either roots or free-standing words, defining a universal description of compounding is difficult (Lieber and Štekauer, 2009). Yet, different scholars have suggested different criteria to

[1] Marchand classifies prefixed words such as replay on the same category as compounds like snowman.

resolve this challenge. Donalies (2004), for example, suggests criteria for defining compounds in Romance, Slavic, modern Germanic, Greek, and Finno-Ugric languages (see Lieber and Štekauer, 2009). Compounds, according to him, are complex, formed without word-formation affixes, spelled together, have a specific stress pattern, are right-headed, inflected as wholes, and syntactically inseparable; they are syntactico-semantic islands, form conceptual units, and include linking elements (p.76). Clearly, many of these principles cannot be applied to all languages, but, according to Lieber and Štekauer (2009), three out of all of the mentioned criteria are the most important ones and will be briefly explained in the following section.

(i) Criteria for recognizing compounds

The syntactic criteria. Lieber and Štekauer (2009) suggest that the most important syntactic criterion for differentiating compounds from phrases in English, which can also be applied to many other languages, is the inseparability principle. It means that no element can be placed between the two parts of a compound word whereas it is usually possible to insert another word into a phrase. Another syntactic principle related to compoundhood is that the first stem of a compound word cannot be inflected (e.g. toothpaste but not *teethpaste), whereas a syntactic construction allows for such modification (e.g., fast walk) (p.9). Other syntactic criteria for identifying compounds are usually found in languages other than English. An example of this is seen in French. Fradin (2009) suggests that in French the word order can provide hints for compoundhood. “If a sequence of lexemes displays an order that cannot be generated for syntactic phrases, we are likely dealing with a compound” (Lieber and Štekauer, 2009, p.9). This is also suggested for English by Marchand (1960) and can be exemplified in a word like grass-green, where it is syntactically impossible for an adjective (green) to be modified by the previous constituent. Therefore, green as grass is a syntactic phrase but grass-green is not.

The behaviour of the complex item with respect to inflection. Here, there are different possibilities based on the language being studied. As mentioned above, if the language has nominal inflection as a major feature, it is usually the head of the compound, and not the modifier that is inflected (e.g., in Dutch, English or German). Another possible situation can be found in Yimas, an endangered polysynthetic language spoken in Papua New Guinea, in which the modifier noun may contain a compound-specific oblique suffix. Yet, in many languages, an inflection marker that is not compound-specific can occur on the modifier (e.g. in Finnish *“kotiintulo”* meaning *"home-coming"* from *"koti"* (nominative, home) + illative inflection (locative-to) and "tulo" (coming). Gloss: to-home-coming). These claims are undermined by compounds such as "overseas investor" and "parks commissioner" (Selkirk, 1982, p.52). Selkirk (1982) believes that, in these cases, the plural marker is added to avoid the potential ambiguity about the plural nature of the modifier but certainly this explanation can't be applied to all cases of this sort. Another interesting example that Selkirk (1982) mentions is "programmes list" (p.52). Considering that a list always includes more than one item, there should be no ambiguity to be solved by the plural marker that is added to "programme".[2]

Another important discussion that is closely related to compound inflection concerns the linking elements. Linking elements are morphemes that join the two parts of a compound without having or adding a meaning to the whole word or each of the constituents. In many languages, these elements cannot be inflectional morphemes. For example, Ralli (2013) explains that in modern Greek, after the first constituent of a compound word, there is always -o, which has no meaning attached to it and can be historically considered as a remainder of a theme vowel that is not existent any more (Lieber and Štekauer, 2009, p.10). In languages like German and Dutch, where the linking elements were historically case or number markers, the debate on whether the linking-elements can be interpreted as inflectional or not still continues. In German "Kinderstube" (literally 'nursery' but also 'upbringing'), "Kinder- "looks like the (irregular) plural form of "Kind" ('child') and "Stube" means room (See Neef, 2009, for a discussion on how to handle the linking elements in German).

Stress and other phonological means. Stress placement can be a relevant principle to rely on when it comes to defining compounds in English. The overall rule in English is that primary stress should be on the first constituent of the compound whereas syntactic phrases usually have their stress on the second constituent, which is the head of the compound. Like any other generalization, many exceptions can be found to this rule. Lieber and Štekauer (2009) point out that the context in which a compound is used, as well as various pragmatic factors can influence the pronunciation of certain compounds. Besides, many linguists believe that the placement of stress can differ based on whether the word is in isolation or is pronounced in the context of a sentence (see for example

[2] There is an increase of use of this particular compound with a singular modifier – "program list." Plural form appears mostly when an adjective is added to the noun phrase as in "full-time programs list."

Kingdon, 1958; and Bauer, 1983). Spencer (2003) points out that sometimes the same compound might have a different reading based on stress placement. As an example, a ‘toy factory can be interpreted as a factory where toys are made, whereas a toy ‘factory is a factory which is a toy (Lieber and Štekauer, 2011, p.5). Giegerich (2004) suggests that most attribute-head NN constructions are phrases and not compounds. This would explain why they often have stress on the second constituent. For example, “steel bridge” is a phrase, in which steel modifies bridge, with the stress falling on bridge (Lieber and Štekauer, 2009, p.6). The complement-head NN constructions (e.g. “body lotion”), on the other hand, are compounds and their stress falls on the first constituents. Another interesting observation in this regard has been made by Sampson (1980, pp. 265–6) who points out that compounds with the first stem describing what the second stem is made of usually bear the stress on the second constituent. Sampson (1980) notes that this is only true for compounds in which the second stem is a solid item (as in “steel bridge”). Therefore, “rubber band” has the stress on the second constituent whereas in “wine stain”, although the first constituent describes what the second one is made of, the stress is not on the second constituent. In a closely related experimental study, Schlücker and Plag (2011) demonstrated how the position of stress in novel compounds can rely on analogy to similar NN constructions that already exists in the mental lexicon of the speaker. Bauer (1983) exemplifies this with *street* and *avenue* compounds. The ones with *street* have their stress on the first constituent (e.g. First Street) whereas the ones with *avenue* are right stressed (e.g. First Avenue). However, there might be a combination of factors responsible for this analogical behaviour. Based on all of the above-mentioned points, it is safe to say that in English, if the placement of stress is on the first element, it can be a sign of compoundhood, but it is not “a necessary or a sufficient condition for distinguishing a compound from a phrase” (Lieber and Štekauer, 2009, p.8).

(ii) Theoretical frameworks on compounding

As mentioned in the previous section, how we interpret compounds semantically has been the focus in the literature on compounding over the past three decades. Different theoretical frameworks have been proposed to address this issue. The main question here is how the meaning of a compound word relates to and can be interpreted from the meaning of its constituents. In order to answer this question, some models have focused on providing criteria for the classification of compounds. An example of such a classification can be a model proposed by Scalise and Bisetto (2009). At the top level of this model, there are three classes of compounds: subordinate, ATAP (attributive or appositive), and coordinative. The first two classes are then divided into smaller branches and finally each branch is split into endocentric and exocentric compounds. Ten Hacken (2016, p.211) argues that although this framework looks as a hierarchical tree, the labels of the nodes don't represent a hierarchy. He also argues that some constructions that are classified as compounds in this framework are ambiguous. An example of this is '*blue-eyed'* that is discussed by Ten Hacken (2010) to be more of a derivation than a compound. In a recent work Ten Hacken (2016) discusses and compares three frameworks for the analysis of compounds. These frameworks will be briefly presented here.

The first theory discussed by Ten Hacken (2016) is the parallel architecture (PA) model proposed by Jackendoff (2002). Although Jackendoff proposes a set of "basic functions" for compounds, he is not interested in establishing a classification of compounds. The model distinguishes thirteen basic functions; they are mainly related to the meaning and use of compounds and are not meant as criteria for classifications of any sort. Thus, Jackendoff's model can be considered as conceptual.

The second model is the lexical semantic framework proposed by Lieber (2004), in which there is a semantic/grammatical skeleton and a semantic/pragmatic body and it's the meaning or a part of the meaning of a compound structure that classifies compounds in this model. The semantic/grammatical skeleton consists of seven binary elements. For example, the feature +/- material distinguishes between concrete (+material) and abstract (-material) nouns. The presence or absence of these elements can also be considered separately. Therefore, this model proposes many different theoretical combinations. Yet, the structure of Lieber's semantic/grammatical skeleton is more than just combining some features. Therefore, different grammatical or syntactic classes can be distinguished in in the skeleton, and so it is possible to analyze whether two compounds belong to the same syntactic category in the skeleton or not. In fact, Ten Hacken (2016) believes that determining if two compounds belong to the same skeleton class in Lieber's (2004) model is less complicated than determining their conceptual structures based on Jackendoff's PA model. The semantic body, in Lieber's (2004) model refers to the range of interpretations that are possible for a given compound and includes both speaker dependent, idiosyncratic interpretations, as well as more universal, "encyclopaedic" elements of meaning. Ten Hacken mentions that "Lieber aims to build up the meaning of a compound from the meaning of its parts, whereas Jackendoff describes the meaning of the compound, referring to the meaning of the parts" (Ten

Hacken, 2016, p.230). In Jackendoff's model, conceptual and spatial structures are different. Both structures are to represent meaning, but the spatial structures deal with the visual, haptic, and action systems, whereas the conceptual structure mainly deals with language. Therefore, Jackendoff's model provides richer information than Lieber's but Lieber's concept of skeleton has a more formal structure than Jackendoff's (see Ten Hacken, 2016, pp, 229-232 for a more detailed comparison of the two frameworks).

The third framework discussed is Štekauer's (2005) onomasiological model. Onomasiology is the study of naming or designation mechanisms. Štekauer views naming as the essential motivation for word formation since the use of word formation rules, in his model, is determined by the necessity to name new concepts. Štekauer's onomasiological model can be considered as the cognitive basis for the "act of meaning" (see *Figure 1-1* for the image taken from Štekauer, 2016, p.56 on onomasiological model of complex words). The onomasiological method analyses the process from conceptualization of an object to how it is labelled linguistically. The relations between semantic categories such as agent, time, instrument, manner, etc., play an important role in the construction of this model. The *"onomasiological base"* is responsible for classifying the object that is to be named. The *"onomasiological mark"*, which is the same as a modifier, identifies the base concept, including both the *"determined"* and the *"determining"* constituent. "The determined constituent always stands for the cognitive category of Action in one of the three modifications (Action proper, Process, and State) and is, as it were, the crucial constituent in terms of interpretation/predictability of the meaning of novel complex words, because it has the capacity to relate the other two constitutes of the onomasiological structure" (Štekauer's, 2016, p.56)[3].

So far compounds have been considered from the point of view of linguistic theory. The following section discusses the main psycholinguistic frameworks for the explanation of compound word processing.

(iii) Morphological processing of compound words

The morphological processing of English compound words has been researched extensively.

[3] A part of this direct quotation that is related to meaning predictability, as mentioned by Štekauer (2016), is taken from principles presented by Štekauer (2009).

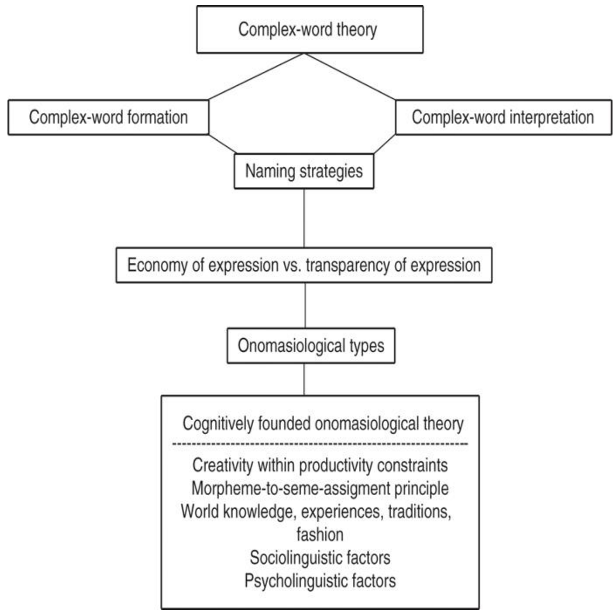

Figure 1-1 An onomasiological model of complex words. Adapted from 'Compounding from an onomasiological perspective,' by P. Štekauer, 2016, in The Semantics of Compounding, p. 67.

The focus has been on the question of whether compounds are accessed as wholes or through their constituents. Forster and Taft (1976) were the first to show that the frequency of the first lexeme can affect lexical decision performance. They proposed a model which suggests that complex words are fully decomposed into their constituents and their stem forms are stored in the mental lexicon. Based on this model (called *full decomposition model*), the stem forms, as targets for lexical search, are required for whole word access. On the other end of the decomposition spectrum is a model by Butterworth (1983), suggesting that compounds are processed and stored as wholes and there is a separate representation for each compound word in the mental lexicon. This model, being referred to as the *full listing hypothesis*, has been challenged by other scholars. Libben (1998) noted that this model would require a huge capacity of the mental lexicon since all compound words would need to be stored as wholes, with no decomposition taking place. Views alternative to this model suggest that activation of constituents happens but only after the whole word has been activated (see for example Giraudo & Grainger, 2001). Data from experimental research does not fully support either theory. *Hybrid models* of complex word processing, also known as *dual route models*, propose both whole word activation and decomposition routes, which in the early versions of these models work in parallel and give rise to separate frequency effects based on the whole word and its constituents. This means that when the whole word frequency is higher, the complex word is processed as a whole but if the constituent morphemes have higher frequency than the whole word, the complex word will be decomposed to its morphemes (Semenza & Luzzatti, 2014; see also Badecker & Caramazza, 1991; Laudanna, Badeckera, & Caramazza, 1992; Schreuder & Baayen, 1995; Pollatsek, Hyönä, & Bertram, 2000). The latest versions of this type of models, the *multiroute models*, suggest a flexible process even from the earliest milliseconds of complex word processing. Based on this model, compound word processing is dependent on both the characteristics of whole words and those of its constituents (Kuperman, Schreuder, Bertram, & Baayen, 2009).

The above-mentioned theories of complex word processing have concentrated on the lexical characteristics of these words. Yet, another main topic of debate are the semantic properties of the complex words and how they interact with the lexical aspects of these words. Libben (1998) underlines that the processing and representation of multimorphemic words in the mental lexicon depends on the language being studied along with many other lexical factors. These include, but are not limited to, word frequency, lexical category (noun, verb, etc.), and type of morphological structure (compounded, derived, inflected). He mentions that another important factor in this context is the semantic relationship between the constituents of complex words, as well as their semantic transparency. Sandra (1990) mentions that although many studies indicate that, when being processed, semantically transparent compounds (e.g. *doorbell*, *blueberry*) are decomposed into their constituents, there are studies reporting that opaque compounds (e.g., *humbug, deadline*) show no priming effects when their constituents are used as primes (e.g., Sandra, 1990). Based on traditional models, it is only in later stages of processing that semantics can have a role in the lexical access (see Libben, 1998) while recent theories highlight the role of semantics even at the very early stages of processing (see Marelli & Luzzatti, 2012).

The "lemma" model (Levelt, Roelofs, & Meyer, 1999) is the key theory relevant for compound word production. A lemma, as defined in this model, is a representation of the syntactic and semantic aspects of a word that are activated after the first purely conceptual stage in word production. The result of the second stage, which incorporates phonological information, is the lexeme. Lemmas, therefore, link the grammatical structures and semantics to their phonological information and are suggested to be the level at which the compound word information is stored. This hypothesis is supported by psycholinguistic experiments (see for example: Badecker, Miozzo, & Zanuttini, 1995; Delazer & Semenza, 1998; Biran & Friedmann, 2011). A similar proposal to lemma theory was an older one introduced by Caramazza (1997). This was called the *independent network model*, suggesting that there is a separate network (not an intermediate level of the production process) for storing the syntactic information of words and that "the selection of grammatical features typically occurs temporally prior to the selection of the specific phonological and orthographic content of a word" (Caramazza, 1997, p.195). The representation of morphologically complex words is viewed differently in the two models. In the independent network model, words are stored as whole forms (Caramazza, 1997; Janssen, Bi, & Caramazza, 2008), whereas in the lemma model, morpheme-based representations are stored as word forms in the mental lexicon (Levelt et al., 1999). It should be noted that Caramazza's model is based on the reported frequency effects for whole words, when no frequency effects for morpheme-based representations of nominal compounds were found, and is not directly related to compound words (see Semenza & Luzzatti, 2014).

As mentioned above, several factors, such as the *lexical characteristics* of compounds, the *frequency of the whole word and its constituents*, and a compounds' *semantic transparency* can all play important roles in the processing of compound words. Another important variable in this context is *headedness*. Headedness is the order of the syntactic roles of the constituents, i.e. the head and the modifier, in noun-noun (NN) compound words. This order can vary based on the language being studied. Compounds can be head initial (left-headed) or head final (right-headed). In most languages the position of the head is fixed. In English, which is an Indo-European, Germanic language, the head of the compound is always the right constituent, whereas some languages, such as Vietnamese and Hebrew, are left-headed, meaning that the head is always the first (left) constituent (see for example Fabb, 1998 for Vietnamese). In some Indo-European, Romance languages, such as Italian and French, the head position in compounds varies (see Libben & Jarema, 2006 for French; Semenza & Luzzatti, 2014 for Italian). Interestingly, Persian, another Indo-European language, also allows for variable head position (see Kahnemuyipour, 2014; Foroodi-Nejad & Paradis, 2009; Kalbasi, 1997; Shariat, 2005). This variability in headedness raises the question of different processing of compounds based on their head position.

If we agree with decomposition taking place, we accept that the constituents of a compound word are necessarily affected by their position-in-the-string, as language unfolds over time. This brings up another general debate in the literature. Is position-in-the-string more important than morphological headedness in compound processing? Early research on English shows that the first constituents of polymorphemic words play a more crucial role than the second ones (see for example Taft and Forster, 1976). However, with the position of the head being always fixed in English, there is no opportunity to investigate this question further without data from other languages. Jarema et al. (1999) conducted a comparative study on the headedness effect in French, a language which also allows for variable head positions, and Bulgarian, to isolate the functional effects of headedness from the position effects. Their primed lexical decision experiment revealed a stronger priming effect for the first constituents than for the second ones in left-headed (LH) but not right-headed (RH) French compounds. Similar study by Jarema et al. (1999) suggests that the head constituent improves the processing more than the modifier. These findings also indicate that compounds are decomposed during processing and that, with the exception of having semantically opaque constituents, compound words go through the decomposition process in the mental lexicon (see Wälchli, 2016; Boutonnet et al., 2014).

(iv) Role of memory in compound processing

The existing literature on the effect of morphological complexity on memory recall tasks is limited (see for example Wälchli, 2016; Service & Maury, 2015; Németh et al., 2011; Service & Tujulin, 2002). Research on Finnish looked at how morphological complexity can influence the recall of word sequences in working memory (WM) experiments involving simple recall or combined processing and recall. Results indicate that memory load was affected by the presence of both inflection and derivation markers (Service & Tujulin, 2002). The findings also indicated that the processing could vary for the different classes of complex word forms, which demonstrates that different word classes have qualitatively different representations. This questions the theories on how cognition interacts with lexical and semantic characteristic of words (see Service & Maury, 2015; Service & Tujulin, 2002). A similar study (Németh et. al., 2011) investigated how verbal short-term memory (STM) can be influenced by morphological complexity. They used Hungarian two-syllable stems and two-syllable complex words as their stimuli. An additional experiment with three syllable words was later added. The findings showed that morphological complexity can negatively affect STM span. The recall was reported to be better for derived words compared to inflected ones. Besides, irregular structures appeared to result in poorer recall than morphologically regular structures. This could suggest that irregular forms added to STM load even if the forms were not readily decomposable. Thus, the added load might result from added semantic complexity (See also Chase & Simon, 1973).

In a different experiment, Matzen & Benjamin (2009) used compound words in a recognition task to investigate how sentence context predicts different forms of false memories. In one experiment, compound words that could form a conjunction lure (e.g. tailspin + floodgate = tailgate) were used as stimuli. They were presented to participants, either as single words in lists or in a sentence context. In a recognition test, participants saw previously presented compounds, compounds that had not been presented, i.e. conjunction and semantic lures. They had to decide for each word if they had seen it among the stimuli before. The findings revealed that when compounds were seen in a sentence context, more semantic lure errors and fewer conjunction lure errors were made. The results for single-word lists were opposite. The authors explained these findings might be related to different strategies in the two contexts. Conjunction lure errors were more common in word lists with minimal protective semantic contexts. Thus, experiments with word lists may be especially sensitive to morphological decomposition.

Several studies address the question of decomposition in visual word recognition. These include constituent-priming tasks that examined if lexeme frequency plays a part in lexical decision times (see for example Andrews, Miller, & Rayner, 2004; Libben, Gibson, Yoon, & Sandra, 2003). A good number of studies in English have reported that when constituent lexeme frequency is manipulated, while the whole word frequency is controlled, processing times can vary. This suggests that constituent lexemes are accessed during compound word recognition (see for example Juhasz, Starr, Inhoff, & Placke, 2003; Inhoff, Starr, Solomon, & Placke, 2008; Andrews, Miller, & Rayner, 2004). Similar results have been reported from research on languages like Finnish, in which compounding is more productive (see for example Pollatsek, Hyönä, & Bertram, 2000). Productivity of word formation generally refers to the potential to make new words for representing a concept with a provided word-formation pattern (Libben & Jarema, 2006). However, when this potential is recognized, it can form the actual word patterns in the language and thus the richness of compounding in any language is correlated with the number of productive patterns of compounding that the given language owns (Libben & Jarema, 2006, p.346). Recent studies in the field have focused on how morphological families and the conceptual relations between the compound constituents play a part in compound word organization in the mental lexicon (e.g., Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009; Schmidtke, Kuperman, Gagné, & Spalding, 2015).

The main goal of this dissertation is to provide more information on the representation and processing of Persian noun-noun compound words in the mental lexicon. Since RH and LH compound words in Persian have different status (Kahnemuyipour, 2014, p.6), I hypothesized that this internal syntactic difference results in different processing of RH versus LH compounds in this language. I hypothesized that the RH compounds, which are the default form in the language and are called "true" compounds (Kahnemuyipour, 2014, p.6) are processed faster by native Persian speakers. To better understand the fundamentals of my dissertation research work, some relevant syntactical features of the Persian language should be described. These are reviewed briefly in the following section.

1.1.2 Persian language

Persian belongs to the southwestern branch of the Iranian language family, which is spoken in Iran (Farsi or Persian), Afghanistan (Dari), and Tajikistan (Tajiki). Dabir-Moghaddam (2019) has proposed that modern Persian is an analytic language. He argues that the large number of prepositions present in modern Persian shows that this language is, morphologically, an analytic language and from a syntactic typological view, a prepositional language. Based on the linear word order of modern standard Persian, previous literature within generative frameworks has suggested the canonical word order of the language to be SOV[4] (see for example Dabir-Moghaddam, 1982; Hajati, 1977). According to Frommer (1981), this order is "(Subject) + (Adverbs) + (Direct Object) + (Indirect Object) + Verb" (Sadeghi & Shabani-Jadidi, 2019, p.54). Greenberg (1966) claims that languages with normal SOV word order are universally postpositional with Persian as one of the exceptions to this rule.[5] The canonical order of a noun phrase in Persian is noun + modifier (where modifier follows the noun), linked by a morpheme known as Ezafe. The dominant form of Persian NN compounds, consisting of a head noun and a modifier noun, is where the modifying noun comes before the head noun, but there are also some compounds (Kahnemuyipour, 2014). I shall discuss this difference between RH and LH compounds further below, where I provide description and examples of Ezafe construction in Persian.

[4] See Marashi (1970) for a different proposal that suggest the SVO order for Persian.
[5] The only postposition in Persian is the direct object marker "ra" (Dabir-Moghaddam, 208)

First, I am providing some important typological parameters on Persian, which can be beneficial for understanding this study. All parameters and examples, provided in this section, are taken from Dabir-Moghaddam (2019, pp.57-61).

- A relative clause always follows its head noun as shown in example (1).

(1) *Mærd-i* *ke* [*diruz* *baa* *šomaa* *sohbæt* *kard-Ø*]

Man-head marker that yesterday with you talk do.PST-3SG

'The man who talked to you yesterday'

- The head noun always precedes a genitive. Example (2) demonstrates this structure. EZ means Ezafe linker, which will be explained after the examples.

(2) *pedær-e minaa*

father-EZ Mina

'Mina's father'

- Modifying adjectives follow their head noun, as in example (3).

(3) *zæn-e* *mehræbaan*

woman-EZ kind

'A kind woman'

- Demonstratives *in* (this) and *an* (that) always precede the noun as exemplified in (4).

(4) *aan* *zæn*

that woman

'that woman'

- Numerals are positioned before the noun. Example (5) illustrates this structure.

(5) *do ketaab*

two book

'two books'

- The indefinite article suffix *-i* 'a(n)' and the colloquial definite article suffix *-e* appear after the noun as shown in examples (6) and (7).

(6) *ketaab-i*

book-INDF

'a book'

(7) *ketaab-e*

book-DEF

'the book'

- The possessive morphemes in Persian are always enclitics. Examples (8) and (9) illustrate this.

(8) *ketaab=æš*

book=3SG.POSS

'his/her book'

(9) *ketaab-e elmi-y= æš*

book-EZ scientific-hiatus filler=3SG.POSS

'his/her scientific book'

The Ezafe construction is known to be one of the most interesting characteristics of Persian noun phrases. Mahmoodi-Bakhtiari (2018) points out that Ezafe is the most important Persian clitic and functions as one of the most special syntactic properties of Persian. According to Karimi (2019), Ezafe is a morphosyntactic element that extends over several constructions within the Determiner Phrase (DP). Ezafe is derived from the Arabic word 'idafa(t)', literally means *'addition'* (Karimi and Brame, 2012). The Ezafe affix -*e* or -*ye* can be added to mass nouns, count nouns, verbal nouns, past participles, pronouns, quantifiers, and adjectives, but not verbs, adverbs, conjunctions, or certain prepositions and can be used for showing possession, origin, modification, specification, material, and more (Karimi, 2019). When there are several modifiers in a noun phrase, each one except the last one can be linked to the previous one by an Ezafe affix. If there is a possessor, it is always the final noun, as in the example below.

(10) *Pirhæn-e æbrišæm-e aabi-e Nimaa*

shirt-EZ silk-EZ blue-EZ Nima

'Nima's blue silk shirt'

There are different views of the syntactic nature of the Ezafe Structure. Samiian (1994) (also Larson and Yamakido 2008) believes that Ezafe affix is a case marker and that the reason why clauses and real prepositions[6] do not let Ezafe follow them is that they don't require a case marker. Ghomeshi (1997) mentions that Persian common nouns, categorized as X^0, do not come with any

[6] Samiian (1994) classifies Persian prepositions into two main groups. She believes that the real ones are the ones that don't allow Ezafe to follow them such as *ba* 'with' or *az* 'from'. The other group such as *Kenaar* 'next to' or *ru* 'on' can go with Ezafe.

complements or specifiers; thus, the only available place for the possessor phrase is the specifier of the DP, on the right of the phrase. She defines the Ezafe element as a linker attached to X^0. Samvelian (2007) views Ezafe as a phrasal affix, which illustrates dependency relations of the head noun with its modifiers and the possessor NP and acts similarly to personal enclitics and the indefinite article *-i*. Kahnemuyipour (2006, 2014) defines the Ezafe structure, based on Cinque's (2010) proposed DP structure, as a reflex of a roll-up movement. Cinque (2010) suggests that the roll-up movement of elements in the phrase is responsible for variations of the word order. Applying this analysis to Persian and assuming that Persian NPs are head-final, Kahnemuyipour (2014) proposes a phrasal movement analysis of Ezafe and explains that the surface structure of orders is the result of the phrasal movements of the NP into the specifier positions of the intermediate functional heads. Here, movements go with overt morphology. Therefore, the prenominal order within the NPs is considered "basic" and the postnominal order, including Ezafe, as a result of the roll-up movement, goes with overt morphology. As mentioned before, Ezafe never appears on elements before the noun, and it only accompanies the elements following it. Thus, prenominal components, which are considered as heads (X^0s) and don't have any roll-up movements, don't need the Ezafe affix. Considering this last point, Kahnemuyipour rejects Ghomeshi's proposal that describes the Ezafe element as a linker connected with X^0.

As mentioned before, the NN left-headed and right-headed compounds in Persian have different internal structures. This might be related to how compounding works in the language. Kahnemuyipour (2014) considers the left-headed compounds in Persian to be 'historical' compounds. In this view, Persian LH compounds "originated as a syntactic phrase in the Ezafe construction", but the frequency of use resulted in the Ezafe vowel being dropped over time (p.6). These LH compounds, therefore, are "syntactically formed" and have "semantically more compositional source" that results in their more transparent meaning (p.6). According to this view, right-headed compounds are considered to be 'true' compounds, "which reflect the base order of modifiers and nouns in Persian":

> This compound formation can be seen as the result of some kind of morphological merger. This morphological merger occurs between two heads, typically A and N, which will be involved in the rest of the syntactic derivation as a single unit [A N]. Given that this

compound formation involves the merger of two heads, phrasal modifiers are barred from this construction… (Kahnemuyipour, 2014. p.6).

In the following section, I first review the main characteristics of simple and compound words in Persian and then discuss some theories on how Persian compound words are structured in the mental lexicon. The next section reviews the existing literature on the processing of Persian NN compounds and different models that have been proposed in this area.

(i) Simple versus compound words in Persian

Persian is a language with concatenative morphology. Words usually consist of a stem and one to five affixes, prefixes or suffixes, that are added to the stem. Affixation as a word-formation process is productive in Persian and morphological processes are the main way of forming new words in the language, except for loan words, which can be considered as exceptions.
A simple word in Persian consists of a root morpheme to which inflectional or derivational morphemes are added (Nojoumian, 2011) as illustrated below:

“ROOT + [derivational morphemes] + [inflectional morphemes].” (p.33)

In Persian, prefixes are mainly attached to verbs. Nojoumian (2011) found 100 suffixes and 20 prefixes in his text corpus of Persian words. The suffixes were divided into 20 inflectional and 80 derivational ones. Compounding, according to Nojoumian, is the main verb-formation process in Persian. This has resulted in an extensive number of compound verbs in this language. Derivation is another word–formation processes which is productive in Persian. This can be seen in many derived stems of the simple verbs. Note that inflectional morphemes in Persian include clitics, which can mark subject-verb agreements. Below is an example of a complex Persian noun with a prefix, verbal root and two derivational morphemes, taken from Nojoumian (2011, p.34):

(11) ناتراوایی (impermeability)

/naa-taraav-aa-yy/

{Inflectional prefix + [[verb root + derivational morph.] + derivational morph.]}”

A word in Persian can have “zero to a maximum of three prefixes and zero to a maximum of five suffixes” (Nojoumian, 2011, p.34). A simple word consists of one lexical morpheme, e.g. “mother” /madær/. Loanwords in Persian are mainly from French, Arabic, English, and Turkish and due to their non-productivity are considered as simple words (i.e. are not decomposed).

A compound word in Persian consists of at least two lexical morphemes. In Persian orthography, there is no consistency in putting a space in between the two constituents of compound words. Although, there is usually a space in between them, there are words with the two constituents attached or with half a space in between the two parts. As previously mentioned, although Persian allows for variable head positions, the dominant pattern for Persian NN compounds is right headedness (see for example Kahnemuyipour, 2014). Here are some examples of the RH compounds in Persian:

(12) *kæmær bænd*

waist band

'belt'

(13) *mosaafer xaane*

traveler house

'inn'

(14) *gol aab*

flower water

'rosewater'

The main difference between this type of compounds in Persian and their LH counterparts is that the latter are made of two nouns combined originally with the Ezafe construction, which was gradually dropped[7] (Kahnemuyipour, 2014). This LH form can be substituted with the productive compounds with the Ezafe (N-Ez Mod) with no change in their meaning. Therefore, almost all the LH compounds in Persian can be expressed with Ezafe and still maintain the same meaning. This is different from RH compounds in Persian. For example, *nun lavaash* meaning *"Lavaash bread"* is an LH compound which means the same thing as *nun-e lavaash* which is a syntactic N-Ez Mod structure, whereas *ketaab-xaane* (an RH compound) meaning "*library*" and *xaane-ye ketaab* are different. The latter means "*the house of books*" and cannot be used to mean library (Kahnemuyipour, 2014). Of course, one can argue that an N-Ez Mod structure such as *aab-e porteqaal* meaning "*the juice of orange*" and *aab porteqaal*, which is an LH compound meaning *"orange juice"*, are not exactly the same. The former can refer to a fresh and more natural orange

[7] It should be noted that we cannot consider a Persian LH compound such as *aab limu* 'lime juice' to be a compound if it is pronounced with Ezafe structure. Therefore, *aab-e limu* 'the juice of lime' is not a compound word.

juice whereas the latter generally refers to any form of orange juice that is produced industrially and is sold in a can or bottle. It is important to note that the order of the constituents in the LH type of compounds is identical to the canonical order in noun phrases (Kalbasi, 1997; Shariat, 2005; Mahoozi, 2006; Arjang, 2006; Kahnemuyipour, 2014). Persian LH compounds are more transparent than the RH ones. Below are some examples of Persian LH compounds:

(15) *aab* *sib*
water apple
'apple juice'

(16) *nun* *lævaaš*
bread Lavaash
'Lavaash bread'

(17) *hælghe ezdevaaj*
ring marriage
'wedding ring'

In Persian X + Y compounds, finding the head results in revealing a possible relationship between X and Y. This relationship according to Foroodi-Nejad and Paradis (2009) is "fairly obvious in adjective–noun compounds such as, (18) *loobiaa sabz* literally *'bean green'* meaning *"green bean"* (p.414) or (19) *pirmard* literally *'old man'* meaning *"old man"* (p.414) in that "loobia sabz can be a green-type of beans and pirmard can be an old-type of men" (p.414).

In Persian noun–noun compounds, the head–modifier relationship is relatively transparent and easy to guess for both RH and LH compounds. For example, (20) *mosaafer xaane* literally *'traveler house'* meaning *"inn"*as a right-headed compound and (21) *hælghe ezdevaaj* literally *'ring marriage'* meaning *"wedding ring"* as a left-headed compound. There are exceptions to this rule when the semantic relationship between the elements of Persian NN compounds appears ambiguous. Foroodi-Nejad and Paradis (2009) illustrate the exceptions with the example of (22) *sofre maahi* literally *'tablecloth fish'* can be interpreted as "*stingray*" ('fish'=head) or a "*tablecloth-with-pictures-of-fish-on-it*" ('tablecloth'=head)" (p.414). The authors explain that this is a situation where we need the context to solve the problem of semantic ambiguity.

As already mentioned, some heads in Persian are flexible enough to be in either left or right position. However, it should be noted that there are also some Persian compound NN structures with fixed head positions. Examples of these are the compounds reflecting kinship relations. If the head position changes in this group of words, the semantic result is a reverse kind of kinship relationship. Foroodi-Nejad and Paradis (2009) mentioned the followings as an example of this group:

(23) *xaahar shohar*
sister husband
'sister in law' (sister of husband) (p.414)

versus

(24) *shohar xaahar*
husband sister
'brother in law' (husband of sister) (p.414).

There are other groups of Persian NN compounds with a so-called preferred head position. This means that certain heads are preferably used in either right or left position. This can be seen in the family of compounds related to concepts such as juices and aches. Thus, compounds for various types of juice are mainly left-headed such as in (25) and (26):

(25) *aab sib*
water apple
'apple juice'

or

(26) *aab aanaanaas*
water pineapple
'pineapple juice'

whereas compounds for ache are mostly right-headed. Examples of the latter include:

(27) *shekæm dard*
stomach pain
'stomach pain'

or

(28) *zaanu dard*

knee pain
‘knee pain.’

Therefore, the interpretation of Persian NN compounds is affected by the context as well as by the semantic family membership of the expressions under consideration (Foroodi-Nejad & Paradis, 2009).

Another point that needs to be added here is that the plural marker “ha” can only be added to the second constituent in Persian NN compounds (example 30.b) whereas in noun phrases it comes after the head noun (example 29.b). To illustrate this, the following examples are taken from Foroodi-Nejad and Paradis (2009, p.415).

(29) a. *morabi- e footbaal khoshaal ast*
coach-SG-EZ soccer happy is -SG
‘the soccer coach is happy’

b. *morabi-ha-ye footbaal khoshaal hast-and*
coach-PL-EZ soccer happy be-PL
‘the soccer coaches are happy’

c. ∗*morabi-e footbaal-ha khoshaal hast-and*
coach-EZ soccer-PL happy be-PL
‘the soccer coaches are happy’

(30) a. *tanhaa yek [ketaab xaane] dar in shar ast*
only one book house in this city is
‘there is only one library in this city’

b. *[ketaab xaane] - [ha]-ye ziyadi dar in shar ast*
book house- PL-EZ many in this city is
‘there are many libraries in this city’

c. ∗*[ketaab -[ha] [xaane] -ye ziyadi dar in shahr ast*

book PL house-EZ many in this city is
'there are many libraries in this city'

Below is an example of this in a left-headed:

(31) a. *Nima tanhaa yek [pesar amu] daarad*
Nima only one son uncle have-SG
'Nima has only one cousin'

(31) b. [*Pesar amu]- [ha]-ye Mahyaar khoshaal hast-and*
Cousin -PL-EZ Mahyaar happy be-PL
'Mahyaar's cousins are happy'

(31) c. *∗[Pesar -[ha] [Amu] -ye Mahyaar khoshaal hast-and*
Son -PL uncle-EZ Mahyaar happy be-PL
'Mahyaar's cousins are happy'

Examples (29) and (30) are stated by Foroodi-Nejad and Paradis (2009) as an indicator of compounds being considered as a single unit in Persian. I shall discuss this further in the following sections where I provide an overview of the theories of how Persian compound words are processed in the mental lexicon. The section starts with the most recent experimental works on Persian words.

(ii) Structure of the mental lexicon in Persian

The concept of a mental lexicon has been a topic of many psycholinguistic studies on different languages, including Persian. There are only few psycholinguistic studies of Persian word processing. Research done by Shabani-Jadidi (2014) and Nojoumian et al. (2006) are examples of this research on Persian. Nojoumian et al. (2006) used Persian compound nouns as stimuli for a lexical decision task and Shabani-Jadidi (2014) examined the processing of transparent and opaque compound verbs under a masked priming paradigm for a lexical decision task. Shabani-Jadidi (2019) concludes that these studies point to the decomposition of opaque, transparent, and pseudo-complex words to their smallest constituting elements (p.412) and proposed that since Persian is a morphologically rich language, native speakers break down words into their constituent elements in a linear order "regardless of whether the remaining element is an existing word or morpheme in the language" (p.412). There are similar arguments to this for other morphologically rich languages like German (see Smolka et al., 2008).

Investigating the structure of words is a topic of interest to theoretical linguists as well as psycholinguistics. Therefore, there are different theories addressing this issue. In her review of psycholinguic work done on Persian, Shabani-Jadidi (2019) discusses the relevant theories for the structure of language and the mental lexicon, concluding that most of the studies on Persian syntax, and more specifically those on complex predicates (CPrs) support a constructionist view (such as that of Marantz, 1997, which is explained below) for Persian CPrs. CPrs consist of more than one word and convey information that can usually be expressed by a simple verb in English, for example *shekast daad***,** literally *'defeat gave'* meaning *"defeated"* (Karimi, 2019, p.167) (for more information on Persian CPrs see Dabir-Moghaddam, 2019; Karimi, 2019; and Ghomeshi, 2019). Lexicalist views, such as that of Pinker (1989), suggest that all of the word information, such as roots, suffixes, and prefixes, as well as the related word-formation rules or the internal structure of various verbs are stored in the lexicon. On the other hand, constructionist theories, (see for example Marantz, 1997) state that the mental lexicon may not have enough capacity for storing all the syntactic rules of all words and their constituents. They grant this function to syntax, proposing that only the smallest units of words, i.e. roots are stored in the lexicon. Marantz bases his view on the merge and move mechanism, i.e. that all the words used by the syntax of a language are constructed by this mechanism of grammar (see also Chomsky, 1992). According to him, the mechanism for constructing words is the same one that is also used for constructing phrases. A recent framework in line with the constructionist view is the Distributed Morphology (DM) theory that postulates that our mental lexicon only stores the neutral roots. These roots can later go through derivation, where all the category-defining functional heads are combined (see Borer 2005). The common ground between constructionist theory and DM framework is that they are both syntax-based. As mentioned before, most of the research on Persian CPrs supports the constructionist view of the structure of the Persian language (See Shabani-Jadidi, 2019; Folli et al, 2005). This view is, of course, in contradiction to the lexicalist approach. The argument is supported by the systematicity and interdependence of complex predicates' nominal and verbal constituents, meaning that the lexical specifications of the "nonverbal (NV) element or the light verb (LV)" cannot simply construct the event structures and syntactic properties of Persian CPrs

(Karimi, 1997), “therefore suggesting that the syntactic and semantic properties of these elements must be determined post-syntactically rather than in the lexicon” (Folli et al., 2005, p.1365). Another argument that Folli et al., (2005) raise in support of the constructionist approach to Persian is how systematicity contributes to identifying event structure and replacement possibilities of whole complex predicates (p. 1369). There are similar studies such, as Megerdoomian (2001), which provide further evidence on a syntax-based full compositional approach to Persian. Here, the author mentions that it is not the lexical entries of the nominal and verbal elements that determine the semantic and syntactic properties of the complex predicate, it is rather the syntactic construction of these elements that play the most important role in this context.

Although most of the psycholinguistic studies in Persian suggest a fully decompositional route for word processing, there are some studies, such as Samvelian and Faghiri (2014), which theoretically challenged the fully compositional route for Persian CPrs and proposed a partial compositional route for them. They explained that “ not only the lexical meaning of Persian CPrs is barely ever fully predictable from the meaning of their component parts, but also that even more abstract properties, such as argument and event structure, cannot be determined a priori, on the basis of solely one component of the CPr regardless of the other one (p.72). They referred to Nunberg et al. (1994) view of idiomatically combining expressions and point that there should be no contradiction between storage and compositionality of CPrs if compositionality is viewed as posteriori. Shabani-Jadidi (2019) argues that this hypothesis is in contradiction to most of the findings of experimental studies on Persian CPrs. An example of this are Shabani-Jadidi’s (2014, 2016) results revealing that when her participants heard or read the word “zamin” which means “earth”, the compound verb “zamin kandan” which means “earth-to dig” and the idiomatic verb “zamin xordan” which means “earth-to eat / to hit”, meaning “to fall’” were activated. She also mentions that the translation equivalents of these words in other languages that the participants knew, as well as in their idiomatic and literal varieties, were activated within milliseconds (Shabani-Jadidi, 2019, p. 413). She believes these results are indicators that support fully decompositional routes in Persian.

Based on the above-mentioned findings, decomposition was assumed in the present study that addressed the hypothesis that the functional roles of the different constituents in Persian NN compounds affect processing.

(iii) Psycholinguistic experiments on Persian compound words

(a) Lexical decision tasks

As mentioned before, up to this date, there are only few experimental studies on the processing of words in the Persian language. Some of these include lexical decision tasks on Persian complex predicates. For example, Shabani-Jadidi (2014) used a masked priming technique to investigate three kinds of relations between the constituents in Persian CPrs. The first type included stimuli that were relatively semantically transparent, for example: compound verb "qazaa-khordan" 'food-to eat', qazaa means food; khordan means to eat; qazaa-khordan means to eat food. The second group were relatively semantically opaque, such as "qasam-khordan", 'oath- to eat' qasam means oath; khordan means to eat; qasam khordan means to swear. The third group of stimuli were orthographically overlapping such as "shenakhtan" 'to recognize', shena means swimming, which the author called "pseudo-compounds". It should be noted that shenakhtan is a word (infinitive) by itself and is not related to "shena" (see Shabani-Jadidi, 2019, p.420 for details of the experimental design). The goal was to examine whether compound verbs in Persian get decomposed into their constituents and, if so, whether this decomposition is on the basis of morphological or orthographical forms or semantic transparency. There were two experiments in this study. The first one was designed to investigate the priming effect of the compounds and pseudo-compounds on their nominal constituents/pseudo-constituents and the second one was focused on those effects on the verbal constituents/pseudo-constituents.[8] The results in both experiments showed significant priming effects. Based on these results, the author claimed, that decomposition was happening at early stages of compound word processing in Persian. The authors didn't mention anything in regards with the possible phonological effects that might have an overlap with their reported results .

In a similar study, Nojoumian et al. (2006) investigated the effect of relatedness on Persian compound nouns. They grouped compound words similarly to the previous study, i.e. into compounds that were transparent, opaque, and orthographically related (pseudo-compounds). An example for the first (transparent) group is a word like "sar-angosht" 'fingertip', with sar meaning tip and angosht meaning finger, and the target word being the second constituent "angosht"

[8] Note that complex verbs in Persian are made up of a non-verbal element, often a noun, added to a light verb. It is worth noting that there are many light verbs in Persian used in the formation of complex predicates.

'finger'. The second condition includes opaque compounds such as "sanjab" 'squirrel' as prime where the target is "*sanj*", meaning measure. The word "*aab*" means water, but it is important to note that "*sanjaab*" is a mono-morphemic word by itself and is not made up of or related to "*sanj*" and "*aab*". The third condition is orthographically related words including prime words such as "badraghe" 'seeing off' with the target being "badr" 'moon' (Nojoumian et al., 2006, pp.31-32). The results, similarly, displayed priming effects in all three conditions. These results were also suggested as further evidence for fully decompositional routes of word processing in Persian. As mentioned above, these effects can also be considered phonological, which were not mentioned to be controlled.

Based on the above-mentioned results, showing opaque, transparent, and pseudo-compound words all being decomposed to their smallest units at the earliest stages of processing (for a similar discussion, see also Boudelaa & Marslen-Wilson, 2000; Smolka et al. 2008), Shabani-Jadidi (2019) explains that in morphologically rich languages like Persian, in which morphologically complex words are not marked, semantic transparency does not play a major role in the processing of compound words. Shabani-Jadidi (2019) describes this to be different in a language like English in which those complex words that are semantically transparent (e.g., *government*) can prime their root (*govern* in this case), whereas opaque complex words (e.g., *apartment*) cannot prime their root (*apart* in this case)[9] (p.419). Smolka et al. (2008) list three main criteria for assessing the morphological richness of a language. These are firstly, how inflectional and derivational morphemes are structured in the language systems, secondly, the compounding productivity in a language, and thirdly, the ratio of semantically opaque versus transparent compounds in the language.

As for the productivity of compounding in Persian, novel words can easily be coined and will generally be transparent to native speakers. An example of this is how novel compound verbs can be created by Persian native speakers by adding a borrowed noun from another language to a verb such as *kardan* ('to do'), e.g., *date-kardan*, "date-to do", meaning 'to date someone'. As for the ratio of transparent and opaque compounds in Persian, Shabani-Jadidi (2019) uses the huge number of idiomatic compound verbs in Persian and the semantic transparency of these

[9] The fact that "apart" is not a true root in English might play a role here. Apartment is a loan word and can be considered a pseudo-derivation.

constructions as evidence in this regard (e.g., *del-dadan* "heart-to-give", meaning to 'fall in love') (p.423).

In addition to semantic transparency, there are non-linguistic factors, like frequency, that can affect the processing of complex words in a language. Different models of complex word processing predict either effects of constituent-morpheme or surface whole-word frequencies. Non-decompositional theories pay more attention to the surface frequency, whereas the decompositional hypothesis focuses more on the morpheme frequency. As mentioned before, due to the lack of consistency in the orthography of Persian compound words, it is very difficult to tag these words in a written corpus (see Ghayoomi et al. 2010 for corpus gathering problems in Persian). Consequently, there is a lack of frequency studies for Persian compound words.

(b) Studies of aphasic patients

Aphasic patients' language errors have been used to provide further support for complex word processing strategies and the structure of the mental lexicon in different languages. There are also studies that focus on the knowledge of compounds stored in the mental lexicon, using aphasic patients' data. The results of these studies show that patients tend to replace simple words with simple words and compounds with compounds. In particular, studies show that even though some patients are not able to make compound words, the knowledge of the compounds structure remains as a part of their mental lexicon (Semenza et al., 1997). Studies also revealed that structural consistency applies to the word-building rules. These studies (Semenza et al., 1997) show that aphasic patients replaced NN compounds with novel well-formed NNs, and compounds with verb–noun structures with similar novel verb-noun compounds. Therefore, the word-building rules and knowledge of the compound structure are inferred to be a part of our mental lexicon (see also Hittmair-Delazer et al., 1994).

In studies on Italian compound words, aphasic patients were substituting semantically proper constituents for the correct ones (Delazer and Semenza, 1998). The authors believe this to prove that the compound substituents are processed separately at the lemma level. As mentioned in section 1-3, lemmas are activated at the level of language production at which word processing is done subconsciously. Semantic processing triggers the activation of lemmas, which in turn activate related syntactic elements. Delazer and Semenza (1998) report that their patients made errors that show they retained the compound structure and that the substitutions they made for the targets were semantically adequate. The patient errors also revealed that word-building rules were retained. Therefore, the authors suggested that the compound structure and the constituent position are retrieved prior to the lemma access. The patient data also showed that both first and second constituents of compounds could be substituted. This suggested that the activation of both constituents works in parallel and simultaneously. The data did not show any frequency effects of the compound constituents and the patients retained first and second constituents equally. However, these studies are in conflict with findings from other studies, such as one by Rochford and Williams (1965), who used a picture naming task for testing aphasic native English speakers to investigate frequency effects. Their results showed reliable effects of only the first constituents' (the modifiers') frequency on the complex word naming accuracy.

Most of the work on Persian aphasic patients in Iran has been done by Nilipour (e.g., Nilipour 1989; 2000; Nilipour et al., 2014). There is no relevant work on compounds in Persian aphasics.

1.2 Thesis overview

In this dissertation, I examined the processing of compound words, with the core question centered around their headedness. This work was conducted on Persian, which is an Indo-European language with SOV word order. As mentioned above, the canonical order of a noun phrase in Persian is where a modifier follows the noun. The dominant form of Persian noun-noun compounds is where the modifying noun comes before the head noun. Persian allows for variable head positions in compound words and, thus, creates an opportunity to investigate if the processing effort is different for head-final (right-headed or RH) versus head-initial (left-headed or LH) compounds.

Previous psycholinguistic experiments on compound headedness in languages that allow for various head positions, such as Italian and French, report that headedness can influence the processing of compound words (see e.g., Arcara, Marelli, Buodo, & Mondini, 2014; Jarema et al., 1999). Arcara et al., (2014), presented Italian head-final (right-headed) and head-initial (left-headed) compounds as visual stimuli to investigate if different head positions in Italian result in different processing of the compounds. Their results indicate that, when decomposition is happening, processing effort for head-final (RH) compounds is higher. This result, they suggest, points to different mental representations for head-initial versus head-final compounds in Italian. El Yagoubi et al., (2008) note that Italian LH compounds are default form in the language and, diachronically, appeared earlier than RH ones. They also follow the canonical order of Italian language, which is noun + modifier. The authors mention that RH Italian compounds are more productive in contemporary Italian and are mainly derived from other languages like Latin and English.

Although, based on the factors mentioned in this chapter, Persian creates a novel opportunity for studying different factors affecting complex word recognition such as constituent order and syntactic roles, there is very little experimental work on the processing of compound words in Persian, up to this date. As mentioned in the previous sections, Shabani-Jadidi (2014, 2019) mainly worked on the processing of CPrs in Persian, with the goal of showing a fully decompositional route for compound processing in the language. Similarly, Nojoumian et al., (2006) used compound word stimuli to work on relatedness in Persian. They reported results confirming their hypothesis of a fully decompositional route for complex word processing in Persian.

Headedness in different languages may need different syntactic-semantic interpretations. Thus, the LH compounds in Persian may have different properties from those in Romance languages. Noun-noun (NN) compounds in Persian consist of a head and a modifier and are mainly right-headed, as in Germanic languages. The LH Persian compounds are more transparent in the language. They were historically shaped out of two nouns combined with the Ezafe construction, mentioned above, and the Ezafe was dropped over time, due to the frequency of use. Kahnemuyipour (2014) called this group of compounds historical compounds. Ezafe (EZ) is an unstressed vowel *-e* (*-ye* after vowels) that connects a noun to its modifier and any following modifiers, if present (the Ezafe construction will be further discussed in the following section). The primary focus of the present research was to investigate, experimentally, if and how the linguistic difference in the internal structures of the RH and LH Persian compound words affects their recognition, recall, and processing times, and, further, to present a model of how Persian compound words are stored and represented in the mental lexicon. Answering these questions can shed light on how the brain of a native Persian speaker processes competing syntactic-semantic structures in the language. More specifically, Persian compound words present an interesting case with the LH compounds being semantically more transparent while having the syntactic Ezafe structure embedded in them, which might make the processing more difficult, and the RH compounds follow the dominant order of NN compounds in Persian.

The first experiment, the second chapter of this thesis, elaborates on a database of NN compounds in Persian that was created for the purpose of this study. This database provides ratings of familiarity, imageability, and age of acquisition (AOA) by native Persian speakers providing us with information about aspects of Persian compound words that are likely to affect their accessibility in the mental lexicon. With the lack of reliable frequency data in Persian, due to its orthography, the information about these lexical variables had to be gathered from rating responses. This information can play an important role in hypothesis formation and the investigation of how certain syntactic structures can affect the recognition and processing of complex words in Persian.

The third chapter of this thesis explains data from a second experiment testing short-term memory (STM) for RH versus LH compounds in Persian. The aim of this chapter is to examine the effects of syntactic roles and order of compound word constituents on STM recall. Assuming that right headedness is the default and left-headed compounds are marked in Persian, I had two main hypotheses in this chapter. In an immediate recall task, (i) RH compounds are recalled more easily than the LH compounds because they follow the dominant pattern for NN compounds in Persian. (ii) Morphological complexity can affect the STM span negatively. This means that the LH compounds, which have Ezafe construction embedded in them and are morphologically more complex than the right-headed ones, are recalled slower than the RH ones. The results presented in this chapter inform us about how long-term knowledge in the mental lexicon supports the maintenance of compound words in the competitive environment of short-term memory.

In the fourth chapter, I address the headedness question using a priming paradigm. Here, I present the results of a primed naming task on Persian compound words to compare how the left and right constituents of Persian NN compounds, as primes, can activate the RH and LH compounds in the language. Considering the internal structure of left-headed compounds in Persian and the major role that the Ezafe construction plays in the creation of these types of compounds in the language, the two main hypotheses of this chapter were: In a primed reaction time experiment when both LH and RH compounds are primed once with their left and once with their right constituents, (i) RH compounds have a shorter reaction times than LH compounds. (ii) There is an interaction between headedness and prime condition to suggest that priming effects are different for these two types of compound words. The results of this chapter provide additional evidence on the role of certain syntactic structures of compounds in the processing and organization of these words in the mental lexicon.

To the best of my knowledge, there are no other experimental studies, up to this date, that have investigated the effects of headedness in Persian NN compounds on their short-term memory recall or naming reaction times. Surface headedness in different languages has different deep structures and these lead to different processing consequences for these languages. For example, Italian also allows for both RH and LH compounds and the results of the study done by Arcara et al., (2014) shows that processing effort for head-final (RH) compounds is higher in this language. As previously mentioned, El Yagoubi et al., (2008) noted that this difference is related to the LH compounds being the default form in the language and that, diachronically, Italian LH compounds appeared earlier than RH ones. They also interpreted these results in the light of the canonical order of this language, which is noun + modifier. The authors also mention that RH Italian compounds are more productive in contemporary Italian and are mainly derived from other languages like Latin and English. Now, I shall compare these interpretations to what can cause the different processing for LH and RH compounds in Persian, analysing their deep structure. Persian LH compounds originated from two nouns combined with Ezafe construction and the Ezafe was dropped over time, due to the frequency of use (Kahnemuyipour, 2014). Ezafe is a unique syntactic construction in Persian noun phrases. As mentioned before, this can lead to the LH compounds being more complex than the RH ones, when going through processing in the brain of a native speaker. The RH compounds in Persian, which are called the ‘true’ compounds, reflect the dominant pattern of NN compounds in the language (Kahnemuyipour, 2014). It should be noted that although the order of Persian noun phrases is where the modifier follows the noun, Kahnemuyipour (2014) proposed that the order of RH compounds in Persian (where the head follows the modifier) should be taken as the “base order” of nouns and modifiers in Persian. This compound formation is considered to be a result of a “morphological merger” (Kahnemuyipour, 2014. p.6). Having the above-mentioned points in mind, we can see how the deep structure of compounds in Persian can propose a unique case for investigation of how syntactic structure of compounds and the order of the elements in the noun phrase of languages can affect the processing of their complex words in the native speakers’ brain.

References

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. https://doi.org/10.1080/09541440340000123

Arcara, G., Marelli, M., Buodo, G., & Mondini, S. (2014). Compound headedness in the mental lexicon : An event-related potential study. *Cognitive Neuropsychology, 31*, 164–183.

Arjang, Gh. (2006). *Dastoor zabane Farsi emrooz [Modern Persian grammar]*. Tehran: Ghatreh Press

Badecker, W., & Caramazza, A. (1991). Morphological composition in the lexical output system. *Cognitive Neuropsychology*, *8*(5), 335–367.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two stage model of lexical retrieval evidence from a case of anomia with selective preservation of grammatical gender. *Cognition, 57*(2), 193–216.

Bauer, L., & Laurie, B. (1983). *English word-formation*. Cambridge university press.

Bauer, L. (2006). 21 Compounds and minor word-formation types. *The handbook of English linguistics*, 483.

Biran, M., & Friedmann, N. (2011). The representation of lexical-syntactic information : Evidence from syntactic and lexical retrieval impairments in aphasia. *CORTEX*, *48*(9), 1103–1127. https://doi.org/10.1016/j.cortex.2011.05.024

Borer, H. (2005). *In name only* (Vol. 1). Oxford University Press on Demand.

Boudelaa, S., & Marslen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing: Evidence from Modern Standard Arabic. In *ISCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes*.

Boutonnet, B., Mcclain, R., & Thierry, G. (2014). Compound words prompt arbitrary semantic associations in conceptual memory. *Frontiers in Psychology*, *5*(222), 1–7. https://doi.org/10.3389/fpsyg.2014.00222

Butterworth, B. (1983). *Lexical representation* (pp. 257-294). na.

Caramazza, A. (1997). How many levels of processing are there in lexical access ? *Cognitive Neuropsychology*, *14*(1), 177–208.

Chase, W. G., & Simon, H. A. (1973). *The mind's eye in chess*. In: William Gary Chase (ed.): Visual information processing, (pp. 215-281). Academic Press, New York.

Chomsky, N. (1992). *A Minimalist program for linguistic theory'*, draft.

Cinque, G. (2010). *The syntax of adjectives: A comparative study* (Vol. 57). MIT press.

Dabir-Moghaddam, M. (1982). Passive in Persian. *Studies in the Linguistic Sciences*, *12*(1), 63-90.

Dabir-Moghaddam, M. (2019). Typological approaches and dialects. In A. Sedighi & P. Shabani-Jadidi (eds), *The Oxford Handbook of Persian Linguistics*. pp. 52-88. Oxford: Oxford University Press.

Delazer, M., & Semenza, C. (1998). The processing of compound words : A study in aphasia. *Brain and Language*, *61*(1), 54–62.

Donalies, E. (2004). Grammatik des Deutschen im europäischen Vergleich: Kombinatorische Begriffsbildung: Teil I: Substantivkomposition.

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.

Fabb, N. (1998). Compounding. In A. Spencer & A. M. Zwicky (eds), *The handbook of morphology*. pp. 66–83. Oxford: Blackwell.

Folli, R., Harley, H., & Karimi, S. (2005). Determinants of event type in Persian complex predicates. *Lingua*, *115*(10), 1365-1401.

Foroodi-Nejad, F., & Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in PersianEnglish bilinguals. *Bilingualism*, *12*(4), 411–427. https://doi.org/10.1017/S1366728909990241

Forster, I., & Taft, M. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of verbal learning and verbal behavior, 15*(6), 607-620.

Fradin, Bernard 2009. 'IE, Romance: French', In: Lieber, R. & Štekauer, P. (eds), *The Oxford Handbook of Compounding*. pp. 417-435. Oxford University Press, Oxford.

Frommer, P. R. (1982). Post-verbal phenomena in colloquial Persian syntax.

Ghayoomi, M., Momtazi, S., & Bijankhan, M. (2010). A study of corpus development for Persian. In *International Journal on ALP*.

Ghomeshi, J. (1997). Non-projecting nouns and the ezafe: construction in Persian. *Natural Language & Linguistic Theory*, *15*(4), 729-788.

Ghomeshi, J. (2019). Other approaches to syntax. In A. Sedighi & P. Shabani-Jadidi (eds), *The Oxford Handbook of Persian Linguistics*. pp. 205-225. Oxford: Oxford University Press.

Ghonchepour, M., & Raghibdoust, S. (2012). Processing of compound nouns: Evidence from Persian-speaking transcortical aphasics. *Procedia-Social and Behavioral Sciences*, *32*, 267-274.

Giegerich, H. J. (2004). Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language & Linguistics, 8*(1), 1-24.

Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, *8*(1), 127–131. https://doi.org/10.3758/BF03196148

Greenberg, J. H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*, ed. by JH Greenberg, 73-113.

Hajati, A. K. (1977). *ke-constructions in Persian: Descriptive and theoretical aspects* (Doctoral dissertation, University of Illinois at Urbana-Champaign). Available from ProQuest Dissertations & Theses Global database.

Hlttmair-Delazer, M., Andree, B., Semenza, C., De Bleser, R., & Benke, T. (1994). Naming by German compounds. *Journal of Neurolinguistics*, *8*(1), 27-41.

Inhoff, A. W., Starr, M. S., Solomon, M., & Placke, L. (2008). Eye movements during the reading of compound words and the influence of lexeme meaning. *Memory and Cognition*, *36*(3), 675–687. https://doi.org/10.3758/MC.36.3.675

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press, USA.

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies : Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, *23*(7–8), 1191–1223. https://doi.org/10.1080/01690960802250900

Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., & Libben, G. (1999). Processing compounds: A cross-linguistic study. *Brain and Language*, *68*(1-2), 362-369

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words. *British Journal of Psychology*, *94*(2), 223–244.

Kahnemuyipour, A. (2006). Persian Ezafe construction: case, agreement or something else. In *Proceedings of the 2nd Workshop on the Persian Language and Computer. Tehran: University of Tehran*.

Kahnemuyipour, A. (2014). ScienceDirect Revisiting the Persian Ezafe construction : A roll-up movement analysis. *Lingua*, *150*, 1–24. https://doi.org/10.1016/j.lingua.2014.07.012

Kalbasi, I. (1997). *Sakhte eshteghaghiye vage dar zabane Farsi emrooz* [The derivational structure of word in modern Persian]. Tehran: Pajuheshgah Press.

Karimi, S., 1997. Persian complex verbs: idiomatic or compositional. *Lexicology* 3, 273–318.

Karimi, S., & Brame, M. (2012). A generalization concerning the EZAFE construction in Persian. *Linguistic Analysis*, *38*(1), 111-144.

Karimi, S. (2019). Generative approaches to syntax. In A. Sedighi & P. Shabani-Jadidi (eds), *The Oxford Handbook of Persian Linguistics*. pp. 161-204. Oxford: Oxford University Press.

Kingdon, R., & Scott, N. C. (1958). *English intonation practice*. Longman.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7–8), 1089–1132. https://doi.org/10.1080/01690960802193688

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876–895.

https://doi.org/10.1037/a0013484

Laudanna, A., Badeckera, W., & Caramazza, A. (1992). Processing inflectional and derivational morphology. *Journal of Memory and Language*, *31*(3), 333–348.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–75.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language*, *61*(1), 30–44. https://doi.org/10.1006/brln.1997.1876

Libben, G. (2014). The nature of compounds : A psychocentric perspective. *Cognitive Neuropsychology, 31*, 8–25.

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

Libben, G., & Jarema, G. (Eds.). (2006). *The representation and processing of compound words*. Oxford University Press on Demand.

Lieber, R. (2004). *Morphology and lexical semantics* (Vol. 104). Cambridge University Press.

Lieber, R., & Stekauer, P. (Eds.). (2009). *The Oxford handbook of compounding*. Oxford University Press.

Mahmoodi-Bakhtiari, B. (2019). Morphology. In A. Sedighi & P. Shabani-Jadidi (eds), *The Oxford Handbook of Persian Linguistics*. pp. 273-299. Oxford: Oxford University Press.

Mahoozi, M. (2006). *Ketab Farsi omoomi [Persian grammar]*. Tehran: Zavar.

Matzen, L. E., & Benjamin, A. S. (2009). Remembering words not presented in sentences: How study context changes patterns of false memories. *Memory & Cognition*, *37*(1), 52-64.

Marantz, A. (1997). No escape from syntax: Don't try morphological analysis in the privacy of your own lexicon. *University of Pennsylvania working papers in linguistics*, *4*(2), 14.

Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds : Modulation of headedness and semantic transparency. *Journal of Memory and Language*, *66*(4), 644–664. https://doi.org/10.1016/j.jml.2012.01.003

Marchand, H. (1965). On the analysis of substantive compounds and suffixal derivatives not containing a verbal element. *Indogermanische Forschungen, 70*, 117.

Marchand, H. (1969). *The categories and types of present-day English word-formation: A synchronic-diachronic approach*. Beck.

Maury, S. (2015). Differential recall of derived and inflected word forms in working memory: examining the role of morphological information in simple and complex working memory tasks. *Frontiers in human neuroscience, 8,* 1064.

Megerdoomian, K. (2001). Event structure and complex predicates in Persian. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *46*(1-2), 97-125.

Neef, M. (2009). IE, Germanic: German. In: Lieber, R. & Štekauer, P. (eds), *The Oxford Handbook of Compounding*. pp. 386-400. Oxford University Press, Oxford University Press.

Németh, D., Ivády, R., Guida, A., Miháltz, M., Peckham, D., Krajcsi, A., & Pléh, C. (2011). The effect of morphological complexity on short-term memory capacity. *Acta Linguistica Hungarica*, *58*(1), 85-107.

Nilipour, R. (1989). Task-specific agrammatism in a Farsi-English bilingual patient. *Journal of Neurolinguistics*, *4*(2), 243-253.

Nilipour, R. (2000). Agrammatic language: Two cases from Persian. *Aphasiology*, *14*(12), 1205-1242.

Nilipour, R., Pourshahbaz, A., & Ghoreyshi, Z. S. (2014). Reliability and validity of bedside version of Persian WAB (P-WAB-1). *Basic and clinical neuroscience*, *5*(4), 253.

Nojoumian, P., Shabani-Jadidi, P., & Tangestanifar, A. (2006). Morphological priming of multi-morphemic Persian words in the Mental Lexicon. In *The second workshop on Persian Language and Computer* (pp. 29-38).

Nojoumian, P., *Towards the development of an automatic diacritizer for the Persian orthography based on the Xerox Finite State Transducer* (Doctoral Dissertation, University of Ottawa). Available from ProQuest Dissertations & Theses Global database.

Olsen, S. (2001). Copulative compounds: a closer look at the interface between syntax and morphology. In *Yearbook of Morphology 2000* (pp. 279-320). Springer, Dordrecht.

Pinker, S. (1989). Learning, development, and conceptual change. *Learnability and cognition*: The acquisition of argument structure.

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), 820–833.

Ralli, A. (2013). Compounding and its locus of realization: Evidence from Greek and Turkish. *Word Structure*, *6*(2), 181-200.

Rochford, G. E. R. A. R. D., & Williams, M. O. Y. R. A. (1965). Studies in the development and breakdown of the use of names. IV. The effects of word frequency. *Journal of neurology, neurosurgery, and psychiatry*, *28*(5), 407.

Samiian, V. (1994). The Ezafe construction: Some implications for the theory of X-bar syntax. *Persian Studies in North America*, 17-41.

Sampson, R. (1980). Stress in English N+ N phrases: A further complicating factor.

Samvelian, P. (2007). A (phrasal) affix analysis of the Persian Ezafe. *Journal of Linguistics*, *43*(3), 605-645.

Samvelian, P., & Faghiri, P. (2014). Persian complex predicates: How compositional are they?.

Sandra, D. (1990). On the representation and processing of compound words : Automatic access to constituent morphemes does not occur On the Representation and Processing of Compound Words : Automatic Access Does Not Occur. *The Quarterly Journal of Experimental Psychology*, *42*(3), 529–567. https://doi.org/10.1080/14640749008401236

Scalise, S., & Bisetto, A. (2009). The classification of compounds. In *The Oxford handbook of compounding*.

Scalise, S., & Vogel, I. (Eds.). (2010). *Cross-disciplinary issues in compounding* (Vol. 311). John Benjamins Publishing.

Schlücker, B., & Plag, I. (2011). Compound or phrase? Analogy in naming. *Lingua*, *121*(9), 1539-1551.

Schmidtke, D., Kuperman, V., Gagné, C. L., & Spalding, T. L. (2015). Competition between conceptual relations affects compound recognition : the role of entropy. https://doi.org/10.3758/s13423-015-0926-0

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. *Morphological aspects of language processing*, *2*, 257-294.

Sedighi, A., & Shabani-Jadidi, P. (Eds.). (2019). *The Oxford handbook of Persian linguistics*. Oxford University Press.

Selkirk, E. O. (1982). The syntax of words (Linguistic Inquiry Monograph 7). *Cambridge, MA*.

Semenza, C., Arcara, G., Facchini, S., Meneghello, F., Ferraro, M., Passarini, L., ... & Mondini, S. (2011). Reading compounds in neglect dyslexia: The headedness effect. *Neuropsychologia*, *49*(11), 3116-3120.

Semenza, C., & Luzzatti, C. (2014). Combining words in the brain : The processing of compound words Introduction to the special issue, *31*, 1–7.

Service, E., & Tujulin, A. (2002). Recall of morphologically complex forms is affected by Mmmory task but not dyslexia. *Brain and Language*, *81*(1_3), 42–54. https://doi.org/10.1006/brln.2001.2505

Shabani-Jadidi, P. (2014). *Processing compound verbs in Persian: a psycholinguistic approach to complex predicates*. p. 189. Leiden University Press (LUP).

Shabani-Jadidi, P. (2019) Psycholinguistics. In A. Sedighi & P. Shabani-Jadidi (eds), *The Oxford Handbook of Persian Linguistics*. pp. 411-435. Oxford: Oxford University Press.

Shariat, M. J. (2005). *Dastoor zabane Farsi* [Persian grammar 1]. Tehran: Asatir Press.

Smolka, E., Komlosi, S., & Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, *24*(3), 337-375.

Spencer, A. (2003). Periphrastic paradigms in Bulgarian. *Syntactic structures and morphological information*, *7*, 249-282.

Shabani-Jadidi, P., & Sedighi, A. (2019). Teaching Persian to speakers of other languages. In A. Sedighi & P. Shabani-Jadidi (eds), *The Oxford Handbook of Persian Linguistics*. pp. 388-408. Oxford: Oxford University Press.

Štekauer, P. (2005). Onomasiological approach to word-formation. In *Handbook of word-formation* (pp. 207-232). Springer, Dordrecht.

Štekauer, P. (2016). Fundamental principles of an onomasiological theory of English word-formation. *A Recollection of 11 Years of Onomasiology Online (2000-2010) All Articles Re-Collected*, 15.

Štekauer, P. (2016). Compounding from an onomasiological perspective. *The semantics of compounding*, 54-68.

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

Ten Hacken, P. (2010). Synthetic and exocentric compounds in a Parallel Architecture. *New impulses in word-formation*, 233-251.

Ten Hacken, P. (Ed.). (2016). *The semantics of compounding*. Cambridge University Press.

Wälchli, Z. (2016). *Recall of compound words in simple and complex span tasks*. (Master's thesis, McMaster University). Available from ProQuest Dissertations & Theses Global database.

2 Experiment One: Ratings of familiarity, imageability and age of acquisition of 149 Persian noun-noun compound words as a function of their headedness

Abstract

Several variables can affect the recognition, recall, and naming of compound words. Some of the variables, that have been found to have an effect on the processing of words in general, and compounds in particular, include age of acquisition (AOA), familiarity, and imageability (cf. Juhasz, Lai, & Woodcock, 2015; Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012; Stadthagen-Gonzalez & Davis, 2006). There are presently data on these variables from English and only a few other languages. To allow more cross-linguistic research, there is a need for such information on languages other than English, and particularly on ones with a different syntactic make up. Data on lexical variables in more languages would allow us to consider the effects of these background variables when studying the role of certain syntactic structures in complex-word recognition and processing in different languages. Persian is a language of interest in the study of compound word processing. It is an Indo-European language that allows variable head positions for compound words, unlike English that prefers right-headed compounds. The present study introduces a database with AOA, familiarity, and imageability ratings for 149 Persian compound noun-noun words and non-words (constructed of real word constituents) collected from 102 participants. The database includes information on the right and left constituents of these compounds, their headedness, meaning in English, length in letters and phonemes, and Google frequencies. There is no reliable frequency data on Persian compound words. Since there were no studies on these lexical variables in Persian compounds, there was a need for collecting such data to allow comparative research with other languages. Also, this database was created as a way to control for frequency in the data analysis of the experiments presented in the next chapters. The data analysis of this chapter indicated that familiarity, imageability, and AOA are highly inter-correlated in both right-headed and left-headed (LH) compounds. The analysis of AOA separated by headedness revealed that words that left-headed compounds were rated to have been learnt earlier in life.

2.1 Introduction

As mentioned in the previous chapter, basic definitions of compounds range from a denial of their lexical status, “compound words do not exist” (Spencer, 2003), to broad inclusion, “any noun-noun construction being considered as a compound” (Olsen, 2001). Compound words as complex lexical units are the results of one of the fundamental strategies for information packaging in our mental lexicon. Jackendoff (2002) highlighted the importance of compounds, suggesting that they may be seen as protolinguistic fossils from which more complex linguistic structures have developed. Libben (2014) noted that although the concept of compound word may seem very simple, insight into the neurological interactions underlying their comprehension is valuable for understanding lexical and morphological processing in general, and cognitive representation of human language at large. For the past four decades, there has been a debate on how compound words are represented and processed in the brain.

There is already a large body of data on morphological processing of English compound words. Forster and Taft (1976) were the first to propose that the frequency of the first lexeme of the compound affects lexical decision times. Since then, many researchers have conducted word recognition experiments on compounds to see whether they are accessed through their constituents or as whole words. This research includes constituent-priming tasks to examine if the constituent lexeme frequency has a role in lexical decision performance (e.g., Andrews, Miller, & Rayner, 2004; Libben, Gibson, Yoon, & Sandra, 2003). A great number of studies in English display differences in processing times when the constituent lexemes’ frequencies were manipulated while the whole word’s frequency was controlled. This suggests that the constituent lexemes are accessed during compound word recognition (cf. Andrews, Miller, & Rayner, 2004 ; Inhoff, Starr, Solomon, & Placke, 2008; Juhasz, Starr, Inhoff, & Placke, 2003). Similar results have been reported from research on languages like Finnish, in which compounding is very productive (e.g., Pollatsek, Hyönä, & Bertram, 2000). Other recent studies have looked at how morphological families and the conceptual relations between the constituents of compounds affect the organization of compounds in the mental lexicon (e.g., Kuperman, Bertram, & Baayen, 2008; Kuperman, Schreuder, Bertram, & Baayen, 2009; Schmidtke, Kuperman, Gagné, & Spalding, 2015).

Based on the existing studies, we know that several variables can affect recognition, recall, and naming of compound words. Some of these variables are the constituents’ lengths, frequencies, neighbourhood sizes, and mutual conceptual relations. The effects of these variables have mainly been studied based on the information available from large corpora of English text, tagged using specialized statistical software. However, the literature also reports psychological variables as significant predictors of word recognition performance. These variables include rated AOA, familiarity, and imageability. The existing databases for these norms are also mainly for English (see, e.g., Stadthagen-Gonzalez & Davis, 2006; Juhasz, Lai, & Woodcock, 2015), which highlights the need for similar subjective ratings for languages other than English. The need for rated lexical information about compound words can be considered even more pressing for languages with syntactic structures different from English, as the availability of psychological word information in these languages would allow the study of how syntactic structures can play a role in complex word processing.

One of the structures that varies across different languages is compound headedness. Headedness refers to the order of the syntactic roles (modifier and head) in noun-noun (NN) compound words. English favors right headedness, while in Hebrew and Vietnamese, for example, the head always precedes the modifier (Fabb, 1998 for Vietnamese). In Romance languages, such as French and Italian, the position of the head can vary (Libben and Jarema, 2006, for French; Semenza & Luzzatti, 2014, for Italian). Interestingly, Persian, an Indo-European language with SOV word order, also allows variable head positions (Kalbasi, 1997; Shariat, 2005; Foroodi-Nejad & Paradis, 2009). Such variability in headedness between languages raises the question of whether it is the functionally determined headedness or simply the surface order of constituents that cause differences in processing compound words. In particular, languages that allow left heads, make it possible to separate effects of constituent position (usually first vs. second) from those of headedness (head vs. modifier).

There are different theories on the processing of compound words and whether they are processed in the mind as whole words or decomposed into their constituents (cf. e.g., Forster & Taft, 1976, for a full parsing model; Butterworth, 1983, for a whole-word listing model; Schreuder & Baayen, 1995 for a dual-route model; and Kuperman et al., 2009, for a multiroute interactive model). The decomposition hypothesis has led to another argument in the literature on compound words: whether the first constituent of the compound is always processed first or whether it is the syntactic make-up of the compound, i.e. their head and modifier, that plays a more important role than constituent position. As mentioned above, in a language like English, the second constituent of a NN compound is always the head. Therefore, conducting experiments on languages with variable head positions allows researchers to distinguish between the effects of surface order and syntactic role. In addition to head position, semantic transparency can play an important role in whether the compound is comprehended through its constituents or if it is accessed as a whole word representation (Libben et al., 2003). For example, a compound like *orange juice* can be comprehended through its constituents more easily than a word like *hotdog*. Therefore, semantically opaque (*hotdog*) words are more likely to be stored as a whole in the mental lexicon.

Persian allows us to investigate whether there are differences in recognition, recall, or naming speed and accuracy for head-initial (left-headed, LH) versus head-final (right-headed, RH) compounds. Such processing effects allow us to further develop linguistic theory, including its fundamental combinatorial features. Frommer (1982) suggested that the canonical word order of formal standard Persian is "(subject) + (adverbs) + (direct object) + (indirect object) +verb" (pp. 30-31). As mentioned in the previous chapter, modifying adjectives in Persian always follow their head noun (Dabir-Moghaddam, 2019). An example of this is "*zæn-e mehræban*": "woman-EZ kind", which means "a kind woman" (p.57). Although the canonical order is noun + modifier in phrases, Persian NN compounds consisting of a head and a modifier are mainly right-headed. The Persian compounds that are left-headed have been proposed to form a diachronically distinguishable subgroup (Kahnemuyipour, 2014). They have been argued to have been historically shaped out of a phrase, i.e. two nouns combined with a so-called Ezafe construction. Ezafe is an unstressed vowel *-e* (*-ye* after vowels) appearing between a noun and its modifier, as well as any subsequent modifiers, if present. Although the LH compounds can only be identified as compounds when they are NOT pronounced as Ezafe structures, these so called historical compounds (Kahnemuyipour, 2014) are semantically more transparent than RH compounds in modern Persian. The availability of psycholinguistically important information about Persian compound words will allow controlling for this in investigations of the processing of compound words in Persian. In the following section, I present a brief literature review on the main lexical variables examined in this study.

2.2 Familiarity

An important factor in the whole-word access of compound words is word frequency. Libben (2006) has observed that the relatively low frequency of English compound words makes it difficult for researchers to conduct experiments with frequency manipulations in this language.

There are other languages, such as Finnish and Dutch, that do not have this problem since compounds are high in frequency in these languages (cf. Kuperman et al., 2008, for Finnish; van Jaarsveld & Rattink, 1988, for Dutch). Another complicating factor in some languages, like Persian, is a lack of consistency in the orthography, i.e., the inclusion of spaces separating the constituents of compound words. This results in problems in assessing the corpus frequency of these words. Megerdoomian (2019) mentions that one major problem in the analysis of Persian language is the existence of a large number of multi-word expressions (MWE). As a subcategory of MWEs, compound words can have their elements separated or attached in writing. Between the two constituents of NN compounds there can be a full, a half, or no space at all. This makes it difficult to do frequency counts in written corpora. However, Gernsbacher (1984) has argued that familiarity ratings are a better predictor of word processing performance than word frequency, especially for words that are not highly frequent. Thus, subjective familiarity ratings are very much needed in the study of compound word recognition (cf., Juhasz et al., 2015; Juhasz & Rayner, 2003; Williams & Morris, 2004). Familiarity questionnaires typically ask participants to rate (usually on a 1–7 scale) how often they have come in contact with certain words. These ratings can be considered a way to evaluate the frequency of individual exposure to the words (Stadthagen-Gonzalez & Davis, 2006). The present study reports such ratings for Persian NN compounds.

2.3 Imageability

A number of studies (see for example Bird, Franklin, & Howard, 2001) show that the concreteness of a word affects its ease of processing. Imageability can be considered a semantic variable and is very similar to concreteness. Imageability assesses how easy it is to produce a mental image of a word. This measure can be used to demonstrate the effects of meaning on word processing and memory. According to Paivio's (1971) dual-coding theory, not all words can arouse both a verbal representation and a nonverbal image. Some words can only be represented through their verbal characteristics, such as their relationship to other words. The ones with dual, both verbal and strong imagery, representations, have an advantage during processing. It is reported that imageability is a good predictor of performance in word recognition experiments, such as word naming tasks (Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004; Cortese & Schock, 2013). Kuperman (2013) has noted that most semantic variables that have been reported to play a role in compound word processing can be considered as relational. This refers to the special relationship between the two constituents of the compounds. This matter has also been discussed under the label *conceptual relations* in compound words (cf. Schmidtke et al., 2015 for the effects of conceptual relations on compound recognition), and is related to the concept of semantic transparency. A compound word can be considered semantically transparent when the meaning of the whole compound word is in line with the meanings of its constituents (e.g., doorbell) (Pollatsek & Hyönä, 2005). Semantic transparency is a semantic variable that can affect compound word recognition in general and it can also affect imageability in a positive way, i.e., more transparent compounds tend to be more imageable. Imageability ratings are elicited by asking participants how fast, or easily, certain words can arouse an image in their mind. This is usually evaluated using a 1–7 scale. Kuperman (2013) found that whole-word imageability, but not constituent-lexeme imageability, is a reliable predictor of lexical decision times for compounds (cf. also Balota et al., 2004; Cortese & Khanna, 2007; Cortese & Schock, 2013, for imageability and lexical decision tasks).

As mentioned earlier, LH compounds in Persian are semantically more transparent. Consider an LH Persian compound such as "*aab sib*" 'water apple' which means 'apple juice' and has two transparent constituents, which make the whole compound more transparent than an RH compound word like "*khær mægæs*" 'donkey fly' which means 'horsefly' or a very large fly. The opacity in this RH compound comes from the meaning of its first constituent '*khær*' not being related transparently to its original meaning. Based on greater semantic transparency, I expected LH compounds in Persian to be more imageable than RH ones.

2.4 AOA

Another variable that has been reported to influence word recognition tasks is the age of acquisition of words. A general assumption is that words acquired earlier in life have stronger semantic representations, or more connections to related concepts in the mental lexicon, and can be processed faster than the ones learned later (cf., Steyvers & Tenenbaum, 2005). Another possible hypothesis explains how words acquired early have a network advantage in terms of greater plasticity due to early exposure (Ellis & Lambon Ralph, 2000). According to this hypothesis, AOA can be related to orthographic, phonological, and semantic representations of words since it can affect all levels of the mental lexicon (see also, Juhasz et al., 2015). Related studies have shown that AOA can significantly predict performance in memory and word processing tasks (Hirsh & Funnell, 1995; Juhasz & Rayner, 2003; Turner, 1998) such as lexical decision and word naming times (see, Cortese & Khanna, 2007, for monosyllabic words; Cortese

& Schock, 2013, for disyllabic words), such that earlier acquired words allow faster responses. AOA assessment instruments ask adults to estimate the age at which they think they learned a certain word for the first time. This can be done on a rating scale, or AOA can be estimated through asking participants to report the actual age at which they think a certain word was learned for the first time (cf. for example Kuperman, Stadthagen-Gonzalez, & Brysbaert, 2012). Since there are no studies of the AOA of Persian compounds, there was a need for collecting such data to allow comparative research with other languages.

The present study aimed to develop a database of three subjective experience dimensions of language users, reporting variables that have a role in speed and accuracy of complex word recognition. More specifically, I provide ratings of familiarity, imageability, and AOA through data collected online. Participants were asked to rate each of the compound words in my database based on questions like “how old do you think you were when you first heard this word?”. Such a database was lacking in Persian.

The database presented here also includes information on compound headedness, the whole word and constituent lengths in letters and phonemes, and Google frequency for whole words and their constituents. The Google frequency data are presented as the best estimates of printed form frequencies, given the above-mentioned problems in identifying compound words with variable orthographic form variants. After discussing the characteristics of my subjective ratings, I will present some reaction time data to demonstrate how my database can benefit researchers in correlation studies and in the selection of stimuli for experiments.

2.5 Research Methodology

2.5.1 Participants

Ratings were collected from 102 participants recruited through a personal network of acquaintances in Iran and across the world. They were asked to follow a link that was sent to them through an email. They were then invited to complete a consent form and complete a McMaster online Lime Survey. Since many participants were living in Iran and would have had confidentiality concerns, no demographic data was collected. In each online survey (Imageability, Familiarity, AOA), 149 of all the NN items (including 31 non-words) were rated by 34 participants. Participants were all native speakers of Persian. There was no compensation for taking part in the surveys and participants were all volunteers, who consented to the online study. The study received ethics clearance from McMaster Research Ethics Board.

2.5.2 The database

I first created a corpus of 200 NN compound words. These words were collected through my introspection as a native speaker, my discussions with Dr. Arsalan Kahnemuyipour, and from the list of words that I found in an online Persian Wiktionary. The database included 130 RH and 70 LH compounds, written both in the original Persian (Arabic) and English alphabetical letters. My corpus also includes the compounds' translations into English, their constituents' literal meanings in English, their length in terms of number of letters and phonemes in Persian, headedness information, and a Google frequency count separately for the whole words and their constituents. Archaic words and ones with a very low frequency in Persian were excluded. The compounds ranged in length from five to 12 letters. The first constituent length ranged from two to nine letters. The second constituent length ranged from two to seven letters. To the best of my knowledge, up to this day there is no other such corpus in the Persian language for NN compound words. A total of 118 words were selected from this corpus to be rated by the (online) participants. The list also included 31 non-words, consisting of non-existent combinations of existing constituent words. This resulted in 149 compound items each rated by 34 participants for each one of the three assessed indicators of word experience: familiarity, imageability, and AOA. The surveys were created with McMaster University's template on the Lime Survey platform.

2.5.3 Procedure

Ratings were collected online over the course of a year. Participants were first provided with a preamble statement, letter of information, and instructions before proceeding to the surveys' first pages. Surveys didn't last more than 30 minutes each and were presented on 6 pages with the maximum of 25 words to be rated on each page. The instructions were adopted from Stadthagen-Gonzalez & Davis (2006) and were adjusted for each survey based on its nature. For familiarity, raters were asked to evaluate each word based on a 7-point Likert scale, where 1 corresponded to words that the participants had never seen and 7 to those words that had been seen very often (nearly every day). For imageability, participants were asked to rate the words based on how easily the words evoked mental images for them. Here again, the ratings were on a 7-point Likert scale with 1 referring to low imageable words and 7 to high imageable ones. For the AOA survey, participants were asked to type the age in years at which they thought they had acquired the word for the first time.

2.6 Results and discussion

All the rated values were moved to an Excel sheet for further data processing. *Table 2-1* shows group descriptive statistics for imageability, familiarity, and AOA as maximums, minimums, means, and medians for the rated variables. *Table 2-2* summarizes the means, standard deviations, and standard errors of the variables separated by compound headedness. I then present the characteristics of familiarity, imageability, AOA, and Google frequencies. This includes analysis of densities of the variables followed by independent t-tests. Then, I briefly discuss the correlations between the rated variables and the Google frequencies of the whole words, left constituents, and right constituents separated by headedness (*Table 2-4*). Finally, I discuss the descriptives of the non-words included in this study (*Table 2-5*).

2.6.1 Characteristics of 149 Persian NN compounds

Familiarity. Ratings of familiarity, as mentioned, can be used as a way to measure subjective frequency. In Persian, due to a lack of consistency in the orthography of compound words, assessing the corpus frequency of these words is very difficult. Thus, having access to familiarity ratings is of even more importance for this language. The averages for familiarity ratings ranged from 1.88 to 6.82 (see *Tables 2-1* and *2-2*). The median, however, falls on 5.82, higher than the scale midpoint of 4. This suggests that participants were familiar with most of the presented compound words. *Figure 2-1* shows the density distribution of the familiarity data. A correlative analysis relating familiarity and the other two studied dimensions showed high correlations between familiarity and imageability ratings, r = .816 (see *Figure 2-2 and Table 2-3*), as well as between familiarity and AOA ratings, r= .637 (see *Table 2-3*). This means that those compounds that were more familiar could evoke more readily mental images and were reported to have been learnt earlier in life. Results also suggest that participants tended to rate the words as either familiar or unfamiliar rather than halfway between the two extremes (see *Figures 2-1 & 2-2*). The analysis of differences in familiarity ratings based on compound headedness did not show significance (see *Table 2-4*).

Imageability. The ratings for imageability ranged from 2.41 to 7, with a median of 6.65, showing that participants found most of the compounds highly imageable (see *Tables 2-1* and *2-2*). *Figure*

2-3 shows the density of the data. As discussed above, there were strong correlations between familiarity and imageability with *r*= .816 (see *Figure 2-2 and Table 2-3*). The correlations between imageability and AOA were also strong with *r*= .727. This indicates that those compounds that aroused mental images more readily were rated as learnt earlier in life. As can be seen in *Table 2-4*, the analysis of imageability based on compound headedness did not show significance.

Table 2-1 Descriptive statistics for the 3 variables calculated over all rated items

	Imageability	Familiarity	AOA
N	118	118	118
Mean	6.40	5.56	10.6
Median	6.65	5.82	10.5
Minimum	2.41	1.88	5.41
Maximum	7.00	6.82	23.1

N =number. AOA=age of acquisition

Table 2-2 Ratings separated by headedness calculated over all rated items

	Headedness	N	Mean	SD	SEM
Familiarity	left	59	5.54	0.97	0.13
	right	59	5.57	1.08	0.14
Imageability	left	59	6.47	0.92	0.12
	right	59	6.32	0.80	0.10
AOA	left	59	9.94	3.50	0.46
	right	59	11.3	3.61	0.47

N =number. AOA=age of acquisition SD =standard deviation. SEM =standard error of mean

AOA. As mentioned in the introduction, AOA has been reported to have an effect on word processing tasks based on the assumption that words acquired earlier in life have stronger semantic representations to related concepts in the mental lexicon and can therefore be processed faster than the ones learned later. Databases with AOA information are needed in different languages to help experimental researchers in stimulus selection for word processing and production studies. *Table 2.1* shows the minimum average over participants for the age, at which participants reported to have learnt the compounds for the first time, to be 5.41 and the maximum to be 23.1 (see *Table 2-*

2 for the analysis separated by headedness). The lower age limit is set by the ability of adult participants to remember events in their early childhood. The upper limit is set by the age of the responding participants. *Figure 2-5* shows the frequency distribution of my AOA data. As discuss-

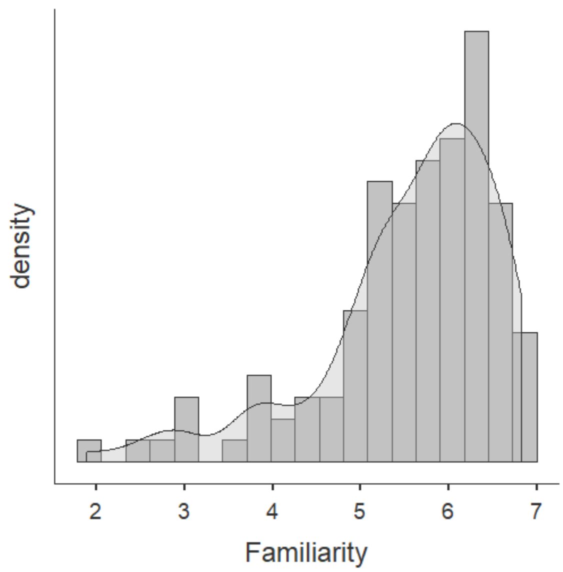

Figure 2-1 Familiarity density distribution

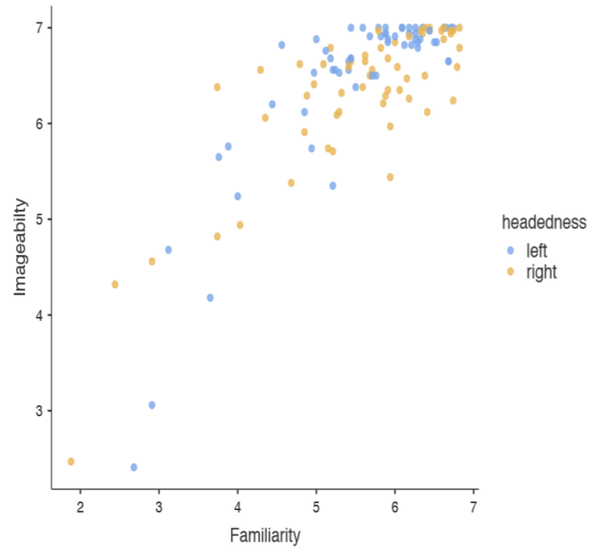

Figure 2-2 Scatterplot of imageability ratings as function of familiarity ratings

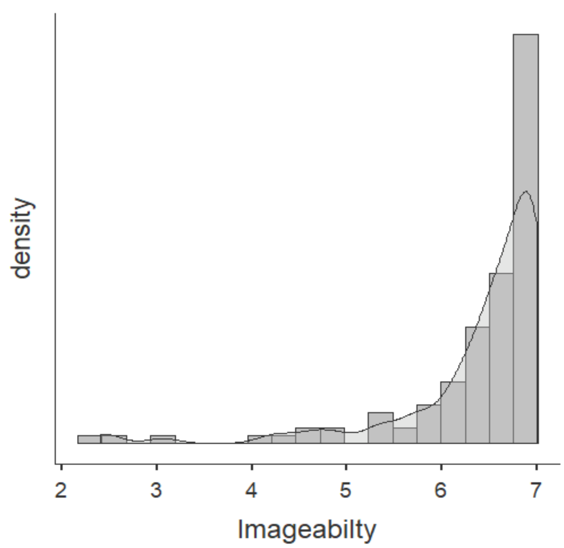

Figure 2-3 Imageability density distribution

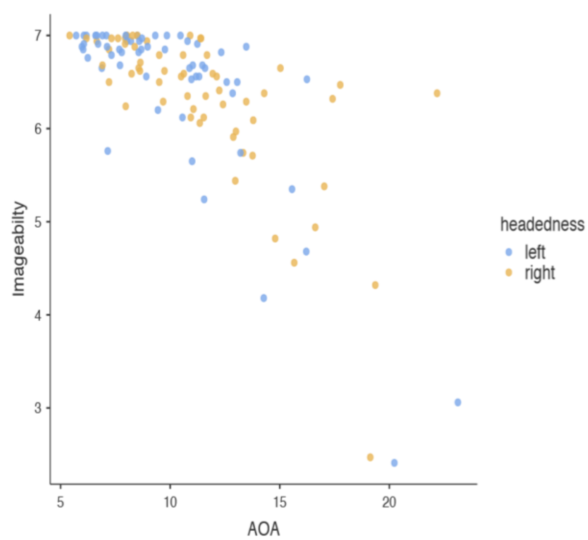

Figure 2-4 Scatter plot of AOA ratings as a function of familiarity ratings

-ed above, AOA had high correlations with both familiarity (*r*= .637) and imageability (*r*= .727), suggesting that compounds which were learnt earlier in life were both more familiar and could induce mental images more readily *(*see *figure 2-4* and *Table 2-3)*. The analysis of AOA separated by headedness shows that the LH compounds were rated to have been learnt earlier in life (cf. *Tables 2-2 and 2-4*). This can be explained by the fact that the left-headed compounds in Persian are considered to be the ones that were historically shaped out of two nouns combined with the Ezafe construction, and that they are semantically more transparent for native Persian speakers (LH compounds have a closer relationship with noun phrases in Persian where each word contributes to the meaning).

Google Frequencies As mentioned before, in Persian, assessing the corpus frequency of compound words is very difficult due to a lack of consistency in the orthography of compound words. As an approximation, I included Google frequencies of the compounds in my database.

Table 2-3 Correlation analysis of familiarity, imageability, and AOA

		Familiarity	Imageability	AOA
Familiarity	Pearson's r	—	0.816	-0.637
	p-value	—	< .001	< .001
	95% CI Upper	—	0.868	-0.516
	95% CI Lower	—	0.745	-0.733
	Spearman's rho	—	0.672	-0.489
	p-value	—	< .001	< .001
Imageability	Pearson's r		—	-0.727
	p-value		—	< .001
	95% CI Upper		—	-0.629
	95% CI Lower		—	-0.802
	Spearman's rho		—	-0.708
	p-value		—	< .001

AOA=age of acquisition

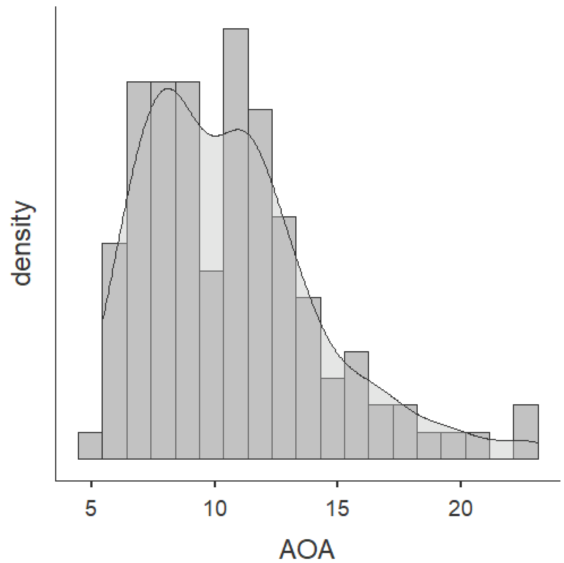

Figure 2-5 AOA Density

Table 2-4 Differences between LH and RH compound words for the three rated variables

		Statistic	df	p	Direction
Familiarity	Student's t	0.16	116	0 .871	n.s.
AOA	Student's t	-2.122	116	0.036	LH < RH
Imageability	Student's t	0.937	116	0.351	n.s.

AOA=age of acquisition. n.s.=not significant.

Table 2-5 shows the descriptive statistics for whole words and each constituents' frequencies separated by headedness. As can be seen in *Figures 2-6* and *2-7*, the frequency of the left constituent is higher in LH compounds than in their right-headed counterparts and the right-constituents' frequency is higher in RH compounds than the left-headed ones. The whole word frequency is higher for RH compounds than LH ones. I also looked at the correlations of whole word frequencies, right constituent frequencies, and left constituent frequencies with the three rated variables, with headedness added as a factor. The analysis did not show significance for any of the rated variables with Google frequencies. It should be noted that the existence of Ezafe structure in Persian, which is usually not written in the orthographic form, makes it impossible to distinguish the LH compounds from constructions of two nouns attached with Ezafe in written

Table 2-5 Descriptive statistics for Google frequencies of the compounds in this study

	Headedness	Google frequency of whole word	Google frequency of left const.	Google frequency of right const.
N	left	59	59	59
	right	59	59	59
Mean	left	340907	2.99e+8	5.77e+7
	right	1.74e+7	1.87e+8	1.87e+8
SEM	left	41868	1.13e+8	1.64e+7
	right	8.30e+6	8.53e+7	6.41e+7
Median	left	348000	6.15e+7	3.35e+7
	right	481000	3.69e+7	1.19e+8
SD	left	321593	8.68e+8	1.26e+8
	right	6.e+7	6.55e+8	4.93e+8

N =number of all items. SD =standard deviation. SEM =standard error of mean

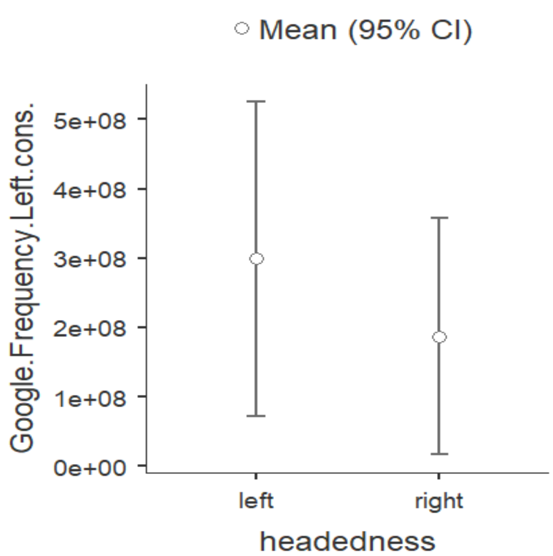

Figure 2-6 Left constituent frequency by headedness

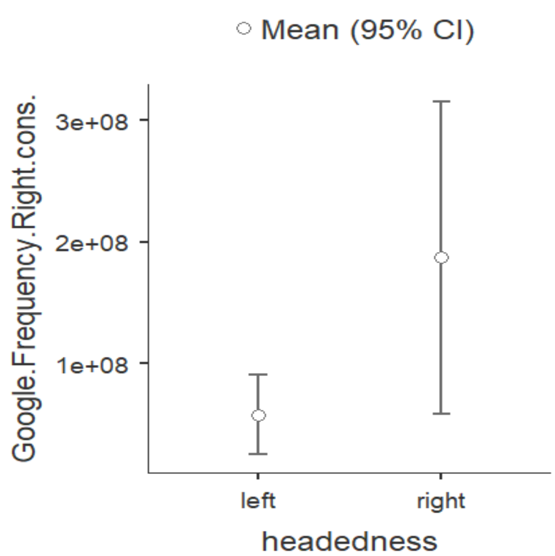

Figure 2-7 Right constituent frequency by headedness

materials. This applies to my Google frequency check or any corpus analysis, making the frequency data unreliable.

Non-words *Table 2-6* shows the descriptives for the non-words used in this study. It is interesting to observe that these items were rated by our participants to be familiar and to have been learnt at a certain age. These results can indicate that participants could falsely rate these words or that they

Table 2-6 Ratings of non-words separated by headedness calculated over all rated items

	Headedness	N	Mean	SD	SEM
Familiarity	left	15	2.51	0.936	0.242
	right	16	2.21	0.402	0.101
Imageability	left	15	4.10	1.32	0.340
	right	16	3.46	0.814	0.204
AOA	left	15	15.8	5.63	1.45
	right	16	17.4	2.16	0.540

N =number. AOA=age of acquisition SD =standard deviation. SEM =standard error of mean

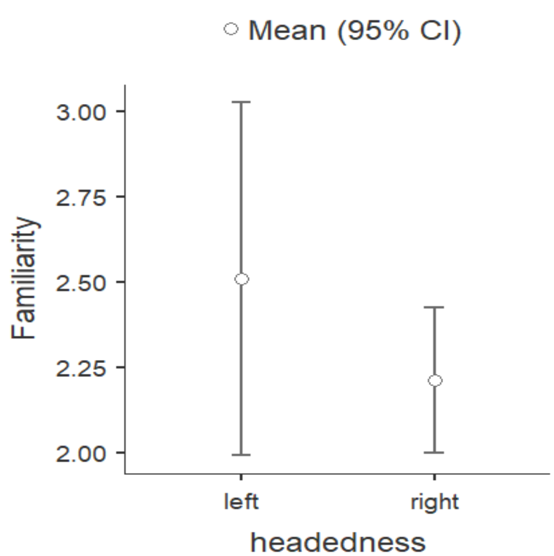

Figure 2-8 Familiarity of non-words based on headedness

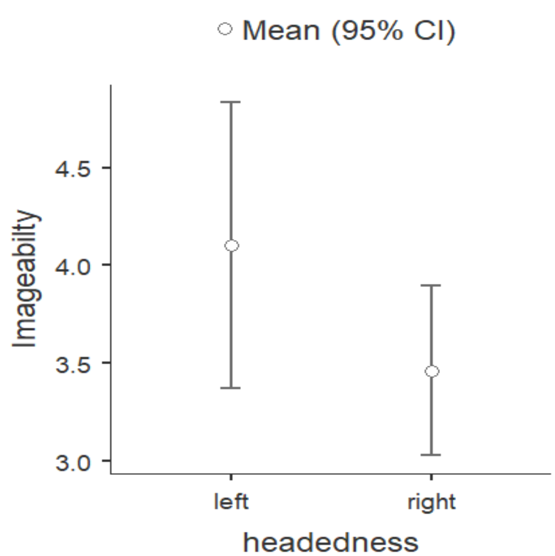

Figure 2-9 Imageability of non-words based on headedness

rated non-words due to a social pressure, i.e., they were afraid to be judged if they don't recognize the words, although they were informed that the participation was anonymous. Another reason might be that our non-words were similar to the real words since they were made up of real word constituents (e.g., ***halghe tavallod***, literally *'ring birthday',* meaning *"birthday ring"* is a non-word made of two real words *halghe* meaning *"ring"* and *tavallod* meaning *"birthday"* and is very similar to the compound ***halghe ezdevaaj*** literally *'ring marriage',* meaning *"wedding ring").* *Figures 2-8, 2-9, and 2-10* show the analysis of non-words for familiarity, imageability, and AOA separated by headedness.

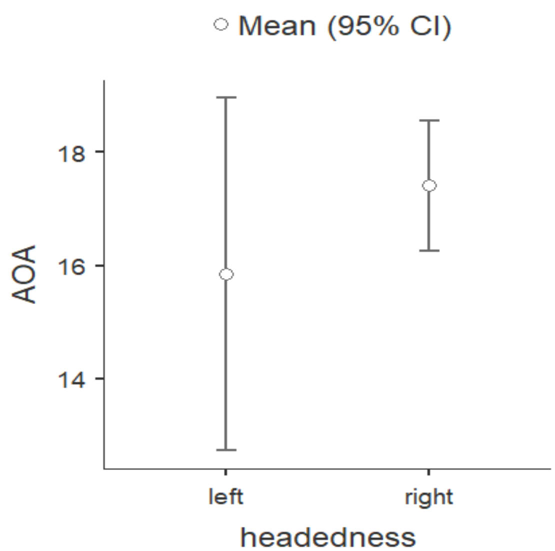

Figure 2-10 AOA of non-words based on headedness

2.7 Conclusion

This chapter reports familiarity, imageability, and age of acquisition ratings for 149 Persian noun-noun compound words. As Persian allows both the left and the right constituent to be the head, the words are tagged for headedness. The lexical variable ratings provided here allow experimental researchers to choose compound word stimuli for their experiments, controlling for the rated variables. The ratings can also be used to explore the effects of these lexical features in different compound word processing tasks such as naming and lexical decision. Kuperman et al., (2012) was the first study that reported the effects of these variables on lexical decision performance for

English and reported a high correlation between AOA and lexical decision times. The lexical variable ratings presented in this chapter allow us to explore factors that affect compound word processing in Persian.

Results showed that familiarity, imageability and AOA are highly intercorrelated. Another major aspect of my data relates to the headedness component added to the matrix of subjective norms. To the best of my knowledge, there have been no other studies up to this date that have provided such ratings in a language that allows different positions for the head of compound words. The AOA results showed that the LH compounds were thought to have been learnt earlier in life. This finding suggests that left-headed or so-called historical compounds (Kahnemuyipour, 2014) that are shaped by the Ezafe construction and are more transparent in Persian, were acquired earlier in life. I also observed that the familiarity, imageability, and AOA ratings were highly intercorrelated. This suggests that compounds that were learnt earlier in life were more familiar and could arouse more mental images.

The database presented in this study also provides the Google frequency counts for the selected compounds and for their constituents. As mentioned earlier, in Persian, there is a lack of consistency in the orthography of compound words and, thus, assessing corpus frequencies of compound words is difficult. I inspected the correlations of my rated variables with Google frequencies of the compounds and the results showed no significance. The existence of Ezafe structure in Persian doesn’t allow for distinguishing the LH compounds from two nouns connected with Ezafe in written materials. Therefore, Google frequency data are not reliable in Persian.

There are, of course, many other variables that have been reported to have an effect on the processing and recognition of compound words. These can include the conceptual relations between the two constituents of a compound word (see Schmidtke, Kuperman, Gagné, & Spalding, 2016), morphological family sizes and frequencies of the constituents (see Juhasz & Berkowitz, 2011; Kuperman et al., 2008), as well as valence and arousal (see Kuperman, 2013). Future research can use the present database to incorporate the above-mentioned variables along with other potential ones to shed more light on compound word recognition and representation in the mental lexicon.

References

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. https://doi.org/10.1080/09541440340000123

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283–316. https://doi.org/10.1037/0096-3445.133.2.283

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers*, *33*(1), 73-79.

Butterworth, B. (1983). *Lexical representation* (pp. 257-294). na.

Cortese, M. J., & Khanna, M. M. (2007). Age of acquisition predicts naming and lexical-decision performance above and beyond 22 other predictor variables: An analysis of 2,342 words. *Quarterly Journal of Experimental Psychology*, *60*(8), 1072–1082. https://doi.org/10.1080/17470210701315467

Cortese, M. J., & Schock, J. (2013). Imageability and age of acquisition effects in disyllabic word recognition. *Quarterly Journal of Experimental Psychology*, *66*(5), 946–972. https://doi.org/10.1080/17470218.2012.722660

Dabir-Moghaddam, M. (2019). Typological approaches and dialects. In In A. Sedighi & P. Shabani-Jadidi (eds), *The Oxford Handbook of Persian Linguistics*. pp. 52-88. Oxford: Oxford University Press.

Ellis, A. W., & Lambon Ralph, M. A. (2000). Age of acquisition effects in adult lexical processing reflect loss of plasticity in maturing systems: Insights from connectionist networks. *Journal of Experimental Psychology: Learning, memory, and cognition*, *26*(5), 1103.

Fabb, N. (1998). Compounding. In A. Spencer & A. M. Zwicky (eds), *The handbook of morphology*. pp. 66–83. Oxford: Blackwell.

Foroodi-Nejad, F., & Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in PersianEnglish bilinguals. *Bilingualism*, *12*(4), 411–427. https://doi.org/10.1017/S1366728909990241

Forster, I., & Taft, M. (1976). Lexical storage and retrieval of polymorphemic and polysyllabic words. *Journal of verbal learning and verbal behavior*, *15*(6), 607-620.

Frommer, P. R. (1982). Post-verbal phenomena in colloquial Persian syntax.

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281. https://doi.org/10.1037/0096-3445.113.2.256

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods & Instrumentation*, 12(4), 395-427.

Hirsh, W. K., & Funnell, E. (1995). Those old, familiar things: Age of acquisition, familiarity and lexical access in progressive aphasia. *Journal of Neurolinguistics*, *9*(1), 23–32.

Inhoff, A. W., Starr, M. S., Solomon, M., & Placke, L. (2008). Eye movements during the reading of compound words and the influence of lexeme meaning. *Memory and Cognition*, *36*(3), 675–687. https://doi.org/10.3758/MC.36.3.675

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford University Press, USA.
Juhasz, B. J., & Berkowitz, R. N. (2011). Effects of morphological families on English compound word recognition : A multitask investigation, *26*. https://doi.org/10.1080/01690965.2010.498668
Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods*, *47*(4), 1004–1019.
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.
Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words. *British Journal of Psychology*, *94*(2), 223–244.
Kahnemuyipour, A. (2014). ScienceDirect Revisiting the Persian Ezafe construction : A roll-up movement analysis. *Lingua*, *150*, 1–24. https://doi.org/10.1016/j.lingua.2014.07.012
Kalbasi, I. (1997). *Sakhte eshteghaghiye vage dar zabane Farsi emrooz* [The derivational structure of word in modern Persian]. Tehran: Pajuheshgah Press.
Kuperman, V. (2013). Accentuate the positive : Semantic access in English compounds, *4*(April), 1–10. https://doi.org/10.3389/fpsyg.2013.00203
Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7–8), 1089–1132. https://doi.org/10.1080/01690960802193688
Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876–895. https://doi.org/10.1037/a0013484
Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior research methods*, *44*(4), 978-990.
Libben, G., & Jarema, G. (2006). The representation and processing of compound words. *Word Knowledge and Word Usage*, 336.
Libben, G. (2005). Everything is psycholinguistics: Material and methodological considerations in the study of compound processing. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *50*(1-4), 267-283.Libben, G. (2014). The nature of compounds : A psychocentric perspective. *Cognitive Neuropsychology*, *31*, 8–25.
Libben, G., Gibson, M., Yoon, B., & Sandra, D. (2003). Compound fracture : The role of semantic transparency and morphological headedness q. *Brain and Language*, *84*(1), 50–64.
Megerdoomian, K. (2019). Typological approaches and dialects. In *The Oxford Handbook of Persian Linguistics*. (pp. 461-480). Oxford: Oxford University Press.
Olsen, S. (2001). Copulative compounds: a closer look at the interface between syntax and morphology. In *Yearbook of Morphology 2000* (pp. 279-320). Springer, Dordrecht.
Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinheart & Winston.
Paivio, A. 1986. Mental representation: A dual-coding approach.
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.

Pollatsek, A., & Hyönä, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive processes*, *20*(1-2), 261-290.

Schmidtke, D., Kuperman, V., Gagné, C. L., & Spalding, T. L. (2016). Competition between conceptual relations affects compound recognition: the role of entropy. *Psychonomic Bulletin and Review*, *23*(2), 556–570. https://doi.org/10.3758/s13423-015-0926-0

Schreuder, R., & Baayen, R. H. (1995). Modeling morphological processing. *Morphological aspects of language processing*, *2*, 257-294.

Semenza, C., & Luzzatti, C. (2014). Combining words in the brain : The processing of compound words Introduction to the special issue, *31*, 1–7.

Shariat, M. J. (2005). *Dastoor zabane Farsi [Persian grammar 1]*. Tehran: Asatir Press.

Shoolman, N., Andrews, S., Kinoshita, S., & Lupker, S. (2003). Masked priming: The state of the art.

Spencer, A. (2003). Periphrastic paradigms in Bulgarian. *Syntactic structures and morphological information*, *7*, 249-282.

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition , imageability , and familiarity. *Behavior Research Methods*, *38*(4), 598–605.

Steyvers, M., & Tenenbaum, J. B. (2005). The Large-Scale Structure of Semantic Networks : Statistical Analyses and a Model of Semantic Growth, *29*, 41–78.

Turner, J. E. (1998). Contrasting effects of age of acquisition and word frequency on auditory and visual lexical decision, *26*(6), 1282–1291.

van Jaarsveld, H. J., & Rattink, G. E. (1988). Frequency effects in the processing of lexicalized and novel nominal compounds. *Journal of Psycholinguistic Research*, *17*(6), 447–473. https://doi.org/10.1007/BF01067911

Williams, R. S., & Morris, R. K. (2004). Eye movements, word familiarity, and vocabulary acquisition. *European Journal of Cognitive Psychology*, *16*(1–2), 312–339. https://doi.org/10.1080/09541440340000196

3 Experiment Two: Short-term memory for Persian compound words: Effects of constituent order and syntactic role

Abstract

Morphological processing has mainly been studied in reading, lexical decision and naming tasks. Here, I present data from a study of compound word maintenance in short-term memory. Headedness refers to the order of the syntactic roles (modifier and head) in noun-noun compound words. English compounds are right-headed (RH) with modifiers preceding heads. Results from morphological processing experiments in variable-headed Italian have suggested a central role for transparent right heads in constituent access and semantic combination. Persian allows both left and right head positions of its compounds. I tested 34 native Persian speakers using memory span tasks and looking for different recall of lists consisting of left- and right-headed compounds. Because right headedness is the default in Persian (Kahnemuyipour, 2014), I hypothesized that RH compounds would also be remembered better than their LH counterparts. I formed lists of five different types of auditory stimuli: the first two types of lists included compounds that shared a constituent. All the compounds in the Shared as RH condition were right-headed (head-final) and the ones in the Shared as LH condition were left-headed (head-initial). The head placed in the second position in the Shared as RH condition (e.g., *ordak* ***maahi,*** literally *'duck fish' meaning "duck fish"*), were the same as the heads in the initial position of the compounds in the Shared as LH condition (e.g., ***maahi*** *Qobaad*, literally *'fish Qobaad',* meaning *"Qobaad fish"*). The next two conditions were named Common RH and Common LH list types. They consisted of lists with some common right-headed (e.g., *dast band*, literally 'hand band', meaning "*bracelet*") and left-headed (e.g., *shaal garden*, literally *'shawl neck'*, meaning *"winter scarf"*) compounds in the Persian language. There was no constituent overlap between these conditions. The last created condition (Always RH) included right-headed compounds from families of compounds in which the position of the head is usually fixed so that the head constituents never occurs in initial position (e.g., family of pain such as *kamar dard*, literally *'back pain', meaning "back pain"*). The planned comparison analysis found only one significant effect: better recall of the Common RH compounds compared to the Common LH ones. The results suggest that processing of an irregular or a marked structure can negatively affect STM span in Persian. This supports a syntax-based decomposition view of the mental lexicon in Persian since it reveals that it is not the constituent order but the syntactic structure of the language that plays a major role in the processing of compounds.

3.1 Introduction

Headedness refers to the order of the syntactic components of noun-noun compound words. Unlike in English, where the second constituent of a noun-noun compound word is nearly always the head (e.g. a *firehose* is a hose for extinguishing fire), Persian also allows compounds with left heads (Shariat, 2005; Foroodi-Nejad & Paradis, 2009). These are compounds with the constituent order head–modifier (e.g., *halqe ezdevaj*, literally 'ring marriage' meaning *"wedding ring"*). Such variability in headedness raises the question of whether processing of compounds is different for different internal structures. Despite a significant body of research on the processing of compound words in general, there are only a few studies on the effects of headedness on the processing of compounds and how the brain processes competing syntactic-semantic structures within words. Furthermore, studies of Italian (Marelli & Luzzatti, 2012; Arcara, Marelli, Buodo, & Mondini, 2014), another language with variable headedness, have suggested influence of language-specific syntactic structure, which needs to be studied in other languages in which constituent position and syntactic role can be separated.

Memory experiments on compound words allow researchers to understand the role of morphology in representation and recall of morphologically complex words. Their representations might vary based on the morphological processes (e.g., derivation, inflection, compounding) that have been applied to the words. Morphological effects on short-term memory (STM) performance can challenge word comprehension theories by showing how syntax and semantics forming language structure interrelate with cognition for language understanding and remembering (Service & Maury, 2015). Studying memory tasks rather than reading or word processing tasks should especially maximize sensitivity to spreading activation and competition processes in the mental lexicon, as these tasks force the simultaneous representation of several words as well as their morphological constituents.

Previous research on how various lexical variables of compound words in English can affect their recall, reported that for right-headed compounds "with larger left family sizes and higher left constituent frequencies the modifier was used as a retrieval cue at the semantic level when accessing whole word forms" (Wälchli, 2016, p.59). Considering that compounds in English are always right-headed, it is interesting to investigate the effects of compound constituents on processing in languages in which the position of the head can vary. Most Romance languages, including Italian, allow for such variability in compound words. Arcara et al. (2014) used both lexical decision task and event-related potential (ERP) methods to investigate if position of the head affects the processing of Italian noun-noun (NN) compounds. They presented their stimuli in two different conditions, as whole words or decomposed into their constituents. Their behavioural findings, which were in line with those of el Yagoubi et al. (2008), demonstrated that in the lexical decision task, their decomposition condition affected the head-initial (left-headed) compounds less than the head-final (right-headed) ones, which showed a higher processing load. This means that when the constituents of the compound words were written as two separate words, processing of head-final compounds required higher amounts of working memory resources. Therefore, Arcara, et al. (2014) concluded that headedness can affect the processing of compound words in Italian. The authors believe that the reason for this difference in the behaviour of head-initial versus head-final compounds in Italian is related to the fact that head-final compounds are a marked case in Italian. Left headedness is considered to be the default in most Romance languages, including Italian, since it is the canonical order inside NPs in the language. As mentioned in chapters one and two, the default for Persian compound words is right headedness. Although this is not in line with the canonical order of elements of NPs in the language, most of the compound words in Persian are right-headed. The left-headed ones, as discussed before, consist of historically two words joined by the Ezafe structure, which was gradually dropped in the language and resulted in left-headed compound words. As Persian acts differently from Italian with regard to the markedness of its compounds, conducting research on compounds in this language can add to our understanding of the potential role of the canonical order of NPs and underlying differences in the deep structure of compounds, such as a dropped Ezafe construction. We can ask whether STM capacity for compound words depends on constituent order or constituent roles, or perhaps more subtle compound-internal structure featuring in this language.

3.2 Morphological processing of complex words

In the next section, I will provide an overview of different lexical and semantic theories on the processing of morphologically complex words. Since compound words are considered to be under the umbrella of complex words, these theories are all applied to the processing of compound words as well. The data of this chapter can shed more light on the processing of compound words in languages other than English. Different recall for RH versus LH compounds in Persian can extend or challenge the existing understanding, pointing to decomposition routes for the processing of complex words.

3.2.1 Lexical and semantic theories

Over the past few decades, there has been a general debate concerning the processing of morphologically complex words. The main question is whether these words are stored and accessed as a whole in the mental lexicon or if they are decomposed into their morphological constituents. Different theories have been developed to address this matter. The full listing theories (Butterworth, 1983) suggests the words are stored as a whole and that each word has a separate representation in the mental lexicon. Libben, (1998) mentions that this theory would pre-suppose a large capacity for the mental lexicon since words need to be stored as a whole, listing all inflected forms. The parsing theories, on the other hand, propose full decomposition of complex words into their constituents (Forster & Taft, 1976). According to these theories, the stem forms of multimorphemic inflected words are stored in the mental lexicon. Therefore, when reading, the target for a lexical search is this stem form which allows the matching of the stimulus letter strings with it. The theory by Forster and Taft (1976) requires parsing of multimorphemic words into their morphological constituents for access to the whole word forms. There are also variants of these theories that allow whole word activation prior to the activation of the constituents (see, e.g., Giraudo & Grainger, 2001).

Libben (1998) has suggested that the representation and processing of multimorphemic words depends on the language in question, as well as various lexical factors, such as part of speech (noun, verb, etc.), morphological type (compounded, derived, inflected), word frequency, and the semantic relationship between the constituents of the multimorphemic words. Another important factor in this context is the role of semantic transparency. It is suggested that transparent compounds, whose meaning is related to that of their constituents (e.g., blueberry, teaspoon), are decomposed during input processing. In contrast, opaque compounds (e.g., deadline, buttercup) have shown no effect when primed with their constituents (see, e.g., Sandra, 1990). Neither the full-listing nor the full decomposition theory can fully explain the findings reported to date. Dual route theories of processing put forward that there are both whole-word and decomposition routes for morphologically complex words. However, the mechanisms underlying these theories, which are also being referred to as hybrid theories (see for example Pollatsek, Hyönä, & Bertram, 2000**;**

Schreuder & Baayen, 1995; Badecker & Caramazza, 1991; Laudanna, Badeckera, & Caramazza, 1992), are still not clear. The early forms of the dual route theory suggested that the whole-word access and decomposition processes work in parallel, and that their effectiveness is determined by morpheme and word frequencies (see Semenza & Luzzatti, 2014, p.2). Based on this version of the theory, complex words with high whole word frequencies are processed as wholes whereas when the frequency of the whole word is low, but the constituent morphemes are highly frequent, a parsing process is activated (Semenza & Luzzatti, 2014). Newer theories, called multiroute theories, introduce a flexible procedure from the earliest stages of complex word processing that is based on information from both the whole word and its constituents (Kuperman, Schreuder, Bertram, & Baayen, 2009).

All of the above-mentioned theories are focused on the lexical characteristics of complex word processing. Yet, the representation and processing of semantic properties of words and their interaction with lexical aspects are still poorly understood. The traditional theories suggest that semantic features only play a role at the later stages of word processing (Libben, 1998) whereas some more recent theories suggest that semantics can influence lexical access at the very early stages of processing (Marelli & Luzzatti, 2012).

An important theory related to the production of compounds is the "lemma" theory (see for example Levelt, Roelofs, & Meyer, 1999 for a more comprehensive description of lemma and lexical access). According to this serial theory of language production, beginning from a concept, the output of the first selective stage in word production contains both syntactic and semantic properties, and is called the *lemma*. The output of the second stage adds the phonological information of the selected word resulting in a representation named a *lexeme*. Therefore, the lemma level is an intermediate level linking the semantics and grammatical structure of a specific word to its phonological information (Semenza & Luzzatti, 2014). This theory suggests that the information specifying compound words for output is stored at the lemma level. A number of psycholinguistic experiments have supported this hypothesis (e.g., Badecker, Miozzo, & Zanuttini, 1995; Biran & Friedmann, 2011; Delazer & Semenza, 1998). An alternative to the lemma theory is the independent network theory (Caramazza, 1997). This theory rejects the lemma level and suggests that the syntactic information of words is stored in a separate network, which does not form an intermediate level. The two theories have different assumptions about the representation of morphologically complex words. In the lemma theory, morpheme-based representations are stored at the word form (lexeme) level (Levelt et al., 1999) whereas in the independent network theory, it is proposed that words are stored as whole forms (Caramazza, 1997; Janssen, Bi, & Caramazza, 2008). Caramazza's theory is not directly related to the processing of complex words and his proposal is based on the reported frequency effects for whole words and not frequency effects for morpheme-based representations of nominal compounds (see Semenza & Luzzatti, 2014).

3.2.2 Lexical variables related to compound processing

(i) Semantic transparency

Semantic transparency has been shown to play a key role in understanding how multimorphemic words are processed and represented in the mental lexicon (Libben, Gibson, Yoon, & Sandra, 2003**).** Zwitserlood (1994) notes that because there is huge variation in the transparency of compounds as a whole and relative to their constituents, compound words can be used as perfect stimuli to examine the effects of form and meaning in word comprehension and production. Research has separated semantic transparency of the entire multimorphemic string from that of its separate constituents in relation to the full word meaning. Looking at the semantics of individual constituents, Zwitserlood (1994) categorized compound words into three main groups: 1) Compounds consisting of two semantically transparent constituents (e.g., *birdhouse*), 2) compounds consisting of one semantically transparent constituent and one semantically opaque constituent (e.g., *bilberry*), 3) compounds consisting of two semantically opaque constituents (e.g., *hotdog*) (see also Wälchli, 2016). Libben, (2014) notes that English compound words with a semantically transparent first constituent and a semantically opaque second constituent (e.g., *jailbird*) are less common than those with semantically opaque first constituent and semantically transparent second constituent (e.g., *nickname*). This might suggest that the second component is more important in signalling the meaning of the whole word.

(ii) Frequency

Studies have shown that the whole word frequency of compounds as well as the frequency of the constituents can influence word production tasks (see e.g., Blanken, 2000 for data on a naming task by German aphasic patients). Forster and Taft (1976) were the first to show that, for compounds, the frequency of the first lexeme can influence lexical decision times. Since then, a great number of word recognition experiments have used compounds to see whether lexical access is achieved through the constituents of a compound or through the word as a whole. Among these are constituent-priming studies that examined the role of the constituent lexeme frequencies on lexical decision performance (see for example Andrews, Miller, & Rayner, 2004; Libben, Gibson, Yoon, & Sandra, 2003). In line with these are studies in English that manipulated constituent frequency controlling for whole word frequency. The results revealed differences in processing times indicating that the compounds lexical decision times were shorter when the final constituent was a high frequency. This indicates that constituents were accessed during compound word recognition (e.g., Andrews, Miller, & Rayner, 2004; Inhoff, Starr, Solomon, & Placke, 2008; Juhasz, Starr, Inhoff, & Placke, 2003). Research in other languages, like Finnish, with more productive compounding, has reported similar results (e.g., Pollatsek, Hyönä, & Bertram, 2000). More recent studies in this field have been focused on how conceptual and morphological relations between constituents affect the organization of compounds in the mental lexicon (e.g., Kuperman, Bertram, & Baayen, 2008; Kuperman et al., 2009; Schmidtke, Kuperman, Gagné, & Spalding, 2015).

(iii) Headedness

Headedness here refers to the order of the syntactic roles (modifier and head) in noun-noun (NN) compound words. Boutonnet et al. (2014) consider headedness as an important factor in the context of morphological complexity, playing a key role in processing of compound words. Headedness can vary and compounds, depending on the language, can be either head initial (left-headed) or head final (right-headed). Compounds in English are always right-headed, with some rare exceptions, whereas in some languages, like Hebrew and Vietnamese, the head always precedes the modifier (see Fabb, 1998, for Vietnamese). Some Romance languages, such as French and Italian, allow variable head positions in compounds (see e.g., Semenza & Luzzatti, 2014, for Italian; Libben & Jarema, 2006, for French). Interestingly, variable head position also occurs in Persian, which is an Indo-European language that has SOV word order (Kahnemuyipour, 2014; Foroodi-Nejad & Paradis, 2009; Kalbasi, 1997; Shariat, 2005). It is conceivable that this variability in compound headedness between different languages could lead to differences in compound processing between languages with different headedness. Priming studies on compounds and their morpheme constituents indicate that the head constituent as a prime can facilitate the processing of the whole form more than the modifier (see for example Jarema et al., 1999). These sorts of findings point to decomposition of compounds during processing and that the head constituent can be a better access cue to the whole form. This can also be in line with studies suggesting that unless the constituents are semantically opaque, compounds are not represented or processed as a whole in the mental lexicon (e.g., Wälchli, 2016; Boutonnet et al., 2014).

As mentioned in chapters 1 and 2, LH compounds in Persian are semantically more transparent than their RH counterparts. I mentioned in the previous chapter that because of the existence of the Ezafe structure in Persian, which has no orthographic form, it is not possible to separate the left-headed compounds from NN constructions attached with Ezafe in written contents. This applies also to my Google frequency check, presented in the previous chapter, and makes these data unreliable. Therefore, only ratings of familiarity, imageability, and AOA, as lexical variables in this study, were used as the control variables. I was interested to see how the mentioned differences on constituent order in Persian can affect the STM in a memory recall experiment.

3.3 Maintenance and recall mechanisms for verbal information in working memory

3.3.1 Memory for complex words

Most morphological processing tasks have targeted the time course of one or several processes by behavioral or neurophysiological measures. Baddeley (1992) defines working memory as “the temporary storage of information in connection with the performance of other cognitive tasks such as reading, problem solving or learning” (p.311). Working memory (WM) tasks can reveal the cognitive limits for keeping different aspects of linguistic information in a readily accessible state. Most commonly, the ability to store verbal material in short-term memory (STM) for immediate ordered recall has been tested. However, the number of studies on the role of morphological complexity in immediate recall tasks is very small up to this date (see for example Service & Tujulin, 2002; Németh et al., 2011, Service & Maury, 2015; Wälchli, 2016).

Service and Tujulin, (2002) investigated how morphological complexity in Finnish can affect the recall of word sequences in various working memory (WM) experiments. Their results demonstrated that memory load was affected by both derivation and inflection. They showed that morphological complexity can negatively affect both simple STM span and more complicated WM tasks that involved secondary processing tasks in addition to STM storage. Their results also indicated different processing based on different classes of complex word forms (e.g., derived forms from a base form *kirja* meaning “book” such as *kirja* + *sto* [library] were processed differently from inflected forms like *kirja* + *sta* [from book]) (p.45). This speaks to theories of how cognition interacts with the lexical and semantic features of words during language comprehension, storage and production (Service & Tujulin, 2002; Service & Maury, 2015).

In a study of memory for Hungarian morphologically complex words (Németh et al, 2011), participants recalled derived words "(e.g., boy + hood)" better than inflected ones "(e.g., boy + s)" (p.85), and morphologically regular structures better than irregular ones. Service and Tujulin, (2002) reported results that are in line with these results showing that for both visual and auditory stimuli, monomorphemic words were recalled better than suffixed words. They reported that both inflected and derived forms were harder to recall than base forms in their Word Span Task, with derived forms being easier to recall than the inflected forms. This finding can be understood if parsing words into smaller constituents, or decomposing a morphologically complex word into their morphemes, can deplete resources from being used in STM.

In a study of the combinatorial confusability of constituent combinations (Matzen & Benjamin, 2009), compound words as well as single target words were presented in a list or embedded in sentences for later recognition. One objective was to examine the differences between a sentence and a list context in ability to predict different forms of false memories: re-combinations of compound word constituents or semantically similar words. In one experiment, the stimuli were compound words that can make a conjunction lure (e.g., *floodgate* and *taillight* for false recognition of *tailgate*). Participants saw these words either as single words or in a sentence context. After each set was finished, participants were presented a list consisting of previously presented compound words, new compound words, and re-combination lures, and asked to decide if they had seen each compound word before. Participants made fewer conjunction lure errors when compounds had been seen in a sentence context. The authors concluded that these results indicate that the information on words' surface forms, and not that much on their meaning, were maintained by participants, when words do not happen in a sentence context.

Altogether, the information that has to be processed in an immediate recall task or a complex WM task affects recall in addition to other potentially important variables such as phonological factors and factors forming part of lexical knowledge, which are addressed by most theories of verbal STM and WM.

3.3.2 Working memory and its maintenance mechanisms for verbal information

The most influential conceptualization of WM was proposed by Baddeley and Hitch (1974). Based on this model, the most important component of WM is the central executive that is described as an "attentional controller" (Baddeley, 2000; Baddeley, 2003). This component combines information from two modality-specific components, the visuospatial sketchpad and the phonological loop, into consciously available episodes that are coherent. In a newer version of the model of WM, the central executive cannot store information, and a new component, the episodic buffer, has this role. Verbal information in immediate recall tasks is, however, posited to be stored and rehearsed in the phonological loop, a STM component for speech-coded content (Baddeley, 2000).

Barrouillet and Camos have recently introduced their version of a WM model for processing and maintenance of information, called the Time-Based Resource-Sharing (TBRS) model (Barrouillet, Bernardin, & Camos, 2004). This model proposes that over the course of time, memory traces in a list decay (see also Hurlstone, Hitch, & Baddeley, 2014; Camos & Barrouillet, 2014). In complex WM tasks that combine processing and maintenance of information, cognitive load is created by the proportion of time, between stimulus presentation and recall, that task processing occupies, preventing the executive component (the executive loop) from performing maintenance functions on decaying memory traces. In word recall experiments with distracting processing tasks, as the cognitive load of a simultaneous processing or distractor task increases, the accuracy of recall of memorized words declines. The time-based decay of items in memory can be prevented by two kinds of active maintenance: articulatory rehearsal and attentional refreshing. Attention simplifies processing of a target word during processes that have been involved in the maintenance of different kinds of information in WM (Awh, Vogel, & Oh, 2006). Engle, Kane, and Tuholski (1999) emphasized the "interaction of attentional and memorial processes in the working memory system", and argued "that this interaction between attention and memory is an elementary determinant of broad cognitive ability" (p.147). When an attentional WM component called the executive loop is occupied by a distracting task it cannot perform attentional refreshing. Based on the TBRS model, attention plays an important role in WM, and is shared by processing and maintenance of information. Attention The way verbal information is maintained in this model of WM is partly similar to the function of the phonological loop in Baddeley's model (Baddeley, 1983), which relies on silent articulatory rehearsal. Articulatory rehearsal is thought to be less dependent on the executive loop than the other maintenance function, i.e., attentional refreshing (Camos, Mora, & Oberauer, 2011). Based on a study in which the two rehearsal processes were selectively blocked, Camos et al. (2009) suggest that articulatory suppression and attentional demand can have separate negative effects on the maintenance mechanisms of WM. The TBRS model has been challenged by researchers arguing against time-based decay and suggesting that poor memory performance is mostly related to interference in WM (see Lewandowsky, Oberauer, & Brown, 2009; Oberauer & Lewandowsky, 2014). However, the criticism does not concern the assumption of the executive loop as a universal bottleneck of information processing. Thus, the effects of morphological complexity on STM and WM tasks could be explained by morphological processes taking up executive loop time that has to be shared with maintenance processes. The rationale for the present study is based on the assumption that extra processing steps in maintenance of morphological information limits the ability to maintain verbal items in an accessible state.

3.4 The present study

Literature on compound headedness in languages that allow for different head positions, such as Italian, indicates that headedness can influence the processing of compound words (Yagoubi, Chiarelli, Mondini, Danieli, & Semenza, 2008; Arcara, Marelli, Buodo, & Mondini, 2014). In the event-related potential (ERP) study by Arcara et al., (2014), Italian head-initial (left-headed) and head-final (right-headed) compounds were presented as visual stimuli in a lexical decision task to test if compounds in Italian are processed differently based on their head position. The compounds were shown as written together as single words or as artificially split into two words. The results showed that, when the compounds were split and decomposition forced, greater effort resulting in a higher amplitude ERP component (LAN) was seen for processing head-final compared to head-initial compounds. This was interpreted to result from the different mental representations for head-initial and head-final compounds in Italian. In Italian the canonical order in noun phrases is head – modifier, corresponding to the head-initial, LH type for compounds. On the other hand, RH has been suggested to be the default structure of compound words. Thus, these Italian results could be explained by a conflict between the syntactic structure of phrases and that of compound words. Persian is a language that affords an opportunity to investigate headedness in another language with variable headedness but different syntax.

As previously mentioned in this chapter, the canonical order of Persian NPs is noun + modifier. In contrast, noun-noun (NN) compounds, consisting of a head and a modifier are right-headed by default. Left-headed Persian compounds were historically two nouns combined with a special Ezafe construction (see Kahnemuyipour, 2014 for a definition of historical compounds). Ezafe is an unstressed vowel -e (-ye after vowels) appearing between a noun and its modifiers in an NP as well as any subsequent modifiers if present (Kahnemuyipour, 2014). It is important to note that LH compounds in Persian are semantically more transparent (Kahnemuyipour, 2014) and are more productive in the context of compounding. Items in this group can only be considered compounds when they are not pronounced as Ezafe structures (Kahnemuyipour, 2014). A main research question in this study is whether this internal difference between LH and RH Persian compounds can lead to different processing of these two types of compound words in Persian.

In the present study, I aimed to investigate how STM for compound words can be affected based on their headedness. To do so, I designed an immediate serial recall experiment in which lists of auditory compound words were presented in five different conditions. In the first two conditions I compared compounds that shared a constituent. All the compounds in the Shared as RH condition were right-headed (head-final) and the ones in the Shared as LH condition were left-headed (head-initial). This means that for every compound that was put in the Shared as RH condition and used constituent A as a head in the final position (e.g., e.g., *ordak* ***maahi,*** literally *'duck fish' meaning "duck fish"*), there was a compound in the Shared as LH condition which used constituent A as a head in the initial position (e.g., e.g., ***maahi*** *Qobaad*, literally *'fish Qobaad',* meaning *"Qobaad fish"*). Thus, these word pairs shared the head but in different positions. This strict formula limited the set of available stimuli but allowed to control for various lexical features of the head constituent. However, it was not possible to control the characteristics of the modifier constituents at the same time.

Since the compound whole word characteristics could not be controlled when comparing the two patterns Shared as LH and Shared as RH, two other list types were created. The Common RH and Common LH lists consisted of compound words with some common right-headed (e.g., *ketaab khaneh 'book house' "library"*) and left-headed (e.g., *aab porteqaal 'water orange' "orange juice"*) compounds, meaning that the frequency of the RH or LH compound as a whole was high. I added one more condition (Always RH) to my pool. This last list type consisted of right-headed compounds from families of compounds in which the position of the head is usually fixed, i.e., there are no examples of the head nouns occupying the left position (e.g., family of pain such as *kamar dard 'back pain' "back pain"*). There are no families of compounds with fixed left heads.

Since right headedness is the default in Persian, I generally hypothesized that in a memory span task, native Persian speakers would be more likely to correctly recall head-final (RH) compounds than head-initial (LH) ones. Therefore, I expected words in the Shared as RH condition to be recalled better than words in the Shared as LH condition. I also expected the Common RH condition to result in better recall than the Common LH condition. Finally, I was interested to see how the recall for fixed-headed compounds in the Always RH condition would be different from my other conditions as there were no compounds with these heads in a competing left position.

3.5 Methods

3.5.1 Participants

Thirty-four participants (and 2 pilot participants) took part in this experiment (13 females). These individuals were between the ages of 20 and 40; with no visual, language, learning or hearing problems. Participants were all native speakers of Persian. They were compensated with $15 CAD dollars for 1.5 hours of their time.

3.5.2 Word stimuli

The words were taken from a previously created database of Persian compound words described in Chapter 2, with ratings of familiarity, imageability, and age of acquisition (AOA) of 149 Persian NN compound words. This allowed me to control for the potential effects of these lexical variables on recall accuracy in immediate serial recall.

As there are no reliable frequency counts for Persian compound words, I used the lexical variables from this study as control variables. I created five pools of compounds for creation of lists for each one of the five conditions (Shared as RH, Shared as LH, Always RH, Common RH, Common LH). There were 10 lists of each of the list lengths of 3, 4, 5, and 6 words (180 words in each condition). I recorded the auditory stimuli with a voice recorder in the Language, Memory, and Brain Lab in in ARiEAL Research Centre at McMaster University. I, then, checked the loudness of the recorded stimuli with Audacity.

3.5.3 Procedure

Testing took place at the McMaster University Language Memory and Brain Lab in ARiEAL Research Centre at L.R. Wilson Hall. The experiment lasted approximately 80 to 90 minutes. A consent form approved by the McMaster Research Ethics Board was administered and completed prior to data collection. Following the data collection, participants were debriefed on the study's research questions.

Participants sat in front of an iMac computer. The task was immediate recall of lists of compound words. Ten lists of each of the lengths 3, 4, 5, and 6 words were presented to participants auditorily, controlled by SuperLab 5 software (Cedrus Corporation). I used a stimulus-onset asynchrony (SOA) of 1 second. The accuracy scoring was done for each word separately. This means that if a word in a list (for example the first word in the first list of 3 words) was recalled as a correct item in the correct position, the score was 1. Otherwise, the score was 0. All participants listened to all the five different conditions, but the order of conditions was counter-balanced between participants. If the participants required a break, they were allowed to stop between the lists. The study was cleared by McMaster Research Ethics Board.

3.5.4 Data analysis

In the analysis, I considered *list length* and *list type* as fixed variables. My primary variable of interest, that is my dependent variable, was the binomial *recall accuracy* of each presented compound word by each participant. I used generalized linear mixed effects logistic regression models with maximum likelihood (Laplace approximation) goodness of fit statistics from the lme4 R package (lme4, R version 3.5.2, R Foundation for Statistical Computing, 2018). Both participants and items (word IDs) were modelled as random variables. To control for lexical effects, I included fixed effects for the control variables: *familiarity*, *imageability*, and *AOA* for each word. These variables as well as *list length* were scaled in the analyses. To investigate the effects of headedness on recall, I investigated planned pairwise contrasts between list types using multivariate *t*-adjustment. I expected the list types which include right-headed compounds to show better recall than their left-headed counterparts. I then looked at the odds ratio of the following list comparisons: Shared as RH versus Shared as LH, Common RH versus Common LH, Shared as RH versus Always RH, and Common RH versus Always RH.

3.6 Results

The average proportion of recall over participants was 0.749 with the standard deviation of 0.129. This shows that the total recall accuracy over participants was almost 75 %, neither at ceiling nor at floor. *Figure 3-1* shows recall accuracy as a function of list type in lists of different lengths. As the figure shows, there was a ceiling effect for list length of 3. Recall average as a function of list length is shown in *Table 3-1* and of list type in *Table 3-2*. The control variables were mean centered. The effects of the lexical control variables were not statistically significant. However, the more interesting overall effect of list type was statistically significant according to Wald test, X^2 (4) = 37.93, $p < .001$.

Next, I inspected planned contrasts between individual list types to investigate the effects of headedness on recall (see *Table 3-3*). As I made multiple comparisons, the Type 1 errors of the

Table 3-1 Mean proportion of correct recall as a function of list length: data grouped by participants and summarized by mean accuracy of recall.

List length	Recall accuracy	SD
3	0.98	0.03
4	0.91	0.10
5	0.74	0.14
6	0.52	0.19

SD =standard deviation

Table 3-2 Mean proportion of correct recall as a function of list type: data grouped by participants.

List type	Recall accuracy	SD
Shared as RH	0.77	0.14
Shared as LH	0.76	0.12
Always RH	0.73	0.14
Common RH	0.77	0.14
Common LH	0.68	0.13

SD =standard deviation

post-hoc comparisons were controlled by multivariate-*t* adjustment. The pair-wise comparisons based on log odds ratio are shown in *Table 3-3*. The odds ratio scale allows reporting how many times more likely an event is compared to another event. For instance, the odds of recalling a compound word correctly in list type Common RH were 2.10 times as high as the odds for recalling an item in the Common LH list type and the odds of correct item recall in Shared as RH lists were 1.61 times as high as the odds in the Always RH lists. The contrasts between Common RH compounds and Common LH compounds, and between Shared as RH and Always RH were statistically significant as the confidence intervals did not include 1.

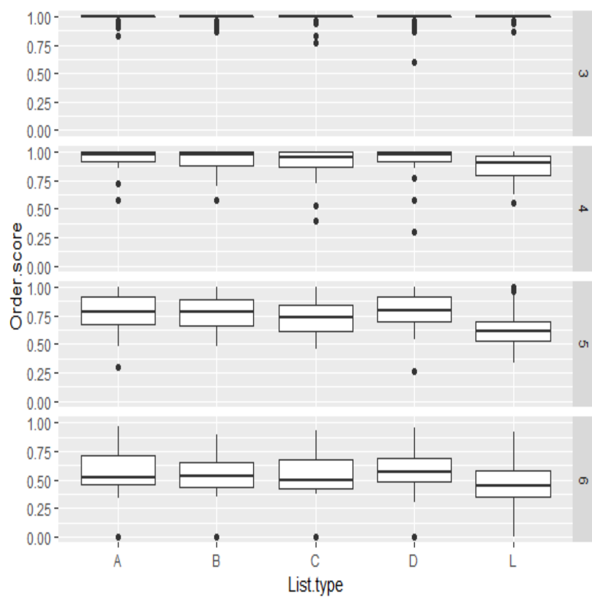

Figure 3-1 Boxplots of recall accuracy proportion as a function of list type, separated by list length. The horizontal bar denotes mean proportion of correct recall. List length is indicated on the right side. A = Shared as RH, B = Shared as LH, C = Always RH, D = Common RH, and L = Common LH.

3.7 Discussion

This study examined the effects of compound headedness on recall in an immediate recall task in Persian. This language allows for variable head positions of its compounds. Right headedness is the default in Persian, whereas a linguistic analysis of left-headed compounds suggests that these may have a different underlying structure involving the so-called Ezafe construction. When Persian compounds were presented in a list, I hypothesized that it would be easier for native Persian speakers to recall head-final (RH) compounds than head-initial (LH) ones in a memory span task. The logistic regression model results showed that the theoretically interesting effect of list type was statistically significant. On the other hand, the lexical control variables of compound familiarity, imageability and age of acquisition were not statistically significant.

Two comparisons contrasting LH and RH compounds were investigated: one between the head-matched lists and one between sets of common LH and RH compounds. The two conditions Shared as RH and Shared as LH were related and were matched for the head constituent (Shared as RH = *constituent x – constituent n;* Shared as LH = *constituent n – constituent y*). I included these two compound types to control for frequency and other lexical characteristics of the head constituent when examining the effects of headedness on processing. With the three compound-level lexical variables controlled, and the heads being shared, I hypothesized the default RH compounds to be easier to recall than the LH compounds which are thought to have an underlying Ezafe structure. This would have provided strong evidence for comparative ease of representing RH compounds in phonological STM.

Table 3-3 Contrasts of interest between list types. Differences between pairs of list types are shown on the odds ratio scale (see text for explanation).

Contrast	**Odds ratio**	**SEM**	**Lower CL asymptote**	**Upper CL asymptote**
1. Shared as RH – Shared as LH	1.30	0.24	0.81	2.08
2. Common RH – Common LH	2.10	0.38	1.33	3.30
3. Shared as RH – Always RH	1.60	0.28	1.03	2.49
4. Common RH – Always RH	1.27	0.22	0.82	1.97

SEM = standard error of the mean, CL = confidence level; confidence intervals that do not include 1 are considered significant

Although there was a trend in the expected direction, I did not obtain statistically reliable evidence from this comparison. This could have resulted from the limited set of compound words available for the strictly head-matched condition. Some of the compound words selected for these two conditions were not common compound words in the language. Also, many of the words in the lists Shared as RH and Shared as LH were from certain families of words such as *fish* (*maahi* in Persian). Therefore, the semantic similarity of the words in the Shared as RH and Shared as LH lists might be another reason for not observing a significant difference between these two conditions.

As mentioned above, because there is a limited set of Persian compound words that share a head in different constituent positions, some of the included words may have been outliers on lexical or non-shared constituent characteristics. For this reason, I included conditions with other types of compounds. In the Common RH and Common LH compounds, only familiar compounds were included. The selection was subjective because the lexical ratings used as control variables in the analyses were not available at the time of stimulus selection. A headedness contrast between the two categories of common compounds showed a significant recall advantage for Common RH over Common LH, as hypothesized based on the linguistic analysis of headedness in Persian (see Kahnemuyipour, 2014). The lexical control variables included in the analysis guarded against the initial subjectivity of the stimulus selection. The results of this contrast provided confirming evidence for the hypothesis that RH compounds are easier to represent in STM than LH compounds. This result is interesting because LH compounds in Persian are semantically more transparent than the RH ones. Therefore, one interpretation of these results can be that transparency in Persian compounds might affect the recall negatively. Kahnemuyipour (2014) proposes that LH compounds in Persian "originated as a syntactic phrase in the Ezafe construction" (p.6) and that the "syntactically formed and hence semantically more compositional source of these compounds, leads to the more transparent meaning noted above" (p.6). Right-headed (head final) compounds, on the other hand, can demonstrate the "*base order* of modifiers and nouns in Persian" (p.6) (Kahnemuyipour, 2014, p.7).

Different processing for LH versus RH compounds in Persian is also in line with the findings of priming studies on compounds and their morpheme constituents that show that a head constituent prime can improve compound processing more than a modifier prime (see for example Jarema et al., 1999) and therefore confirming the decomposition theory of compounds during processing. These studies indicated that unless they have semantically opaque constituents, compounds are not represented or processed as a whole in the mental lexicon (e.g., Wälchli, 2016; Boutonnet et al., 2014).

Other related literature has shown that morphological complexity can negatively interact with STM span and that participants recalled derived words better than inflected ones (Service & Tujulin, 2002; Service & Maury, 2015) and morphologically regular structures better than irregular ones (Németh et al., 2011). As the regular or unmarked state in Persian compounds is right headedness, my results are also in line with the findings of these type of studies. As mentioned earlier, left-headed Persian compounds are shaped out of a structure called Ezafe, which is an unstressed vowel -e (-ye after vowels) appearing between a noun and its modifiers and any subsequent modifiers if they are present (Kahnemuyipour, 2014). Based on my results, we can interpret that processing the LH compounds, which are historically shaped out of Ezafe structures, may create more cognitive load and therefore make the recall slower.

I also investigated a further list condition, Always RH, that included heads that always occur in right position. In these compounds, the constituent order has to be modifier – head based on the lexical identity of the second constituent. I was interested to see how the Always RH compounds would be recalled in comparison with the compounds in the two other RH conditions: Shared as RH and Common RH. If there were an advantage for the Always RH lists, this would suggest a role for the second constituent in facilitating processing by ruling out a LH parse. The familiarity, imageability, and AOA ratings, taken from the data presented in the second chapter of this thesis, were used to control for other compound lexical characteristics. The results did not provide any evidence for an Always RH advantage. The contrast between Common RH and Always RH was not significant. The contrast between Shared as RH and Always RH showed the opposite effect, i.e., better recall of the compounds in the Shared as RH than the Always RH lists. Note that the experiment also included the Shared as LH lists, possibly providing long latency priming for this alternative parse for the head in the Shared as RH lists, and modulating any head-specific effects. Thus, the conclusion must be that no evidence was obtained for the role of the heads always appearing as second constituents to facilitate RH parsing of a NN compound word in Persian. These results can be better understood if we consider that the words in Always RH condition belong to certain families of words in Persian that are always right-headed. The compounds in the Always RH list belonged to semantic families of words, such as the family of pain words or house words. They can be considered as marked in the language. The high semantic similarity of compounds in this list could have negatively affected the processing of these groups of compounds.

In sum, my study is novel in that there are no other studies, to the best of my knowledge, that investigated the effect of compound headedness on individuals recall accuracy in an SOV memory task, rather than reading or lexical decision, to maximize sensitivity to spreading activation and competition processes in the mental lexicon. The results of this study provide more evidence on the decompositional routes for compound processing in Persian and are in line with a syntax-based decomposition approach that is suggested for Persian and is supported by the results of other experimental works in this language (see Shabani-Jadidi, 2019; Nojoumian et al., 2006; Megerdoomian, 2001). The Common RH condition showing significantly better recall than the Common LH condition confirmed my hypothesis and suggests that processing of an irregular or a marked structure can negatively affect STM span in Persian. This is in line with findings in English and some other languages, such as Finnish (Service & Tujulin, 2002) and Italian (Arcara et al., 2014), that point to a higher cognitive load for processing marked structures.

References

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. https://doi.org/10.1080/09541440340000123

Arcara, G., Marelli, M., Buodo, G., & Mondini, S. (2014). Compound headedness in the mental lexicon : An event-related potential study. *Cognitive Neuropsychology*, *31*, 164–183.

Awh, E., Vogel, E. K., & Oh, S. H. (2006). Interactions between attention and working memory. *Neuroscience*, *139*(1), 201-208.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. Retrieved from http://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829–839. https://doi.org/10.1038/nrn1201

Badecker, W., & Caramazza, A. (1991). Morphological composition in the lexical output system. *Cognitive Neuropsychology*, *8*(5), 335–367.

Badecker, W., Miozzo, M., & Zanuttini, R. (1995). The two stage model of lexical retrieval evidence from a case of anomia with selective preservation of grammatical gender. *Cognition*, *57*(2), 193–216.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Biran, M., & Friedmann, N. (2011). The representation of lexical-syntactic information : Evidence from syntactic and lexical retrieval impairments in aphasia. *CORTEX*, *48*(9), 1103–1127. https://doi.org/10.1016/j.cortex.2011.05.024

Blanken, G. (2000). The production of nominal compounds in aphasia. *Brain and Language*, *74*(1), 84–102. https://doi.org/10.1006/brln.2000.2338

Boutonnet, B., Mcclain, R., & Thierry, G. (2014). Compound words prompt arbitrary semantic associations in conceptual memory. *Frontiers in Psychology*, *5*(222), 1–7. https://doi.org/10.3389/fpsyg.2014.00222

Camos, V, Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469. https://doi.org/10.1016/j.jml.2009.06.002

Camos, V., Mora, G., & Oberauer, K. (2011). Adaptive choice between articulatory rehearsal and attentional refreshing in verbal working memory. *Memory & Cognition*, *39*(2), 231-244.

Camos, Valérie, & Barrouillet, P. (2014). Attentional and non-attentional systems in the maintenance of verbal information in working memory : the executive and phonological loops. *Frontiers in Human Neuroscience*, *8*(November), 1–11. https://doi.org/10.3389/fnhum.2014.00900

Caramazza, A. (1997). How Many levels of processing are there in lexical access ? *Cognitive Neuropsychology*, *14*(1), 177–208.

Delazer, M., & Semenza, C. (1998). The processing of compound words : A study in aphasia. *Brain and Language*, *61*(1), 54–62.

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.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex.

Foroodi-Nejad, F., & Paradis, J. (2009). Crosslinguistic transfer in the acquisition of compound words in PersianEnglish bilinguals. *Bilingualism*, *12*(4), 411–427. https://doi.org/10.1017/S1366728909990241

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

Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, *8*(1), 127–131. https://doi.org/10.3758/BF03196148

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

Inhoff, A. W., Starr, M. S., Solomon, M., & Placke, L. (2008). Eye movements during the reading of compound words and the influence of lexeme meaning. *Memory and Cognition*, *36*(3), 675–687. https://doi.org/10.3758/MC.36.3.675

Janssen, N., Bi, Y., & Caramazza, A. (2008). A tale of two frequencies : Determining the speed of lexical access for Mandarin Chinese and English compounds. *Language and Cognitive Processes*, *23*(7–8), 1191–1223. https://doi.org/10.1080/01690960802250900

Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., & Libben, G. (1999). Processing compounds: A cross-linguistic study. *Brain and Language*, *68*(1-2), 362-369.

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words. *British Journal of Psychology*, *94*(2), 223–244.

Kahnemuyipour, A. (2014). Revisiting the Persian Ezafe construction : A roll-up movement analysis. *Lingua*, *150*, 1–24. https://doi.org/10.1016/j.lingua.2014.07.012

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7–8), 1089–1132. https://doi.org/10.1080/01690960802193688

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876–895. https://doi.org/10.1037/a0013484

Laudanna, A., Badeckera, W., & Caramazza, A. (1992). Processing inflectional and derivational Morphology. *Journal of Memory and Language*, *31*(3), 333–348.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–75.

Libben, G. (1998). Semantic transparency in the processing of compounds: Consequences for representation, processing, and impairment. *Brain and Language*, *61*(1), 30–44. https://doi.org/10.1006/brln.1997.1876

Libben, G. (2014). The nature of compounds : A psychocentric perspective. *Cognitive Neuropsychology, 31*, 8–25.

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

Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds : Modulation of headedness and semantic transparency. *Journal of Memory and Language*, *66*(4), 644–664. https://doi.org/10.1016/j.jml.2012.01.003

Matzen, L. E., & Benjamin, A. S. (2009). Remembering words not presented in sentences: How study context changes patterns of false memories. *Memory and Cognition*, *37*(1), 52–64. https://doi.org/10.3758/MC.37.1.52

Megerdoomian, K. (2001). Event structure and complex predicates in Persian. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *46*(1-2), 97-125.

Németh, D., Ivády, R., Guida, A., Miháltz, M., Peckham, D., Krajcsi, A., & Pléh, C. (2011). The effect of morphological complexity on short-term memory capacity. *Acta Linguistica Hungarica*, *58*(1), 85-107.

Nojoumian, P., Shabani-Jadidi, P., & Tangestanifar, A. (2006). Morphological priming of multi-morphemic Persian words in the Mental Lexicon. In *The second workshop on Persian Language and Computer* (pp. 29-38).

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. *Journal of Memory and Language, 73*, 15–30. https://doi.org/10.1016/j.jml.2014.02.003

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), 820–833.

RStudio Team (2018). RStudio: *Integrated Development for R. RStudio*, PBC, Boston, MA URL http://www.rstudio.com/.

Sandra, D. (1990). On the representation and processing of compound words : Automatic access to constituent morphemes does not occur on the representation and processing of compound words : Automatic access does not occur. *The Quarterly Journal of Experimental Psychology*, *42*(3), 529–567. https://doi.org/10.1080/14640749008401236

Schmidtke, D., Kuperman, V., Gagné, C. L., & Spalding, T. L. (2015). Competition between conceptual relations affects compound recognition : the role of entropy. https://doi.org/10.3758/s13423-015-0926-0

Semenza, C., & Luzzatti, C. (2014). Combining words in the brain : The processing of compound words Introduction to the special issue, *31*, 1–7.

Service, E., & Tujulin, A. (2002). Recall of morphologically complex forms is affected by memory task but not dyslexia. *Brain and Language*, *81*(1_3), 42–54. https://doi.org/10.1006/brln.2001.2505

Service, E., & Maury, S. (2015). Morphological knowledge and rehearsal in complex span. Paper presented at the 19th Conference of the European Society for Cognitive Psychology, September 17 to 20th, Paphos, Cyprus.

Shabani-Jadidi, P. (2014). *Processing compound verbs in Persian: a psycholinguistic approach to complex predicates* (p. 189). Leiden University Press (LUP).

Shabani-Jadidi, P. (2019) Psycholinguistics. In *the Oxford handbook of Persian linguistics* (pp. 411-435). Oxford: Oxford University Press.

Wälchli, Z. (2016). *Recall of compound words in simple and complex span tasks* (Master's thesis, McMaster University). Available from ProQuest Dissertations & Theses Global database.

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

4 Experiment Three: Primed naming of Persian compound words: Evidence from a reaction time experiment

Abstract

Persian is an Indo-European language in which the head positions of compound words can vary and, therefore, it allows researchers to investigate different recognition, processing, and representations for the head-initial (left-headed or LH) versus head-final (right-headed or RH) compounds. Although there is a significant body of research on compound word processing, the studies examining the effects of different constituents of the compounds as primes in an SOV language with variable head positions are very few up to this date. Linguistic analysis of Persian suggests that the underlying structure of LH noun-noun compounds is different from that of RH compounds (Kahnemuyipour, 2014). This chapter reports the results of a reaction time experiment in which thirty-three individuals took part in an auditorily primed visual naming task. This experiment investigated the lexical representation of Persian compound words in the mental lexicon and the hypothesis that, since right headedness is the default in Persian, the naming times for the natural RH compounds are faster. I also hypothesized that there is an interaction between headedness and prime condition suggesting that priming effects are different for the two types of words. The theoretically predicted effects of both headedness and prime condition were statistically significant and in line with the hypothesis. These results indicated that the left head or the left modifier as primes both help access to the word beginning. However, priming the modifier appeared not to work for the left-headed compounds. These findings also revealed that only in LH compounds, the head primed better than the modifier. Results of this chapter provide more proof for the decompositional routes for the processing of Persian NN compound words in the mental lexicon. Based on the results of this chapter, the framework proposed by Di Sciullo & Williams (1987), and Kahnemuyipour's (2014) linguistic analysis of LH and RH compounds in Persian, I discussed how the interpretations of "hierarchical morphological constructs" and" syntactic words" can be applied to RH and LH compounds in Persian (see Marelli et al., 2009; Marelli et al., 2012 for a similar argument in Italian).

4.1 Introduction

The purpose of this chapter is to provide information on the representation of Persian noun-noun (NN) compound words in the mental lexicon. As mentioned in the previous chapters, Persian RH compounds are the default form in this language. However, LH compounds are semantically more transparent than the RH ones (Kahnemuyipour, 2014). As for productivity, both RH and LH compound processes are active, but there is no existing data on the frequency of these processes in modern Persian. Because new LH compounds can be shaped out of the existing frequent words connected with Ezafe structures and are frequently used by native speakers, it is possible to consider LH compounds to be more productive than RH ones in modern Persian. In the experiment I am presenting here, the main questions were: how does compound headedness affect word naming reaction times? and can presenting the head constituents as primes for Persian RH compounds affect naming speed differently from priming LH ones with their heads?

As we see or hear words, we can instantly repeat, organize, and name them. When we see a familiar word, our brain activates a range of information through the process of lexical access (Carmichael, Hogan, & Walter, 1932). There are different layers of representation being engaged in this process, leading to the question whether these representations become available all at the same time or whether there are different stages of processing for them? In the case of morphologically complex words, we can ask how the different layers of form and meaning converge and make the processing of complex words happen.

The lexical and semantic theories of complex word processing deal with the general debate of whether such words are stored as one unit or if they are decomposed and activated through their morphological constituents. As covered in the previous chapter, at one end of the spectrum is the full listing hypothesis (see Butterworth, 1983), suggesting that words are stored as a whole and in the form of separate entries in the mental lexicon. At the other end are parsing models (Forster & Taft, 1976), proposing that words are fully decomposed into their constituents and are stored as multimorphemic chains and therefore lexical access requires activation of the constituents of multimorphemic words. Alternative theories to these two traditional models suggest, for instance, that activation of the whole word occurs prior to constituent activation (see Giraudo & Grainger, 2001 for a dual route theory). The early versions of dual-route theories proposed parallel functioning of both decomposition and whole word processes, based on word and constituent frequencies. This means that the highly frequent words would be processed as whole units. However, low-frequency whole words with highly frequent morphemes, would be subjected to a decomposition process (Semenza & Luzzatti, 2014). Some recent versions of these theories suggest a flexible operation even from the earliest moments of the processing of complex words. Recognition, based on such a model, relies on both whole word and constituent information (see Kuperman, Schreuder, Bertram, & Baayen, 2009). All of these theories were presented as explanations of the processing of specific languages being studied. Many different factors have been put forward as involved in the processing of complex words. Such factors are word frequency, semantic transparency, lexical category, morphological type, and the relationship between the constituents of the complex words (Libben, 1998).

Compounding of words represents one of the core strategies of information packaging in our mental lexicon. Many researchers have emphasized the importance of compound words. For example, Jackendoff (2002) considers compounds as protolinguistic fossils from which more complex linguistic constructions have been developed. Libben (2014) states that the knowledge of neurological connections underlying compound word conception is key to identifying the cognitive mechanisms for morphological and lexical processing and cognitive representation of human language in general. In NN compound words, Headedness is defined as the order of the syntactic roles (modifier and head). Depending on the language, compounds can be either head initial (left-headed) or head final (right-headed). If our brain processes the morpheme-based representations of compound words starting from the earliest milliseconds of word processing, we can ask which of the compound's constituents is being recognized and processed first. Is it always the first constituent that is processed first, or does the syntax of the language being studied play a role in this context? This has been studied using priming paradigms.

If complex words are decomposed into their constituents, while going through processing, then presenting their parts briefly before full presentation can facilitate their processing. This effect is called the priming effect. If a word is repeated (repetition priming) or if it is preceded by a semantic associate (semantic priming), the word recognition performance has been found to be facilitated. In constituent priming paradigms, a target compound (e.g., blackberry) is preceded by one of its constituents' presentation (e.g., berry). In masked priming paradigms, a visual prime is shown very briefly and is not consciously accessible. Masked priming is a great method to assess the hypothesis that visual word recognition relies on a competitive process (Davis, 2003). Masked priming eliminates any conscious recognition of the relationship that exists between the target and its prime. Studies with constituents as masked primes preceding compound words and their morpheme constituents have revealed that the head has a more influential role on processing than the modifier although the first constituent has also a special role because of its word-initial position (see e.g., Jarema, Busson, Nikolova, Tsapkini, & Libben, 1999).

Studying compounds in a language like English, in which the position of the head is always fixed, does not provide an opportunity to examine the effects of headedness, i.e. the potentially separate effects of constituent position and syntactic role, on the processing of complex words.[10] There are only a few psycholinguistic studies investigating the effects of headedness on compound word processing in languages that allow for variable head positions (see Duñabeitia, Perea, & Carreiras, 2007, for Basque and Spanish; Jarema et al., 1999, for French). Some studies have found that the effects of constituent position are more important; some report headedness to play the major role (head more important), and there are other studies that reported interactions (it depended on the structure of the compounds). Duñabeitia et al. (2007), examined frequency effects in Basque and Spanish. Their results revealed that the second constituent frequency effects were similar in both languages. Considering that Spanish compounds are mainly right-headed and Basque compounds are mainly left-headed, they chose to propose second position (and no headedness) effects. The results of the study by Jarema et al. in French showed significant effects of priming on facilitating compound recognition in French, revealing a greater priming effect of the initial constituent than the final constituent in head initial (LH) French compounds. These results point to the combined effects of headedness and constituent position.

As mentioned in chapter three, research in Italian confirms the influence of headedness on compound word processing (Arcara, Marelli, Buodo, & Mondini, 2014; El Yagoubi, Chiarelli, Mondini, Perrone, Danieli, & Semenza, 2008; Marelli et al., 2009). Marelli et al., (2009) found greater effects of head primes than modifier primes in right-headed compounds. They mentioned that this might be due to the assumed default right headedness in Italian, which means that RH compounds in this language are more efficiently accessed than LH ones. This argument assumes that the head-modifier structure is different from the canonical order of the NPs in this language

[10] See Hyönä & Pollatsek (1998) for results showing first constituent is more important and Juhasz, Starr, Inhoff, & Placke (2003) for results showing that the second constituent is more important.

(where the modifier usually follows the head). Arcara et al., (2014) studied Italian head-initial (LH) and head-final (RH) compounds as visual stimuli in an event-related potential (ERP) experiment to investigate if head position affects processing of compound words. They found enhancement of ERP responses suggesting that processing head-final (RH) compounds required more effort, i.e., there was an extra processing cost, when the compounds were presented as two words with a space between the constituents. This was thought to result from their internal compounded modifier-head structure contrasting with a syntactic NP-like structure for the LH compounds. Accepting these authors' argument that right headedness is syntactically marked but the standard structure for NN compounds in Italian[11], these results are in line with my findings for Persian, presented in chapter three. My memory recall experiment data revealed that the recall of head final (RH) ones, which is the dominant pattern of compounds in Persian, was higher than of head initial (LH) compounds. Both results in Persian and Italian also indicate that the mental representations for head-initial and head-final compounds are different.

Now, how are lexical representations structured, processed, and stored in Persian? Persian presents a case of compound word characteristics that differ from those of Germanic or Romance languages. The default structure in Persian NN compounds is right headedness. However, as mentioned in the introduction, surface headedness in different languages has different deep structures and these lead to different processing consequences for these languages. Marelli et al., (2009) mentioned, based on the framework proposed by Di Sciullo and Williams (1987), that in Romance languages the head-initial (LH) compounds are "syntactic words" with morphologically flat structural representations. The authors explained that these compounds are "syntactic strings imported into the lexicon, a juxtaposition of words without a real morphological hierarchy" (Marelli et al., 2009, p. 445), whereas their observed priming-effect pattern for head-final compounds (i.e., head primes facilitate more than the modifier primes) suggests an underlying internal head-modifier hierarchy for RH compounds showing that only RH compounds in Romance languages are hierarchical morphological constructs (see also Marelli et al., 2012). Let us compare this with Persian. The word order in noun phrases of Modern Persian is one in which the modifier follows the head noun. Persian LH compounds have been argued to be shaped out of

[11] Marelli and Luzzatti (2012) mention that the common assumption of right headedness being the default in Romance languages, such as Italian, is contradictory. However, both RH and LH compounds in Italian are equally standard and neither one can be considered exceptional. This conflicts with Arcara et al., (2014) and Scalise (1984) claims, which mention left-headed NN compounds are the dominant pattern in Italian.

two nouns attached with Ezafe construction where the frequency of use has caused the Ezafe to be dropped over time (Kahnemuyipour, 2014). Therefore, this group of compounds in Persian have more syntactic structure and hierarchy than their RH counterparts. However, it is important to note that the existence of Ezafe, which is a unique construction in Persian noun phrases, in the foundation of Persian LH compounds makes them different from LH compounds in Romance languages. This embedded unique structure in Persian LH compounds can be the reason for their slower processing than the RH ones. The right-headed compounds in Persian are considered as 'true' compounds since they reflect the dominant pattern of NN compounds in the language (Kahnemuyipour, 2014). However, Kahnemuyipour (2014) also proposed that the constituent order of RH compounds in Persian (where the head follows the modifier) is the syntactic "base order" of nouns and modifiers in this language. He considers this compound formation as the result of a "morphological merger" (see Kahnemuyipour, 2014. p.6 for more details). Therefore, although less than in LH compounds, there is also a syntactic structure in Persian RH compounds. Based on the above mentioned points, neither RH or LH compounds in Persian can be considered as flat structures with no hierarchy (see Di Sciullo and Williams 1987; Marelli & al., 2009; Marelli & al., 2012, for this argument in Romance languages).

All of the above mentioned points show that the deep structure of compounds in Persian provides a unique opportunity to examine how the syntactic structure of compounds, the order of the elements in Persian noun phrases, and the surface constituent order of compounds affect the processing of compounds in the native speakers' brain. I hypothesize that the higher semantic transparency and the underlying Ezafe construction, can make the processing of LH compounds more complex than the RH ones in Persian. The results of my memory recall experiment (chapter three) showed that for native Persian speakers, head final NN compounds in Persian were easier to remember. Here I aim to investigate the role of constituent position and headedness in output creation: which constituent of RH and LH compounds in Persian produces a stronger priming effect on the whole word representations. To do so, I designed a naming experiment with cross-modal priming in which participants' reaction times to LH versus RH compounds were analyzed.

4.2 Naming latencies and cross-modal priming

Observation of naming latencies is one among several methods that can be used for word recognition research. Balota et al., (2004) note that there are disadvantages to other well-known methods such as the use of eye tracking in reading experiments. When the recognition of isolated words is studied, one limitation of reading methods is that, when reading, there are multiple sources of information available. These can include syntactic, semantic, and parafoveal visual information (Balota et al., 2004, p.283). Therefore, psycholinguistic researchers usually rely on lexical decision or naming performance. In visual naming tasks, participants are asked to name the presented visual word or non-word, as quickly and accurately as they can. Jared and Seidenberg (1990) mention that naming latencies are interesting tasks as when we learn how to pronounce the visual words aloud, we learn how to read (see also West & Stanovich, 1986). Another reason, they mention, is that impairments of word naming are often associated with acquired dyslexia that is developed after a brain injury. Most of the existing naming literature in visual word recognition has been focused on monosyllabic words, naming latencies of multimorphemic words has only become popular more recently.

For the present experiment, I chose to use a naming task, not a lexical decision task, because of the existence of the Ezafe structure in Persian. As I mentioned before, the Ezafe structure has a phonological form, which is a short vowel added in between the two words attached, and is usually not written in Persian. Because of that, it is impossible to distinguish in written text left-headed compound words from constructions of two nouns attached with Ezafe. For the same reason, I used cross-modal priming with auditory primes and visual target stimuli.

Research on priming shows that prior experience with a certain stimulus can lead to unconscious facilitation of speed, accuracy, and efficiency of stimulus recognition in the perceptual system (see Greene, Easton, & LaShell, 2001; Jacoby, 1983). This effect is evident for haptic, visual, and auditory perception and is called priming (see Segal, 1966; Easton, Srinivas, & Greene, 1997). Previous priming work has provided evidence of constituents' activation during compound processing (see for example Isel, Gunter, & Friederici, 2003, for a review). Monsell (1985) was the first to report equal repetition priming effects for both constituents of compounds. Sandra (1990) showed a facilitation effect of primes that are semantically related to one of the constituents of transparent compounds (e.g., milkman). Similarly, Zwitserlood (1994) designed an experiment in which compounds were presented for 300 ms and were followed by one of their constituents (e.g., milk after milkman). This immediate partial repetition priming paradigm resulted in similar priming effects for the first and second constituent. The researcher then used a semantic priming paradigm and replicated the results. It is important to note that these results can all be considered as morphological or orthographic priming effects. In a similar study, Jarema, Busson, Nikolova, Tsapkini, and Libben (1999) examined the processing of French and Bulgarian compound words. French compounds are mostly left-headed whereas Bulgarian compounds are right-headed. The researchers reported that the first constituent as a prime had a greater facilitatory effect on recognition than the second constituent for the French words. The authors thought this resulted from the added effects of first position and head role. Libben, Gibson, Yoon, and Sandra (2003) carried out a similar study in English and reported similar priming effects of first and second constituent on compound word recognition.

In line with these findings are results from eye tracking studies, pointing to serial activation of compounds (e.g., Juhasz et al., 2003; Andrews et al., 2004). These results are from reading experiments showing that the first constituent is accessed first and then the second one is accessed in a serial way. These results add to the body of evidence that shows greater priming effects of the first constituent of compound words compared to the second one (see also Niswander-Klement & Pollatsek, 2006; Bertram & Hyönä, 2003). It is important to note that in a language like English, which is concatenative and in which there is a linear manner of adding prefixes and suffixes to the root, semantic transparency plays an important role in the processing of complex words. As mentioned in chapter one, in English, more semantically transparent complex words (e.g., *government*) can prime their root (*govern* in this case), but their opaque counterparts (e.g., *apartment*) can't prime their root (*apart* in this case). However, Libben (1994) has proposed that in English, semantic transparency works at the level of lexical representation and not morphological parsing (see also Libben et al., 2003). Therefore, in the case of compound words, the existence of semantic transparency can be defined as a relationship between the whole word representation and the representation of its constituents as separate units at the lexical representation level. Hence, in English, multimorphemic strings (e.g., *discoverable*) act quite differently from compound words. Libben (2003) explains that even for semantically transparent compounds in English (e.g., *football*), the whole word meaning cannot be predicted from the constituent meanings.

It is reported that in languages like Arabic, which is non-concatenative, even in unmasked priming designs, semantic transparency does not have an influence on compound processing (Shabani-Jadidi, 2019; see also Boudelaa & Marslen-Wilson, 2000). This means that even for opaque compounds, there is a facilitatory priming effect in Arabic on the target in cross-modal or auditory-auditory priming paradigms. In contrast, in English opaque pairs, the facilitatory effect can only be observed in masked priming paradigms (Boudelaa 2013). The observed priming effect in both semantically opaque and transparent Arabic compounds can be explained by the morphological richness of the language and that the high number of complex morphological words in such languages makes complex words less marked and more frequent (e.g., Shabani-Jadidi, 2014 for Persian and Smolka et al., 2009 for German). Shabani-Jadidi (2019) mentions that since Persian is a morphologically rich language, native speakers break the words into their constituents in a linear manner irrespective of the remaining element (i.e., whether it is a word or a morpheme existing in the language). In a masked priming study on complex predicates in Persian (Shabani-Jadidi, 2014), the author points to an early decomposition route for Persian complex structures. She explains that as the parser encounters any decomposable element in the language, it separates that unit and proceeds to the next element. In Persian, semantic transparency does not seem to play a significant role in compound word processing because complex words are not marked due to the morphological richness of the language (Shabani-Jadidi, 2019).

Based on the above-mentioned points, we can see that there is still a debate on which constituent of the compound words plays a more important role on their processing. Most of the reported studies on compound words have been done in languages in which the position of the head is usually fixed. Persian is a morphologically rich language from the same family as many European languages. However, it has a different internal order of the elements and a different compound structure than most of the other languages in this family. Furthermore, the existence of the unique Ezafe structure, which has a surface form in noun phrases, but may also provide a deep structure for LH compound words, makes the study of NN compounds in this language especially interesting. Naming latency experiments with constituent priming can inform about the role of syntactic structure in compound NN processing in Persian and more generally.

I used a cross-modal priming paradigm, where participants heard a prime word and then saw a target word that they were asked to name. Cross-modal priming offers an opportunity to examine perceptual interactions before integration of concepts happens. There has been a debate on the mechanisms underlying cross-modal facilitation of orthographic and phonological features. Previous research reports the transferability of some but not all information across modalities (see for example Greene, Easton, & LaShell, 2001; Kirsner, Dunn, & Standen, 1989). Although visual and auditory words have different perceptual features (orthographic for visual words and phonological for auditory words), there are studies proposing that cross-modal facilitation can be explained by a perceptual sharing mechanism, which means modalities interact and there is information-sharing (Greene, Easton, & LaShell, 2001, for an argument on theories of cross-modal priming). Other studies have claimed that lexical and phonological codes have a mediating effect on visual–auditory priming before semantic processing occurs (e.g., Komatsu & Naito, 1992; Curran, Schacter, & Galluccio, 1999). These studies base their hypothesis on perceptual representations assumed not to be specific to one modality or processing system. The mentioned processes and the nature of those representations are still under debate. Although some neurophysiological theories assume processing streamed through different perceptual systems, there are many nervous system structures that can respond to objects and events in different modalities (see Stoffregen & Bardy, 2001; Fishman & Michael, 1973). While there might be different explanations for cross-modal priming, the most parsimonious theory is suggested by Greene, Easton, & LaShell (2001). They explain that when there is an advantage in sharing, at the level of surface feature processing, the perceptual systems can interact.

In my experiment, the primes were either the first or second constituents of the target compound words presented auditorily to the participants (e.g., *aab* or *porteqaal* in *aab porteqaal,* “water orange”, meaning ‘orange juice’). Based on the assumption of different syntactic and morphological structures in RH and LH compounds in Persian, I was interested to see if the head in LH compounds can prime better than the head in RH compounds or vice versa. I mentioned that RH compounds are default in Persian. In my STM experiment, RH compounds were significantly better recalled than the LH ones. Therefore, I expected my RH compounds to be faster to name. The results of this experiment can also shed more light on whether headedness or constituent order plays a more important role in the processing of compound words. In addition, the experiment can reveal different patterns for RH and LH compounds based on their deep structure: an interaction between headedness and prime constituent position (first or second) would be expected. Some previous research has suggested that the first constituent in a NN compound can facilitate lexical access. There may, therefore, be a main facilitating effect of first constituent priming. However, second constituent priming could depend on compound headedness. The results may also provide support for decomposition processes in accessing the representations of Persian NN compound words in the mental lexicon.

4.3 Methods

4.3.1 Participants

There were 33 participants in the experiment (15 females). The participants were all native Persian speakers between the ages of 20 and 40; and reported to have no visual, language, learning or hearing problems. They were compensated with $15 CAD dollars for one hour of their time.

4.3.2 Word stimuli

I used 118 Persian NN compounds as my stimuli. These words were taken from a database of Persian compound words, described in chapter two, which provided ratings of familiarity, imageability, and age of acquisition (AOA) for these compounds (see chapter two for more details). In the data analysis, the rated lexical variables were used as control variables. Each one of the 118 compounds was once primed auditorily with their left and once with their right constituent. This means that each participant encountered 236 compound stimuli, presented in one of four different orders (four different scripts). Half of the stimuli were left-headed and the other half right-headed. The auditory primes were recorded in the Centre for Advanced Research in Experimental and Applied Linguistics' (ARiEAL) phonetic lab at McMaster University. All the sound files were recorded by a native Persian speaker and were controlled for acoustic loudness, fade-in, and fade-out effects.

4.3.3 Procedure of naming experiment

Testing was performed in the Language Memory and Brain Lab in the ARiEAL Research Centre, McMaster University. The length of the experiment was between 45 and 60 minutes. Printed instructions for the task were presented on paper. Participants completed and signed a consent form that was approved by the McMaster Research Ethics Board. After the experiment, they received a debriefing information form about the study's main research hypothesis.

The experiment was a cross-modal auditory-visual primed naming task controlled by an iMac computer. Before seeing each one of the NN compounds, in the center of the screen, participants heard either the first or the second constituent of that word from the computer speakers. They were instructed to verbally name the word as soon as they saw it on the screen, which was placed approximately 60 cm from the participants. Participants were asked to respond as fast and as accurately as possible. Participants were informed that they are allowed to take a break and stop in between the stimuli. There were programmed breaks after every 50 words.

The experiment was programmed with Super lab 5 by Cedrus Corporation. The stimulus-onset asynchrony (SOA) was 1 second between the onsets of the auditory primes and the presentation of the visual stimuli. A voice key was employed to record the naming latencies. However, because of a programming error, the voice key data could not be used. However, the trials with the auditory primes and the participants' voice were also recorded using Audacity software as a backup procedure. The reaction times were measured from these visualized sound files.

4.3.4 Data analysis

The reaction time data for each real compound word was merged with the lexical variables for that word, taken from chapter two, which includes ratings of familiarity, imageability, and AOA in Persian NN compounds. The first measure was to average the RTs, which were grouped by participants, for descriptive purposes. Then I merged the trial-by-trial RTs with the data with the mean ratings of the fixed lexical variables to allow the investigation of the effects of these control variables on the compound word naming reaction times. Then I employed a linear mixed effects model to examine the effects of headedness and prime conditions on the naming RTs. Since right headedness is the default in Persian and based on the data presented in chapter three, I expected faster naming times for the natural RH compounds and an interaction between headedness and prime condition to suggest that priming effects are different for the two types of words.

4.4 Results

Table 4-1 shows the average naming reaction times (RTs) over participants to be 652 with the standard deviation of 112. Based on the 3rd Quartile (Q3) presented in *Table 4-1*, 75% of participants had a mean reaction time below 718 ms. *Figure 4-1* shows the density of the reaction time data.

Table 4-1 Descriptive statistics of naming time data grouped by participants

	1st Qu.	Mean	SD	3rd Qu.
Reaction time	575	652	112	718

SD = standard deviation. 1st Qu.= First quartile(Q1). 3rd Qu. = Third quartile(Q3).

The naming RT data were analysed using a linear mixed effects model (lme4, R version 3.5.2, R Foundation for Statistical Computing, 2018). Cut-off points of 400 ms as minimum and 1000 ms as maximum were used for filtering the reaction time data (reaction time > 400 and < 1000). The fixed variables in the analysis were *headedness* (LH vs. RH) and *prime condition (Left constituent vs. Right constituent)*. The random variables in my analysis were the *participants* and the *compound items*. The effects of some known lexical variables were controlled by including the control variables (fixed lexical variables), i.e., ratings of *familiarity*, *imageability*, and *AOA* for each compound stimulus. The mixed model fit was done by restricted maximum likelihood (REML) statistics. The model R^2 was 0.43. The results of this analysis are reported in *Table 4-4*.

Table 4-2 shows the marginal mean RTs based on headedness. The results demonstrate that the average of RTs for RH compounds is lower than for LH compounds. These results add evidence to my hypothesis that, due to right-headedness being the default in Persian, RH compounds are processed faster than their LH counterparts in this language. One reason for RH compounds being processed faster than LH ones can be that this group of compounds are processed more as a whole while the higher semantic transparency of LH compounds, which is due to the embedded Ezafe structure in them, leads to the decomposition of constituents in this group of compounds and this can lead to an overall slower processing.

Table 4-3 and *Figure 4-2* show the reaction time results based on the different headedness and prime conditions. A post hoc analysis with Bonferroni corrections confirmed that the right-headed compounds primed with their right (head) constituents or left (modifier) constituents had significantly shorter reaction times than the left-headed compounds primed with their right (modifier) constituents, (see *Table 4-3* and *Figure 4-2*). The difference between the right-headed

Table 4-2 Estimated marginal means showing the significant effect of compound headedness on naming times

	Headedness	**Mean**	**SEM**	**df**
Reaction time	right	642	11.9	49.0
	left	663	11.9	48.3

df= degrees of freedom. SEM= standard error of the mean

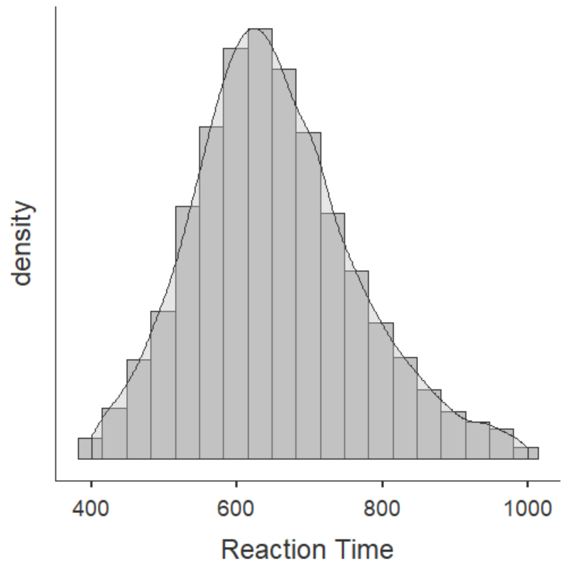

Figure 4-1 Naming time density distribution

compounds primed by their right heads and the left-headed compounds primed by their left heads was not significant. Of the lexical control variables, only familiarity had a significant effect on naming times. There was a significant interaction between headedness and prime condition. Based on the previous studies, we can assume that priming effects here are all facilitative, making naming faster, not slower. When the left constituent is the prime, it looks like there is priming irrespective of whether it is the head or not. When the right constituent is the prime, it makes the naming faster if it is the head of the compound (see *Figure 4-2)*.

Table 4-3 Reaction time results for the compound word targets based on Headedness and Prime condition

Headedness	Prime Condition	M	SD
left	Left Constituent	656	114
left	Right Constituent	671	117
right	Left Constituent	638	105
right	Right Constituent	642	109

SD = standard deviation. M = mean

Table 4-4 Fixed Effect Parameter Estimates

				95% Confidence Interval				
Names	**Effect**	**Estimate**	**SEM**	**Lower**	**Upper**	**df**	**t**	**p**
(Intercept)	(Intercept)	652.17	11.31	630.00	674.34	39.9	57.647	< .001
Prime Condition1	Right Constituent - Left Constituent	8.80	2.04	4.80	12.79	7471.8	4.318	< .001
Headedness1	left - right	20.75	7.53	5.99	35.50	149.7	2.756	0.007
Familiarity	Familiarity	-13.68	6.39	-26.20	-1.16	117.3	-2.141	0.034
Imageability	Imageability	11.58	8.46	-5.00	28.17	116.4	1.369	0.174
AOA	AOA	1.54	1.58	-1.57	4.64	116.6	0.969	0.335
Prime Condition1 ✻ Headedness1	Right Constituent - Left Constituent ✻ left - right	17.19	4.06	9.23	25.14	7479.2	4.236	< .001

SEM= standard error of the mean. df= degrees of freedom. p= p-value

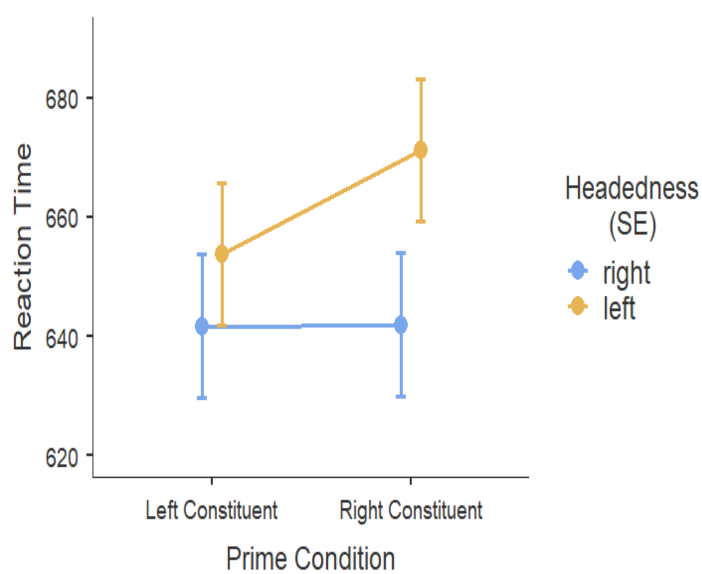

Figure 4-2 Linear mixed model results showing the effects of headedness and prime condition on reaction times. Error bars are standard errors of the mean.

4.5 Discussion

This study examined the effects of compound headedness and prime condition (left vs. right constituent prime) on naming latencies for Persian NN compound words. Persian allows for different head positions in NN compound words. Left headedness is marked in the language and a linguistic analysis of left-headed compounds suggests that these compounds have a different underlying structure involving the so-called Ezafe construction (see Kahnemuyipour, 2014). The naming reaction times presented in this chapter revealed that their overall mean was significantly shorter for RH compounds than LH ones. These results were expected and confirmed my hypothesis. Based on the linguistic analysis of RH and LH compounds in Persian, I hypothesized that right-headed compounds will be processed faster than left-headed ones.

I also hypothesized that there will be an interaction between headedness and prime condition suggesting that these two different types of compounds in Persian will be affected differently by constituent priming. The theoretically expected effects of both headedness and its interaction with prime condition were statistically significant. The results could be interpreted as suggesting that the right head as a prime boosts access to the whole compound. The left head or the left modifier as primes both help access to the word beginning. However, priming the modifier did possibly not work for the left-headed compounds because it is not as solidly attached to the head as the modifier is in the "real" RH compound words that do not have an underlying Ezafe structure.

As I mentioned in the introduction of this chapter, surface headedness in different languages originates from different deep structures and these lead to different processing consequences for these languages. For example, Italian also allows for both RH and LH compounds. Marelli et al., (2009) followed the framework proposed by Di Sciullo and Williams (1987) and proposed that in Romance languages the head-initial (LH) compounds can be considered as "syntactic words" with flat structural representations. They mentioned that LH compounds in Romance languages are "syntactic strings imported into the lexicon, a juxtaposition of words without a real morphological hierarchy" (Marelli et al., 2009, p. 445). Their results indicated that only in head-final compounds, do head primes facilitate more than the modifier primes. They interpreted these results based on Di Sciullo and Williams's (1987) framework suggesting an underlying internal head-modifier hierarchy for RH compounds showing that only RH compounds in Romance languages are hierarchical morphological constructs (see also Marelli et al., 2012). Persian LH compounds are originally developed out of two nouns connected with Ezafe construction where the Ezafe is assumed to have been dropped due to the frequency of use (Kahnemuyipour, 2014). I mentioned in the introduction of this chapter that this group of compounds in Persian can be considered to have more syntactic structure than the RH ones. However, the existence of Ezafe, which is a unique construction in Persian noun phrases, in the foundation of Persian LH compounds creates a different case for LH compounds in Persian than LH compounds in Romance languages. As mentioned before, priming the modifier could possibly not work for the left-headed compounds because it is not as solidly attached to the head as the modifier is in the "real" RH compound words that do not have an underlying Ezafe structure. My results appear to demonstrate the opposite of the findings of Marelli et al. (2009), who reported that heads primed more than modifiers for head final compounds. My findings showed that only for head initial compounds, did heads prime more than modifiers. However, as my task was naming rather than lexical decision, phonological facilitation could have played a role in the first constituent effects whereas morphological structure differences would have been better revealed in the second constituent effects. My theoretical assumption that the head final or RH compounds in Persian are considered as 'true' compounds since they represent the dominant pattern of NN compounds in the language (Kahnemuyipour, 2014) remains in line with the shorter response times for the RH compound primed with their heads compared to the LH compounds primed with their modifiers when there is no phonological priming of the word beginning. In Persian noun phrases, as in those in Italian, the modifier follows the head noun. However, Kahnemuyipour (2014) has proposed that the constituent order of RH compounds in Persian (where the head follows the modifier) is also the "base order" of nouns and modifiers in the phrasal syntax of this language. According to him, RH compound formation is the result of a "morphological merger" (see Kahnemuyipour, 2014. p.6 for more details). Therefore, Kahnemuyipour's analysis of compounds in Persian can be considered in line with Di Sciullo and Williams (1987); Marelli et al., 2009; Marelli et al., 2012 showing that RH compounds can also be considered as hierarchical morphological constructs.

Considering that LH compounds in Persian are semantically more transparent than RH ones (Kahnemuyipour, 2014), the results of this chapter, showing that RH compounds had faster naming than LH ones, are in line with Shabani-Jadid (2019) proposal pointing that in morphologically rich languages like Persian, semantic transparency does not have a significant role in compound words processing. I propose, based on the reaction times and memory recall results presented in chapter three, that although LH compounds are semantically more transparent in Persian, due to the existence of Ezafe structure embedded in LH compounds, this group of compounds are associated with a heavier cognitive load and, therefore, their processing is slower in the language.

Different naming latencies for Persian LH versus RH compounds are also in line with other compound studies in Persian that point to a decompositional route during processing (Shabani-Jadidi, 2019; Nojoumian et al., 2006; Megerdoomian, 2001). Similar suggestions are made by studies in English, revealing that unless the constituents are semantically opaque, compound words are not processed as whole words in the mental lexicon (e.g., Wälchli, 2016; Boutonnet et al., 2014).

In sum, this study is novel in that there are no other priming studies, to the best of my knowledge, that investigated the effects of different compound constituents as primes in a naming latency task in Persian. I used auditory primes to avoid the possible interference of Ezafe structure in left-headed compounds. My findings show different processing for RH versus LH compound words in Persian and add more evidence to the decompositional route of compound processing in this language, which was also suggested by other experimental results in Persian (e.g., Megerdoomian, 2001; Nojoumian et al., 2006; Shabani-Jadidi, 2019). Also, the longer reaction times for LH compounds, as a marked structure in Persian NN compounds, and in line with my results from chapter three, showing a poorer recall for this group of compounds, adds more evidence to the existing literature in other languages, suggesting that the processing of marked structures is cognitively harder than that of regular ones. (see for example Service & Tujulin, 2002 for Finnish and Arcara et al., 2014 for Italian). My priming results revealed that unlike the results for RH compounds, in LH compounds, the head primed better than the modifier. This is a novel finding for Persian compound words. This also means that priming by the modifier may not work for the left-headed compounds. This can be explained if the modifier is not solidly attached to the head, as the modifier is in the "real" RH compound words that do not have an underlying Ezafe structure. Findings of this chapter also shed more light on the importance of the Ezafe structure as a unique construction in Persian noun phrases, which causes syntactic complexity in left-headed compounds and, therefore, results in LH compound words in Persian being harder to access, represent in working memory, and process (see Kahnemuyipour, 2006; 2014; Ghomeshi 1997; Karimi & Brame, 2012 for more details on Persian Ezafe construction).

References

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. https://doi.org/10.1080/09541440340000123

Arcara, G., Marelli, M., Buodo, G., & Mondini, S. (2014). Compound headedness in the mental lexicon : An event-related potential study. *Cognitive Neuropsychology, 31*, 164–183.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283.

Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when reading short and long Finnish compounds. *Journal of memory and language*, *48*(3), 615-634.

Boudelaa, S., & Marslen-Wilson, W. D. (2013). Morphological structure in the Arabic mental lexicon: Parallels between standard and dialectal Arabic. *Language and cognitive processes*, *28*(10), 1453-1473.

Boudelaa, S., & Marslen-Wilson, W. D. (2000). Non-concatenative morphemes in language processing: Evidence from Modern Standard Arabic. In *ISCA Tutorial and Research Workshop (ITRW) on Spoken Word Access Processes*.

Butterworth, B. (1983). *Lexical representation* (pp. 257-294). na.

Carmichael, L., Hogan, H. P., & Walter, A. A. (1932). An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of experimental Psychology*, *15*(1), 73.

Curran, T., Schacter, D. L., & Galluccio, L. (1999). Cross-modal priming and explicit memory in patients with verbal production deficits. *Brain and Cognition*, *39*(2), 133-146.

Davis, C. J. (2003). Factors underlying masked priming effects in competitive network models of visual word recognition. *Masked priming: The state of the art*, 121-170.

Di Sciullo, A. M., & Williams, E. (1987). *On the definition of word* (Vol. 14). Cambridge, MA: MIT press.

Duñabeitia, J. A., Perea, M., & Carreiras, M. (2007). The role of the frequency of constituents in compound words: Evidence from Basque and Spanish. *Psychonomic Bulletin & Review*, *14*(6), 1171-1176.

Easton, R. D., Greene, A. J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin & Review*, *4*(3), 403-410.

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.

Fishman, M. C., & Michael, C. R. (1973). Integration of auditory information in the cat's visual cortex. *Vision research*, *13*(8), 1415-1419.

Forster, I., & Taft, M. (1976). Lexical Storage and Retrieval and Polysyllabic. *Journal of Verbal Learning and Verbal Behavior*, *15*(6), 607--620.

Ghomeshi, J. (1997). Non-projecting nouns and the ezafe: construction in Persian. *Natural Language & Linguistic Theory*, *15*(4), 729-788.

Giraudo, H., & Grainger, J. (2001). Priming complex words: Evidence for supralexical representation of morphology. *Psychonomic Bulletin and Review*, *8*(1), 127–131. https://doi.org/10.3758/BF03196148

Greene, A. J., Easton, R. D., & LaShell, L. S. (2001). Visual–auditory events: cross-modal perceptual priming and recognition memory. *Consciousness and Cognition*, *10*(3), 425-435.

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.

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.

Jacoby, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of verbal learning and verbal behavior*, *22*(5), 485-508.

Jared, D., & Seidenberg, M. S. (1990). Naming multisyllabic words. *Journal of Experimental Psychology: Human perception and performance*, *16*(1), 92.

Jarema, G., Busson, C., Nikolova, R., Tsapkini, K., & Libben, G. (1999). Processing compounds: A cross-linguistic study. *Brain and Language*, *68*(1-2), 362-369

Juhasz, B. J., Starr, M. S., Inhoff, A. W., & Placke, L. (2003). The effects of morphology on the processing of compound words. *British Journal of Psychology*, *94*(2), 223–244.

Kahnemuyipour, A. (2006). Persian Ezafe construction: Case, agreement or something else. In *Proceedings of the 2nd Workshop on the Persian Language and Computer. Tehran: University of Tehran*.

Kahnemuyipour, A. (2014). ScienceDirect Revisiting the Persian Ezafe construction : A roll-up movement analysis. *Lingua*, *150*, 1–24. https://doi.org/10.1016/j.lingua.2014.07.012

Karimi, S., & Brame, M. (2012). A generalization concerning the EZAFE construction in Persian. *Linguistic Analysis*, *38*(1), 111-144.

Kirsner, K., Dunn, J. C., & Standen, P. (1989). Domain-specific resources in word recognition. *Implicit memory: Theoretical issues*, 99-122.

Komatsu, S. I., & Naito, M. (1992). Repetition priming with Japanese Kana scripts in word-fragment completion. *Memory & cognition*, *20*(2), 160-170.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic

Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 876–895. https://doi.org/10.1037/a0013484

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. https://doi.org/10.1006/brln.1997.1876

Libben, G. (2003). Morphological parsing and morphological structure. In *Reading complex words* (pp. 221-239). Springer, Boston, MA.

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

Marelli, M., Crepaldi, D., & Luzzatti, C. (2009). Head position and the mental representation of nominal compounds: A constituent priming study in Italian. *The Mental Lexicon*, *4*(3), 430-454.

Marelli, M., & Luzzatti, C. (2012). Frequency effects in the processing of Italian nominal compounds: Modulation of headedness and semantic transparency. *Journal of Memory and Language*, *66*(4), 644-664.

Megerdoomian, K. (2001). Event structure and complex predicates in Persian. *Canadian Journal of Linguistics/Revue canadienne de linguistique*, *46*(1-2), 97-125.

Monsell, S., & Ellis, A. W. (1985). Progress in the psychology of language.

Nikolova, R., Tsapkini, K., Busson, C., & Libben, G. (1999). Processing compounds : A cross-linguistic study. *Brain and Language*, *68*(1–2), 362–369.

Niswander-Klement, E., & Pollatsek, A. (2006). The effects of root frequency, word frequency, and length on the processing of prefixed English words during reading. *Memory & Cognition*, *34*(3), 685-702.

Nojoumian, P., Shabani-Jadidi, P., & Tangestanifar, A. (2006). Morphological priming of multi-morphemic Persian words in the Mental Lexicon. In *The Second Workshop on Persian Language and Computer* (pp. 29-38).

RStudio Team (2018). RStudio: *Integrated Development for R. RStudio,* PBC, Boston, MA URL http://www.rstudio.com/.

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. https://doi.org/10.1080/14640749008401236

Scalise, S. (1984). Generative Morphology. Dordrecht: Foris.

Segal, S. J. (1971). Processing of the stimulus in imagery and perception. In Imagery (pp. 69-100). Academic Press.

Semenza, C., & Luzzatti, C. (2014). Combining words in the brain : The processing of compound words Introduction to the special issue, *31*, 1–7.

Service, E., & Tujulin, A. (2002). Recall of Morphologically Complex Forms Is Affected by Memory Task but Not Dyslexia. *Brain and Language*, *81*(1_3), 42–54. https://doi.org/10.1006/brln.2001.2505

Shabani-Jadidi, P. (2014). *Processing compound verbs in Persian: a psycholinguistic approach to complex predicates* (p. 189). Leiden University Press (LUP).

Shabani-Jadidi, P. (2019) Psycholinguistics. In *The Oxford Handbook of Persian Linguistics* (pp. 411-435). Oxford: Oxford University Press.

Smolka, E., Komlosi, S., & Rösler, F. (2009). When semantics means less than morphology: The processing of German prefixed verbs. *Language and Cognitive Processes*, *24*(3), 337-375.

Stoffregen, T. A., & Bardy, B. G. (2001). On specification and the senses. *Behavioral and Brain Sciences*, *24*(2), 195-213.

Wälchli, Z. (2016). Recall of compound words in simple and complex span tasks (Master's thesis, McMaster University). Available from ProQuest Dissertations & Theses Global database.

West, R. F., & Stanovich, K. E. (1986). Robust effects of syntactic structure on visual word processing. *Memory & Cognition*, *14*(2), 104-112.

5 Summary and Conclusions

The main objective of this dissertation was to examine the effects of headedness on compound word processing. This work was done on Persian, which allows for variable head positions of compounds, with right headedness being the default form in this language. First, I developed a database of noun-noun (NN) compound words in Persian. This database provides information on headedness, phoneme length, letter length, literal meaning, English translation equivalent, whole word Google frequency, and constituent Google frequency of two hundred NN compound words in Persian. The stimuli used for the three experiments, presented in this dissertation, were taken from this database. The first experiment provides ratings for familiarity, imageability, and AOA. Since there is a lack of frequency information on Persian compound words, due to their orthography, these three lexical variables were used as controls for the results of the second and third experiments. To address my main research question, the following specific questions were asked: (i) in a memory recall experiment, are head-final (RH) compounds recalled better than head-initial (LH) ones?, (ii) Does the internal structure of the compounds play a role on their processing?, (iii) in a primed naming experiment, do natural RH compounds have faster reaction times than LH compounds?, (iv) in a primed naming experiment, is there an interaction between headedness and prime condition to suggest that priming effects are different for the two types of words?

The following section outlines the most important content, findings, significance, and implications of each chapter. After this, some limitations of the study and suggestions for future research are listed.

5.1 Thesis summary and findings

This thesis consists of three experiments. The main hypothesis and findings of each one is listed below.

Familiarity, imageability and age of acquisition ratings of Persian right and left-headed NN compounds

Chapter 2 provided information on age of acquisition (AOA), familiarity, and imageability as lexical variables that may have an effect on the processing of compound words (cf. Stadthagen-Gonzalez & Davis, 2006; Juhasz, Lai, & Woodcock, 2015). The existing data on these variables are from English and only a few other languages. To allow more cross-linguistic research, there is a need for such information on languages other than English, and particularly on ones with a different syntactic make up. This chapter introduced a database of AOA, familiarity, and imageability ratings for 149 Persian compound words and non-words (constructed of real word constituents) collected from 102 participants. The data analysis indicated that familiarity, imageability, and AOA are highly inter-correlated in both right-headed and left-headed compounds.

Another major aspect of the data in this chapter relates to the headedness component added to the matrix of subjective norms. This is the first study to provide such ratings in a language that allows for different head positions of compound words. The findings revealed for AOA that the LH compounds were thought to have been learnt earlier in life than the RH compounds. This can be due to the fact that left-headed or so-called historical compounds are historically shaped by the Ezafe construction and are more transparent in Persian.

Effects of constituent order and syntactic role of compounds on short-term memory

Chapter 3 presents data from a study of compound word maintenance in short-term memory. Results from morphological processing experiments in variable-headed Italian have suggested a central role for transparent right heads in constituent access and semantic combination. In this chapter, I used memory span tasks and looked for different recall of lists consisting of left- and right-headed compounds. Because right headedness is the default in Persian, I hypothesized that RH compounds would be remembered better than their LH counterparts. I formed lists of five different types of auditory stimuli. The data analysis confirmed better recall of the Common RH compounds versus the Common LH ones. The results show that processing of an irregular or a marked structure can negatively affect STM span in Persian (for similar results in other languages see Service & Tujulin, 2002 for Finnish and Németh et al., 2011 for Hungarian) and confirms a syntax-based decomposition approach for the mental lexicon in Persian.

The study presented in this chapter is the first experiment that shows the effects of compound headedness on working memory span in Persian.

The role of the different compound constituents as primes on reaction times in naming

Chapter 4 reports the results of a primed reaction time experiment on compound words in Persian. There is a significant body of research on compound word processing, but the number of studies examining the effects of different constituents of the compounds as primes in an SOV language with variable head positions are very few up to this date. I designed this experiment to investigate the lexical representation of Persian compound words in the mental lexicon and the possible influence of the compound heads as primes on whole-word naming. This is the first naming study in Persian that used such design to show the effects of different head constituents, as primes, on naming latencies. The results demonstrated different processing for RH versus LH compound words in Persian and added more evidence to the decompositional route of compound processing in this language, which was also suggested by other experimental results in Persian (e.g., Megerdoomian, 2001; Nojoumian et al., 2006; Shabani-Jadidi, 2019). Results also indicated longer reaction times for LH compounds, as a marked structure in Persian NN compounds, and are in line with results from chapter three revealing poorer recall for LH compounds. These results added more evidence to the existing literature in other languages, suggesting that the processing of marked structures is cognitively harder than that of regular ones. (see for example Service & Tujulin, 2002 for Finnish and Németh et al., 2011 for Hungarian). My priming results revealed that in LH compounds, the head can prime better than the modifier. This is a novel finding for Persian compound words. This also suggests that priming by the modifier did not work for the left-headed compounds.

5.2 Conclusions

I started this dissertation with three main questions about NN compounds in Persian. (i) in a memory recall experiment, do head-final (RH) compounds have a better recall than head-initial (LH) ones?, (ii) Does the internal structure of the compounds play a role on their processing?, (iii) in a primed naming experiment, do natural RH compounds have faster reaction times than LH compounds?, (iv) in a primed naming experiment, is there an interaction between headedness and prime condition to suggest that priming effects are different for the two types of words?

Chapter 2 findings indicated that familiarity, imageability, and AOA are highly inter-correlated in both right-headed and left-headed compounds. The AOA results demonstrated that the LH compounds were rated to have been learnt earlier in life. I suggest that this is due to the fact that left-headed or so-called historical compounds are shaped by the Ezafe construction and are more transparent in Persian. The findings suggested that the internal structure of Persian NN compounds affects their lexical ratings by native speakers.

The main conclusions of chapter 3 were that lists with common RH compounds were significantly better recalled than the common LH compounds. This finding confirmed my second hypothesis, indicating that RH compounds, as the default form in the language, have more memorable structures than their LH counterparts. This result also proves that the internal structure of compounds plays a significant role in their processing. Another important conclusion following from this finding is that an irregular or a marked structure can negatively affect STM span in Persian. These results align with findings in English and some other languages, such as Finnish (Service & Tujulin, 2002) and Italian (Arcara et al., 2014), indicating that processing marked structures results in a higher cognitive load.

Chapter 4 confirmed shorter reaction times for RH compounds, indicating that LH compounds in Persian are processed more slowly, as a marked structure. The priming analysis revealed that unlike RH compounds, the head in LH compounds primed better than the modifier. This finding is novel and can be interpreted based on the embedded Ezafe structure in Persian, which might mean that the modifier is not solidly attached to the head, as the modifier is in the "real" RH compound words that do not have an underlying Ezafe structure. Therefore, findings of this chapter also reveal the importance of the Ezafe structure as a unique construction in Persian noun phrases, which causes syntactic complexity in left-headed compounds and, therefore, results in LH compound words in Persian being harder to access, represent in working memory, and process (see Kahnemuyipour, 2006; 2014; Ghomeshi 1997; Karimi & Brame, 2012 for more details on Persian Ezafe construction).

5.3 Limitations

One major limitation of this study was the lack of corpus frequencies of NN compounds in Persian. This was inevitable due to the ambiguous orthography of Persian, in which the space between the two constituents of the compound words can be one, half, or no space at all. However, I used the rated lexical variables provided in the second chapter of this thesis to control for lexical effects on results of the memory and naming latency experiments.

Another limitation of this work was the lack of a neutral prime in experiment 3 and the lack of semantic transparency ratings for compounds used for this experiment.

5.4 Path for future research

To my best knowledge, my experiments are all novel, in methodology and design, for providing information on the influence of headedness on NN compound word processing in Persian. My experiments open the path for follow-up experiments using different methods such as eye tracking to provide more data on the role compound constituents, as primes, play in their processing.

Another possibility for a follow-up experiment can be to use the whole compound word to prime a constituent. It would be interesting to see if similar results, regarding headedness, can be replicated with such a design.

Also, for future primed experiments in Persian compounds, I suggest controlling for phonological effects of primes to add more evidence on the priming effects being syntactic and to eliminate the phonological aspects. One way to do this would be to create one condition with phonologically similar primes that are not compound constituents, e.g. priming “milkman” with “mill” in English.

References

Arcara, G., Marelli, M., Buodo, G., & Mondini, S. (2014). Compound headedness in the mental lexicon: An event-related potential study. *Cognitive Neuropsychology*, *31*, 164–183.

Ghomeshi, J. (1997). Non-projecting nouns and the ezafe: construction in Persian. *Natural Language & Linguistic Theory*, *15*(4), 729-788.

Juhasz, B. J., Lai, Y.-H., & Woodcock, M. L. (2015). A database of 629 English compound words: Ratings of familiarity, lexeme meaning dominance, semantic transparency, age of acquisition, imageability, and sensory experience. *Behavior Research Methods*, *47*(4), 1004–1019.

Kahnemuyipour, A. (2006). Persian Ezafe construction: Case, agreement or something else. In *Proceedings of the 2nd Workshop on the Persian Language and Computer. Tehran: University of Tehran*.

Kahnemuyipour, A. (2014). Revisiting the Persian Ezafe construction: A roll-up movement analysis. *Lingua*, *150*, 1–24. https://doi.org/10.1016/j.lingua.2014.07.012

Karimi, S., & Brame, M. (2012). A generalization concerning the EZAFE construction in Persian. *Linguistic Analysis*, *38*(1), 111-144.

Németh, D., Ivády, R., Guida, A., Miháltz, M., Peckham, D., Krajcsi, A., & Pléh, C. (2011). The effect of morphological complexity on short-term memory capacity. *Acta Linguistica Hungarica*, *58*(1), 85-107.

Service, E., & Tujulin, A. (2002). Recall of morphologically complex forms is affected by memory task but not dyslexia. *Brain and Language*, *81*(1_3), 42–54. https://doi.org/10.1006/brln.2001.2505

Stadthagen-Gonzalez, H., & Davis, C. J. (2006). The Bristol norms for age of acquisition, imageability, and familiarity. *Behavior Research Methods*, *38*(4), 598–605.

Appendix A: Database of Persian NN compounds used for the experiments

COMP Persian	COMP English	Word ID	L CONS	R CONS	Literal translation L CONS	literal translation R CONS	English Translation	head	COMP Count	LCONS Count	RCONS Count	Phonetic transcription	LCON 1st Syl Onset	RCONS 1st Syl Onset	Phoneme Count COMP	Phoneme Count LCONS	Phoneme Count RCONS	Frequency COMP	Frequency LCONS	Frequency RCONS
آب آلبالو	aab aalbalu	aab aalbalu (morello cherry)	aab	aalbalu	water	morello cherry	cherry juice	left	8	2	6	a:b a:lba:lu:	NA	NA	8	2	6	97200	178000000	2750000
آب آناناس	aab aanaanaas	aab aanaanaas (pineapple juice)	aab	aanaanaas	water	pineapple	pineapple juice	left	8	2	6	a:b a:na:na:s	NA	NA	8	2	6	92500	178000000	1800000
آب انار	aab anaar	aab anaar (pamegranate juice)	aab	anaar	water	pamegranate	pamegranate juice	left	6	2	4	a:b ænɒ:r	NA	NA	6	2	4	381000	178000000	10900000
آب انبه	aab anbeh	aab anbeh (mango juice)	aab	anbeh	water	mango	mango juice	left	6	2	4	a:b ænbe	NA	NA	6	2	4	79500	178000000	1870000
آب انگور	aab angur	aab angur (grape juice)*	aab	angur	water	grape	grape juice	left	7	2	5	a:b angu:r	NA	NA	7	2	5	271000	178000000	6620000
آب هویج	aab havij	aab havij (carrot juice)	aab	havij	water	carrot	carrot juice	left	6	2	4	a:b hævi:d͡ʒ	NA	h	7	2	5	361000	178000000	4220000
آب لیمو	aab limu	aab limu (lime juice)	aab	limu	water	lime	lime juice	left	6	2	4	a:b li:mu:	NA	l	6	2	4	419000	178000000	6040000
آب پرتقال	aab porteqaal	aab porteqaal (orange juice)*	aab	porteqaal	water	orange	orange juice	left	6	2	8	a:b porteqɒ:l	NA	p	10	2	8	486000	178000000	6340000
آبغوره	aab qure	aab qure (verjuice)	aab	qure	water	grape	verjuice	left	6	2	4	ɒ:b qu:re	NA	q	6	2	4	830000	178000000	2090000
آب رنگ	aab rang	aab rang (watercolor)	aab	rang	water	color	watercolor	right	5	2	3	ɒ:b ræng	NA	r	6	2	4	1630000	178000000	123000000
آب ریواس	aab rivaas	aab rivaas (non-word)	aab	rivaas	water	rhubarb	non-word	left	7	2	5	a:b riva:s	NA	r	7	2	5	0	0	0
آب شکلات	aab shokolaat	aab shokolaat (non-word)	aab	shokolaat	water	chocolate	non-word	left	7	2	5	a:b ʃokolɒ:t	NA	ʃ	9	2	7	0	0	0
آب سیب	aab sib	aab sib (apple juice)	aab	sib	water	apple	apple juice	left	5	2	3	a:b si:b	NA	s	5	2	3	348000	178000000	24300000
آب ذغال	aab zoghaal	aab zoqaal (non-word)	aab	zoghaal	water	charcoal	non-word	left	6	2	4	a:b zoɣɒ:l	NA	z	7	2	5	0	0	0
آش رشته	aash reshte	aash reshte (a type of thick soup)	aash	reshte	thick soup	noodle	a type of thick Iranian soup	left	6	2	4	ɒ:ʃ reʃte	NA	r	7	2	5	456000	61100000	75000000
آشپزخانه	aashpaz khaane	aashpaz khaane (kitchen)*	aashpaz	khaane	chef	house	kitchen	right	8	4	4	a:ʃpæz xɒ:ne	NA	x	9	5	4	30900000	3900000	203000000
آیین نامه	aayeen naame	aayeen naame (edict)	aayeen	naame	rule	letter	regulations	right	8	4	4	ɒ:ji:n nɒ:me	NA	n	8	4	4	481000	27400000	131000000
انگشت بند	angosht band	angosht band (non-word)	angosht	band	finger	band	non-word	right	8	5	3	ængoʃt bænd	NA	b	10	6	4	0	0	0

بانک ملت	baank mellat	baank mellat (Nation bank)*	baank	mellat	bank	nation	name of a bank in Iran	left	7	4	3	baːnk mellæt	b	m	10	4	6	394000	90300000	51200000
باربند	baar band	baar band (carrier)	baar	band	cargo	band	fixture	right	6	3	3	bɒːr bænd	b	b	7	3	4	1790000	170000000	62000000
بارنامه	baar naame	baar naame (bill of landing)	baar	naame	cargo	letter	bill of landing	right	7	3	4	bɒːr nɒːme	b	n	7	3	4	534000	170000000	131000000
بخشنامه	bakhsh naame	bakhsh naame (bylaw)	bakhsh	naame	section	letter	regulation	right	7	3	4	bæxʃ nɒːme	b	n	8	4	4	319000000	168000000	131000000
برف پاک کن	barf paakkon	barf paakkon (wiper)	barf	paakkon	snow	eraser	wiper	right	8	3	5	bærf pɒːkkon	b	p	10	4	6	124000	27300000	332000
بستنی خانه	bastani khaane	bastani khaane (non-word)	bastani	khaane	icecream	house	non-word	right	9	5	4	bæstæniː xɒːne	b	x	11	7	4	0	0	0
چادرنماز	chaador namaaz	chaador namaaz (prayer chador)	chaador	namaaz	chador	prayer	praying chador	left	8	4	4	t͡ʃʰɒːdor næmɒːz	t͡ʃʰ	n	10	5	5	236000	13600000	38200000
چای خانه	chaay khaane	chaay khaane (tea house)	chaay	khaane	tea	house	teahouse	right	7	3	4	t͡ʃʰɒːj xɒːne	t͡ʃʰ	x	7	3	4	122000	29400000	296000000
چراغ مطالعه	cheraagh motalee	cheraagh motalee (desk lamp)	cheraagh	motalee	lamp	study	desk lamp	left	10	4	6	t͡ʃʰerɒːɣ motaːleʔe	t͡ʃʰ	m	13	5	8	2400000	53800000	91200000
چراغ خواب	cheraaq khaab	cheraaq khaab (night lamp)	cheraaq	khaab	lamp	sleep	night lamp	left	8	4	4	t͡ʃʰerɒːɣ xɒːb	t͡ʃʰ	x	8	5	3	411000	31600000	76100000
دامپزشک	daam pezeshk	daam pezeshk (veterinarian)*	daam	pezeshk	animals	doctor	veterinarian	right	7	3	4	dɒːm pezeʃk	d	p	9	3	6	1010000	37900000	26400000
دانشیار	daanesh yaar	daanesh yaar (associate professor)	daanesh	yaar	knowlegde	fellow	associate professor	right	7	4	3	dɒː neʃ jɒːr	d	y	8	5	3	7380000	96700000	131000000
داروخانه	daaru khaane	daaru khaane (drugstore)*	daaru	khaane	drug	house	drugstore; pharmacy	right	8	4	4	dɒːruː xɒːne	d	x	8	4	4	7100000	21800000	203000000
دبیرخانه	dabir khaane	dabir khaane (secretariat)	dabir	khaane	teacher	house	secretariat	right	8	4	4	dæbiːr xɒːne	d	x	9	5	4	7590000	21200000	203000000
دفتر مدرسه	daftar madrese	daftar madrese (non-word)	daftar	madrese	notebook	school	school notebook	left	9	4	5	dæftær mædrese	d	m	13	6	7	0	0	0
دفتر مشق	daftar mashgh	daftar mashgh (assignment notebook)	daftar	mashgh	notebook	homework	homework notebook	left	7	4	3	dæftær mæʃq	d	m	10	6	4	223000	4090000000	3410000
دفتر نقاشی	daftar naqaashi	daftar naqaashi (painting notebook)	daftar	naqaashi	notebook	painting	painting notebook	left	7	4	3	dæftær næqqɒːʃiː	dæf	naq	13	6	7	428000	107000000	561000
دفتر نقشه	daftar naqshe	daftar naqshe (non-word)	daftar	naqshe	notebook	map	non-word	left	8	4	4	dæftær næqʃe	d	n	11	6	5	0	0	0
دندان پزشک	dandaan pezeshk	dandaan pezeshk (dentist)	dandaan	pezeshk	tooth	doctor	dentist	right	9	5	4	dændɒːn pezeʃk	d	p	12	6	6	472000	22800000	26400000
درس نامه	dars naame	dars naame (non-word)	dars	naame	lesson	letter	non-word	right	7	3	4	dærs nɒːme	d	n	8	4	4	0	0	0
دست بند	dast band	dast band (bracelet)*	dast	band	hand	band	bracelet	right	6	3	3	dæst bænd	d	b	8	4	4	392000	252000000	61800000

دستگیره	dast gire	dast gire (handle)*	dast	gire	hand	handle	handle	right	7	3	4	dæst giːre	d	g	8	4	4	2910000	252000000	129000000
دست مزد	dast mozd	dast mozd (wage)	dast	mozd	hand	wage	wage	right	6	3	3	dæst mozd	d	m	8	4	4	89500	252000000	119000000
دسته در	daste dar	daste dar (door knob)	daste	dard	knob	door	door knob	left	6	4	2	dæste dær	d	d	8	5	3	280000	3730000000	825000000
دستمال گردن	dastmaal gardan	dastmaal gardan (neck scarf)	dastmaal	gardan	napkin	neck	neck scarf	left	10	6	4	dæstmɒːl gærdæn	d	g	13	7	6	338000	13000000	18500000
دستمال کمر	dastmaal kamar	dastmaal kamar (non-word)	dastmaal	kamar	handkerchief	waist	non-word	left	9	6	3	dæstmɒːl kæmær	d	k	12	7	5	0	0	0
دستمال توالت	dastmaal tovaalet	dastmaal tovaalet (toilet paper)	dastmaal	tovaalet	napkin	toilet	toilet paper	left	11	6	5	dæstmɒːl tuːɒːlet	d	t	13	7	6	441000	13000000	7640000
دوست دختر	dust dokhtar	dust dokhtar (girlfriend)	dust	dokhtar	friend	girl	girlfriend	left	8	4	4	duːst doxtʰær	d	d	10	4	6	381000	127000000	98200000
دوست پسر	dust pesar	dust pesar (boyfriend)	dust	pesar	friend	boy	boyfriend	left	7	4	3	duːst pesær	d	p	9	4	5	223000	4090000000	3410000
فلکه پارک	falake paark	falake paark (park square)	falake	paark	square	park	Park Square; famous in Iran	left	8	4	4	fælæke paːrk	f	p	10	6	4	31200	5730000	33500000
فرهنگ نامه	farhang naame	farhang naame (encyclopedia)	farhang	naame	culture	letter	encyclopedia	right	9	5	4	færhæŋg nɒːme	f	n	11	7	4	350000	102000000	131000000
گاوصندوق	gaav sanduq	gaav sanduq (safe box)*	gaav	sanduq	cow	box	safe	right	8	3	5	gɒːv sænduːq	g	s	9	3	6	1970000	10100000	745000000
گاوزنبور	gaav zanbur	gaav zanbur (redbee)	gaav	zanbur	cow	bee	redbee	right	8	3	5	gɒːv zænbuːr	g	z	9	3	6	172	10100000	5080000
گواهی نامه	gavaahi naame	gavaahi naame (certificate)*	gavaahi	naameh	certificate	letter	certificate	right	9	5	4	gævɒːhiː nɒːme	g	n	10	6	4	1620000	17100000	169000000
قلم خانه	ghalam khaane	ghalam khaane (non-word)	ghalam	khane	reed pen (stylus)	house	non-word	right	7	3	4	qælæm xɒːne	q	x	9	5	4	0	0	0
قلب پزشک	ghalb pezeshk	ghalb pezeshk (non-word)	ghalb	pezeshk	heart	doctor	non-word	right	7	3	4	qælb pezeʃk	q	p	10	4	6	0	0	0
گیره سر	gire sar	gire sar (hair clamp)*	gire	sar	clamp	head	hair clamp	left	6	4	2	giːre sær	g	s	7	4	3	125000	129000000	196000000
گلاب	gol aab	gol aab (rosewater)*	gol	aab	flower	water	rosewater	right	4	2	2	golaːb	g	NA	5	3	2	7710000	115000000	178000000
گلدسته	gol daste	gol daste (finial)	gol	dasteh	flower	bouquets	finial	right	6	2	4	gol dæste	g	d	8	3	5	727000	115000000	3730000000
گوش بند	gush band	gush band (non-word)	gush	band	ear	band	non-word	right	6	3	3	guːʃ bænd	g	b	7	3	4	0	0	0
گوش پاک کن	gush paakkon	gush paakkon (cotton swab)*	gush	paakkon	ear	eraser	cotton swab	right	8	3	5	guːʃ paːkkon	g	p	9	3	6	704000	43500000	464000
حلقه تولد	halghe tavallod	halghe tavallod (non-word)	halghe	tavallod	ring	birthday	non-word	left	8	4	4	hælqe tævællod	h	t	13	5	8	0	0	0
حلقه ازدواج	halqe ezdevaaj	halqe ezdevaaj (wedding ring)*	halqe	ezdevaaj	ring	marriage	wedding ring	left	10	4	6	hælqe ezdevɒːd͡ʒ	h	NA	12	5	7	428000	34900000	57400000
جوجه ماهی	juje maahi	juje maahi (non-word)	juje	maahi	chick	fish	non-word	right	8	4	4	d͡ʒuːd͡ʒe mɒːhiː	d͡ʒ	m	8	4	4	0	0	0

کاغذ کادو	kaaqaz kaado	kaaqaz kaado (giftwrap)*	kaaqaz	kaado	paper	gift	giftwrap	left	8	3	4	kɒːɣæz kɒːdo	k	k	9	5	4	213000	24400000	5980000
کاغذ نما	kaaqaz namaa	kaaqaz namaa (non-word)	kaaqaz	namaa	paper	façade	non-word	right	7	4	3	kɒːɣæz namaː	k	n	9	5	4	0	0	0
کارخانه	kaar khaane	kaar khaane (factory)	kaar	khaane	work	house	factory	right	7	3	4	kɒːr xɒːne	k	x	7	3	4	37800000	289000000	203000000
کارت دعوت	kaart davat	kaart davat (invitation card)	kaart	davat	card	invitation	invitation card	left	8	4	4	kaːrt dævæt	k	d	9	4	6	467000	68500000	46700000
کارت خنده	kaart khande	kaart khande (non-word)	kaart	khande	card	laughter	non-word	left	8	4	4	kaːrt xænde	k	x	10	4	5	0	0	0
کارت تبریک	kaart tabrik	kaart tabrik (greeting card)*	kaart	tabrik	card	greeting	greeting card	left	9	4	5	kaːrt tæbriːk	k	t	10	4	6	546000	68500000	38900000
کباب آلو	kabaab aaloo	kabaab aaloo (non-word)	kabaab	aaloo	kebab	plum		left	7	4	3	kæbɒːb ɒːluː	k	NA	8	5	3	0	0	0
کباب برگ	kabaab barg	kabaab barg (Leaf kebab)*	kabaab	barg	kebab	leaf	leaf kebab	left	7	4	3	kæbɒːb bærg	k	b	9	5	4	148000	9360000	27900000
کیک پلو	kaik polo	kaik polo (non-word)	kaik	polo	cake	rice	non-word	right	6	3	3	kejk polo	k	p	8	4	4	0	0	0
کمربند	kamar band	kamar band (belt)	kamar	band	waist	band	belt	right	6	3	3	kæmær bænd	k	b	9	5	4	14600000	22500000	107000000
کمر درد	kamar dard	kamar dard (back pain)*	kamar	dard	back	pain	back pain	right	6	3	3	kæmær dærd	k	d	9	5	4	454000	11900000	52400000
کتاب دعا	ketaab doa	ketaab doa (prayer book)*	ketaab	doa	book	prayer	prayer book	left	7	4	3	ketɒːb doʔɒː	k	d	9	5	4	287000	227000000	64000000
کتاب خواب	ketaab khaab	ketaab khaab (non-word)	ketaab	khaab	book	sleep	non-word	left	8	4	4	ketɒːb xɒːb	k	x	8	5	3	0	0	0
کتابخانه	ketaab khaane	ketaab khaane (library)*	ketaab	khaane	book	house	library	right	8	4	4	ketɒːb xɒːne	k	x	9	5	4	27200000	142000000	203000000
کتاب مسابقه	ketaab mosaabeqe	ketaab mosabeqe (non-word)	ketaab	mosabeqe	book	contest	non-word	left	10	4	6	ketɒːb mosɒːbeqe	k	m	13	5	8	0	0	0
خرمگس	khar magas	khar magas (horse-fly)	khar	magas	donkey	fly	horsefly	right	5	2	3	xær mægæs	x	m	8	3	5	239000	7600000	865000000
کلاه گیس	kolaah gis	kolaah gis (wig)	kolaah	gis	hat	hair	wig	right	7	4	3	kolɒːh giːs	k	g	8	5	3	386000	20000000	2630000
کوهپایه	kuh paaye	kuh paaye (mountain range)*	kuh	paaye	mountain	base	foothill	right	7	5	4	kuːhpɒːjæ	k	p	7	3	4	1220000	1570000000	66000000
کوکو بادنجان	kuku baadenjaan	kuku baadenjaan (eggplant frittata)	kuku	baadenjaan	frittata	eggplant	eggplant frittata	left	11	4	7	kuku bɒːdemd͡ʒɒːn	k	b	12	4	8	3800	3780000	60300000
کوکو سبزی	kuku sabzi	kuku sabzi (herb frittata)	kuku	sabzi	frittata	herb	herb frittata	left	8	4	4	kuːkuː sæbziː	k	s	9	4	5	484000	3780000	542000000
کوکو سیب	kuku sib	kuku sib (non-word)	kuku	sib	keesh	apple	non-word	left	7	4	3	kuku siːb	k	s	7	4	3	0	0	0
کوکو سیب‌زمینی	kuku sib-zamini	kuku sib-zamini (potato fritata)	kuku	sib-zamini	frittata	potato	potato frittata	left	12	4	8	kuːkuː siːbzæmiːniː	k	s	13	4	9	298000	4600000	14300000
کوسه ماهی	kuse maahi	kuse maahi (shark)	kuse	maahi	shark	fish	shark	right	8	4	4	kuːse mɒːhiː	k	m	8	4	4	125000	2170000	38600000

لباس خواب	lebaas khaab	lebaas khaab (sleeping dress)*	lebaas	khaab	clothes	sleep	sleeping dress	left	8	4	4	lebɒ:s xɒ:b	l	x	8	5	3	559000	102000000	76200000
ماهی کباب	maahi kabaab	maahi kabaab (fish kebab)*	maahi	kabaab	fish	kebab	Fish Kebab	right	8	4	4	mɒ:hi: kæbɒ:b	m	k	9	4	5	102000	38600000	9360000
ماهی قباد	maahi qobaad	maahi qobaad (A type of Fish)*	maahi	qobaad	fish	Ghobaad	A sort of Fish	left	8	4	4	mɒ:hi: qobɒ:d	m	q	9	4	5	16700	38600000	39700000
ماهی شیر	maahi shir	maahi Shir (lionfish)	maahi	shir	fish	lion	lionfish	left	7	4	3	mɒ:hi: ʃi:r	m	ʃ	7	4	3	141000	38600000	39700000
ماهی تن	maahi ton	maahi ton (tuna fish)	maahi	ton	fish	tuna	tuna fish	left	6	4	2	mɒ:hi: ton	m	t	7	4	3	270000	38600000	67300000
مار ماهی	maar maahi	maar maahi? (snakehead fish)	maar	maahi	snake	fish	snakehead fish	right	7	3	4	mɒ:r mɒ:hi:	m	m	7	3	4	69700	18200000	38600000
ماشین مسابقه	maashin mosaabeqe	maashin mosaabeqe (racecar)	maashin	mosaabeqe	machine	race	racecar	left	11	5	6	mɒ:ʃi:n mosɒ:beqe	m	m	13	5	8	237000	102000000	43600000
ماشین تحریر	maashin tahrir	maashin tahrir (typewriter)	maashin	tahrir	machine	writing	typewriter	left	10	5	5	mɒ:ʃi:n tæhri:r	m	t	11	5	6	118000	102000000	40600000
مهمان خانه	mehmaan khaane	mehmaan khaane (guesthouse)	mehmaan	khaane	guest	house	guesthouse	right	9	5	4	mehma:n xɒ:ne	m	x	10	6	4	366000	25100000	203000000
میخانه	mey khaane	mey khaane (bar)	mey	khaane	wine	house	pub	right	6	2	4	meʃxɒ:ne	m	x	7	3	4	1290000	462000000	203000000
میدان آزادی	meydaan aazaadi	meydaan aazaadi (freedom square)	meydaan	aazaadi	square	freedom	name of a famous square in Iran	left	10	5	5	mejdɒ:n a:za:di:	m	NA	11	6	5	529000	61500000	56500000
میز خانه	miz khaane	miz khaane (non-word)	miz	khaane	desk	house	non-word	right	7	3	4	mi:z xɒ:ne	m	x	7	3	4	0	0	0
میز تحریر	miz tahrir	miz tahrir (office desk)	miz	tahrir	desk	writing	office desk	left	8	3	5	mi:z tæhri:r	m	t	9	3	6	517000	36200000	40600000
معلّم ورزش	moallem varzesh	moallem varzesh (sports teacher)	moallem	varzesh	teacher	sports	sports teacher; physical education teacher	left	8	4	4	moællem værzeʃ	m	v	13	8	6	232000	47900000	87800000
مسافرخانه	mosaafer khaane	mosaafer khaane (inn)*	mosaafer	khaane	traveler	house	inn	right	9	5	4	mosɒ:fer xɒ:ne	m	x	11	7	4	710000	25300000	203000000
مو درد	mu dard	mu dard (non-word)	mu	dard	hair	pain	non-word	right	5	2	3	mo: dærd	m	d	6	2	4	0	0	0
نان لواش	naan lavaash	naan lavaash (Lavash bread)	naan	lavaash	bread	thin	Lavash bread	left	7	3	4	na:n læva:ʃ	n	l	8	3	5	349000	24700000	669000
نان نخودچی	naan nokhodchi	naan nokhodchi (chickpea cookie)	naan	nokhodchi	bread	chickpea	chickpeas cookie	left	9	3	6	na:n noxodtʃʰi:	n	n	10	3	7	124000	45400000	1730000
نان سنگک	naan sangak	naan sangak (Sangak bread)	naan	sangak	bread	small stone	Sangak bread	left	7	3	4	na:n sængæk	n	s	9	3	6	350000	24700000	905000
نمایشنامه	namaayesh naame	namaayesh naame (play)	namaayesh	naame	play	letter	play	right	9	5	4	næmɒ:jeʃ nɒ:me	n	n	11	7	4	4030000	132000000	131000000
علوم پایه	olum paaye	olum paaye (fundamental science)	olum	paaye	science	fundament	fundamental science	left	8	4	4	ʔolu:m pɒ:jæ	ʔ	p	9	5	4	415000	126000000	66000000

اردک ماهی	ordak maahi	ordak maahi (pike)*	ordak	maahi	duck	fish	pike	right	8	4	4	ordæk mɒ:hi:	NA	m	9	5	4	63600	2980000	38700000
پا درد	paa dard	paa dard (leg pain)	paa	dard	leg	pain	leg pain	right	5	2	3	pɒ: dærd	p	d	6	2	4	251000	54700000	54700000
پارک ملت	paark mellat	paark mellat (nation park)	paark	mellat	park	nation	name of an urban park in northern Iran	left	7	4	3	pa:rk mellæt	p	m	10	4	6	406000	33500000	51200000
پایان نامه	paayaan naame	paayaan naame (thesis)*	paayaan	naame	end	letter	thesis	right	9	5	4	pɒ:jɒ:n nɒ:me	p	n	9	5	4	412000	149000000	131000000
پایه صندلی	paaye sandali	paaye sandali (chair base)*	paaye	sandali	base	chair	chair base	left	9	4	5	pɒ:jæ sændæli:	p	s	11	4	7	131000	66600000	28600000
پایه تابلو	paaye taablo	paaye taablo (non-word)	paaye	taablo	leg	picture frame	non-word	left	9	4	5	pɒ:jæ tɒ:blo	p	t	9	4	5	0	0	0
پشه بند	pashe band	pashe band (mosquito net)*	pashe	band	mosquito	band	mosquito net	right	6	3	3	pæʃe bænd	p	b	8	4	4	367000	2220000	61800000
پرسشنامه	porsesh naame	porsesh naame (questionnaire)	porsesh	naame	question	letter	questionnaire	right	8	4	4	porseʃ nɒ:me	p	n	10	6	4	9440000	4860000000	131000000
پست نامه	post naame	post naame (non-word)	post	naame	post	letter	non-word	right	7	3	4	post nɒ:me	p	n	8	4	4	0	0	0
رگ درد	rag dard	rag dard (non-word)	rag	dard	artery	pain	non-word	right	5	2	3	ræg dærd	r	d	7	3	4	0	0	0
رودخانه	rud khaane	rud khaane (river)	rud	khaane	river	house	river	right	7	3	4	ru:d xɒ:ne	r	x	7	3	4	15200000	5500000	203000000
روزنامه نگار	ruznaame -negaar	ruznaame-negaar (journalist)*	ruznaame	negaar	newspaper	writer	journalist	right	11	7	4	ru:znɒ:me negɒ:r	r	n	12	7	5	555000	62700000	16300000
سبزی آش	sabzi aash	sabzi aash (vegetable for thick soup)	sabzi	aash	vegetable	thick soup	vegetable for thick soup	left	6	4	2	sæbzi: a:ʃ	s	NA	7	5	2	105000	542000000	60900000
سبزی قورمه	sabzi qorme	sabzi qorme (vegetable for gourmet)	sabzi	qorme	vegetable	gourmet	vegetable for gourmet	left	9	4	5	sæbzi: qorme	s	q	10	5	5	51300	542000000	763000
صندوق پست	sanduq post	sanduq post (postbox)*	sanduq	post	box	post	postbox	left	8	5	3	sændu:q post	s	p	10	6	4	162000	745000000	105000000
سر درد	sar dard	sar dard (headache)*	sar	dard	head	pain	headache	right	5	2	3	sær dærd	s	d	7	3	4	425000	196000000	52400000
سربازخانه	sarbaaz khaane	sarbaaz khaane (barracks)	sarbaaz	khaane	soldier	house	barracks	right	9	5	4	særbɒ:z xɒ:ne	s	x	10	6	4	152000	16000000	203000000
شاهکار	shaah kaar	shaah kaar (masterpiece)*	shaah	kaar	king	work	masterpiece	right	6	3	3	ʃɒ:h kɒ:r	ʃ	k	6	3	3	142000000	36900000	289000000
شاهرگ	shaah rag	shaah rag (artery)	shaah	rag	king	vessel	artery	right	5	3	2	ʃɒ:hrag	ʃ	r	6	3	3	406000	36900000	8120000
شاهرود	shaah rud	shaah rud (great river)	shaah	rud	king	river	the great river	right	6	3	3	ʃɒ:h ru:d	ʃ	r	6	3	3	7070000	36900000	55200000
شاه توت	shaah tut	shaah tut (blackberry)	shaah	tut	king	berry	blackberry	right	6	3	3	ʃɒ:h tu:t	ʃ	t	6	3	3	348000	36900000	13900000
شال گردن	shaal gardan	shaal gardan (neck scarf)*	shaal	gardan	scarf	neck	winter scarf	left	7	3	4	ʃɒ:l gærdæn	ʃ	g	9	3	6	460000	14300000	18500000
شام خانه	shaam khaane	shaam khaane (non-word)	shaam	khaane	dinner	house	non-word	right	7	3	4	ʃɒ:m xɒ:ne	ʃ	x	7	3	4	0	0	0
شانه بند	shaane band	shaane band (non-word)	shaane	band	shoulder	band	non-word	right	7	4	3	ʃɒ:næ bænd	ʃ	b	8	4	4	0	0	0

شلوار جین	shalvaar jean	shalvaar jean (jean pants)*	shalvaar	jean	pants	jeans	jean pants	left	8	5	3	ʃælva:r d͡ʒi:n	ʃ	d͡ʒ	9	6	3	374000	22000000	12200000
شلوار کتان	shalvaar kataan	shalvaar kataan (cotton pants)	shalvaar	kataan	pants	cotton	cotton pants	left	9	5	4	ʃælvɒ:r kæta:n	ʃ	k	11	6	5	368000	22000000	3330000
شلوار پوست	shalvaar pust	shalvaar pust (non-word)	shalvaar	pust	pants	skin	non-word	left	9	5	4	ʃælva:r pu:st	ʃ	p	10	6	4	0	0	0
شمشیر ماهی	shamshir maahi	shamshir maahi (swordfish)	shamshir	maahi	sword	fish	swordfish	right	9	5	4	ʃæmʃi:r mɒ:hi:	ʃ	m	10	6	4	37400	305000000	38600000
شناسنامه	shenaas naame	shenaas naame (birth certificate)*	shenaas	naame	ID	letter	birth certificate	right	8	4	4	ʃenɒ:s nɒ:me	ʃ	n	9	5	4	9340000	8380000	131000000
شیشه نوشابه	shishe nushabe	shishe nushabe (coke glass bottle)	shishe	nushabe	glass	coke	coke glass bottle	left	10	4	6	ʃi:ʃe nu:ʃɒ:be	ʃ	n	10	4	6	579000	66400000	12600000
شیشه سرکه	shishe serke	shishe serke (non-word)	shishe	serke	glass	vinegar	non-word	left	8	4	4	ʃi:ʃe serke	ʃ	s	9	4	5	0	0	0
سیرترشی	sir torshi	sir torshi (garlic pickle)*	sir	torshi	garlic	pickle	garlic pickle	right	7	3	4	si:r torʃi:	s	t	8	3	5	64000	24600000	4990000
صبح نامه	sobh naame	sobh naame (non-word)	sobh	naame	morning	letter	non-word	right	7	3	4	sobh nɒ:me	s	n	8	4	4	0	0	0
سفره ماهی	sofre maahi	sofre maahi (flounder)	sofre	maahi	table-cloth	fish	flounder	right	8	4	4	sofre mɒ:hi:	s	m	9	5	4	136000	32000000	38600000
سوراخ دعا	suraakh doa	suraakh doa (prayer place)?	suraakh	doa	hole	prayer	prayer place	left	8	5	3	su:rɒ:x doʔɒ:	s	d	9	5	4	423000	9960000	47700000
طاق نما	taaq namaa	taaq namaa (arcades)*	taaq	namaa	vault	facing	arcades	right	6	3	3	tɒ:q nama:	t	n	7	3	4	116000	3210000	32100000
تخته پاک کن	takhte paakkon	takhte paakkon (board eraser)	takhte	paakkon	board	eraser	board eraser	right	9	4	5	tæxte pa:kkon	t	p	11	5	6	82100	12700000	462000
تلفن بانک	telephon baank	telephon baank (telehone bank)*	telephon	baank	telephone	bank	telehone bank	right	8	4	4	telefon ba:nk	t	b	11	7	4	198000	85000000	90000000
ترشی لیته	torshi lite	torshi lite (mashed pickle)*	torshi	lite	pickle	eggplant	mashed eggplant pickle	left	8	4	4	torʃi: li:te	t	l	9	5	4	137000	5020000	201000
ترشی سیر	torshi sir	torshi sir (garlic pickle)	torshi	sir	pickle	garlic	garlic pickle	left	7	4	3	torʃi: si:r	t	s	8	5	3	58300	5020000	24700000
زانو درد	zaanu dard	zaanu dard (knee pain)	zaanu	dard	knee	pain	knee pain	right	7	4	3	zɒ:nu: dærd	z	d	8	4	4	446000	7950000	52400000
زنبور عسل	zanbur asal	zanbur asal (honey bee)	zanbur	asal	bee	honey	honey bee	left	8	5	3	zænbu:r ʔæsæl	z	ʔ	11	6	5	394000	5080000	36500000
زندگینامه	zendegi naame	zendegi naame (biography)	zendegi	naame	life	letter	biography	right	9	5	4	zendegi: nɒ:me	z	n	11	7	4	357000000	175000000	131000000

In Word ID column: * = word was used in the memory experiment

Appendix B: Example of the scoring sheet for the memory recall task

Simple Span A3 **PARTICIPANT NUMBER: ______________________**

TRIAL	WORD	NOTES
3.1.1 3.1.2 3.1.3	دستگیره گلاب اردک ماهی	
3.2.1 3.2.2 3.2.3	گلاب گاو صندوق کوه پایه	
3.3.1 3.3.2 3.3.3	کوه پایه گلاب تلفن بانک	
3.4.1 3.4.2 3.4.3	ماهی کباب سیر ترشی اردک ماهی	
3.5.1 3.5.2 3.5.3	اردک ماهی ماهی کباب تلفن بانک	
3.6.1 3.6.2 3.6.3	گلاب ماهی کباب تلفن بانک	
3.7.1 3.7.2 3.7.3	کوه پایه تلفن بانک ماهی کباب	

3.8.1 3.8.2 3.8.3	گلاب اردک ماهی کوه پایه	
3.9.1 3.9.2 3.9.3	اردک ماهی تلفن بانک دستگیره	
3.10.1 3.10.2 3.10.3	اردک ماهی دستگیره سیر ترشی	

Simple Span A4

TRIAL	WORD	NOTES
4.1.1 4.1.2 4.1.3 4.1.4	تلفن بانک کوه پایه سیر ترشی گلاب	
4.2.1 4.2.2 4.2.3 4.2.4	سیر ترشی دستگیره تلفن بانک اردک ماهی	
4.3.1 4.3.2 4.3.3 4.3.4	سیر ترشی تلفن بانک اردک ماهی کوه پایه	

4.4.1 4.4.2 4.4.3 4.4.4	گلاب اردک ماهی گاو صندوق سیر ترشی	
4.5.1 4.5.2 4.5.3 4.5.4	ماهی کباب اردک ماهی کوه پایه تلفن بانک	
4.6.1 4.6.2 4.6.3 4.6.4	تلفن بانک کوه پایه دستگیره گلاب	
4.7.1 4.7.2 4.7.3 4.7.4	ماهی کباب گلاب گاو صندوق دستگیره	
4.8.1 4.8.2 4.8.3 4.8.4	گلاب سیر ترشی دستگیره تلفن بانک	
4.9.1 4.9.2 4.9.3 4.9.4	اردک ماهی سیر ترشی کوه پایه ماهی کباب	
4.10.1 4.10.2 4.10.3 4.10.4	گلاب گاو صندوق ماهی کباب دستگیره	

Simple Span A5

TRIAL	WORD	NOTES
5.1.1 5.1.2 5.1.3 5.1.4 5.1.5	گاو صندوق سیر ترشی اردک ماهی کوه پایه تلفن بانک	
5.2.1 5.2.2 5.2.3 5.2.4 5.2.5	کوه پایه اردک ماهی سیر ترشی گلاب گاو صندوق	
5.3.1 5.3.2 5.3.3 5.3.4 5.3.5	ماهی کباب تلفن بانک سیر ترشی اردک ماهی گلاب	
5.4.1 5.4.2 5.4.3 5.4.4 5.4.5	گلاب سیر ترشی کوه پایه دستگیره ماهی کباب	
5.5.1 5.5.2 5.5.3	تلفن بانک گاو صندوق دستگیره کوه پایه	

5.5.4 5.5.5	سیر ترشی	
5.6.1 5.6.2 5.6.3 5.6.4 5.6.5	گلاب گاو صندوق سیر ترشی اردک ماهی ماهی کباب	
5.7.1 5.7.2 5.7.3 5.7.4 5.7.5	گاو صندوق تلفن بانک دستگیره ماهی کباب گلاب	
5.8.1 5.8.2 5.8.3 5.8.4 5.8.5	گاو صندوق کوه پایه سیر ترشی دستگیره ماهی کباب	
5.9.1 5.9.2 5.9.3 5.9.4 5.9.5	گاو صندوق گلاب اردک ماهی ماهی کباب سیر ترشی	
5.10.1 5.10.2 5.10.3 5.10.4 5.10.5	تلفن بانک گاو صندوق اردک ماهی سیر ترشی ماهی کباب	

Simple Span A6

TRIAL	WORD	NOTES
6.1.1	کوه پایه	
6.1.2	گلاب	
6.1.3	دستگیره	
6.1.4	تلفن بانک	
6.1.5	سیر ترشی	
6.1.6	اردک ماهی	
6.2.1	گلاب	
6.2.2	سیر ترشی	
6.2.3	کوه پایه	
6.2.4	گاو صندوق	
6.2.5	تلفن بانک	
6.2.6	ماهی کباب	
6.3.1	کوه پایه	
6.3.2	دستگیره	
6.3.3	گاو صندوق	
6.3.4	تلفن بانک	
6.3.5	اردک ماهی	
6.3.6	گلاب	
6.4.1	دستگیره	
6.4.2	گلاب	
6.4.3	سیر ترشی	
6.4.4	کوه پایه	
6.4.5	اردک ماهی	
6.4.6	تلفن بانک	

6.5.1 6.5.2 6.5.3 6.5.4 6.5.5 6.5.6	ماهی کباب گلاب گاو صندوق دستگیره اردک ماهی تلفن بانک	
6.6.1 6.6.2 6.6.3 6.6.4 6.6.5 6.6.6	سیر ترشی تلفن بانک اردک ماهی کوه پایه گلاب ماهی کباب	
6.7.1 6.7.2 6.7.3 6.7.4 6.7.5 6.7.6	گلاب دستگیره تلفن بانک ماهی کباب سیر ترشی گاو صندوق	
6.8.1 6.8.2 6.8.3 6.8.4 6.8.5 6.8.6	اردک ماهی کوه پایه دستگیره گلاب ماهی کباب گاو صندوق	
6.9.1 6.9.2 6.9.3 6.9.4	دستگیره کوه پایه گلاب تلفن بانک سیر ترشی ماهی کباب	

6.9.5 6.9.6		
6.10.1 6.10.2 6.10.3 6.10.4 6.10.5 6.10.6	تلفن بانک اردک ماهی کوه پایه گاو صندوق گلاب دستگیره	