Language and Cognition: Effects of Grammatical Gender on the Categorisation of Objects

Fatimah Almutrafi

A Thesis Submitted for the Degree of Doctor of Philosophy (Integrated)

School of Education, Communication, and Language Sciences

Newcastle University

May, 2015

Abstract

One of the key testing grounds for investigating linguistic relativity is to study the effects of grammatical gender on speakers of two languages (bilinguals) who have the category of gender present in only one of their languages. Previous studies have shown that speakers of grammatically gendered languages think of objects as being either masculine or feminine according to the grammatical gender of the objects' nouns. This study investigates the possible effects of grammatical gender on Arabic-English bilinguals and on two 'control' monolingual speakers of Arabic and English. Specifically, two cognitive experiments were carried out in order to investigate gender effects with variations in task instructions and task demands (categorisation vs. similarity ratings). In the first experiment, the bilingual and monolingual participants were asked to attribute masculine and feminine voices to pictures of inanimate items. The results show that the English speakers assigned voices arbitrarily, whereas the Arabic monolinguals attributed more masculine voices to objects whose noun is grammatically masculine in Arabic and more feminine voices to objects whose noun is grammatically feminine in that language, showing the strong effects of the Arabic grammatical gender system. The bilinguals were not greatly affected by the gender system and their voice attributions were somewhere between the two monolingual groups. In the second experiment, the monolinguals and Arabic-English bilinguals were asked to rate similarities between pairs on seven-point scales. The rating task used only pictorial stimuli in an attempt to prevent any strategic use of grammatical gender. Results show that all groups rated the pairs similarly and did not significantly diverge from each other. Overall, these studies suggest that conceptual organisation seems to be free from the effect of grammatical gender and that ways of accessing cognitive representations differ with the modalities tested and with the demands of the task.

Acknowledgements

During this long journey, I have accumulated numerous debts, many of which I will never be able to repay. First of all, I owe a special debt of heartfelt thanks to my husband, Oudah, for his unconditional love, support and encouragement throughout the years of my study. I would not have made it this far without him. Thanks will never be enough.

There are no proper words to convey my sincerest gratitude and thanks to my supervisor, Professor Vivian Cook, for his advice and invaluable feedback at every stage of the research process. This thesis would not have developed in the way that it has without his clear insight and support.

My sincere thanks must also go to my second supervisor Dr. Panos Athanansopoulos for his guidance and much appreciated advice and support during the design of the experiments and his valuable feedback on the overall project.

I sincerely thank my thesis examiners, Emanuel Bylund and Nick Riches for the effort and time they spent reading and commenting on my work.

My deepest appreciation is further extended to my parents and my brothers and sisters for their continued prayers, love and support.

I also thank my wonderful children: Firas and Samer, for always making me smile and for their understanding during those holiday periods when I was writing this thesis instead of playing with them. I hope that one day they can read this thesis and understand why I spent so much time in front of my computer.

Last but not least, special thanks go to all the participants who voluntarily took part in the study, their input is greatly acknowledged and valued.

Above all, I owe it all to Almighty Allah for granting me the health and strength to undertake this research project and enabling me to bring it to completion.

Dedication

To my beloved children Firas, Samer and the coming one, with love

Table of Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Introduction

The latest edition of *Ethnologue* (a guide to the world's languages published by the Summer Institute of Linguistics) reports that there are some 7,105 living languages in the world (Lewis et al., 2013). These languages differ in many ways; such as the range of phonemes, word forms, phonotactic rules and sentence-level syntax. Many researchers claim that such differences between languages should lead to large differences in experience and thought. This relationship between the language we speak and the way we think and perceive the world is known as the 'Linguistic Relativity Hypothesis' (LRH). This idea was first proposed by an American anthropological linguist Edward Sapir (1884-1936) and his student Benjamin Lee Whorf (1897–1941). Although Sapir and Whorf's ideas have been seen as controversial, they have also exerted a strong influence on most scientific thinking. Since their publication, there has been a great deal of debate concerning the ways in which language affects thought. As a result, varying beliefs and different arguments have developed with regard to the relationship between language and thought. Some researchers strongly support linguistic relativity (e.g. Brown & Lenneberg, 1954; Lucy, 1997; Sera, Berge, Castillo, 1994; and Kay & Kempton, 1984; Boroditsky, Schmidt, & Philips, 2003). Other researchers, however, hold a 'Universalist' view of cognition and put forward evidence against the relativity hypothesis (e.g. Martinez & Shatz, 1996; and Takano, 1989; January and Kako 2007; Chen, 2007). Early evidence taken from research into colour perception in the 1970s (e.g. Berlin and Kay, 1969; Heider, 1972[1]) appeared to refute the hypothesis. It was labelled as circular, non-testable and probably wrong by a number of prominent scholars (e.g. Fodor, Bever & Garrett, 1974). Advocates of the Universalist view argue that non-linguistic concepts are formed independently from the words that label them. Although languages differ in their grammatical structure and/or lexical properties, their conceptual structures are the same across languages and cultures. Both of these views are discussed further in chapter 2.

[1] These studies represent the second wave of research on colour perception; it refuted the first wave which did show support for linguistic relativity (Brown & Lenneberg, 1954; Lenneberg & Roberts, 1956; Lenneberg, 1953).

Languages differ in terms of their representation of aspects of reality in areas such as time, space, number, colour and classification of objects and substances at a lexical and grammatical level. Several studies have demonstrated that the habitual use of a language can affect its speakers' thinking (e.g. Boroditsky et al., 2003; Lucy, 1997; Roberson, Davies, & Davidoff, 2000; Sera et al., 1994; Whorf, 1956). One of the most remarkable ways in which languages differ is in whether or not they assign a grammatical gender to object nouns. 'Grammatical gender' is a system in which all nouns are classified as belonging to a certain gender (e.g. masculine, feminine, or neuter) with which other elements in the sentence connected to that noun (such as articles or adjectives) must agree. In English, for example, nouns with a natural gender such as *boy* or *girl* must agree in their grammatical gender with any pronouns used to represent them. For example, 'he is a nice girl' is incorrect in English. English is considered, however, as a 'natural' gender language as it rarely assigns gender to nouns referring to inanimate entities. Other languages have much more extensive and complex systems of grammatical gender. In languages such as Arabic, French and Spanish, all nouns are either masculine or feminine. Other languages, such as German and Russian, have three gender classes; namely masculine, feminine and neuter. Furthermore, languages such as Zande and Dyirbal have four or more genders (Corbett, 1991) - see section 2.2.1 for a review of gender in languages.

Researchers have argued that if languages differ in so many aspects, then speakers of different languages must think in different ways. If a language really does affect thinking and categorisation, would speakers of languages that inflect nouns in terms of grammatical gender (e.g. Arabic) categorise nouns with no biological gender as either feminine or masculine according to their grammatical gender? Take, for example, *spoon* 'melaqah, *fork* 'shawkah' and *table* 'tawelah' are all grammatically feminine in Arabic, whereas *chair* 'korsi', *door* 'bab' and *scissors* 'maqas' are grammatically masculine. Arabic grammatical gender was, therefore, chosen as a test case for linguistic relativity as it cannot be replaced by other lexicalisation patterns and its assignment is arbitrary - except in the case of natural gender (a male/female distinction). To speakers of gender-marked languages, gender is an essential part of comprehension and is employed in many different ways. Research into this aspect of language can give insights into both the way in which we perceive the world around us and how pervasive or limited the role of language could be. Grammatical gender is also psychologically important to speakers as it helps them to predict forthcoming constructions and helps the comprehension process for complex noun phrases. Such experimental research will

shed light on how language is stored and structured in the mind of a speaker and how it can go on to shape their thinking.

It has become common in recent times for many people to speak more than one language (see Grosjean, 2010; Jenkins, 2009). It is therefore interesting to look at the 'Relativity Principle' in relation to both bilinguals and learners of more than one language. Research on bilingual cognition has adopted the view that second language users differ from monolinguals in many ways, particularly by having a different knowledge of the native language in terms of syntax, phonology, vocabulary and pragmatics amongst other areas.

1.2 Focus of the research

The overall issue addressed by the present study is the relationship between language and thinking in human cognition. More specifically, it investigates the effect of Arabic grammatical gender on cognitive processing in both monolingual and bilingual speakers of Arabic. It should be noted, however, that the word 'gender' has distinct meanings in both linguistics and biology and that the gender systems of many languages divide nouns into classes that have no relation whatsoever to biological gender. Linguistically speaking, gender is considered the most puzzling of the grammatical categories as in some languages it is essential and pervasive, while in others it is completely absent (Corbett, 1991). The aim here is to compare the performances of Arabic speakers against those of English speakers through cognitive tasks in the Arabic and English languages (voice attribution and similarity ratings). It is important that cognitive experiments keep the use of language to a minimum by using visual stimuli rather than verbal ones. Such tasks should therefore avoid overt reference to gender - whether natural or grammatical - as this may prompt participants to access grammatical gender to perform the task, rather than simply reflecting the effect of language on their cognition.

In addition, it should be noted that research on grammatical gender has been undertaken mostly with regard to Indo-European languages (e.g. Italian, French, Russian, German and Spanish), meaning that "further studies involving non-Indo European languages are necessary to assess the generality of these findings" (Boroditsky, 2003: 78). As languages tremendously differ from one another, one cannot be confident to generalise results of studies conducted on some language families to represent all languages of the world. Each language, therefore, should be studied in and

for itself and not to be explained by findings of others which might be more appropriate to certain languages. Arabic, for example, is a Semitic language which has received little attention in this respect. The only study to involve Arabic speakers was conducted over thirty years ago by Clarke et al. (1981). Furthermore, that study was seriously questioned with regard to a number of methodological issues (e.g. the task used directed participants to think clearly of the language). It can be argued that when a task is linguistically mediated, there is uncertainty as to whether it is informative about anything other than using language as a strategy to complete the given task.

Another novel dimension of this study is that it involves both monolingual and bilingual speakers, particularly as most of the world's population is now bilingual (Cook, 2003). The bilingual group includes speakers from two different levels i.e. intermediate and advanced. The aim of including these two groups was to see whether the effect of Arabic grammatical gender (if there is any) would change as a result of learning a language with no (or few) gender markers.

Two cognitive tasks were used in this study to investigate the effect of Arabic grammatical gender on the categorisation of objects by Arabic speakers. These experiments manipulate task instruction and task modality and only use non-linguistic stimuli. The performance of Arabic speakers is investigated under different task demands; namely categorisation and similarity ratings. Using different types of tasks can give a clear idea of whether the effect of grammatical gender on the cognition of speakers is at the level of conceptual or semantic representation. The hypothesis is, therefore, that if an effect of grammatical gender is found across these cognitive tasks, this will strongly support the effect of grammatical gender at a conceptual level. If the effect is only found within some tasks, however, this would suggest that such an effect is likely to arise only at a semantic level and is due to task demands. The current study aims, therefore, to show whether (a) Arabic grammatical gender has an effect on the categorisation of objects, (b) Arabic-English bilinguals categorise objects differently from Arabic monolinguals. In other words, does learning a second language lead to a restructuring in the bilingual mind? If so, to what extent.

1.3 Organisation of the thesis

The next chapter presents the background for this research and is divided into two main parts. The first presents general information about the linguistic relativity hypothesis and the second discusses a number of issues with regard to grammatical gender. Chapter

3 describes the methodology and plan for the study as well as the design stages for the experiments; it also gives details about the population studied in all tasks. Chapters 4 and 5 present details of the two experiments used in this study (the voice-attribution task and similarity rating task, respectively). Both chapters include two sections (A and B); section A deals with monolingual data and section B with bilingual data. Each of these sections consists of aims, hypothesis, method, participants, materials and procedure. A report of the results and discussions will end each section. Chapter 6 comprises a general discussion on the findings of the two experiments in relation to previous research. Finally, chapter 7 concludes the thesis by stating a conclusion, advising on the limitations of the current study and making proposals and recommendations for future research.

Chapter 2. Literature Review

2.1 The Linguistic Relativity Hypothesis

This chapter reviews the background literature for the research and is divided into two main parts; the first describes relevant literature on the linguistic relativity hypothesis while the second reviews literature on grammatical gender.

The first section starts by describing the historical background of the relationship between language and cognition. It goes on to describe the emergence of the Sapir-Whorf hypothesis and the notions of linguistic determinism and linguistic relativity are presented. This is followed by sections looking at earlier and more recent studies on linguistic relativity. The final section of part 1 explores the relationship between language and cognition in relation to bilingualism.

2.1.1 Historical Background on Language and Cognition

The relationship between language and thought has been discussed by a range of different scholars throughout time. In the eighteenth century, German scholars such as Machaelis, Johann Gottfried Herder and Johann Georg Hamann were concerned with ideas of language and thought (Lucy, 1992). These ideas continued in the work of the German philosopher Wilhelm von Humboldt who fostered the belief that speakers of different languages have varying views of the world they live in. Humboldt, however, viewed language as the thought of the community rather than of the individual; he argued that "thought and language are therefore one and inseparable from each other" ([1836] 1988: 54). The anthropologist Franz Boas (1858-1942) later took a less deterministic approach to linguistic diversity and observed that language "determines those aspects of each experience that must be expressed" (1938: 127). He made further reference to differences across languages, argued that language played a part in culture and summarised this belief in three ways:

(a) languages classify experience,

(b) different languages will have a different classification of the same experience; meaning that it can lead to different experiences of the same event,

(c) these varying experiences of the same events - due to language - remain unobserved by the speakers of a language because of the automatic nature of language.

Humboldt and Boas were followed by American anthropological linguist Edward Sapir (1884-1936) and his student Benjamin Lee Whorf (1897–1941) who proposed what has

now become known as the 'Linguistic Relativity' or 'Sapir-Whorf Hypothesis'. The section that follows presents the work of Sapir and Whorf in more detail.

2.1.2 Emergence of the Sapir-Whorf Hypothesis

The status of linguistics as a science, an article by Sapir (1958 [1929]: 69) contains one of the most frequently cited quotes about language and thought, namely:

> Human beings do not live in an objective world alone, nor alone in the world of social activity as ordinarily understood, but are very much at the mercy of the particular language which has become the medium of expression for their society It is quite an illusion to imagine that one adjusts to reality essentially without the use of language and that language is merely an incidental means of solving specific problems of communication or reflection. The fact of the matter is that the 'real world' is to a large extent unconsciously built up on the language habits of the group. No two languages are ever sufficiently similar to be considered as representing the same social reality.

Whorf (1940/1956: 213) also wrote the following influential passage:

> We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds--and this means largely by the linguistic systems in our minds.

Additionally, in another essay Whorf (1940/1956: 221) said:

> …users of markedly different grammars are pointed by their grammars toward different types of observations and different evaluations of externally similar acts of observation, and hence are not equivalent as observers but must arrive at somewhat different views of the world.

The quotations above from the writings of Sapir and Whorf present their views on the relationship between language and cognition. They argue that different languages divide reality in different ways. More specifically, the structures of language influence the way we think about the world. Sapir and Whorf based their claims on personal experience of the languages and cultures they described. They studied the languages and cultural practices of a selection of Native American tribes (such as Hopi and Navaho) in the first half of the twentieth century.

Whorf's interest in linguistics stemmed from his field work as a chemical engineer - more specifically an inspector - for the Hartford Fire Insurance Company, where his job was to analyse reports about circumstances that caused fires on factory premises. Whorf investigated an explosion caused by a worker who threw a cigarette butt into an 'empty' petrol drum. Whorf described the situation; in the workplace there

were two rooms for storing petrol drums, one for full drums and the other for 'empty' ones. Because of the presence of flammable vapour the 'empty' drums were actually more dangerous than the full ones. The workers, however, perceived them as less dangerous and smoked in the room with the empty petrol drums being unaware of the risk of their behaviour. Their concept of 'empty' had rendered them unable to see that the space was full of dangerous fumes. From this experience, Whorf concluded that linguistic terms along with physical conditions lead people to adopt certain behaviour in different situations. This was the starting point of Whorf's idea that other linguistic items and grammatical categories (e.g. plurality, tenses, or gender) could have an effect on how people think and interact in different conditions. Whorf (1956: 138) studied Hopi - an American Indian language spoken in Arizona - in an attempt to answer his hypothetical questions about language and thought. He considered (1) are our own concepts of "time", "space" and "matter" given in substantially the same form by experience to all men, or are they in part conditioned by the structure of particular languages? (2) … Are there traceable affinities between (a) cultural and behavioural norms and (b) large-scale linguistic patterns? Whorf compared Hopi with what he called the Average Standard European (ASE)[2] languages (e.g. English, French, and German) to test his hypothesis. In Hopi, plurals only apply to physical objects and not periods of time, therefore in Western languages one might say, "I spent two days writing" representing each individual day, while in Hopi the period of time would be included as a whole, "I finished writing on the second day" (Carroll, 1956: 139). Whorf found that Hopi speakers did not experience the passage of time in a way that is similar to speakers of languages that have time terms and concluded that the concepts of time and matter are not the same to all people, but rather that these concepts depend on the language in which they are developed.

Several scholars have, however, doubted Whorf's findings and argued that the accuracy of his analysis of Hopi and ASE is questionable. For example, Pinker (1994) pointed out that the Hopi language includes words/phrases that refer to time of day (sunrise), human ageing (child, old man) and season time (harvest). Consequently, even though the Hopi language does not have specific terms for hours of the clock and months of the year, the Hopi Indians do experience the passage of time. It seems that Whorf's assertions that people who speak different languages think differently are based upon word-for-word translations of those languages into English which Pinker (1994)

[2] This term was introduced by Whorf to group the modern, Indo-European languages of Europe as having many related features.

considers inadequate and lead to inaccurate conclusions, such as an Apache sentence meaning "He invites people to a feast" would be translated into English as "He, or somebody, goes for eaters of cooked food" (Pinker, 1994: 60). Whorf considered examples like that as a proof that Apache thinking must differ essentially from his own. These findings have been taken as evidence against the Whorfian hypothesis and as a result, the idea that language affects cognition lost its popularity and credibility in linguistics due to a lack of strong and convincing evidence.

The decline in the acceptance of the linguistic relativity hypothesis seemed to allow other opposing views (e.g. Chomsky's Universal Grammar and Piaget's theory of development) to spread and become more popular. These theories emphasised that the human conceptual structure is relatively similar across cultures and that it is combined with a semantic structure (Gentner & Goldin-Meadow, 2003). Such universal views received support from a range of different scholars. For example in cross-cultural studies Rosch [Heider] (1972) offered empirical evidence of universality among languages. Rosch published an influential paper showing that although the Dani people in New Guinea have only two basic colour terms (cool/dark and warm/light) - compared to eleven in English - they performed cognitive tasks as though their colour categories were similar to the English system (even though accuracy was much lower). Rosch found that the Dani's similarity groupings corresponded better with English colour terms than with their own. Furthermore, Dani speakers found learning new categorisation tasks easier when the categories were grouped around the English main colours. The implication of this being that perception of colour is not determined by the language learned, but rather by the biology of human colour perception. Such contradictory findings have encouraged researchers to critically review this hypothesis. Pinker (1994: 49) stated that "The idea that thought is the same thing as language is an example of what can be called a conventional absurdity". Language is therefore perceived as a translation of concepts rather than having any effect on their representation, "knowing a language, then, is knowing how to translate mentalese into strings of words and vice versa. People without a language would still have mentalese" (Pinker, 1994: 78, also see Fodor, 1975).

Lenneberg (1953) published a detailed criticism of a line of thought that had been essential for Sapir and Whorf. Lenneberg's main point was that Whorf's work had never shown the causality between a linguistic phenomenon and thought or behaviour, but merely assumed it to be there. Lenneberg and his colleague Roger Brown (Brown & Lenneberg, 1954) started the work of proving or disproving the existence of linguistic

relativity experimentally. They identified the two tenets of the Whorf hypothesis as (a) "the world is differently experienced and conceived in different linguistic communities" and (b) "language causes a particular cognitive structure" (Brown and Lenneberg, 1954: 455- 457). Brown (1976: 128) developed those concepts into the "weak" and "strong" formulations of linguistic relativity. In Brown's (1976: 128) summary, "Whorf appeared to put forward two hypotheses:

- Structural differences between language systems will, in general, be paralleled by non-linguistic cognitive differences, of an unspecified sort, in the native speakers of the language.
- The structure of anyone's native language strongly influences or fully determines the worldview he will acquire as he learns the language."

Brown's two formulations have become commonly known and attributed to Whorf and Sapir. The second, however, verging on linguistic determinism was advanced by neither Sapir, nor Whorf (see Alford, 1980).

Furthermore, Lakoff (1987), as one of those who adopted a more Whorfian approach, identified four factors on which researchers differed in their views of what constitutes linguistic relativity, the first being the degree and depth of linguistic relativity. Some researchers are satisfied that a few examples of superficial differences in language and associated behaviour are sufficient to show the existence of linguistic relativity. Others, however, assert that only deep differences that permeate both the linguistic and cultural system can be considered as evidence. A second factor is whether conceptual systems are to be seen as absolute or can be expanded or changed during people's lifetime. The third aspect concerns whether translatability is accepted as evidence of similarity/difference between concept systems; or whether it is instead the actual habitual use of linguistic expressions that is to be examined. The final consideration is whether linguistic relativity is viewed to be contained in language or in the mind, although this factor implies that language is distinct from the mind and there is no clear-cut definition of this notion. Lakoff (1987) concluded that since many of Whorf's critics had criticised him for using definitions of linguistic relativity that Whorf did not himself use, their criticisms were often ineffective.

It seems that a great deal of the confusion about the Whorfian hypothesis has resulted from Whorf's lack of specifics with regard to three aspects (Slobin, 1979). The first concerns the kind of linguistic facts that are being referred to; Whorf seemed to combine the lexical and grammatical levels of languages. The second involves the kinds

of other mental phenomena with which language is being connected. Specifically, in which way should language be seen? For example, should it be related to feeling and perception (e.g. colour division) or more closely connected to 'higher level' processes (e.g. memory and global world view)? The final aspect relates to the nature of the connection, should the connection be 'strong' or 'weak'. Consider Slobin's (1979) main view, which concerns the lack of evidence supporting the Whorfian hypothesis.

Based on the division of the linguistic relativity hypothesis into 'strong' and 'weak' versions, the weak version may not be easily rejected. Hakuta (1986: 77) is of the opinion that this view has more to do with "guiding our choice of alternative than with rigid determination". The effects of language, however, could be either permanent or temporary; the former takes place at the deep level of everyday 'habitual thought' and the latter presents at the moment of language use. For some researchers, this temporary effect seems to conform to Slobin's (1987-1996) idea of 'thinking for speaking'. According to this view, people depend on categories introduced in language in order to partition reality at the moment of speaking, reading, writing and listening. Slobin (1996) suggested a different phrasing by replacing *language* and *thought* with *speaking* and *thinking*. This substitution, according to Slobin, draws attention to mental processes which occur when formulating an utterance. It was also argued that this gave the advantage of allowing us to distinguish between linguistic and non-linguistic thought. In the main cognitive processes involved in accessing and choosing words, placing them in grammatical structures and planning speech are all examples of thinking for speaking. The process of thinking for speaking differs among languages. In English, for example, when planning to say a verb, one needs not to think of the grammatical gender agreement between the verb and the subject of the sentence. While in many grammatical gender languages - such as Arabic - speakers do need to plan in advance to construct a correct sentence and thus their thinking for speaking will be different from that of English speakers.

As the historical debate over the effect of language on thought continues, researchers should not focus their attention on the details of Whorf's analysis which were probably not accurate, but can be considered as of their period. They should rather focus on his general approach and ideas to investigate the linguistic relativity principle.

2.1.3 Language and Cognition: determinism vs. relativism

The Sapir-Whorf Hypothesis has changed the way many people look at language, has influenced many researchers and also opened up large areas of study. The extent to which language affects cognition is, however, still hotly debated, making it hard to confirm or refute the hypothesis. The two versions of the hypothesis proposed (strong and weak) have greatly contributed to constant arguments about linguistic relativity. The strong version, also known as linguistic determinism, views language as having a profound impact on human cognition. The weaker version, or linguistic relativism, suggests a moderate view for this relationship, arguing that language has some influence on thought. Some Universalists (such as Pinker, 1994), however, consider the hypothesis trivial, believing that the underlying mental representations are the same across all languages and cultures and that any observable differences in behaviour between speakers of different languages can be related to the use of language itself. Using the deterministic view as a basis, language acquisition shapes mental representations which then produce various differences among languages and cultures. Language therefore strongly affects and modulates representations at a neurological level. Nonetheless, there have long been concerns over such a strong view and several scholars have questioned whether it has actually ever been held by any researchers - including Whorf himself (Boroditsky, 2001; Levinson, 2003).

It is undeniable that human beings are biologically endowed with universal linguistic and cognitive abilities. In some languages different concepts may have different linguistic terms, while in others they may have only one linguistic term. This does not mean that speakers of different languages are unable to understand other terms that do not exist in their language. For example, according to Whorf, 'Eskimo' languages (Yupik and Inupiat) have multiple words for the English word *snow*; this by no means, however, meant that English speakers cannot distinguish the difference between these words. Nevertheless, findings obtained from several studies of language and thought such as Ervin (1962), Brown and Lenneberg (1954) and Brown (1957), even though they can be considered weak, have fascinated a wide range of researchers from a wide range of disciplines. Further expansion of this idea suggests that although cognitive universals exist among humans, there may still be room for remarkable ways in which differences in language may lead to differences in cognition. Instead of refuting the hypothesis absolutely, we can say that some aspects of language do indeed affect some aspects of thought (see Majid, Bowerman, Kita, Haun, & Levinson, 2004). Such a moderate view of the relationship between language and thought has encouraged

recent researchers to empirically investigate this hypothesis across different domains and using a variety of methods in an attempt to reach a consensus.

This recent research focuses on the effect of language on different cognitive processes emphasising the effects of language on non-linguistic processing, yet it does not argue for the effect to be innate or permanent. Fundamentally, it says that some thought processes are more likely to happen in some languages than in others (Hunt and Agnoli, 1991). For example if certain features do not exist in a language, speakers of that language can still be able to make them, either by putting in more effort, or by learning new ways and processes to do so. An extensive body of research (e.g. Brown, 1957; Carroll and Casagrande, 1958; Lenneberg and Roberts, 1956; Brown and Lenneberg, 1954; Ervin Tripp, 1961b; Rosch [Heider], 1972; Levinson, 1996; Imai and Gentner, 1997; Slobin, 1996; Kousta et al., 2008; Athanasopoulos, 2006) exists charting investigations into linguistic relativity showing evidence that either supports or refutes the weak version of the hypothesis. Sections (2.1.5) and (2.1.6) contain an analysis of these studies from the 1950s to the present day.

2.1.4 Investigating the Linguistic Relativity Hypothesis

The hypothesis of linguistic relativity has been widely investigated across different domains. In addition to the traditional colour domain (Rosch [Heider], 1972), research has included areas such as spatial relation, time, motion events, number systems, grammatical gender and shapes. Investigation into these different domains has provided supportive evidence for linguistic relativity and challenged the Universalist view so powerfully believed in the past. To that end, Gopnik (2001: 45) was right when she stated that "after decades of obloquy, Benjamin Whorf is back". In order to examine linguistic relativity, however, one needs to provide an unbiased frame of reference to compare languages. Lucy (1997- 2011)[3] identified three strategies for research into linguistic relativity: the structure-centred, domain-centred and behaviour-centred strategies.

a. The Structure-centred strategy

This strategy selects some grammatical structures (e.g. number or gender), asks how it varies across languages and how reality might be presented differently from each

[3] In a later article by Lucy (2011), he introduced only the first two strategies and excluded the behaviour-centred strategy.

relevant system. Structure-centred strategy is based on a long tradition of comparative work in linguistics; when different linguistic forms in a specific area of meaning are observed, then different interpretations of reality may be discovered. Implementing this strategy is, however, difficult because comparing categories across languages requires extensive linguistic work in terms of both local description and typological framing. This can be disrupted by blindness to categories very different from one's own and may not easily "yield referential entailments suitable for an independent assessment of cognition" (Lucy, 2011: 49).

Conversely, a structure-centred strategy holds the most potential for respecting linguistic differences and therefore assures the greatest promise for identifying structural differences and directing the investigation of cognitive influences in suitable ways. A well-known example of this approach is Whorf's (1941, 1956) comparison of number marking pattern in Hopi and English. Another example of structure-centred research is Lucy's (1992b) study comparing the relationship between grammatical number marking and cognition among speakers of American English and Yucatec Maya. The findings showed that Mayan speakers classified objects according to material, rather than shape as preferred by English speakers. The present study adopts this strategy as a framework for investigating the relativity principle by choosing a grammatical gender system to examine its possible effects on cognition (see chapter 3 for details).

b. The Domain-centred strategy

This strategy selects a semantic domain (e.g. colour, space, or time) and compares it across languages and cultural groups in an attempt to find correlations between language and cultural behaviour. The aim of this is first to ask how individual languages treat the domain and then explore how speakers will treat that domain cognitively during some activities (Lucy, 2011). A domain-centred strategy endeavours to solve comparison problems by enquiring how different languages partition the same domain of reality. This strategy suffers, however, from two drawbacks; the first being that the domain is usually represented through a single linguistic and cultural tradition. One would wonder whether the domain or its representations are commonly recognised. Some researchers acknowledge this problem and tend to describe the domain using well-established scientific concepts in order to guarantee objectivity and neutrality. Second, using one vision of reality as the standard for comparison essentially prioritises the original language and culture from which it arose. This, according to Lucy, leads to

many demonstrations of differences in which a hierarchy gently (re)appears. The best-known type of research adopting the domain-centred approach is that on colour (see section 2.1.7, for a discussion of some of these studies). It was originally stated through the work of psychologists Brown and Lenneberg (1954) which addressed challenges to the Whorfian methodology. Since the late sixties and hitherto, work on the domain of colour has continued and some findings have gone against the Whorfian hypothesis, such as those obtained by Berlin and Kay (1969). Other recent studies, however, have revisited the Whorfian hypothesis and new evidence relating to the effect of language on thought has been shown (e.g. Athanasopoulos, 2009).

c. The Behaviour-centred strategy

This strategy starts by observing different behaviour between linguistic groups and then searching for possible reasons for that behaviour in terms of that language. Bloom (1981- 1984) adopted this strategy by noticing behavioural differences between speakers of English and Mandarin Chinese when answering counterfactual questions - which refers to a mode of thinking that is literally in opposition to a fact - in questionnaires on political situations. Bloom (ibid) noted that the Chinese language does not differentiate between counterfactual and other implicational relations. He carried out counterfactual experiments on speakers of Chinese and English using two versions of a counterfactual story. His findings indicated that Chinese speakers had more difficulty interpreting counterfactual premises due to the way in which counter-factuality is marked grammatically in the Chinese language. Additionally, Stromnes (1974a), examined the reasons for the higher occurrence of occupational accidents in Finnish factories than in similar Swedish ones (Salminen and Hiltunen, 1993, cited in Lucy, 1997). Stromnes claimed that cognitive differences between the grammatical usage of Swedish prepositions and Finnish cases might have led organisers of Swedish factories to focus more on the work process, whereas Finns organise the workplace in a way that focuses on the individual worker. Such a study is an excellent example of a behaviour-centred approach that compares a practical behavioural difference between groups and seeks to examine it in a known language difference (Lucy, 1997).

2.1.5 Early Studies of Linguistic Relativity Hypothesis

Most criticism and confusion about linguistic relativity research is due to misinterpretation of its original proposals. Hill and Mannheim (1992) asserted that

researchers including Whorf and Sapir admit the existence of linguistic and cognitive universals and the aim of their research is to find any possible effects of language on thought, even though these universals exist. Further research is, therefore, still needed in this arena to provide insight into the different circumstances under which language effects appear to be greater or weaker.

One of the early researchers who empirically investigated the effect of language on cognition was Brown (1957). He studied the effects of some grammatical categories on individual behaviour. Although his study only included English-speaking children, it revealed an effect of lexical categorisation on the inferred meaning of new words. Pictures of actions, objects and substances were used as stimuli; each picture showed an action being performed on a substance in a container; children were then shown three additional pictures and asked to choose those most similar to the first picture, either one with the same action, one with the same substance or one with the same container. It was found that children selected the pictures with the same action, discovering semantic implications of verbs, mass nouns and count nouns in their language. Brown linked the results of his experiments to the linguistic relativity hypothesis arguing that grammatical categories are clearly different in different languages, so speakers of those languages may have quite different cognitive categories. That suggests that the grammatical categories of a language would probably affect the cognition of the speakers of that language; nevertheless, "it remained to determine how seriously and how generally thought is affected by these semantic distinctions" (Brown, 1957: 5).

Other studies compared non-linguistic behaviour between speakers of different languages. Carroll and Casagrande (1958) studied the performance of adult speakers of Hopi and English using tasks that did not require the use of language to be performed. Participants were asked to classify action pictures, in which verbal descriptions varied between Hopi and English, by deciding which of the three pictures were more similar. Categorisation responses were expected across the two groups according to the naming patterns of their respective languages. Each set of items included three pictures, with one set including (A) a picture of a person spreading topping over a cake, (B) a picture of a person painting a picture on a vase, (C) a picture of a person painting a wall. As predicted, English speakers chose B and C because both pictures depicted the action of painting. The Hopi speakers paired A and C as they both showed the Hopi action *leluwi* which in English means 'to apply or to spread over a surface'. The authors argued that different verb meanings between Hopi and English essentially affected how their speakers grouped pictures depicting actions.

In another experiment, Carroll and Casagrande (1958) compared the performance of English and Navaho children using triad pictures of objects. Navaho speakers, when talking about handling an object, have to add a suffix called a 'classifier' to the verb corresponding to the shape or some other attributes to the objects. The authors hypothesised that this obligatory use of a shape classifier might make Navaho speakers group pictures of objects by shape more than by size and colour, English speakers were expected to perform the opposite. Children were shown a pair of objects varied in size and form, e.g. *Yellow rope and blue stick,* then they were asked to which of the two objects should they place a *blue rope?* Findings went hand in hand with the author's hypothesis i.e. Navaho-dominant Navaho children were more likely to group objects according to the shape (70% selected the yellow rope), whereas English-dominant Navaho children preferred the opposite pattern grouping the objects based on their size and colour (40% selected the yellow rope). Monolingual American English speaking children from Boston, Massachusetts, were tested using the same task. Unexpectedly 80% of these children chose the *yellow rope* (form), which means that their responses were highly similar to the Navaho dominant group (grouping objects based on the shape). This part of the study went against the Whorfian hypothesis and the authors commented that preference for shape increased with age, that is Navaho children favoured shape at around three to four years of age, while English speakers did not until nine years of age. In an attempt to explain this unpredicted result, Carroll and Casagrande (1958) maintained that the American English children's preferences might be attributed to their experiences with toys of various shapes. They further argued that shape and colour preferences could be influenced by experience in using various objects alongside shape-class in language. Such an interpretation, however, may not be strong enough to support the linguistic relativity hypothesis, meaning that further replication of their study would definitely give better insight into the nature of this effect.

A common argument for the Sapir-Whorf Hypothesis is perception of colour across languages. Lenneberg and Roberts (1953) studied Zuni (a Native American language) colour terms and colour memory. Their study compared Zuni speakers and English speakers and provided supporting evidence to the relativity hypothesis. The Zuni language has only one word for the English *yellow* and *orange*; unlike English speakers, Zuni speakers often confused the two colours in the recognition task. Furthermore, the performance of Zuni-English bilinguals was somewhere in the middle between monolingual speakers of English and Zuni.

In another study, Brown and Lenneberg (1954) investigated the effect of codability on recognition, more specifically on English colour terms and colour memory. They expected that colours described with shorter terms might be remembered and recognised more easily than other colours with longer names. In the experiment, they showed English participants colour chips for a few seconds and then the chips were removed. After a short time, participants were asked to find the same colour chips in an array of 120 Munsell chips (a standardised set of chips of different colours) mounted on a card. The findings showed that colour chips which were more codable (those which have short, reliable and well-known names were accurately remembered. Brown and Lenneberg concluded that the availability of basic colour terms in a language affected the remembering of these colours in recall tasks. They related their findings to those of Lenneberg and Roberts' (1953) and pointed out that language had an important role to play in shaping cognition across languages. Nevertheless, Lenneberg (1961) reconsidered his earlier work with Brown (1954) and found that when a different array of colour chips was used, the relationship between codability and memory was wiped out. Lenneberg concluded that language effects on cognition did not seem to be a general phenomenon. Based on this reconsideration of Brown and Lenneberg's work, support for linguistic relativity was deemed to be weak and unreliable, yet 'codability' was valued and had a great influence on later research.

Other research studied the effects of colour on memory by including both monolingual and bilingual participants. For example Ervin-Tripp (1961) conducted a study on Navaho-English bilinguals, Navaho monolinguals and English monolinguals focusing on colour names and pictures recall. The Navaho language has only one colour name for the English *green, blue* and *purple*. Data was collected on Munsell colour chip naming and on reaction times in naming from both monolingual and bilingual groups. The findings revealed that Navaho-English bilinguals, who were dominant in English, called a certain colour *green* considerably more often than the other groups, who called the same colour stimulus *yellow*. Ervin (ibid.) argued that it was an effect of recent acquisition of English colour names and commented that there was a semantic change among Navaho-English bilinguals who performed in a way that is in between Navaho monolingual and English monolingual groups. This study provided evidence for the linguistic relativity hypothesis in monolinguals and bilinguals alike.

Although the effects of language on cognition in the domain of colour were found in the above studies, a series of cross-cultural studies by Rosch [Heider] (1972a, 1973, 1975a) found no language effect and rather that these studies indicated universality

across languages. Rosch compared the performance of Dani and English speakers using colour categorisation tasks. The findings showed that even though Dani speakers have only two basic terms for all English colour terms, their cognitive organisation of colour did not differ from English speakers. Rosch argued that these results indicated universal perception of focal colours despite the fact that the performance of Dani speakers on memory task was significantly low (5%) compared to English speakers (28%).

It is clear that the domain of colour has been widely investigated in relation to linguistic relativity; it was studied from the 1950s to the 1980s using a variety of methods across different languages. Using a different methodology, investigating differences in colour perception and languages, Kay and Kempton (1984) studied English and Tarahumara speakers of Mexico using a matching-pairs triad task. Tarahumara speakers use the same word to identify the colours *green* and *blue*, so they lack a lexical boundary at this position in colour space. Kay and Kempton (1984) used a task that included a set of colour stimuli ranging across the boundaries between the English two colours *blue* and *green* in equally spaced steps and asked participants to observe three stimuli (*green, bluish green* and *blue*) and then to decide which two samples were most similar. The findings revealed that English speakers perceived colours that cross the lexical boundary between *green* and *blue* to be less similar than Tarahumara speakers which indicated that they were systematically affected by the lexical boundary of English.

Moving into another area of investigation, Bloom (1981) studied the effect of counterfactual expressions[4] on the behaviour of Mandarin Chinese and English speakers when he noticed some behavioural differences in questionnaire responses by Chinese and English speakers. Bloom (ibid.) found that unlike English speakers, Chinese speakers had more difficulty in distinguishing factual and counterfactual events as well as in interpreting them. Bloom's findings were, however, open to question and were criticised by many researchers (e.g. Au, 1983; Liu, 1985; and Takano, 1989). To give an example, Au (1983) examined the stories that Bloom used and argued that the study had several serious flaws; the Chinese translations of the stories were not written in a similar idiomatic way to the original English ones, which may have contributed to the lower rates of counterfactual interpretation by Chinese participants.

Linguistic relativity was also investigated in the domain of emotion. Ervin-Tripp (1964) conducted a study using the Thematic Apperception Test[5] (TAT). The test

[4] For example, 'if it had been fine yesterday, we would have had a barbecue'.
[5] The Thematic Apperception Test is a projective test generally used in psychological assessment.

consisted of a series of neutral pictures of people who expressed different emotions. The participants were two groups of adult bilinguals in the United States; one group was Japanese-English and the other French-English. The test was given to the groups in both of their languages and their answers varied according to the language used in each session. For example, in the Japanese session with the Japanese-English group, it was found that they used more emotional expressions and themes associated with family relationships. On the contrary, the English session with the same group produced abstract and cold stories about formal relationships between people. Similarly, the responses of the French-English group differed in the two sessions (the test was a picture of a couple featuring a twenty-seven year old French woman who spoke English with her husband and child). In the French session, the picture elicited a variety of themes of aggression and striving for autonomy, while in English the heroine supports the husband striving for achievement (Ervin Tripp, 1964: 504).

2.1.6 Recent Research on Linguistic Relativity

One language feature that has no clear relation with meaning is grammatical gender. The focus of this research project is on the Arabic grammatical gender system and evidence concerning the effect of grammatical gender categories on thought is separately and extensively presented in the second part of this chapter (2.2).

The effect of language on cognition has also been examined in different domains; most common are the domains of colour and shape in which linguistic relativity seems plausible. Interest in the domain of colour might be attributed to the fact that in the physics of light, there are no stated rules for drawing boundaries between colours at one place rather than another. Different languages have colour terms which divide the colour spectrum differently. It has been argued, therefore, that since there is nothing that can determine how people should think of colours; differences in their colour cognition are likely to be caused by their languages. Davidoff, Davies, and Roberson (1999) replicated earlier work by Rosch [Heider] (1972) investigating the domain of colour by comparing English and Bernimo - a small tribe in Papua New Guinea - speakers, whose language has five basic colour terms compared to eleven in English. Their findings indicate the strong effect of language on the cognitive processing of colour and shape. It was found that constant differences in a variety of perceptual and memory tasks were systematically related to colour categories in each language and culture (see also Roberson, Davies & Davidoff, 2000). In another cross cultural study, Roberson,

Davidoff, Davies and Shapiro (2005) studied the linguistic categories of colour in Himba and English and showed that Himba participants indicate categorical perception only for their own linguistic categories which are different from English as well as Berinmo (previously studied by Roberson et al., 1999). The authors observed that although two languages may appear to be very similar, speakers of the two languages encode, remember and categorise colour stimuli differently. There has, however, been an increasing debate over whether these effects are actually perceptual. For example, Munnich and Landau (2003) pointed out that the participants in Roberson's study were engaged in explicit speech practice of colour names during the thirty second interval before testing their memory, so the task was verbally mediated. This, according to Munnich and Landau (2003), did not provide any support for linguistic relativity (showing effect of language on non-linguistic forms of cognition); rather it only revealed language effects on verbally mediated tasks.

In a similar line of research, Winawer et al. (2007) compared the performance of Russian and English speakers using a colour discrimination task. Russian language differentiates *dark blue* 'siniy' from *light blue* 'goluboy', a difference that resulted in colour discrimination. Unlike English speakers, Russian speakers were faster on a matching task when the colours belonged to different linguistic categories than when they belonged to the same category. Other research by Gilbert and colleagues (2006), however, showed that these cross-linguistic differences disappeared under conditions of verbal interference. As language is mainly processed in the left side of the brain, and each visual hemi-field is perceived by the contralateral cerebral hemisphere (the right side of the brain receives information from the left optic nerve), Gilbert et al. (ibid.) argued that colour perception in the right visual hemi-field might be more affected by language than that in the left visual hemi-field. Their results showed that English speakers were faster to find a target when its linguistic category was different from that of the surrounding distractors (e.g. a green among blues) and slower when the target and distractors had the same linguistic category (e.g. a green among other shades of green); nonetheless, that was only when the target was presented in the right visual field. The study suggested that when using a forced left visual field rather than the right visual field, possessing verbal memory could lessen the impact of verbal classifications on perceptual memory because they most probably need left side processing. Gilbert et al (2006) argue that, unlike other studies which tended to look for a simple yes/no answer to the Whorfian question; their findings suggest a more complex picture, based on the

functional organisation of the brain. This is raised in the study by Winawer et al. (2007), showing that the effect of language was eliminated by a verbal interference task.

In addition to the colour domain, plenty of evidence has shown the effect of language on cognition by comparing different ways to mark grammatical numbers in different languages. Most notable in this vein is Lucy (1992), who studied differences in grammatical number marking in both Yucatec and English. His experiments involved sets of pictures describing different scenes of everyday life. Each set of pictures consisted of six drawings - one original and five alternates - which were different from the original drawing. The participants were asked to choose which of the five alternate drawings was most similar to the original. Lucy (ibid.) predicted that if differences are reflected in the speakers' cognitive processing, then English speakers would pay more attention to the number of inanimate objects than Yucatec speakers. The findings revealed that Yucatec speakers were sensitive to the number of animate objects, but not to the number of inanimate objects or to the amount of substance. English speakers, however, were sensitive to the number of animate and inanimate objects, but not to the amount of substance.

Imai and Gentner (1997) carried out experiments on the same lines using triads. The study considered speakers of English and Japanese at different ages to observe any cognitive development and language effects. They analysed differences in noun forms between English, Japanese and Yucatec. English participants pluralise most animate and inanimate discrete objects obligatorily, whereas Japanese and Yucatec alike only pluralise humans and some animals optionally. Furthermore, the authors prepared three types of standard, or pivot objects and presented them with nonsense word names. The three types were simple objects (e.g. a pyramid), complex objects (e.g. a wood whisk), and a substance (e.g. sand in an S-shape). Participants from both language groups chose an alternative based on shape for the complex objects. Their responses were, however, considerably different in the simple object and substance trials. English speakers across different age groups, from infancy to adulthood, treated the simple objects like complex ones. This was unlike the Japanese speakers who treated them as between the complex objects and the substance. The substance trials found that Japanese speakers - with the exception of the two-year-old group - constantly showed material-based categorisation, while the English groups did not show such preference patterns. Imai and Gentner (1997) commented that children generally preferred shape-based categorisation when they are at the very beginning stage of learning their first words (at the age of two),

however, their categorisation should be affected by more aspects of language as they acquire it.

In addition, the effect of language on thought has been found in the domain of artefacts. Labels given to objects vary cross-linguistically and such variations were found to affect speakers in object naming tasks which could reflect conceptual similarities between those objects. Malt, Sloman and Gennari (2003) studied how sets of common household objects such as *bottles*, *jars* and *containers* are labelled by speakers of American English, Argentinean Spanish and Mandarin Chinese. The data showed that there were some general similarities in the naming patterns given to these objects and that there were some systematic differences in the labels given to objects in different languages. These differences did not, however, indicate that some languages make better distinctions between types of objects than others. Based on these findings, the authors argued that some languages may use very different linguistic categories "forming their categories around different dimensions or combinations of dimensions, or simply following some language - or culture idiosyncratic paths in the evolution of their linguistic category membership that the end result is substantially divergent category membership" (Malt, Sloman and Gennari , 2003: 22). The findings obtained from different languages in the domain of artefacts did not provide strong evidence of the effect of language on thought, unlike the domain of colour which showed the strong effect of language on colour perception.

Additionally, one of the ways in which languages differ greatly is in describing spatial locations. For example, languages such as English and Dutch use relative spatial terms in describing the relative locations of objects (e.g. *left/right, back/front*); whereas other languages use an absolute reference (the equivalent of *north*/*south*) such as Tzeltal (a Mayan language in Mexico) which has mainly two absolute reference terms '*uphill/downhill'* which in English will roughly mean '*north/south'*. This use of absolute direction is similar to Guugu Yimithirr (an Australian language) which only uses absolute directions to describe spatial relations i.e. objects are described by using the cardinal directions. For instance, a speaker of English may describe a person as being *in front of the house,* whereas a speaker of Guugu Yimithirr describes a person as being *north of the house*. It has been argued that this difference makes speakers of Guugu Yimithirr better at finding and describing locations in open land, whereas speakers of English perform better in tasks describing objects relative to the speaker.

To investigate whether these differences between languages have a cognitive effect, Levinson (1996) compared speakers of Dutch and Tzeltal via a series of spatial

tasks. He found that Tzeltal speakers relied heavily on absolute reference in their spatial descriptions, while Dutch speakers used a relative description of spatial relations, showing parallel responses to the directional expression of their languages. This provides further evidence of the effect of language on cognition in the domain of space. These findings were, however, questioned by Li and Gleitman (2002), who argued that Levinson's tasks were carried out in different environmental conditions (laboratory vs. outdoors) which could have constrained participants to choose a particular frame of reference over others. Using the same tasks with English speakers (which is similar to Dutch), findings showed that English speakers did use absolute directions when tasks were carried out in outdoor contexts rather than in laboratory conditions. This data suggests that the frame of reference available in some languages may impose vital constraints on the spatial thinking of speakers. Haviland (1993) reported that English speakers are never able to describe how a ship has turned[6] exactly from one direction to another by using north, south, east and westerly directions; conversely, Guugu Yimithirr speakers would be able to do this. Consequently, the effects of spatial language on cognition should not be overlooked, even though Li and Gleitman's (2002) study showed different results with English speakers using both relative and absolute directions in certain contexts. Conversely, Levinson, Kita, Haun and Rasch (2002) responded to these findings by arguing that Li and Gleitman's (2002) study oversimplified the original tasks which may have allowed the participants to think the aim of the task was to test their memory and spatial direction. In a further experiment, it was found that task instructions have a crucial effect on results. Levinson et al. (2002) demonstrated that access to linguistic information happened when there was a requirement to keep information in mind in order to be able to provide linguistic descriptions when needed.

Another important theme in linguistic relativity research is the semantic theory of motion and manner put forward by Leonard Talmy (1985). Talmy proposed that languages fall into two types on the basis of how they encode two aspects of motion – 'manner' and 'path'. These types are verb-framed and satellite-framed languages, a distinction which refers mainly to how verb phrases incorporate the meanings of the path of motion or the manner of motion in different languages. English, for example, is a satellite-framed language. It usually includes manner as part of the meaning of a verb such as 'walked' and uses particles to show the direction of motion (e.g. 'run into', 'go

[6] In English, ship-direction is described in terms of 'port' and 'starboard' instead of the left/right sides respectively.

out', 'fall down'). On the other hand, verb-framed languages such as Arabic, Korean, Spanish and Greek make heavy use of verbs of motion, e.g. Arabic ‘dakala’, ‘kharaja’ (go in, go out) and express the manner of motion through the use of a participle e.g. ‘dakala rakidan’, literally meaning ‘he entered running’.

Choi and Bowerman (1991) studied how common motion verbs differ between Korean and English and how children from both language groups learn to express motion events. Korean is a verb-framed language as suggested by Talmy’s typological distinction. For instance, to describe a situation in which a cassette tape is placed into its case, Korean speakers usually use the verb *kkita* to refer to the ‘put in’ relation, whereas English speakers would say ‘we put the tape *in the case*’. Although both verbs *put in* and *kkita* can be used to describe an act of an object in a location, the Korean word *kkita* does not have the same extension as English *put in*. This means that while the English word *put in* can be used for all cases of containment (e.g. flowers in a vase, fruit in a bowl); *kkita* is only used to express a tight fit. The findings revealed that Korean children and English children learn to talk about motion events in different ways. Both groups realised their respective language-specific classifications of such motion relations and events, which was clear by their usage and comprehension (Choi and Bowerman, 1991).

In an experiment investigating mental imagery of motion events, Slobin (2000) asked English monolinguals and Spanish-English bilinguals to summarise a fragment of a novel. English participants specified the actions of the protagonist and reported the story by using a number of manner verbs, while only a few Spanish participants did so. Furthermore, Spanish participants reported the events in a static manner although clear descriptions of the scenes were provided in the story. Such findings showed that speakers of different languages focus on different aspects of motion in these tasks. The way that Spanish-English bilinguals used manner verbs in the reports was found to depend on the language of the retelling. This suggests that language does play a role in how one thinks about motion events by focusing attention on those aspects that are encoded in language more saliently.

Conversely, the effect of language on cognitive tasks was found to be limited and constrained in other domains, e.g. Gennari, Sloman, Malt and Fitch (2002) who found language-specific effects in similarity judgments in triads by Spanish speakers. This effect occurred only, however, when participants verbally described target motion events before recording their similarity judgments. Similarly in the domain of artefacts, Malt, Sloman and Gennari (2003) found some general similarities in the naming

patterns given to objects by speakers of English, Spanish and Chinese. These findings suggest that the effect of language on thought varies between different domains, that is from a stronger effect in the case of the colour domain to a more limited effect in the domain of artefacts, where the effect might be temporary and depend on task demands. It seems that if linguistic categorisations have a clear relation to meaning (e.g. spatial concepts in Korean and English), the effect of language on thought might be stronger and if there is no direct relation between them the effect could be limited and transitory (e.g. grammatical gender).

In the light of the studies mentioned above, it can be noted that enough evidence has shown that some aspects of language influence some aspects of human cognition. This effect seems, however, to depend on many factors, such as the type of domain under investigation as well as the instructions (to be discussed further in sections 2.2.6 and 2.2.7) used in the cognitive tasks. The effect of language on thought seems quite complex. Researchers should consider various important factors before investigating the relationship between language and cognition. Cognitive tasks, for example, should not include linguistic stimuli and task instructions should be kept to the minimum. In addition, recent research has shown that other variables might affect participant performance in cognitive experiments: these include age, cultural environment, language competence, bi/multilingualism, cultural exposure and many others. Most previous studies investigated the relationship between language and thought by testing two monolingual communities to determine whether a particular linguistic difference between the two communities influences cognitive performance, often overlooking the crucial role of bilingualism in this field. It should also be noted that research considering this issue has recently been appearing across different domains and languages. The next section briefly outlines some of these recent studies on bilingual cognition.

2.1.7 Linguistic Relativity and Bilingualism

There is a sufficient body of evidence regarding the relationship between language and thought across various languages and cognitive domains. Most notable is the domain of colour, which suggests a strong effect of language on categorical colour perception across different languages (including Russian, Bernimo, Himba and Greek) in comparison with English. Most recent studies suggest that a clear influence of language on colour processing is quite plausible, though this does not necessarily mean that

speakers of different languages see different colours. Such studies (e.g. Roberson and colleagues, 2000, 2004, 2005; Athanasopoulos, 2009) have contributed to the revival of the field of linguistic relativity. Furthermore, the domain of colour should be investigated across other language families such as the Semitic[7] languages to better understand the extent to which we can generalise the existing findings.

Earlier research on the relationship between language and cognition was mostly carried out by means of a linguistic task using monolingual participants. In such tasks, the relevant linguistic dimension seems to produce the anticipated effect. Kousta et al. (2008), however, adopted a relatively different approach to investigating the effects of grammatical gender: they tested both monolingual and proficient bilingual Italian speakers using the same linguistic task (an error-induction experiment), after which they made inferences about the effect of language on cognition. This was based on the extent to which the performance of the bilingual speaker patterns with that of monolingual speakers.

In another paradigm of research, Athanasopoulos (2006) conducted a study on grammatical numbers with a different language (Japanese) and with bilinguals. The Japanese language does not obligatorily mark plurality. Plural forms often refer to humans and seldom to animals, but the plural is never used with words having inanimate referents. Consequently, Japanese monolingual speakers are more sensitive to changes in the number of humans and animals but not to artefacts. The situation is different in English where plural morphology is used with animate as well as inanimate referents, with substances as the single exception to pluralisation. Monolingual English speakers are therefore more sensitive to changes in the number of both animate and inanimate entities. In his study, Athanasopoulos (ibid.) asked monolingual English, monolingual Japanese and Japanese-English bilinguals (at intermediate and advanced levels of proficiency) to match pictures according to their similarities. The results revealed that advanced Japanese-English bilinguals tended to behave in a similar way to the English monolinguals, whereas intermediate bilinguals acted like the Japanese monolinguals. Athanasopoulos, however, was very careful about interpreting these findings as evidence for a role of language on non-linguistic cognition. It was pointed out that it is possible that language was used to mediate the non-linguistic task because the language of instruction seems to play a crucial role in the response to non-linguistic tasks (see Athanasopoulos, 2001; Cook et al., 2006). Furthermore, if the task was

[7] This includes Arabic, Hebrew, Aramaic, Amharic and Akkadian.

linguistically mediated, there is uncertainty as to whether it is informative about anything other than using language as a strategy to completing the task at hand.

A recent study on motion by Czechowska and Ewert (2011) included Polish and English speakers. Polish, like English, is a satellite-framed language, but has fewer path verbs and lacks the very common motion verbs in English *go, come,* and *get*. The Polish language also uses verb prefixes as satellites. The study explored the effects of language on cognition through two satellite-framed languages; it is therefore unique in that it sees degrees of 'satelliteness' for languages rather than a case of either/or. The results show that English speakers focused on the path of motion more than monolingual speakers of Polish. The findings also indicated that L2 users of English shifted towards an L2-based conceptual representation when compared to those who were only learners of English rather than users. They would then pay more attention to the path of motion than monolinguals of their native language. Czechowska and Ewert (2011) concluded that when bilinguals use their second language, they change the way they think to some extent, independently from the language they speak.

More research on bilingual cognition in relation to grammatical gender is presented in the next section (2.2.7). It seems that the majority of research carried out on language has shown that it plays an essential role in people's lives; in addition to allowing communication, it provides different ways to mentally store and remember information. Differences in languages can entail differences in the ways in which information is stored in the human mind.

2.2 Grammatical Gender

2.2.1 Introduction

The term *gender* most commonly refers to classes of nouns within a language which are "reflected in the behaviour of associated words" (Hockett, 1958:231). In some languages, biological and grammatical gender are closely related and in some others they are completely unrelated. Grammatical gender is considered by many researchers (e.g. van Berkum, 1996) as the property of individual nouns, not their referents. Gender assignment, particularly to inanimate nouns, is largely arbitrary and independent of the referents' conceptual properties (ibid.).

One of the testing grounds for investigating the possible effects of language on cognition exploits the fact that grammatical gender systems vary considerably between languages in terms of the number of gender distinctions individual languages make, as well as the degree to which grammatical gender correlates with biological gender across languages (de Groot, 2011). A language may have two or more classes of nouns that are considered as genders, but some languages may lack a distinct gender 'system'. In fact, the notion of gender as introduced by linguists is significantly more general. Hockett's (ibid.) definition can generally encompass all noun categories that linguists consider to be *genders*, whether they are labelled 'masculine', 'feminine', 'neuter', 'common', or 'class IV'(Foundalis, 2002: 304).

Corbett (1991) differentiates between languages with semantic gender systems and those with formal gender systems. English and Mandarin are examples of languages with semantic gender systems where the gender is encoded in linguistic elements only for referents in terms of biological sex. In English, for example, grammatical gender mostly plays a role in choosing the third person pronouns *he*, *she*, and *it*, largely based on natural gender (De Groot, 2011). There are, however, some exceptions such as *ships*, *cars* and *the Moon*, which are all feminine in some styles and babies which are frequently *it*. Some other nouns are lexically assigned to refer to male/female entities (e.g. *brother–sister*). These gender assignments are maintained in the pronominal system as a distinction between obviously male and female entities (e.g. *he/she*) and everything else (*it*) (Vigliocco et al., 2005). Conversely gender assignments in languages with formal gender systems apply to all types of nouns whether their referents have or do not have a biological gender. In such instances, gender assignment - especially for inanimate - nouns seems to be mainly arbitrary. We may find therefore that some languages assign a different grammatical gender to what seem to be the same entities (Boroditsky et al., 2003), as is the case for the word *village* which is feminine in

Arabic *qariyah*, masculine in Hebrew *kfar*, and neutral in Russian *selo*. This arbitrary assignment of grammatical distinctions for items that do not have a biological gender can provide a test case for the study of the relationship between language and cognition. If this assignment is really arbitrary, then classifying a noun as feminine or masculine should relate neither to its semantic meaning nor to the conceptual representation of its referent. Investigating this area through a variety of cognitive tasks that involve speakers from different languages might, therefore, reveal whether such an arbitrary assignment of gender will have long term effects on the semantic and conceptual representations related to those nouns.

Previous research on grammatical gender focused mainly on two key areas. One area concerned the degree to which processing grammatical gender in a speaker's native language might influence the way they perceive and categorise physical objects (e.g. Clarke, Losoff, McCracken, and Still, 1981; Sera, Berge, and Castillo Pintado, 1994; Boroditsky, Schmidt, and Philips, 2003). The other investigated the effects of grammatical gender on language processing (see Costa, Alario, and Sebastian-Galles, 2007) and suggested that grammatical gender is not included in semantic or conceptual representations, but is rather accessed during linguistic processing. Additionally, there have been some studies (e.g. Vigliocco et al., 2005) that focused on investigating these two areas together to obtain a clearer understanding of the possible effects of grammatical gender on cognitive processing. Most studies that used language in cognitive tasks have, however, been criticised by researchers into linguistic relativity who argue that using linguistic stimuli in cognitive experiments is unreliable, as this might measure the speakers' knowledge of the language rather than the effect of the language on their thinking.

The present study addresses the question of whether the grammatical gender system in languages has an invasive and lasting effect on speakers' cognitive representations of objects. It looks at investigating this issue in bilinguals to see whether learning a second language can restructure the bilinguals' mind and lead them to think differently from their monolingual counterparts.

The next section explores the role of grammatical gender in languages and is followed by a separate section on the Arabic grammatical gender system. A third section reviews some of the literature on the effects of grammatical gender on cognition. Then, a fourth section describes the effect of grammatical gender on bilingual cognition. Finally, the motivation for the current study is highlighted and concludes this chapter.

2.2.2 The Role of Grammatical Gender in Languages

Vygotsky ([1934] 1962) demonstrated the effects of the interdependence of grammatical and semantic aspects of language by giving two examples of how changes in formal and grammatical structure lead to profound changes in meaning. The first example is in the fable, 'The Dragonfly and the Ant'. Krylov substituted the dragonfly for La Fontaine's grasshopper while retaining the inapplicable epithet "the jumper". In French, the word grasshopper is feminine, so it is a suitable term to embody the image of a carefree attitude and feminine light-headedness. In Russian, however, the grammatical gender of "grasshopper" is masculine, so this nuance of meaning critical to the illustration of frivolity would have disappeared had the fable been translated literally (Vygotsky, 1962, p: 221-222). Consequently, Krylov took grammatical gender over actual meaning and substituted the dragonfly for the grasshopper while preserving characteristics of the grasshopper such as jumping and singing that are obviously not characteristic of the dragonfly. To adequately translate the sense of the tale, therefore, feminine grammatical gender had to be preserved.

In addition, Vygotsky ([1934] 1962) cited a similar case in the Russian translation of Heine's poem 'The Fir and the Palm'. In German, *fir* 'tanne' is masculine, meaning that the poem represents love for women. To preserve this sense for the German text, Tiutchev substituted a *cedar* for the *fir*, since in Russian *cedar* (kedrovogo dereva) is masculine. In contrast, by translating the poem literally, Lermontov lost this sense. Accordingly, his translation gives the poem a deeply different sense, one that is more abstract and generalised. According to Vygotsky (1962, p: 221-222) "a change in a single, seemingly insignificant, grammatical detail can lead to a change in the whole meaningful aspect of speech". What Vygotsky meant by grammatical detail is in fact grammatical gender which is a property of nouns and whose functions are mostly syntactic and morphological. It is a grammatical category which in some languages is both essential and pervasive (as shown in the examples) whereas in others it is completely absent (Corbett, 1991).

Speakers of different languages must therefore deal with/encode remarkably different aspects of the world in order to use their language accurately (Sapir, 1921; Slobin, 1996). People communicate with each other using a variety of languages which differ from one another in numerous ways (for instance from clear differences in vocabulary and pronunciation to subtler differences in grammar) (Boroditsky at al., 2003). In languages with grammatical gender systems all nouns whether referring to animate or inanimate referents are assigned a gender. Some of these languages

demonstrate a less clear assignment based on semantics. For example, Zande, a language spoken largely in the Democratic Republic of the Congo, usually assigns nouns to four genders: masculine, feminine, animal, and neuter. There are, however, about eighty exceptions, including such concepts as metal and heavenly objects and edible plants, which fall into the animal gender (Corbett, 1991). Additionally, an Australian Aboriginal language, called Dyirbal, has four genders, indicated by 'class I, II, III, and IV'. Dixon (1972) explains that male humans and non-human animates belong to class I; female humans, water, fire, and fighting to class II; non-flesh food to class III and everything else to class IV. These rules are semantic but non-obvious (cited in Foundalis, 2002: 305). It has been said, however, that since Dixon published the Dyirbal grammar in (1972), the language has grown steadily closer to extinction as younger community members fail to learn it.

Additionally, the Tamil language, a member of the Dravidian family in south India, divides nouns into *rational* (referring to people and gods) and *non-rational* (referring to animals, and other entities) and further subdivides rational gender into *masculine* and *feminine* (Corbett, 1991: 8–10). Tamil can be said to employ a "natural gender system", which means that one can predict the gender of a noun by being given its semantics and vice versa. With regard to Indo-European languages, Foundalis (2002) believes that a smaller dependence on semantics can be identified. Nouns referring to humans (assigned to masculine or feminine gender according to sex) are in a minority. The exceptions (non-sexed objects assigned to either of those two genders) are the majority; semantic association is therefore a rather worthless predictor for the gender of a noun. In some cases, however, the word form may predict the gender of the noun, e.g. in Arabic feminine words are formed from the masculine nouns by suffixing 'ah', so they tend to be longer than the masculine nouns which sometimes makes it easier to identify gender.

2.2.3 Arabic Grammatical Gender System

Arabic is one of the Semitic languages and the native language of majority groups in countries ranging from Mauritania to Oman and from Iraq to Sudan. It is also widely studied in the non-Arabic-speaking Muslim world. Knowing the gender of a noun in gender-marking languages is necessary for correct sentence construction in both spoken and written forms. Arabic is among those languages with formal gender systems in which all nouns must be either grammatically masculine or feminine with a few exceptions. The general rule is to suffix the *'ah'* known as 'Ta Marbutah' to masculine

noun or adjective forms to derive the feminine ones. For example, the word *student*, is 'Talib' (Male), 'Talibah' (Female); likewise, the adjective *new*, 'Jadeed' (Male), 'Jadeedah' (Female). Specifically in nouns referring to humans, there is almost always a clear relationship between the gender of the noun and the sex of the referent which can be recognised by using different words (e.g. 'Rajol/Emrah' *man/woman*) as well as through the use of a derivational and inflectional affix that turns masculine nouns into feminine (e.g. 'Momathel/Momathelah' *actor-actress*; 'Tefel/Teflah' *male child-female child*). Also, as mentioned above, one can identify some formal differences between masculine and feminine nouns where feminine nouns are longer than the masculine nouns as they are derived from them. This has led some researchers to assume that masculine nouns are easier to acquire than feminine nouns. Gass and Selinker (2001: 160) offered a good explanation of noun marking:

> If we consider words denoting professions, avocations, or societal roles, we see that male terms are the basic ones (e.g., *actor, poet, host, hero*), whereas the female counterparts have suffixes added on to the male term (*actress, poetess, hostess, heroine*). The male term is taken to be the basic one (unmarked) and the female term is the marked derivative.

According to the Markedness Differential Hypothesis, structures that are simple or common in language are assumed to be unmarked, whereas those which are complex and less common are assumed to be marked (Archibald, 1998). If this claim is true, then female nouns in Arabic can be classed as difficult to learn and perhaps harder to process. All other nouns (such as animals, substances, artefacts, abstract entities, nouns describing actions and events) are marked for gender too. Some exceptions exist, however, where the suffix '*ah*' is used in certain ancient Arabic male proper names (such as, Hamzah, Talhah, Moawiyah) (Jiyad, 2006: 8). Additionally, in Arabic there is no gender-neutral pronoun (i.e. there is no equivalent of the English *it*) and everything must be a *he* or a *she*. It is not possible to construct a gender-neutral sentence as it is in English. Like many languages, the assignment of grammatical gender in Arabic is semantically arbitrary. A good example of its arbitrariness can be exemplified by the fact that we can have two words that refer to the same entity but are assigned different genders. For example, the word 'window' can be masculine as in the word 'Shobbak', or feminine as in the word 'Nafethah'.

2.2.4 The Acquisition of Grammatical Gender

It seems to be spontaneous in native speakers to acquire grammatical gender from an early age, e.g. children master the grammatical gender of their first language relatively early (see Slobin, 1985). L2 learners of grammatically gendered languages may, however, face difficulties in mastering grammatical gender. This might affect meaning for adult speakers and these effects must come about as a result of language-learning mechanisms. Alternatively, such effects may take place during language development. One possible hypothesis is the *similarity and gender hypothesis,* suggested by Vigliocco et al. (2005), in which these effects arise as a consequence of a very common learning mechanism used by children to acquire aspects of meaning from linguistic input (Fisher and Gleitman, 2002). The central idea is that words that share the same syntactic and morphophonological properties are likely to have similar meanings. Nouns with the same gender are used in similar linguistic contexts, such as the Arabic words *Kateb* 'writer' and *Ketab* 'book' which differ from other contexts where nouns of a different gender are used. The differences in linguistic contexts can be seen at both the syntactic and morphophonological level. The former refers to words with a similar gender which need gender agreement with determiners, pronouns and adjectives in sentences. The latter refers to the relationship between languages in terms of the syntactic specification of gender and how it is recognised in phonological and morphological forms for determiners, adjectives, pronouns and inflectional affixes of nouns. This argument has been proposed in the literature for different syntax-to-meaning mappings (Vigliocco et al., 2005).

According to the *similarity and gender hypothesis*, gender effects are not based on creating a relationship between the grammatical gender of nouns and sex of the referents. These effects may be found for languages with two genders (such as Arabic) in addition to other languages with three or more genders. That is because aspects of similarity in linguistic contexts are found across languages in spite of whether the genders of nouns are classified into the two sexes. This occurs as long as the languages are morphologically rich and therefore provides a sufficient number of gender-marked sentence contexts. In one language, gender effects should be present for all words irrespective of the type of referent, as they are based on general aspects of similarity rather than merely on whether the sex of referents is associated with gender of words.

Conversely, effects of grammatical gender depend on establishing associations between gender of nouns and sex of the referent, which is called the *sex and gender hypothesis* (Vigliocco et al., 2005). There is a relationship between gender of nouns and

the sex of their referents for humans. Children who learn a language with a grammatical gender system will notice this relationship between gender and sex (male or female features) for nouns referring to humans. Furthermore, they expand this rule to include other nouns which have no direct relationship but still refer to entities with a biological sex, such as animals. This view leads to similarity effects: that is words of the same gender being more similar among themselves than words of different gender by virtue of having male/female-like semantic properties. Another means of explanation is a relationship being established between the gender of nouns referring to humans and the sex of referents because of the co-occurrence of linguistic (gender of nouns) and conceptual (sex) features. If this relationship is recognised for words referring to humans, then it could be generalised to other nouns that have both linguistic and conceptual features. *The sex and gender hypothesis* supposes differential gender effects across languages and these effects are stronger for languages with the most association between the sex of the referents and the gender of nouns referring to humans, because the higher degree of clear association significantly simplifies the learning task (e.g. in Romance languages, there are only two genders and few exceptions to the consistent mapping between the sex of the referents and the gender of nouns referring to humans). The effects will, however, be weaker in languages with multiple genders or in those where nouns referring to humans fall into more than two classes. In such cases, it is challenging for the language-learner to establish a relationship between the sex and gender of human referents (e.g. the German word for *girl* is 'das Mädchen' which is grammatically neuter and not feminine as one might expect). Within a language, gender effects should only be found for animate entities and not for other entities that lack relevant conceptual features (those related to sex) (Vigliocco et al., 2005).

Nevertheless, it is essential to note that a less constrained version of the sex and gender hypothesis is possible too. The strong relationship between the gender of nouns and the sex of human referents could be more than enhancing the male/female-like conceptual properties of other sexuated[8] referents. This leads to the assignment of male/female-like conceptual properties, including entities in which sex is not a relevant conceptual dimension, merely by virtue of masculine or feminine gender marking on the related nouns (Vigliocco et al., ibid). Additionally, in the less constrained version of the *sex and gender hypothesis*, gender effects within a language should be extended to all words. In light of both hypotheses, i.e. *the similarity and gender* and *the sex and*

[8] The term 'sexuated' was used by Vigliocco et al. (2005) to refer to entities which have biological sex such as animals.

gender, the effects of gender are expected in verbal tasks, where lexical and conceptual information are retrieved.

Conversely, there is some evidence that children assume a relationship between the grammatical gender of nouns and the gender of their referents. The story of a three-year-old Italian girl cited by Chini (1995) helps prove such an assumption. The Italian little girl refused to accept that the *dress* (*vestito* in Italian) is masculine because it belonged to her doll (*bambola*) although it is grammatically masculine (Bassetti, 2007). A similar story happened with the two children of the researcher, aged six and seven who both categorised the *car* as masculine although it is grammatically feminine in Arabic i.e. *sayarah*. Their categorisation might be a result of cultural issues; women are not allowed to drive in our country (Saudi Arabia). They refused to accept the idea that a *car* is feminine. In fact, the association between grammatical gender and referents' gender could be explained to older children by explicit grammar teaching. Adults might, however, seek some logical explanations of the gender assignment of their native language as in the case of Arabic speakers mentioned in Clarke et al. (1981). Those speakers explained that *beard*, a typical male attribute, is grammatically feminine in Arabic because it is soft and pliable. The conflict between grammatically feminine nouns which have masculine connotation is justified by saying that beards have some feminine characteristics. Clearly there are many ways grammatical gender can infiltrate people's perception of entities.

2.2.5 How Grammatical Gender Affects Cognition

There has been relatively little research on the effects of grammatical gender on thought when compared to other cognitive domains such as colour, time and space. The available literature, however, suffices to show how an arbitrary property of some languages - such as grammatical gender - affects the conceptual representation of speakers. Some researchers maintain that children's categorisation of entities in the real world might be affected by the categorisations reflected in their languages. For example, Bowerman (1985: 1285) commented that "children are prepared from the beginning to accept linguistic guidance as to which distinctions – from among the set of distinctions that are salient to them – they should rely on in organizing particular domains of meaning". There seem to be different ways for the effects of grammatical gender to influence thought. Some researchers (e.g. Boroditsky and colleagues, 2003) argue that speakers of gendered languages begin to assign male and female properties to objects that do not have a sex as a result of acquiring the gender systems of their languages

which differentiate object nouns into feminine and masculine. The gender system then leads people to focus on some property of the noun's referent. For example, if the word for *sun* is feminine in your language, you might focus on its warming and nourishing qualities. If, on the other hand, the word for *sun* is masculine, you may try to conceive it in terms of what are perceived as stereotypically masculine properties like power and threat. Sera et al. (1994), however, stated that speakers might store the grammatical gender of nouns as an extra feature of their conceptual representation of the object, especially in languages with two gender classes. Both views suggest the profound effect of grammatical gender on thought as it can change other universal conceptual representations of objects. Both the studies by Boroditsky et al. (2003) and Sera et al. (1994) suggest that grammatical gender plays a role in affecting the mental representation of objects. They further asserted that when two concepts or objects share labels of the same gender, this increases their semantic similarity.

Clarke et al. (1981) and Konishi (1993) show that speakers of languages with masculine and feminine genders were affected by this grammatical category when asked to rate the similarity between pairs of words in relation to masculine or feminine properties. According to others, however, (e.g. Gennari et al., 2002) these effects can only influence speaker performance in tasks where the use of this knowledge may be strategic or mandatory in order to accomplish the task. We cannot ignore the possibility that when speakers of grammatically gendered languages are asked to assess artificial concepts/objects in terms of their gender classifications, they might feel prompted to rely on the linguistic markings of these items, particularly when there is no better way of completing the task. Clarke et al. (1981) argued, however, that participants in their task did not use grammatical gender as a strategy because they did not rate all grammatically masculine words as either a hundred per cent masculine or grammatically feminine words as one hundred per cent feminine. They considered this evidence of the effects of grammatical gender on speakers' judgements.

Konishi (1993) tried to avoid an explicit reference to gender when employing a list of high-frequency words which were grammatically feminine in Spanish and grammatically masculine in German, or vice versa. In that study Spanish and German participants were asked to rate words in their language for potency - a characteristic determined to correlate with masculinity. The results showed an effect of gender on the participants' ratings, which were consistent with the grammatical categorisations of their respective languages. Konishi (1993) considered these findings an indication of grammatical gender effects on speakers' perceptions. If Konishi's (1993) interpretation

is true, then all objects that share the same grammatical categories (e.g. masculine) should be perceived as more similar when compared to objects of another category (e.g. feminine), as they would share those feminine or masculine features. Therefore, similarity effects should also be seen on similarity ratings between concepts of the same semantic groups whose label carries the same gender. One serious weakness of these studies is that the authors used words as stimuli and asked participants to explicitly rate words on masculinity/femininity scales. In gendered languages, there are often some formal qualities of nouns that show the type of gender. Using purely linguistic stimuli to study the effects of language on cognition might be argued to only measure participant knowledge of grammatical gender of their languages, rather than its effects on their cognition. The findings of these studies would have been more convincing if the authors had used a variety of cognitive - as well as non-cognitive tasks - to study the same issue.

Another view, however, suggests that the effects of grammatical gender on thought are caused by strategic access to this grammatical property in the language of the speaker to provide them with an additional feature to accomplish cognitive tasks (Vigliocco et al., 2002; Bowers et al., 1999). Based on this view, semantic and conceptual representations would not differ between languages and might be mostly independent of linguistic representations. Nevertheless, some languages were found to have a powerful role in influencing thought in tasks where no language was included or required, such as picture categorisation tasks.

A final view considers any effects of grammatical gender as a result of the implicit use of language to perform cognitive tasks. This should not be taken as evidence of language effects on thought as it barely shows any effect on linguistic encoding (Munnich and Landau, 2003). According to this view, tasks that involve obligatory language processing such as naming tasks should reveal language effects on thought. In order to better understand this effect, a variety of tasks that do not include linguistic processing should be carried out, such as picture categorisation and picture similarity rating tasks. Finding evidence of the effects of grammatical gender on cognition can offer us a better understanding of the role that our languages play on our thinking.

2.2.6 Methodological Approaches for Investigating Grammatical Gender

A substantial body of empirical research has been conducted to investigate the effects of grammatical gender on cognition across languages with mixed findings. There seem to be three types of approach to studying this issue.

One set of studies used a *semantic differential task* which provides a quantitative measure of the connotative meaning of concepts (Osgood, Suci, and Tannenbaum, 1957). It asks participants to judge the meaning for a given concept on a series of scales between bipolar adjective pairs. For example, participants are presented with stimuli that differ in terms of grammatical gender and then required to produce or rate masculine and feminine characteristics. Those ratings provide a measure of three factors: evaluation, potency and activity. One of the earliest studies on grammatical gender using this method was carried out by Ervin (1962). In that study, Italian speakers were taught nonsense words that possessed masculine/feminine Italian affixes, then asked to rate these nonsense words which differed in their ending vowel. The findings showed that the Italian speakers rated the nonsense words with masculine affixes as more like men and the words with feminine affixes as more like women. They attributed masculine connotations such as 'big' and 'strong' to stimuli ending with an Italian masculine marker (*-o*) and feminine connotations such as 'weak' and 'good' to stimuli ending with an Italian feminine marker (*-a*). Some studies have, however, failed to suggest such an effect, as revealed by Hofstatter (1963, cited in Zubin and Köpcke, 1984) who studied the perception of gender among German and Italian speakers using a semantic differential attribute scale. Studies which investigated grammatical gender effects yielded inconsistent results using this method.

Supporting evidence for the effect of grammatical gender on cognition was reported for Arabic speakers by Clarke, Losoff, McCracken and Still (1981) using a simple masculinity/femininity scale. Clarke and colleagues extended previous research which found no effect of grammatical gender on Hebrew speakers' rating of objects (Guiora and Sagi, 1978; Guiora and Acton, 1979). In their study, participants were asked to evaluate physical objects along a masculine/feminine scale. The results from comparing Arabic and English groups reveal that the gender of nouns in Arabic affected the responses of Arabic participants. For example, nouns with a masculine gender in Arabic such as *necklace* 'oqed' and *perfume* 'oter' received a higher masculine rating from the Arabic group than from the English group. It is possible, however, that Arabic speakers might consciously use grammatical gender in their language to do that task, leaving these results somewhat open to question.

Furthermore, a study which examined gender assignment to animals and objects by English and German adults and children showed that the assignment of gender correlated greatly with German grammatical gender and pronoun use in English (Mills, 1986). This is despite some of the responses not fitting precisely into the German neuter gender. In fact, it was not known whether the participants accessed the gender system, or just perceived gender-related attributes in the objects. For this reason, a semantic differential scale was used and participants rated animals and artefacts on fifteen semantic differential scales (such as 'strong'-'weak', 'large'-'small', 'tense'-'relaxed') which were shown to correlate with masculine and feminine attributes. Mills' conclusion stated that grammatical gender was the main influence on the choice of sex.

Along the same lines, Konishi (1993) asked German and Spanish speakers to rate words on a semantic differential scale and revealed that grammatically masculine words were rated higher on semantic dimensions that have masculine connotations - such as 'power' - but not evaluation or activity. Essentially, speakers of German and Spanish differed in their ratings for words that had a contrasting gender in the two languages. Based on these findings, grammatical genders and conceptual representations of words are closely associated. Konishi (1993) commented that grammatical gender affects meaning because the participants' perception of the characteristics of the assumed inanimate entities associates with the grammatical gender of the nouns which represent these entities. These results strongly support the linguistic relativity hypothesis, meaning that language may shape the way speakers perceive the world.

Some studies used various scales related to masculinity and femininity, rather than a simple masculinity-femininity scale. Tong, Chiu, & Fu (2001) studied Chinese speakers to investigate the relationship between the grammatical and conceptual gender. The Chinese language does not inflect or vary pronouns for gender, so the marking of gender is less obvious there. It was found that pseudo-words with the semantic radical[9] for *woman* were rated lower on potency and activity, compared to those with the radical human being. Semantic differential scales were also used to study abstract concepts; for example, German speakers rated affect nouns such as *sadness* 'Traurigkeit' and *courage* 'mut' (Zubin & Kopcke, 1984), grammatically masculine nouns rated higher on extroversion and grammatically feminine affect nouns were higher on introversion (a feminine characteristic). These results were marred, however, by the lack of a control group with a different language background. Hofstatter (1963) used a twenty four

[9] In modern Chinese about 80% to 90% of the characters are composed of two components: the radical and the stem.

semantic differential scale which focused on comparing German and Italian monolinguals' concepts of *moon* and *sun*. The study showed that although the words *moon* and *sun* have opposite genders in German (*mond*, *sonne*, respectively) and Italian (*luna*, *sole*, respectively), both groups of participants chose similar descriptions and no effect of grammatical gender was reported. From these studies, we can conclude that grammatical gender has little effect on the perception of objects and that perception was determined mainly by the attributes of the objects themselves.

The previously mentioned studies used words as targets in semantic differential tasks, but some of the studies also used pictures. On various two-point scales, Flaherty (1999) found the effects of grammatical gender on French and Spanish adult ratings of pictures of objects and animals, but no such effects found in English and Japanese. Replicating the study with English and Spanish children, Flaherty (2001) reported grammatical gender effects on children above the age of ten; unlike other types of tasks (e.g. name attribution) where grammatical gender effects were found in children at the age of eight. Flaherty, however, posits the possibility that language affects only on-line cognitive processing not cognitive representations.

Taken together, the studies mentioned above have used semantic differential tasks to investigate the effect of grammatical gender on cognition. Although some failed to show strong effects, the majority provided evidence of grammatical gender effects on speaker performance across languages.

A second set of studies used *picture categorisation tasks* which asked participants to categorise pictures in order to reduce the possibility that speakers overtly referred to language in their categorisations. Sera et al. (1994) compared the classification of the pictured objects into masculine and feminine between speakers of Spanish and English. The results indicated that Spanish speakers categorised pictures of objects as masculine or feminine according to the Spanish grammatical gender system, whereas English speakers assigned gender arbitrarily. To reduce obligatory grammatical gender access, the authors subsequently used a voice-attribution task with Spanish and English adults, which - according to Sera and colleagues - should have led the attention of Spanish speakers away from explicitly referring to grammatical gender. Participants were asked to assign a feminine or masculine voice to a proposed cartoon animation of the pictured objects. The findings showed that Spanish speakers continued to classify pictures according to Spanish gender where eighty five per cent of their choices reflected the grammatical gender of the object labels, compared to fifty three per cent of the choices of the English speakers. The third experiment in the same study looked at the

development of gender categories by presenting the same categorisation task to English and Spanish adults and children from different age groups. Results showed that older Spanish children did the task in a way that was similar to Spanish adults, but that the categorisation patterns of the English speakers did not differ between kindergarten, second grade, fourth grade and adult. The authors point out that learning a grammatically gendered language restructures basic mental representations; hence if the objects share grammatical gender, they are more likely to be represented as more alike, albeit they may be semantically discrete. Sera and colleagues concluded that these findings, together with previous similar findings, strengthen the claim of a meaningful link between the grammatical and conceptual organisations of gender. Furthermore, they offer evidence that the relationship between grammatical and conceptual classifications of gender is not arbitrary, but instead involves a coupling of thought to language and development. In a subsequent study, Sera et al. (2002) replicated the voice-attribution task to investigate the correlation between grammatical and conceptual gender. Adult and children monolingual speakers of English, German, Spanish, and French participated in a series of tasks to show the grammatical gender effects on categorisation of inanimate objects. The results revealed that gender effects on categorisation were found for the French and Spanish participants, but not for the German speakers. The authors reported that the distinction between masculine and feminine may be unclear for languages with a neutral gender system such as German.

In a third set of studies, some researchers have used tasks that lessen the possibility that participants could use the grammatical gender of their languages in a conscious manner. Martinez and Shatz (1996) used *a free-classification task* in which they asked children to sort pictures of people and objects into groups. The effects of Spanish grammatical gender were found in three to four year old children where six out of eighteen sorted items were identified according to Spanish grammatical gender.

Nonetheless, these studies did not analyse how task instructions and presentation modality restrain gender effects. Using the same task (free categorisation) with Spanish and English speaking children, Vigliocco, Vinson, Paganelli and Dworzynski (2005) found that the semantic effects of grammatical gender are limited to animate categories that have sex as a semantically related property. Italian speaking children were found to judge animal names of the same gender as more similar in meaning than English speaking children, but no difference was found in judgements for artefact names. Additionally they made more same-gender semantic errors - in comparison to English speakers - in semantic substitution errors tasks (e.g. saying *tiger* when *leopard* is

intended), proposing that semantic similarity increases between words of the same gender. This effect was restricted to animal names and was "mediated by an association between gender of nouns and male or female-like aspects of meaning[10]" rather than directly affecting representations (Vigliocco et al., 2005: 506).

Using a *similarity rating task* to study the effect of the Spanish grammatical gender system on Spanish speakers' performance, Degani (2007) pointed out that pairs of nouns that matched in grammatical gender (e.g. *camisa* (feminine) – *mesa* (feminine), shirt & table, respectively) did not elicit higher semantic similarity ratings when compared with unmatched pairs, furthermore those pairs were not processed more quickly or accurately in a primed naming task.

Other studies simply used linguistic tasks to study the possible effects of grammatical gender. For example, in a recent study by Vuksanovic et al. (2014), the effect of grammatical gender on mental representation was examined with musical instruments. Serbian participants were provided with a list of twenty two pseudo-words indicating different grammatical genders, along the information that they are names of musical instruments. Participants were asked to select adjectives that best describe each of the given instruments. The results showed the strong effect of Serbian grammatical gender in shaping participants' notions about objects. The methodology used in this study, however, raises some questions about the nature of this effect and how pervasive it is in influencing cognitive representations. More specifically, participants were explicitly asked to select either feminine or masculine adjectives to describe words referring to musical instruments. In such a case, participants are more likely to make their selection in accordance with the grammatical gender system of their language in order to construct correct sentences. This study also lacks a control group which might give more convincing evidence of the effects of this linguistic property.

Most of these studies were conducted on monolingual communities and the majority showed the effect of language on cognition. The question is then; what happens with all those who know more than one language? Most of the world's population are bilingual and many researchers now consider bilingual to be the norm and monolingual the exception (Harris & McGhee Nelson, 1992). Cook (2002) argues that there are only "a handful of isolated pockets of 'pure' monolinguals and they are now hard to find even in the mountains of Papua New Guinea". Widespread

[10] In Italian masculine nouns - although referring to animals of both sexes - may be considered as more male-like by Italian speakers because a feminine version of these nouns is possible. Feminine nouns (also referring to both sexes), instead may be considered as more neutral because they cannot be changed into masculine.

immigration and learning of foreign languages in schools has increased globalisation in modern societies, so it seems that monolinguals are "a species likely to become extinct in no time" (De Groot, 2011: 342). Furthermore, including bilinguals in research on linguistic relativity allows us to answer more questions about the nature of language effect as well as bout the extent to which our cognition can be restructured by learning another language. To this end, a substantial body of research has been undertaken on the effects of grammatical gender on the cognition of bilinguals using a number of methods.

2.2.7 Grammatical Gender and Bilingual Cognition

One of the reasons for studying bilingual cognition in relation to gender is to gain a better insight into the nature of bilingual cognition itself. If language has an effect on non-linguistic cognition, then bilingual speakers might be expected to perform similarly in both their languages and not to differ from monolingual speakers of their first language. If language affects only the semantic representations of that same language, then it would be expected that there will be a significant difference in performance between the bilingual speakers' languages. In grammatically gendered languages (e.g. Arabic), nouns referring to humans almost constantly correspond to their gender, but gender assignments for inanimate nouns are mainly arbitrary. It is therefore possible that learning a second language with no gender system (e.g. English) changes the arbitrary nature of gender assignments in the bilingual speakers' first language and may show up cognitive differences between monolingual and bilingual people. This idea is referred to as the Multi-Competence theory by Cook (1991, 2002, 2003) who argues that the L2 user is a unique individual who knows more than one language and that an L2 learner's mind is not the same in nature as the mind of a person who knows only one language.

Research on the effects of grammatical gender on bilingual cognition started with the work of Ervin (1962) who studied the role of Italian grammatical gender and classification in Italian monolinguals and Italian-English bilinguals. Participants were taught nonsense words that possessed masculine/feminine Italian affixes and were asked to rate these words on four scales (they were asked to rate the form of the words). Italian-dominant bilinguals rated the nonsense words with masculine affixes as more like men and the words with feminine affixes as more like women. English-dominant bilinguals - who had acquired English at an early age - were not affected, however, by the Italian gender system. This study demonstrates that learning a second language

earlier in life - even if it is a grammatically genderless one - can remove the effects of the grammatical gender of the native language.

The extent to which the grammatical gender of nouns influences people's perception of the cognitive category of biological gender or sex was examined by Boroditsky and Schmidt (2000). Their findings showed that English speakers' intuition about the gender of some nouns (animals) connect with the gender assigned to those nouns in languages such as German and Spanish. They taught groups of Spanish, German and English speakers proper names for 24 objects (e.g. an apple may have been called Patrick), then tested their memory for these object-name pairs. The results showed that the Spanish and German speakers' memory for object-name pairs were better for pairs when the object proper name was consistent with the grammatical gender of the object name (in their native language) than when the two genders were inconsistent. English speakers, also, performed the task in a way that was similar to the two groups even though their language does not have a grammatical gender system. This led the authors to argue that inanimate objects do appear to have conceptual gender and this gender is consistent with the grammatical gender assigned by some languages. Essentially they found that people's ideas about the assumed biological gender (sex) of objects are influenced to a large extent by the grammatical gender of those objects in their native language. Foundalis (2002) argued, however, that the interpretation supplied by Boroditsky and Schmidt (2000) is unwarranted and that they combined the concepts of both biological gender (sex) and formal gender which is used by most Indo-European languages - unlike the 'natural gender' in English.

Other studies used non-linguistic stimuli to study bilinguals with a grammatical gender L1 and a natural gender L2. Boroditsky et al. (2003) identified some limitations in the previous research which questions its findings. First they argue that speakers of different languages were tested only in their native languages and any differences in these comparisons only reveal the effect of a language on thinking for that particular language. The consequences are that such studies cannot show whether experience with a language affects language-independent thought. Second, comparing studies conducted in different languages causes a deeper problem, as there is no way to be certain that the stimuli and instructions are truly the same in both languages. The final limitation reported was that all tasks described in the previous research asked participants to provide some subjective judgement which requires participants to decide on a strategy to complete the task. This means that they might consciously decide to follow the grammatical gender divisions of their languages. Boroditsky et al. (2002) tested

Spanish-English and German-English bilinguals' memory for object-name pairs in their L2. Participants were taught proper names for twenty four objects; for example, an *apple* might be called *Patrick*. Half the objects were consistent with the grammatical gender of the object's name in the participants' native language and the other half were inconsistent. Both Spanish and German speakers remembered the object-name pairs where the gender of the proper name was consistent with the grammatical gender of the object name (in their language) better than when the genders were inconsistent. As this study was conducted in English, grammatical gender was argued to shape the participants' underlying conceptual representations, rather than affecting only online linguistic processing (Boroditsky et al., 2003). Although Boroditsky and colleagues tried to avoid the limitations that they raised in relation to previous studies, there was an overt reference to natural gender in their task through the use of proper names which could have prompted participants to use their knowledge about grammatical gender as an extra memory cue. It is therefore necessary for researchers to keep the use of language to a minimum in order to show pure cognitive effects, if there are any.

In a similarity rating task, Phillips and Boroditsky (2003) asked Spanish and German bilingual speakers to rate the similarity of unlabelled pictures depicting objects and people. The objects were chosen on the basis that they have an opposite grammatical gender in Spanish and German[11] and all the objects were then compared to pictures of a number of biological males and females. The results showed that differences in the participants' similarity judgements correlated with linguistic experience. To explain further, Spanish-German bilinguals with more Spanish experience rated their similarity like native Spanish speakers; whereas bilinguals with more German experience rated their similarity like native German speakers. In addition, the same results were obtained when participants made their similarity ratings while performing a verbal interference task. The interference task was assumed to affect any on-line processing that might occur during the similarity ratings, meaning that any observed differences between the Spanish and German groups should only reflect differences in their mental representations. Phillips and Boroditsky (2003) emphasised that their findings revealed substantial evidence of a non-linguistic effect acquiring a grammatical gender system on mental representations.

Along the same lines, Boroditsky, Schmidt, and Philips (2003) conducted a study on grammatical gender systems in Spanish and German and found that grammatical

[11] The German language has a three gender system, the authors only mention choosing German masculine and feminine noun classes.

gender may change non-linguistic representations and that objects have a conceptual gender. German and Spanish participants were presented with a series of object pictures and were asked to describe them using three adjectives in L2 English. The results revealed that participants produced feminine adjectives for referents whose nouns were feminine in their first languages and masculine adjectives for referents whose nouns were masculine (e.g. the word for *bridge* is feminine in German and masculine in Spanish whereas the word for *key* is feminine in Spanish and masculine in German). German participants described *bridges* as beautiful, peaceful and elegant; whereas Spanish participants described them as strong, big and towering. As for the word *keys*, German speakers said they are 'hard', 'heavy' and 'jagged'; whereas Spanish speakers described them as 'little', 'tiny' and 'lovely'. The authors commented that since the responses were given in English and were affected by the grammatical gender of the participants' first languages, these findings indicated that conceptual information is shaped by gender with some semantic features. The authors further argued that as all stimuli had an opposite grammatical gender in German and Spanish, this can show that these are language effects rather than effects of referent characteristics. This interpretation does not, however, consider the fact that Spanish and German participants might still access the grammatical gender information from their native languages even though they performed the task in English. Effects of the first language on the second have been shown by researchers such as Malt and Sloman (2003) who mentioned that native language naming patterns affect speakers' performance in the second language even if those speakers had reached an advanced level in their second language.

The effects of grammatical gender systems of two languages - Italian and German - were investigated by Bassetti (2007) who studied monolingual and bilingual Italian children. A voice-attribution task was used with pictures of artefacts that had an opposite grammatical gender in German and Italian. The results showed that Italian-German bilinguals were not affected by Italian grammatical gender compared to Italian monolinguals (whose voice assignments were consistent with Italian grammatical gender). These results indicated that "when the two languages of a bilingual represent a specific aspect of reality differently, the bilingual may develop different concepts from a monolingual" Bassetti (2007: 251). It appears that knowledge of two grammatically gendered languages reduces the grammatical gender effect of the first language. Such studies on bilinguals with two grammatically gendered languages demonstrate that the effects of grammatical gender on cognition vary between monolinguals and bilinguals whose two languages assign opposite genders to the same object.

Some researchers have a contrasting view of the effect of Italian grammatical gender on the conceptual representations of Italian speakers. Kousta et al. (2008: 855) argued that "Italian grammatical gender cannot logically have an effect on the nonlinguistic, conceptual representations of bilingual speakers". The authors used a continuous naming task where pictures were presented at a fast rate and then analysed semantic substitution errors. Their results revealed that Italian monolinguals showed a considerably higher proportion of gender-preservation errors than English monolinguals. Italian-English bilinguals also performed like Italian monolinguals when the task was in Italian and like English monolinguals when the task was in English. These findings therefore demonstrate the importance of the language used to conduct the task as it may greatly affect the results even if the same task is used. The authors assert that grammatical gender increases semantic similarity between words that share the same gender when compared to words that do not.

Another type of research looked at the effects of grammatical gender on bilinguals who acquire a gendered language at an older age. Kurinski and Sera (2011) asked English-speaking learners of Spanish to perform a voice-attribution task where they had to attribute gender to inanimate objects. The findings indicated that Spanish grammatical gender affected English-Spanish bilinguals' responses to the voice-attribution task; the effects were not, however, observed for all kinds of objects and did not increase with the learners' proficiency of Spanish. The effects of grammatical gender seem to be limited, in that adult learners of Spanish reached a level beyond which changes in categorisations do not occur (Kurinski and Sera, 2011).

In a recent study, Nicoladis and Foursha-Stevenson (2012) found that French grammatical gender affects children's classification of objects as boys or girls in English. The study included French-English bilingual children in two different age groups (three to five and eight to ten) who were compared to French monolingual children to control for possible cultural effects. In addition, English speaking adults were asked to do the same task. The results showed that bilingual children aged eight to ten were affected more by French grammatical gender, while the English controls and preschool children were not. The effects of language were small, however, when compared to the cultural effect.

To sum up, although there may be a substantial body of research that investigates the effect of grammatical gender on cognition, the available evidence is still inconsistent and open to different interpretations. The majority of researchers mention that grammatical gender affects non-linguistic tasks, yet there are divergent views about the

nature of these effects. For example, Sera et al. (1994) and Boroditsky et al. (2003) argue for the effect to be at a deep level of cognitive representation even when language is not involved. Others, such as Flaherty (2001), argue that grammatical gender affects cognitive processing rather than cognitive representations. A final view by Vigliocco et al. (2005) considers the effects of grammatical gender to be mediated by implicit or explicit linguistic processing.

2.2.8 Factors Affecting Bilingual Thinking

It is worth noting that results from studies on the relationship between grammatical gender and categorisation have provided relatively little information about the participants included in those studies. Much other linguistic and sociocultural information needs to be considered in the study of language and bilingual cognition. A good description of those variables has been provided by Athanasopoulos (2011) who nicely describes the position of each variable on the continuum, illustrated in figure 2.1, below.

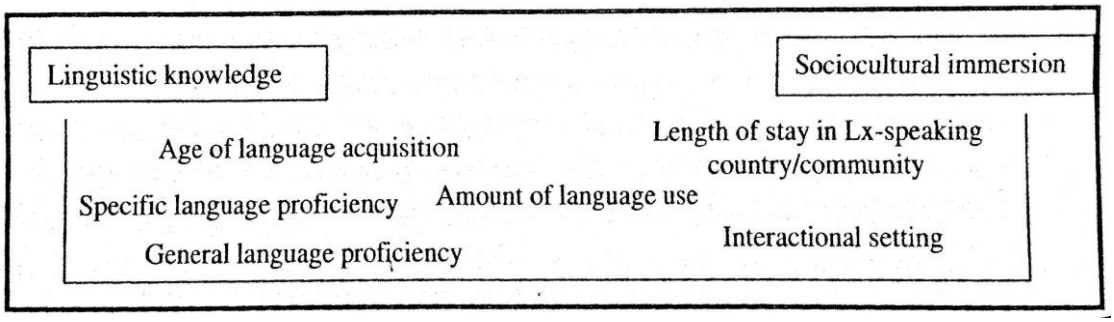


Figure 2.1 Continuum of linguistic and sociocultural variables that may affect bilingual cognition, adopted from Athanasopoulos (2011: 37)

A closer look at the figure above shows that age of language acquisition is placed towards the linguistic end of the continuum. The reason for this, according to Athanasopoulos (2011), is that development and mastery of language could depend on maturational constraints, either due to the critical period for language acquisition, or because of the continuing decline of learning mechanisms with increasing maturation. Thus, the effect of this variable may not be directly noticeable, but rather "it may be a mediating variable in the relationship between language proficiency and degree of cognitive restructuring" (p: 37). One of the most controversial questions in bilingualism research is whether there is a best specific age for second language acquisition[12]. Some

[12] Lenneberg (1967) proposed that natural language acquisition through exposure can only happen during the critical period (age two - puberty). Before the age of two the brain has not developed enough, and

researchers (e.g. Gleitman and Newport, 1995) suggest a sensitive or optimal period rather than a critical one; e.g. optimal ages of around seven to eight years and ten to twelve years. The strongest evidence for the critical period hypothesis in second language learning is in the study of accent. Much of the research shows that people beyond the age of puberty do not acquire a native-like accent but - of course - exceptions do exist (Brown, 2000, but also see Abrahamsson and Hyltenstam, 2009, for discussion on age of onset and native-likeness in a second language). Johnson and Newport (1989) pointed out that seven to twelve years old is the critical period for successful second language acquisition. Others suggested the age of five years to be the point at which any additional language learning might be considered second language learning. Recently, Singleton (2014) summarised the different claims about the critical period hypotheses for L1 acquisition and L2 learning/acquisition and stressed the importance of age as a factor in L2 learning alongside other *age-related* factors that play a role in second language learning. Singleton (ibid.) favoured childhood as being the most favourable time to begin to be exposed to a second language at least in 'naturalistic' circumstances (p: 38), but the cut-off point for second language acquisition is still controversial.

Two other variables are placed on the linguistic end of the continuum, namely specific language proficiency and general language proficiency. The former means knowledge of the specific linguistic property under investigation, which could be elicited through a variety of ways such as free narratives or controlled conversations in picture description tasks, or through written tasks such as grammaticality judgments, to name a few possibilities. The latter refers to the general level of proficiency in a range of language areas through the use of independent language tests or by simply asking participants to rate their proficiency on a questionnaire[13].

Moving to the other end of the continuum, we can see sociocultural variables, which include the length of stay in the country where the second language of the person is spoken as a native language and their interactional settings. Specifically the length of stay may lead the speaker to follow - consciously or unconsciously - the behaviour of the target-language community and interactional settings might promote

after puberty it is has developed too much, with the loss of "plasticity" and the completion of "lateralisation" of the language function.

[13] Although such introspective techniques can be used to determine bilingual proficiency, many researchers argue that it is more reliable to use independent language tests e.g. Nation Vocabulary Test, Oxford Placement Test, or even more specific language tests which test all language components (Athanasopoulos, 2011).

bilinguals to behave differently in different settings, a notion called ‘language mode’ by Grosjean (2001, 2010).

Speaking of the amount of language use, it sits in the middle of the continuum since it can be categorised by linguistic as well as sociocultural factors. The use of one of the bi/multi-lingual languages depends on interactional settings as well as the degree of immersion in a particular community. Therefore, this use will in turn increase proficiency in the most-used language, which will eventually provide the speaker with native-like competence in specific linguistic features (Athanasopoulos, 2011).

Nevertheless, it is worth noting that bilingualism is dynamic in nature so the aforementioned variables may be found to affect bilinguals’ performance in cognitive tasks. Athanasopoluos (ibid.) argues that if the effect of these variables is mediating or moderating rather than directing, it may be difficult to show in experimental contexts. Studies that investigate the relationship between language and cognition should therefore take linguistic and extra-linguistic variables into account. However, no studies in the domain of grammatical gender have considered all of these factors, meaning that accounting for these multitudinous factors was one of the aims of the current study. Specifically, this study aims to examine the potential effects of Arabic grammatical gender on the categorisation of objects by monolingual and bilingual Arabic speakers using categorisation and similarity rating tasks. There are two main hypotheses, (a) speakers of Arabic and English will differ significantly in their performance on these cognitive tasks; (b) bilingual speakers are expected to perform the tasks differently from monolinguals showing an effect of bilingualism on their cognition.

Chapter 3. Methodology

This chapter starts by considering the research questions. It then moves on to explain the design stage of the experiments used in this study (a voice-attribution task and a similarity rating task) by describing the stimuli and criteria followed in selecting them. General information about the population of participants and more detailed information will be provided in the following chapters. This chapter then continues with a report on a pilot study where a thorough description of the process of the design of the experiments is presented. We conclude with a section in which the methodological considerations are presented.

3.1 Research Questions

The literature surveyed in the previous chapter showed that further research was needed to answer such questions on how language can affect our thinking in general and specifically how grammatical gender systems can show evidence of such effects. Investigating bilingualism was a point of interest too. Consequently, the following overall research questions were proposed:

1. Does Arabic grammatical gender have an effect on the categorisation of objects?
2. Would learning another language change the cognitive performance of bilinguals and make them categorise objects differently from monolinguals? If the answer is yes, then to what extent?

The specific hypotheses are discussed in chapter 4 (section 4.1.3) and chapter 5 (section 5.1.3).

3.2 Stimuli used in the experiments

This study is aimed at investigating the possible effects of Arabic grammatical gender on the cognition of Arabic monolingual speakers, taking English monolingual speakers as a baseline for comparison. Furthermore, it investigates the performance of Arabic-English bilinguals to find out whether they differ from their monolingual counterparts and whether they change any cognitive aspect as a consequence of learning English. The investigation uses different cognitive experiments and this section describes the selection process for the stimuli used in the experiments. It

presents the sources from which the stimuli were chosen, along with the criteria followed in each selection.

3.2.1 Voice Attribution Task (VAT)

A voice-attribution task has been widely used in previous research (e.g. Sera et al., 1994; Sera et al., 2002; Bassetti, 2007; Ramos & Roberson, 2010) with the aim of investigating grammatical gender effects on participant voice attributions to pictures of different entities. In this study, stimuli for the VAT were taken from two sources; the first was the study by Snodgrass and Vanderwart (1980) in which they present a set of 260 black-and-white line drawings which have been standardised according to four variables of central relevance to memory and cognitive processing: name agreement, image agreement, familiarity and visual complexity (Snodgrass and Vanderwart, 1980: 182). The other source was the International Picture Naming Project (IPNP) database (Szekely et al., 2004-2005) which provides 520 common objects and 275 transitive and intransitive actions which can be used in cross-linguistic research. The picture stimuli and related reaction time norms are available for browsing and downloading in seven languages (English, German, Mexican, Italian, Bulgarian, Hungarian and Mandarin). The cross-language database includes information about the norming study together with available lexical information (e.g. frequency, age of acquisition) for the associated target names.

A number of criteria were considered when choosing the stimuli. All were everyday familiar objects in order to make it easier for the participants to identify them and they fall into five semantic groups, namely; 'body parts', 'clothes', 'vehicles', 'food' and 'household items'. It is worth mentioning that the group 'musical instruments' was not included in the current study, even though it has been used in previous research. The reason behind this was that the milieu of investigation is religiously dominated and music is not part of religious or cultural life in Saudi Arabia. This decision was taken to avoid any difficulty in identifying musical instruments. The stimuli only represent inanimate objects and do not include any pictures of animals. This is because in Arabic all nouns referring to animals have either masculine or feminine forms according to their biological gender, therefore attributing gender to pictures of animals may not be arbitrary, as it is for inanimate objects.

Furthermore, half of the chosen stimuli depict artificial objects (e.g. *house, door*) and the other half depicted naturally occurring objects (e.g. *sun, banana*). For each of the artificial and natural objects, half were masculine in gender and the other

half feminine. Another factor taken into consideration was that the selected stimuli did not include pictures of objects that have a gender connotation to their referents and share the same grammatical gender at the same time. An example is the noun *lipstick* 'Homrah' is grammatically feminine in Arabic and the item is generally used by females, so items such as this were not included (a full list is presented in section 4.1.3.2).

3.2.2 Similarity Rating Task (SRT)

The similarity rating task has been used in various previous studies (e.g. Boroditsky, Schmidt, and Phillips, 2003; Phillips & Boroditsky, 2003; Degani, 2007; Ramos & Roberson, 2010). In these studies participants were asked to match pictures of animals and objects with pictures of female or male humans. Results from the first two studies (Boroditsky and her colleagues) showed the effects of grammatical gender on the ratings of objects with Spanish and German speakers. The other studies did not, however, show the effects of gender on the participants' ratings of similarity (see sections 2.2.6 and 2.2.7).

For this study, sixty pairs of stimuli were used in the SRT experiment taken from five semantic groups, namely; 'body parts', 'clothes', 'vehicles', 'food' and 'household items'. As in the voice attribution experiment, the stimuli were taken from Snodgrass and Vanderwart (1980) and Szekely et al. (2004). The primary criterion for this task was the semantic connection of the stimuli. Thirty pairs were from the same semantic groups: ten were of the same gender and twenty of a different gender. The other thirty were from different semantic groups: ten were of the same gender and twenty of a different gender.

This experiment used a seven-point scale as that seems to provide a good balance; offering enough points of discrimination without maintaining various response options. The psychometric literature suggests that seven-point scales are better than those with five or eleven points and that although having more scale points is better, there is a weakening return after around eleven points (Nunnally, 1978). Five-point scales were avoided in this experiment for the following reasons; participants might feel forced to choose the 'next best' alternative if there are not enough response options, which may in turn introduce measurement error. For example, if they think a '1' is too high and '2' is too low, then they will be forced to select an option that is higher or lower than what they actually think because they cannot choose a '1.5'. Additionally, a study by Finstad (2010) tested some

participants in a five-point scale condition while others were tested using a seven-point scale, counting the number of times participants could not decide between two points. With a five-point condition, participants were allowed to 'interpolate' or choose between points such as a '4.5'. Results showed that in 2.5% of the 858 responses, participants chose responses between two points (1.6% - 3.9%). Conversely, in the seven-point condition, participants did not interpolate in any of the 840 ratings. Such findings were taken as convincing evidence to always use seven-point over five-point scales. Finstad (2010) concluded that seven-point Likert scales provide a more accurate measure of a participant's true evaluation.

3.3 Populations

All participants in the experiments were adult native speakers of Arabic and English, some were monolingual and others bilingual. The term monolingual is used throughout this research to describe those who are only proficient in one language with no or a low level in any other language. The term bilingual is used to describe a person who uses two or more languages to communicate[14]. It should be noted that the bilingual participants speak the same dialect as the Arabic monolingual baseline. All participants were individually recruited by the author at local universities in Saudi Arabia and the United Kingdom. Further information about them is provided in chapters 4 and 5.

3.4 Pilot study

A pilot study is a small experiment designed to test plans and methods for a research study in order to improve the quality and efficiency of the large main study (Waite, 2002). One of the reasons to undertake a pilot study is to reveal deficits in the design of a planned experiment or procedure so that these can then be improved or changed before time, effort, and resources are spent on large scale experiments. A pilot study is usually small and will provide only limited information in comparison with the main study; it can, however, provide vital information that helps researchers to assess the feasibility of their research (van Teijlingen & Hundley, 2001).

[14] In fact, the definition of bilingualism is problematic since individuals with varying bilingual characteristics may be classified as bilingual, ranging from a minimal proficiency in two languages, to an advanced level of proficiency which allows the speaker to function and appear as a native-like speaker of two languages (see Grosjean, 2010).

In the light of this, the voice-attribution experiment went through many different designs before the final design was completed. Piloting the experiment revealed some deficiencies with the design of the task and the choice of stimuli. The pilot study involved twelve participants: seven native speakers of Arabic (five women and two men) and five native speakers of English (three women and two men), all of whom were living in the UK. The task included forty four pictures; four control items which were used to ensure that participants fully understood the task; namely that they should assign voices according to the referents' natural genders (biological sex) and forty inanimate test items. The first design was based on Sera et al. (1994) and Ramos and Roberson's (2010) studies which presented the stimuli in a booklet with six items per page, a similar method to Kurinski and Sera (2011), which presented the pictures on individual hand-outs with eight pictures per page.

Therefore, the first design of the current task was presented in a booklet with six pictures per page (one control and five test items). Preliminary analysis showed that some participants seemed to be affected by the gender of the control item on each page and that they assigned the voices for the test items according to the gender of the control item. For example, one page included an *old lady* as a control and *iron*, *ball*, *flag*, *apple* and *table* as test items. Some participants (particularly the English group) assigned all the test items a feminine voice, using the gender of the control item as a clue. One solution for such a problem was to have one picture per page in order not to confuse the participants. That was to be avoided, however, because of the number of pages the task would comprise; having forty four items would mean forty four pages plus the introduction, instruction page, two consent forms, the language proficiency test and another two pages for the test of knowledge of the gender of the items. It was thought that having such an amount of pages would be both tedious and time consuming and might affect participants' performance. As a result, that design was substituted with a Power Point presentation with one picture per slide which was a quick and convenient way to present items. This design was also used by Boutonnet & Athanasopoulos (2011). Each picture was presented for five seconds, followed by a blank screen for three seconds and then an asterisk appeared to indicate that the next picture was coming up. This design was piloted with eight speakers of Arabic and English, who were all tested individually. There were also some issues with the order of the stimuli. In the first design, the order of the stimuli unintentionally tended to follow a consistent pattern (feminine, masculine) following the Arabic grammatical gender system. The stimuli were

therefore randomised using Excel to eliminate the effect of order on participant responses.

3.5 Methodological considerations

This section is divided into two sub-sections; the first presents methodological considerations of the experimental tasks with the second describing the questionnaires and their design.

3.5.1 Experiments

Investigating the effect of grammatical gender on cognitive tasks (e.g. categorisation and similarity ratings) has produced abundant evidence of the effect of language on cognition. Such findings led many researchers (e.g. Boroditsky et al., 2005; Sera et al., 1994; Sera et al., 2002, Kurinsky and Sera, 2010) to believe in the effect of this grammatical category on mental representations, as described in sections (2.2.6 and 2.2.7). Nonetheless, it has been frequently argued that these patterns of behaviour may be produced by the demands of specific tasks. Such task demands prompt speakers to use grammatical gender as a strategy to complete the given task (Bowers et al., 1999). Likewise, some types of tasks (e.g. linguistic stimuli, verbal processing) may allow participants to use a wide range of linguistic information and different cognitive strategies. Other researchers (e.g. Munnich and Landau, 2003; Pilling et al., 2003) state that the influence of language may result from implicit verbal coding strategies, which tell us nothing about non-linguistic thought.

For this reason, different experiments were used in the present study to examine whether the effect of grammatical gender continues in tasks which minimise the strategic use of linguistic information. For example, the experimental tasks were organised in such a way that there was no explicit reference to natural gender. The aim of this was to examine whether the relationship between natural and grammatical gender is part of the speakers' mental representations of objects or a result of the task demand. Also, stimuli in the experiments did not include animals as the relationship between grammatical gender and natural gender in Arabic is not arbitrary for nouns that refer to animate entities (only two forms are there for referents of biological sex). Some researchers (e.g. Vigliocco et al., 2005) suggest that different grammatical gender systems across languages might have contributed to the inconsistency of the findings obtained from previous research. Including such a category was therefore

thought to lead to an overestimation of the effects of grammatical gender on categorisation. Also, apart from task instructions, no linguistic input was used or required in the experiments. It can be argued, however, that language cannot be totally avoided in cognitive experiments as participants may implicitly use language without verbalising it. These experiments have tried to keep language usage to a minimum and did not ask for any language production during or after the task.

The voice-attribution task replicated the studies of Sera et al. (1994- 2002) by adopting the same experimental task with another language (Arabic), but using unlabelled pictorial stimuli that represent inanimate objects. Participants were expected to use linguistic information (e.g. word forms) less while performing the task. The stimuli were chosen on the basis that half of them were artificial objects (e.g. *flag* 'alam'*)* and the other half natural objects (e.g. *apple* 'tofahah'*)*. Also, half of each group were grammatically feminine in Arabic (e.g. *table* 'tawleah') and the other half grammatically masculine (e.g. *chair* 'korsi').

The similarity rating task replicated the Boroditsky et al. (2002) study, in which participants were asked to judge the extent to which two inanimate objects were similar. This task did not provide any direct reference to gender, whether grammatical or natural, to block any strategic use of grammatical information during the task. The aim was to examine whether effects of grammatical gender are only found in tasks where judgements of the participants' own representations are required. Some researchers (e.g. Sera et al., 1994; Konishi, 1993) investigated grammatical gender effects on the mental representations of Spanish speakers and argued that participants often attribute male and female-like properties to objects with masculine or feminine nouns, indicating that they link the grammatical gender of object nouns and a corresponding natural gender (see section 2.2.6). Nevertheless, if grammatical gender does indeed affect the speakers' mental representation of objects, this effect should be observed with pictures of objects that share the same gender. This is because these stimuli share some defining features (e.g. male/female-like), compared with other objects of a different grammatical gender. The similarity judgement task measures the extent to which natural gender is likely to be part of the participants' representation of the given objects, as a result of the grammatical gender of the noun to which the object refers. Specifically, participants were asked to rate the similarity between pairs of inanimate objects that have either a similar or different grammatical gender, with no reference whatsoever to gender.

Finally, interactional settings were taken into account. Grosjean (2001, 2010) referred to this as a 'language mode' which is defined as a state of activation in the bilingual's languages and language processing mechanisms. According to this principle, bilinguals may behave like native speakers of their L1 if tested in their native language and like native speakers of their L2 if tested in their second language (L2). Instructions were therefore given in the native language of the participants - Standard Arabic - since bilinguals might behave differently, depending on the social and interactional setting in which they are engaged.

3.5.2 Participant Demographic Information and Language Background

Linguistic and sociocultural information are important in the study of language and bilingual cognition. For example, it is essential to know the number of languages participants understand, as a little knowledge of other languages might produce some differences among speakers from the same language group. Other useful information may be the age of language acquisition, proficiency, daily amount of language use and cultural immersion as reported by Athanasopulos (2011). These factors exerted influence on cognition in studies investigating other domains (e.g. grammatical number, colour).

Although studying all these background variables is unlikely to be possible in one study, the present work aims to obtain as much information as possible about the participants. Careful measurement allows for correlational studies between cognitive performance and these socio-cultural variables (Athanasopoulos, 2011). Questionnaires were designed to cover a range of linguistic and sociocultural data with a mix of closed and open-ended questions. These included areas such as the age of language acquisition, where participants were asked to state all the languages they know, the age at which they acquired each of their languages and their proficiency in each one. Participant language proficiency was measured by asking participants to rate their proficiency on a six point scale where '1-2' means beginner, '3-4' means intermediate and '5-6' means advanced (see appendices 1A, 1B). Although such self-rating language measures can be a valuable tool for measuring participant proficiency and are also considered by some researchers (e.g. Bachman and Palmer,1989) to be indicative of linguistic ability, other researchers claimed they might measure participant perception about what they think their levels are, rather than measuring their actual levels. Therefore, participant proficiency was also measured by an independent language test - the Oxford Quick Placement Test (2001) - which measures performance on a range of language features

such as grammar, syntax and vocabulary. This test was used in previous studies (e.g. Athanasopoulos, 2006, 2007) to divide participants into different groups.

Moreover, the amount of language use was examined, with this variable combining linguistic and sociocultural information. The degree to which people use one of the languages they know depends on the interactional context and their degree of immersion in a specific country. As noted by Athanasopoulos (2011), increasing the opportunity to use language due to these factors leads to an increase in the knowledge of the language used and this in turn provides the person with target-like examples of some linguistic features. To this end, the questionnaires included questions about the length of stay in the L2 speaking country. This variable might promote the participants to follow the linguistic and non-linguistic behaviour of the L2 community in which they live, or have previously lived.

3.6 Summary

This chapter has described the criteria used for selecting the stimuli for the experiments and reported general information about the populations involved. It has also given detailed information about the pilot study and concluded by presenting methodological considerations taken into account during the design stage.

Chapter 4. Voice Attribution Task

4.1 Experiment 1A: The Effects of Grammatical Gender System on Voice Assignments by Monolingual Speakers of Arabic and English

4.1.1 Aims

The focus of this experiment was to examine the potential effects of Arabic grammatical gender on the categorisation of objects by monolingual Arabic speakers. This experiment built upon the work of Sera et al. (1994), who compared the classifications of pictured objects according to their Spanish grammatical gender and the natural/artificial division among speakers of Spanish and English. The Arabic grammatical gender system is mostly similar to that of Spanish (see section 2.2.3). The assignment of grammatical gender is mainly arbitrary for inanimate objects and parallels natural gender for animate objects, additionally every noun must be assigned either a masculine or feminine gender. Therefore, the findings of the current experiment are expected to be similar to those obtained from the study of Sera and colleagues (1994) - as discussed in section 2.2.6 - and to be taken as reliable controls to the completion of the next experiment with bilingual speakers (Experiment 1B).

In this experiment, participants were asked to attribute the voice of either a man or a woman to each pictured object. Two types of stimulus item were presented. Control pictures constituted objects that possess a clear natural gender (e.g. *girl* and *boy)*; they were included as test materials to ensure the participants had understood the task. The classification of these objects as either masculine or feminine was hypothesised to be easily identifiable by speakers of both languages. The other type of stimuli were test pictures which consisted of objects that were only grammatically gendered as masculine or feminine in Arabic, in that they lack any natural gender (e.g. *house, book* and *banana*).

It is worth mentioning, however, that in addition to grammatical gender which might guide the voice assignments, there may be other more universal conceptual distinctions revealed in the participants' classifications. As stated by Ortner (1974), there is a common conceptual distinction that strongly affects the role of males and females in society by associating males with artificial objects and females with natural ones. Other empirical studies have provided evidence that children's

classifications follow this tendency (Sera et al., 1994). Mullen (1990) showed that English speaking children (six to seven years-old) more frequently assigned artificial objects to a male category and natural objects to a female category, suggesting that such associations somehow exist from a relatively early age.

The aim of this task was therefore to investigate whether Arabic grammatical gender would have an effect on Arabic speakers' categorisation of objects. This was of interest to discover whether Arabic speakers would show the gender effects observed for speakers of gendered languages in previous research (e.g. Spanish and French, by Sera et al., 1994 and Italian by Bassetti, 2007). It examined the attribution of male and female voices to pictured objects by monolingual speakers of Arabic and English. The English speakers' performance was taken as a baseline for comparison as speakers of a language that possesses few grammatical gender markers.

4.1.2 Hypothesis

There were two hypotheses:

- Speakers of Arabic and English will differ significantly in their voice assignments to pictures of inanimate objects. Arabic participants will assign voices to inanimate objects according to the grammatical gender in their language. The proportion of same-gender voice assignments will be significantly higher for both masculine and feminine objects.
- Speakers of Arabic and English will follow a female-natural/male-artificial distinction in their voice assignments.

4.1.3 Method

4.1.3.1 Participants

Thirty monolingual speakers of Arabic took part in the study (13 men and 17 women, mean age = 25.55, age range: 18-38 years). As English language is taught in all Saudi schools as a formal school subject from the age of 12[15], participants were asked to take the Oxford Placement Test (2001) in order to measure their level of English. Only thirteen (out of thirty) agreed to do the test and their scores were under seventeen (out of 60, mean = 12.15). The other seventeen participants reported that

[15] Educational policies changed in 2011 and English is now taught in the 4th grade, at around 9-10 years old.

they did not understand a word of English and if they had to do the test, their answers would be the result of chance and guesswork, so they were treated as monolinguals. The majority of participants were first exposed to English at the age of twelve, except for three participants who studied English at the ages of four and five. Participants were students at Saudi universities[16] and were recruited through personal contacts of the author. There were both undergraduates (16), postgraduates (14) and all were born and raised by Arabic-speaking parents. Eight of the thirty had lived in English-speaking countries, five for less than a year, one for three years and two for more than three years, but they never used English for daily communication because of their low level. All participants were tested individually; some were tested in their universities, some in cafés and others in their own houses.

Thirty monolingual speakers of English also participated in the study (12 men, 18 women, mean age = 21.81 years, age range: 18-38 years). Participants were either undergraduates (11) or postgraduates (19) at Newcastle University and were tested individually in libraries[17] and at the Resource Centre at Newcastle University. Eighteen reported that they had a very minimum knowledge of other languages but they never used them for any kind of communication due to their low level (12 knew French, 4 Spanish, 1 Japanese and 1 Irish). One participant was replaced because of a language background inappropriate for this study (he reported intermediate level in Arabic). All participants from both groups participated in the study voluntarily.

4.1.3.2 Materials

Participants were shown forty four pictures which included two types of items, four controls and forty test items (see appendix 2). The four control items consisted of two pictures of men and two of women (grandmother, grandfather, girl and boy). These control items were used to ensure that the participants had clearly understood the instructions of the task and that their judgements reflected their attribution of natural gender, namely that participants should assign voices according to the referents' natural gender (biological gender). For example the picture of a grandmother should be assigned a woman's voice by all participants (see table 4.1, below).

[16] Princess Nora University, and King Saud University.
[17] The Robinson Library and City Library in Newcastle-upon-Tyne.

Male	Female
Grandfather ‘jadd’	Grandmother ‘jaddah’
Boy ‘walad’	Girl ‘bent’

Table 4.1 List of control items

Masculine		Feminine	
Artificial	**Natural**	**Artificial**	**Natural**
Flag ‘allam’	Mushroom ‘feter’	Vacuum cleaner ‘meknasah’	Apple ‘tofahah’
Key ‘moftah’	Watermelon ‘jeh’	Ball ‘korah’	Sun ‘shams’
Chair ‘korsy’	Nose ‘anf’	Star ‘najmah’	Banana ‘mozah’
Bus ‘bass’	Moon ‘qamar’	Toothbrush ‘forshah’	Corn ‘thorah’
Pencil ‘qalam Rosas’	Mouth ‘fam’	Spoon ‘melaqah’	Strawberry ‘farawlah’
House ‘bait’	Thumb ‘ebham’	Table ‘tawlah’	Tree ‘shajrah’
Heart ‘qalb’	Arm ‘theraa’	Car ‘sayarah’	Flower ‘zahrah’
Dress ‘thobe’	Mountain ‘jabal’	Basket ‘sallah’	Ear ‘othun’
Door ‘bab’	Lettuce ‘khas’	Handbag ‘shantah’	Hand ‘yadd’
Book ‘ketab’	Head ‘raas’	Traffic light ‘esharah’	Eye ‘ayn’

Table 4.2 List of test items

The test pictures were intended to test the effect of grammatical gender on the voice assignments of Arabic participants. Of the forty test items that illustrated inanimate objects, ten depicted artificial objects that are feminine in Arabic (e.g. *basket* ‘sallah’), ten artificial objects that are masculine in Arabic (e.g. *flag* ‘allam’), ten naturally occurring objects that are feminine in Arabic (e.g. *apple* ‘tofahah’) and ten naturally occurring objects that are masculine in Arabic (e.g. *mushroom* ‘feter’) (see table 4.2). Pictures were taken from Snodgrass and Vanderwart (1980) and Szekely et al. (2004), all stimuli chosen were presented as black and white drawings and edited to be of the same size in order to eliminate any potential variable effect such as colour or size (Flaherty, 2001). Stimuli were presented on PowerPoint slides one item at time using a 17-inch screen with a resolution of 1600 X 900 pixels. Each picture was presented for five seconds, followed by a blank screen for three seconds, before an asterisk appeared to indicate that the next picture is coming up. The stimuli were randomised to avoid any possible effects of order and there was one order used with all participants. All pictures were numbered from ‘1’ to ‘44’ and an answer sheet was provided for each participant (see Appendices 3A & 3B). Figure 4.1shows examples of six of the black and white pictures that were used as stimuli.

Grammatically feminine	Grammatically masculine
Examples of the control items	
	
Examples of natural entities	
	
Examples of artificial items	
	

Figure 4.1 Black and white copies of sample pictures used as stimuli

In addition to the task, biographical and language background was obtained using a questionnaire which asked about the participants' age group, gender, the number of languages known and the age of acquisition of each language (see appendices 1A & 1B). They were also asked to rate their proficiency level in each of their languages. Other questions looked at the average time of language usage, cultural exposure and the participants' qualifications. There were two versions, one written in Arabic and the other in English.

4.1.3.3 Procedure

The whole task lasted between twenty and thirty minutes and participants were tested individually. Each session started by giving participants an ethics form (see appendices 4A & 4B), then assigning two consent forms, one for the researcher and the other for the participants to keep (see appendices 5A & 5B). Participants filled out their biographical and language background, then the task procedure was explained and participants were informed that their task was to assign men's or women's voices to pictures of inanimate objects. The instructions for the experiment were all given in the participants' native language and all participants were tested individually.

As suggested by Kurinski and Sera (2011), it was important to avoid using the words *gender*, *masculine* and *feminine* in the instructions to keep participant attention away from the focus of the task, that is grammatical gender; so they were

never told about covert gender manipulation. The participants were informed that there was no right or wrong answer for this task and that their categorisations should be according to their opinions. Finally, the participants were told before starting that they should tell the experimenter to stop the task at any time they need to clarify something or if they were simply getting tired. One Arabic monolingual participant was replaced as she withdrew in the middle of the experiment, although the aims of the task were explained before starting the task, this participant was reluctant to complete the task as she thought the task would be measuring how good her thinking skills were. The exact instructions were adopted from Sera et al. (1994) and were as follows:

> We are thinking of making a new movie in which some everyday objects come to life, sing and dance. You will see a series of pictures of these objects and will need to determine whether each item should have a man's/boy's voice or a woman's/girl's voice. If you decide that an object should have a female voice please circle "F" in the column named "VOICE" on your answer sheet. If you decide that it should have a male voice, then circle "M". You will see the image for 5 seconds, it will be followed by a blank for 3 seconds, and then an "asterisk" will indicate that the next picture is coming up. Press the SPACE BAR to start the experiment. Please make sure that picture numbers correspond to the numbers on your answer sheet.

The specific Arabic instructions were back-translated again by two bilingual speakers to confirm they were as similar as possible. They were:

> اننا نفكر بعمل فيلم جديد حيث أن كل الأشياء تصبح حية و تغني و ترقص. سترى سلسلة من الصور لهذه الأشياء و عليك أن تحدد ما إذا كان صورة الشئ يجب أن تعطى صوت امرأة وبنت أو صوت رجل وولد. إذا قررت أن الشئ في الصورة يجب أن يأخذ صوت رجل، أحط الحرف (ر) في العمود المسمى (صوت) في ورقة إحابتك. وإذا قررت أن الشئ في الصورة يجب أن يأخذ صوت امرأة، أحط الحرف (م). من فضلك تأكد بأن رقم الصورة يتفق مع الأرقام الموجودة في ورقة الإجابة.

Finally, as an extra step in ensuring that Arabic participants had no problem identifying the gender of the presented items, their knowledge of grammatical gender for the test items was tested (see appendices 6A & 6B).

4.1.4 Results

The percentage of times each picture was assigned a voice according to grammatical gender in Arabic was calculated (raw data are provided in appendix 7). Table 4.3 below shows the mean percentage of times (standard deviations in brackets) control and test pictures were assigned a voice consistent with Arabic grammatical gender by monolingual speakers of Arabic and English. Figure 4.2 present these percentages.

Language Group	Pictures		F	P
	Control	Test		
English (*N*=30)	100.0	56.83 (*SD*=6.33)	130.196	0.000
Arabic (*N*=30)	100.0	84.00 (*SD*=11.40)		

Table 4.3 Percentage of times English and Arabic monolingual speakers' voice assignments honoured grammatical gender of the Arabic language

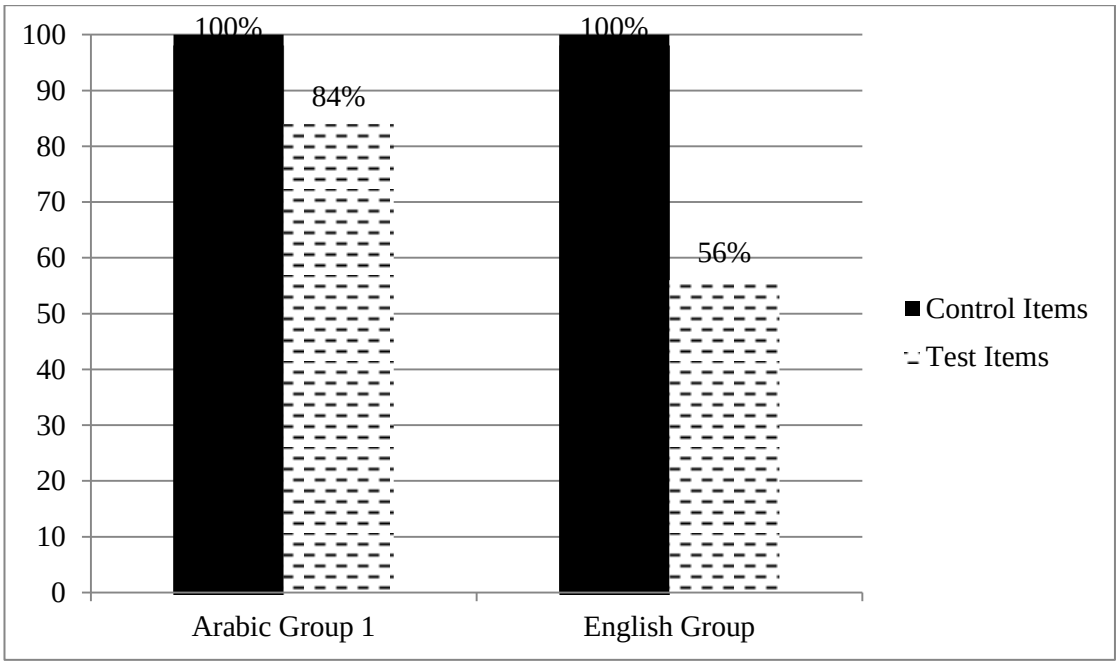


Figure 4.2 Voice assignments of control and test items by Arabic and English speakers

All participants assigned voices to the control pictures according to natural and grammatical gender a high percentage of the time (100%). For the test items, a two-factor ANOVA was conducted to examine the percentage of times they were classified according to the Arabic gender system with Language (Arabic vs. English) as a between-subjects factor and Arabic Grammatical Gender (masculine vs. feminine) as a within-subject factor. This showed the main effect of Language [F (1, 59) = 130.196, p <0.00], with Arabic speakers assigning items in accordance with

Arabic grammatical gender, (M= 84, *SD*= 6.33) significantly more than English speakers, (M= 56, *SD*= 11.40). The Cohen's d of 3.03 indicated that this was a large effect size difference between the groups. This result confirmed the first hypothesis concerning the significant difference between voice assignments by speakers of Arabic and English[18].

To test the second hypothesis, that is the effect of female-natural/male-artificial distinction, a three-way ANOVA[19] was performed in order to examine whether the differences between masculine and feminine items found in the English group could be related to different types of items, both natural and artificial. This was done with Language (Arabic vs. English), Arabic Grammatical Gender (masculine vs. feminine) and Conceptual Class (natural vs. artificial) on the mean percentage of items that were classified according to the Arabic gender system by monolingual Arabic speakers.

Many reliable effects were observed: a main effect of Language, F (1,240) = 189.568, $p < 0.00$; a Language × Gender interaction, F (1,240) = 9.247, $p < 0.03$; a Language × Concept interaction, F (1,240) = 4.110, $p < 0.04$; a Gender × Concept interaction, F (1,240) = 42.303, $p < 0.00$; and a three-way interaction between Language, Gender, Concept, F (1,240) = 7.770, $p < 0.00$. Language by Gender interaction indicated that English monolingual speakers assigned a man's voice to masculine items more often than to feminine items (58% vs. 54%). Arabic monolingual speakers classified grammatically feminine items as having a woman's voice more often than grammatically masculine items (88% vs. 80%).

There was, however, no significant effect for Gender in analysis by items (F (I, 240) = 1.027, p= 0.312 > 0.5). This means that grammatically consistent voice assignments to grammatically feminine items did not significantly differ from masculine items. The mean average for grammatically consistent voice assignments made by both groups of participants to masculine items was 69.41 (*SD*= 1.39) and to feminine items was 71.41 (*SD*= 1.39).

Furthermore, Language × Concept interaction, through simple main effects by participant, indicated that both Arabic and English speakers made more grammatically consistent voice assignments to masculine artificial items and feminine natural items,

[18] The English speakers performed the task at a chance level (56%), a binomial probability test finds a critical value of 25 out of 40 (62.5%).

[19] Three-way ANOVA is a statistical test used to determine the effect of three nominal predictor variables on a continuous outcome variable. The test analyses the effect of independent variables on the expected outcome along with their relationship to the outcome itself.

suggesting that both groups were following masculine-artificial/feminine-natural distinction in their assignments. Simple main effects by item indicated that both groups made grammatically consistent voice assignments to masculine artificial items 75% (*SD*= 19.08) rather than to feminine artificial items 64% (*SD*= 26.51), effect size (cohen's d was small *d*= 0.47, and to feminine natural items 78% (*SD*= 16.82) rather than to masculine natural items 63% (*SD*= 20.05), with a large effect size, *d*= .81. Analysing voice assignments by conceptual categories, Arabic monolingual speakers assigned a woman's voice to feminine natural items more frequently than to feminine artificial items (90% vs. 86%); for example, the Arabic speakers assigned a woman's voice to a *strawberry* more than to a *table* although both *strawberry* and *table* are grammatically feminine. A similar pattern was observed for the English group who assigned a woman's voice to feminine natural items more frequently than to feminine artificial items (66% vs. 43%). Furthermore, both groups assigned a man's voice to masculine artificial items more often than to masculine natural items (85% vs. 74%) for Arabic speakers and (65% vs. 52%) for English speakers. See figure 4.3 below and table 4.4 which explains these values.

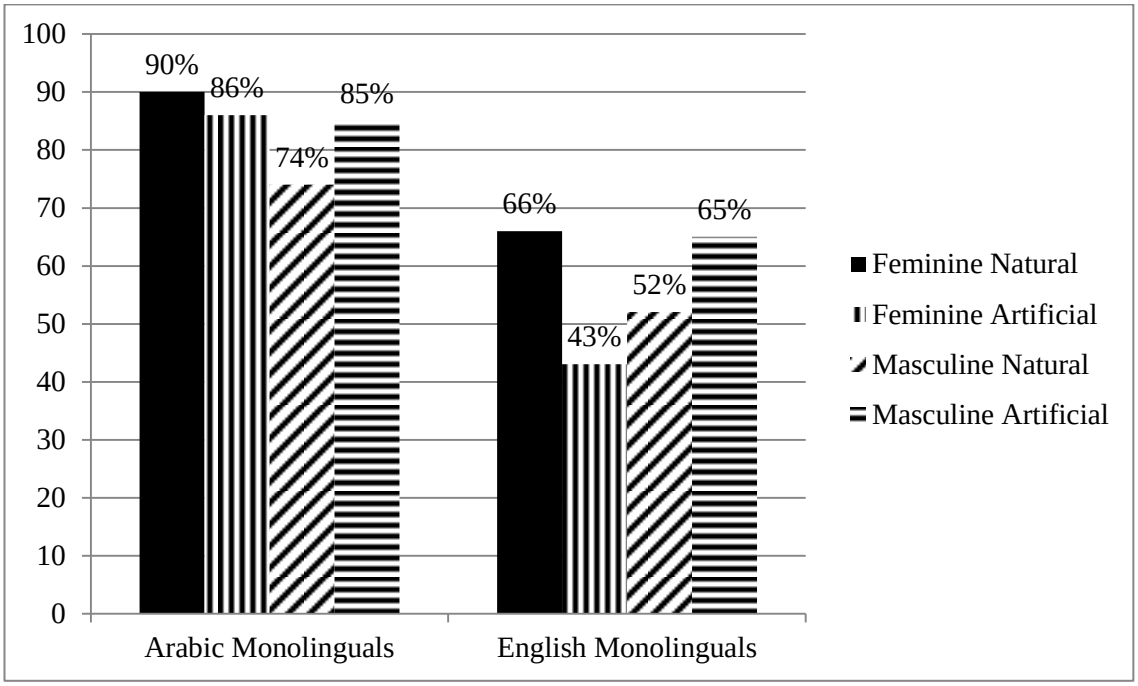


Figure 4.3 Mean percentage of grammatically consistent voice assignments for artificial and natural categories for masculine and feminine grammatical gender by monolingual speakers of Arabic and English

Language	Grammatically Feminine		Grammatically Masculine	
	Natural	Artificial	Natural	Artificial
Arabic	90.00 (13.64)	86.00 (12.75)	74.67 (19.25)	85.33 (15.02)
English	66.33 (9.99)	43.33 (18.06)	52.00 (13.493)	65.67 (17.75)

Table 4.4 Percentage of times grammatically feminine and masculine natural and artificial items were assigned a voice according to Arabic grammatical gender

It can be seen that English speakers assigned a woman's voice to natural items that were grammatically feminine in Arabic more often than to natural items that were grammatically masculine (66% vs. 52%); similarly, they assigned a man's voice to artificial items that were grammatically masculine more often than to artificial items that were grammatically feminine (65% vs. 43%), suggesting a non-arbitrary relationship between grammatical and conceptually masculine items. It seems that both Arabic and English speakers were sensitive to feminine-natural /masculine-artificial distinction in their voice assignments. This result confirmed the second hypothesis that both speakers of Arabic and English follow the feminine-natural/masculine-artificial distinction when assigning voices to inanimate objects. Such results show that both grammatical gender and conceptual category influenced the decisions of Arabic and English monolingual speakers. For instance, Arabic monolinguals assigned a woman's voice to 88% of grammatically feminine items, but 90% of the time to natural grammatically feminine items and 86% of the time to grammatically feminine artificial items. Similarly, English speakers assigned a man's voice to 58% of the grammatically masculine items, but 65% of the time to artificial grammatically masculine items and 52% of the time to grammatically masculine natural items. For Arabic monolinguals, the grammatical gender classifications yielded significantly different judgments overall and for English monolinguals the natural/artificial distinction yielded reliably different judgments overall. Therefore, Arabic grammatical gender was significant in influencing categorisation among Arabic monolinguals whereas conceptual category was an influence on the categorisation of English monolinguals.

4.1.5 Discussion

The main finding of this study is that grammatical gender was shown to affect voice assignment to inanimate objects. The first hypothesis that Arabic speakers would assign voices according to Arabic grammatical gender for both grammatically masculine and

grammatically feminine objects was therefore confirmed. These results support previous findings reported by Sera et al. (1994); Sera et al. (2002); Flaherty (2001); Phillips & Boroditsky (2003) and Bassetti (2007) on the effect of grammatical gender on categorisation. In this experiment, Arabic monolinguals made more voice assignments to objects that lack a natural gender according to the grammatical gender assignments of Arabic language to the noun names of these objects. English speakers, however, made significantly fewer grammatically consistent voice assignments than Arabic monolinguals. These results are very important in showing the effects of language on our thinking, so they could be used as a basis for studying the potential effects of bilingualism by testing the bilingual group on the same task to see whether or not they differ from other monolinguals of their L1 (Experiment 1B below focuses on this issue).

Furthermore, voice assignments by English participants were at a coincidental level for masculine and feminine objects. A possible explanation for this is that English speakers made their voice assignments according to a certain method e.g. assigning a man's voice to artificial items and a woman's voice to natural items. This in turn made their voice assignments - more than half the time - consistent with the Arabic gender system. Both speakers of Arabic and English tended to assign a female voice to natural objects that are grammatically feminine in Arabic and a male voice to artificial objects that are grammatically masculine. In fact, the English speakers assigned masculine voices to grammatically masculine artificial objects more often than to natural objects. A similar pattern was found for the feminine natural objects which were assigned a woman's voice more often than feminine artificial ones.

Arabic speakers, however, made more grammatically consistent voice assignments than the English group to all types of objects, suggesting that they were reliably influenced by the grammatical gender system of their language. Previous research by Sera et al. (2002) compared voice assignments between Spanish and English speakers and showed that the English group assigned a man's voice to masculine objects more frequently than to feminine objects. Such a tendency led the authors to assume that Spanish grammatical gender of inanimate objects captured a universal conceptual tendency which also predicts the judgements of the English speakers.

Another view is that the English speakers' voice assignments might be consistent with male-artificial/female-natural distinction, suggested by Ortner (1974; Sera et al., 1994) whereas Arabic monolingual speakers tended to follow the grammatical gender of their language to perform the task. The current findings support this view as both

speakers of Arabic and English assigned more masculine voices to artificial objects than to natural objects even though both of these objects were grammatically masculine in Arabic. Similarly, they assigned more female voices to natural objects than to artificial objects although both were grammatically feminine in Arabic. Arabic monolingual speakers' classifications were, however, significantly above average for all inanimate objects, suggesting a strong effect of the grammatical gender of Arabic language on the speakers' performance in the voice-attribution task.

These findings revealed, by and large, that monolingual speakers of Arabic and English were different in their voice assignments. Arabic speakers' responses conform to the gender system of their language significantly more than those of English speakers. A close analysis of artificial-masculine, artificial-feminine, natural-masculine and natural-feminine confirms this tendency. Sloman and Malt (2003) argued that artificial categories are not stable, meaning that the effects of language and culture might increase for artificial objects more than they would for natural ones. Nevertheless, the findings of this study reveal that Arabic monolinguals were reliably influenced by the gender system, regardless of the object's conceptual category, be it artificial or natural.

The aforementioned findings provide evidence that some grammatical features of a language - such as grammatical gender - could affect cognition and this offers vital insights into how humans think. These findings provide a convincing answer to the first research question about the effects of a grammatical gender system on the categorisation of objects for Arabic monolingual speakers. The fact of the matter is, however, that not many people remain monolingual throughout their life. On an international basis, bilingualism is very common and much more the rule than the exception in many countries (Cook, 1997, 2002). If different languages perceive reality and the world in different ways and if linguistic experience can deeply influence our cognitive processes about reality and the world, then how are different categorical features in languages reconciled in the mind of bi/multilingual speakers? To this end, the next experiment aimed to investigate the effect of language on cognition among bilingual speakers. More specifically, it focuses on the effect of Arabic grammatical gender on Arabic-English bilinguals' performance in a cognitive task.

4.2 Experiment 1B: The Effects of Grammatical Gender on Voice Assignments by Arabic-English Bilinguals

4.2.1 Aims

As mentioned in previous chapters, a growing body of evidence suggests a correlation between grammatical and conceptual gender, supporting the view that languages have an impact on cognition. Most of the earlier investigations in the field of grammatical gender, however, involved comparisons of two or more monolingual populations, as was seen in Experiment 1A. For example, Sera et al. (1994) compared Spanish and English monolinguals speakers; while Ramos and Roberson (2010) tested monolingual speakers of Portuguese and English. As most of the world population is either bilingual or multilingual, it is worth investigating whether bilingualism affects the way bi/multilingual people think. The question of interest here is to ask how bilinguals think when they know two languages; particularly when their two languages represent the same object or event differently. Bilinguals may have access to two different concepts and consequently think about this object or event in a way that is different from monolingual speakers of either language (Bassetti, 2007).

The present study attempts to investigate the effects of grammatical gender on the thinking of Arabic-English bilinguals by comparing their categorisation of objects to that of monolingual speakers of Arabic and English (as found in Experiment 1A). The categorisation task was a voice-assignment task adopted from Sera et al. (1994, 2002) that was designed to provide evidence on the conceptual gender perception of inanimate entities by all participants. It further sought to avoid the shortcomings of earlier work on this area by testing speakers' performance on non-linguistic tasks and taking into consideration a number of linguistic and sociocultural variables that might play a role in the bilinguals' performance, (e.g. culture, language usage, proficiency level, age of acquisition and academic qualifications, see section 2.2.8).

The aim of this study is to investigate whether Arabic-English bilinguals would show the same effect of grammatical gender as monolinguals, or would develop new concepts that are in between the concepts of their two languages. Would they even do something different from either, supporting the notion of multi-competence proposed by Cook (1991, 2002) as a consequence of bilingualism? The study involved bilinguals with different proficiency levels - intermediate and advanced - in order to investigate the extent to which learning another language with no grammatical gender system

would change the bilinguals' performance. To summarise, the experiment aimed to answer the following question:

- Would learning another language change the bilinguals' cognitive performance and make them categorise objects differently from monolinguals? If yes, to what extent would this occur?

4.2.2 Hypothesis

There were two hypotheses for this study:

- Arabic-English bilinguals will categorise inanimate objects differently from Arabic monolingual speakers.
- Intermediate and advanced bilinguals will differ in their voice assignments according to proficiency level in their L2.

4.2.3 Method

4.2.3.1 Participants

In addition to the thirty English-speaking monolinguals and thirty Arabic-speaking monolinguals who participated in Experiment 1A, two groups of Arabic-English bilingual speakers participated in this experiment: one group consisted of thirty advanced Arabic L2 learners of English and the other consisted of thirty intermediate Arabic L2 learners of English. The two bilingual groups were divided according to how they scored on the Oxford Placement Test (2001); the mean averages being 38.5 and 52 (out of 60) for the intermediate and advanced groups respectively. Participants voluntarily participated in the experiment and were recruited by personal contact from four universities, Newcastle and Northumbria in the UK and King Saud and Northern Borders in Saudi Arabia. Table 4.5 below shows participant profiles, detailing mean age, gender, proficiency category, language usage, time spent in the UK and academic qualifications.

	Language Groups			
	Arabic monolinguals (*N*=30)	**Intermediate bilinguals (*N*=30)**	**Advanced bilinguals (*N*=30)**	**English monolinguals (*N*=30)**
Demographics				
Age	25.55 (age range: 18-38)	25.75 (age range: 20-34)	27 years (age range: 22-38)	21.81 years (age range: 18-38)
Male	13	10	13	12
Female	17	20	17	18
Proficiency Category				
Early bilinguals	N/A	14	8	N/A
Late bilinguals		16	22	
Daily Language Usage (Average time)				
Arabic	All the time	12 hrs	10.5 hrs	None
English	None	3 hrs	4.75 hrs	all the time
Living in the L2 Country (England)				
Yes	8	18	26	7
No	22	12	4	23
Academic Qualifications				
Undergraduates	16	19	4	8
Postgraduates	14	11	26	22

Table 4.5 Participants' demographic and linguistic background

4.2.3.2 Materials

The same materials were used as in Experiment 1A: forty four pictures of inanimate items, with four controls and forty test items (see section 4.1.4.2). In addition, the Oxford Quick Placement Test (2001) was used in order to divide the bilingual participants into intermediate and advanced groups. This test is divided into two parts; with the first containing forty questions, while the second has twenty questions. Only advanced bilinguals were expected to proceed to the second part as it was meant to be more difficult than the first part, however, all bilingual participants actually answered the two parts. There was no time limit for the test so the participants took their time to

complete it, ranging from twenty to forty minutes. With regard to the questionnaire, the bilingual groups were asked to fill in the Arabic version (see appendix 1B) and Arabic was also the language of instruction.

4.2.3.3 Procedure

Similar procedures were used to the previous experiment but the whole session lasted longer, between forty and fifty minutes. Participants were tested individually at their universities. Each session started with the signing of two consent forms. The task procedure then started with identical instructions to Experiment 1A (see p.64- 65), and explained to the bilingual participants in their native language by a native speaker of Arabic. The participants were next asked to complete the questionnaire and language test in Arabic. To conclude, a short interview took place where participants were asked what they were thinking about during the voice-assignment task.

4.2.4 Results

The data was treated in the same way as for Experiment 1A. First, participants' voice assignments to the control items were examined. All participants overwhelmingly assigned a female voice to the natural gender female control items (100%) and a male voice to the natural gender male control items (100%). This suggests that the participants fully understood the task as their categorisations reflected their attribution of natural-like gender properties. The number of times each test item was categorised according to Arabic grammatical gender was then examined through a one-way ANOVA[20]. Figure 4.4 below shows the mean percentage of times test pictures were categorised according to Arabic grammatical gender by all language groups (raw data is attached as appendix 8). Table 4.6 shows the mean and standard deviations of the voice assignments for the test items by all groups.

[20] One-way ANOVA is used in research to determine whether there are any significant differences between the means of three or more independent (unrelated) groups.

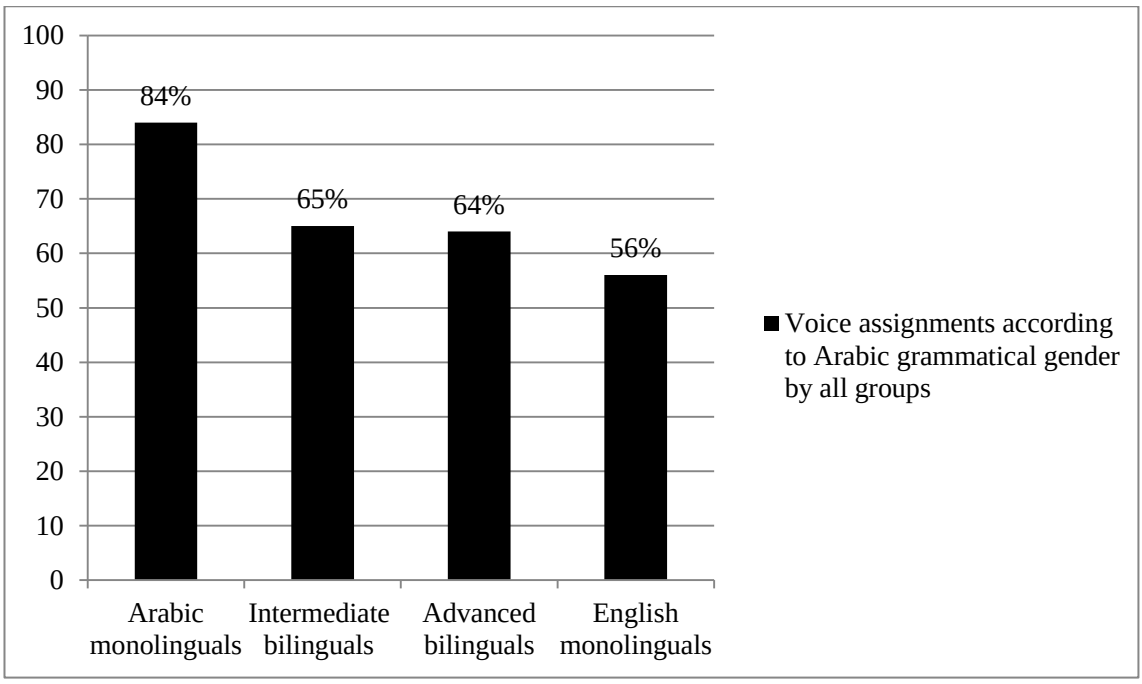


Figure 4.4 Voice assignments to all test pictures by monolingual speakers of Arabic and English and intermediate and advanced Arabic-English bilinguals

Language Group	Voice assignment to all test items
Arabic monolinguals	84.00 (*SD*=11.40)
Intermediate bilinguals	65.08 (*SD*=17.30)
Advanced bilinguals	64.33 (*SD*=16.07)
English monolinguals	56.83 (*SD*=6.33)

Table 4.6 Mean (standard deviations in brackets) of the voice assignments made by all groups to all test items

As can be seen in the graph, Arabic monolinguals made same-voice assignments more often than the other groups (M= 84.00, *SD* =11.40), English monolinguals made fewer same-voice assignments (M= 56.83, *SD* =6.33) than the other groups, intermediate bilinguals and advanced bilinguals were slightly similar in their voice assignments which came between the two monolingual groups (M= 65.08, *SD*=17.30; M= 64.33, *SD* =16.07, respectively). This analysis shows a significant main effect of language, *F1* (3, 479) = 61.713, $p < .05$.

Although the one-way ANOVA showed a significant difference between the groups, this test cannot tell us which specific groups were significantly different from each other. In order, therefore, to determine which of these groups differs from each

other, a Tukey (HSD) post-hoc test[21] was conducted. This test revealed that English speakers significantly differed from Arabic monolinguals, $p < .001$ and intermediate and advanced bilinguals, $p < .000$. Similarly, Arabic monolinguals differed from the English speakers and two bilingual groups, $p < .000$. Intermediate bilinguals and advanced bilinguals, however, did not differ from each other, $p < .706$. This analysis indicated an effect of change from learning another language. This change does not, however, increase with proficiency as both bilingual groups performed similarly in the voice-attribution task, meaning that there was no statistically significant difference between them.

There was no statistical significant effect of the gender of the participants on their voice assignments, female (M (SD) = 67.84 (14.51) and male (M (SD) = 67.13 (12.84), $F1$ (1,119) = 076, p =.784).

Analysis of Grammatical Gender (feminine vs. masculine items)

In order to examine the participants' categorisations according to types of items, a two-way ANOVA was conducted with Language (English speakers vs. Arabic monolinguals vs. intermediate bilinguals vs. advanced bilinguals) as between-subjects factor and Arabic Grammatical Gender (masculine vs. feminine) as a within-subject factor. The analysis yielded a significant main effect of language [F1 (3, 472) = 62.229, $p < .000$]. Arabic monolingual and bilingual groups made more feminine voice assignments to grammatically feminine items (M= 68.25, *SD*= 17.31) than to grammatically masculine items (M= 66.87, *SD*= 14.99). Although this suggests that the feminine is more salient, this effect was not significant in the analysis by gender [F1 (1,479) =.982, p < .322]. The Language by Gender interaction did not reach significance [F1 (3,479) = 2.203, p < .087]. Table 4.6 below presents the percentage of times same-gender voice assignments occurred for feminine and masculine items in the four groups. Figure 4.6 visualises the difference between Arabic monolinguals and Arabic-English bilinguals.

[21] Post-hoc tests are used to compare the mean of groups that have been determined to have some overall statistically significant differences and additional exploration of the differences is needed to provide specific information on which means differ significantly from each other.

Language groups	Arabic grammatical gender	
	Masculine	Feminine
Arabic monolinguals	80.00 (15.48)	88.00 (10.95)
Intermediate bilinguals	64.83 (13.22)	65.33 (14.85)
Advanced bilinguals	63.83 (10.96)	64.83 (12.14)
English monolinguals	58.83 (11.34)	54.83 (11.48)

Table 4.7 Percentage of times (standard deviation in brackets) same-gender voice assignments made for feminine and masculine items by all groups

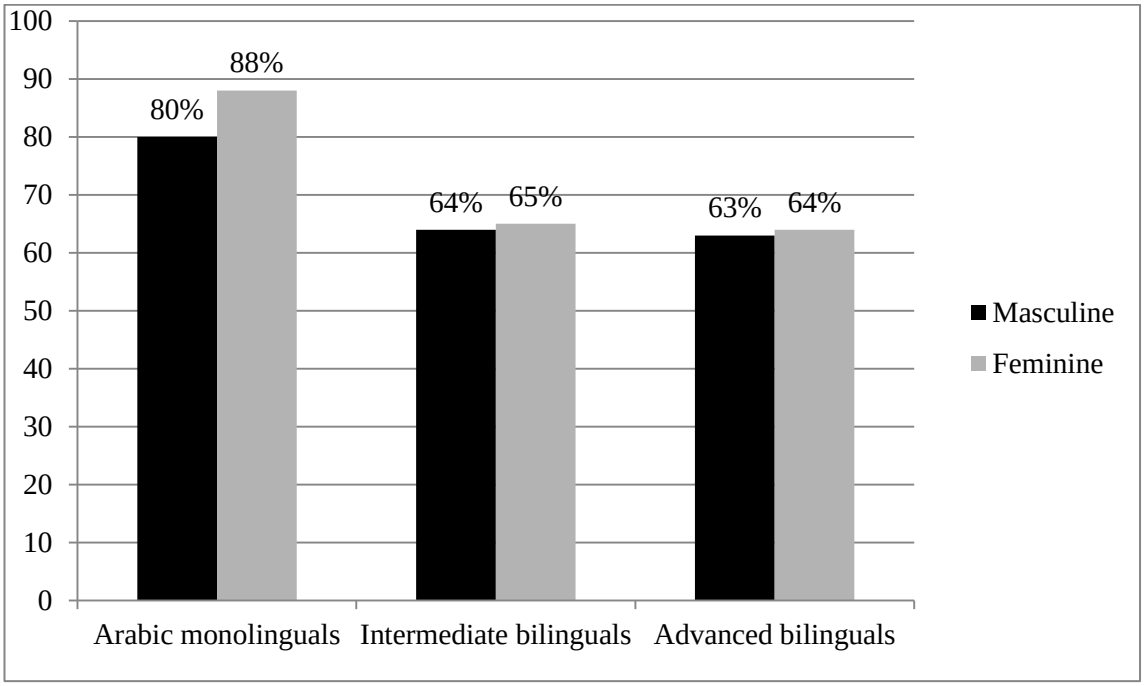


Figure 4.5 Voice assignments made for grammatically feminine and masculine items by Arabic monolinguals and Arabic-English bilinguals

Analysis of Conceptual Categories (artificial vs. natural items)

A two way ANOVA analysis of voice assignments to conceptual categories revealed that the four groups made same-gender voice assignments to artificial and natural items equally often (M= 67.64, *SD* = 20.09) for artificial items and (M= 67.29, *SD* = 18.60) for natural items. This shows that this effect was not significant [*F1* (1,479) = 009, $p <$ .925], see figure 4.7 below.

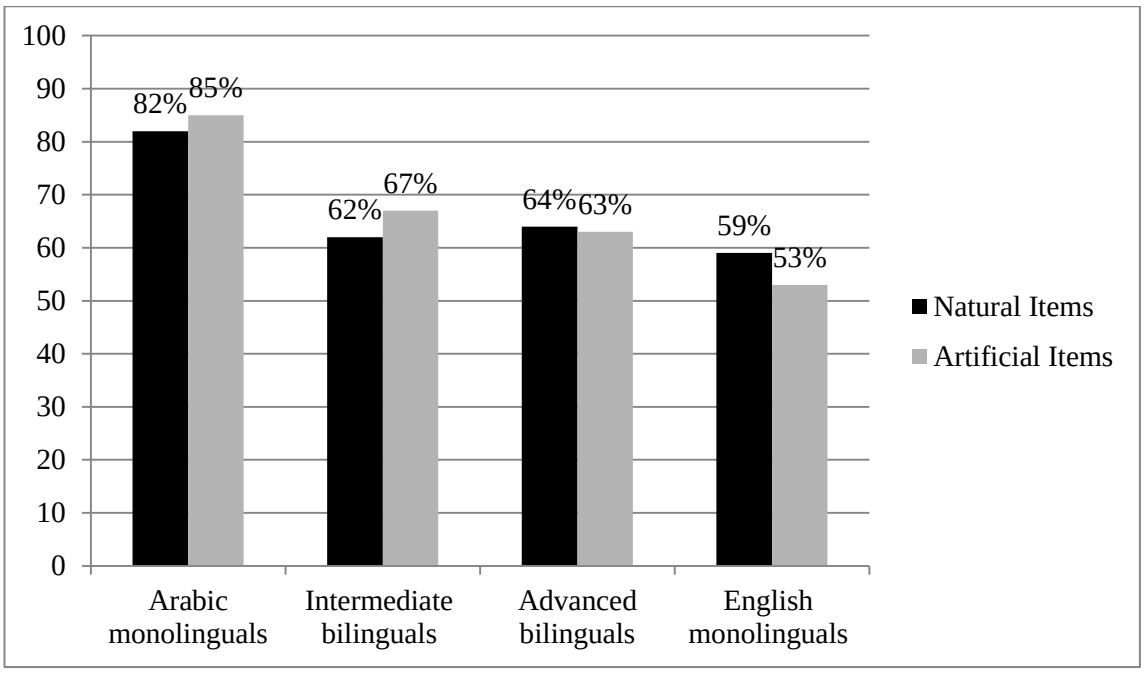


Figure 4.6 Participants' voice assignments according to conceptual categories: natural vs. artificial

To be more specific, natural items were assigned voices that were consistent with their Arabic gender more consistently by Arabic monolinguals (M= 82.33, *SD* = 18.26), as compared to intermediate bilinguals (M = 62.83, *SD*= 18.13), advanced bilinguals (M= 64.83, *SD* = 15.01), and English monolinguals (M= 59.17, *SD* = 13.81). Similar assignments were made to artificial items which were assigned voices consistent with their Arabic grammatical gender by Arabic monolinguals (M= 85.67, *SD* = 13.82); intermediate bilinguals (M= 67.33, *SD* = 16.45) and advanced bilinguals (M= 63.83, *SD* = 17.18), but a lower mean was obtained from the English monolingual group (M= 53.00, *SD* = 17.78). There was a significant interaction between language and concept, *F1* (3,479) = 2.612, $p < .05$. This effect tells us that voice assignment across all language groups was different for natural and artificial objects.

Further analysis of the interaction between Conceptual Categories and Grammatical Gender showed that all groups made more grammatically consistent voice assignments to masculine artificial items and to feminine natural items, indicating that they followed natural-feminine/artificial-masculine distinction in their categorisations. This shows us that participants from all groups showed a trend to assign more male voices to masculine artificial items (M= 71.75, *SD* = 16.27) than to feminine artificial items (M= 61.25, *SD* = 18.17). Similarly, all participants assigned more female voices to feminine natural items (M= 73.33, *SD* = 17.06) than to feminine artificial items (M= 63.17, *SD* = 22.56). There was a significant interaction between grammatical gender and concept, *F1* (1, 479) = 53.915, $p < .000$), this effect tells us that voice assignments

differed according to the objects' grammatical gender and conceptual category. That is to say, voice assignments to masculine and feminine objects differed when these objects were natural or artificial. However, the Language × Gender × Conceptual Categories interaction was not significant (*F1* (3, 480) = 2.443, $p < .064$). Analysis of these means is shown in table 4.7 below.

Language	**Conceptual Categories**			
	Artificial		**Natural**	
	Arabic Grammatical Gender			
	Masculine	Feminine	Masculine	Feminine
Arabic monolinguals	85.33 (15.02)	86.00 (12.75)	74.67 (19.25)	90.00 (13.64)
Intermediate bilinguals	71.67 (12.88)	63.00 (18.59)	58.00 (18.45)	67.67 (16.75)
Advanced bilinguals	67.33 (16.59)	60.33 (17.31)	60.33 (13.25)	69.33 (15.52)
English monolinguals	62.67 (11.12)	43.33 (18.06)	52.00 (13.49)	66.33 (9.99)

Table 4.8 Mean percentages of same-gender voice assignments (standard deviation in brackets) for artificial and natural items by language and Arabic grammatical gender

Age of L2 Acquisition and Categorisations (early vs. late bilinguals)

It is possible that bilingual participants who acquired proficiency in their second language earlier may have differed from those who acquired the language later (see section 2.2.8). The current experiment divided the two bilingual groups, intermediate and advanced, into early and late bilinguals following Johnson and Newport (1989). They were pooled together at first ($n = 60$) because previous analysis did not show significant differences between their performances. They were subsequently divided on the basis that early L2 acquisition bilinguals acquired L2 proficiency before the age of seven and late L2 acquisition bilinguals acquired L2 proficiency after the age of seven.

To test this possibility, a one way ANOVA was conducted. Late bilinguals made slightly more same gender voice assignments (M= 65.26, *SD*= 16.55) than the early bilinguals (M= 63.75, *SD*= 17.04), but this effect was not significant, *F1* (1, 240) = 456, $p < .500$. In addition, a three-way ANOVA was conducted with (Age of L2 Acquisition: early vs. late) × 2 (Gender: masculine vs. feminine) × 2 (Concept: natural vs. artificial). Again the analysis revealed that there was no main effect of age of L2 acquisition on the bilinguals' performance, nor did it reliably interact with any other variable. A linear regression was performed on the age of L2 acquisition as a predictor variable and voice assignments by all bilingual groups as an outcome variable. The results showed that there was no significant correlation between those two variables F (1,239) = 456, $p < .500$, $R^2 = .044$. Therefore, early or late bilingualism did not affect Arabic bilinguals' performance in the voice-attribution task, $b = 1.513$, t = .675, $p < .500$.

L2 proficiency and categorisation

The bilinguals' L2 proficiency was also taken into consideration as a variable that might affect their performance. A linear regression was conducted on this variable as a predictor variable and the bilinguals' voice assignments to test items as an outcome variable. The results revealed that there was no statistically significant correlation between the bilinguals' L2 proficiency and their voice assignments (R^2 = .038, b = .750, t = .292, p < .771).

Length of stay in L2-speaking country

Previous research found that the length of stay in the bilinguals' L2-speaking countries affected their performance. In a study that investigated shape and material preferences among Japanese-English bilinguals and monolinguals**,** Cook et al. (2006) found that Japanese-English bilinguals who had lived in England for more than three years had moved some way towards the English preference. This indicated that cultural immersion in the L2 country may affect categorisation preferences in bilinguals.

In order to test the effect this variable might have on the bilinguals' categorisations, bilingual participants were divided into three groups according to length of stay in their L2 countries, that is bilinguals who had lived in an English speaking country[22] for more than three years (n = 34), bilinguals who lived in an English speaking country for less than three years (n = 10) and bilinguals who did not live in an English speaking country (n = 16). A one way ANOVA indicated that bilinguals who stayed in an English speaking country for less than three years attributed voices in accordance with Arabic grammatical gender (M= 66.75, *SD*= 15.91) slightly more than those who had lived in an English speaking country for more than three years (M= 63.75, *SD*= 16.73). This effect does not, however, reliably reflect any effect of cultural immersion, because an investigation of categorisations of bilinguals who never lived in an English speaking country showed that their grammatically consistent voice assignments were slightly fewer than those had who lived in an English country for less than a year (M= 65.47, *SD*= 17.26). Figure 4.8 below represents the mean percentages of voice assignments by bilinguals, based on their length of stay in their L2 countries.

[22] All bilinguals lived or had been living only in the UK at the time of the study. Just one participant reported living in the United States for three years and the UK for four years.

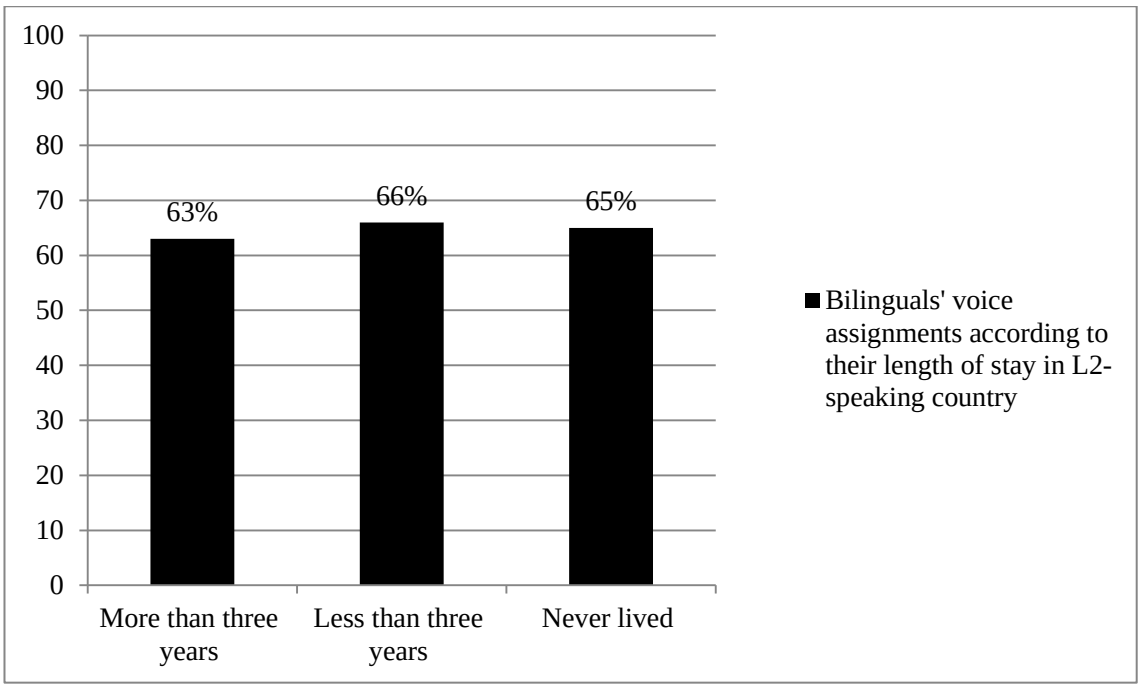


Figure 4.7 Mean percentages of bilinguals' voice assignments based on their length of stay in L2-speaking countries

There was no simple main effect of cultural immersion on the categorisation of objects in terms of the bilinguals' performance (*F1* (2, 240) = 548, $p < .579$). Furthermore, a correlation test was conducted on this variable in which cultural immersion and same-gender voice assignments were entered into a linear regression model. Cultural immersion was used as a predictor variable and same-gender voice assignments as an outcome variable. The regression was not statistically significant, F (1,239) = 473, $p < .493$, $R^2 = .045$. Consequently, the length of stay in any L2 country was not significant in affecting the bilinguals' performance, $b = -.581$, t = -687.

Language usage

Regarding the amount of language usage, this variable was treated in the same way as the other variables in terms of its correlation with the bilinguals' categorisations of voices. First, the bilinguals were separated into three groups according to how much time they were using each of their languages (question 6 in appendices 1A & 1B). Thirty seven bilinguals (M= 65.53, *SD* = 9.79) reported that they use Arabic substantially more than English, nineteen bilinguals (M= 62.18, *SD* = 10.71) reported that they use both languages equally often but tend to use Arabic slightly more than English and only four bilinguals (M= 65.62, *SD* = 5.15) reported that they use English substantially more than Arabic. There was no significant difference between those groups (F1 (2, 58) = .776, $p < .465$). From this information, it seems that bilinguals'

usage of L1 or L2 has no effect on their assignments of voices. A linear regression analysis also confirmed that there was no significant correlation between language usage and the bilinguals' performance, $R^2 = .045$, $b = - 922$, $t = -688$, $p < .492$. As the sample size of the groups was not equal, Levene's Test of Equality of variances was used[23]and it confirmed this results ($p = .346$).

Analysis of items

The way each particular test item was categorised by all participants was examined by using Chi-square tests[24]. In more detail, for each item the test compared the number of participants from each language group that categorised the item in accordance with the Arabic language relative to the number that would be expected by chance. The purpose of these tests was to examine (a) the degree to which each item was categorised according to chance, (b) whether English speakers, Arabic monolinguals and Arabic-English bilinguals agree on the gender of any items and - if so - which ones and finally (c) whether English speakers, Arabic monolinguals and Arabic-English bilinguals disagree on the gender of any items, and again - if so - which ones.

A number of findings were obtained from these item analyses. First, more than half of the items (31 out of 40) were assigned voices that were consistent with the Arabic grammatical gender by all groups. The total number of participants was one hundred and twenty, with each of these items being assigned a voice in accordance with Arabic grammatical gender by more than sixty of the participants. Only nine out of the forty items (*ball, mouth, table, car, heart, dress, ear, traffic light,* and *head*) were categorised at levels equal to chance by the four groups. Five items were grammatically feminine in Arabic and four were grammatically masculine. The feminine items *ball, car, traffic light, table,* were artificial items except for *ear* and were strongly assigned male voices by all groups. For these items, participants relied more on the artificial-masculine/natural-feminine distinction rather than on the Arabic gender system. The items, *ball, car, traffic light,* have cultural connotations for men - especially for Arabic speakers who were all Saudis living in a country where women are not allowed to drive cars and where female football teams do not exist. It seems that for items with strong cultural associations, participants' reliance on their language disappeared. As for the

[23] This test is used to test the assumption of homogeneity of variance; if the p value is greater than .05 then group variances can be treated as equal. If p is less than 0.05, however, we have unequal variances and have violated the assumption of homogeneity of variance.

[24] Chi-square tests investigate whether variations in data are due to chance or due to other variables, in our case the Arabic grammatical gender system.

four grammatically masculine items that were assigned female voices, half were natural (i.e. *mouth, head*) and the other half were artificial (*dress, heart*). For *mouth* and *head,* voice assignments were in line with feminine-natural distinction, but for *dress* and *heart*, again cultural effects were stronger on the speakers' perceptions of these items.

Unexpectedly, English speakers were consistent in their voice assignments to eleven of the test items. On six of the eleven items, they agreed with the Arabic groups on grammatically feminine items (*apple, sun, strawberry, basket, flower,* and *eye*) for which the Arabic gender almost always agrees with the artificial-male/natural-feminine distinction, except for the item *basket* which is artificial but the Arabic language categorises as 'female-like', as did all the four groups. The English group also agreed with Arabic groups on Arabic grammatically masculine items (*key, pencil, nose, mountain,* and *door*) for which the Arabic gender agrees with the artificial-male/natural-feminine distinction on the items, *key, pencil,* and *door* but not on *nose* and *mountain* which are naturally occurring kinds. Interestingly, these two items *mountain* and *nose* are grammatically masculine in Arabic and were assigned male voices at significantly higher rates than chance by the majority of participants, a hundred and nineteen for *mountain* and a hundred and ten for *nose,* out of one hundred and twenty.

These tests revealed that unlike Arabic monolinguals, voice assignments made by English speakers and Arabic-English bilinguals differed from chance and disagreed with the Arabic gender system on the items, *mouth* and *head* as they attributed female voices to these items which are grammatically masculine in Arabic. These items are naturally occurring, so the English group seemed to follow the artificial-male/natural-feminine distinction in their voice assignments to these items (n= 28 for *mouth*, and n= 29 for *head*, out of 30). Similarly, when these two organisations diverge, the voice assignments of the Arabic-English bilinguals to these items were directed by artificial-natural distinction more than Arabic grammatical gender (n= 24 for both *mouth* and *head*). We see that Arabic bilinguals did not follow grammatical gender for these two items but rather the artificial-male/natural-feminine distinction.

Participants' justifications

Following Kurinski and Sera (2011), participants were asked about their responses in order to better understand what they were thinking during the task and whether or not they used any strategies to perform it. After completion of the voice assignment task, they were asked how they assigned voices to inanimate pictures of objects. Twenty two of the English group indicated that they assigned voices based on their feelings and

personal experiences and eight reported that they were unaware of the basis for their assignments. The majority of Arabic monolinguals, twenty five of the thirty, reported that they associated voices according to Arabic grammatical gender. They stated that when the noun that refers to a picture has a feminine ending (*ah*), they assigned feminine voices to those items and when the nouns are grammatically masculine, they assigned masculine voices to those items (even though the pictures were not labelled). There were, however, five Arabic monolinguals who did not state their reliance on the Arabic gender system and said that they made their choices according to the pictures not to their noun referents. It should be noted, however, that their responses were mostly consistent with the Arabic gender system. With regard to the bilingual groups, they reported similar justifications; some associated some of the pictures with personified characters from children's cartoons. For instance, one mentioned that she assigned a masculine voice to the *bus* not because it is grammatically masculine in Arabic, but because she remembered the voice of an old man who acted as the voice of the bus in a cartoon called 'The Bus Driver'. In addition, three advanced bilinguals indicated that they assigned female voices to all plants and eatable items as cartoons reminded them of the soft feminine voice that acted the role of the flower and strawberry. One advanced participant stated that at some point he tried to remember the Arabic words for some pictures to assign the voice according the gender system but was unable to do so as he could not recall the Arabic word during the experiment, so he just did it randomly.

4.2.5 Discussion

The aim of this study is to determine whether learning a second language with a natural gender system affects bilinguals' categorisations of objects. More specifically it has asked whether Arabic-English bilinguals categorise objects differently from Arabic monolinguals as a result of learning English. The main finding was that the bilinguals' performance significantly differed from that of monolingual speakers of Arabic and English in that they behaved in between the two monolingual groups. It appears therefore that learning a second language without a grammatical gender system seems to change cognitive representations in bilinguals, since there was a significant difference between Arabic monolinguals and Arabic-English bilinguals. The current findings are in line with previous findings by Bassetti (2007; 2011) in relation to the effects of the Italian and German grammatical gender systems on the performance of bilingual children in terms of cognitive tasks. They also support Boroditsky's studies with

colleagues (Philips and Boroditsky, 2003; Boroditsky and Schmidt, 2000), who investigated the effects of grammatical gender with bi and multilingual speakers.

Conversely, however, these findings contradict those obtained by Kousta et al. (2008), who argue that learning a second language without a grammatical gender system does not seem to affect semantic representations in the gendered first language, because in their study monolingual and bilingual Italian participants did not differ in their performance during the tasks. One possibility for this contradiction might be that Kousta et al. (2008) used a purely linguistic task - speech error induction - and tested the bilingual speakers in two conditions: one in their first language (Italian) and the other in their second (English). This had the effect of the bilingual participants behaving like monolingual speakers of their first language when tested in their native language and like monolinguals of their second language when tested in that second language. Therefore, such results could be predicted according to Grosjean's (2001, 2011) idea of 'language mode' which states that bilinguals may behave differently depending on the particular social setting in which they are engaged.

Furthermore, these results are consistent with the idea that speakers of a language without grammatical gender such as English have a "folk theory" of gender which enables them to assign gender to objects that lack natural gender (Sera et al., 1994: 287). For speakers of gendered languages - Arabic in this study - grammatical gender creates a reference point within that speakers' folk theory of gender, which also makes consistent and non-arbitrary categorisations. Although this study showed an effect of change from learning another language, this change did not increase in the advanced bilinguals; they performed similarly to the intermediate group in the voice-attribution task. This finding is in line with the study by Kurinski and Sera (2011) which demonstrated that learning a second language could change learner cognition, yet this change does not increase with learner proficiency. It seems that even though bilingualism affects cognition, it does so to only a limited extent. Kousta et al. (2008) argue for a different interpretation of such a result; they stated that if bilinguals who acquired their second language after their first show evidence of change toward the monolingual norm of their second language, then their first language has a very limited effect on cognition. This was not the case in this study, however, because even though bilinguals differed in their categorisations from monolinguals of their L1, they did not perform exactly like monolinguals of their L2, but rather seemed to reconcile some elements from both languages, supporting Cook's (1991, 1994) multi-competence theory.

Another important finding was that all language groups, except Arabic monolinguals, made more grammatically consistent voice assignments to masculine artificial items and to feminine natural items, indicating that they followed natural-feminine/artificial-masculine distinction in their categorisations. Regardless of the existence of grammatical gender and bilingualism, there was a significant difference between voice assignments to natural-feminine vs. natural-masculine items as well as to artificial-masculine vs. artificial-feminine items; this is in complete agreement with the Mullen (1990) and Sera et al. (1994) results. Arabic monolinguals assigned same-gender voices to natural and artificial items more reliably according to Arabic grammatical gender compared with Arabic-English bilinguals who were less affected by this grammatical feature.

In addition, this study looked at the effect of early and late bilingualism on the bilinguals' categorisations of objects. The results revealed that although late bilinguals differed from early bilinguals - they assign voices slightly more consistent with the Arabic gender system than early bilinguals - the difference was not at all significant and did not affect the bilinguals' performance. This finding is out of step with Boroditsky's (2001) study which examined the conceptualisation of time between early and late Chinese-English bilinguals. That study showed the bilinguals' thinking with regard to time depended on how young they were when they started to acquire English and that older bilinguals followed a Chinese pattern of thinking more than younger bilinguals. A possible interpretation is that learning an L2 earlier in life might alter bilinguals' cognition toward the L2 in certain domains (e.g. time) rather than others (e.g. grammatical gender).

Length of stay in the bilinguals' L2 country was a variable of interest in this study. The current findings do not show any effect of length of stay on the bilinguals' categorisations and this enhances previous research by Athanasopoulos (2007) which showed that length of stay in the bilinguals' L2 country was not a significant predictor of restructuring in bilingual cognition in the domain of grammatical number. These findings did not, however, conform to those obtained by Cook et al. (2006) in which Japanese-English bilinguals who had lived in their L2 country for more than three years tended to behave like monolinguals of their L2. In fact, Cook et al. (ibid.) did not collect data from monolingual speakers of both languages; rather they compared their bilinguals' results to other findings obtained by Imai and Gentner (1997), so the stimuli and language of instructions were more likely to be different and that may have contributed to the observed effect of culture on bilinguals' performance. Alternatively,

target-language culture may play a role in bilingual cognition in some domains (e.g. grammatical number) rather than others (e.g. grammatical gender).

As length of stay in their L2 country does not seem to have a role in affecting the bilinguals' way of thinking about objects, this may rule out the possibility that the observed difference between monolinguals and bilinguals, in the domain of grammatical gender, were due to L2 cultural effects rather than language effects. In some cases, however, native culture might have a strong effect on cognition over language, as was the case in Carroll and Casagrande's (1958) study (mentioned in section 2.1.5), in which preferences of Navaho and English children were compared using triad pictures of objects and the results of their study contradicted the prediction that the Navaho language would direct speakers towards shape rather than English. Their results were considered to be attributable to the children's environment. Nevertheless, item analysis confirmed that the majority of the test items were mostly assigned voices according to Arabic gender. Monolingual and bilingual speakers were nearly systematic and followed the Arabic grammatical gender in their categorisations. A closer analysis of the items that were assigned voices in contradiction with Arabic gender revealed, however, that although native culture could play a role in affecting people's perceptions of the world, this role is very limited. It only affected categorisations of items that have a strong cultural relationship with men and women in society (e.g. car, ball and traffic light). As the number of items assigned in accordance with Arabic gender outweighed those assigned according to the speakers' culture, this indicated the limited role that a person's culture has on their thinking.

Heeding the call of some other studies (e.g. Bassetti, 2011; Athanasopoulos, 2011) for more research on the effects of other variables such as academic qualifications and language usage on the bilinguals' cognition, this study investigated the effect of these variables in relation to bilingualism. The findings showed that neither the bilinguals' academic qualifications, nor their language use played a part in the bilinguals' categorisation. These variables did not seem to affect the bilinguals' cognitive restructuring in the domain of grammatical gender. These findings also challenge previous research in the domain of spatial reasoning. Li and Gleitman (2002) pointed out that cognitive differences between Tenejapan and English speakers could be due to differences in education and environment rather than entirely language-related differences. Furthermore, Mazuka and Friedman (2000) doubted the results obtained by Lucy (1992), arguing that the cognitive differences between the two language groups could be because of differences in their educational backgrounds and lifestyles.

Nonetheless, these non-linguistic variables may be controlled and reduced to a minimum in bilingual research, as is the case in this study, where different proficiency levels of bilinguals were compared and the results showed cognitive differences.

To sum up, this study has avoided the use of language in the tasks by using only pictorial stimuli; it has also avoided some gender-related words (e.g. gender, feminine and masculine) in order to keep the participants' attention away from the aim of the task. The main drawback of the present experiment and other studies that used the voice attribution task is, however, that it cannot guarantee whether grammatical gender system affects cognition at a deep level of cognitive representation (that the observed effects are caused by the conceptual consequences of grammatical gender) or whether it is used by participants as a strategy for carrying out the task. Some researchers (e.g. Vigliocco et al., 2005; Vigliocco et al., 2002; Bowers et al., 1999) argue that in these tasks where participants are explicitly asked to link natural gender (human voices) and objects, it seems natural to use language as a clue to connect these two types of information. If the effects of grammatical gender reflect a deeper level, however, its effect should be found in a range of tasks. The next experiment therefore used a task that minimised the possibility that participants might use grammatical gender as a strategy to perform the task; it investigated the possible effects of the Arabic grammatical gender system in a similarity rating task. The aim of this experiment was to see if this grammatical category would have a strong effect on the speakers' ratings where there is no explicit reference to natural gender.

Chapter 5. Similarity Rating Task

5.1 Experiment 2A: The Effects of Grammatical Gender on Similarity Ratings by Monolingual Speakers of Arabic and English

5.1.1 Introduction

The voice attribution task used in the previous chapter revealed that Arabic monolingual speakers were significantly affected by the grammatical gender system of their language. This experiment investigated the similarity ratings of inanimate items made by monolingual speakers of Arabic and English. The aim was to determine whether the effects of grammatical gender observed in the voice assignment experiment would also appear in tasks where no overt reference to gender is made or required. This will lead to conclusions about the extent to which this grammatical property influences Arabic speakers' cognition.

Two opposing theories have been put forward by researchers with regard to the possible effects of grammatical gender on cognition. According to one view, grammatical gender may shape people's attitude towards things on an unconscious level (Jakobson, 1966) and that then leads to reorganisation at a representational level during language acquisition (Sera et al., 1994). Another view, put forward by Bowers et al. (1999), is that knowledge of grammatical gender may lead speakers to access it when they are required to perform certain tasks. The former view seems to propose deep permanent effects of grammatical gender on cognitive representation, whereas the latter sees it only as a strategy that can be adopted sometimes by speakers to help them accomplish a given task.

Some studies have suggested that grammatical gender affects semantic representations, e.g. items with feminine labels being perceived to have more feminine qualities (Konishi, 1993; Boroditsky et al., 2003). If these effects were pervasive in all grammatically gendered languages, however, then models of semantic representations would expect to show the effects of grammatical gender. Such effects could appear as a result of the activation of related meanings. Lucas (2000) asserted that related meanings share a number of the same semantic features; hence earlier activation of a given meaning facilitates the processing of another related meaning, as a result of the activation of features shared by the two meanings. Sera et al. (1994, 2002) argued that during the acquisition of grammatical gender in a language, masculine or feminine properties become related to representations of objects consistent with the grammatical

gender of their referents. If this is the case, however, then activating information about an object (e.g. watch-feminine), should activate the (feminine) gender features connected to it, combining and presenting both sources of information. For instance, a picture of a *watch* paired with a picture of a *girl*, should lead to greater activation in the cognitive system and probably make these pictures easier to remember. This can be compared with different types of pictures that do not share the same grammatical gender (e.g. a picture with a masculine label presented with feminine information, such as a picture of a *chair* and a *girl*).

This experiment required participants to think about possible relationships between objects by accessing their own conceptual and semantic representations and to elicit related features from the pairs of items in comparison. Using semantic similarity rating tasks was thought to be a way of studying whether semantic representations of objects are affected by grammatical gender among Arabic speakers.

5.1.2 Aims

The aim of this experiment was to study whether similarity ratings made by Arabic speakers would be affected by grammatical gender categorisations. Similarity-rating tasks have been used in previous research, e.g. Boroditsky et al. (2003), Degani (2007) and Ramos & Roberson (2010) (see sections 2.2.6 and 2.2.7). This experiment set out to investigate whether the grammatical gender system is part of - or closely related to - conceptual and semantic representations and whether Arabic speakers access and use grammatical gender information spontaneously in tasks that do not overtly refer to gender classes. Specifically it tried to answer the following question:

- Does the Arabic grammatical gender system affect Arabic speakers' similarity ratings and - if so - in what way?

Monolingual speakers of Arabic and English were asked to rate the similarity between pairs of pictures of both the same and different semantic groups. If the grammatical gender of nouns referring to objects leads speakers to extend the attribution of natural gender features to inanimate objects during language acquisition, Arabic speakers should show a preference towards grammatical gender. This would be indicated by increased similarity ratings for pairs of object nouns with the same grammatical gender when compared to English speakers. As in previous studies (e.g. Vigliocco et al., 2005) the performance of English speakers was taken as a baseline for comparison since in English nouns referring to objects are not assigned a gender. Their

ratings can therefore be considered to be free from any linguistic bias in classifying words into different grammatical categories.

5.1.3 Hypothesis

There were two hypotheses for this study:

- Arabic speakers will rate pairs that share the same grammatical gender as being more similar, whereas English speakers will not follow the same pattern.
- Speakers of both Arabic and English will rate pairs from the same semantic groups as more similar and pairs from different semantic groups as less similar, showing the effect of semantic homogeneity.

5.1.4 Methods

5.1.4.1 Participants

Thirty monolingual speakers of Arabic took part in the experiment. Individually recruited by the author from universities in Saudi Arabia[25] and ranging in age from eighteen to thirty five, (mean age =27; 11 were male and 19 female) they were all residents of the Saudi Arabian capital Riyadh, where English is not used for daily communication. Twelve were undergraduates, six postgraduates and twelve had recently graduated from the same universities. With regard to their academic qualifications, twelve had reached secondary school level, fifteen bachelor level and three had postgraduate degrees.

In addition, thirty monolingual speakers of English were recruited from different universities in the UK[26], ranging in age from eighteen to forty four (mean age = 26.9; 9 were male and 21 female). Ten reported that they knew a second language (7 stated French, 2 German and 1 Spanish) but that their proficiency levels in these languages were very low and they never use them for communication. The age at which they acquired their second languages varied with six participants at the age of nine and the remaining four between four and six years old. Eleven were undergraduates and nineteen postgraduates (11 had A-levels, 8 had a bachelor degree and 11 had a postgraduate degree).

For convenience, the experiment was conducted at the participants' own universities in quiet rooms where only the experimenter and one participant were

[25] Princess Norah and King Saud universities
[26] Newcastle, Northumbria and Durham universities

present during each session. All participants were assured that the data would be kept confidential and anonymous and only used for the purposes of the study. All participants participated in the experiment voluntarily.

5.1.4.2 Materials

Sixty pairs of inanimate items were taken from Snodgrass and Vanderwart (1980) and Szekely et al. (2004). The stimuli chosen came in the form of black and white drawings and drawn from five semantic groups; 'body parts', 'clothes', 'vehicles', 'food' and 'household items'. Half came from the same semantic groups and half from different semantic groups. For example, in the body parts group pairs were either congruent or incongruent in terms of grammatical gender (*nose-head* 'masculine' vs. *nose-hand* 'masculine-feminine'). In each semantic group, ten pairs were of the same gender: five grammatically feminine pairs and five grammatically masculine pairs (e.g. in the clothes group, *hat-skirt* was 'feminine' and *jacket-pants* 'masculine'. Twenty pairs carried different grammatical genders but shared the same semantic group (e.g. *jacket-hat* 'masculine vs. feminine'). Likewise with the other thirty pairs from different semantic groups, ten pairs shared the same grammatical gender (e.g. *basket-car*, are 'household-vehicle' and both are feminine). Twenty pairs carried a different gender (e.g. *mushroom-table*, are 'food-household items' and masculine- feminine, respectively. A full list of these items is provided in table 5.1, below. A sample picture is presented in figure 5.1 and a full list of these pictures is attached as appendices 10A & 10B.

Same grammatical gender and same semantic groups		
Grammatical Gender	**Pairs**	**Semantic Group**
Feminine	Table-Basket	Household
	Eye-Hand	Body parts
	Car-Airplane	Vehicle
	Hat-Skirt	Clothes
	Apple-Strawberry	Food
Masculine	Cup-Chair	Household
	Nose-Thumb	Body parts
	Train-Bus	Vehicle
	Jacket-Pants	Clothes
	Mushroom-Grapes	Food
Different grammatical gender and same semantic groups		
Grammatical Gender	**Pairs**	**Semantic groups**
Masculine /feminine	Nose-hand	Body parts
	Nose-eye	

Feminine/masculine	Hand- thumb	
	Eye-thumb	

Masculine /feminine	Chair-table	Household items
	Cup-basket	
Feminine/masculine	Table-cup	
	Basket-chair	

Masculine /feminine	Pants-skirt	Clothes
	Jacket-hat	
Feminine/masculine	Hat-pants	
	Skirt-jacket	

Masculine /feminine	Train- car	Vehicles
	Bus- airplane	
Feminine/masculine	Car-bus	
	Train-airplane	

Masculine /feminine	Grapes-apple	Food
	Mushroom-strawberry	
Feminine/masculine	Apple-mushroom	
	Strawberry-grapes	

Same grammatical gender and different semantic groups

Grammatical Gender	Pairs	Semantic Group
Feminine	Hat-Eye	Clothes-Body part
	Table-Skirt	Household-Clothes
	Hand-Strawberry	Body part-Food
	Basket-Car	Household- Vehicle
	Apple-Airplane	Food-Vehicle
Masculine	Train-Mushroom	Vehicle-Food
	Bus-Pants	Vehicle-Clothes
	Nose-Chair	Body part-Household
	Chair-Grapes	Household-Food
	Cup-Thumb	|Household-Body part

Different grammatical gender and different semantic groups

Grammatical Gender	Pairs	Semantic groups
Masculine /feminine	Thumb-strawberry	Body part-food
	Mushroom-table	Food-household
	Jacket-hand	Clothes-body part
	Grapes-hand	Food-body part
	Bus-skirt	vehicle- clothes
	Bus-basket	Vehicle-household
	Chair-hat	Household-clothes
	Grapes-airplane	Food-vehicle
	Pants-basket	clothes-household
	Nose-skirt	Body part-clothes

Feminine/masculine	Eye-mushroom	Body part-food
	Table-pants	Household-clothes
	Basket-grapes	Household-food
	Apple-nose	Food-body part
	Car-chair	Vehicle-household
	Table-train	Household-vehicle
	Eye-train	Body part-vehicle
	Airplane-cup	Vehicle- household
	Apple-bus	Food-vehicle
	Strawberry-jacket	Food- clothes

Table 5.1 Pairs of stimuli listed according to grammatical gender and semantic groups divisions

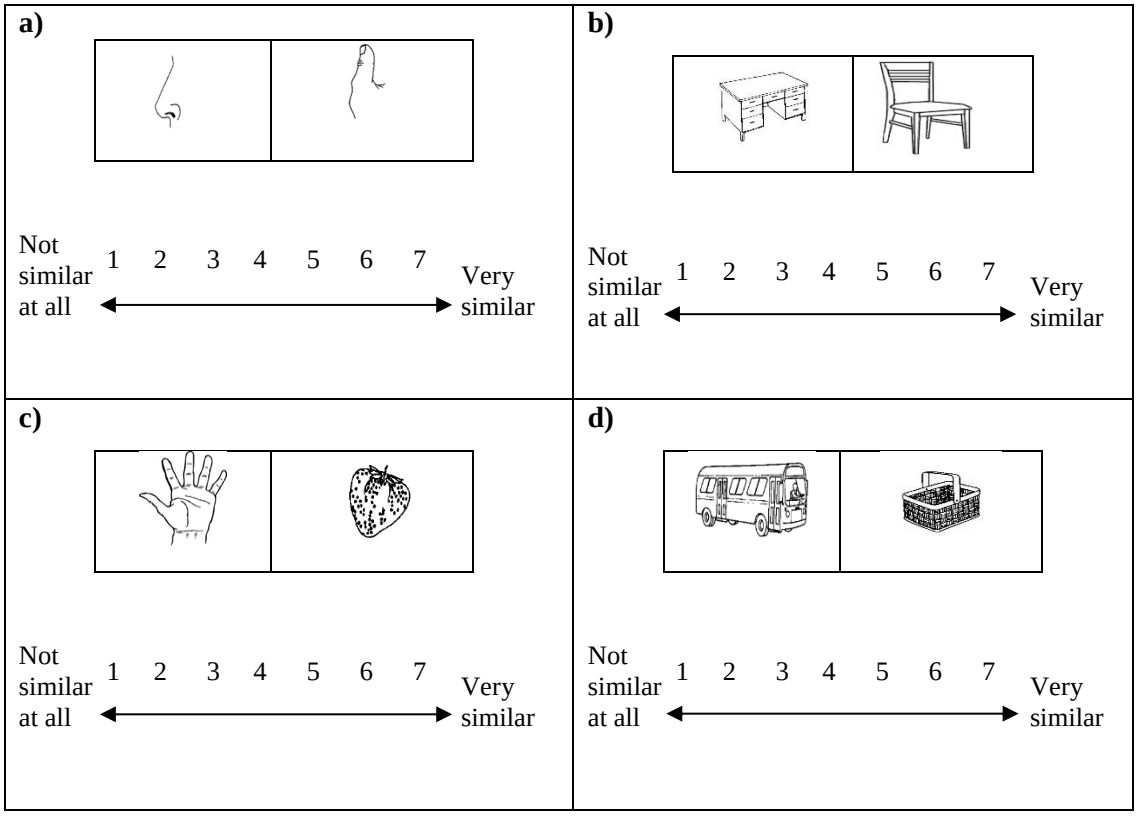


Figure 5.1 (a) pairs that share the same grammatical gender and same semantic group, (b) pairs from different grammatical genders but sharing the same sematic group, (c) pairs that share same grammatical gender but different sematic groups, and (d) pairs from different grammatical gender and different semantic groups

5.1.4.3 Procedure

The whole task lasted between twenty and thirty minutes and all participants were tested individually. Each session started by giving participants ethics forms (appendices 9A & 9B) and signing two consent forms; one for the researcher and the other for the participants to keep (appendices 5A & 5B). The task was then explained and participants were informed that their task was to rate the similarity between each pair of pictured objects on seven-point scales where (7) means very similar and (1) means not similar at all (see section 3.2.2, for the thinking behind the use of 7-point scales). The

instructions for the experiment were all given in the participants' native language; English with the English group and Arabic with the Arabic group. A practice session was performed before the actual task took place to make sure that the participants understood the task. Stimuli were all edited to fit the same sized frame, eliminating any potential size variable. They were presented one pair at a time using PowerPoint presentation software. Each pair of items was presented for five seconds, followed by a white screen for three seconds, then an asterisk appeared at the centre of the screen for one second, indicating that the next pair was about to be shown (see figure 5.2). It was confirmed in the pilot study that a five-second exposure to each pair was appropriate. The PC, which had a 17-inch screen with a resolution of 1600 X 900 pixels, was set approximately 50 cm in front of the participants. All the stimulus pairs were numbered from one to sixty and an answer sheet was provided for each participant containing sixty rating scales (see appendices 10A & 10B). The experiment was undertaken in the presence of the researcher who could deal immediately with any questions that participants might have. When the task was completed, the participants were asked to fill in the questionnaire (see appendices 1A & 1B), followed by some informal questions about how they decided on their ratings, giving an opportunity to report on any strategies or feelings they experienced during the task.

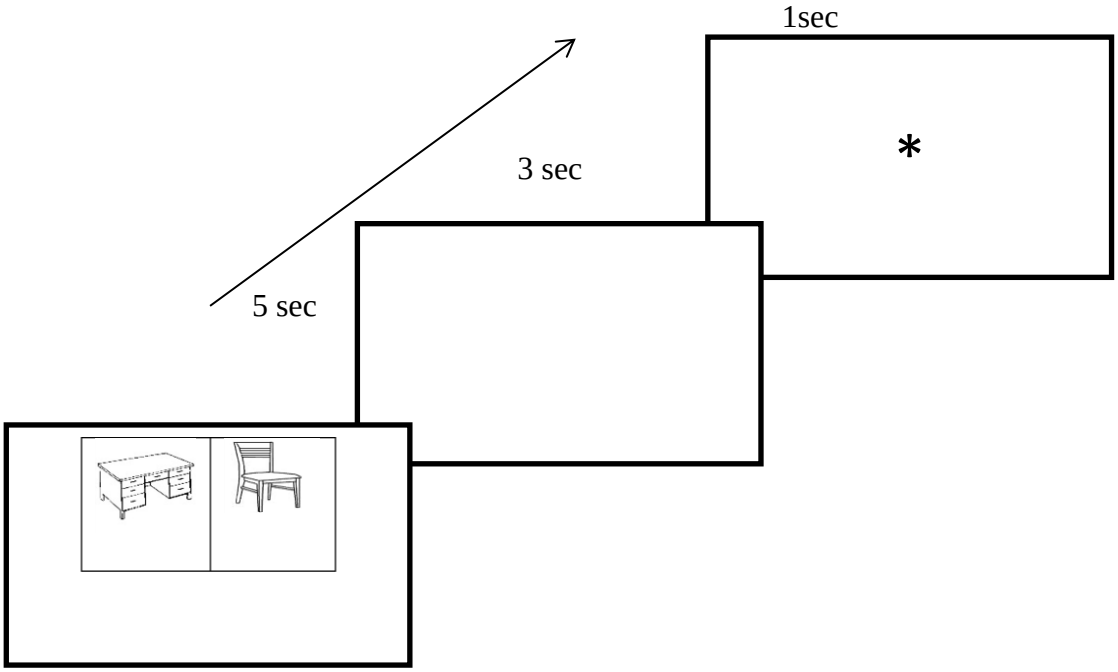


Figure 5.2 Illustration of how the pairs were presented using a Power Point presentation

The English instructions were:

> You will see pairs of pictures of different objects. Each pair will be presented for five seconds, it will be followed by a blank for three seconds and then an asterisk "*" will indicate that the next picture is coming up. Your task is to rate the similarity between the two pictures on the seven-point scales in your answer sheet. Please note that (7) means very similar and (1) means not similar at all. Please make sure that picture numbers correspond to the numbers on your answer sheet. Press the SPACE BAR to start the experiment.

The specific Arabic instructions were:

> سوف ترى/ين صورتين لشيئين مختلفين ستعرض كل صورتين لمدة ٥ ثواني، يتبعها صفحة فارغة لمدة ٣ ثواني، ثم علامة نجمة ستظهر لتدل على أن الصورتين التاليتين سوف تظهر. كل ما عليك أن تقيم/ين التشابه بين الصورتين المعروضتين على مقياس من سبع نقاط على ورقة اجابتك حيث أن ٧ يعني متشابه جدا و ١ تعني غير متشابه على الاطلاق. من فضلك تأكد بأن رقم الصور يتناسب مع الارقام في ورقة الاجابة. اضغط زرمسافة لتبدأ التجربة.

5.1.5 Results

Sample characteristics

There has been an on-going debate over how Likert scales should be analysed statistically (see Carifio & Perla, 2008 for a discussion). Murray (2013) provided empirical evidence that the type of statistical tests conducted on data from Likert scales do not affect the conclusion. In that study, Likert scale data obtained from 111 participants was analysed using parametric and non-parametric tests and obtained similar conclusions. For this reason, a Shapiro-Wilk's test and a visual inspection of their histograms and normal Q-Q plots showed that the mean ratings were approximately normally distributed for both language groups (see appendix 11 for more data), with a Shapiro-Wilk (p= .413), a skewness of -.341 (SE= .427) and a kurtosis of .726 (SE= .833) for English speakers and a Shapiro-Wilk (p= .607), a skewness of .131 (SE= .427) and a kurtosis of -.045 (SE= .833) for Arabic speakers. Therefore, as the data appeared to be approximately normally distributed, they were analysed using parametric tests, following Carifio and Perla (2008: 1151), who stated that it is "perfectly appropriate to summarise the ratings generated from Likert scales using means and standard deviations, and it is perfectly appropriate to use parametric techniques like Analysis of Variance to analyse Likert scales".

Similarity ratings analysis

For each participant, a mean rating was calculated from their ratings of the sixty pairs on seven-point scales. Ratings by the English group were taken as a baseline for comparing Arabic speakers' ratings according to Arabic grammatical gender. Higher ratings reflect a higher similarity between the pairs.

Effects of group **–** a three-way ANOVA was used to compare the ratings of English and Arabic monolingual speakers. See figure 5.3 and table 5.2 below. Both groups rated all pairs with the mean (*SD*) = 3.51 (.67) and they did not significantly differ from each other [English speakers: mean (*SD*) = 3.61(.65); Arabic speakers: mean (*SD*) = 3.42 (.70), [*F* (1, 59) = 1.254, *p* = .267].

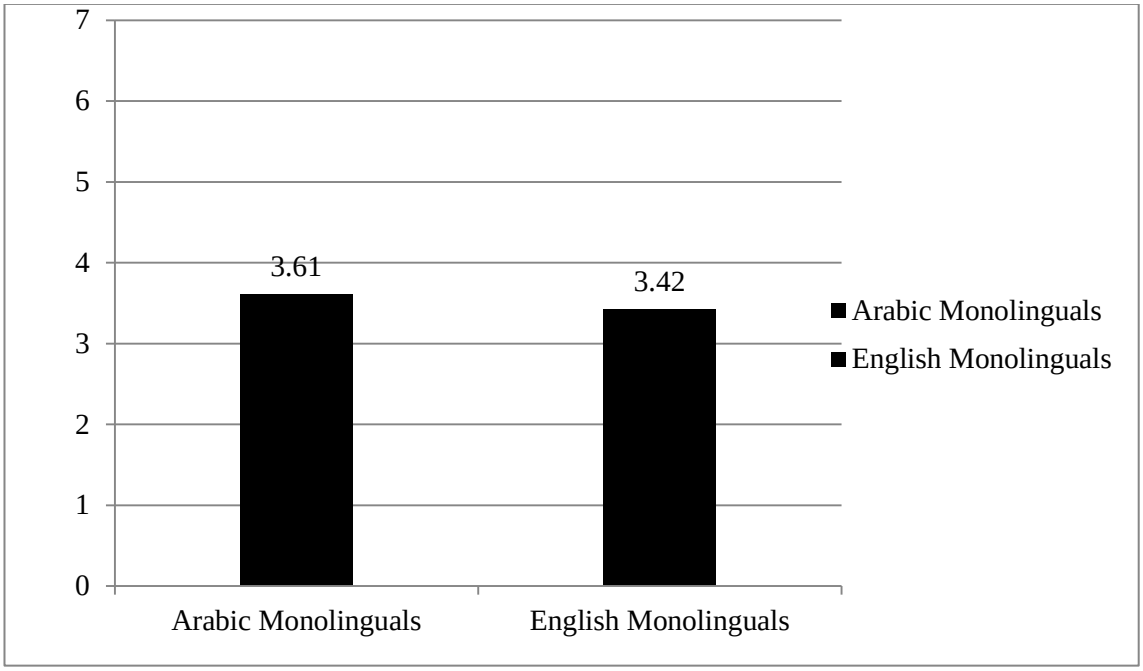


Figure 5.3 Arabic and English speakers' ratings to all pairs

Language groups	Mean of all pairs ratings
Arabic monolinguals	3.42 (.70)
English monolinguals	3.61 (0.65)

Table 5.2 Mean (standard deviations in brackets) of the participants' ratings to all pairs

Furthermore, all pairs were divided into two groups on the basis of their grammatical gender consistency or inconsistency. Participants' ratings to pairs that share the same grammatical gender were calculated separately from ratings to pairs of a different grammatical gender. The means indicated that the two language groups [English: mean

(*SD*) = 3.48 (.65), Arabic: mean (*SD*) = 3.40 (.76)] rated pairs of the same grammatical gender pairs similarly [mean (*SD*) = 3.44 (.70), *p* = .645]. A similar pattern was observed for pairs of different grammatical gender [English: mean (*SD*) = 3.68 (.66), Arabic: mean (*SD*) = 3.42 (.69), overall mean (*SD*) = 3.55 (.68), *p* = .158]. Figure 5.3, below, shows that the Arabic speakers did not rate pairs of same grammatical gender as more similar in comparison to the English speakers' ratings, nor did they give lower ratings to pairs of different grammatical gender. Their overall ratings did not, therefore, diverge from those of the English speakers.

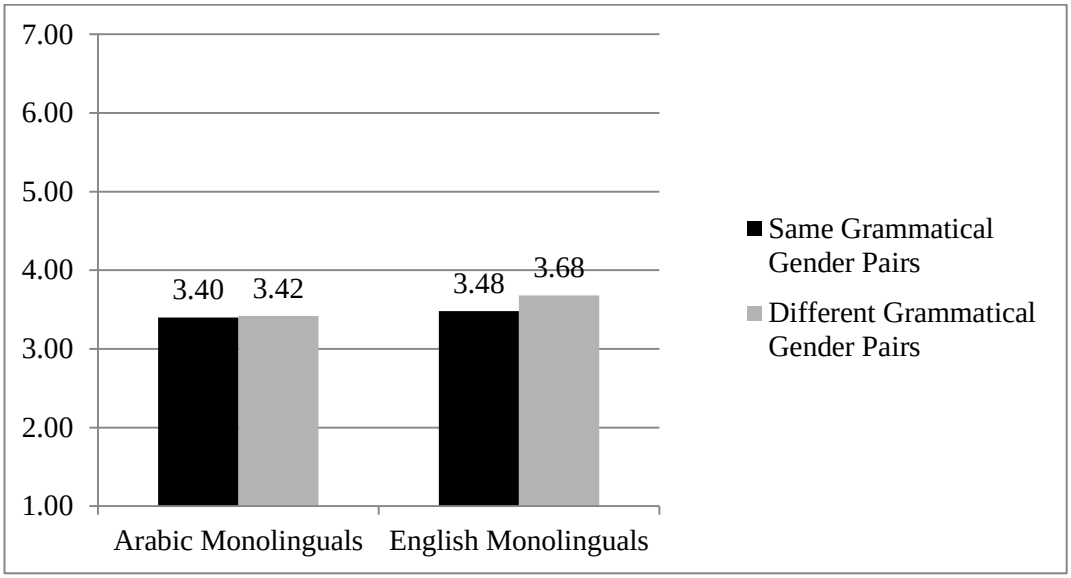


Figure 5.4 Groups' rating for same and different grammatical gender pairs

Effects of Type of Pairs (grammatical gender) - All same grammatical gender pairs were divided into two categories: (a) same grammatical gender and same semantic group pairs and (b) same grammatical gender but different semantic group pairs. The aim of this division was to see if Arabic and English speakers' ratings would differ when the pairs were from the same or different semantic groups. The mean of all pairs of the same grammatical gender by the two language groups was calculated [mean (*SD*) = 3.44 (.70)], for pairs of the same grammatical gender and same semantic group [mean (*sd*) = 4.63 (.96) and for pairs of the same grammatical gender but a different semantic group [mean (*SD*) = 2.25 (1.09)]. See table 5.2 below.

Language groups	Same Gender Same Semantic pairs	Same Gender Different Semantic pairs
Arabic (*N*= 30)	4.54 (1.10)	2.25 (0.94)
English (*N*= 30)	4.72 (0.81)	2.25 (1.23)

Table 5.3 Means of participants' ratings (standard deviation in brackets) of pairs of same grammatical gender and same semantic groups and pairs of same grammatical gender and different semantic groups

This table shows that Arabic and English monolingual speakers did not differ in their ratings relating to pairs sharing the same grammatical gender. This means that both groups gave higher ratings to pairs of the same grammatical gender which were from the same semantic groups than to pairs of the same grammatical gender but were from different semantic groups. There was no significant difference between ratings of the two groups to pairs of the same grammatical gender and the same semantic groups (p (2-tailed) = .484), nor to pairs of the same grammatical gender but different semantic groups (p (2-tailed) = .981). Although the two groups rated pairs of the same grammatical gender and same semantic groups as more similar than pairs of the same gender but different semantic groups, the correlation between these same gender same semantic pairs and same gender different semantic pairs was not statistically significant (r = -.060, p = .647). This means that although ratings of the two types of same grammatical gender pairs differ in the two groups according to semantic group, the difference was not significant. When pairs shared the same grammatical gender, the performance of the English and Arabic groups did not differ no matter how congruent or incongruent they were in terms of semantic relatedness. This result did not support the study hypothesis that speakers of Arabic and English differ in their ratings with regard to pairs that share the same grammatical gender in Arabic.

A similar procedure was followed with pairs of a different grammatical gender, they were divided into two categories according to their semantic groups: (a) different grammatical gender but same semantic group pairs and (b) different grammatical gender and different semantic group pairs. The intention was to examine whether the two groups' ratings of different grammatical gender would differ if they were from the same or different semantic groups. Table 5.3 below shows that the Arabic speakers rated pairs that were of different grammatical gender - but the same semantic group -slightly lower than the English speakers. This difference, however, did not reach a level of significance (p (2-tailed) = .07). Both language groups rated pairs of different gender

and different semantic groups as equally similar, so the difference between their ratings was non-significant (*p* (2-tailed) = .970).

Language groups	Different Gender Same Semantic pairs	Different Gender Different Semantic pairs
Arabic (*N*= 30)	4.52 (1.20)	2.33 (0.88)
English (*N*= 30)	5.04 (0.93)	2.32 (1.16)

Table 5.4 Means of participant ratings (standard deviations in brackets) of pairs of different grammatical gender but same semantic groups and pairs of different grammatical gender and different semantic groups

Effect of semantic groups **-** Since the Arabic grammatical gender system did not play a part in the speakers' similarity ratings, it was of interest to examine whether ratings given by the two language groups would be affected by semantic relationships. All pairs were therefore divided into two groups, but this time the division was according to their semantic groups, essentially pairs from the same semantic groups vs. pairs from different semantic groups. The means of same and different semantic groups pairs were then calculated for both groups [same semantic: mean (*SD*) = 4.53 (1.14), different semantic: mean (*SD*) = 2.30 (.88)]. The English speakers rated same semantic pairs as more similar [mean (*SD*) = 4.93 (.88)] which were slightly higher than the Arabic speakers' ratings [mean (*SD*) = 4.53 (1.14)]; the difference between the two language groups was not, however, statistically significant (*p* (2-tailed) = .135). Likewise, both the Arabic and English speakers' ratings of pairs from different semantic groups were quite similar and did not diverge from each other [English group: mean (*SD*) = 2.29 (1.17), Arabic group: mean (*SD*) = 2.30 (.88), *p* (2-tailed) = .974].

The participants' ratings were further analysed using a one-sample test with the aim of detecting any significant differences between the ratings (of the same and different semantic pairs) within each language group. Arabic monolinguals' ratings to same semantic group pairs significantly differed from their ratings to pairs from different semantic groups. In other words, they gave higher ratings with regard to similarities to pairs that share the same semantic groups and lower ratings to pairs that did not come from the same semantic groups [mean (*SD*) = 4.53(1.14) vs. mean (*SD*) = 2.30 (.88), respectively) and the difference was statistically significant (*t* (29) = 21.74, *p* = .00), the effect size (Cohen's d) was large *d*= 2.18. Likewise, the English speakers' ratings followed the same pattern, they gave higher ratings to pairs from the same semantic groups [mean (*SD*) = 4.93(.88)] and lower ratings to pairs from different semantic groups [mean (*SD*) = 2.29 (1.17)]. The difference was also significant (*t* (29) =

30.46, p = .00), the effect size (Cohen's d) was large d= 2. 55. These tests revealed that semantic groups had a significant effect on the participants' ratings of pairs regardless of their language background (see figure 5.4 below).

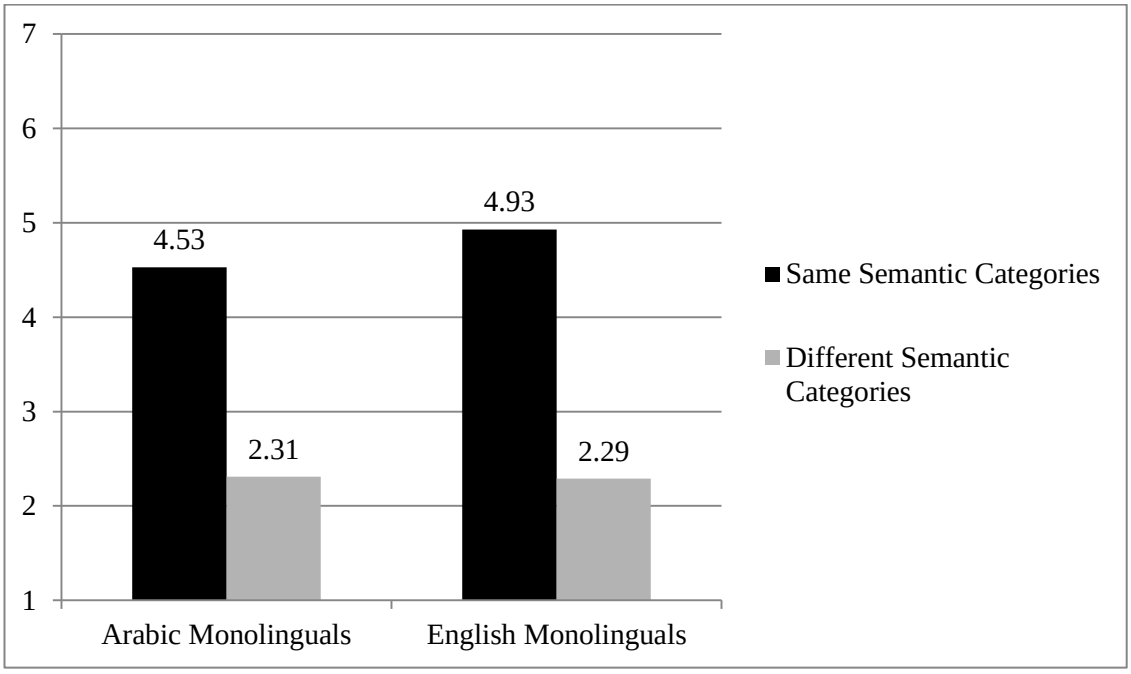


Figure 5.5 Participants' ratings to pairs according to semantic groups

Test of individual pairs - Finally, the mean of each particular pair was examined in order to understand the ways in which they were rated by the speakers of each language. The aim of this analysis was to find out: (a) which pairs were perceived as most similar by Arabic and English speakers? (b) Which pairs were perceived as dissimilar by the two groups? (c) Did the Arabic grammatical gender system play a role in Arabic speakers' ratings for some particular pairs? If yes, what are they? Finally, (d) did the semantic relationships of some pairs affect the speakers' ratings? If yes, what are they? It should be noted that an analysis for each language group was performed separately, with very similar findings being discovered (see appendix 12, for a full list of statistical analysis of the ratings of all pairs); this section presents the data for the two groups together.

A number of findings emerged from these analyses. Arabic and English speakers did not rate any pairs as *very similar* (none of the pairs were rated 7 on the scales). Such a result was expected as most people tend to avoid extreme response categories or endpoints. Both language groups gave a maximum similarity ratings of six for only two pairs that were perceived as similar (train-bus, mean (*SD*) = 6.11 (1.07), and apple-strawberry, mean (*SD*) = 6.01 (1.30). These two pairs shared the same grammatical

gender in Arabic (train-bus are both grammatically masculine and apple-strawberry are both grammatically feminine), as well as each pair being from the same semantic groups (vehicles and food, respectively). As these pairs shared the same grammatical gender and same semantic groups, it was difficult to determine which categories (grammatical vs. semantic) the participants followed in their ratings. Therefore, pairs that were rated quite similar (5 on the scales) by both language groups were analysed. Only ten pairs out of sixty were perceived to be quite similar by speakers of both Arabic and English [mean (*SD*) = 5.31(1.74)]. All of these pairs were from the same semantic groups with nine of them were of different grammatical gender in Arabic and only one pair being of the same grammatical gender (jacket-pants, both clothes and grammatically masculine). A total of ten pairs were rated as extremely dissimilar by the two groups [mean (*SD*) = 1.84(1.50)]. All of these pairs received the lowest ratings and were from different semantic groups, half of them (5 pairs) were of the same grammatical gender in Arabic and the other half were of a different grammatical gender. These results showed that grammatically consistent pairs did not receive higher ratings, particularly if they were from different semantic groups. Therefore we can say that this grammatical feature did not have any effect on the speakers' ratings, even when pairs were tested separately.

5.1.6 Discussion

The overall similarity between Arabic and English speakers' ratings indicated that ratings by the two groups did not differ from each other, suggesting that they neither increased with pairs of same gender, nor decreased with pairs of different gender. These results are consistent with Degani (2007) and Ramos & Roberson (2010) who studied the effects of grammatical gender using similarity rating tasks - similar to the one used in the present study - and did not find any difference between speakers of Spanish and English (Degani, 2007), or speakers of English and Portuguese (Ramos & Roberson, 2010). Through the use of similarity rating tasks - with only pictorial stimuli - to investigate the possible effects of grammatical gender on speakers' ratings, most research has not reported any effect.

Furthermore, these results are in line with early evidence on the limited effects of grammatical gender by Hafstatter (1963) who studied both German and Italian speakers. Hafstatter assumed different behaviours between the two groups based on the grammatical differences in their gender systems. Southern Europeans, for example, would perceive the sun as "powerful, but also threatening", whereas northern Europeans would perceive the sun as a "comfortably warm, mother-like womanly sun". He tested

this idea using a semantic differential test with twenty four bipolar adjectives, and concluded that neither grammatical gender nor geographical differences (northern Germany vs. Palermo, Italy) had an effect on the participants' ratings. Using similar scales, Mills (1986) tested English and German speakers on six nouns with animate referents and four with inanimate referents. The findings revealed that grammatical gender does not have an effect on ratings made by people. Likewise, the present findings speak against the idea that speakers of grammatically gendered languages perceive inanimate items that share the same grammatical gender as more similar and then attribute natural gender properties (masculine and feminine features) to these items during language development.

Gentner and Markman (1994) noted that the process of determining the similarity of a pair of items is central to various mental processes and that a pair's similarity increases with its commonalities (the elements of the matching representational structure) and decreases with its differences. Asking participants to quantitatively judge the similarities between two items was, however, not an easy task as most participants were unsure about properties that can be appropriate to make accurate ratings. Participants were, therefore, advised to use the whole range of values on the provided scales and to use whatever dimension to assess similarities between the pairs. The main findings showed that Arabic grammatical gender was not a useful property in assessing similarities between two items of the same grammatical gender, meaning that these types of pair were rated equally similar by both Arabic and English speakers. All participants mentioned that they performed the task without thinking of grammatical gender.

We should also consider two studies, by Clarke et al. (1981) and Boroditsky et al. (2003), which showed the effects of gender on participant ratings. The inconsistent conclusions reached by these two studies and other research - including the present study - might be due to the different types of stimuli used in the experiments. For example, Clarke et al. (1981) only used words and explicitly asked participants to assess the words on masculine-feminine scales. In such cases, it seems natural for speakers of gendered languages to use gender information in order to construct grammatically correct sentences, thereby showing linguistic knowledge rather than cognitive influence of this grammatical feature. A possible interpretation of their findings might be found in the *Similarity and Gender Hypothesis,* mentioned earlier (in section 2.2.4), which assumes that words that share a gender will be perceived as more similar in meaning and thus would behave as semantically related by virtue of their shared linguistic

contexts. With regard to Boroditsky et al. (2003), they used pictures of people and animals which may allow the use of grammatical gender information to match the biological gender of the stimuli.

To sum up, this property does not seem to have a pervasive role in conceptual representation. The effects of grammatical gender on speakers' categorisations has therefore, appeared only in tasks that cannot be completed without accessing this knowledge. For example in voice and name attribution tasks, participants are likely to use any available knowledge in order to better give accurate responses and grammatical gender could be one of them. The similarity rating task, however, prevented participants from drawing on grammatical gender information in a strategic manner, meaning that no effect was found. The discrepancy between these findings and those obtained from the voice-attribution task indicates that the effects of grammatical gender do not arise at a deep conceptual level; rather the effect is task-dependent. This explanation is consistent with Bowers et al. (1999) and Gennari et al. (2002) who are proponents of the strategic use of grammatical gender in certain tasks, suggesting that grammatical gender does not affect people's cognition at a deep conceptual level and that such effects are not apparent in all circumstances. Martinez & Shatz (1996) pointed out that only six (out of 18) Spanish speaking children sorted pictures of people and objects based on the grammatical gender of their language and that small number did not differ from being pure chance.

After verifying that both monolingual speakers of Arabic and English behaved in similar ways in the rating task, it was of interest to see what the bilinguals who speak these two languages would think about similarities between these pairs. Therefore, the aim of the next section (5.2) is to study how bilinguals would rate the pairs in question and whether or not they differ from their monolingual counterparts, taking into account a number of factors that might affect their cognition.

5.2 Experiment 2B: The Effects of Grammatical Gender on Similarity Ratings of Arabic-English Bilinguals

5.2.1 Aims

This experiment further examined whether grammatical gender can affect the ratings of similarities among bilingual speakers. The same similarity rating task used with the monolinguals was utilised to discover whether similar results would be obtained. Arabic-English bilinguals were expected to rate these pairs in a similar way to monolingual speakers of Arabic and English, supporting the view that grammatical gender has a very limited effect on peoples' ratings. Alternatively, if they developed a new way of thinking that is different from both monolingual groups, this would indicate that learning another language produced an effect regardless of its grammatical system. Therefore, the main question to be answered through this experiment is:

- Does bilingualism lead to a change in any specific area of Arabic speakers' cognition and as a consequence would they think differently from their monolingual counterparts? If this is the case, in what way do they differ?

5.2.2 Methods

5.2.2.1 Participants

Thirty Arabic-English bilinguals[27] took part in the similarity rating task. Because the voice-attribution experiment (chapter 4) did not show a significant difference between intermediate and advanced bilinguals, this experiment only included advanced bilinguals. All participants were students at Newcastle University in the United Kingdom - three were recent graduates from this university - and were living in the UK at the time of the study. Two bilingual participants were substituted because their test scores on the Oxford Placement Test (2001) fell into the low intermediate level (one scored 35 and the other 37). All participants voluntarily participated in the experiment and were recruited by personal contacts of the author. Table 5.4 below presents their profiles in detail.

[27] To rule out any dialectical effects, only Saudi participants were included.

Bilinguals' profiles	
Age	30. 46 (4.11), age range, 25-45
Gender	17 male 13 female
Languages (Proficiency)	30 speakers of Arabic (native) 30 speakers of English as their L2 (advanced) 1 know French as L3 (beginner)
Age of L2 acquisition	12 Early bilinguals 18 Late bilinguals
Length of stay in L2 country	19, more than 3 years 11, between 1-3 years
Academic qualifications	16 postgraduates 14 undergraduates
Test score	50.7 (2.30)

Table 5.5 Profiles of the Arabic-English bilinguals

5.2.2.2 Materials

The same materials were used as in Experiment 2A. Sixty pairs of inanimate items were presented along with seven-point scales where (1) means not similar at all and (7) means very similar. Participants were told they were taking part in an experiment seeking to analyse how people rate similarities between two inanimate items and were advised that there was no right or wrong answer to the task. Although all participants were students at a UK university, which means that they all passed an IELTS test prior to their studies with a minimum score of 6.5, they were asked to do the Oxford Placement Test (2001) to confirm that they all had similar levels of English at the time of the experiment. The test mean score for all participants was 50.7 (2.30), and two were substituted as their scores did not reach the advanced level.

5.2.2.3 Procedure

Procedures followed the same practices as Experiment 2A, with Arabic as the language of instruction for the experiment.

5.2.3 Results

The mean proportion of all the pair ratings (standard deviation in brackets) was calculated for Arabic-English bilinguals, 3.30 (.78). They gave slightly lower ratings to all pairs, but they did not significantly differ from their monolingual counterparts (F (89) = 1.44, p = .242). Figure 5.5 shows a subtle difference across the three groups,

monolingual speakers of both Arabic and English and the Arabic-English bilinguals. Table 5.6 shows the mean and standard deviations of the pairs' ratings by all the groups.

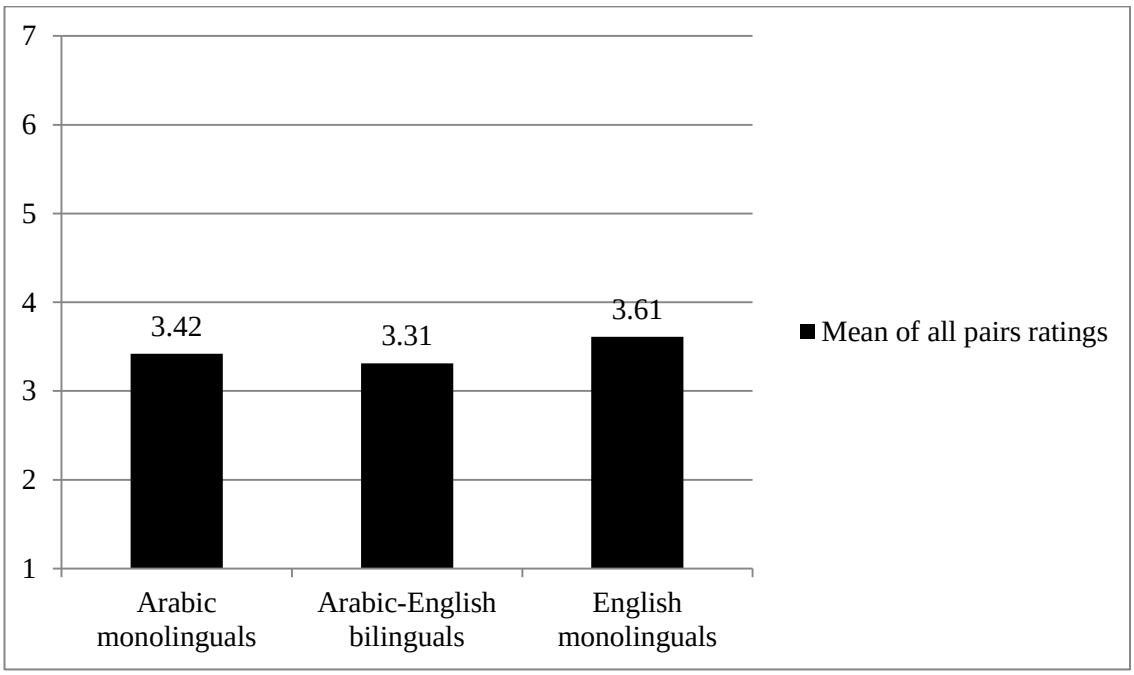


Figure 5.6 Mean of all pairs ratings by monolingual and bilingual groups

Language groups	Mean of all pairs ratings
Arabic monolinguals	3.42 (.70)
English monolinguals	3.61 (0.65)
Arabic-English bilinguals	3.30 (.78)

Table 5.6 Mean (standard deviations in brackets) of the pairs' ratings by all the groups

Effects of Type of Pairs (grammatical gender) - All pairs were divided into two groups on the basis of their grammatical gender consistency or inconsistency. Bilinguals' ratings of pairs that share the same grammatical gender were calculated separately from their rating of pairs of a different grammatical gender. The means indicated that Arabic-English bilinguals did not give higher ratings to same gender pairs [mean (*SD*) = 3.20 (.76)], or lower ratings to different gender pairs [mean (*SD*) = 3.35 (.81). The results revealed similar patterns to those found with the two monolingual groups of Arabic and English speakers. A one-way ANOVA was conducted to compare the bilinguals' ratings of same and different gender pairs against those of the monolingual groups. The bilinguals' overall rating was used as the dependent variable. Again, no significant difference was found between the three groups' ratings of same gender pairs (*F1* (89) =

1.17, p = .314) or to different gender pairs ($F1$ (89) = 1.64, p = .199). As the one-way ANOVA test cannot tell us whether some specific groups differ from others, post-hoc comparisons - using the Tukey adjustment - were conducted to examine whether any of these groups would differ from each other. This test revealed that all the groups' ratings of same gender pairs (p = .299), as well as those of different gender pairs (p = .199) did not reach significance. See figure 5.6 below.

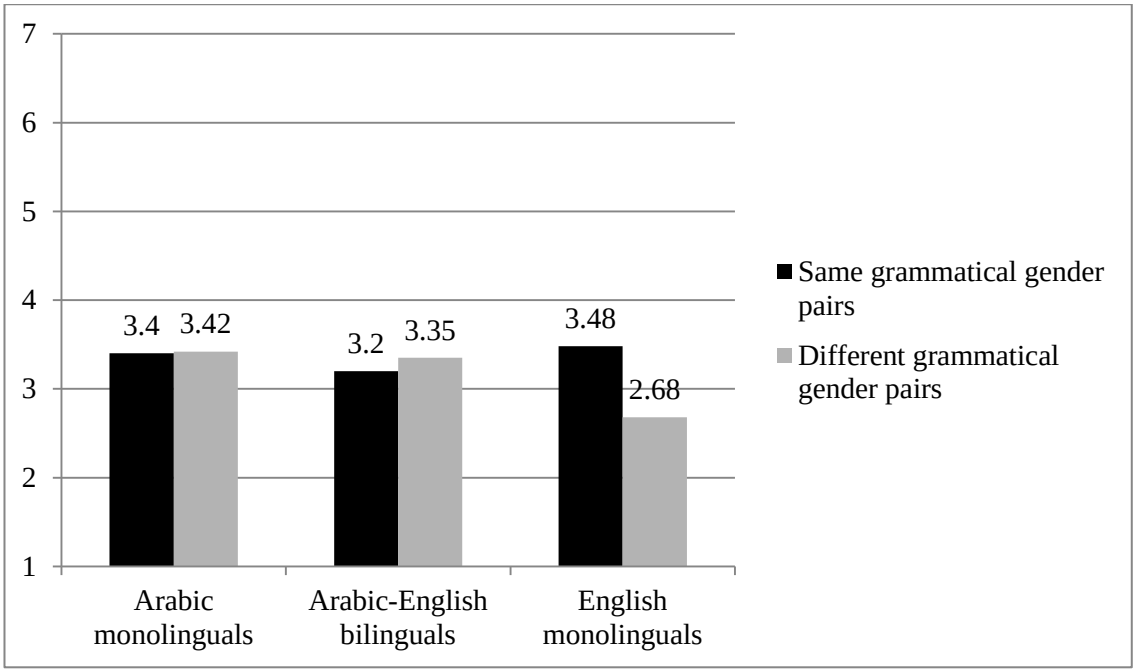


Figure 5.7 Ratings of same grammatical gender pairs vs. different grammatical gender pairs by monolingual and bilingual groups

Effect of semantic groups - All the pairs were divided into two groups according to their semantic relationships, that is pairs from the same semantic groups vs. pairs from different semantic groups. It was of interest to examine whether the bilinguals' ratings would be affected by semantic relationships as was the case with the monolingual groups and whether or not they differ from monolinguals. The means of both the same and different semantic group pairs were calculated. Bilinguals rated the same semantic pairs between the two monolingual groups [mean (SD) = 4.64 (1.09)], but their ratings to different semantic pairs were slightly lower than the two monolingual groups [mean (SD) = 1.96 (1.09)], the effect size was large (Cohen's d= 2.55), this indicated a big difference between the bilinguals' ratings to same and different semantic group pairs. Although there were slight differences across the monolingual groups and the bilingual group, none of them was statistically significant ($F1$ (89) = 1.16, p = .317, for same semantic pairs and $F1$ (89) = 1.01, p = .368 for different semantic pairs). A paired sample t-test showed the difference between the bilinguals' ratings to pairs from same

and different semantic groups to be significant (t (29) = 23.20, p = .00). This indicated that, like the two monolingual groups, Arabic-English bilinguals were significantly affected by the semantic relationships between the pairs and thus they gave higher ratings of similarities to pairs from same semantic groups (e.g. both pairs were household items) and lower ratings to pairs from different semantic groups (e.g. food vs. vehicle). See figure 5.7 below.

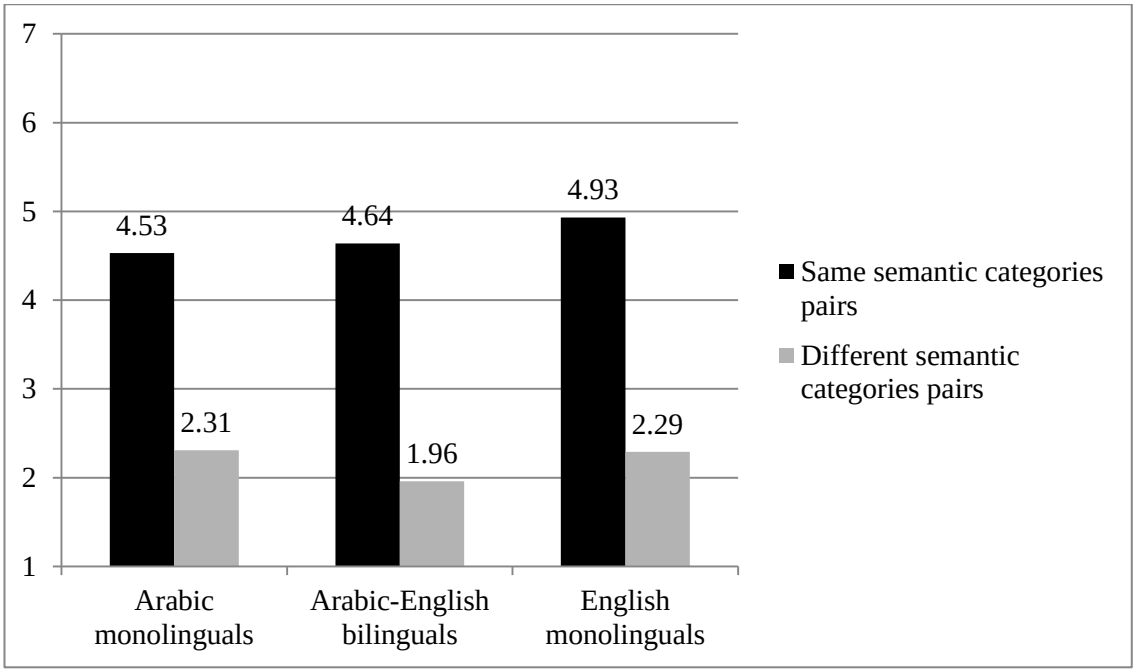


Figure 5.8 Mean ratings of same semantic groups vs. different semantic groups by monolingual and bilingual groups

Test of individual pairs - As in Experiment 2A, the mean of each particular pair was examined in order to understand the ways in which they were rated by Arabic-English bilinguals. This examined which pairs were perceived to be very similar and which were perceived to be very dissimilar by the bilingual group. Similar findings to those obtained by the monolingual groups were observed. The bilingual group seemed reluctant to give higher ratings of similarities, they did not rate any pairs as *very similar* (none of the pairs rated 7 on their scales); their maximum higher ratings were six for two pairs which both happened to be from the same semantic groups (strawberry-grapes and apple-grapes, both of which are food) and of different grammatical gender (feminine vs. masculine), mean (*SD*) = 6.06 (1.17). See appendix 12 for a full list of all pair ratings. This indicated that Arabic grammatical gender did not have an effect on the bilinguals' ratings, whereas semantic relationships between the pairs played a significant role in their ratings.

As only two pairs received ratings higher than six we may not, however, be certain from such a small number that the main factor that influenced the bilinguals' ratings was semantic relationships. For this reason, pairs that were rated quite similar (5 on the scales) were examined. Ten pairs out of sixty were rated quite similar by the bilingual group [mean (*SD*) = 5.46(1.70)]. All of these pairs were from the same semantic groups and seven were of different grammatical gender in Arabic, only three pairs were of same grammatical gender. Turning to the bilinguals, they were more open to providing very low ratings of the pairs. That was clear when they rated seventeen pairs as not similar at all [mean (*SD*) = 1.71(1.38)], all of them from different semantic groups, eleven were of a different grammatical gender and six were of same grammatical gender. These results show that the same grammatical gender pairs mostly did not receive higher ratings if they were from different semantic groups. This grammatical feature again did not have any effect on the speakers' ratings even when the pairs were tested separately.

Of further interest was an investigation into whether other variables affected the bilinguals' ratings in this experiment, therefore separate independent sample t-tests were used to seek any such possibilities.

Effects of Age **-** Bilinguals were divided into two groups based on their ages, a younger group aged twenty nine or under and an older group of thirty or over. The mean ratings were compared using an independent sample t-test. Analysis showed that both age groups provided very similar ratings (younger group (*N*= 13): mean (*SD*) = 3.39 (.66), older group (*N*= 17): mean (*SD*) = 3.23 (.88)]. The slight difference in the mean ratings was not at all significant in affecting the bilinguals' similarity ratings (t (28) = .515, p = .610). A correlation test was conducted on this variable, the age of participants and overall similarity ratings were entered into a linear regression model. This was completed with the participants' age as a predictor variable and all similarity ratings as an outcome variable. The regression was not statistically significant [F (1, 29) = 266, p < .610, R^2 = .097. Consequently, we can conclude that age was not significant in affecting participant performance, b = -.152, t = -.515].

Furthermore, bilinguals were again divided into two groups based on the age at which they learned or acquired their L2. Early bilinguals acquired L2 proficiency before the age of seven and late bilinguals acquired their L2 proficiency after the age of seven. An independent sample t-test was used to detect any differences between the means of the two groups. The results showed that early bilinguals [(*N*= 12), mean (*SD*) = 3.32

(.84)] did not rate the pairs differently from older bilinguals [(*N*= 18), mean (*SD*) = 3.28 (.77)], meaning that there was no statistical significance between them (*t* (28) = .135, *p* = .894). A correlation test confirmed this result [*F* (1, 29) = .018, *p* < .894, R^2 = .025. Consequently, age was not significant in affecting participants' performance, *b* = -.040, t = -.135].

Effect of Culture - Length of stay in the L2 speaking country and amount of L2 usage were additional points of interest. An independent sample t-test was conducted to examine the possible effects of bilinguals' L2 culture on their pair ratings. Eleven participants reported that they had been in the UK (L2 country) for fewer than three years [mean ratings (*SD*) = 3.23 (.77)], while nineteen participants had spent more than three years [mean ratings (*SD*) = 3.34 (.81)] and none had been in an English speaking country for less than a year. Comparing the mean ratings of these two groups did not reach significance (*F* (2, 28) = 342, t = -.349, *p* = .730). It appears that length of stay in the bilinguals' L2 country does not affect their performance, as those who stayed longer did not rate pairs differently from those who had a shorter stay.

Furthermore, participants were asked to choose among seven statements about language usage the one that they think most applicable to them. Nine participants used their native language (Arabic) substantially more frequently than English [mean (*SD*) = 3.36 (.84)], thirteen participants used both languages more or less equally often [mean (*SD*) = 3.23 (.78)], while eight reported that they use both languages more or less equally often, but they tend to use English slightly more often [mean (*SD*) = 3.34 (.82)]. There was no significant difference between the ratings of bilinguals who either use their L1 more, L2 more or even both languages equally often (*F* (29)= .077, *p* = .926) . Therefore, the amount of language usage does not seem to influence the bilinguals' performance in cognitive tasks such as similarity ratings.

5.2.4 Discussion

The main aim of the similarity rating experiment was to investigate the strength and pervasiveness of language-specific effects on cognition by comparing the performance of monolingual and bilingual speakers using the same task (similarity rating) and identical stimuli. Participants were asked to rate a number of pairs on seven-point scales in terms of similarities in order to examine whether the Arabic gender system would play a role in the participants' ratings. The findings of this experiment revealed that Arabic grammatical gender did not serve as an organising dimension in the conceptual

representation of Arabic speakers and that grammatical gender effects were not observed in using the similarity rating task. This indicated that Arabic speakers do not perceive pairs of the same grammatical gender as more similar than pairs that do not share this syntactic property. Bilinguals did not show increased ratings to same gender pairs or lower ratings to different gender pairs, showing that their ratings were not affected by the gender system of their language. The central variable that induced higher ratings of similarities across groups was the semantic relationships between the compared pairs. To put this simply, all participants gave higher ratings when the two pairs were from the same semantic groups and lower ratings when they were from different semantic groups.

Furthermore, investigating other variables such as bilingual's age, gender, languages known, age of L2 acquisition, L2 culture, amount of language use and academic qualifications did not appear to influence the participants' ratings. These results, as with the ones obtained from the monolingual groups, support a number of other studies about the limited effect of grammatical gender when using a methodology that made the variable of interest (grammatical gender) immune to the use of strategies (e.g. similarity ratings of pictorial stimuli). As mentioned earlier (in section 5.1.6), nearly all the studies that used similarity rating experiments reached the conclusion that knowledge of grammatical gender does not affect speakers' ratings in any significant way (e.g. Degani, 2007; and Ramos and Roberson, 2010).

In addition, the current results give strong evidence that effects of grammatical gender on people's ratings found in previous studies (e.g. Konishi, 1993; Clarke et al., 1981), in which participants were required to rate words in terms of feminine and masculine attributes, were task-dependent, that is participants in such tasks were specifically directed to attend to the masculine/feminine classes available in their languages. Clarke et al. (1981) did not rule out the possibility that participants in their study rated the pairs by referring to linguistic markings in their own languages. The present results show that grammatical gender was not an informative linguistic feature in rating the pictures and was not taken into consideration. Therefore, participants from the two different language groups have seemed to use similar clusters of features to make their ratings. These inconsistencies might be attributed to the different tasks used in these studies (e.g. using verbal vs. pictorial stimuli), and raise further questions about the nature of the effects of grammatical gender and how they creep into perception during cognitive and linguistic development.

Recent evidence suggests that effects of grammatical gender could emerge under specific conditions and produce semantic influences. For example, Cubelli et al. (2011) studied the performance of Italian, Spanish, and English speakers in categorisation tasks. The findings showed that speakers of both gendered languages - but not the English group - were faster when the pair of pictures shared the same grammatical gender in their native language than when the genders of the two pictures were of different gender. These results suggest that participants' performance was affected by the congruity of grammatical gender in their native language. One possible interpretation for such effects might be that the dependent variable in Cubelli's study is reaction time. Grammatical gender effects may manifest themselves in the ease of access to conceptual features of objects, but not on the representations themselves, as revealed by the present experiment. Another possibility is that if grammatical gender plays a semantic role, it may be limited to representation of some nouns but not others. For example, it might be shown with animate nouns (e.g. animals) but not with inanimate nouns (e.g. artefacts). The effects of grammatical gender may also be present in some languages but not others (e.g. Spanish and Italian, but not Arabic as the effect may not be strong enough to manifest itself in every task).

Chapter 6. General Discussion

The aim of this study was twofold. First, it set out to investigate the general question of whether the Arabic grammatical gender system affects cognitive representation and processes. The second aim was to discover the possible effects of bilingualism on cognition; that is whether learning a second language actually changes cognitive representation in bilinguals. The nature of the effect of language on the thoughts of its users has been summarised, in the main, from two approaches. The most widely investigated view is the linguistic relativity hypothesis, which has come to be associated with Benjamin Lee Whorf (1939, 1941). The other view, which is a version of the linguistic relativity hypothesis, was put forward by Gennari et al. (2003) and emphasises that language affects cognitive processes when it is used strategically in cognitive experiments. Gaining evidence of the influence of language on cognition offers us a test-bed to investigate bilingual cognition and find out whether bilingualism changes cognitive representations and processes and additionally the extent of this change. We set out to discover whether the influence of bilingualism is strong, meaning that bilinguals differ from monolinguals of their native language and behave like monolinguals of their second language. An alternative would be to discover that it is limited, meaning that bilinguals and monolinguals behave similarly.

The voice-attribution and similarity rating tasks both tried to investigate these questions in the domain of grammatical gender. The first issue investigated whether learning a grammatically gendered language leads to the restructuring of conceptual and semantic representations in monolingual speakers, as suggested by Sera et al. (1994) and Sera et al. (2002). The second issue examined the effect of bilingualism on Arabic-English bilinguals (e.g. Boroditsky et al., 2003; Bassetti, 2007, 2011; Kousta et al., 2008).

Effects of grammatical gender on monolingual cognition

The results of the first experiment, the *voice-attribution task*, revealed that Arabic monolingual speakers were strongly affected by the gender system of their language when required to assign voices to pictures of inanimate objects compared to English speakers who did not follow the same pattern. Arabic monolinguals' voice assignments followed the Arabic grammatical gender classifications considerably more than would be expected by chance, irrespective of the conceptual features of the inanimate objects.

Voice assignments by English speaker were, however, in conjunction with the items' conceptual nature, that is natural vs. artificial distinction. The number of inanimate objects for which Arabic monolinguals did not assign voices in association with the Arabic gender system was very small in comparison to the total number of the task items. Furthermore, the proportion of same gender voice assignments by Arabic monolinguals significantly outweighed those made by English speakers. One might argue that grammatical gender was used strategically in the voice-attribution task. If this were the case, however, Arabic monolinguals would have given one hundred per cent same gender voice assignments. In fact their overall voice assignments were eighty four per cent, which is still statistically significant i.e. performance was above chance. This indicated that grammatical gender strongly affected their voice assignments. It might not, however, be the only factor affecting their performance.

It is worth noting that English speakers assigned voices in consistency with the Arabic gender system more than half the time, suggesting that Arabic monolinguals relied on grammatical gender in their voice assignments, taking into account other features of the items which are more likely to be shared across languages. This indicated that, in Arabic, grammatical gender assignments for inanimate objects are not totally arbitrary. The present findings also provide evidence that voice assignments made by English speakers were influenced by the male-artificial/female-natural distinction (Ortner, 1974). These results support previous research by Sera et al. (1994, 2002); Bassetti (2007) and Ramos and Roberson (2010).

The second experiment, the *similarity rating task*, showed that in a task where strategic access to linguistic information is blocked and reference to the relationship between grammatical and natural gender is avoided, there was no evidence of a strong association between these two sources of information for Arabic monolingual speakers. In this task, Arabic and English monolinguals were asked to rate the similarity between pictured pairs of items on seven-point scales. Only pictorial stimuli were used as test materials providing a more reliable test of the linguistic relativity hypothesis, as suggested by (Johnson et al., 1996); this manipulation helped to access conceptual information about the presented objects without retrieving their linguistic labels. The similarity rating task may explain the role that encoding processes and task demands play in the participants' performance. The findings revealed that Arabic monolinguals were not affected by the grammatical gender system in the similarity rating task, suggesting that their knowledge of grammatical gender was not taken into account while rating the similarity between pairs of items. They did not rate gender-congruent

pairs as more similar and did not give lower ratings to gender-incongruent pairs. A different view, however, is suggested by Lupyan (2012) who introduces the label-feedback hypothesis. Lupyan (2012) argues that verbal labels are co-activated when we activate conceptual representations. Specifically, the label-feedback hypothesis views language to be always involved in categorising unless it is disrupted (e.g. by verbal interference or aphasia), which means that, in everyday situations, language is affecting our non-linguistic cognition. Nevertheless, this was not the case in the current study as it was not an interference-based experiment and no language effect was observed.

Another important finding of this experiment was the strong effect of semantic homogeneity on the ratings of both Arabic and English monolingual speakers. More specifically, pairs that were from the same semantic groups were perceived as more similar by the two language groups and these ratings were independent of the effect of grammatical gender. The performance of both Arabic and English groups was therefore comparable in the similarity task, indicating that they both relied on similar processing strategies. This pattern fits well with previous research (e.g. Ramos & Roberson, 2010) into the significant role of semantic homogeneity on participants' ratings.

We can see that Arabic grammatical gender is neither a central part of the overall representation of objects, nor a naturally significant property in situations where people are asked to rate how similar two items are. These findings go against the findings of deep effects of grammatical gender on our conceptual representations (e.g. Sera and colleagues, 1994, 2002; Boroditsky et al., 2003) according to which there should be an observable effect of grammatical gender across different tasks. It could be argued, however, that such effects would have been found had the experiment been carried out with verbal stimuli rather than pictures. Nevertheless, verbal materials are best avoided in studies that test the possible effects of language on cognition as they may bias participants in favour of verbal coding and lead to an exaggeration of linguistic effects on conceptual representations (Munnich and Landau, 2003). Moreover, different results might have been obtained had the task instructions not avoided any reference of gender, or if the task had used pictures of animate objects. A simple change in task instructions may lead to different outcomes; in other words, any change in the experiment may lead participants to perform a given task in a way that either accords or does not accord with an expected conclusion (Levinson et al., 2002).

In fact, the quantitative judgments required for the similarity rating task may not be that easy as they demand a precise decision about the commonalities between the presented pairs. If Arabic grammatical gender was used in this task, it would have

facilitated disambiguation and made it easier for Arabic monolinguals as was the case with voice attribution task where grammatical gender knowledge was made relevant by task demands, even though linguistic encoding was not required. Medin et al. (1993) point out that when participants are asked to assess the similarity between a set of items, it is unlikely that they would access all their knowledge about the pairs to be compared. Medin et al. (ibid.) further argued that only a small set of features is likely to be activated during the task, meaning that there would be no guarantee that all participants would choose the same features and perform the task in a similar way. This view suggests that the different results obtained from the two experiments could be due to the fact that participants adopted different criteria and relied on different strategies to perform cognitive tasks.

Effects of grammatical gender on bilingual cognition

The results obtained from the *voice-attribution task* with Arabic-English bilinguals revealed an effect of learning a second language (L2) on the cognition of the bilinguals. In this experiment, bilinguals assigned voices differently from monolinguals of their native language L1 (Arabic) and their voice assignments were in-between the two monolingual groups. These results extend previous cross-linguistic research (e.g. Sera et al., 1994; Sera et al., 2002; Ramos and Roberson, 2010) by using the same voice-attribution task with different test items and with bilingual as well as monolingual samples. The difference between voice assignments made by Arabic monolinguals and Arabic-English bilinguals indicates that learning a second language (L2) with a natural gender system foregrounds the arbitrary nature of gender assignment in the bilinguals' native language (L1) and leads to a restructuring of semantic representations. These findings also support the idea of multi-competence (Cook 1991; 1994) which asserts that bilinguals (or L2 users) think or perceive things differently from monolinguals. Although the terms *think* and *perceive* have a broad scope of meanings, they both include a variety of cognitive activities such as judgement, memory, inference and classification (Lucy, 1997) and categorisation of objects is one of the key cognitive activities of human beings (Murphy, 2004). These results are in agreement with the results of various studies that have investigated the relationship between language and thought in bilinguals regarding the domains of colour (Athanasopoulos, 2009), emotion vocabulary (Pavlenko and Driagina, 2007) and grammatical number and object categorisations (Athanasopoulos, 2007; Cook et al., 2006), which all suggest that learning a second language leads to restructuring in the bilingual mind. Although these

studies addressed different bilingual domains (grammatical categories, colour, emotion vocabulary and object categorisations) and provided supportive evidence, more research is necessary to better understand the nature of the effects of language on our thinking in a more general context.

Such effects may be due to other linguistic and socio-cultural factors underpinning bilinguals' cognitive behaviour, such as age of L2 acquisition, number of languages known, amount of language usage and length of residence in the countries where the L1 and L2 are spoken. Although it would be difficult to take all these variables into consideration when conducting psychological research, the present study has tried to include as much information as possible about participants from all language groups. The aim behind studying all these background variables was to allow for correlational studies between cognitive performance and the socio-cultural variables (Athanasopoulos, 2011). This study did not, however, find any effects from all these variables on the overall performance of the bilinguals.

For example, in order to rule out the possibility that bilinguals were thinking in their monolingual L2 mode, interactional settings (see Grosjean, 2010) were carefully considered when testing bilinguals. More specifically, the whole experiment was conducted in the participants' native language and they were tested by a native speaker of their L1. This step clearly showed that observed differences between monolingual and bilingual performances were due to the effects of language rather than other factors. Stronger effects of interactional settings might have been found had the bilinguals been tested in their second language L2 (English), as they might show an increased shift towards the cognitive patterns of monolingual speakers of English. Future research could test the same group of bilinguals in their L2 and then assess the extent to which language mode affects the bilinguals' performance.

Furthermore, the present findings reveal that non-linguistic socio-cultural variables such as length of stay in the L2-speaking country do not have any significant role in bilingual cognitive restructuring in the domain of grammatical gender. In some studies, this variable was reported to affect performance; for example, Cook et al. (2006) showed that Japanese L2 users of English living in an English speaking country moved some way towards the English preference. Nevertheless, the present study did not show such an effect. The length of stay in an English speaking country was not strong enough to affect the performance of Arabic-English bilinguals. Although these findings differ from Cook et al. (2006), they are in agreement with Athanasopoulos (2006) and Athanasopoulos and Kasai (2008) who implemented a triads matching task,

comparing similarity judgements between English and Japanese monolinguals and Japanese-English bilinguals and found that length of stay in the UK did not influence the Japanese bilinguals' cognition. A possible explanation for such inconsistent evidence between these studies could be attributed to methodological differences and areas of investigation, e.g. the effects of this variable might be more apparent in some domains (grammatical number) rather than others (grammatical gender).

Moving on to the similarity rating task, the results did not reveal any differences between the Arabic and English monolingual groups and the Arabic-English bilinguals did not differ from them either. As mentioned above, when gender is not made relevant by task demand, the effects of this grammatical property disappear. Similar to the monolingual groups, bilinguals were affected by the semantic-homogeneity of the pairs; they gave higher ratings to pairs from the same semantic groups (e.g. body parts, household items, food, vehicles and clothes) and lower rating to pairs from different semantic groups.

The use of different tasks to study the same issue helps shed light on the nature of grammatical gender effects. It helps us to disentangle the real effect of this grammatical property on cognition from strategically using it to perform the task at hand. Researchers need to bear in mind that in studying the cognitive differences between speakers of different languages, or the effects of language on bilingual cognition, it is necessary to investigate differences as well as to acknowledge the universal elements of cognition. Although language plays a vital role in the categorisation of people, both relativity and universals interact in language use and cognition at the same time (Imai, 2000; Gelman, 2004). We cannot, therefore, ignore the fact that we all have a universal aspect in the greater part of the cognitive processes in categorisation, yet there are some areas more specific to speakers of certain languages than to others. These specific aspects of a language may guide its speakers to pay more attention to some specific sections of events, based on linguistic form.

It appears that grammatical gender effects on judgement tasks relating to conceptual and semantic knowledge in Arabic are more likely to arise from an access to linguistic information in tasks where this information is relevant. In such a case, the effects were noticeable even with inanimate items, showing that grammatical gender effects on categorisation are not limited to classes of objects with natural gender such as animals (e.g. Vigliocco et al., 2005). Nevertheless, the effects of the Arabic gender system found in the voice-attribution task do not seem to appear as the result of the importance of male and female-like properties for items for which sex is an irrelevant

feature. It could be argued that knowledge of grammatical gender is taken into account in tasks in which the use of this linguistic knowledge facilitates a decision that might be difficult to make on the basis of other properties. For Vigliocco et al. (ibid), generalising the idea of male and female-like properties to referents of nouns with the corresponding gender occurs only for items for which these properties are relevant. The findings of the present voice-attribution task, however, showed that when the use of grammatical gender is made relevant by task demands, its effects strongly appear on judgments about inanimate objects. As the effects of gender information were not observable across the tasks, it seems that this property complements rather than dominates conceptual and semantic representations. On an alternative view, "conceptual representations are dynamic assemblies that are a function of a prior knowledge as well as current task demands" (Lupyan, 2012: 10). That is to say, the effect of language on ongoing processing may be present in some tasks and non-existent in others. This view provides a useful way for thinking about the interaction of language with other processes. Furthermore, the present experiments show good examples of the fact that significant grammatical gender differences do exist between languages as well as there being many similarities in the performance of speakers of both Arabic and English.

It is worth noting that previous studies (e.g. Sera et al., 1994, 2002; Boroditsky et al., 2003) on the effects of grammatical gender on cognition have been mainly concerned with a static view of the influence of language on cognition. We can say that these studies supposed a dominant effect of grammatical gender systems at a deep conceptual and semantic representational level. The present findings, however, revealed that grammatical gender can be taken into account in some tasks even when linguistic access is not required or allowed and that other types of task can be performed without any access to this linguistic property. Such results suggest that the underlying conceptual and semantic representations of items are not always affected by knowledge of grammatical gender, yet cognitive judgements made on those representations might be affected by this property under certain situations. Therefore, the effects of grammatical gender on our thinking are better to be viewed as both a dynamic and a controlled effect of language on cognition.

An interpretation made by Hunt and Agnoli (1991: 378) of the weak version of the linguistic relativity hypothesis asserts that "language differentially favours some thought processes over others". In the current experiments, Arabic speakers seem to be affected by gender knowledge in tasks where this linguistic feature was made relevant by task demands such as an explicit reference to natural gender - as in the voice-

attribution task. In the similarity rating task, which was completed without taking gender into consideration, there was a lack of evidence on the role of grammatical gender on the participants' performance. We can conclude that these different results fit with the view that some thought processes are favoured by the characteristics of the language. In the case of Arabic, explicitly accessing grammatical gender information and its classifications as an extra feature together with other strategies would be more favoured in some tasks rather than others. For example, while the effect was obvious in the voice-attribution task, in the rating task it was not strong enough to represent a pattern of habitual thought (Whorf, 1956).

Chapter 7. Conclusion

7.1 Conclusion

This study set out to investigate the relationship between the language we speak and the way we think about reality. It has further sought to examine whether learning a second language can change bilinguals' thinking and make them think differently from those who know only one language. This issue, also known as linguistic relativity (Whorf, 1956; Lucy, 1992), cuts across the fields of linguistics, psychology and cognitive sciences. The general theoretical literature on this topic and specifically in the domain of grammatical gender is inconclusive on a number of vital questions within the language thought relationship. This study sought to answer two of these questions namely; (a) does the Arabic grammatical gender system affect the categorisation of objects and (b) does learning a second language change bilinguals' thinking and if so, to what extent? The study reviewed some of the previous research on the linguistic relativity hypothesis as well as on grammatical gender, which mostly made claims regarding the effects of language on cognition.

In a way that is similar to most studies in linguistic relativity, this study has adopted a comparative approach. This is believed to be the most effective approach to linguistic relativity as comparisons allow researchers to establish similarities and differences between the studied languages. Grammatical gender was selected as an area of investigation for the linguistic relativity hypothesis because this grammatical category is both salient and pervasive in Arabic, but absent in English. Additionally, this study tried to avoid some of the methodological pitfalls found in previous research such as the use of linguistic stimuli and examining linguistic behaviour which showed linguistic diversity rather than unequivocally the effects of language on thinking. Therefore, the present study used non-linguistic stimuli (pictorial) and tasks which were cognitive (voice-attribution and similarity ratings), so here neither stimuli nor tasks involved any linguistic interaction. The aim of this manipulation was to find out whether cognitive perception parallels linguistic structure without relying on language itself by making sure that participants' cognitive behaviour is not biased by language during the task performance. Additionally, the study addressed the relative impact of a number of variables that might affect the bilinguals' cognitive representations.

The main findings of this piece of research demonstrate that the Arabic grammatical gender system is not a central part of Arabic speakers' mental

representations; rather it is used as a strategy to assist in performing certain tasks. For example, as shown in the voice-attribution task (chapter 4) this grammatical property was strongly significant in affecting Arabic monolinguals' categorisations. The study further gave evidence to the idea of male-artificial/female-natural distinction (Ortner, 1974; Sera et al., 1994) as being an important factor in affecting the performance of both language groups. This distinction originally suggested by Ortner (1974) who makes her famous argument that culture is associated with men, whereas women are more aligned with nature. Ortner argues that a woman's body and its functions keep her closer to nature more than a man's physiology, allowing him more freedom to work in culture. That is, women are biologically constructed to enable them to perform the reproduction of the human species, whereas men are not constantly occupied with this role, so they seek distraction within culture. According to this view, the association of females with natural objects and males with artificial ones is a universal conceptual distinction that greatly affects the role of males and females in society. Although these claims have been criticised, many studies, including this one, reported that speakers (either of grammatical or natural gender languages) more often assign natural objects to a female category and artificial objects to a male category. In this study, Arabic and English speakers both assigned more masculine voices to artificial objects than to natural objects even though both of these objects were grammatically masculine in Arabic. Likewise, they assigned more female voices to natural objects than to artificial objects although both were grammatically feminine in Arabic. Nevertheless, Arabic speakers' voice assignments were significantly above the possibility of chance for all inanimate objects, indicating a strong effect of the Arabic grammatical gender system on their cognition. It is worth noting that, in some experiments (e.g. voice and name attribution tasks), participants are likely to use any available knowledge in order to assist with their decisions and grammatical gender seems to be a suitable strategy to use.

On the other hand, when using a more cognitive task (similarity ratings) that blocked access to language in general and gender information in particular, grammatical gender did not seem to exert any influence on the participants' cognitive representation and processing. These results show that this grammatical feature does not have a pervasive role in conceptual representation; rather that its effects only appear in tasks that allow access to linguistic knowledge. The discrepancy between these findings and those obtained from the voice-attribution task indicates that the effects of grammatical gender do not arise at a deep conceptual level rather the effect is task-dependent as it depends on the use of this linguistic information according to task demands. This

explanation goes in line with Bowers et al. (1999) and Gennari et al. (2002) who all believe in the strategic use of grammatical gender in certain tasks, thus grammatical gender does not affect people's cognition at a deep conceptual level.

It is apparent that speaking differently does not necessarily cause different thinking. Athanasopoulos and Bylund (2013) offered two possible reasons for the association between grammatical aspects and cognition; (a) language permanently shapes the way speakers of different languages perceive and interpret the world, (b) the effect of language is more dynamic and flexible, becoming apparent only in certain contexts where the strategic use of language is allowed in the task at hand. The first view would be acceptable if we could find differences between speakers of different languages across a range of cognitive tasks including similarity rating tasks. The present study does not support this view, however, as differences between Arabic and English speakers were only found in the voice-attribution task and these differences disappeared in the similarity rating task. These results strongly support the latter view that language is used strategically to carry out the task.

The issue of bilingualism was another main concern of this thesis. Potential effects of learning a second language with a natural gender system (e.g. English) on the bilinguals' categorisations of objects were examined. The study showed that Arabic-English bilinguals' voice assignments were not significantly influenced by the Arabic grammatical gender system as was the case with the Arabic monolinguals and their performance came in between the two monolingual groups. These results adduced evidence in support of the effect of bilingualism on our cognition, showing that by knowing more than one language, a person will no longer think as a monolingual (see Cook, 1991). The effects of bilingualism were found in a number of studies in different domains such as colour perception (Athanasopoulos, 2009), object categorisation (Cook et al., 2006), motion event (Malt et al., 2003; Slobin, 2003) and grammatical gender (Bassetti, 2007, 2010; Kousta et al., 2008)[28].

This study, therefore, adds to this line of research by offering supporting evidence to the effect of learning a second language on the cognition of the bilingual. Furthermore, the study showed that intermediate bilinguals behaved similarly to advanced bilinguals, so that their voice assignments did not differ from each other. Such results indicate that learning a second language might shift our thinking towards that of the monolinguals' L2 even if advanced proficiency is not yet achieved. Learning

[28] Also, see Cook & Bassetti (2011) for a number of recent studies from various disciplines of the relationship between language and cognition, with a focus on bilinguals. They contrast macro effects which occur regardless of languages with micro effects which occur for a specific pair of languages.

another language may force us to see things in a different way, according to Finch (1998-2007: 277) "if learning another human language can be compared to opening a window on the world, then learning an alien language may open the door on the universe. We will never be the same person again".

This in turn, has valuable implications to both second language learning and second language research in general. Further research in this area will surely help to shed light on the nature as well as the extent of the effects of language on cognition. It further helps individuals to realise that by learning another language, they are developing a unique conceptual system that is different from their monolingual counterparts. Therefore, they might "cut nature up, organize it into concepts and ascribe significance" by using the language they speak (Whorf, 1956: 213).

In addition, the study examined the age of L2 acquisition as a potential factor in affecting bilinguals' performance. The results showed that although late bilinguals, who acquired/learned English L2 after the age of seven, were more consistent with Arabic grammatical gender than early bilinguals who acquired English L2 before the age of seven, they did not significantly differ from each other as they both shifted their behaviour as they acquired an L2. Bilingualism at any age, therefore, could modulate the effects of grammatical gender in certain contexts, e.g. late bilinguals, like early bilinguals, differed from the monolinguals' L1. These findings have implications for second language learning as learning a second language, even later in life, seems to change our thinking and reduce language-induced biases in our mental representations of the world. These results also accord with the views of Whorf (1956) who alleged that the solution to language biases in our views about the world was to learn more than one language. The correlational analyses revealed that other non-linguistic variables such as acculturation, language use and participants' academic qualifications did not play a role in the bilinguals' cognitive change, indicating that the observed changes in bilinguals were due learning another language per se. This has implications for the relationship between language and cognition in the bilingual mind. Learning another language with different concepts from the speakers' native language may lead to reorganisation in the bilingual's cognition and the extent of this organisation is open to further research (see Kroll et al., 2014, for a thorough discussion of these organisations).

7.2 Limitations and Recommendations for Future Research

The nature of this project adopted an experimental approach and the findings of such research may not be generalised to the whole population. Specifically, the study has

certain limitations which should be addressed in future research. There are many varieties of Arabic language - such as Moroccan Arabic and Lebanese Arabic - but the present study only focused on Saudi Arabic to discover a precise answer on the nature of the language effects; the current findings may not therefore be generalised to cover all Arabic speakers.

Another limitation is that this study only used two experiments to examine gender effects on Arabic speakers due to the I-PHD programme restrictions on word-count. Invaluable light would have been shed if more experiments were undertaken on the same populations. Future research may look at whether grammatical gender information is accessed during conceptual tasks that suppress the strategic use of linguistic knowledge through the use of speeded tasks with cognitive overload. Such tasks would help to clarify the extent to which strategic processes are used in accessing linguistic information during these tasks.

The findings of this piece of research answer some general questions about the type of materials used in research on language and cognition. Hampton and Dubois (1993) questioned whether asking participants to provide judgements about objects while using labels as test stimuli would really reveal their underlying conceptual representations, or whether linguistic information would certainly affect their performance to some degree. They argue that researchers might separate conceptual content from language by conducting comparative research across languages and using non-linguistic materials and by doing so researchers then could ask the question "to what extent is conceptual organisation language-free" (Hampton and Dubois, 1993: 26). The findings of the present project advocate that conceptual organisation is free from the effect of grammatical gender, especially for inanimate objects. Nevertheless, these findings suggest that variations in task demands and materials would lead participants to use different strategies for accessing cognitive representations. Future research should, therefore, take these issues into careful consideration.

Furthermore, the study has investigated adult participants because they supposedly have a more complete first language acquisition and are more likely to be better at understanding and performing cognitive activities. Future research could investigate other age groups such as children and young adults which could enrich our understanding of gender effects across age groups.

Finally, this study looked at the effects of learning a natural gender language on bilinguals' cognition; it would be interesting if future research could investigate the effects of learning a grammatically gendered language (e.g. French, Italian or Spanish)

on native speakers of Arabic. Such research may show a different pattern from learning a language whose gender category is empty. This may be undertaken by examining any potential gender effects on categorisation in native Arabic speakers learning French, Spanish or Italian. Further research would reveal whether speakers whose first language has a grammatical gender category would be less prone to be influenced by the second language's grammatical gender pattern. Whorf (1956: 225) referred to this as the 'binding power' of a language which might be different to the present study as English does not have grammatical gender system; therefore, the participants were not 'bound' to an opposite cognitive representation of the same reality. It would be interesting to investigate the binding powers of bilinguals' second language in the domain of grammatical gender when the speakers' native language does have a grammatical gender system. Bilinguals may realise the semantic arbitrariness of L1 grammatical gender and could then not be influenced by the grammatical gender of one language. This is one of the implications of bilingualism proposed by Cook (1999, 2002) as part of his multi-competence theory, according to which bilinguals may develop new concepts that are in-between the concepts of their two languages or different from either. Alternatively, effects might be limited to those objects that have opposite grammatical gender in the bilinguals' two languages. Future research may compare categorisation of objects that have opposite gender in the two languages and objects that have the same gender. This would explain whether learning a grammatically gendered second language has a general effect of reducing the effects of grammatical gender of the first language, or simply influences the perception of masculinity-femininity for objects that have opposite grammatical gender in the two languages.

Appendices

Appendix 1A: Questionnaire (English version)

For the use of the experimenter ID:

1. **Name** (optional):

2. **Age group:**
   - 20-29
   - 30-39
   - 40-49
   - 50-59

3. **Gender:**
   - Male
   - Female

4. **Please list all the languages you can speak (including your first language), the age at which you started learning each one and your proficiency in each of them** on a scale from 1 to 6, with 1= "very basic" and 6= "very advanced"

Language	**Age when you started learning it**	**Proficiency**
		1 2 3 4 5 6
		1 2 3 4 5 6
		1 2 3 4 5 6
		1 2 3 4 5 6
		1 2 3 4 5 6

5. **How much time do you think you currently spend each week using each language (in hours per day)?** (Include in your estimate things like socialising, listening to the radio, watching TV/films and reading/studying in each language):

_____

_____

_____

6. **Which of the following statements applies to you at the moment?**

a) I use Arabic only
b) I use Arabic substantially more frequently than English
c) I use both languages more or less equally often, but tend to use Arabic slightly more often
d) I use both languages more or less equally often
e) I use both languages more or less equally often, but tend to use English slightly more often
f) I use English substantially more frequently than Arabic
g) I only use English

7. **Have you lived in the country of your second language? If yes, for how long?**

_____

8. **What is your most recent academic qualification (e. g. school leaving qualification, BA, other)?**

_____

_____

9. **Are you studying anything at the moment? If so, which institution and course?**

_____

Thank you

Appendix 1B: Questionnaire (Arabic version)
الاستبانة اللغوية

لاستخدام الباحثة **الرقم:**

١ـ الاسم: _____

٢ـ الفئة العمرية :

٢٠ـ٢٩

٣٠ـ٣٩

٤٠ـ٤٩

٥٠ـ٥٩

٣ـ الجنس:

ـ ذكر

ـ أنثى

٤ـ أرجو التكرم بذكر كافة اللغات التي تتحدثها (بما في ذلك لغتك الأم)، و العمر الذي بدأت فيه تعلم كل لغة و كذلك تقييم مستوى اتقانك لكل لغة باستخدام المقياس التالي: من ١ـ٦ حيث أن ١ تعني (مبتدئ جداً) و ٦ (تعني متقدم جداً)

اللغة	في أي عمر بدأت تعلمها	مستوى اتقانها
		١ ٢ ٣ ٤ ٥ ٦
		١ ٢ ٣ ٤ ٥ ٦
		١ ٢ ٣ ٤ ٥ ٦
		١ ٢ ٣ ٤ ٥ ٦
		١ ٢ ٣ ٤ ٥ ٦

٥- كم ساعة تقضيها في الأسبوع باستخدام كل لغة حالياً؟ (و هذا يشمل الأمور التقديرية كالتواصل مع الناس و الاستماع للراديو و مشاهدة التلفاز و الأفلام و القراءة و المذاكرة في كل لغة)

٦- أي من الجمل التاتلية تنطبق عليك حالياً؟

أ. استخدم العربية فقط.

ب. استخدم العربية بشكل أكبر بكثير من استخدامي للانجليزية.

ج. غالباً ما استخدم اللغتين بشكل متساوٍ تقريباً و لكن أجدني استخدم العربية بصورة أكثر نسبياً.

د. غالباً ما استخدم اللغتين بصورة متساوية.

ه. غالباً ما استخدم اللغتين بشكل متساوٍ تقريباً و لكن أجدني استخدم الانجليزية بصورة أكثر نسبياً.

و. استخدم الانجليزية بشكل أكبر بكثير من استخدامي للعربية.

ز. استخدم الانجيليزية فقط.

٧. زرت هل البلد الذي يتحدث لغتك الثانية؟ اذا كان الجواب نعم فما هي المدة؟

---

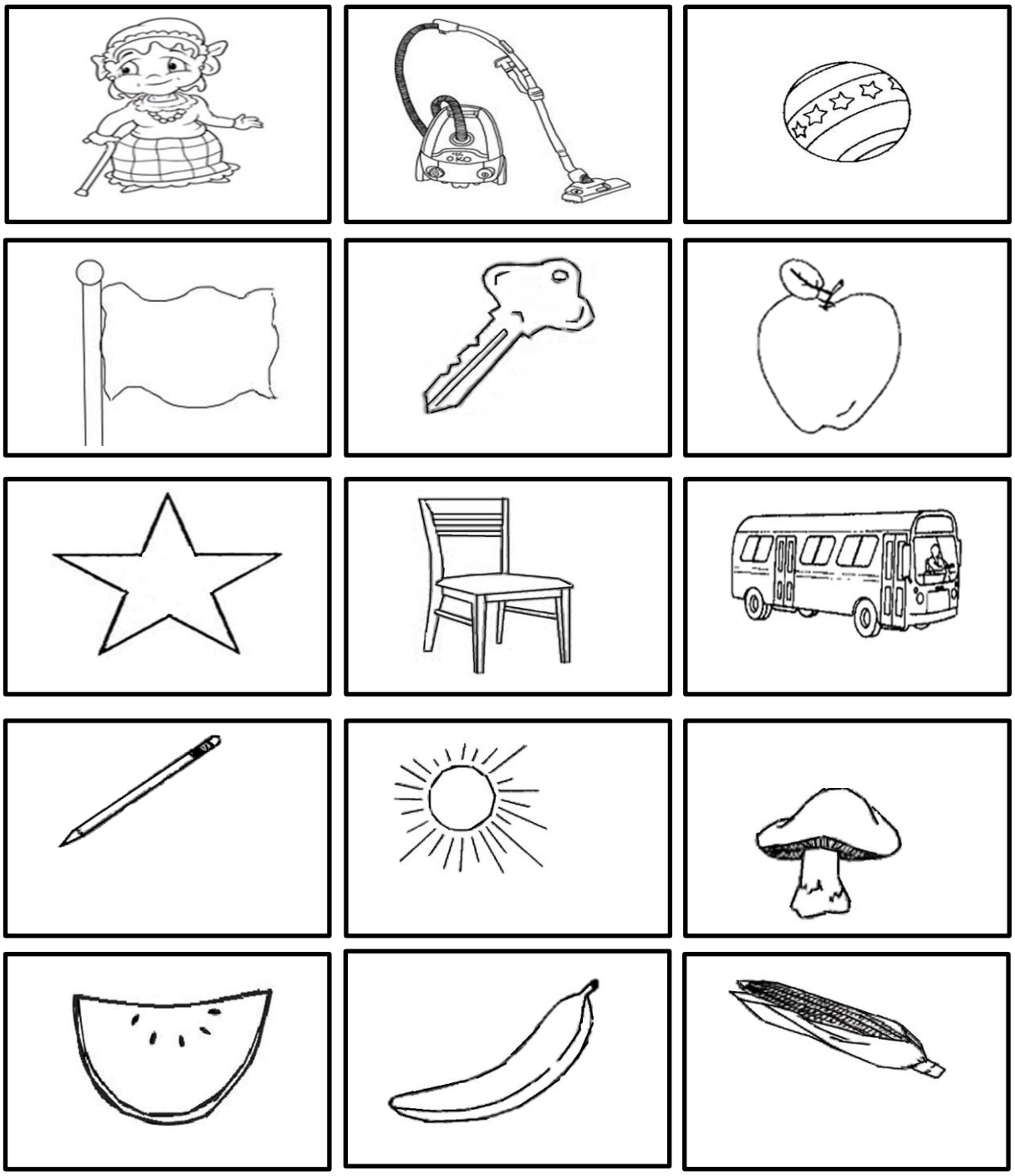
٨. ماهو آخر مؤهل حصلت عليه (بكالريوس، مؤهل آخر)؟

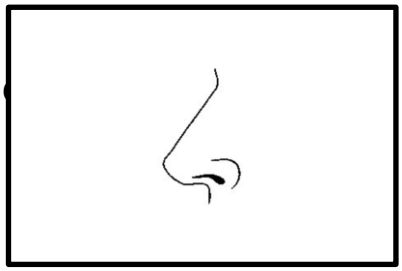
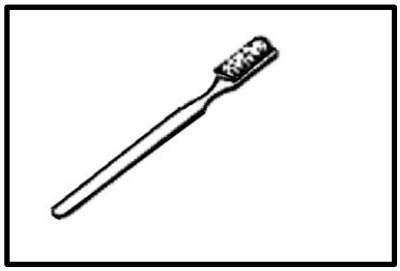
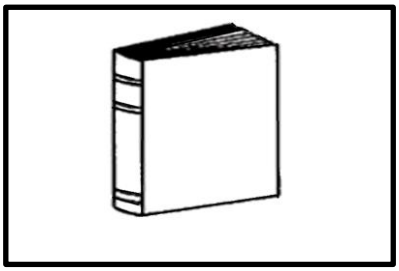
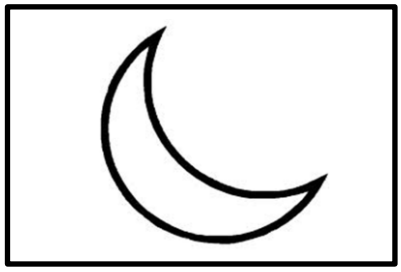
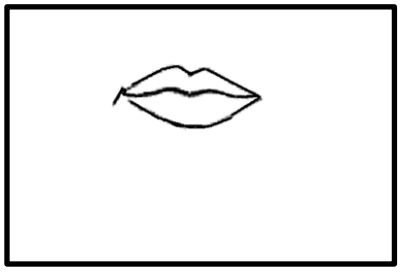
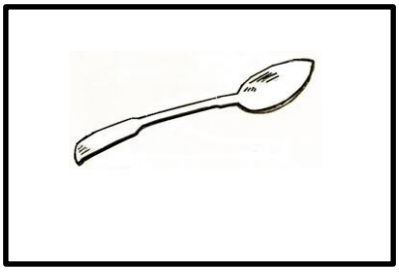
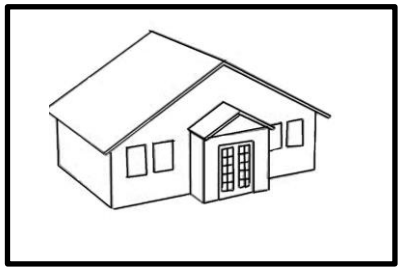
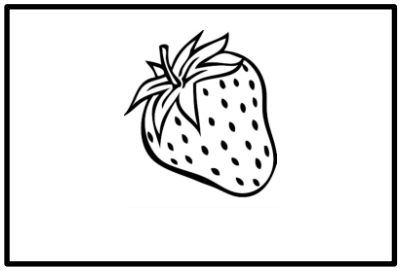
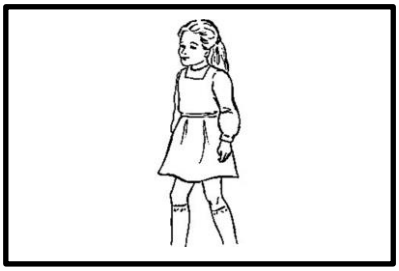
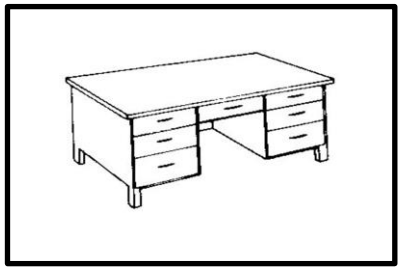
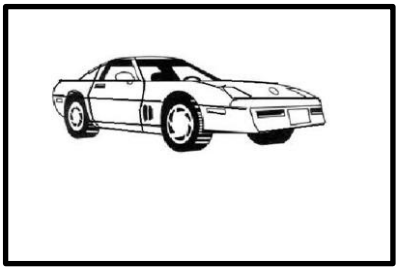
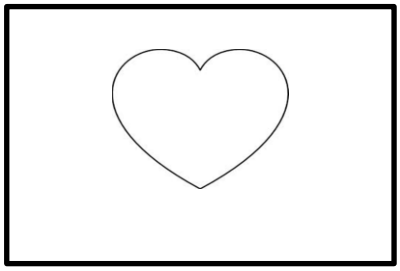
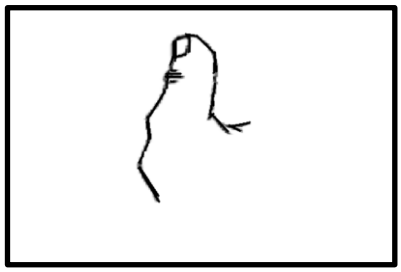
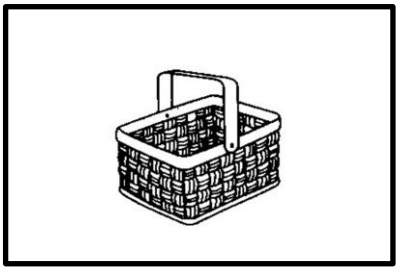
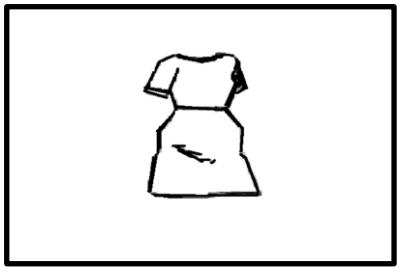
---

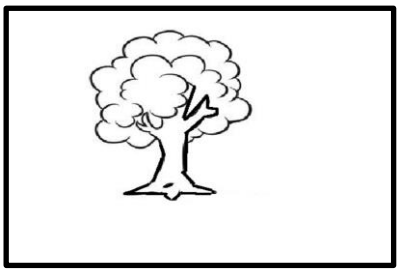
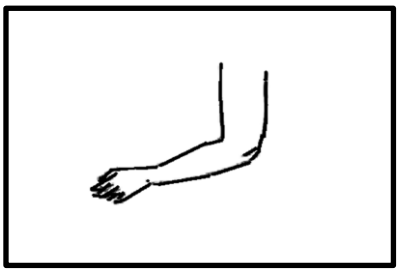
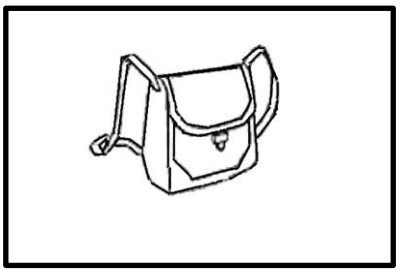
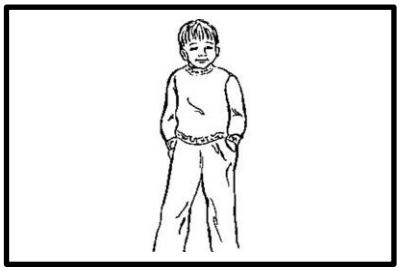
٩. هل تدرس حالياً؟ (اذا كان كذلك، أرجو ذكر المعهد و البرنامج)؟

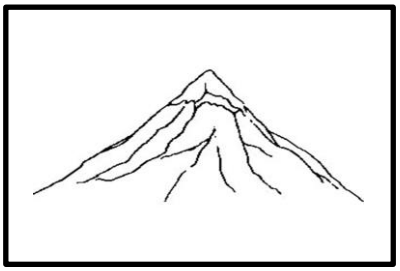
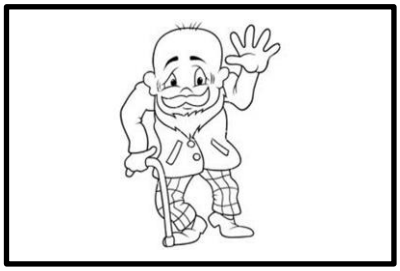
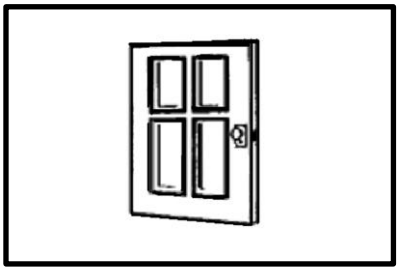
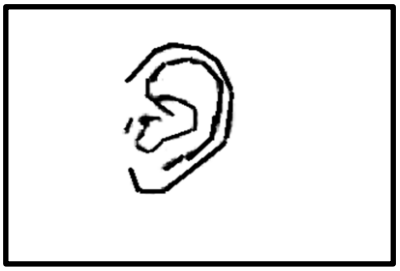
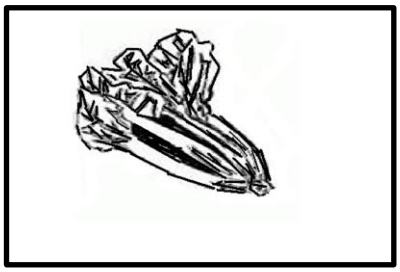
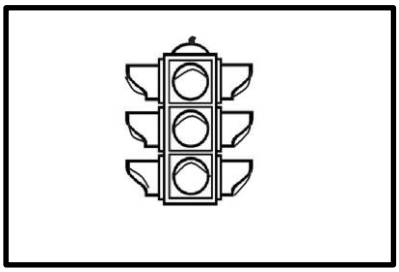
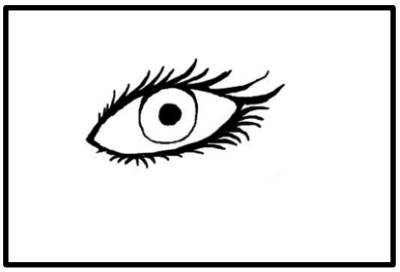
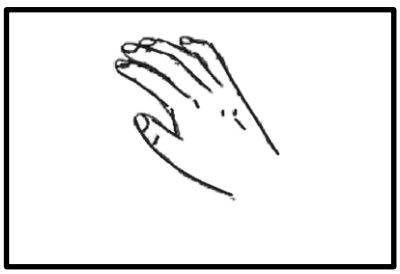
شكراً جزيلاً لك

Appendix 2: Stimuli of the Voice Attribution Experiment



Appendix 3A: Answer Sheet for Voice Attribution Experiment (English version)

Should the pictures shown on the screen have the voice of a man/boy or the voice of a woman or girl? If you decide that the object should have a female voice, please circle *F* in the column named "VOICE". If you decide that it should have a male voice, then circle *M*.

ITEM	**VOICE**	
1.	M	F
2.	M	F
3.	M	F
4.	M	F
5.	M	F
6.	M	F
7.	M	F
8.	M	F
9.	M	F
10.	M	F
11.	M	F
12.	M	F
13.	M	F
14.	M	F
15.	M	F
16.	M	F
17.	M	F
18.	M	F
19.	M	F
20.	M	F
21.	M	F
22.	M	F
23.	M	F
24.	M	F
25.	M	F
26.	M	F
27.	M	F
28.	M	F
29.	M	F
30.	M	F
31.	M	F
32.	M	F
33.	M	F
34.	M	F
35.	M	F
36.	M	F
37.	M	F
38.	M	F

39.	M	F
40.	M	F
41.	M	F
42.	M	F
43.	M	F
44.	M	F

Appendix 3B: Answer Sheet for Voice Attribution Experiment (Arabic version)

هل الصور التي تراها على الشاشة يجب أن تعطى صوت مرأة أو رجل؟ من فضلك ضع دائرة حول حرف (ر) اذا كنت ترى أن الصوت يجب أن يكون صوت رجل، و ضع دائرة حول حرف (م) اذا كنت ترى بأن الصوت يجب أن يكون صوت مرأة.

الصورة	الصوت	
١	ر	م
٢	ر	م
٣	ر	م
٤	ر	م
٥	ر	م
٦	ر	م
٧	ر	م
٨	ر	م
٩	ر	م
١٠	ر	م
١١	ر	م
١٢	ر	م
١٣	ر	م
١٤	ر	م
١٥	ر	م
١٦	ر	م
١٧	ر	م
١٨	ر	م
١٩	ر	م
٢٠	ر	م
٢١	ر	م
٢٢	ر	م
٢٣	ر	م
٢٤	ر	م
٢٥	ر	م
٢٦	ر	م
٢٧	ر	م
٢٨	ر	م
٢٩	ر	م
٣٠	ر	م
٣١	ر	م
٣٢	ر	م
٣٣	ر	م
٣٤	ر	م
٣٥	ر	م
٣٦	ر	م
٣٧	ر	م
٣٨	ر	م

٣٩	ر	م
٤٠	ر	م
٤١	ر	م
٤٢	ر	م
٤٣	ر	م
٤٤	ر	م

Appendix 4A: Ethics for Voice Attribution Experiment (English version)

Dear participant

My name is Fatimah Almutrafi and I am a postgraduate student in the school of Education, Communication and language sciences at Newcastle University in the United Kingdom. I am currently doing a research project to complete my PhD. In this study, I focus on the effect of language on cognition, a principle known as 'linguistic relativity'.

Your participation is very important to the success of this research project. In this experiment, we are thinking of making a new movie in which some everyday objects come to life, sing and dance. You will see a series of pictures of these objects and will need to determine whether each item should have the voice of a man/boy or that of a woman or girl. If you decide that an object should have a female voice please circle "F" in the column named "VOICE" on your answer sheet. If you decide that it should have a male voice, then circle "M".

Please make sure that picture numbers correspond to the numbers on your answer sheet.

I would like to assure you that all your responses will be confidential and anonymous. The people who will have access to the data will be myself - the researcher - and my supervisors; Professor Vivian Cook and Dr. Panos Athanasopoulos. The data will be kept safe in an archive during the research period. All the data will be destroyed immediately after the completion of my PhD thesis.

If you agree to take part in this research, please sign the consent form.

Should you require additional information, please do not hesitate to contact me:

Fatimah Almutrafi, 25 Ascot Walk, Newcastle upon Tyne

United Kingdom; or via email: f.a.almutrafi@ncl.ac.uk.

Your participation is greatly appreciated

Fatimah Almutrafi

Appendix 4B: Ethics for Voice Attribution Experiment (Arabic version)

Fatimah Almutrafi

25 Ascot Walk

Newcastle upon Tyne

NE3 2UF, United Kingdom

f.a.almutrafi@ncl.ac.uk.

بسم الله الرحمن الرحيم

عزيزي/عزيزتي المشارك/ة في هذه الاستبانة ... ،

السلام عليكم ورحمة الله وبركاته ...

أحب في البداية أن أقدم لك تعريفا بنفسي. أنا فاطمة المطرفي وأدرس في جامعة نيوكاسل بالمملكة المتحدة في مرحلة الدكتوراة في تخصص اللغويات. أعكف حاليا على إعداد بحث كجزء من الدرجة العلمية التي أدرسها. يتمحور البحث مدى تأثير اللغة التي نتدحثها على عقولنا.

لمشاركتك في هذا البحث أهمية بالغة في نجاحه، و فكرة التجربة هي اننا نفكر بعمل فيلم جديد حيث أن كل الأشياء تصبح حية و تغني و ترقص. سترى سلسلة من الصور لهذه الأشياء و عليك أن تحدد ما إذا كان صورة الشئ يجب أن نعطى صوت امرأة وبنت أو صوت رجل وولد.

إذا قررت أن الشئ في الصورة يجب أن يأخذ صوت رجل، أحط الحرف (ر) في العمود المسمى (صوت) في ورقة إحابتك. وإذا قررت أن الشئ في الصورة يجب أن يأخذ صوت مرأة، أحط الحرف (م).

من فضلك تأكد بأن رقم الصورة يتفق مع الأرقام الموجودة في ورقة الإجابة.

للمعلومية فإن هذه البيانات والمعلومات التي ستقوم بالإدلاء بها ستحظى بالسرية التامة، وأن مشاركتك فيها اختيارية بحتة ويمكنك الانسحاب من المشاركة في أي وقت دون إبداء الأسباب. وسوف لن يصل للمعلومات سوى أنا والمشرفين على بحثي وهم البروفسور فيفان كوك و الدكتور باناص اثناسبولس. وسوف يتم حفظ البيانات في مكان آمن خلال فترة البحث.

أرجو منك التكرم بتوقيع نسختك ونسختي من النموذج في الصفحة التالية حال موافقتك المشاركة، وإن كان لديك أي تساؤل أو أردت معلومات أكثر أرجو عدم التردد في مراسلتي على البريد الالكتروني **f.a.almutrafi@ncl.ac.uk** أو على عنواني الموجود في أعلى الصفحة.

شكرا لتعاونك معي للقيام بهذا البحث

فاطمة المطرفي

Appendix 5A: Consent Forms (English version)

Fatimah Almutrafi

25 Ascot Walk

Newcastle upon Tyne

NE3 2UF, United Kingdom

f.a.almutrafi@ncl.ac.uk.

(Researcher's Copy)

The purpose of this experiment is to see whether or not the language we speak has an effect on the way we think. Your participation is voluntary and you can withdraw at any point without giving any explanation. Your responses and identity will remain confidential and anonymous at all times. The data will be kept safe in an archive during the research period.

AGREEMENT

I agree to participate in this study and that the data I provide may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Researcher: _____

Signature of Participant: _____

Date: _____

Fatimah Almutrafi

25 Ascot Walk

Newcastle upon Tyne

NE3 2UF, United Kingdom

f.a.almutrafi@ncl.ac.uk.

(Participant's Copy)

The purpose of this experiment is to see whether or not the language we speak has an effect on the way we think. Your participation is voluntary and you can withdraw at any point without giving any explanation. Your responses and identity will remain confidential and anonymous at all times. The data will be kept safe in an archive during the research period.

AGREEMENT

I agree to participate in this study and that the data I provide may be:

1. Held in Newcastle University archives.
2. Made available to bona fide researchers.
3. May be quoted in published work or used in public performance in full or in part.
4. Used for teaching purposes.

Signature of Researcher: _____

Signature of Participant: _____

Date: _____

Appendix 5B: Consent Forms (Arabic version)

Fatimah Almutrafi

25 Ascot Walk

Newcastle upon Tyne

NE3 2UF, United Kingdom

f.a.almutrafi@ncl.ac.uk.

بسم الله الرحمن الرحيم

نموذج الموافقة على المشاركة في البحث

(نسخة الباحث)

الهدف من مشاركتك في هذا البحث هو معرفة رأيك وماتعتقده بخصوص موضوع البحث وهو تأثير اللغة على العقل. تعد مشاركتك في هذا البحث اختيارية بحتة ولك الحق في الانسحاب في أي وقت دون إبداء الأسباب. سيتم التعامل مع هويتك كمشارك وإجاباتك دائما بسرية تامة، وسوف يتم حفظ البيانات المجمعة في مكان آمن طوال مدة البحث.

إقرار بالموافقة على المشاركة في البحث

أوافق على المشاركة في هذا البحث وكذلك على النقاط التالية:

١. قد يتم حفظ البيانات في أرشيف جامعة نيوكاسل.
٢. قد يتم توفير البيانات للباحثين الأخرين.
٣. قد يتم الاقتباس جزئيا أو كليا من البيانات في أعمال منشورة.
٤. قد يتم استخدام البيانات لأغراض تعليمية.

توقيع الباحث: _____

توقيع المشارك: _____

التاريخ: _____

Fatimah Almutrafi

25 Ascot Walk

Newcastle upon Tyne

NE3 2UF, United Kingdom

f.a.almutrafi@ncl.ac.uk.

بسم الله الرحمن الرحيم

نموذج الموافقة على المشاركة في البحث

(نسخة المشارك)

الهدف من مشاركتك في هذا البحث هو معرفة رأيك وماتعتقده بخصوص موضوع البحث وهو تأثير اللغة على العقل. تعد مشاركتك في هذا البحث اختيارية بحتة ولك الحق في الانسحاب في أي وقت دون إبداء الأسباب. سيتم التعامل مع هويتك كمشارك وإجاباتك دائما بسرية تامة، وسوف يتم حفظ البيانات المجمعة في مكان آمن طوال مدة البحث.

إقرار بالموافقة على المشاركة في البحث

أوافق على المشاركة في هذا البحث وكذلك على النقاط التالية:

١. قد يتم حفظ البيانات في أرشيف جامعة نيوكاسل.
٢. قد يتم توفير البيانات للباحثين الأخرين.
٣. قد يتم الاقتباس جزئيا أو كليا من البيانات في أعمال منشورة.
٤. قد يتم استخدام البيانات لأغراض تعليمية.

توقيع الباحث: _____

توقيع المشارك: _____

التاريخ: _____

Appendix 6A: Testing participants' knowledge of the grammatical gender of test items (English version)

Items	Grammatical Gender	
1. Grandmother	Masculine	Feminine
2. Vacuum cleaner	Masculine	Feminine
3. Ball	Masculine	Feminine
4. Flag	Masculine	Feminine
5. Key	Masculine	Feminine
6. Apple	Masculine	Feminine
7. Star	Masculine	Feminine
8. Chair	Masculine	Feminine
9. Bus	Masculine	Feminine
10. Pencil	Masculine	Feminine
11. Sun	Masculine	Feminine
12. Mushroom	Masculine	Feminine
13. Watermelon	Masculine	Feminine
14. Banana	Masculine	Feminine
15. Corn	Masculine	Feminine
16. Nose	Masculine	Feminine
17. Toothbrush	Masculine	Feminine
18. Book	Masculine	Feminine
19. Moon	Masculine	Feminine
20. Mouth	Masculine	Feminine
21. Spoon	Masculine	Feminine
22. House	Masculine	Feminine
23. Strawberry	Masculine	Feminine
24. Girl	Masculine	Feminine
25. Table	Masculine	Feminine

26. Car	Masculine	Feminine
27. Heart	Masculine	Feminine
28. Thumb	Masculine	Feminine
29. Basket	Masculine	Feminine
30. Dress	Masculine	Feminine
31. Tree	Masculine	Feminine
32. Arm	Masculine	Feminine
33. Bag	Masculine	Feminine
34. Boy	Masculine	Feminine
35. Flower	Masculine	Feminine
36. Mountain	Masculine	Feminine
37. Grandfather	Masculine	Feminine
38. Door	Masculine	Feminine
39. Ear	Masculine	Feminine
40. Lettuce	Masculine	Feminine
41. Traffic light	Masculine	Feminine
42. Eye	Masculine	Feminine
43. Hand	Masculine	Feminine
44. Head	Masculine	Feminine

Appendix 6B: Testing participants' knowledge of the grammatical gender of test items (Arabic version)

التصنيف النحوي للاسماء

الاسم	التصنيف النحوي	
١- عجوز	مذكر	مؤنث
٢- مكناسة	مذكر	مؤنث
٣- كورة	مذكر	مؤنث
٤- علم	مذكر	مؤنث
٥- مفتاح	مذكر	مؤنث
٦- تفاحة	مذكر	مؤنث
٧- نجمة	مذكر	مؤنث
٨- كرسي	مذكر	مؤنث
٩- باص	مذكر	مؤنث
١٠- قلم رصاص	مذكر	مؤنث
١١- شمس	مذكر	مؤنث
١٢- مشروم	مذكر	مؤنث
١٣- جج	مذكر	مؤنث
١٤- موزة	مذكر	مؤنث
١٥- ذرة	مذكر	مؤنث
١٦- أنف	مذكر	مؤنث
١٧- فرشاة أسنان	مذكر	مؤنث
١٨- كتاب	مذكر	مؤنث
١٩- قمر	مذكر	مؤنث
٢٠- فم	مذكر	مؤنث
٢١- ملعقة	مذكر	مؤنث
٢٢- منزل	مذكر	مؤنث
٢٣- فراولة	مذكر	مؤنث
٢٤- فتاة	مذكر	مؤنث
٢٥- طاولة	مذكر	مؤنث
٢٦- سيارة	مذكر	مؤنث
٢٧- - قلب	مذكر	مؤنث
٢٨- ابهام	مذكر	مؤنث
٢٩- سلة	مذكر	مؤنث
٣٠- فستان	مذكر	مؤنث
٣١- شجرة	مذكر	مؤنث
٣٢- ذراع	مذكر	مؤنث
٣٣- شنطة	مذكر	مؤنث
٣٤- ولد	مذكر	مؤنث
٣٥- زهرة	مذكر	مؤنث
٣٦- جبل	مذكر	مؤنث
٣٧- شايب	مذكر	مؤنث
٣٨- باب	مذكر	مؤنث
٣٩- اذن	مذكر	مؤنث
٤٠- خس	مذكر	مؤنث
٤١- اشارة	مذكر	مؤنث

٤٢- عين	مذكر	مؤنث
٤٣- يد	مذكر	مؤنث
٤٤- راس	مذكر	مؤنث

Appendix 7: Actual data for the voice attribution task by monolingual speakers of Arabic and English

Group Statistics					
	Group of participant	N	Mean	Std. Deviation	Std. Error Mean
Total Test Item (40)	English monolinguals	30	22.73	2.53	.462
	Arabic monolinguals	30	33.60	4.56	.832
Feminine (20)	English monolinguals	30	10.96	2.29	.419
	Arabic monolinguals	30	17.60	2.19	.400
Masculine (20)	English monolinguals	30	11.76	2.26	.414
	Arabic monolinguals	30	16.00	3.09	.565
Masculine artificial (10)	English monolinguals	30	6.56	1.77	.324
	Arabic monolinguals	30	8.53	1.50	.274
Masculine Natural (10)	English monolinguals	30	5.20	1.34	.246
	Arabic monolinguals	30	7.46	1.92	.351
Feminine Artificial (10)	English monolinguals	30	4.33	1.80	.329
	Arabic monolinguals	30	8.60	1.27	.232
Feminine Natural (10)	English monolinguals	30	6.63	.99	.182
	Arabic monolinguals	30	9.00	1.364	.249

Appendix 8: Actual data for voice assignments by Arabic-English bilinguals (intermediate and advanced)

Group Statistics					
	Group of participant	N	Mean	Std. Deviation	Std. Error Mean
Total test items (40)	Intermediate bilinguals	30	25.73	3.382	.61762
	Advanced bilinguals	30	26.03	4.498	.82139
Masculine (20)	Intermediate bilinguals	30	12.76	2.192	.40024
	Advanced bilinguals	30	12.96	2.645	.48301
Feminine (20)	Intermediate bilinguals	30	12.96	2.428	.44330
	Advanced bilinguals	30	13.06	2.970	.54231
Masculine artificial (10)	Intermediate bilinguals	30	6.73	1.659	.30299
	Advanced bilinguals	30	7.16	1.288	.23530
Masculine Natural (10)	Intermediate bilinguals	30	6.03	1.325	.24204
	Advanced bilinguals	30	5.80	1.845	.33699
Feminine Artificial (10)	Intermediate bilinguals	30	6.03	1.731	.31617
	Advanced bilinguals	30	6.30	1.859	.33954
Feminine Natural (10)	Intermediate bilinguals	30	6.93	1.552	.28338
	Advanced bilinguals	30	6.76	1.675	.30582

Appendix 9A: Ethics for Similarity rating experiment (English version)

Dear participant

My name is Fatimah Almutrafi and I am a postgraduate student in the school of Education, Communication and language sciences at Newcastle University in the United Kingdom. I am currently doing a research project to complete my PhD. In this study, I focus on the effect of language on cognition, a principle known as 'linguistic relativity'.

Your participation is very important to the success of this research project. In this experiment, you will see pairs of pictures of different objects and your task is to rate the similarity between the two pictures on the seven-point scales in your answer sheet. Please note that '1' means not similar at all and '7' means very similar. Please make sure that picture numbers correspond to the numbers on your answer sheet.

I would like to assure you that all your responses will be confidential and anonymous. The people who will have access to the data will be myself - the researcher - and my supervisors: Professor Vivian Cook and Dr. Panos Athanasopoulos. The data will be kept safe in an archive during the research period. All the data will be destroyed immediately after the completion of my PhD thesis.

If you agree to take part in this research, please sign the consent form.

Should you require additional information, please do not hesitate to contact me:

Fatimah Almutrafi, 25 Ascot Walk, Newcastle upon Tyne

United Kingdom; or via email: f.a.almutrafi@ncl.ac.uk.

Your participation is greatly appreciated

Fatimah Almutrafi

Appendix 9B: Ethics for Similarity rating experiment (Arabic version)

Fatimah Almutrafi

25 Ascot Walk

Newcastle upon Tyne

NE3 2UF, United Kingdom

f.a.almutrafi@ncl.ac.uk.

بسم الله الرحمن الرحيم

عزيزي/عزيزتي المشارك/ة في هذه الاستبانة ... ،

السلام عليكم ورحمة الله وبركاته ...

أحب في البداية أن أقدم لك تعريفا بنفسي. أنا فاطمة المطرفي وأدرس في جامعة نيوكاسل بالمملكة المتحدة في مرحلة الدكتوراة في تخصص اللغويات. أعكف حاليا على إعداد بحث كجزء من الدرجة العلمية التي أدرسها. يتمحور البحث مدى تأثير اللغة التي نتدحثها على عقولنا.

لمشاركتك في هذا البحث أهمية بالغة في نجاحه، في هذه التجربة سوف ترى صورتين لشيئين مختلفين و كل ما عليك أن تقيم/ين التشابه في المعنى بين الصورتين المعروضتين على مقياس من سبع نقاط على ورقة اجابتك حيث أن ١ يعني غير متشابه على الاطلاق و ٧ تعني متشابه جدا. من فضلك تأكد بأن رقم الصور يتناسب مع الارقام في ورقة الاجابة.

للمعلومية فإن هذه البيانات والمعلومات التي ستقوم بالإدلاء بها ستحظى بالسرية التامة، وأن مشاركتك فيها اختيارية بحتة ويمكنك الانسحاب من المشاركة في أي وقت دون إبداء الأسباب. وسوف لن يصل للمعلومات سوى أنا والمشرفين على بحثي وهم البروفسور فيفان كوك و الدكتور بانص اثناسبولس. وسوف يتم حفظ البيانات في مكان آمن خلال فترة البحث.

أرجو منك التكرم بتوقيع نسختك ونسختي من النموذج في الصفحة التالية حال موافقتك المشاركة، وإن كان لديك أي تساؤل أو أردت معلومات أكثر أرجو عدم التردد في مراسلتي على البريد الالكتروني **f.a.almutrafi@ncl.ac.uk** أو على عنواني الموجود في أعلى الصفحة.

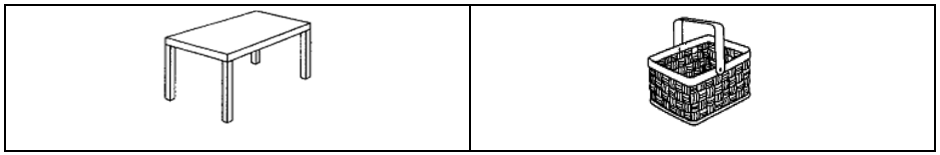
شكرا لتعاونك معي للقيام بهذا البحث

فاطمة المطرفي

Appendix 10A: Answer Sheet for Similarity Rating Experiment (English version)

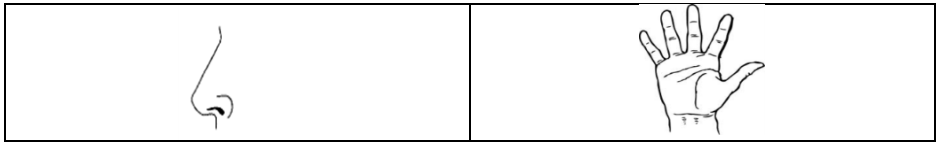
Rate the similarity between the pairs on the screen on the seven-point scale where '1' means not similar at all and '7' means very similar. Please make sure that picture numbers correspond to the numbers on your answer sheet.

1.



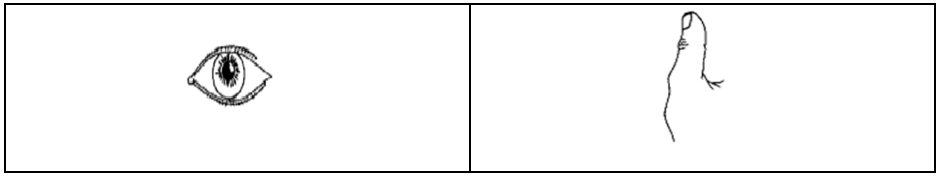
Not similar at all **1 2 3 4 5 6 7** **Very similar**

2.



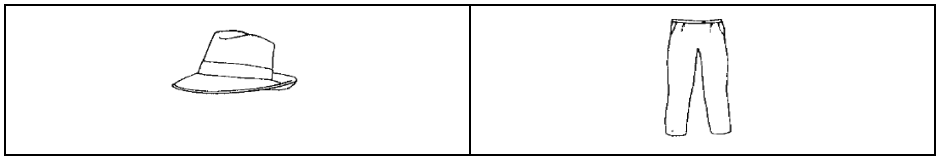
Not similar at all **1 2 3 4 5 6 7** **Very similar**

3.



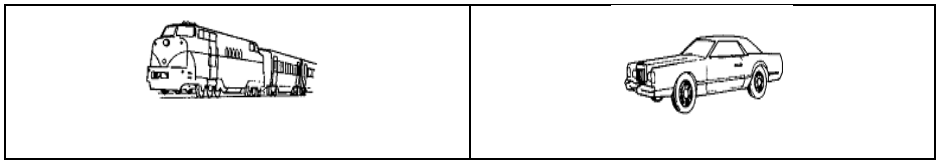
Not similar at all **1 2 3 4 5 6 7** **Very similar**

4.



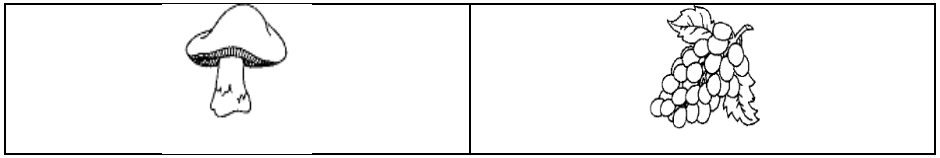
Not similar at all 1 2 3 4 5 6 7 **Very similar**

5.



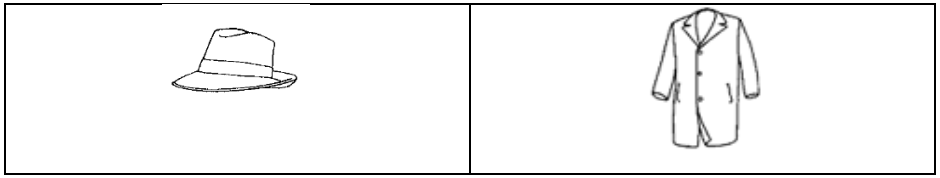
Not similar at all 1 2 3 4 5 6 7 **Very similar**

6.



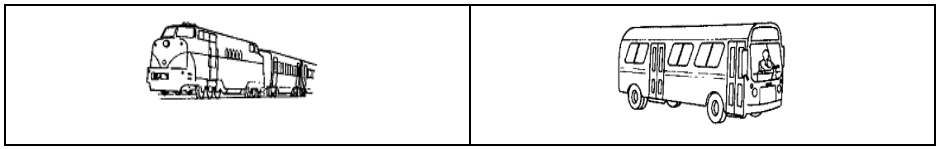
Not similar at all 1 2 3 4 5 6 7 **Very similar**

7.



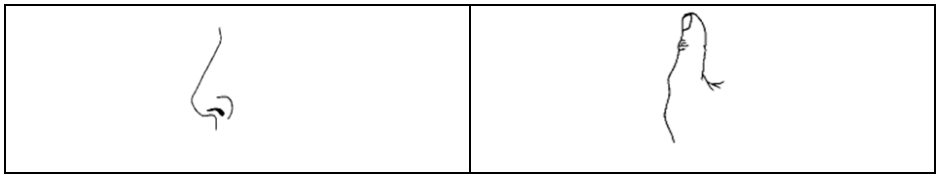
Not similar at all 1 2 3 4 5 6 7 **Very similar**

8.



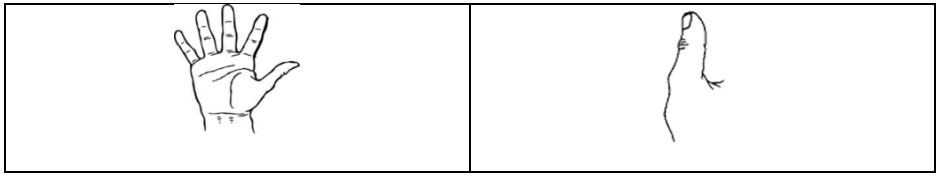
Not similar at all 1 2 3 4 5 6 7 **Very similar**

9.



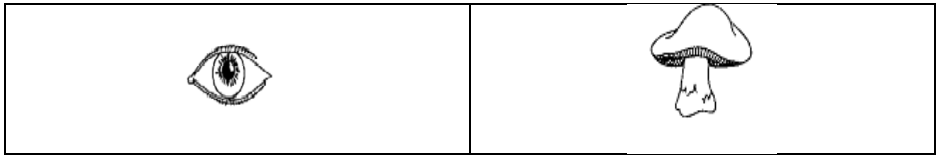
Not similar at all 1 2 3 4 5 6 7 **Very similar**

10.



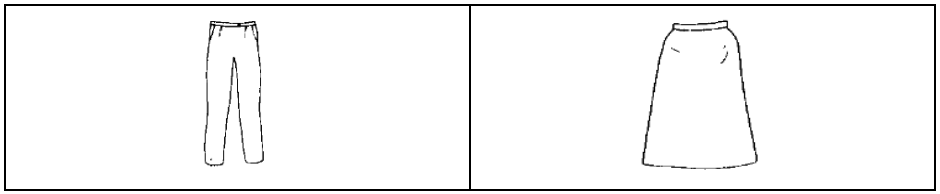
Not similar at all 1 2 3 4 5 6 7 **Very similar**

11.



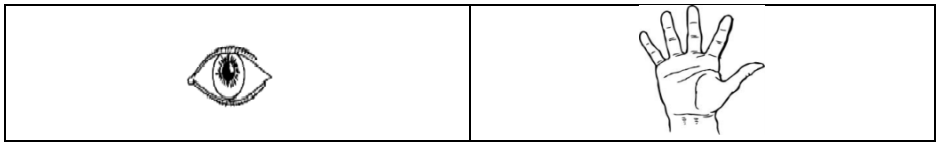
Not similar at all 1 2 3 4 5 6 7 **Very similar**

12.



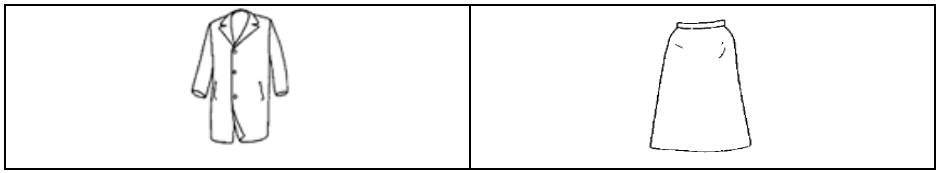
Not similar at all 1 2 3 4 5 6 7 **Very similar**

13.



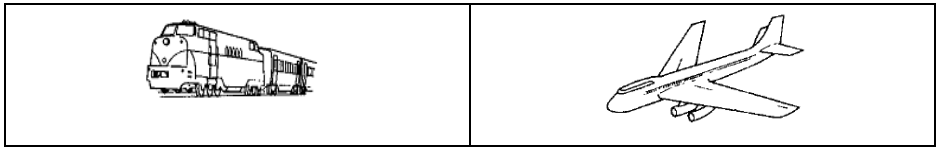
Not similar at all 1 2 3 4 5 6 7 **Very similar**

14.



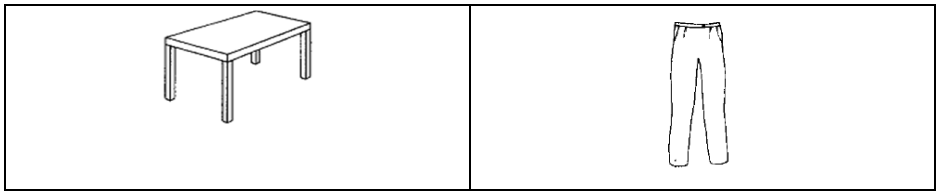
Not similar at all 1 2 3 4 5 6 7 **Very similar**

15.



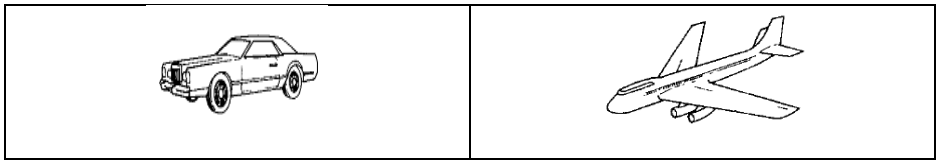
Not similar at all 1 2 3 4 5 6 7 **Very similar**

16.



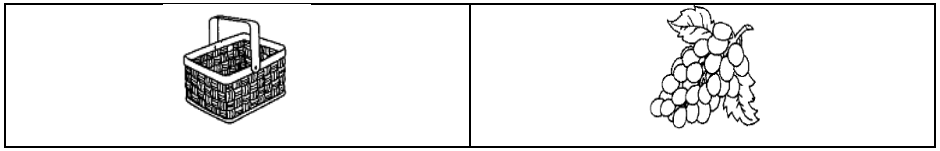
Not similar at all 1 2 3 4 5 6 7 **Very similar**

17.



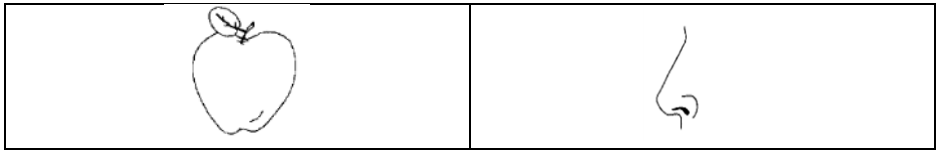
Not similar at all 1 2 3 4 5 6 7 **Very similar**

18.



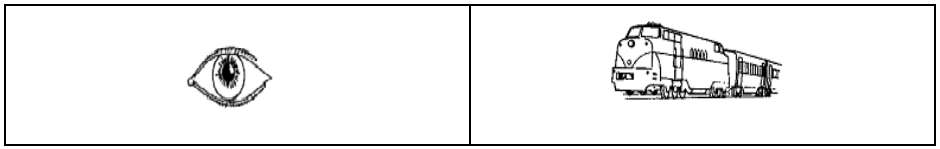
Not similar at all 1 2 3 4 5 6 7 **Very similar**

19.



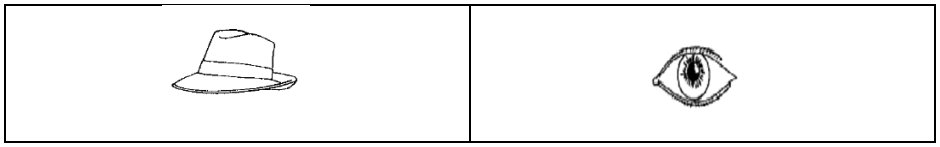
Not similar at all 1 2 3 4 5 6 7 **Very similar**

20.



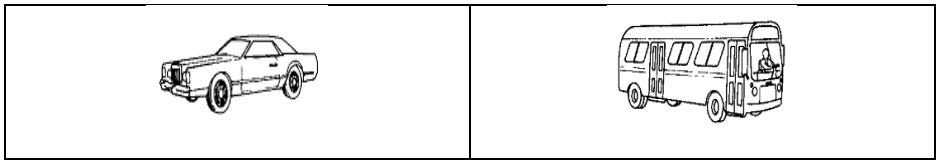
Not similar at all 1 2 3 4 5 6 7 **Very similar**

21.



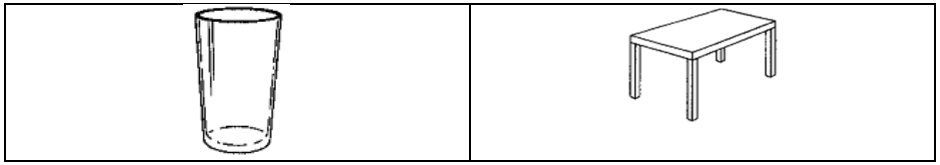
Not similar at all 1 2 3 4 5 6 7 **Very similar**

22.



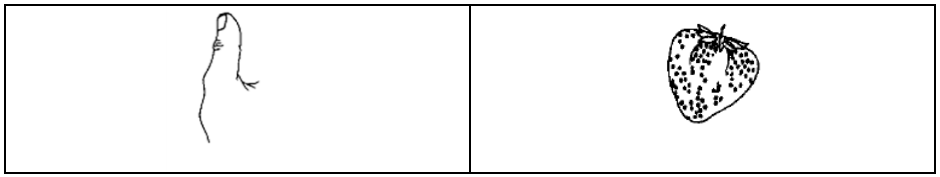
Not similar at all 1 2 3 4 5 6 7 **Very similar**

23.



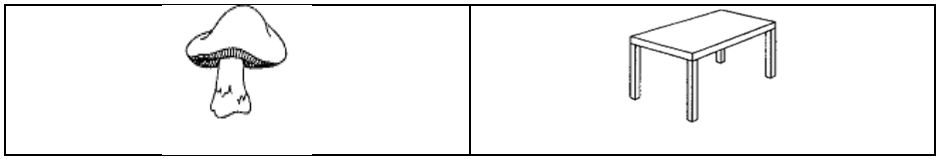
Not similar at all 1 2 3 4 5 6 7 **Very similar**

24.



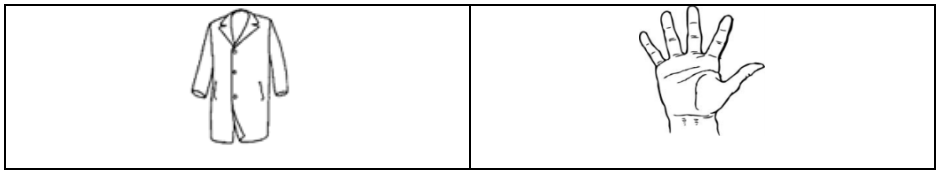
Not similar at all 1 2 3 4 5 6 7 **Very similar**

25.



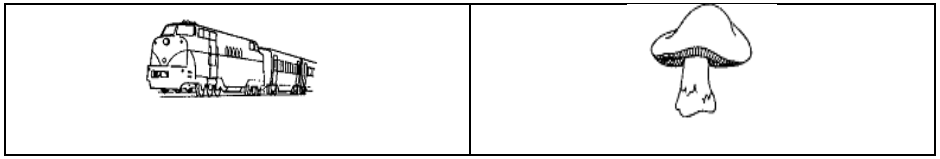
Not similar at all 1 2 3 4 5 6 7 **Very similar**

26.



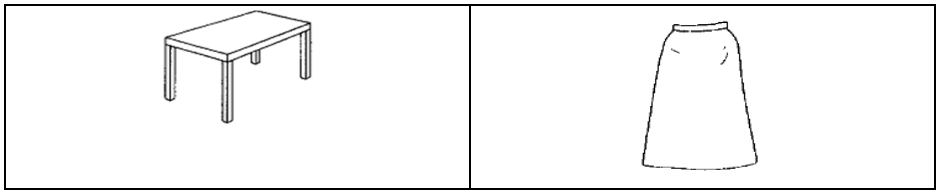
Not similar at all 1 2 3 4 5 6 7 **Very similar**

27.



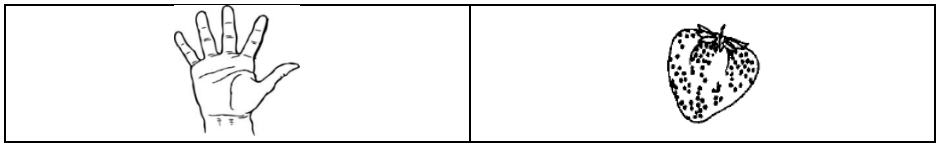
Not similar at all 1 2 3 4 5 6 7 **Very similar**

28.



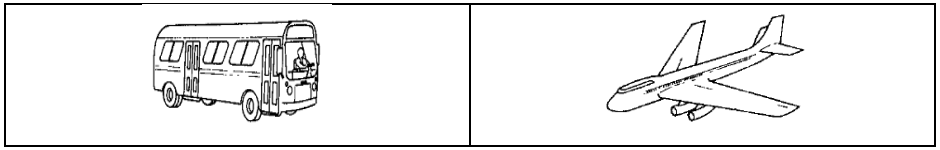
Not similar at all 1 2 3 4 5 6 7 **Very similar**

29.



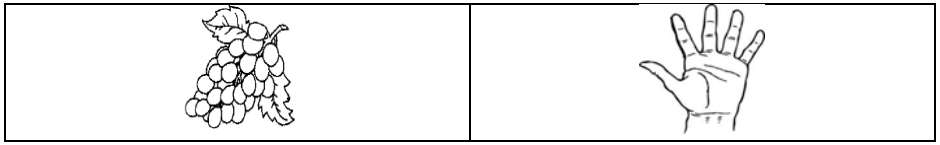
Not similar at all 1 2 3 4 5 6 7 **Very similar**

30.



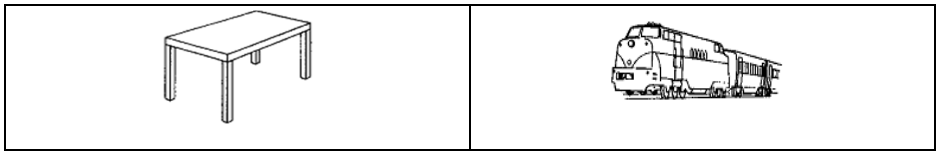
Not similar at all 1 2 3 4 5 6 7 **Very similar**

31.



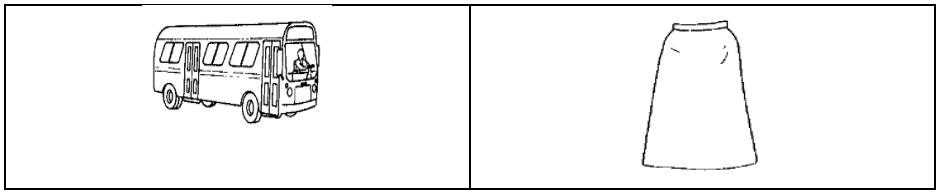
Not similar at all 1 2 3 4 5 6 7 **Very similar**

32.



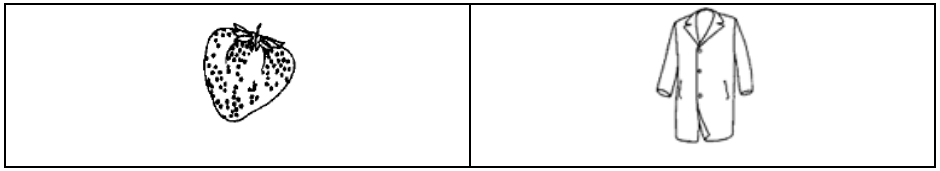
Not similar at all 1 2 3 4 5 6 7 **Very similar**

33.



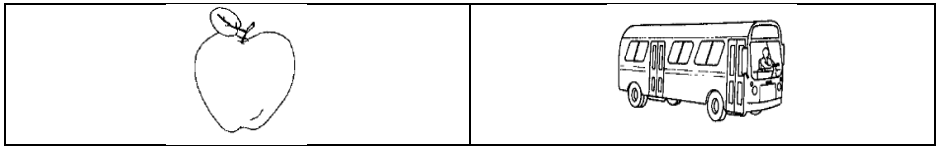
Not similar at all 1 2 3 4 5 6 7 **Very similar**

34.



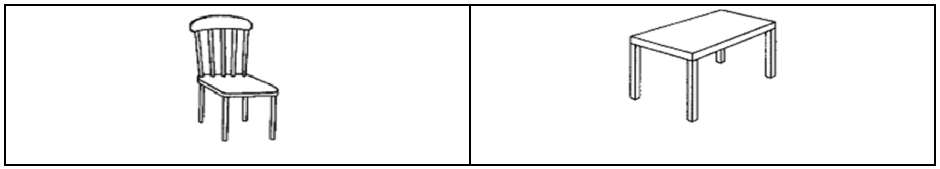
Not similar at all 1 2 3 4 5 6 7 **Very similar**

35.



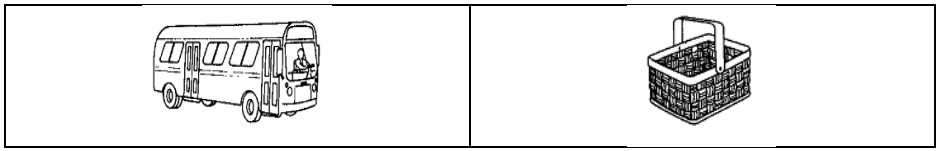
Not similar at all 1 2 3 4 5 6 7 **Very similar**

36.



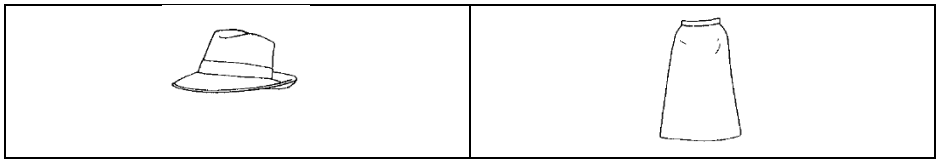
Not similar at all 1 2 3 4 5 6 7 **Very similar**

37.



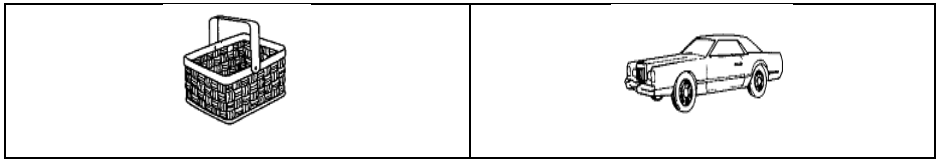
Not similar at all 1 2 3 4 5 6 7 **Very similar**

38.



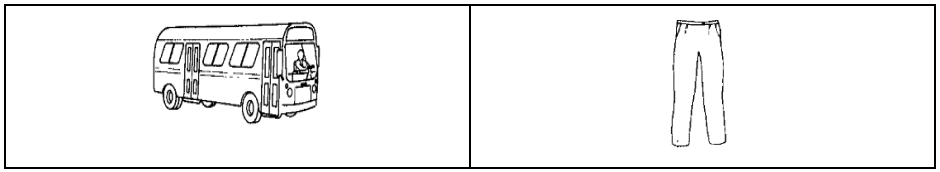
Not similar at all 1 2 3 4 5 6 7 **Very similar**

39.



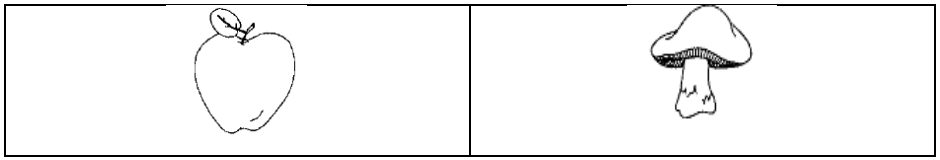
Not similar at all 1 2 3 4 5 6 7 **Very similar**

40.



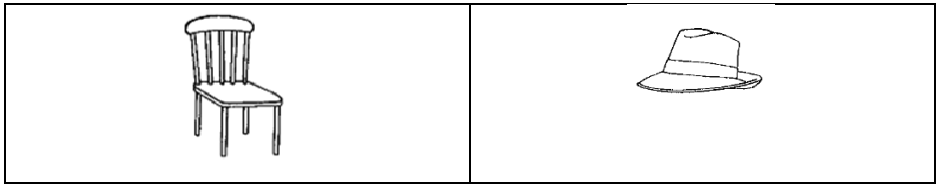
Not similar at all 1 2 3 4 5 6 7 **Very similar**

41.



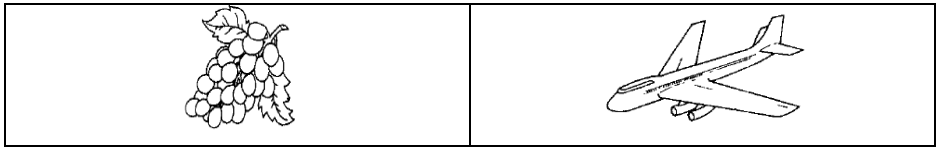
Not similar at all 1 2 3 4 5 6 7 **Very similar**

42.



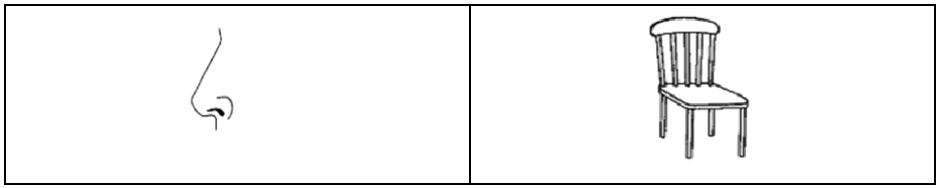
Not similar at all 1 2 3 4 5 6 7 **Very similar**

43.



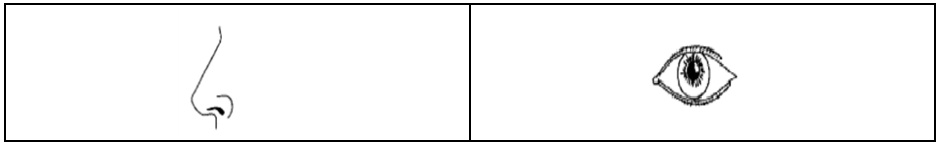
Not similar at all 1 2 3 4 5 6 7 **Very similar**

44.



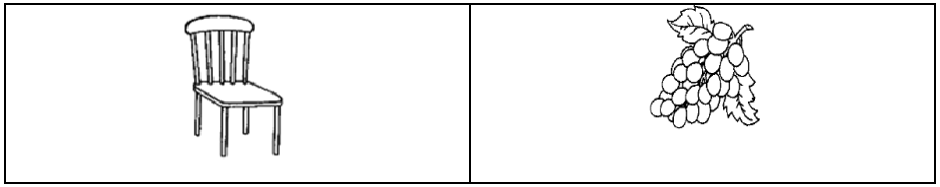
Not similar at all 1 2 3 4 5 6 7 **Very similar**

45.



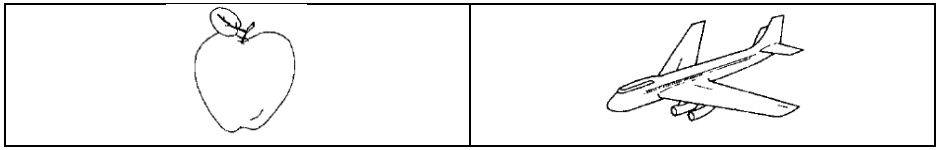
Not similar at all 1 2 3 4 5 6 7 **Very similar**

46.



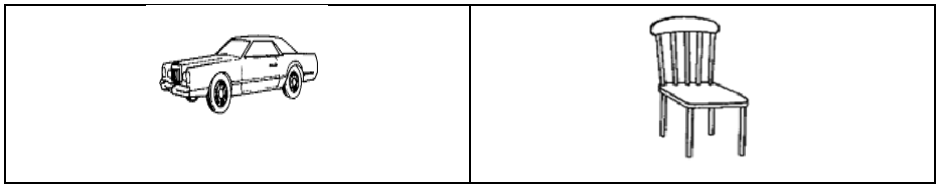
Not similar at all 1 2 3 4 5 6 7 **Very similar**

47.



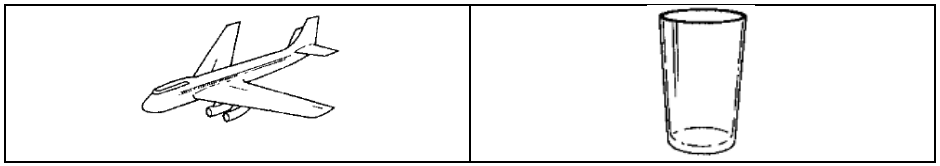
Not similar at all 1 2 3 4 5 6 7 **Very similar**

48.



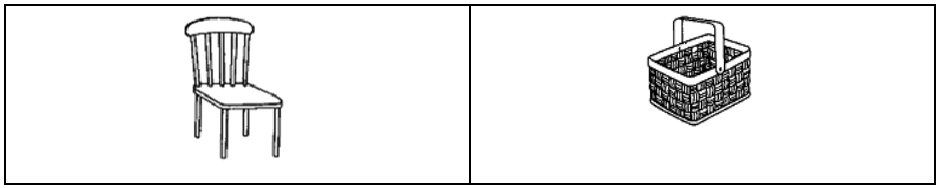
Not similar at all 1 2 3 4 5 6 7 **Very similar**

49.



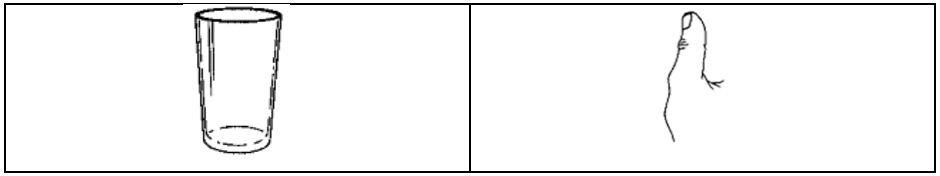
Not similar at all 1 2 3 4 5 6 7 **Very similar**

50.



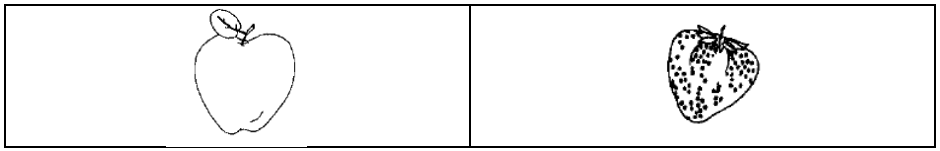
Not similar at all 1 2 3 4 5 6 7 **Very similar**

51.



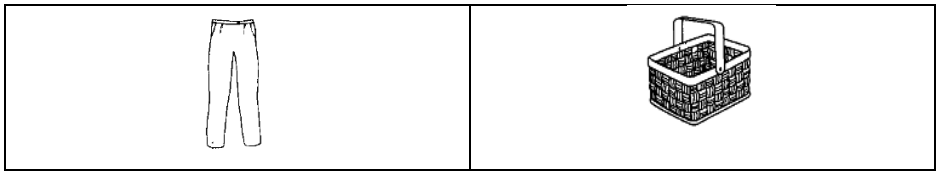
Not similar at all 1 2 3 4 5 6 7 **Very similar**

52.



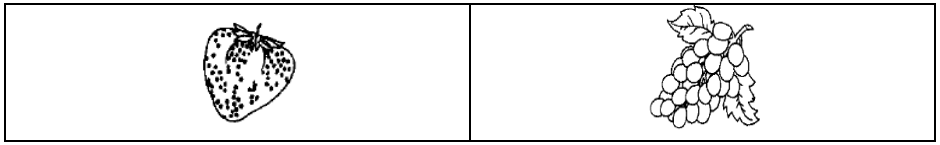
Not similar at all 1 2 3 4 5 6 7 **Very similar**

53.



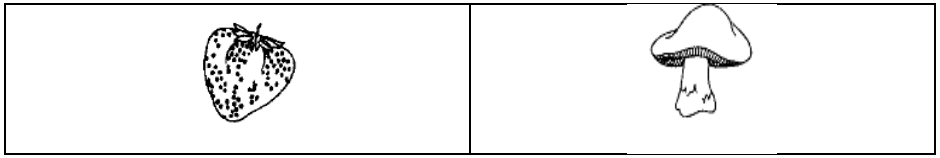
Not similar at all 1 2 3 4 5 6 7 **Very similar**

54.



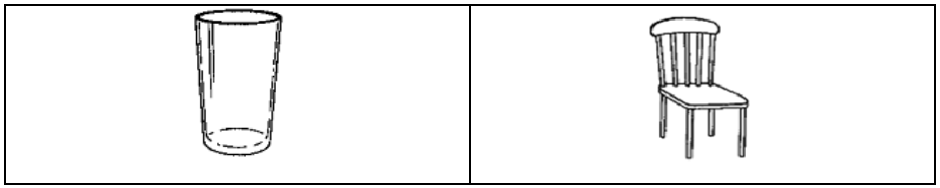
Not similar at all 1 2 3 4 5 6 7 **Very similar**

55.



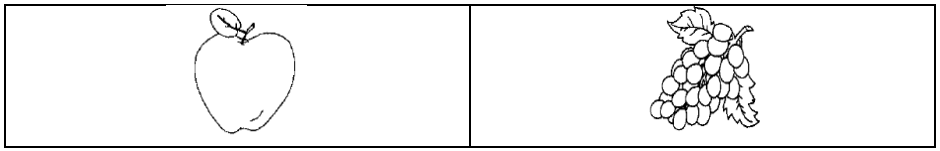
Not similar at all 1 2 3 4 5 6 7 **Very similar**

56.



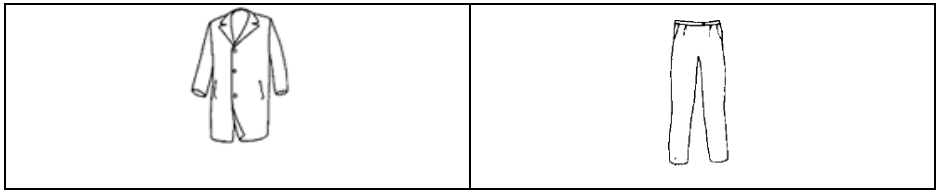
Not similar at all 1 2 3 4 5 6 7 **Very similar**

57.



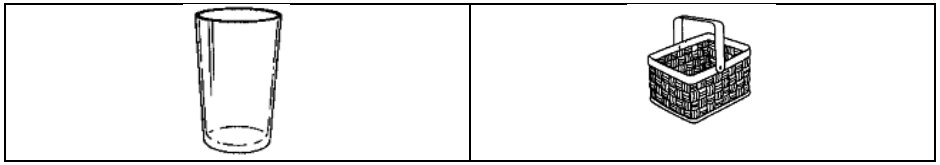
Not similar at all 1 2 3 4 5 6 7 **Very similar**

58.



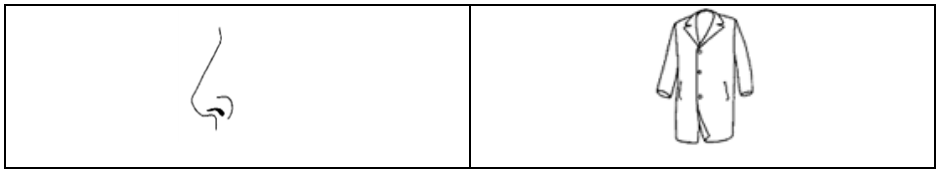
Not similar at all 1 2 3 4 5 6 7 **Very similar**

59.



Not similar at all 1 2 3 4 5 6 7 **Very similar**

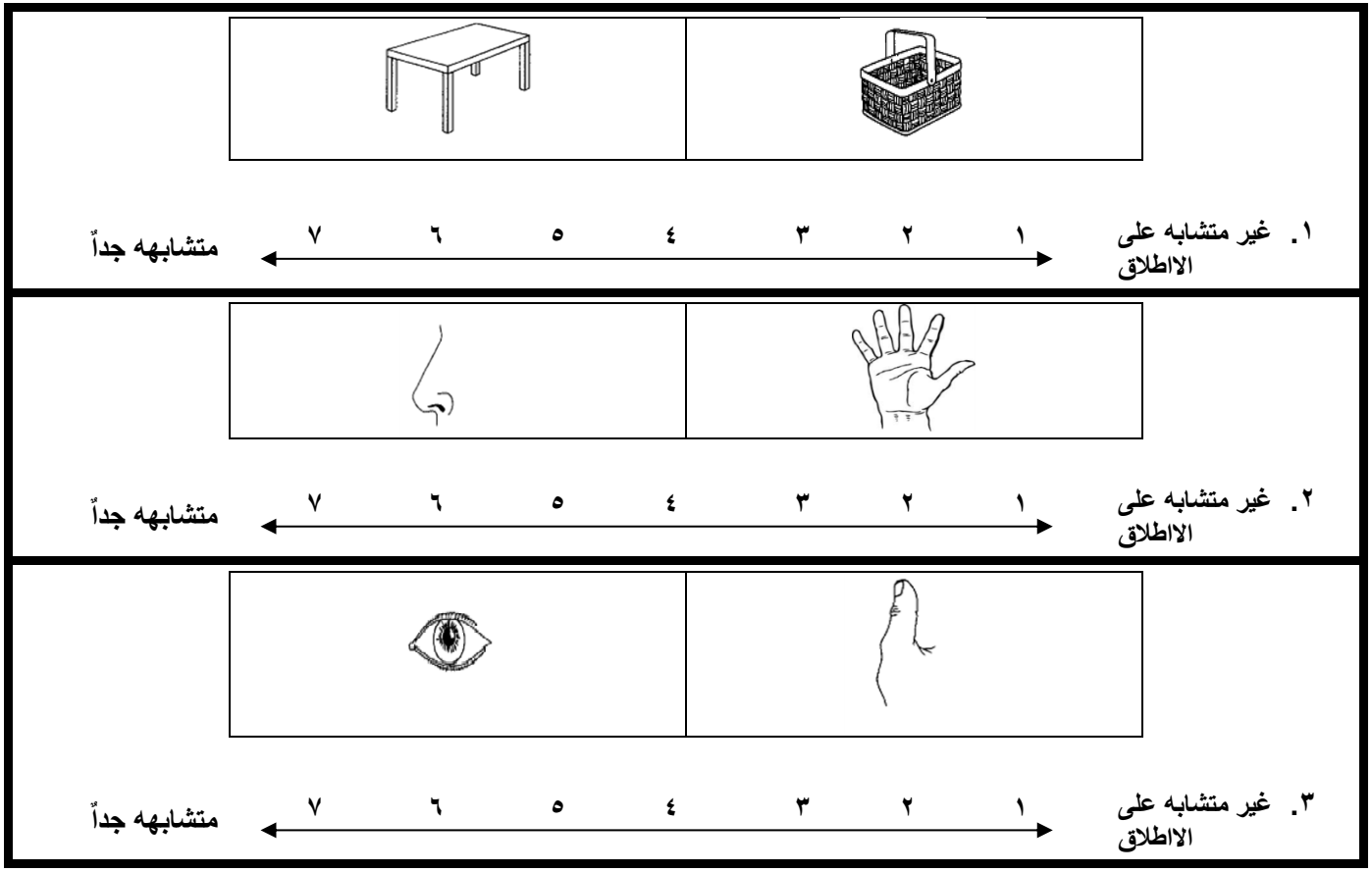
60.

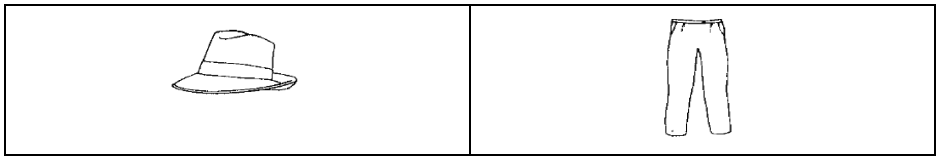


Not similar at all **1** **2** **3** **4** **5** **6** **7** **Very similar**

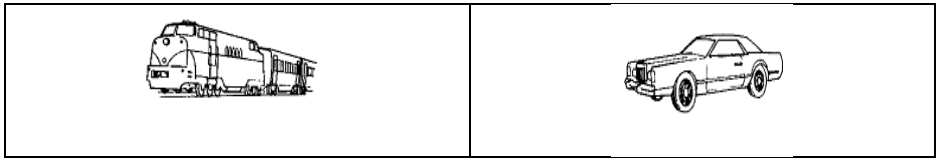
Appendix 10B: Answer Sheet for Similarity Rating Experiment (Arabic version)
ورقة الاجابة لتجربة تقييم التشابة

من فضلك تأكد **قيم التشابه بين الصورتين المعروضتين على الشاشة أمامك** على مقياس من سبع نقاط على ورقة اجابتك حيث أن ١ يعني غير متشابه على الاطلاق و ٧ تعني متشابه جدا.
بأن رقم الصور يتناسب مع الارقام في ورقة الاجابة

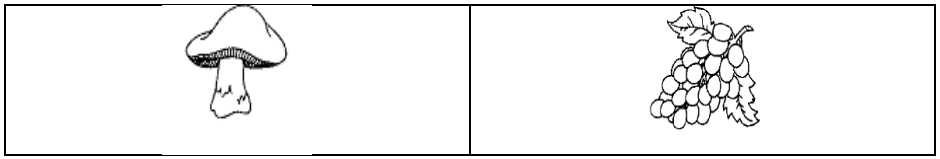




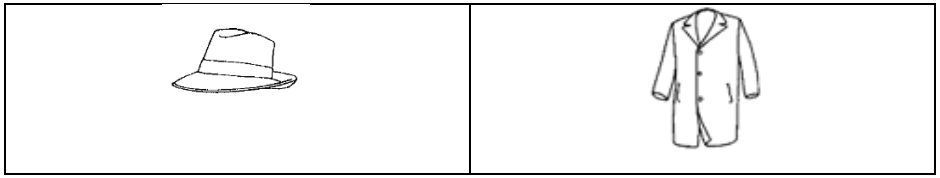
٤. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



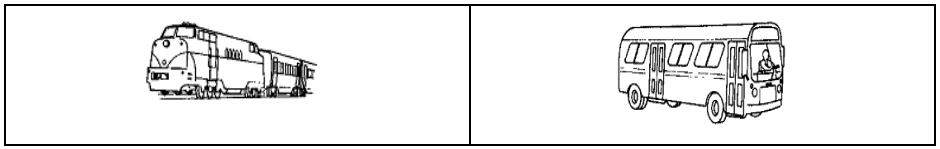
٥. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



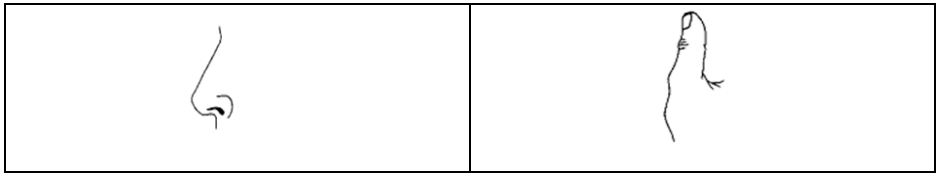
٦. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



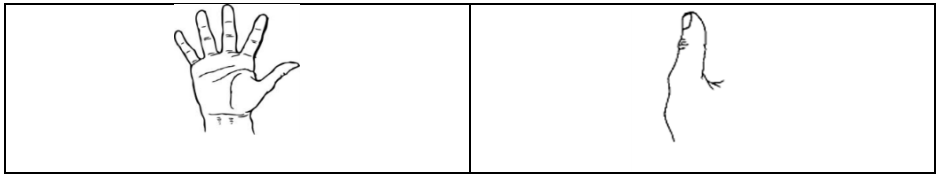
٧. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



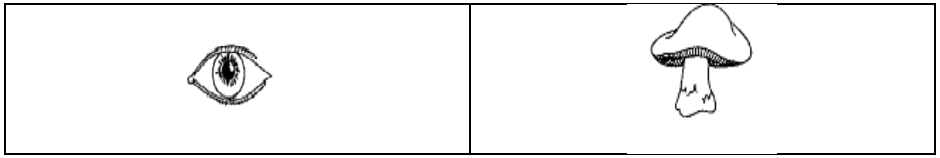
٨. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



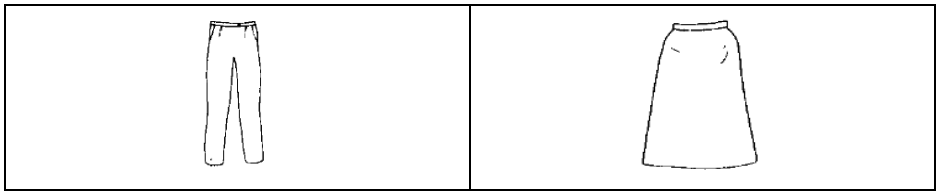
٩. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



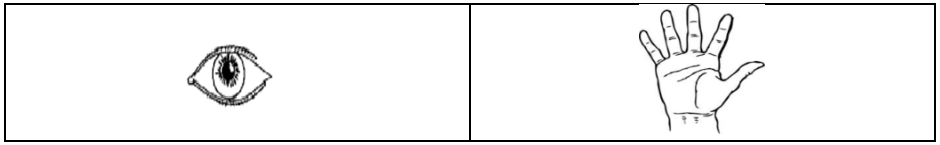
١٠. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



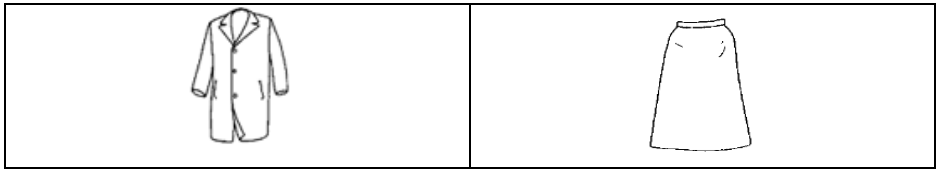
١١. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



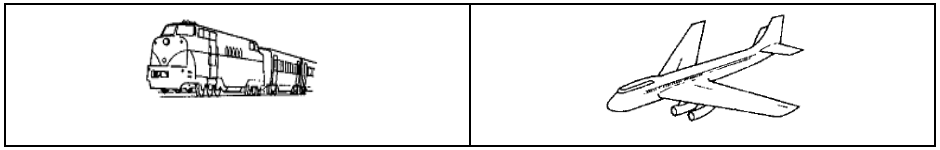
١٢. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



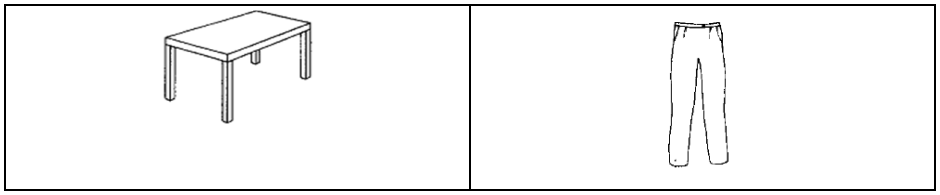
١٣. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



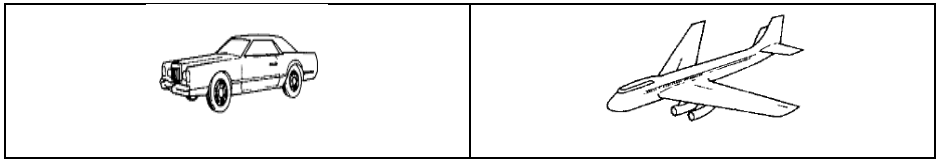
١٤. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



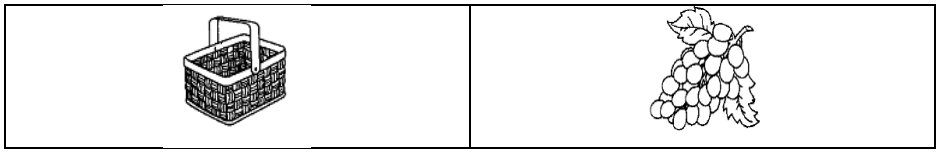
١٥. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



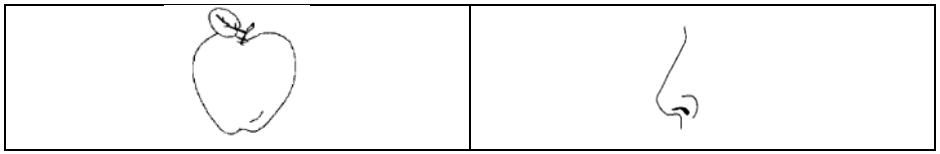
١٦. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



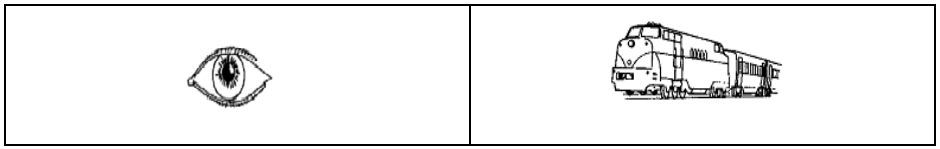
١٧. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



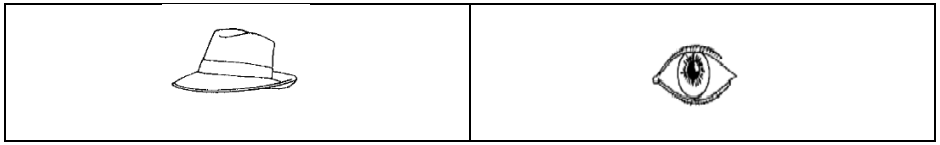
١٨. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



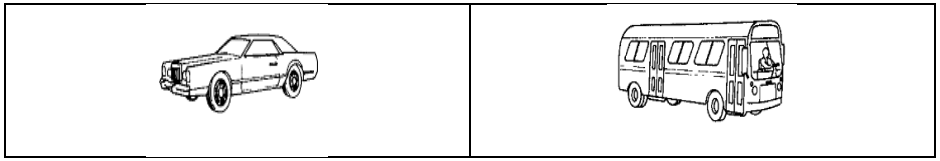
١٩. غير متشابه على الاطلاق ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



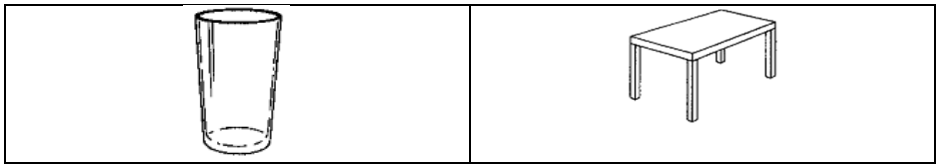
٢٠. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



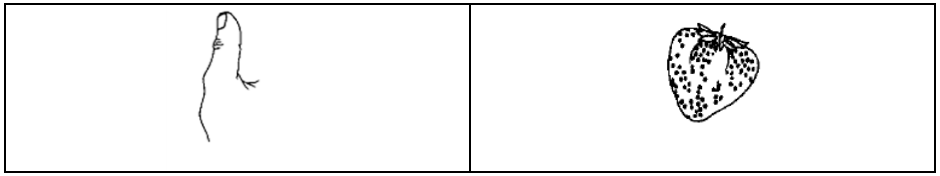
٢١. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



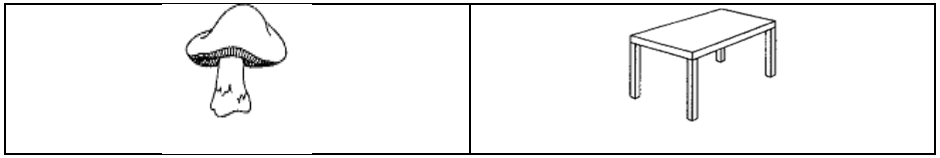
٢٢. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



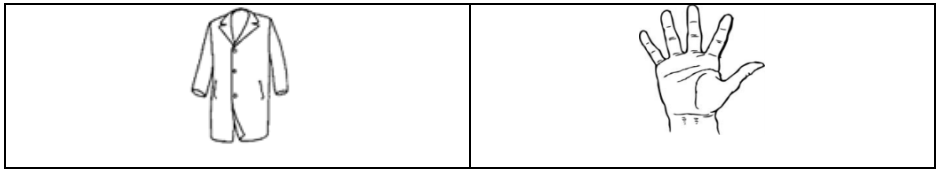
٢٣. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



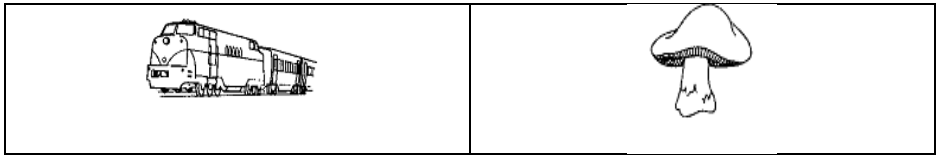
٢٤. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



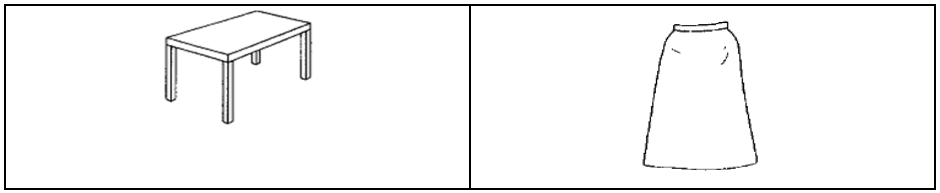
٢٥. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



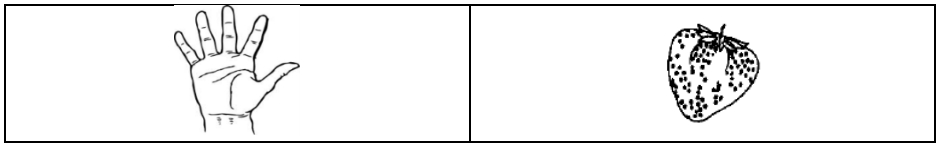
٢٦. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



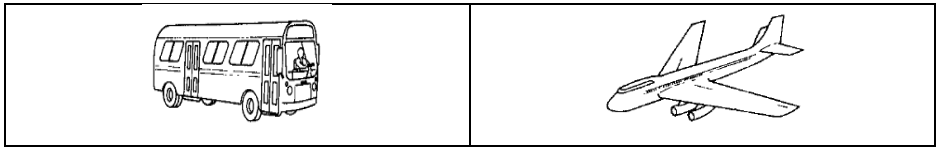
٢٧. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



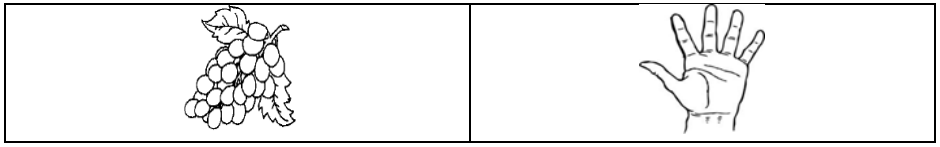
٢٨. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



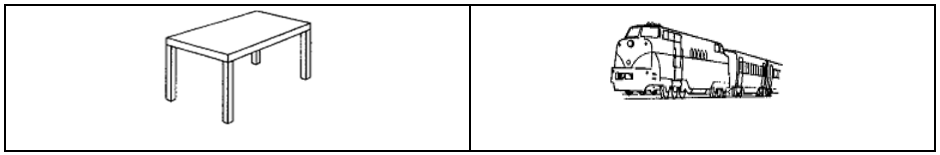
٢٩. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



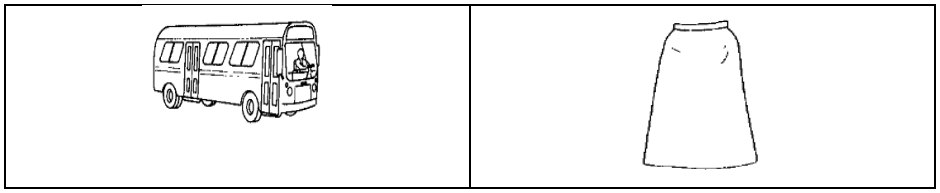
٣٠. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



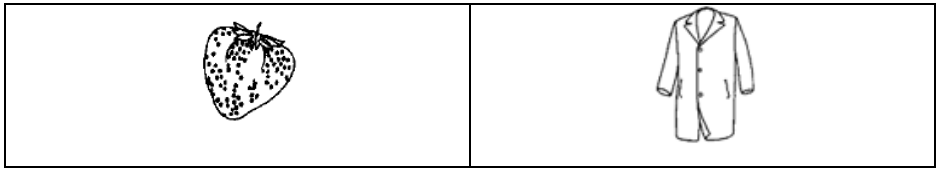
٣١. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



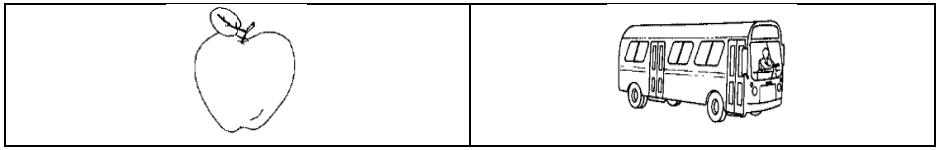
٣٢. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



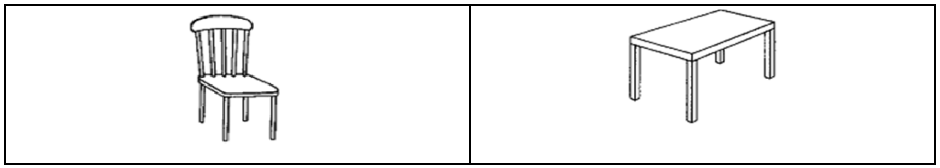
٣٣. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



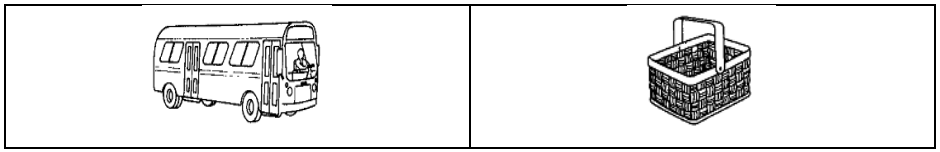
٣٤. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



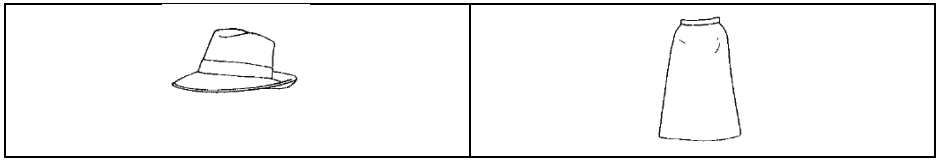
٣٥. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



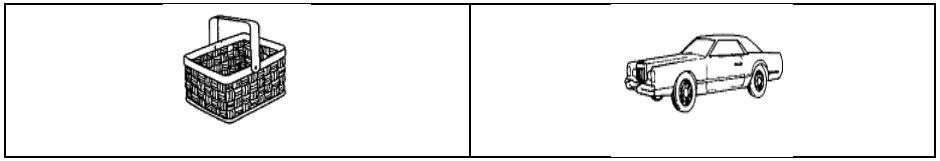
٣٦. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



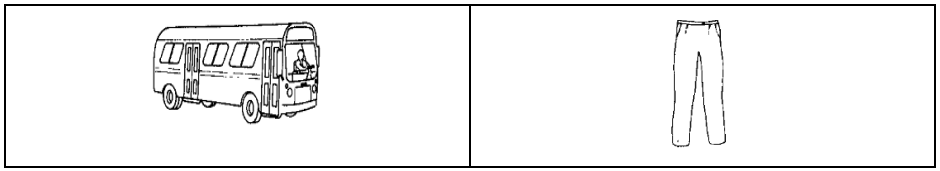
٣٧. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



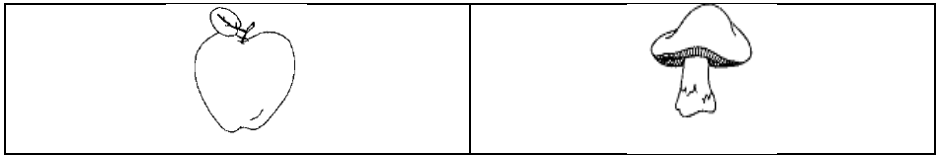
٣٨. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



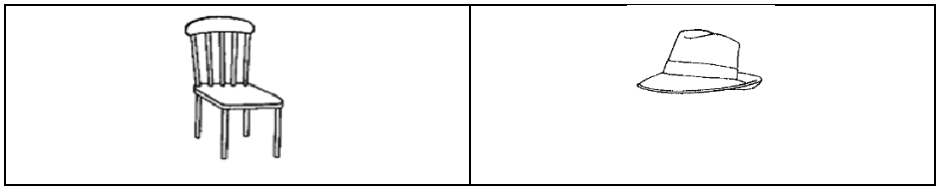
٣٩. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



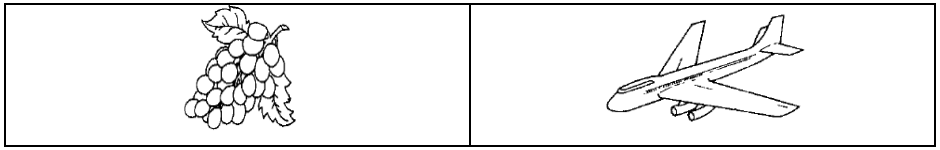
٤٠. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



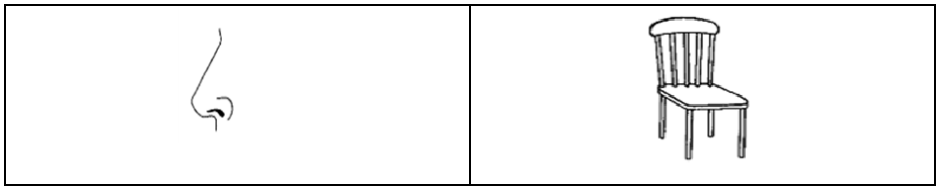
٤١. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



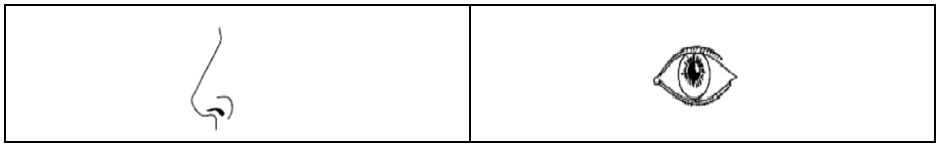
٤٢. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



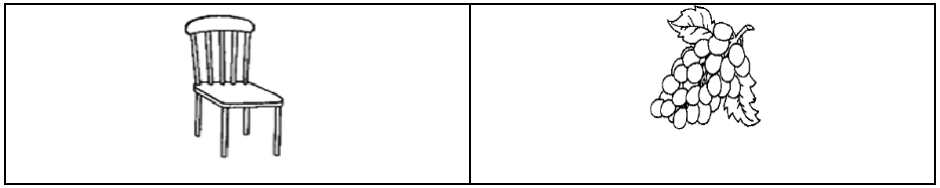
٤٣. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



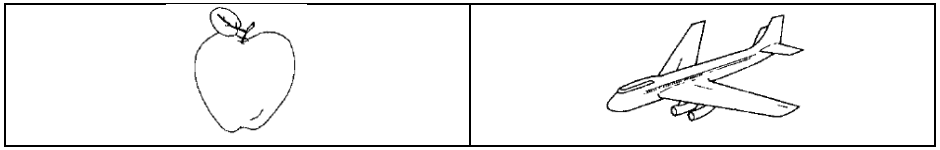
٤٤. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



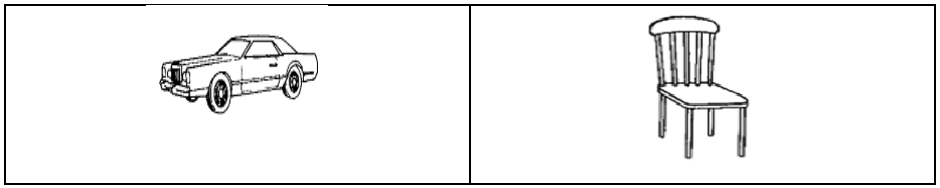
٤٥. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



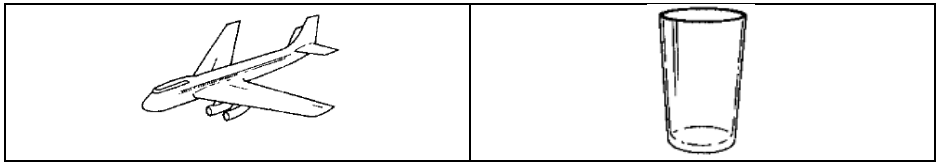
٤٦. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



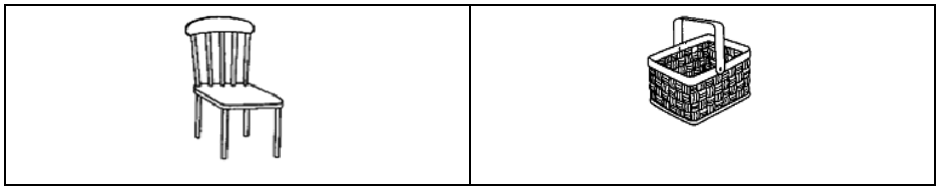
٤٧. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



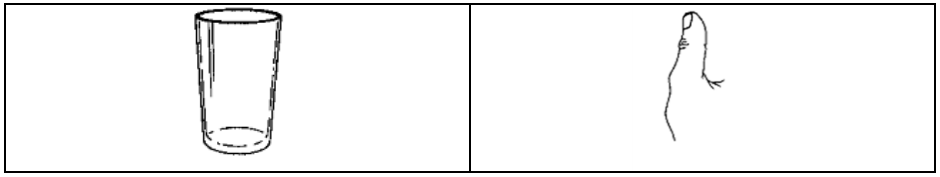
٤٨. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



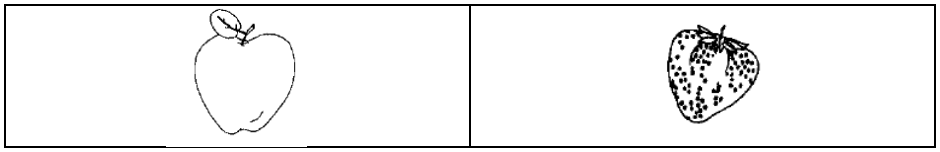
٤٩. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



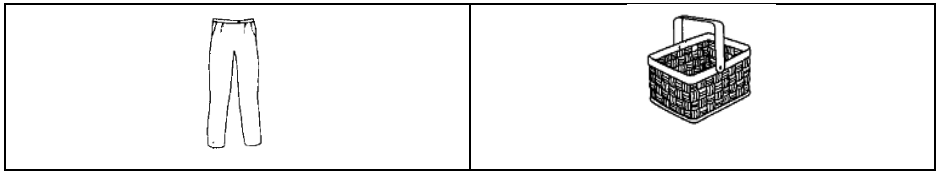
٥٠. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



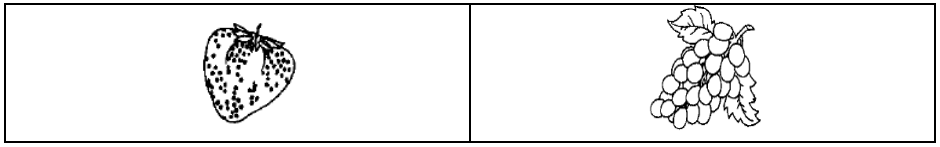
٥١. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



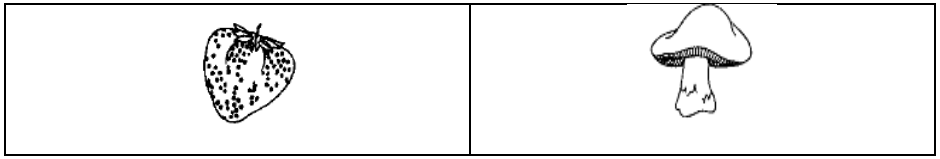
٥٢. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



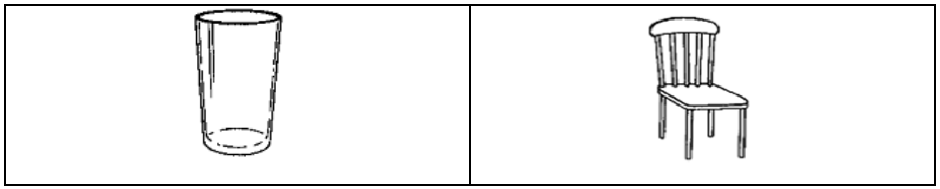
٥٣. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



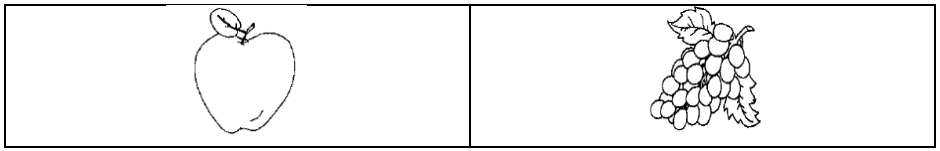
٥٤. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



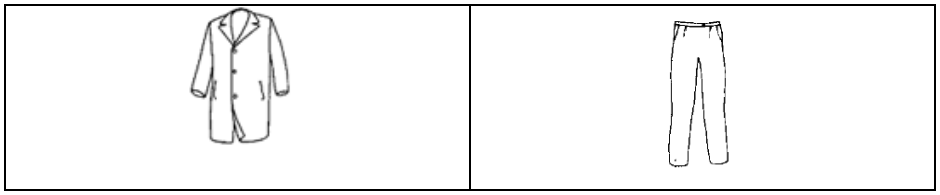
٥٥. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



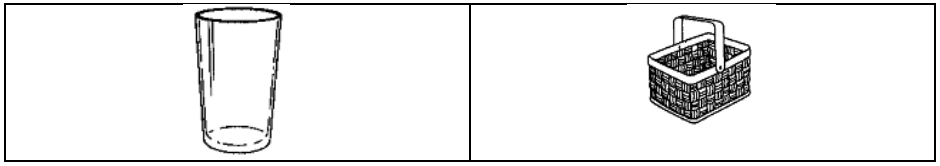
٥٦. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



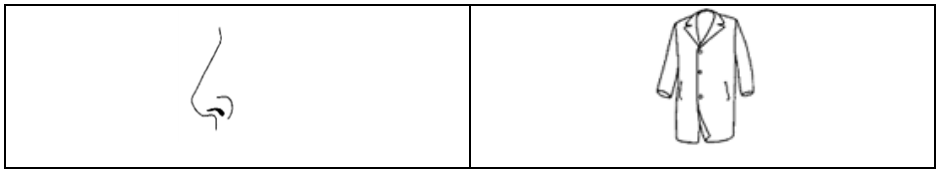
٥٧. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



٥٨. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



٥٩. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً



٦٠. غير متشابه على الاطلاق ١ ٢ ٣ ٤ ٥ ٦ ٧ متشابهه جداً

Appendix 11: Test of Normality

Descriptives					
	Group of participant			Statistic	Std. Error
All pair ratings	English monolinguals	Mean		7.0042	.16093
		95% Confidence Interval for Mean	Lower Bound	6.6750	
			Upper Bound	7.3333	
		5% Trimmed Mean		7.0171	
		Median		6.9125	
		Variance		.777	
		Std. Deviation		.88147	
		Minimum		4.75	
		Maximum		8.95	
		Range		4.20	
		Interquartile Range		.96	
		Skewness		-.201	.427
		Kurtosis		.689	.833
	Arabic monolinguals	Mean		6.8308	.25980
		95% Confidence Interval for Mean	Lower Bound	6.2995	
			Upper Bound	7.3622	
		5% Trimmed Mean		6.8176	
		Median		6.9000	
		Variance		2.025	
		Std. Deviation		1.42297	
		Minimum		3.88	
		Maximum		9.80	
		Range		5.93	
		Interquartile Range		1.83	
		Skewness		.265	.427
		Kurtosis		.095	.833
	Arabic English bilinguals	Mean		6.5933	.19499
		95% Confidence Interval for Mean	Lower Boun	6.1945	

			d		
			Upper Bound	6.9921	
		5% Trimmed Mean		6.5644	
		Median		6.6125	
		Variance		1.141	
		Std. Deviation		1.06799	
		Minimum		4.88	
		Maximum		8.88	
		Range		4.00	
		Interquartile Range		1.76	
		Skewness		.191	.427
		Kurtosis		-.664	.833

Tests of Normality

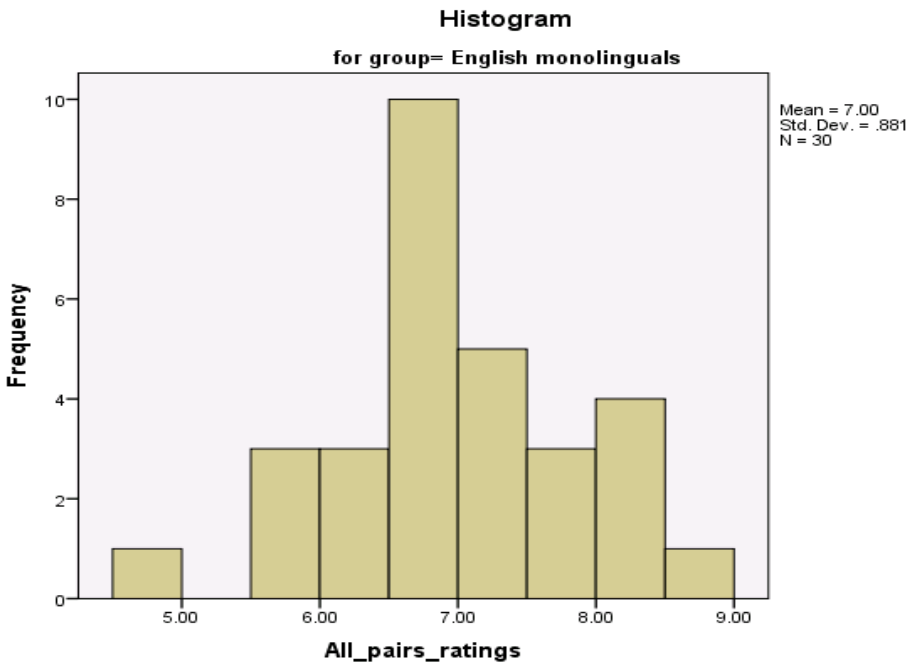
	Group of participant	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
All pairs ratings	English monolinguals	.132	30	.195	.979	30	.785
	Arabic monolinguals	.105	30	.200*	.968	30	.497
	Arabic English bilinguals	.101	30	.200*	.965	30	.401

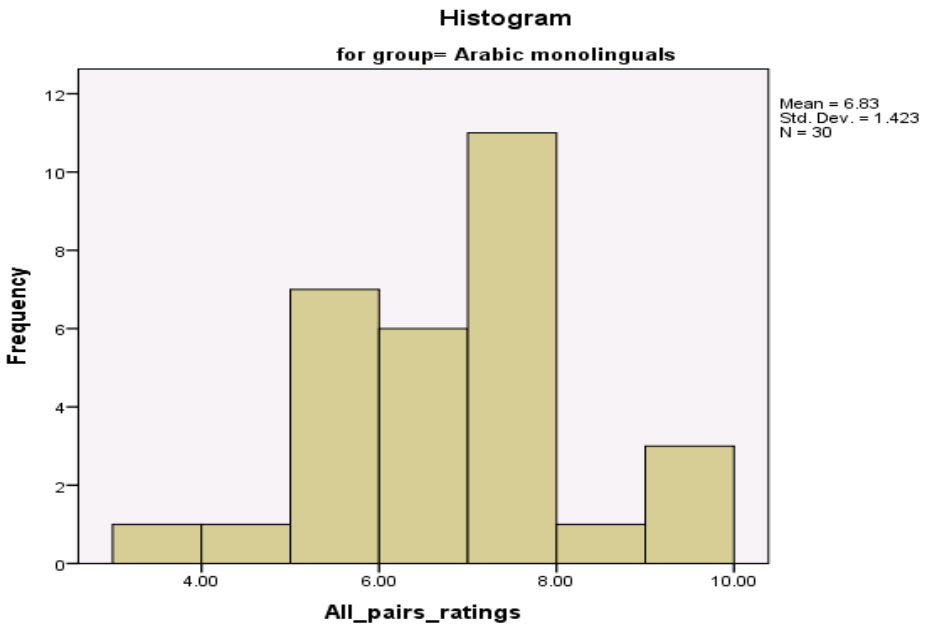
*. This is a lower bound of the true significance.

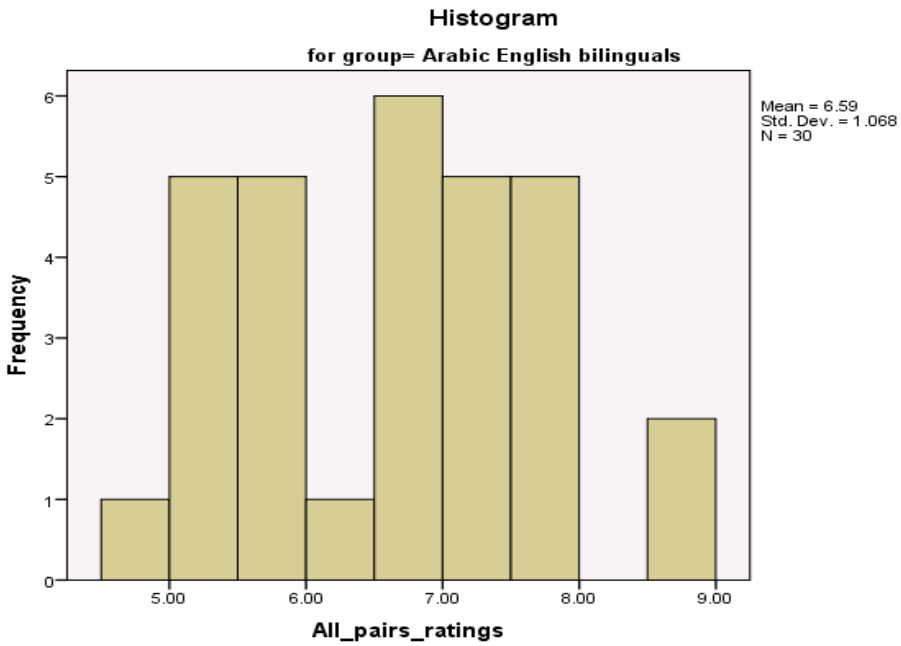
a. Lilliefors Significance Correction

All pair ratings

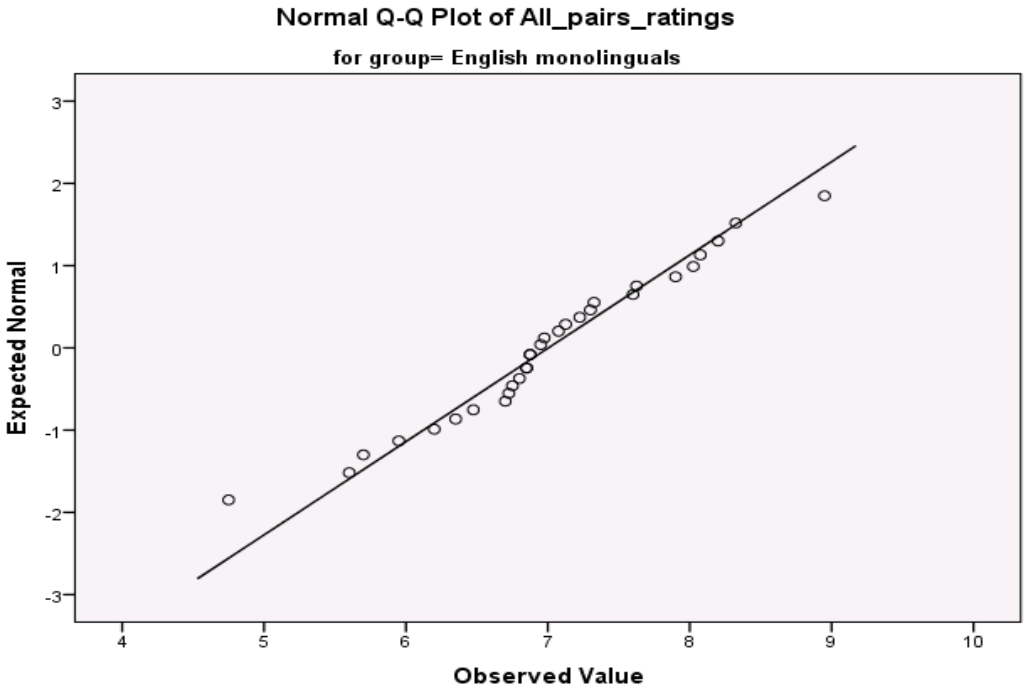
Histograms

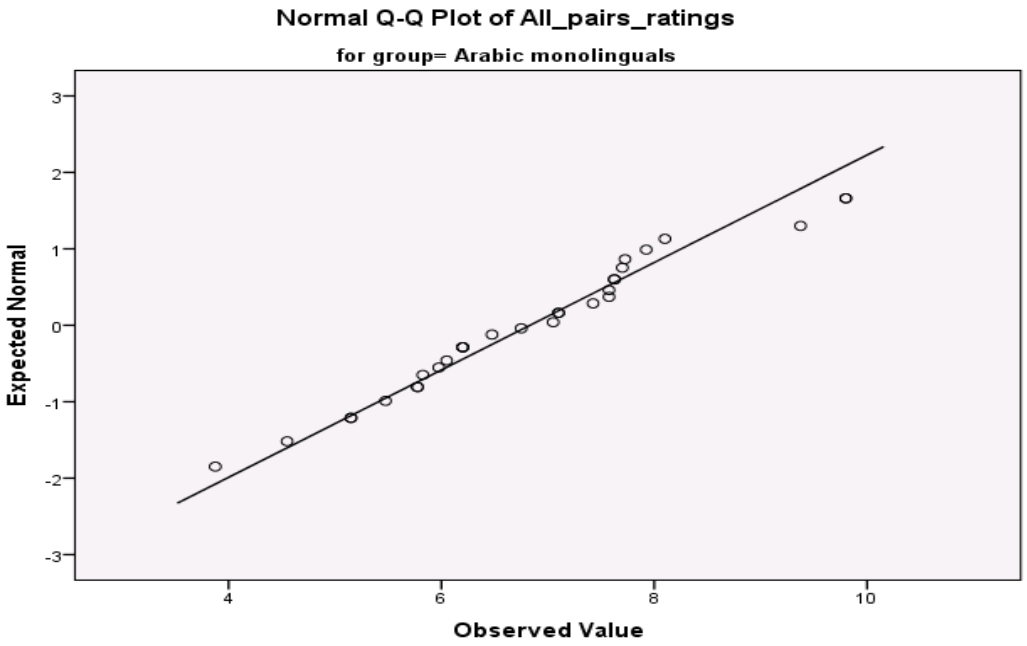


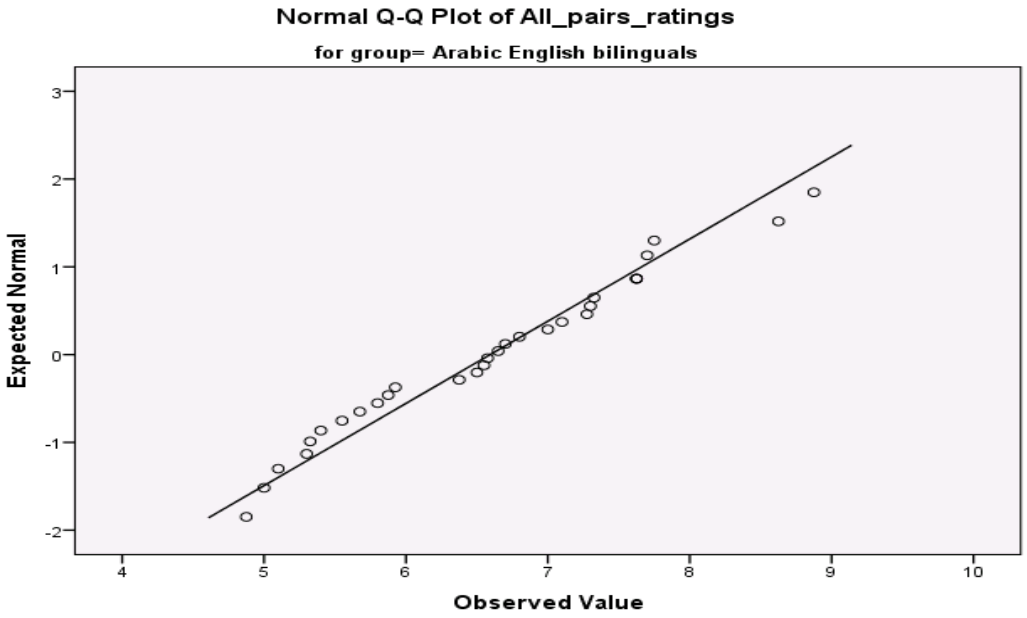




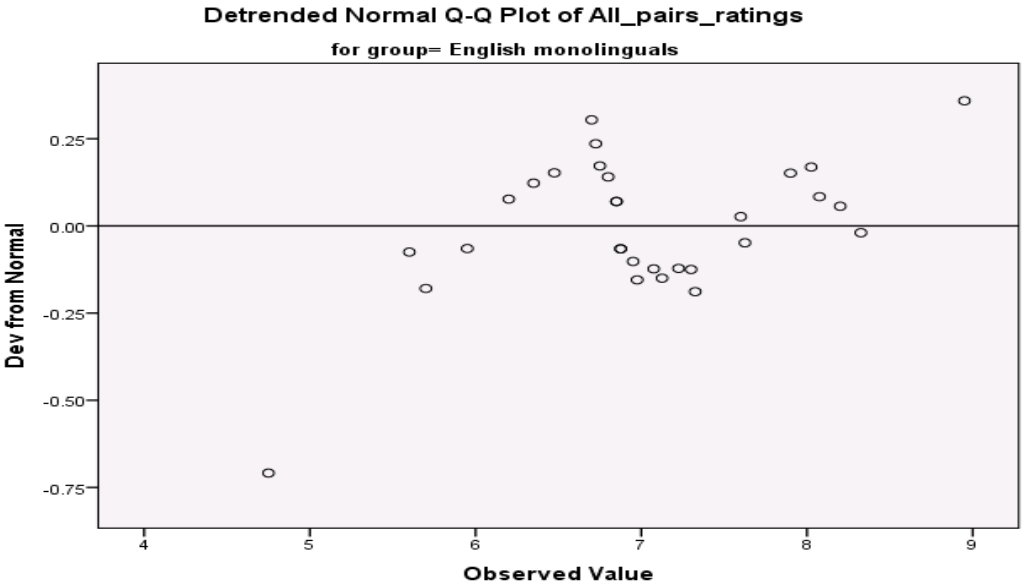
Normal Q-Q Plots

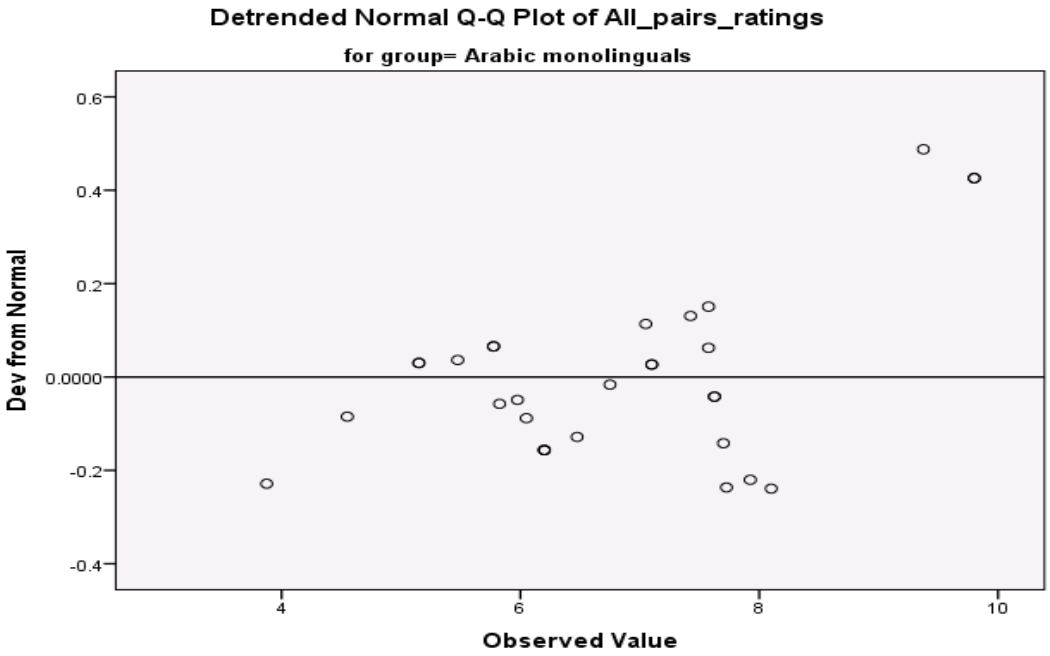


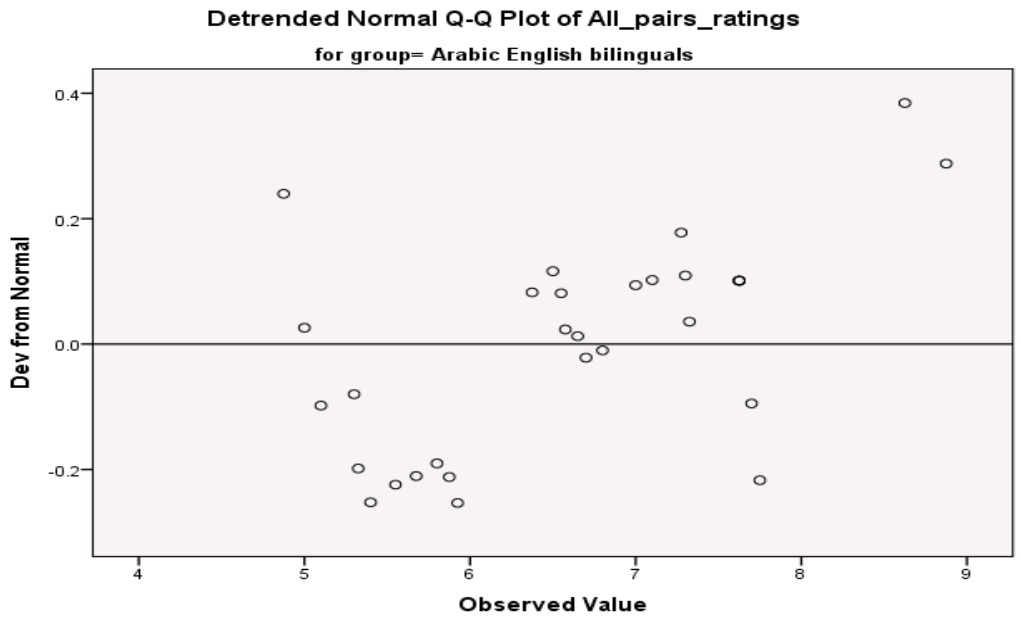

Normal Q-Q Plot of All_pairs_ratings
for group= Arabic monolinguals
Expected Normal
Observed Value
3
2
1
0
-1
-2
-3
4
6
8
10

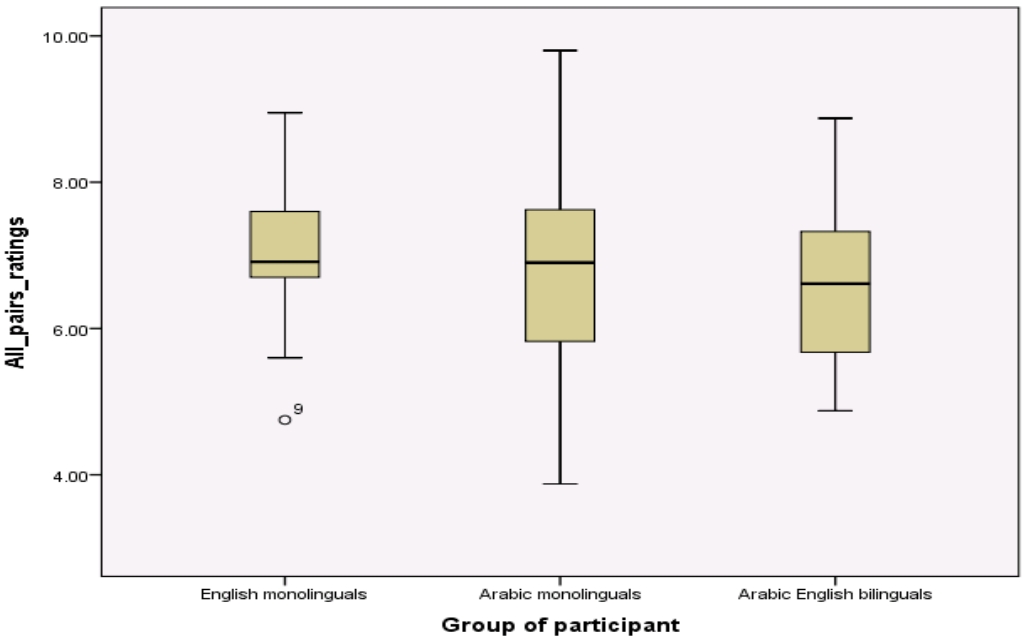

Normal Q-Q Plot of All_pairs_ratings
for group= Arabic English bilinguals
Expected Normal
Observed Value
3
2
1
0
-1
-2
4
5
6
7
8
9
10

Detrended Normal Q-Q Plots






Detrended Normal Q-Q Plot of All_pairs_ratings
for group= Arabic English bilinguals
Dev from Normal
0.4
0.2
0.0
-0.2
4
5
6
7
8
9
Observed Value


All_pairs_ratings
10.00
8.00
6.00
4.00
9
English monolinguals
Arabic monolinguals
Arabic English bilinguals
Group of participant

Analysis by Non Parametric Tests

NPar Tests

Descriptive Statistics					
	N	Mean	Std. Deviation	Minimum	Maximum
Mean of all pairs ratings	60	3.5178	.67780	1.88	4.98
Same Grammatical Gender Pairs	60	3.4442	.70699	2.10	5.25
Different Grammatical Gender Pairs	60	3.5546	.68489	1.78	4.95
Same Semantic Category Pairs	60	4.7344	1.03380	2.17	6.43
Different Semantic Category Pairs	60	2.3011	1.03463	1.00	6.17
Same Grammatical Gender Same Semantic Categories Pairs	60	4.6350	.96776	2.60	6.40
Same Grammatical Gender Different Semantic Categories Pairs	60	2.2533	1.09102	1.00	6.50
Different Grammatical Gender Same Semantic Categories Pairs	60	4.7842	1.10248	1.95	6.50
Different Grammatical Gender Different Semantic Categories Pairs	60	2.3250	1.02153	1.00	6.00
Group of participant	60	1.50	.504	1	2

Mann-Whitney Test

Ranks				
	Group of participant	N	Mean Rank	Sum of Ranks
Mean of all pairs ratings	English monolinguals	30	33.33	1000.00
	Arabic monolinguals	30	27.67	830.00
	Total	60		
Same Grammatical Gender Pairs	English monolinguals	30	31.68	950.50
	Arabic monolinguals	30	29.32	879.50

	Total	60		
Different Grammatical Gender Pairs	English monolinguals	30	34.07	1022.00
	Arabic monolinguals	30	26.93	808.00
	Total	60		
Same Semantic Category Pairs	English monolinguals	30	33.15	994.50
	Arabic monolinguals	30	27.85	835.50
	Total	60		
Different Semantic Category Pairs	English monolinguals	30	30.15	904.50
	Arabic monolinguals	30	30.85	925.50
	Total	60		
Same Grammatical Gender Same Semantic Categories Pairs	English monolinguals	30	31.92	957.50
	Arabic monolinguals	30	29.08	872.50
	Total	60		
Same Grammatical Gender Different Semantic Categories Pairs	English monolinguals	30	29.80	894.00
	Arabic monolinguals	30	31.20	936.00
	Total	60		
Different Grammatical Gender Same Semantic Categories Pairs	English monolinguals	30	34.18	1025.50
	Arabic monolinguals	30	26.82	804.50
	Total	60		
Different Grammatical Gender Different Semantic Categories Pairs	English monolinguals	30	29.85	895.50
	Arabic monolinguals	30	31.15	934.50
	Total	60		

Test Statistics[a]

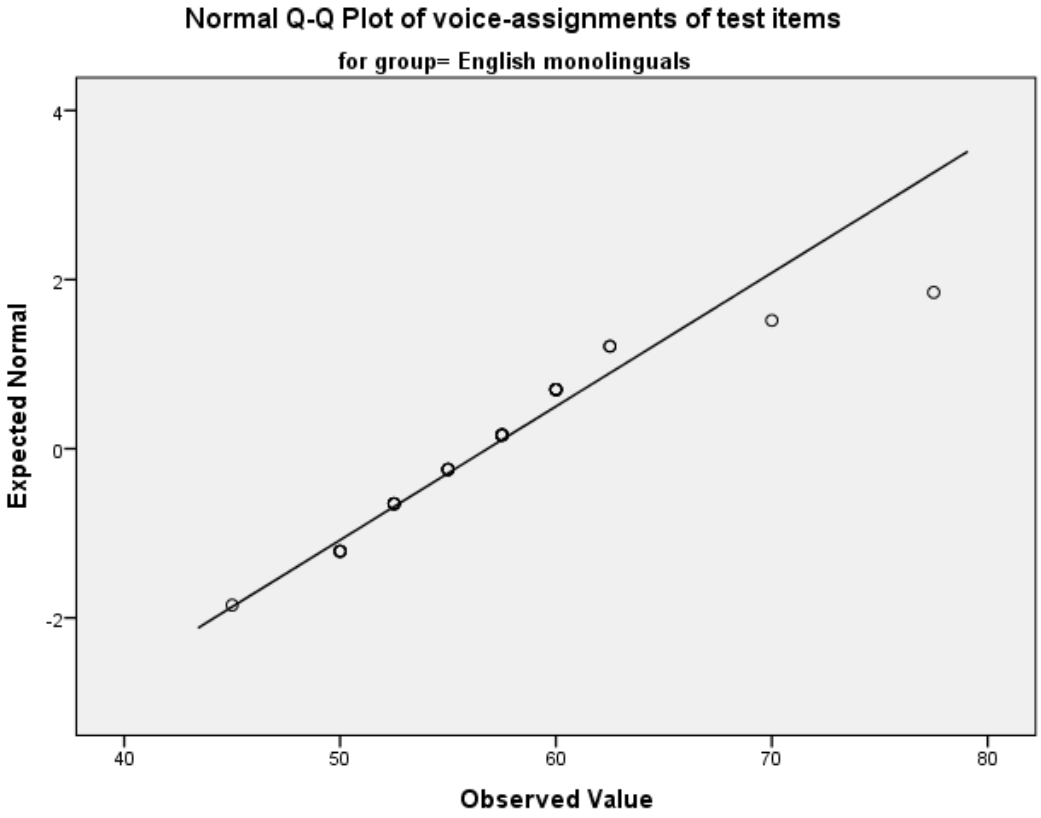
	Mean of all pairs ratings	Same Grammatical Gender Pairs	Different Grammatical Gender Pairs	Same Semantic Category Pairs	Different Semantic Category Pairs
Mann-Whitney U	365.000	414.500	343.000	370.500	439.500
Wilcoxon W	830.000	879.500	808.000	835.500	904.500
Z	-1.257	-.525	-1.583	-1.176	-.155
Asymp. Sig. (2-tailed)	.209	.599	.113	.240	.877

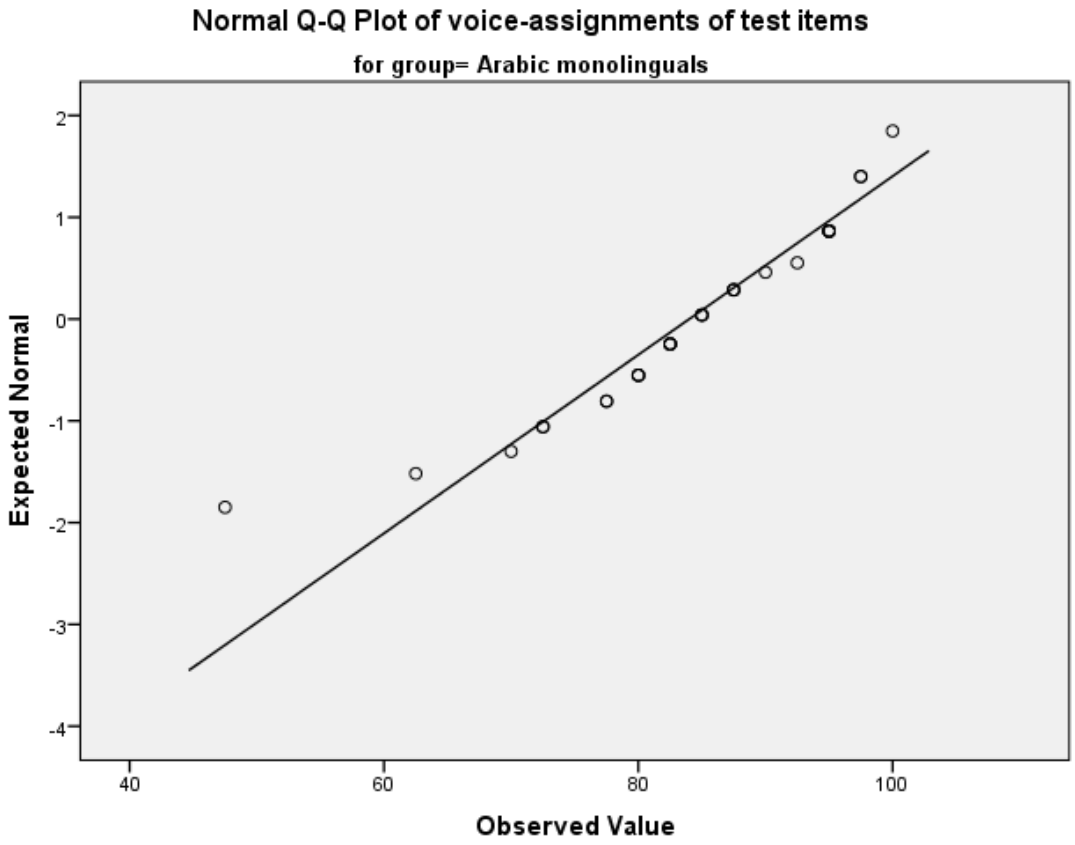
	Same Grammatical Gender Same Semantic Categories Pairs	Same Grammatical Gender Different Semantic Categories Pairs	Different Grammatical Gender Same Semantic Categories Pairs	Different Grammatical Gender Different Semantic Categories Pairs
Mann-Whitney U	407.500	429.000	339.500	430.500
Wilcoxon W	872.500	894.000	804.500	895.500
Z	-.629	-.311	-1.635	-.289
Asymp. Sig. (2-tailed)	.529	.756	.102	.773

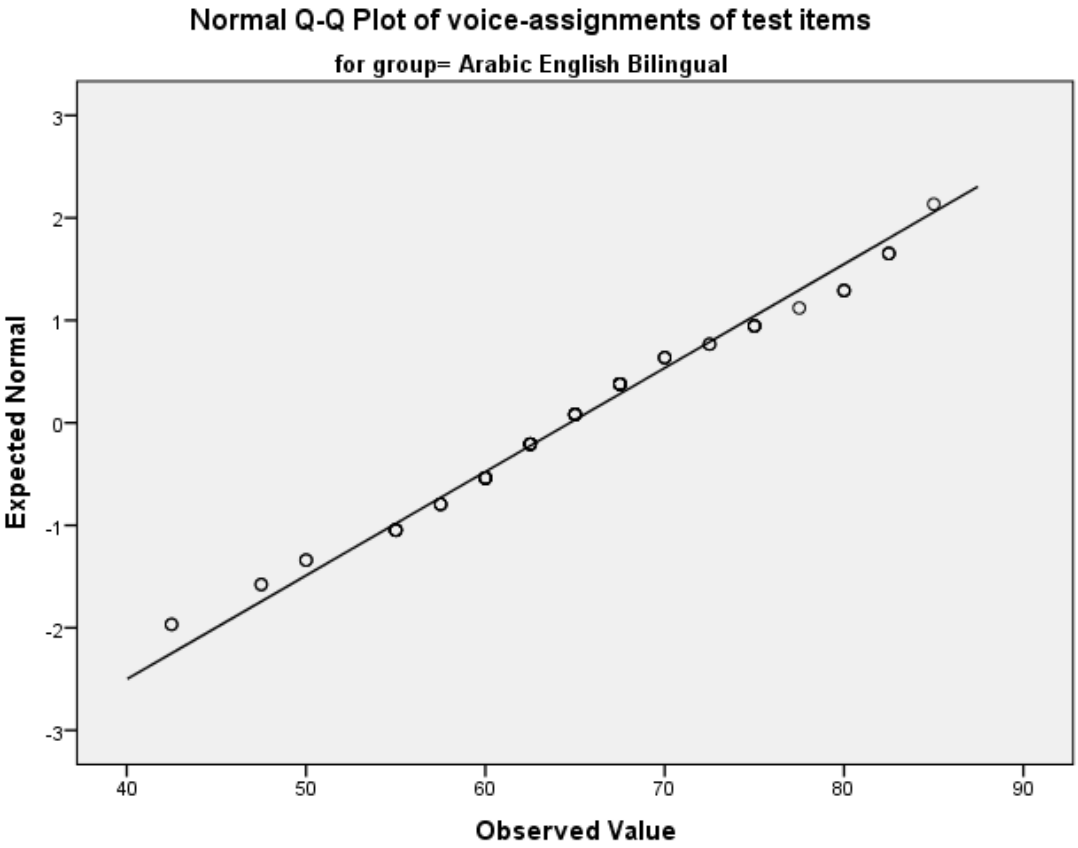
a. Grouping Variable: Group of participant

Q-Q plot of the voice attribution experiment

Normal Q-Q Plots







Appendix 12: Test of individual pairs

(A) Arabic monolinguals

Descriptive Statistics

	N	Mean	Std. Deviation
Table-Basket	30	3.0000	1.55364
Nose-Hand	30	4.5000	1.90734
Eye-Thumb	30	4.0667	2.18037
Hat-Pants	30	5.2667	1.91065
Train-Car	30	5.8667	1.38298
Mushroom-Grape	30	4.6333	1.56433
Hat-Jacket	30	4.8667	1.73669
Train-Bus	30	6.2667	.69149
Nose-Thumb	30	4.4667	1.67607
Hand-Thumb	30	5.8333	1.36668
Eye-Mushroom	30	1.9667	1.54213
Pants-Skirt	30	5.7333	1.28475
Eye-Hand	30	4.4000	1.65258
Jacket-Skirt	30	4.8333	1.14721
Train-Airplane	30	5.3667	1.24522
Table-Pants	30	1.5333	.57135
Car-Airplane	30	5.0000	1.68154
Basket-Grape	30	3.5667	2.06253
Apple-Nose	30	2.0000	.90972
Eye-Train	30	1.6333	.99943
Hat-Eye	30	1.8667	1.27937
Car-Bus	30	5.4333	1.43078
Cup-Table	30	2.9667	1.49674
Thumb-Strawberry	30	2.5667	1.69550
Mushroom-Table	30	2.1667	1.82101
Jacket-Hand	30	2.3667	1.77110
Train-Mushroom	30	1.7667	1.50134
Table-Skirt	30	1.7333	1.14269
Hand-Strawberry	30	2.3333	1.42232
Bus-Airplane	30	5.4333	1.73570
Grapes-Hand	30	2.9000	1.53914
Table-Train	30	2.1667	1.51050
Bus-Skirt	30	2.0667	1.48401
Strawberry-Jacket	30	1.9333	1.52978
Apple-Bus	30	1.7667	1.52414
Chair-Table	30	5.7000	1.82228
Bus-Basket	30	2.4333	1.61210
Hat-Skirt	30	4.3333	1.78757
Basket-Car	30	2.6000	1.73404
Bus-Pants	30	2.7000	2.16795
Apple-Mushroom	30	4.7000	1.96784
Chair-Hat	30	3.1000	1.93605

Grapes-Airplane	30	2.1667	1.70361
Nose-Chair	30	1.8667	1.61316
Nose-Eye	30	5.0667	1.89251
Chari-Grapes	30	1.9333	1.52978
Apple-Airplane	30	1.8333	1.53316
Car-Chair	30	2.3333	1.60459
Airplane-Cup	30	3.0000	1.68154
Chair-Basket	30	3.9333	1.63861
Cup-Thumb	30	3.8667	2.02967
Apple-strawberry	30	6.1333	1.10589
Pants-Basket	30	2.4333	1.54659
Strawberry-Grapes	30	6.2667	1.17248
Strawberry-Mushroom	30	5.1667	1.46413
Cup-Chair	30	3.1333	1.92503
Apple-Grapes	30	5.9333	1.41259
Jacket-Pants	30	5.8667	1.61316
Cup-Basket	30	3.8667	1.87052
Nose-Skirt	30	2.3000	1.78403
Valid N (listwise)	30		

(B) English monolinguals

Descriptive Statistics

	N	Mean	Std. Deviation
Table-Basket	30	3.6333	2.23581
Nose-Hand	30	4.1000	2.27959
Eye-Thumb	30	3.7667	2.11209
Hat-Pants	30	4.5667	1.97717
Train-Car	30	5.1667	1.93129
Mushroom-Grape	30	4.6000	1.67332
Hat-Jacket	30	4.6333	1.88430
Train-Bus	30	5.9667	1.35146
Nose-Thumb	30	4.8667	1.63440
Hand-Thumb	30	4.5667	2.07918
Eye-Mushroom	30	3.0333	1.99107
Pants-Skirt	30	5.2333	1.95965
Eye-Hand	30	4.0667	1.98152
Jacket-Skirt	30	4.9333	1.92861
Train-Airplane	30	4.8000	1.71001
Table-Pants	30	2.1333	1.79527
Car-Airplane	30	4.5667	1.95965
Basket-Grape	30	3.5000	2.02995
Apple-Nose	30	1.8000	1.12648
Eye-Train	30	1.9333	1.11211
Hat-Eye	30	3.0333	1.92055
Car-Bus	30	4.7000	2.19953
Cup-Table	30	2.3333	1.12444
Thumb-Strawberry	30	2.4000	1.61031
Mushroom-Table	30	1.6667	1.12444
Jacket-Hand	30	2.3333	1.34762
Train-Mushroom	30	1.6667	.84418
Table-Skirt	30	2.1333	1.73669
Hand-Strawberry	30	2.8667	1.50249
Bus-Airplane	30	4.9000	1.82606
Grapes-Hand	30	3.9667	1.88430
Table-Train	30	2.3333	2.15492
Bus-Skirt	30	2.1000	1.88186
Strawberry-Jacket	30	1.6667	.99424
Apple-Bus	30	1.8333	1.17688
Chair-Table	30	5.0000	1.94759
Bus-Basket	30	2.3000	1.70496
Hat-Skirt	30	3.8667	2.01260
Basket-Car	30	2.2667	1.74066
Bus-Pants	30	1.5333	.97320

Apple-Mushroom	30	4.5333	1.67607
Chair-Hat	30	2.0000	1.64002
Grapes-Airplane	30	2.0000	1.50860
Nose-Chair	30	1.9000	1.44676
Nose-Eye	30	4.7667	2.12835
Chari-Grapes	30	1.9667	1.54213
Apple-Airplane	30	1.7667	.93526
Car-Chair	30	2.8000	1.58441
Airplane-Cup	30	1.9333	1.31131
Chair-Basket	30	3.1000	1.80707
Cup-Thumb	30	3.4333	2.31462
Apple-strawberry	30	5.9000	1.49366
Pants-Basket	30	2.8000	2.13993
Strawberry-Grapes	30	5.9000	1.26899
Strawberry-Mushroom	30	4.8333	1.91335
Cup-Chair	30	2.5667	1.69550
Apple-Grapes	30	5.4000	1.83077
Jacket-Pants	30	5.4333	2.02882
Cup-Basket	30	3.3333	2.00574
Nose-Skirt	30	2.0667	1.38796
Valid N (listwise)	30		

(C) Both monolingual groups

Descriptive Statistics

	N	Mean	Std. Deviation
Table-Basket	60	3.3167	1.93532
Nose-Hand	60	4.3000	2.09357
Eye-Thumb	60	3.9167	2.13360
Hat-Pants	60	4.9167	1.95969
Train-Car	60	5.5167	1.70236
Mushroom-Grape	60	4.6167	1.60604
Hat-Jacket	60	4.7500	1.80042
Train-Bus	60	6.1167	1.07501
Nose-Thumb	60	4.6667	1.65362
Hand-Thumb	60	5.2000	1.85765
Eye-Mushroom	60	2.5000	1.84575
Pants-Skirt	60	5.4833	1.66206
Eye-Hand	60	4.2333	1.81675
Jacket-Skirt	60	4.8833	1.57407
Train-Airplane	60	5.0833	1.51032
Table-Pants	60	1.8333	1.35505
Car-Airplane	60	4.7833	1.82350
Basket-Grape	60	3.5333	2.02917
Apple-Nose	60	1.9000	1.02014
Eye-Train	60	1.7833	1.05913
Hat-Eye	60	2.4500	1.72150
Car-Bus	60	5.0667	1.87641
Cup-Table	60	2.6500	1.35077
Thumb-Strawberry	60	2.4833	1.64153
Mushroom-Table	60	1.9167	1.52150
Jacket-Hand	60	2.3500	1.56037
Train-Mushroom	60	1.7167	1.20861
Table-Skirt	60	1.9333	1.47138
Hand-Strawberry	60	2.6000	1.47522
Bus-Airplane	60	5.1667	1.78664
Grapes-Hand	60	3.4333	1.78854
Table-Train	60	2.2500	1.84689
Bus-Skirt	60	2.0833	1.68031
Strawberry-Jacket	60	1.8000	1.28617
Apple-Bus	60	1.8000	1.35046
Chair-Table	60	5.3500	1.90294
Bus-Basket	60	2.3667	1.64643
Hat-Skirt	60	4.1000	1.90183
Basket-Car	60	2.4333	1.73075
Bus-Pants	60	2.1167	1.76685

Apple-Mushroom	60	4.6167	1.81418
Chair-Hat	60	2.5500	1.86334
Grapes-Airplane	60	2.0833	1.59758
Nose-Chair	60	1.8833	1.51927
Nose-Eye	60	4.9167	2.00247
Chari-Grapes	60	1.9500	1.52299
Apple-Airplane	60	1.8000	1.25954
Car-Chair	60	2.5667	1.59837
Airplane-Cup	60	2.4667	1.58880
Chair-Basket	60	3.5167	1.76108
Cup-Thumb	60	3.6500	2.16932
Apple-strawberry	60	6.0167	1.30827
Pants-Basket	60	2.6167	1.86030
Strawberry-Grapes	60	6.0833	1.22532
Strawberry-Mushroom	60	5.0000	1.69746
Cup-Chair	60	2.8500	1.82102
Apple-Grapes	60	5.6667	1.64334
Jacket-Pants	60	5.6500	1.83030
Cup-Basket	60	3.6000	1.94152
Nose-Skirt	60	2.1833	1.58907
Valid N (listwise)	60		

(D) Arabic-English bilinguals

Descriptive Statistics

	N	Mean	Std. Deviation
Table-Basket	30	3.1667	1.83985
Nose-Hand	30	4.3000	2.35108
Eye-Thumb	30	3.7000	2.18380
Hat-Pants	30	4.3667	2.02541
Train-Car	30	5.6000	1.65258
Mushroom-Grape	30	4.2333	2.04574
Hat-Jacket	30	4.6000	2.14315
Train-Bus	30	5.9000	1.56139
Nose-Thumb	30	4.7333	1.87420
Hand-Thumb	30	5.5333	2.12916
Eye-Mushroom	30	2.2333	2.01175
Pants-Skirt	30	4.9667	1.69143
Eye-Hand	30	4.0667	2.09981
Jacket-Skirt	30	4.1667	1.62063
Train-Airplane	30	5.0667	1.43679
Table-Pants	30	1.8667	1.85199
Car-Airplane	30	4.8333	1.80198
Basket-Grape	30	3.4000	2.35767
Apple-Nose	30	2.0000	1.53128
Eye-Train	30	1.4667	1.07425
Hat-Eye	30	1.9333	1.72073
Car-Bus	30	5.4333	1.47819
Cup-Table	30	2.7667	1.59056
Thumb-Strawberry	30	2.1333	1.47936
Mushroom-Table	30	1.7333	1.57422
Jacket-Hand	30	2.0000	1.46217
Train-Mushroom	30	1.5333	1.25212
Table-Skirt	30	1.9333	1.41259
Hand-Strawberry	30	2.1667	1.62063
Bus-Airplane	30	5.2000	1.66919
Grapes-Hand	30	2.2333	1.61210
Table-Train	30	1.8333	1.39168
Bus-Skirt	30	1.8000	1.32353
Strawberry-Jacket	30	1.6000	1.22051
Apple-Bus	30	1.6667	1.37297
Chair-Table	30	5.3667	2.12511
Bus-Basket	30	2.2000	1.56249
Hat-Skirt	30	4.1333	1.77596
Basket-Car	30	2.2333	1.75545
Bus-Pants	30	1.7000	1.39333
Apple-Mushroom	30	4.7333	2.11617

Chair-Hat	30	1.9000	1.21343
Grapes-Airplane	30	1.6333	1.49674
Nose-Chair	30	1.6667	1.44636
Nose-Eye	30	5.1667	1.93129
Chari-Grapes	30	1.6000	1.19193
Apple-Airplane	30	1.4667	1.13664
Car-Chair	30	2.1000	1.64736
Airplane-Cup	30	2.0333	1.44993
Chair-Basket	30	2.9333	1.70057
Cup-Thumb	30	2.7667	2.02882
Apple-strawberry	30	5.7667	1.73570
Pants-Basket	30	2.1667	1.41624
Strawberry-Grapes	30	6.1333	1.07425
Strawberry-Mushroom	30	4.7333	2.03306
Cup-Chair	30	2.6000	1.81184
Apple-Grapes	30	6.0000	1.28654
Jacket-Pants	30	5.6667	1.37297
Cup-Basket	30	3.5000	2.16158
Nose-Skirt	30	1.9333	1.57422
Valid N (listwise)	30		

References

Archibald, J. (1998) 'Second Language Phonology, Phonetics, and Typology', *Studies in Second Language Acquisition*, 20, pp. 189-212.

Athanasopoulos, P. (2006) 'Effects of the Grammatical Representation of Number on Cognition in Bilinguals', *Bilingualism: Language and Cognition*, 9, pp. 89-96.

Athanasopoulos, P. (2007) 'Interaction between Grammatical Categories and Cognition in Bilinguals: The Role of Proficiency, Cultural Immersion, and Language of Instruction', *Language and Cognitive Processes*, 22, pp. 689-699.

Athanasopoulos, P. (2009) 'Cognitive Representation of Colour in Bilinguals: The Case of Greek Blues', *Bilingualism: Language and Cognition*, 12 pp. 83-95.

Athanasopoulos, P. (2011) 'Cognitive Restructuring in Bilingualism', in Pavlenko, A. (ed.) *Thinking and Speaking in Two Languages*, Bristol: Multilingual Matters, pp. 29-65.

Athanasopoulos, P. and Bylund, E. (2013) 'Does Grammatical Aspect Affect Motion Event Cognition? A Cross-Linguistic Comparison of English and Swedish Speakers', *Cognitive Science*, 37, pp. 286-309.

Athanasopoulos, P. and Kasai, C. (2008) 'Language and Thought in Bilinguals: The Case of Grammatical Number and Nonverbal Classification Preferences', *Applied Psycholinguistics*, 29, pp. 105-123.

Au, T. (1983) 'Chinese and English Counterfactuals: the Sapir-Whorf Hypothesis Revisited', *Cognition*, 15, pp. 155-187.

Bachman, L. and Palmer, A. (1989) 'The Construct Validation of Self-Ratings of Communicative Language Ability', *Language Testing*, 6(1), pp. 14-29.

Bassetti, B. (2007) 'Bilingualism and Thought: Grammatical Gender and Concepts of Objects in Italian-German Bilingual Children', *International Journal of Bilingualism*, 11, pp. 251-273.

Bassetti, B. (2011) 'The Grammatical and Conceptual Gender of Animals in Second Language Users', in Cook, V. and Bassetti, B. (eds.) *Language and Bilingual Cognition*. Oxford, UK: Psychology Press, pp. 357–384.

Berlin, B. and Kay, P. (1969) *Basic Color Terms: their Universality and Evolution*. Berkeley, CA: University of California Press.

Bloom, A. (1981) *Linguistic Shaping of Thought: a Study in the Impact of Language on Thinking in China and the West*. Hillsdale, NJ: Lawrence Erlbaum.

Bloom, A. (1984) 'Caution – the Words you Use may Affect What you Say: a Response to Au', *Cognition*, 17, pp. 275-287.

Boas, F. (1938) 'Language', in Boas, F. (ed.) *General Anthropology*. Boston: New York: D. C. Heath and Company, pp. 124-145.

Boroditsky, L. (2001) 'Does Language Shape Thought? Mandarin and English Speakers Conceptions of Time', *Cognitive Psychology*, 43, pp. 1-22.

Boroditsky, L., Schmidt, L. (2000) *Proceedings of the 22nd Annual Meeting of the Cognitive Science Society*. Mahwah, NJ. Erlbaum Associates.

Boroditsky, L., Schmidt, L. and Phillips, W. (2003) 'Sex, Syntax, and Semantics', in Gentner, D. and Goldin-Meadow, S. (eds.) *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press, pp. 61-79.

Boutonnet, B., Athanasopoulos, P. (2011) *Bilingualism and Thought: Uncovering the mechanisms of conceptual restructuring. ISB 8,* Oslo, Norway.

Boutonnet, B., Athanasopoulos, P. and Thierry, G. (2012) 'Unconscious Effects of Grammatical Gender during Object Categorisation', *Brain Res*, 1479, pp. 72–79.

Bowerman, M. (1985) 'What Shapes Children's Grammars?', in Slobin, D. (ed.) *The Crosslinguistic Study of Language Acquisition*. Hillsdale, N.J: Lawrence Erlbaum, pp. 1257-1319.

Bowers, J., Vigliocco, G., Stadthagen-Gonzalez, H. and Vinson, D. (1999) 'Distinguishing Language from Thought: Experimental Evidence that Syntax is Lexically Rather than Conceptually Represented', *Psychological Science*, 10, pp. 310-315.

Brown, H. (2000) *Principles of Language Learning & Teaching*. 4th ed. New York: Longman.

Brown, R. (1957) 'Linguistic Determinism and the Part of Speech', *Journal of Abnormal and Social Psychology*, 55, pp. 1-5.

Brown, R. (1976) 'In Memorial Tribute to Eric Lenneberg', *Cognition* 4(2), pp. 125-153.

Brown, R. and Lenneberg, E. (1954) 'A Study in Language and Cognition', *Journal of Abnormal and Social Psychology*, 49, pp. 454-462.

Carifio, L. and Perla, R. (2008) 'Resolving the 50 year Debate around Using and Misusing Likert Scales', *Medical Education*, 42, pp. 1150–1152.

Carroll, J. and Casagrande, J. (1958) 'The Function of Language Classifications in Behavior', in Maccoby, E., Newcomb, T. and Hartley, E. (eds.) *Readings in Social Psychology, 3rd Ed*. New York: Holt, Reinhart & Winston, pp. 18-31.

Chen, J. (2007) 'Do Chinese and English Speakers Think about Time Differently? Failure of Replicating Boroditsky (2001)', *Cognition*, 104(2), pp. 427-436.

Chini, M. (1995) *Genere Grammaticale e Acquisitione. Aspetti Della Morfologia Nominale in Italiano L2*. Milano, Italy: Franco Angeli.

Choi, S. and Bowerman, M. (1991) 'Learning to Express Motion Events in English and Korean: The Influence of Language-Specific Lexicalization Patterns', *Cognition*, 41, pp. 83-121.

Clarke, M., Losoff, A., Dickenson, M. and Mccracken, J. (1981) 'Gender Perception in Arabic and English', *Language Learning*, 31, pp. 159 – 169.

Cook, V. (1991) 'The Poverty-of-the-Stimulus Argument and Multi-Competence', *Second Language Research*, 7(2), pp. 103-117.

Cook, V. (1997) 'The Consequences of Bilingualism', in de Groot, A. and Kroll, J. (eds.) *Tutorials in Bilingualism: Psycholinguistic Perspectives*. Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 279-299.

Cook, V. (2002) 'Background to the L2 User', in Cook, V. (ed.) *Portraits of the L2 User*. Clevedon: Multilingual Matters.

Cook, V. (2007) 'The Nature of the L2 User', *EUROSLA Yearbook 7* Amsterdam. John Benjamins.

Cook, V., Bassetti, B., Kasai, C., Sasaki, M. and Takahashi, J. (2006) 'Do Bilinguals have Different Concepts? The Case of Shape and Material in Japanese L2 Users of English', *International Journal of Bilingualism*, 10, pp. 137-152.

Cook, V., Iarossi, E., Stellakis, N. and Tokumaru, Y. (2003) 'Effects of the L2 on the Syntactic Processing of the L1', in Cook, V. (ed.) *Effects of the Second Language on the First*. Clevedon: Multilingual Matters, pp. 193-213.

Corbett, G. (1991) *Gender*. Cambridge: UK Cambridge University Press.

Costa, A., Alario, F. and Sebastian-Galle´s, N. (2007) 'Cross-Linguistic Research on Language Production', in Gaskell, M. (ed.) *The Oxford Handbook of Psycholinguistics*. Oxford, UK: Oxford University Press, pp. 531-546.

Cubelli, R., Paolieri, D., Lotto, L. and Job, R. (2011) 'The Effect of Grammatical Gender on Object Categorisation', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), pp. 449-460.

Czechowska, N. and Ewert, A. (2011) 'Perception of Motion by Polish-English Bilinguals', in Cook, V. and Bassetti, B. (eds.) *Language and Bilingual Cognition*. Hove, UK: Psychology Press, pp. 287 – 314.

Davidoff, J., Davies, I. and Roberson, D. (1999) 'Colour Categories of a Stone-Age Tribe', *Nature* 398, pp. 203-204 & 402, 604.

de Groot, A. (2011) *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York-Hove: Psychology Press.

Degani, T. (2007) *The Semantic Role of Gender: Grammatical and Biological Gender Match Effects in English and Spanish*. Master's Thesis. University of Pittsburgh.

Dewaele, J. and Verronique, D. (2000) 'Relating Gender Errors to Morphosyntactic and Lexical Systems in Advanced French Interlanguage', *Studia Linguistica*, 54(2), pp. 212-224.

Dixon, R. (1972) *The Dyirbal Language of North Queensland*. Cambridge: Cambridge University Press.

Ervin, S. (1961) 'Learning and Recall in Bilinguals', *American Journal of Psychology*, 74, pp. 446-451.

Ervin, S. (1961) 'Semantic Shift in Bilingualism', *American Journal of Psychology*, 74, pp. 233-241.

Ervin, S. (1962) 'The Connotations of Gender', *Word*, 18, pp. 249-261.

Ervin, S. (1964) 'Language and Thematic Apperception Test Content in Bilinguals', *Journal of Abnormal and Social Psychology* 68, pp. 500-507.

Ewert, A. and Bromberek-Dyzman, K. (2008) 'L2 Users' Sociopragmatic and Pragmalinguistic Choices in L1 Acts of Refusal', *EUROSLA*. Amsterdam. John Benjamins.

Finstad, K. (2010) 'Response Interpolation and Scale Sensitivity: Evidence Against 5-Point Scales', *Journal of Usability Studies*, 5(3), pp. 104-110.

Finch, S (1998-2007). Reading the bones. Reprinted in Finch, S (2007). The guild of xenolinguists. Urbana, IL: Golden Gryphon Press.

Fisher, C. and Gleitman, L. (2002) 'Breaking the Linguistic Code: Current Issues in Early Language Learning', in Gallistel, C. (ed.) *Stevens' Handbook of Experimental Psychology, (Vol. 3)*. New York: John Wiley & Sons.

Flaherty, M. (1999) 'The Influence of a Language Gender System on Perception', *Tohoku Psychologica Folia*, 58, pp. 1-10.

Flaherty, M. (2001) 'How a Language Gender System Creeps into Perception', *Journal of Cross Cultural Psychology*, 32, pp. 18-31.

Fodor, J., Bever, T. and Garrett, M. (1974) *The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar*. New York: McGraw-HilI.

Forbes, J., Poulin-Dubois, D., Rivero, M. and Sera, M. (2008) 'Grammatical Gender Affects Bilinguals' Conceptual Gender: Implications for Linguistic Relativity and Decision Making', *The Open Applied Linguistics Journal*, 1, pp. 68–76.

Foundalis, H. (2002) 'Evolution of Gender in Indo-European Languages', in Gray, W. and Schunn, C. (eds.) *Proceedings of the 24th Annual Meeting of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 304 – 309.

Gass, S. and Selinker, L. (2001) *Second Language Acquisition: An Introductory Course*. 2nd edn. Mahwah, NJ: Lawrence Erlbaum.

Gelman, S. (2004) 'Learning Words for Kinds: Generic Noun Phrases in Acquisition', in Hall, D. and Waxman, S. (eds.) *Weaving a Lexicon*. Cambridge, Mass.: MIT Press, pp. 445-484.

Gennari, S., Sloman, S., Malt, B. and Fitch, T. (2002) 'Motion Events in Language and Cognition', *Cognition*, 83, pp. 49-79.

Gentner, D. and Goldin-Meadow, S. (eds.) (2003) *Language in Mind: Advances in the Study of Language and Thought.* Cambridge, MA: The MIT Press.

Gentner, D. and Markman, A. (1994) 'Structural Alignment in Comparison: No Difference Without Similarity', *Psychological Science*, 5(3), pp. 152-158.

Gilbert, A., Regier, T., Kay, P. and R, I. (2006) 'Whorf Hypothesis is Supported in the Right Visual Field but not the Left', *Proceedings of the National Academy of Sciences*, 103, pp. 489–494.

Gleitman, L. and Newport, E. (1995) 'The Invention of Language by Children: Environmental and Biological Influences on the Acquisition of Language', in Gleitman, L. and Liberman, M. (eds.) *Invitation to Cognitive Science (Vol. 1).* Cambridge, MA: MIT Press.

Gopnik, A. (2001) 'Theories, Language, and Culture: Whorf without Wincing', in Bowerman, M. and Levinson, S. (eds.) *Language Acquisition and Conceptual Development*. Cambridge University Press, pp. 45-69.

Grosjean, F. (2001) 'The Bilingual's Language Modes', in Nicol, J. (ed.) *One Mind, Two Languages: Bilingual Language Processing*. Oxford: Blackwell, pp. 1-22.

Grosjean, F. (2010) *Bilingual: Life and Reality*. Cambridge, Mass: Harvard University Press.

Guiora, A., & Acton, W. R (1979) 'Personality and language: A restatement', *Language Learning*, 29, pp. 193-204.

Guiora, A. and Sagi, A. (1978) 'A Cross-Cultural Study of Symbolic Meaning: Developmental Aspects', *Language Learning*, 28(2), pp. 381 – 386.

Hakuta, K. (1986) *Mirror of language: The Debate on Bilingualism*. New York: Basic Books.

Hampton, J. and Dubois, D. (1993) 'Psychological Models of Concepts', in Van Mechelen, I., Hampton, J., Michalski, R. and Theuns, P. (eds.) *Categories and Concepts: Theoretical Views and Inductive Data Analysis* London: Academic Press, pp. 11-34.

Harris, R. and McGhee-Nelson, E. (1992) 'Bilingualism: Not the Exception Any More', in Harris , R. (ed.) *Cognitive Processing in Bilinguals*. Amsterdam: Elsevier, pp. 3–14.

Haviland, J. (1993) 'Anchoring, Iconicity and Orientation in Guugu Yimithirr Pointing Gestures', *Journal of Linguistic Anthropology*, 3(1), pp. 3-45.

Heider, E. (1972) 'Universals in Color Naming and Memory', *Journal of Experimental Psychology.*, 93, pp. 10-20.

Hill, J. and Mannheim, B. (1992) 'Language and World View', *Annual Review of Anthropology* 21, pp. 381-406.

Hockett, C. (1958) *A Course in Modern Linguistics*. New York: Macmillan.

Hofstätter, P. (1963) 'Uber sprachjliche Bestimmungsleistungen: Das Problem des grammatikalischen Geschlechts von Sonne und Mond. Zeitschrift für experimentelle und angewandte', *Psychologie*, 10, pp. 91-108.

Hunt, E. and Agnoli, F. (1991) 'The Whorfian Hypothesis: A Cognitive Psychology Perspective', *Psychological Review*, 98, pp. 377-389.

Imai, M. (2000) 'Sapia-Wafu-kasetsu saikou: Shikou-keisei ni okeru gengo-no yakuwari, sono soutaisei to fuhensei [Rethinking the Sapir-Whorf hypothesis: Role of Language in Shaping Thought]', *The Japanese Journal of Psychology*, 71(5), pp. 415-433.

Imai, M. and Gentner, D. (1997) 'A Crosslinguistic Study of Early Word Meaning: Universal Ontology and Linguistic Influence', *Cognition*, 62, pp. 169-200.

Jakobson, R. (1966) 'On Linguistic Aspects of Translation', in Brower, R. (ed.) *On translation*. New York: Oxford University Press, pp. 232-239.

January, D. and Kako, E. (2007) 'Re-Evaluating Evidence for the Linguistic Relativity Hypothesis: Response to Boroditsky (2001). , 104 (2), 417–426.', *Cognition*, 104(2), pp. 417-426.

Jenkins, J. (2009) *World Englishes: a resource book for students, 2nd edition*, London, UK, Routledge.

Jiyad, M. (2006) *A Hundred and One Rules! A Short Reference for Arabic Syntactic, Morphological & Phonological Rules for Novice & Intermediate Levels of Proficiency*. Available at: http://www.scribd.com/doc/4931924/Arabic-Grammar-in-English (Accessed: 2/06/2012).

Johnson, C., Paivio, A. and Clark, J. (1996) 'Cognitive Components of Picture Naming', *Psychological Bulletin*, 120(1), pp. 113-139.

Johnson, J. and Newport, E. (1989) 'Critical Period Effects in Second Language Learning: The Influence of Maturational State on the Acquisition of English as a Second Language', *Cognitive Psychology*, 21, pp. 60-99.

Kay, P. and Kempton, W. (1984) 'What is the Sapir-Whorf Hypothesis? ', *American Anthropologist*, 86, pp. 65-79.

Konishi, T. (1993) 'The Semantics of Grammatical Gender: A Cross-Cultural Study', *Journal of Psycholinguistic Research*, 22, pp. 519-534.

Kousta, S., Vinson, D. and Vigliocco, G. (2008) 'Investigating Linguistic Relativity Through Bilingualism: The Case of Grammatical Gender', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, pp. 843-858.

Kroll, J., Bobb, S. and Hoshino, N. (2014) ‘Two languages in mind: bilingualism as a tool to investigate language, cognition, and the brain’ *Current Directions in Psychological Science*, 23(3), pp. 159–163.

Kurinski, E. and Sera, M. (2010) 'Does Learning Spanish Grammatical Gender Change English-Speaking Adults' Categorization of Inanimate Objects?', *Bilingualism: Language and Cognition*, 14(2), pp. 203–220.

Lakoff, G. (1987) *Women, Fire, and Dangerous Things*. Chicago: The University of Chicago Press.

Lenneberg, E. (1953) 'Cognition in Ethnolinguistics', *Language* 29, pp. 463-471.

Lenneberg, E. (1961) 'Color Naming, Color Recognition, Color Discrimination: A Re-Appraisal', *Perceptual and Motor Skills*, 12, pp. 375-382.

Lenneberg, E. (1967) *Biological Foundations of Language*. New York: Wiley.

Lenneberg, E. and Roberts, J. (1953) 'The Denotata of Color Terms', *Paper Presented at the Linguistic Society of America*. Bloomington, Indiana.

Lenneberg, E. and Roberts, J. (1956) *The Language of Experience: A Study in Methodology*. No. Memoir 13 in Indiana University Publications in Anthropology and Linguistics: Waverly Press.

Levinson, S. (1996) 'Frames of Reference and Molyneuxs Question: Crosslinguistic Evidence', in Bloom, P., Oederson, M., Nadel, L. and Garrett, M. (eds.) *Language and Space*. Cambridge, MA: MIT Press, pp. 109-169.

Levinson, S. (2003) 'Language and Mind: Lets Get the Issues Straight', in Gentner, D. and Goldin-Meadow, S. (eds.) *Language in Mind: Advances in the Study of Language and Cognition*. Cambridge, MA: MIT Press, pp. 25-46.

Levinson, S., Kita, S., Haun, D. and Rasch, B. (2002) 'Returning the Tables: Language Affects Spatial Cognition', *Cognition*, 84, pp. 155-188.

Lewis, M.e. (2013) *Ethnologue: Languages of the World, Seventeenth Edition*. Available at: http://www.ethnologue.com (Accessed: March/2014).

Li, P. and Gleitman, L. (2002) 'Turning the Tables: Language and Spatial Reasoning', *Cognition*, 83, pp. 265-294.

Liu, L. (1985) 'Reasoning Counterfactually in Chinese: Are there any Obstacles? ', *Cognition*, 21(3), pp. 239-270.

Lucus, M. (2000) 'Semantic Priming without Association: A Meta-Analytic Review', *Psychonomic Bulletin & Review*, 7(4), pp. 618–630.

Lucy, J. (1992) *Grammatical Categories and Cognition: A Case Study of the Linguistic Relativity Hypothesis*. Cambridge: Cambridge University Press.

Lucy, J. (1997) 'Linguistic Relativity', *Annual Review of Anthropology*, 26(291-312).

Lucy, J. (2011) 'Language and Cognition: The View from Anthropology', in Cook, V. and Bassetti, B. (eds.) *Language and Bilingual Cognition.* New York and Hove, UK: Psychology Press, pp. 43-68.

Lupyan, G. (2012) 'Linguistically modulated perception and cognition: the label feedback hypothesis' *Frontiers in Cognition*, 3(54), pp. 1-13.

Majid, A., Bowerman, M., Kita, S., Haun, D. and Levinson, S. (2004) 'Can Language Restructure Cognition? The Case for Space', *Trends in Cognitive Sciences*, 8(3), pp. 108-114.

Malt, B. and Sloman, S. (2003) 'Linguistic Diversity and Object Naming by Non-Native Speakers of English', *Bilingualism: Language and Cognition*, 6, pp. 47-67.

Malt, B., Sloman, S. and Gennari, S. (2003) 'Speaking vs. Thinking about Objects and Actions', in Gentner, D. and Goldin-Meadow, S. (eds.) *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press, pp. 81-111.

Martinez, I. and Shatz, M. (1996) 'Linguistic Influences on Categorization in Preschool Children: A Cross-Linguistic Study', *Journal of Child Language*, 23, pp. 529-545.

Mazuka, R. and Friedman, R. (2000) 'Linguistic Relativity in Japanese and English: Is Language the Primary Determinant in Object Classification?', *Journal of East Asian Linguistics*, 9, pp. 353-377.

Medin, D., Goldstone, R. and Gentner, D. (1993) 'Respects for Similarity', *Psychological Review*, 100, pp. 254-278.

Mills, A. (1986) *The Acquisition of Gender: A study of English and German*. Berlin: Springer-Verlag.

Mullen, M. (1990) 'Children's Classification of Nature and Artifact Pictures into Female and Male Categories', *Sex Roles*, 23, pp. 577 – 587.

Munnich, E. and Landau, B. (2003) 'The Effect of Spatial Language on Spatial Representations: Setting Some Boundaries', in Gentner, D. and Goldin-Meadow, S. (eds.) *Language in Mind: Advances in the Study of Language and Thought*. Cambridge, MA: MIT Press.

Murphy, E. (2004) ' Identifying and Measuring Problem Formulation and Resolution in Online Asynchronous Discussions', *Canadian Journal of Learning and Technology*, 30(1), pp. 5-20.

Murray, J. (2013) 'Likert Data: What to Use, Parametric or Non-Parametric?', *International Journal of Business and Social Science*, 4(11), pp. 258-264.

Nicoladis, E. and Foursha-Stevenson, C. (2012) 'Language and Culture Effects on Gender Classification of Objects', *Journal of Cross-Cultural Psychology*, 43, pp. 1095–1109.

Nunnally, J. (1978) *Psychometric Theory*. New York: McGraw-Hill. Rayner.

Ortner, S. (1974) 'Is Female to Male as Nature is to Culture?', in Rosaldo, M. and Lamphere, L. (eds.) *Woman, Culture and Society* Stanford, CA: Stanford Press, pp. 67-87.

Osgood, C., Suci, G. and Tannenbaum, P. (1957) *The Measurement of Meaning*. Urbana, IL: University of Illinois.

Oxford University Press. U.o.C., & Association of Language Testers in Europe (2001) *Quick Placement Test Paper and Pen Test*. Oxford: Oxford University Press.

Pavlenko, A. (2005) 'Bilingualism and Thought', in de Groot, A. and Kroll, J. (eds.) *Handbook of Bilingualism: Psycholinguistic Approaches*. Oxford: Oxford University Press, pp. 433-453.

Pavlenko, A. (2011) *Thinking and Speaking in two Languages*. Clevedon: Multilingual Matters.

Pavlenko, A. and Driagina, V. (2007) 'Russian Emotion Vocabulary in American Learners' Narratives', *Modern Language Journal*, 91(2), pp. 213-234.

Phillips, W. and Boroditsky, L. (2003) *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

Pilling, M., Wiggestt, A., Özgen, E. and Davies, I. (2003) 'Is Color "Categorical Perception" Really Perceptual?', *Memory and Cognition*, 31, pp. 538-551.

Pinker, S. (1994) *The Language Instinct: How the Mind Creates Language*. New York: Morrow.

Ramos, S. and Roberson, D. (2010) 'What Constrains Grammatical Gender Effects on Semantic Judgements? Evidence from Portuguese', *European Journal of Cognitive Psychology*, pp. 1-16.

Roberson, D., Davidoff, J., Davies, I. and Shapiro, L. (2004) 'The Development of Color Categories in Two languages: a Longitudinal Study', *Journal of Experimental Psychology: General*, 133, pp. 554-571.

Roberson, D., Davidoff, J., Davies, I. and Shapiro, L. (2005) 'Colour Categories: Evidence for the Cultural Relativity Hypothesis', *Cognitive Psychology*, 50, pp. 378-411.

Roberson, D., Davies, I. and Davidoff, J. (2000) 'Colour Categories are not Universal: Replications and New Evidence from a Stone-Age Culture', *Journal of Experimental Psychology: General*, 129, pp. 369-398.

Rogers, M. (1987) 'Learners' Difficulties with Grammatical Gender in German as a Foreign Language', *Applied Linguistics*, 8, pp. 48-74.

Rosch, E. (1973) 'Natural Categories ', *Cognitive Psychology*, 4, pp. 328-350.

Rosch, E. (1975) 'Cognitive Representations of Semantic Categories', *Journal of Experimental Psychology: General*, 104, pp. 192-233.

Sapir, E. (1921) *Language: An Introduction to the Study of Speech*. New York: Harcourt, Brace and company.

Sapir, E. (1929) 'The Status of Linguistics as a Science', *Language*, 5, pp. 207-214.

Seigneuric, A., Zagar, D., Meunier, R. and Spinelli, E. (2007) 'The Relation between Language and Cognition in 3- to 9-year-olds: The Acquisition of Grammatical Gender in French', *Journal of Experimental Child Psychology*, 96, pp. 229–246.

Sera, M., Berge, C. and del Castillo-Pintado, J. (1994) 'Grammatical and Conceptual Forces in the Attribution of Gender by English and Spanish Speakers', *Cognitive Development*, 9, pp. 261-292.

Sera, M., Elieff, C., Forbes, J., Burch, M., Rodriguez, W. and Dubois, D. (2002) 'When Language Affects Cognition and when it does not: An Analysis of Grammatical Gender and Classification', *Journal of Experimental Psychology: General*, 131, pp. 377-397.

Slobin, D. (1979) *Psycholinguistics*. 2nd edn. Glenview: IL: Scott, Foresman, & Co.

Slobin, D. (1985) 'Crosslinguistic Evidence for the Language- Making Capacity', in Slobin, D. (ed.) *The Crosslinguistic Study of Language Acquisition: Vol. 2. Theoretical Issues*. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 1157-1256.

Slobin, D. (1987) 'Thinking for Speaking', in Aske, J., Beery, N., Michaelis, L. and Filip, H. (eds.) *Proceedings of the Thirteenth Annual Meeting of the Berkeley Linguistics Society*. Berkeley, CA: Berkeley Linguistics Society, pp. 435-444.

Slobin, D. (1996) 'From Thought and Language to Thinking for Speaking', in Gumperz, J. and Levinson, S. (eds.) *Rethinking Linguistic Relativity*. Cambridge, England: Cambridge University Press, pp. 70-96.

Slobin, D. (2000) 'Verbalized Events: A Dynamic Approach to Linguistic Relativity and Determinism', in Niemeier, S. and Dirven, R. (eds.) *Evidence for Linguistic Relativity*. pp. 107 - 138.

Slobin, D. (2003) 'Language and Thought Online: Cognitive Consequences of Linguistic Relativity', in Gentner, D. and Goldin-Meadow, S. (eds.) *Language in Mind: Advances in the Study of Language and Thought* Cambridge, MA: MIT Press, pp. 157-191.

Sloman, S. and Malt, B. (2003) 'Artifacts are not Ascribed Essences, nor are they Treated as Belonging to Kinds', *Language and Cognitive Processes*, 18, pp. 563-582.

Snodgrass, J. and Vanderwart, M. (1980) 'A Standardized Set of 260 Pictures: Norms for Name Agreement, Image Agreement, Familiarity, and Visual Complexity', *Journal of Experimental Psychology: Human Learning and Memory*, 6, pp. 174-215.

Strømnes, F. (1974) 'No Universality of Cognitive Structures? Two Experiments with Almost-Perfect One-Trial Learning of Translatable Operators in a Ural-Altaic and Indo-European Language', *Scandinavian Journal of Psychology*, 15, pp. 300-309.

Szekely, A., Jacobsen, T., D'Amico, S., Devescovi, A., Andonova, E., Herron, D., Lu, C.-C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, G., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., Tzeng, O., Arévalo, A., Vargha, A., Butler, A., Buffington, R. and Bates, E. (2004) 'A New On-Line Resource for Psycholinguistic Studies', *Journal of Memory and Language*, 51, pp. 274-250.

Takano, Y. (1989) 'Methodological Problems in Cross-Cultural Studies of Linguistic Relativity', *Cognition*, 31, pp. 141-162.

van Berkum, J. (1996) *The Psycholinguistics of Grammatical Gender: Studies in Language Comprehension and Production*. Nijmegen: Nijmegen: University Press.

van Teijlingen, E. and Hundley, V. (2001) *The Importance of Pilot Studies*. Available at: http://sru.soc.surrey.ac.uk/SRU35.html (Accessed: 15 January).

Vigliocco, G., Vinson, D., Damian, M. and Levelt, W. (2002) 'Semantic Distance Effects on Object and Action Naming', *Cognition*, 85(3), pp. B61-B69.

Vigliocco, G., Vinson, D., Indefrey, P., Levelt, W. and Hellwig, F. (2004) 'Role of Grammatical Gender and Semantics in German Word Production', *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, pp. 483-497.

Vigliocco, G., Vinson, D., Paganelli, F. and Dworzynski, K. (2005) 'Grammatical Gender Effects on Cognition: Implications for Language Learning and Language Use', *Journal of Experimental Psychology: General*, 134, pp. 501-520.

Vygotsky, L. (1962) *Thought and Language*. Cambridge MA: MIT Press.

Vuksanovic, J. Bjekic, J. and Radivojevic, N. (2014) 'Grammatical Gender and Mental Representation of Objects: The case of Musical Instruments', *Journal of Psycholinguistics Research,* pp. 1-15.

Waite, M. (2002) *Concise Oxford Thesaurus*. 2nd ed. Oxford, England: Oxford University Press.

Whorf, B. (1941/1956) 'The Relation of Habitual Thought and Behavior to Language', in Spier, L. (ed.) *Language, Culture and Personality. Reprinted in J.B. Carroll (Ed.), Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. Cambridge,MA: MIT Press, pp. 134-159.

Whorf, B. (1956) *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf [J. Carroll, ed.]*. Cambridge, MA: MIT Press.

Winawer, J., Witthoft, N., Frank, M., Wu, L., Wade, A. and Boroditsky, L. (2007) 'Russian Blues Reveal Effects of Language on Color Discrimination', *Proceedings of the National Academy of Sciences*, 104, pp. 7780 –7785.

Zubin, D. and Klaus, M. (1984) 'Affect Classification in the German Gender System', *Lingua*, 63(41-96).