Online Sarcasm and its Perception by Second Language Learners: The Case of Iraqi EFL Learners in Iraq and the UK

Dheyaa AL-Fatlawi

Declaration

I hereby declare that this thesis is the product of my own work and has not been submitted in any form for the award of a higher degree elsewhere.

 D.H. A
Dheyaa Al-Fatlawi

Abstract

Although many studies have been written within L2 pragmatics, very few have dealt with L2 irony and sarcasm. The main purpose of this study is to investigate how EFL learners recognize written British English sarcasm. For this purpose, an L2 pragmatics study was designed and applied to two groups of L2 learners of English. Another purpose of the study is to pragmatically analyse online sarcasm, and see how it is used and by what features it is characterized. A corpus study was conducted for the latter purpose.

Regarding data, this study used naturally-occurring sarcasm from real-life situations. The data was collected from a football forum (Manchester United forum; likely used by males) and two parenting forums (Mumsnet and Netmums; likely used by females). These different forums were targeted to ensure a rough gender balance. Sarcasm was identified within these forums by means of a metalanguage strategy. This strategy involved searching for the metalinguistic labels *sarcasm* and *sarcastic*, and then extracting and analyzing the antecedent discourses these labels are referring to. Those discourses were considered a potential environment for sarcasm.

One hundred and forty two sarcasm-containing threads were collected via the metalanguage strategy. First, the data was pragmatically investigated to reveal the general pragmatic characteristics of sarcasm (e.g. Contradiction: *saying something and meaning the opposite* or Insincerity: *flouting the Gricean quality maxim*), as well as its pragmalinguistic characteristics (e.g. hyperbolic expressions, capitalization, exclamation marks) that are used in the data. This is the 'Corpus Study'. Second, the analysed data served as an item pool for the judgment task of the L2 pragmatics study. From that pool, 30 items were ultimately selected as stimuli for that L2 study.

Two groups of Iraqi EFL learners participated in the L2 pragmatics study. Each group contained 30 participants. The members of the first group were studying L2 English at home (Iraq) and had never been to any English-speaking country. The second group involved learners who received their BA and/or MA degree(s) in English from Iraq and were pursuing

MA or PhD degrees at different UK universities. Members of the latter group had 1-4 years sojourn in the UK. A Control group was also provided by 30 British-English native speakers. The 30 stimuli, derived from the online data of the Corpus study, were placed in a two-fold judgement task. The task was designed to: (1) test the participants' recognition of sarcasm within the given texts (threads) on a 7-point Likert scale, and (2) reveal what they consider as 'sarcastic' within those texts by highlighting the potentially sarcastic part(s) in them.

Results of the corpus study revealed that general pragmatic characteristics bear the greatest load in creating/indicating online sarcasm. Among these characteristics 'Insincerity' seems to be the most fundamental or prototypical one. As for pragmalinguistic characteristics, they appear to play only a minor role in triggering and comprehending online sarcasm. 'Hyperbole' seems to be the most prototypical one among pragmalinguistic characteristics. Regarding the L2 pragmatics study, ANOVA results reveal that both learners' groups are significantly different from English Native Speakers. Thus, Iraqi EFL learners appear not to have reached the native level of sarcasm perception. Results also indicate no effect of studying abroad or L2 proficiency upon the sarcasm recognition of those learners. Another finding of the L2 study is that the more characteristics (general pragmatic or pragmalinguistic) available the easier the comprehension of sarcasm turns out to be for both native speakers and learners. However, learners seem to be more sensitive to pragmalinguistic characteristics than English native speakers. They are found to identify sarcasm at the sight of these features more than the native speakers do. More interestingly, the current study has also found out that sarcasm does not always express a negative attitude. Sometimes, it can be used to express a positive emotion in a friendly way.

This study encourages further research on L2 sarcasm, particularly with regard to the kind and amount of L2 input the learners are exposed to. It also focuses attention on the necessity of developing the learners' L2 pragmatic competence in general and their competence of L2 sarcasm in particular in order to bridge the gap between their performance and that of the native speakers.

بسم الله الرحمن الرحيم

يَرْفَعِ اللَّهُ الَّذِينَ آمَنُوا مِنكُمْ وَالَّذِينَ أُوتُوا الْعِلْمَ دَرَجَاتٍ

(المجادلة: ١١)

Allah will rise up to ranks those of you who believe and who have been granted Knowledge

(Sura: Al-Mujadalah, verse 11)

Acknowledgements

In the name of Allah, the Most Gracious and the Most Merciful

Alhamdulillah, all praises to Almighty Allah for giving me the strength, courage, and patience to undertake this thesis and to persevere and complete it satisfactorily. Without His blessings, this achievement would not have been materialized.

First and foremost, I would like to express my special appreciation and thanks to my supervisors Professor Jonathan Culpeper and Dr. Patrick Rebuschat for being tremendous mentors for me. Thank you for encouraging me and providing me with valuable guidance and advice throughout the whole period of my study.

I am also grateful to the government of my country represented by the Iraqi Ministry of Higher Education and Scientific Research for funding this project.

Special thanks go to my mother. I cannot express in words how grateful I am to your toil in life which has made me what I am today and to your prayer which has sustained me thus far. I would like to express appreciation to my beloved wife for the undivided care and attention she provided to me. Thank you for being always ready to take my hand when I stumble and push me forward in the moments when I was about to give up. I would like also to thank all of my friends and colleagues who supported and encouraged me to achieve my goal.

And finally, I owe debt of gratitude to everyone who has helped me, in a way or another, during the last four years to reach this point of thesis completion.

Contents

List of tables

List of figures

List of abbreviations

ANOVA   Analysis of Variance

BrE   British English

CCSARP   Cross-Cultural Speech Act Realisation Project

CEFR   Common European Framework of Reference for Languages

DCT   Discourse Completion Task

EFL   English as a Foreign Language

ESL   English as a Second Language

(E)NS   (English) Native Speaker

H_0   Null Hypothesis

H_1   Alternative Hypothesis

ILP   Interlanguage Pragmatics

FTA   Face-Threatening Act

IELTS   International English Language Testing System

L1   First Language

L2   Second Language

MET   Multimedia Elicitation Task

NNS   Non-native speaker

P   Participant

SLA   Second Language Acquisition

SPSS   Statistical Package for the Social Sciences

SST   Stimuli Selection Task

Chapter One

Introduction

1.1 The Present Study: Statement of the Problem and Rationale for the Study

This study is situated within the domain of second language (L2) pragmatics, although parts of it also contribute to first language (L1) pragmatics. More specifically, its main aim is to investigate the recognition of English sarcasm by L2 learners of English. Since its introduction in late 1970s, L2 pragmatics was an interface wherein pragmatic issues were investigated in SLA (Second Language Acquisition) contexts. Influenced by the mainstream of pragmatics in 1980s and 1990s, L2 pragmatics has been mainly concerned with investigating speech acts and politeness aspects (especially Brown and Levinson's (1987) model of politeness and the reactions it received from proponents and detractors) (e.g. see Kasper and Rose, 1999 review). This trend has continued in L2 pragmatics, though less severely, up to the present (e.g. see Cutrone, 2011 and Halenko, 2016). In parallel with this trend, little work has been conducted within L2 pragmatics to address L2 impoliteness (but see Félix-Brasdefer and McKinnon, 2016; Iwasaki, 2011; Mugford, 2012). L2 impoliteness is part of the L2 social interaction system which L2 learners have to develop competence of in their learning process. Thus, just as L2 pragmatics research needs to investigate learners' competence and performance of L2 politeness, it also needs to do the same thing for the other end of the scale, i.e., L2 impoliteness. L2 learners experience impoliteness (production or comprehension) as part of their L2 interaction. We need to investigate how they understand and use L2 impoliteness. This will give us a clearer picture and a better understanding of the process of learning L2 pragmatics.

In fact, even within L1 sociopragmatics, impoliteness has just appeared as a rapidly growing area of research. It is still a young field despite the scholars' attempts to establish a firm framework for it (e.g. see Culpeper, 1996, 2005 and 2011), and remains dwarfed by work on politeness. However, in comparison with L2 impoliteness, L1 impoliteness seems to be in a better situation, as it has received more attention from pragmaticists (e.g. see Culpeper et al., 2017).

Irony and sarcasm are two popular subjects within impoliteness which have received proportionally more attention from researchers (e.g., see Wilson and Sperber, 1992, 2012; Colston, 1997; Kreuz and Roberts, 1995) in comparison with other impoliteness aspects. However, the main body of irony and sarcasm research is within the domain of L1. As was mentioned earlier, L2 pragmatics has invested little in impoliteness including irony and sarcasm. For example, in the literature review of the current study, only four previous L2 irony/sarcasm studies were found and reviewed (see 3.5). The current study contributes to filling the L2 impoliteness deficit by investigating L2 English sarcasm.

This study has a number of characteristics that improve on earlier studies on L2 irony and sarcasm. These characteristics also plug gaps in previous research. The first characteristic is that, unlike the previous L2 studies, the current study uses naturally-occurring data, which has the advantage of yielding more naturalistic and real-life-representing results and findings. In addition, naturally-occurring data presents for research a first-order (layperson-centred) concept of sarcasm in contrast to second-order (researcher-centred) sarcasm which has dominated previous studies. First-order-based analysis reveals how the native speakers of a language use and/or recognize a certain construct in question. The results of such an analysis should form the basis and the first steps for our understanding of a pragmatic phenomenon even when conducting a later second-order study of the same phenomenon. Sarcasm is no exception. Furthermore, the naturally-occurring data used in this study provides a further bonus for the current study which is not available in many other irony/sarcasm studies: it provides contexts for studying sarcasm in real interaction not as single and isolated utterances produced or recognized in imaginary situations (e.g. see Colston, 1997).

The second characteristic is that the present study involves online British-English sarcasm. There is a growing interest within linguistic research in computer-mediated communication and the language of the internet. Its importance has been increasing since the introduction of emails (mid of 1990s), the social communication platforms of Facebook (2004) and Twitter (2006), and the use of them for communication worldwide. However, when it comes to sarcasm, the available literature of online sarcasm tackles it within the domain of L1 (see 2.8). To my knowledge, no previous study has investigated online sarcasm as an L2 construct. Furthermore, few L2 pragmatics studies have dealt with comprehension rather than production

(e.g. Bardovi-Harlig and Dörnyei, 1998; Kim, 2014; Schauer, 2006). This study will help redress that balance.

The third characteristic of this study is that it involves Arab L2 learners of English, specifically, Iraqis. The preceding L2 pragmatics studies dealing with L2 learners of English mainly focused on Japanese (e.g. Takahashi and DuFon, 1989; Togame, 2016), Chinese (e.g. Halinko, 2016) and European learners (e.g. Schauer 2009). Arab learners, including Iraqis, are hardly ever involved in such studies. Thus, this study contributes to making up for this shortage in the literature.

Finally, the fourth characteristic is that the current study adopts a prototype view of sarcasm. That is to say, it views sarcasm as a pragmatic phenomenon consisting of a set of prototypical characteristics. It seeks to find out which of those characteristics are more prototypical and which are less according to their frequency of use (see chapter four). Furthermore, the study uncovers the influence of those characteristics on the sarcasm recognition of both native speakers and learners. To my knowledge, no previous L2 irony/sarcasm study has investigated that.

1.2 Research Questions

The study seeks to answer the following research questions:

1. Can Iraqi L2 learners of English recognize written sarcasm in British English?

2. If so, how does Iraqi L2 learners' ability to recognize written sarcasm compare to that of native speakers of English?

3. What factors influence Iraqi L2 learners' ability to recognize written sarcasm (age, gender, L2 proficiency, study abroad)?

4. What are the general pragmatic and pragmalinguistic characteristics of sarcasm that English native speakers and Iraqi L2 learners of English draw on in the process of recognizing written sarcasm? Which characteristics are more prototypical and which are less?

1.3 Structure of the Study

   This thesis falls into three parts. Part one presents the literature review of the study and consists of two chapters. Chapter two reviews some of the literature on irony and sarcasm, probing the different definitions and accounts of them. The main purpose is to locate and extract the general pragmatic and pragmalinguistic characteristics of sarcasm discussed in those accounts and begin to build up a prototypical definition of sarcasm. Chapter three reviews some of the available literature on L2 pragmatics. At the beginning, it provides an overview of the L2 pragmatics field. Then, it reviews some of the previous studies in this field. The studies are mainly arranged according to the topics they focus on: L2 proficiency and/or study abroad. Afterwards, the chapter reviews a handful of L2 pragmatics studies which have already investigated irony and sarcasm. These are the only studies found in the literature that tackle irony and sarcasm from an L2 angle. Finally, the chapter discusses some of the popular methods of data collection used in L2 pragmatics (e.g. Judgment task, Multiple-Choice questionnaire, DCT and Role-play) and highlights which one is suitable for the current study.

   Part two discusses the first study conducted in this thesis, i.e., the corpus study. This part consists of two chapters: chapter four and chapter five. Chapter four deals with how the corpus study was conducted. First, it mentions which online sources the data was collected from and why these sources were chosen in particular. Second, it also details how the metalanguage strategy, which consists in identifying sarcasm via a metalinguistic remark (e.g. *I was being sarcastic*), was operationalized for locating the required data. Third, chapter four also provides an initial analysis for the collected data. The purpose of that is to find out which general pragmatic and pragmalinguistic characteristics of sarcasm drawn from the literature review are used and how often. The more frequent characteristics are considered the more prototypical characteristics of sarcasm, whereas the less frequent ones are less prototypical. Chapter five discusses the stimuli selection task (SST). This experiment is a rating task mainly designed to select a testable amount of stimuli from the collected data for the main L2 pragmatics study of the thesis. Another purpose of the SST is to validate the sarcasm in the data by means of the native speakers' judgments (ratings). The chapter first explains how the total data (142 excerpts collected from online sources) was systematically reduced by half using the length

filter. This is because the SST is unable to accommodate all the 142 excerpts as stimuli. Then, the chapter states how the SST is piloted and how its final version is conducted. Finally, the chapter expounds the procedure followed for selecting the final stimuli for the main L2 pragmatics study.

Part three is concerned with the L2 pragmatics study of this thesis. In this part, chapter six gives details of how the L2 pragmatics study was piloted with small groups of native speakers and Iraqi EFL learners using the material selected via the SST. Chapter seven details the methods of the main L2 pragmatics study. Chapter eight presents the results of the main study and provides discussion for those results in the light of the literature reviewed in chapters two and three and the research questions. Finally, chapter nine is the concluding of this thesis in which the research questions are answered with some discussion according to the results of the corpus study and L2 pragmatics study. The chapter also summarizes the theoretical and methodological contributions of the thesis and its pedagogical implications as well. In addition, it describes the limitations of the study and provides some directions for future research.

PART 1

Literature Review

Chapter Two

Sarcasm as a Pragmatic Phenomenon: Definitions and Characteristics

2.1 Introduction

This chapter reviews some of the literature which deals with sarcasm as a pragmatic phenomenon. Section (2.2) answers the question 'What is sarcasm?' by providing some lexical definitions (2.2.1) and academic definitions (2.2.2) for sarcasm. Section (2.3) presents some of the metalinguistic terms used by scholars and researchers, on the one hand, and those used by laypeople, on the other hand, for referring to sarcasm. Some of these terms are used later in the study for collecting the required data. Section (2.4) surveys a number of verbal irony approaches which provide different views for irony as a broader phenomenon comprising sarcasm as a subtype. These approaches are the traditional approach (2.4.1), the Gricean approach (2.4.2), the echoic-mention approach (2.4.3), pretence theory (2.4.4) and the (im)politeness approach (2.4.5). Sections (2.5) and (2.6) respectively enumerate and discuss the general pragmatic and pragmalinguistic characteristics of sarcasm which are extracted mainly from the accounts already discussed. These characteristics will be checked for availability and prototypicality status later in the study. In (2.7) the researcher presents his own prototype definition of sarcasm which is mainly based on the pragmatic and pragmalinguistic characteristics of sarcasm discussed in (2.5) and (2.6). Section (2.8) reviews some of the available studies on online sarcasm.

2.2 What is sarcasm?

Several definitions have been put forward to cover different aspects of the pragmatic phenomenon of sarcasm. This section will touch upon some dictionary and academic definitions of sarcasm to demonstrate the tenor of sarcasm.

2.2.1. Dictionary definitions of sarcasm

The term sarcasm is etymologically traceable to the Greek term *σαρκασμός* (sarazein) which means to speak bitterly or to tear flesh like dogs, (Oxford English Dictionary, retrieved from http://www.oed.com/). Dictionaries typically define sarcasm as verbal irony with a

victim (Jorgensen, 1996). For instance, sarcasm is defined as "the activity of saying or writing the opposite of what you mean or speaking in a way intended to make someone else feel stupid or show them that you are angry" (Macmillan English Dictionary, retrieved from http://www.macmillandictionary.com/). Similarly, sarcasm is described as "a way of using words that are the opposite of what you mean in order to be unpleasant to somebody or to make fun of them" (Oxford Learner's Dictionary, retrieved from http://www. Oxfordlearners dictionaries.com/). In its definition of sarcasm, the Merriam-Webster dictionary (retrieved from http://www.merriam-webster.com/) mentions that sarcasm refers to sharp utterances which are satirical and ironic in nature and designed to cut or give pain. What these definitions agree upon is that sarcasm is an aggressive communicative activity directed against a victim. The next section will provide some academic definitions of sarcasm by a number of scholars and researchers.

2.2.2 Academic definitions of sarcasm

This subsection will present some second-order definitions of sarcasm, i.e., how sarcasm is seen and defined by theorists, academics and researchers. Sarcasm, as it will be stated below, is generally defined as a subtype of verbal irony. Therefore, we, first, need to know what verbal irony is.

Verbal irony is differentiated from other kinds of nonverbal irony (e.g., situational irony, dramatic irony, "irony of Fate") in being done by people by means of speaking or writing, rather than concerning events (see Barbe 1995, Leech 2014). The classic account of verbal irony, which dates back to Aristotle era, defines it as a one kind of trope in which the figurative meaning is contrary or contradictory to the literal meaning (Wilson and Sperber, 2012, p. 123). In pragmatics, pragmaticists differentiate between the *sentence meaning* and the *speaker meaning*. The former is the literal meaning of the utterance out of context which results from the combination of the semantic meaning of its words (also referred to as the *surface meaning*), whereas the latter refers to the intended meaning with which the speaker uses the utterance in a certain context (also referred to as the *underlying meaning*) (see Barbe, 1995, p. 15-16). In the case of irony, sentence meaning is the opposite of speaker meaning. For example, if someone says *What nice weather today for a picnic!* in a downpour, she

intends the opposite of her proposition: the weather is bad for a picnic. Grice (1989) adopts a view of irony that is similar to the classic one. For him, irony is also a kind of trope in which the speaker says something and intends the opposite. He accounts for irony within his Co-operative Principle (CP). He believes that verbal irony is created by flouting the Quality Maxim of the CP (Do not say what you believe to be false). The different approaches of verbal irony will be discussed in details below (see 2.4).

Coming to sarcasm, it is conceived by many researchers as a subtype of verbal irony (e.g. see Kreuz and Glucksberg,1989, p. 374). Scholars and researchers qualify this general definition with other aspects. Some researchers assert the negative nature of sarcasm. For example, Hancock (2004, 453) states that sarcasm is a type of verbal irony "in which the speaker intended the pragmatic opposite of what was said in an effort to convey a negative attitude." Colston (1997) and Toplak and Katz (2000) also point out that sarcasm is used to convey criticism or enhance negativity in general. Others assert the negative or critical nature of sarcasm along with the presence of a victim. For instance, Cheang and Pell (2008, p. 366) define sarcasm as "verbal irony that expresses negative and critical attitudes toward persons or events." Similarly, McDonald (1999, p. 486-87) states that sarcasm is "a form of ironic speech commonly used to convey implicit criticism with a particular victim as its target." Wilson (2013, p. 43) asserts that "sarcasm in particular often has a specific 'target' or 'victim': the person who is the object of the speaker's hostile or derogatory judgment". Bowes and Katz (2011, p. 219) also argue that "sarcasm– but not irony–conveys some negative attitude or appraisal and involves a victim of the verbal barb." A more extreme view is adopted by Rockwell (2000) who defines sarcasm as "a sharply mocking or contemptuous ironic remark intended to wound another" (p. 485).

Regarding the intended meaning of sarcasm, some researchers believe that since sarcasm is a subtype of verbal irony, it always involves an intended meaning that is the opposite of the literal meaning. Capelli et al. (1990, p.1824-25) write that in "ironic sarcasm…the intended meaning is the opposite of the literal meaning." Rockwell (2006, p. 3) also states that "sarcasm represents the opposite of what speakers actually mean." These definitions seem to derive from the classical account of verbal irony. However, Wilson (2006) believes that it is not always the case that sarcasm communicates the opposite. She contends that the intended

meaning of sarcasm could be just different from and not necessarily the opposite of the literal expression. A number of researchers (e.g. Kovaz et al., 2013) agree with Wilson (2006) in her view. Other scholars deal with sarcasm from a purely (im)politeness perspective. Culpeper (1996) defines sarcasm as an act of mock politeness (i.e., politeness that remains on the surface-meaning level and not really intended by the speaker) intended to cause face-threat and cause social disharmony. Likewise, Leech (2014, p. 100) also defines sarcasm as solely mock politeness and adds (p. 233) that it is "more or less limited to the snide remarks" intended to hurt others.

Although the general view holds that sarcasm expresses a negative attitude, several researchers argue that sarcasm can also convey other attitudes. For example, in their definition of sarcasm, Kovaz et al. (2013, p.599) write that sarcasm is a "subtype of verbal irony and frequently involves negativity and humor." Kim (2014, p.1) posits that "Negative emotions such as contempt, anger, dislike and frustration may lead a speaker to use harsh and bitter sarcasm, whereas positive emotions can trigger a speaker to yield light-hearted sarcasm in a friendly way."

In conclusion, researchers mention the following properties in their definitions of sarcasm:

1. Sarcasm is a subtype of verbal irony.
2. Sarcasm is double–levelled as regards meaning: it has a literal meaning and a figurative meaning.
3. The figurative meaning is the opposite of or different from the literal meaning.
4. It has a target or a victim.
5. It expresses a negative attitude.
6. It can be used to convey other than negative attitudes (upheld by a few researchers only).

One goal of the current study is to test the validity of these second-order features by means of investigating them within real-life data of sarcasm. By the time the data of this study is analysed, we shall see which of these features are supported and whether there are others that may appear and need to be used in shaping the definition of sarcasm. In fact, in order to gain a good list of second-order sarcasm features, we need to surf the different approaches to

sarcasm and extract those features from them. However, the literature does not provide separate approaches to sarcasm. Rather, as noted earlier, it deals with sarcasm within the general framework of verbal irony as being a subtype of it. Thus, we shall review the different approaches of irony for that purpose. But before that, I will briefly mention the metalanguage used by researchers to refer to sarcasm. This will be beneficial for two reasons: (1) knowing this metalanguage will help reveal some of the characteristics of sarcasm, and (2) the current study undertakes a metalanguage–reliant procedure in identifying sarcasm within the investigated corpus. Therefore, having a look at the scholars' metalanguage for sarcasm can give an insight into what terms can be used in the data collection procedure.

2.3 Sarcasm metalanguage

Within (im)politeness theory, a distinction is made between the layperson's definition or conception of a construct and that of the theorist/researcher. Watts et al. (1992, p.3) term the former *first-order politeness*, which they define as "the various ways in which polite behaviour is perceived and talked about by members of socio-cultural groups", whereas they term the latter *second-order politeness*, which refers to a "theoretical construct, a term within a theory of social behaviour and language usage", (p.3).

Regarding sarcasm, we shall start with the second-order terms used by different scholars and researchers to refer to sarcasm as a pragmatic phenomenon. Some of them were mentioned within the academic definitions in 2.2.2. The term "sarcasm" is the principal term used in almost all the studies for this pragmatic phenomenon (e.g. see Attardo, 2000 and Kim, 2014). In addition, a number of other terms are used in the definitions or descriptions given to sarcasm. For example, Leech (2014, p.233) describes sarcasm as the "snide remarks" intended to hurt others. Bowes and Katz (2011, p. 226) speak about sarcasm as "caustic comments" used by the aggressor to attack the victim's social standing. Ball (1965, p. 191) refers to sarcasm as a "form of biting communication". Rockwell (2000, p.485) defines sarcasm as a "mocking" or "contemptuous" ironic remark intended to do offence to others. Finally, Culpeper (1996, 356) uses the term "mock politeness" as an equivalent to sarcasm in his model of impoliteness.

As for the first-order sarcastic terms, not much has been written about how laypeople refer to sarcasm. The only study I found in this regard is Taylor (2015) in which the author traced the metalinguistic expressions used in informal internet language to refer to mock politeness in general including sarcasm. At the end of her study, she listed a number of metalinguistic terms used in the investigated online forum (mumsnet.com) for this purpose. Some of them are: "impolite, rude, rudeness, rudely, ironic, ironical, ironically, irony, sarcasm, sarcastic, sarcastically, sarky, laugh at, mimic, mock, tease, bitchy, catty, condescending, passive aggressive, patronise, put down, biting, cutting, caustic" (for a full list of terms, see Taylor, 2015, p.139-40).

Metalanguage is used in the current study as a strategy for pinpointing sarcasm within the searched corpus. This will be detailed in the methodology section of the corpus study (see 4.2).

2.4 Approaches of Verbal Irony

2.4.1 Traditional approach

The story of irony dates back to the ancient philosophers Socrates and Aristotle as well as to the Roman rhetorician Quintilian. For Socrates, irony is "a particular form of conversation in which one participant feigns ignorance in order to expose the ignorance of his interlocutors", (Barbe, 1995, p. 62). This definition admits an element of duality and indicates a discrepancy between appearance and reality. Aristotle, on his part, also concedes this dual nature in his description of irony. In *Rhetoric on Alexander* attributed to Aristotle (4th century BC), he sees irony as a device for blame-by-praise or praise-by-blame. Hence, irony can be evenly used to criticize or praise somebody. From this view of Aristotle, the traditional or classic definition of irony *saying something but meaning the opposite* is derived. This definition is still the dominant one to represent the traditional view of verbal irony, (see Barbe, 1995, p. 62; Knox, 1973, p. 22, cited in Burgers, 2010, p. 19).

As it is shown in Aristotle's definition, the traditional view of verbal irony maintains that irony communicates the opposite of what is stated, and the contradiction between the literal meaning and the intended meaning is essential for crystallizing irony. In fact, it is this contradiction that distinguishes it from other kinds of figurative language such as metaphor or

metonymy (see Katz, 2000 and Kim, 2014). For instance, when I say *You are a brave man!* to a cowardly person, I mean that 'the person is not brave'. The traditional view is also referred to as the *negation theory*, because the intended meaning is always the negative of the literal meaning (see Attardo, 2000, p. 797). According to this view, the hearer adopts a two-stage strategy for recognizing the ironic meaning. In her attempt to reach the intended meaning, the hearer processes both the literal and ironic meanings when encountering the ironic utterance. Due to the incongruity between the utterance and the context, the literal meaning is cancelled and dropped from consideration, whereas the ironic meaning is enhanced and identified as the intended meaning (see Attardo, 2000, p.797; Partington, 2007, p. 1549).

The traditional view of irony has garnered much criticism. The first known criticism seems to come from the Roman rhetorician Quintilian who lived in the first century AD. For him, verbal irony is a rhetorical figure and a kind of trope in which "the intention of the speaker is other than what he actually says" (Booth 1974:49). In this definition, Quintilian asserts that intended meaning of irony could be something "other than" the literal meaning, not necessarily the opposite of it. On her part, Myers Roy (1977, cited in Barbe, 1995, p. 64) criticizes the traditional definition for being inadequate because it does not explain all irony cases. She (1977, p. 171, cited in Barbe, 1995, p. 64) adds that the traditional definition of irony "give[s] no insight into why language should permit such an apparently perverse means of communication".

Wilson and Sperber (1992) believe that the traditional definition of irony does not do justice to the rich and varied nature of irony. They contend, seemingly in congruence with Quintilian, that the intended meaning of irony can be just different from and not necessarily the opposite of the literal expression. They give the following example to illustrate their point:

> You have invited me to visit you in Tuscany [in Italy]. Tuscany in May, you write, is the most beautiful place on earth. I arrive in a freak cold spell, wind howling, rain lashing down. As you drive me home along flooded roads, I turn to you and exclaim:
>
> (1) Ah, Tuscany in May!   (Wilson and Sperber 1992, p. 55-56)

In this example, the speaker succeeds in making this exclamation ironic. However, it is hard to account for this irony via the traditional view. It is difficult to conceive a meaning which is the opposite of the exclamatory utterance and might be the intended meaning.

Despite the above criticism, I still believe that the traditional view cannot be dropped altogether because of its inability to cover all the cases of irony. It can, in fact, still handle many ironic utterances in which the ironic meaning is indeed the opposite of the literal meaning. In terms of irony/sarcasm characteristics, the main characteristic the traditional approach provides is 'Contradiction' (see 2.5).

2.4.2 Gricean approach

Grice (1975, 1978 and 1989) handles irony within his two well-known models of the Co-operative Principle (CP) and conversational implicature. He adopts the traditional view that any ironic utterance has an implicit underlying meaning which is in opposition to its explicit literal meaning. Grice framed this opposition in terms of flouting the Quality Maxim of the CP (i.e., Try to make your contribution one that is true, see Grice 1975, p. 46). For example, if *Peter is a genius!* is said ironically, the ironist intends the utterance to be understood as 'Peter is not genius at all'. By so doing, s/he flouts the Quality Maxim and makes an insincere or untruthful utterance on the surface-meaning level. The hearer, on her part, is supposed to reach the intended ironic meaning by means of implicature triggered by the flout. Kaufer (1981, p. 499) considers Grices's construal a significant advance over the classical account of irony as it spells out how the opposition between the literal and ironic meanings takes place.

Grice (1975) considers the opposition between the literal and the underlying meanings of an utterance and the flouting of the Quality Maxim as two 'necessary-and-sufficient' conditions for irony to arise (see, Burgers, 2010). In a later work, he (1978, 1989) admits that "there was certainly something missing" in his early account of irony. It needs to be amended somehow in order to add more precision to it. Thus, he adds a third condition: "irony is intimately connected with the expression of a feeling, attitude, or evaluation" (1978, p. 124). To illustrate this condition, he mentions the following example:

A and B are walking down the street, and they both see a car with a shattered window.

B says:

(2) *Look, that car has all its windows intact.*

A is baffled. B says "*You didn't catch on; I was in an ironical way drawing your attention to the broken window*". (Grice, 1978, p. 124)

Grice implies that the absurdity of this example stems from the lack of evaluation which makes it difficult for A to grasp the utterance as a case of irony. Hence, Grice stipulates that for irony to be successful it should invariably be accompanied by a kind of evaluation. He also asserts that the main purpose of irony is to do criticism or to express a negative attitude; "I cannot say something ironically unless what I say is intended to reflect a hostile or derogatory judgment or a feeling such as indignation or contempt" (Grice, 1989, p. 53-54).

Grice's account was critiqued by a number of researchers for being inadequate although it was inspiring for them. An initial criticism is directed to his adoption of the traditional view of irony (see Wilson and Sperber, 2012). But the important criticism relates to the incapability of Grice's account to cover all cases of irony. Not all ironic utterances show flouting of the Quality Maxim. There are ironic cases which can flout other CP maxims such as Relevance and Quantity. For instance, Myers Roy (1978, p. 17-18) mentions the following situation, which is quoted by several later studies, to illustrate her critical point against Grice's account. A and B are in a car. A is driving and all of a sudden he takes a left turn without signaling. B ironically says:

(3) I love people who signal.

In this example, B expresses a kind of praise to drivers who uses the signal indicator before making a turn. The irony arises from the fact that A did not use the indicator in this situation and B' praise seems to be irrelevant to this event. Thus, B shows her disapproval to A's behaviour by means of producing an ironic utterance that flouts the Relevance Maxim. Example (3) cannot be accounted for by Grice's original approach as it is not intended to be insincere or untruthful. By saying (3), B expresses a truthful general opinion about liking people who use the signal indicators. Attardo (2000) attempts to expand Grice's approach by adding what he terms *the maxim of inappropriateness* in order to accommodate ironic cases such as (3) above. In fact, this is an unnecessary expansion as (3) can already be accommodated by the Relevance

Maxim of the original CP. What we need is an expansion to Grice's account of irony to include not only the flouting of the Quality Maxim, but the flouting of other CP maxims as well. And, this is what several researchers called for (e.g., Kaufer, 1981; Leech, 1983; Myers Roy, 1978).

To take Leech as another example, he (1983) draws upon Grice's account of irony (flouting the Quality Maxim) and attempts to broaden its scope to include the Quantity Maxim ("Make your contribution as informative as is required", Grice, 1975, p. 45). He writes "insincerity …may take the form of a breach of the Maxim of Quantity… or more often a breach of the Maxim of Quality", (p. 142). He (1983) mentions the case of ironic understatement and exaggeration to show how flouting the Quantity Maxim may create irony. To start with understatements, consider the following example (taken from Wilson and Sperber, 1992):

(4) You can tell he is upset. (said about somebody who is blind with rage)

In example (4), the irony is triggered by using an expression (i.e., upset) which is less informative than required to describe the state of the person in question. And, this is clearly a case of flouting Quantity. In fact, I agree with Leech that flouting the Quantity Maxim may result in irony, but I do not believe that it would be a case of insincerity. If the Quantity Maxim is flouted, the utterance would remain truthful and sincere. In example (4) above, both upsetness and rage refer two different degrees of anger. By understating rage with upsetness, the ironist is still referring to anger, not to the opposite. Then, sincerity is preserved to a degree in the utterance. A better term to describe such a case (i.e., ironic understatements) would be "uninformativeness", a term surprisingly used by Leech himself (see Leech, 1983, p. 143). On the other hand, Leech (1983) also notes that exaggerative statements (or overstatements) can also yield irony by means of flouting Quantity. Consider the following example (adapted from the American movie *The Ghost*):

> A has got angry with a woman downstairs shouting to her friend on the 5th floor. A addresses the woman ironically saying "Hey, there is an invention called TELEPHONE invented by Graham Bell in 1876 which people use to communicate. Have you ever heard of it?!".

In this utterance, A mentions unnecessary facts about a well-known invention like the telephone, its inventor, year of invention, and what it is used for. It would be absurd to mention all such details in this context unless the speaker wants to ironically express his anger with the way the woman is trying to communicate with her friend. This hyperbolic overstatement is an obvious case of flouting Quantity as it mentions more information than required.

The Gricean approach attests two irony/sarcasm characteristics. First, it confirms the 'Contradiction' characteristic proposed by the traditional approach. Second, it suggests the 'Insincerity' characteristic to account for the contradiction in irony (see 2.5).

2.4.3 Echoic mention approach

Within their relevance theory, Sperber and Wilson (1981) developed one of the most influential accounts of verbal irony, i.e., the echoic mention account, which focuses on the allusive nature of irony (revised in Sperber and Wilson, 1986 and later works). Their account comes as a reaction against what they consider the shortcomings of the traditional and the Gricean approaches. They challenge the classical and Gricean tenet that the literal meaning of an ironic utterance is substituted by its underlying opposite meaning. In this regard, they write "what irony essentially communicates is neither the proposition literally expressed nor the opposite of that proposition" (Wilson and Sperber, 2012, p. 125). Instead, Wilson and Sperber (2012, p. 129) argue that irony is a special type of echoic use. In their analysis of utterances, they differentiate between the *use* and the *mention* of an utterance or what they later term as the *descriptive uses* and the *attributive uses* of language (Wilson and Sperber, 2012, p. 128). In the descriptive use, the speaker initiates an utterance of her own to describe "an actual or possible state of affairs" (p. 128). On the other hand, in the attributive use of language, the speaker reports a thought which she "attributes to some source other than herself at the current time" (p. 128). For Wilson and Sperber (2012), echoic use is a subtype of attributive use which is not primarily intended to provide information or a thought of the speaker's own. Rather, it is used to mention a thought attributed to someone else and to convey the speaker's reaction or attitude to that attributed thought. Echoic uses can be used to convey a variety of attitudes. Wilson and Sperber (2002, 271) provide the following example to illustrate how the same echoic utterance can express a range of attitudes:

(5) Peter: That was a fantastic party.

(6) Mary: a. [happily] Fantastic.
b. [puzzled] Fantastic?
c. [scornfully] Fantastic!

In (6 a-c), Mary *echoes* Peter's description of the party as "fantastic" in the previous utterance. However, she echoes Peter's utterance differently. In (6a), Mary confirms Peter's opinion that the party was fantastic. In (6b), Mary expresses surprise about Peter's judgment and questions its validity. In (6c), Mary disagrees with Peter that the party was fantastic.

Wilson and Sperber consider irony as a special kind of echoic use. But how can irony be differentiated from other echoic uses? For this purpose, they qualify their definition of irony further with the condition that irony should implicitly express a negative attitude towards the attributed thought and "to those who might hold or have held it" (Wilson and Sperber, 2012, p. 125). In this case, only (6c) can be an ironic utterance as it echoes a previous thought and implicitly conveys a negative attitude against it. Wilson and Sperber seem to derive the condition of 'negative attitude' from Grice (1978). In that paper, Grice acknowledges that irony should reflect a derogatory judgment or a feeling of indignation or contempt. However, he does not attempt to integrate this 'negative attitude' condition into his definition of irony (see Wilson and Sperber, 2012, p. 127). Another qualification Wilson and Sperber add to the definition of verbal irony is that the speaker *dissociates* herself from the thought she mentions in the ironic utterance. In this concern, they write that in verbal irony "the speaker echoes a thought she attributes to someone else, while dissociating herself from it with anything from mild ridicule to savage scorn" (Wilson and Sperber, 1992, p. 60). On her part, the hearer can grasp irony as a result of considering three factors that work together:

1. "a recognition of the utterance as echoic",

2. "identification of the source of the opinion echoed",

3. "recognition that the speaker's attitude to the opinion echoed is one of rejection or

disapproval" (Sperber and Wilson,1986, p.240).

At its inception, Sperber and Wilson's account of irony was called *The Mention Theory*. It involved the mention (not the use) of a thought previously expressed by someone else other than the ironist and the ridiculing that thought (Sperber and Wilson, 1981). For example:

(7) The botany class was incredibly easy!

This utterance would be ironic if said, say, by a student after experiencing a great deal of difficulty in the botany class, and as a reaction to a friend's advice to take it who claimed that it is useful and easy. In this utterance, the speaker is echoing what has been proposed by somebody else with the aim of dissociating herself from the thought mentioned and showing disapproval of it (negative attitude). However, not all ironic cases allude to thoughts expressed previously by people. Hence, and due to some criticism (e.g. Kreuz and Glucksberg, 1989), Sperber and Wilson revised their account and presented the *Echoic Mention Theory*. In this revised version, the echoic theory expanded the scope of antecedent to further include generalities such as social norms, cultural aspirations, general human hopes and expectations, etc., (see Wilson and Sperber, 1992, p. 60 and Wilson and Sperber, 2012, p. 130). For example:

(8) That is really a perfect room!

If this utterance is said ironically by a tourist after arriving at the hotel she booked in and finding out the room to be terrible, the ironist in this case is not echoing any thought expressed by some other individual. Rather, she just alludes to a general failed expectation that the hotel should provide decent enough rooms to customers.

The echoic mention account of irony has the advantage of accommodating ironic cases which are not accommodated in the classic or Gricean accounts. For example, let us reconsider example (1) mentioned earlier:

(1) Ah, Tuscany in May!

The speaker was invited by a friend to come to Tuscany in Italy in May claiming that it has wonderful weather at this time of the year. Upon arrival in terrible weather, the speaker says

this utterance as a reaction against her friend's claim. Neither the traditional nor the Gricean approaches can account for the irony in this utterance as it is an exclamatory remark with no conceivable opposite. However, being an echoic utterance that mentions a previous thought of somebody else's, and in this particular case also the claim expressed that the weather is wonderful, the utterance can easily be considered ironic according to the echoic approach as it expresses a negative attitude to the echoed thought (and expressed claim) as well. The ironist criticizes the friend's claim about the beautiful weather. The echoic approach can handle similar echoic cases of irony, especially those which are not/not complete declarative sentences (see Burgers, 2010).

The echoic mention theory, however, is not devoid of criticism. The first criticism is that it is not a comprehensive theory of irony as it cannot account for all ironic cases. There are cases of spontaneous or novel irony which can hardly be seen to echo a previous thought or even some failed generality or expectation (see Barbe, 1995). Consider, for example, "You are very tall, Harry! Why don't you join our basketball team?" said ironically by a school bully to a new student (Harry), who was very short, upon meeting him for the first time. Here, the speaker did not meet the hearer before and was making no reference to any previous remark, nor was he referring to any failed general norm or expectation about how Harry was supposed to look like. Despite that, the speaker was successful in using irony in this situation without the need to echo any kind of antecedent.

The notion of dissociation enclosed in the echoic account definition of irony also received some criticism. The echoic account stipulates that the ironist dissociates herself from the attributed thought she mentions in the ironic utterance. Actually, this cannot be applied to all ironic cases. For example, when the speaker ironically says "I really like punctual trains!" upon the late arrival of a train, she does not dissociate herself from liking trains which are punctual (see Martin, 1992; Hamamoto, 1998). The third criticism is against the stipulation that irony is invariably used to express a negative attitude. Kim (2014) contends that sarcasm, which is a subtype of irony, can be used to express a *light-hearted* positive attitude. In addition, I found in several excerpts of my data sarcastic utterances used with some laughing markers (e.g. lol) or smileys which indicate that the sarcasm is being used in a friendly way and with no negativism. Consider the following example (actual pseudonyms retained):

(9)

Meditrina:

[I] wonder why the head of the "Mothers' Union" is a man?

ResurrectionByChocolate:

Yep I thought it was odd too.

DontCallMePeanut:

It's because us women need a man to speak up for us... 😁 <--- PLEASE note sarcasm before flaming starts...

 (sarcasm italicized)

In this example, the sarcasm which resides in the last utterance is accompanied with a laughing emoticon to indicate that the sarcasm is performed in a friendly and humorous way.

Sperber and Wilson's conditions of irony (i.e., echoing a previous thought, dissociation, and expressing negative attitude) appear not to be 'necessary and sufficient' conditions. However, they could be more or less prototypical features of irony the existence of which is typical rather than essential for triggering irony. Furthermore, Sperber and Wilson's account is seen to place excessive focus on the external context in which irony occurs neglecting the ironic utterance per se. Perhaps, Sperber and Wilson want to cover something in their theory that is not considered before (i.e., external context) and show its significance in crystallizing irony. Rather than being a comprehensive theory of irony in itself, I do believe that the echoic mention theory can complement Grice's theory (the modified account that includes all the CP maxims) and overcome some of its deficiencies. If considered together, both theories can result in a more efficient and sophisticated approach for handling irony.

Regarding sarcasm characteristics, the echoic mention theory puts forward two main characteristics. First, it stresses that irony/sarcasm alludes to some antecedent (whether a

previous remark or a general norm). Second, it asserts that irony/sarcasm conveys a negative attitude against the echoed thought (see 2.5).

2.4.4 Pretence theory

Clark and Gerrig (1984) introduced the pretence theory of irony as an alternative account to the echoic mention theory. Pretence theory handles irony as an act of pretence involving an imaginary speaker and an imaginary hearer interacting in some imaginary context, and irony arises through the contrast between the imaginary environment and the reality. It views the ironist as someone ''pretending to be an injudicious person speaking to an uninitiated audience'', (Clark and Gerrig, 1984, p.121). Consider the following example given by Clark and Gerrig (1984, p. 122):

A: "Trust the weather Bureau! See what lovely weather it is: rain, rain, rain"

In "See what lovely weather it is: rain, rain, rain", A is pretending to be an "unseeing person, perhaps a weather forecaster, exclaiming to an unknown audience how beautiful the weather is" (Clark and Gerrig, 1984, p. 122). Through the pretence, the speaker wants the addressee to realize the speaker's criticism to anyone who would issue or accept the utterance (exclamation) as a sincere act.

 Pretence theory may be put as follows:

> Suppose S is speaking to A, the primary addressee, and to A',
>
> who may be present or absent, real or imaginary. In speaking
>
> ironically, S is pretending to be S' speaking to A'. What S' is
>
> saying is, in one way or another, patently uniformed or injudic-
>
> ious…A' in ignorance, is intended to miss this pretense, to take
>
> S as speaking sincerely. But, A… is intended to see everything –
>
> the pretense, S''s injudiciousness, A''s ignorance, and hence S's
>
> attitude towards S', A' and what S' said.
>
> (Clark and Gerrig, 1984, p. 122)

According to pretence theory, there are two kinds of victims. The first is S', the unseeing or injudicious person the ironist is pretending to be. The second is A', the ignorant and uncomprehending audience who take(s) the ironic utterance sincerely, (Clark and Gerrig, 1984, p.122).

Some detractors (e.g. Sperber, 1984 and Utsumi, 2000) propose significant criticisms against pretence theory. First, they question whether hearers really identify S' (unseeing speaker) and A' (uncomprehending audience) when interpreting irony and consider them victims. For both Sperber (1984) and Utsumi (2000), the weather example above is likely to have no victims. Moreover, Sperber (1984) argues that pretence is not a sufficient property of irony or a distinctive feature of it. It cannot distinguish irony from non-ironies which contain pretence such as parody. For example, when one says "Tank you veddy much" in parody of an Indian accent of English, s/he also pretends to be someone else. However, the remark does not perform irony at all. Hence, Sperber (1984) asserts that "what they [Clark and Gerrig(1984)] offer as a theory of irony is a straightforward theory of parody" (p. 135). Similarly, Utsumi (2000) argues that pretence is a property of all forms of indirect speech, not only irony. Kreuz and Glucksberg (1989, p. 384) refer to this fact before Utsumi, and give the following illustrative example of indirect language which involves pretence:

(10) Can you pass the salt?

In this example, the speaker pretends to be ignorant of the hearer's physical ability to pass the salt and ask about it. However, the speaker expects the hearer to recognize the absurdity of the literal meaning of the utterance in the current situation and relies on the hearer's implicature competence to consider the utterance as a polite request for passing the salt. Example (10) meets the definition of irony within the pretence account, yet it is by no means an instance of irony.

In reaction to criticisms, Clark (1996) released a modified version of the pretence theory in which he argues that irony is seen as a joint pretence. This view assumes an imaginary situation, rather than an imaginary person, in which the speaker of irony performs a serious communicative act directed at the addressee. Irony comes into being as a result of the joint pretence of both the speaker and the addressee in the actual situation that the event in the

imaginary situation is taking place. Thus, for example, when the speaker of utterance (7) above says "The botany class was incredibly easy!", both the speaker and the addressee jointly pretend to be in an imaginary situation in which the utterance is meant literally. But this version also has some limitations of its own. First, it assumes that the addressee of irony must share the ironic intention with the speaker beforehand in order to pretend jointly. However, this is not the case in many instances of irony. Second, this version, like the original pretence theory, is also incapable of distinguishing irony from parody and any other non-ironies containing pretence. Finally, the 'joint pretence' theory states nothing about how to treat the victims of irony although it seems more capable of explaining victimless irony than the original theory (Utsumi, 2000, 1782-83).

2.4.5 (Im)politeness approach

As was mentioned in the academic definitions of sarcasm (2.2.2), some scholars handle irony and sarcasm from a purely politeness perspective. The earliest scholar, to the best of my knowledge, who wrote about the relationship between irony and politeness is Leech (1983). He (1983) devised the *Politeness Principle* (PP) to aid Grice's CP and account for how comity and harmony are preserved among people. For Leech, verbal irony is an implicit, rather than explicit, act of violation to the PP which is used to cause offense (impoliteness). To state his view of verbal irony, he proposes the *Irony Principle*[1] which reads as follows:

> If you must cause offence, at least do so in a way which doesn't
> overtly conflict with the PP, but allows the hearer to arrive at the
> offensive point of your remark indirectly, by way of
> implicature. (1983, p. 82)

In this delineation of irony, Leech indicates that irony: (1) is used to cause offence, (2) is overtly in harmony with the PP, (3) is an indirect act, and (4) has a victim who grasps the offence by implicature. Culpeper (1996) argues that the IP form as it is does not depart much from Brown and Levinson's (1978, 1987) 'Off-record' strategy of politeness, which is proposed as a way to lessen face-threat and maintain social harmony. However, Leech (1983) later in his book makes it clear that "the IP, by enabling us to bypass politeness, promotes the

---

[1] Here, I handle the irony principle as it is (i.e., exclusive to irony) although I disagree with calling it " Irony Principle" because it is too broad to include irony only. With the current formula, it may apply to any act of impoliteness that can be conveyed indirectly with no overt conflict with the PP (e.g. the act of taunting can be done in accordance with this principle).

'antisocial' use of language" (p. 142). And, this is definitely the opposite purpose of the off-record politeness strategy (see Culpeper, 1996, p. 357).

Consistent with irony being overtly in harmony with the PP, Leech (1983, p. 144) also defines irony as merely *mock politeness*, "an apparently friendly way of being offensive". Culpeper (1996, p.356) builds on Leech's account of irony and also speaks about mock politeness in his model of impoliteness. He (1996, p. 356) defines mock politeness as the use of obviously insincere politeness strategies, which remain surface realisations, to perform an FTA (Face-Threatening Act). For example:

(11) What a genius you are! (said sarcastically to someone stupid)

In this utterance, the speaker pretends to show admiration to the hearer through 'mocking' a polite form (an exclamation containing the semantically positive noun *genius*). The polite language remains on the propositional semantic level, whereas the intended pragmatic meaning is something else (e.g. the opposite of the literal proposition). Culpeper (1996) prefers to associate, or actually equate, mock politeness with sarcasm rather than irony as Leech does. He argues that irony is a general category which can encompass comic as well as offensive acts, whereas sarcasm is a subtype of irony which consists in using mock politeness to cause offence and social disharmony (see Culpeper, 1996, p. 357). Being an offensive act, sarcasm is seen by Culpeper (1996) to invariably have a victim which is the person the sarcasm is directed against.

In a later work, Leech (2014, p. 100) uses the term 'conversational irony' and 'sarcasm' interchangeable to solely mean mock politeness. This definition, which is also adopted by Culpeper (1996, 2005), implies that sarcasm cannot be other than mock politeness. However, I disagree with this opinion and believe that sarcasm is a more complex pragmatic phenomenon than being merely confined or equated with mock politeness. There are cases of sarcasm which hardly involve mock politeness. To illustrate the point, consider this Chinese example along with its translation cited from Leech (2014, p. 101):

(12) 你起得太早了，天都还没亮呢！

"You got up so early! It's still dark outside" (said sarcastically by a Chinese father to his

son who woke up very late).

Although Leech mentions (12) as an example of conversational irony (or sarcasm) as mock politeness, the example seems problematic in terms of being polite on the surface level. It just consists of two statements which have no politeness implications, in terms of grammar and/or semantics, to be exploited for mock politeness. The example is neither polite nor impolite and falls into the neutral zone on the politeness scale.

Note that I am not denying or underplaying the role of mock politeness in creating sarcasm. On the contrary, I believe that mock politeness is one of the outstanding characteristics of sarcasm that can perhaps occur in most everyday cases. However, I am against the view that confines sarcasm to mock politeness only as there are sarcastic cases, such as (12) above, which have no politeness on the surface level that can be mocked. In partial support, Taylor (2015) also argues against equating sarcasm with mock politeness and comes to the conclusion that sarcasm is only one realisation of mock politeness.

Other researchers who shed light on mock politeness include Haugh (2014) and Taylor (2015). Haugh (2014) speaks about mock politeness implicatures, which he defines as "an ostensibly 'polite' stance, which is indicated through the occurrence of a (non-)linguistic form or practice that would in other circumstances be associated with a polite attitude, masks or disguises an 'impolite' stance that arises through implicature" (p.278). In this definition, Haugh refers to mock politeness as an act of implicature (an act within the hearer's domain) and, more interestingly, to the possibility that mock politeness can be a non-linguistic behaviour. Taylor (2015) also speaks about mock politeness as an act involving implicature. She writes "mock politeness occurs when there is an im/politeness mismatch leading to an implicature of impoliteness", (Taylor, 2015, p.130). Taylor (2015) claims that her definition is broader than Haugh's (2014) as it talks about im/politeness mismatch rather than just politeness that "masks" or "disguises" impoliteness.

Taylor (2015) attempts to position mock politeness within Culpeper's (2011) model of *implicational impoliteness* which is the following:

(1) *Form-driven*: the surface form or semantic content of a behaviour is marked.

(2) *Convention-driven*:

    (a) Internal: the context projected by part of a behaviour mismatches that projected by

         another part; or

    (b) External: the context projected by a behaviour mismatches the context of use.

 (3) *Context-driven*:

    (a) Unmarked behaviour: an unmarked (with respect to surface form or semantic

       content) and unconventionalised behaviour mismatches the context; or

    (b) Absence of behaviour: the absence of a behaviour mismatches the context.

<div align="center">Adapted from Culpeper, 2011, p.155-156 (italics in original)</div>

She (2015, p.129) argues that mock politeness falls within the second category of 'convention-driven impoliteness'. She adds that what is important about this model is that it accounts for two kinds of im/politeness mismatches: internal (co-textual) mismatch and external (contextual) mismatch.

   Taylor's view can gain support. For example, Utterance (11) above displays a contextual mismatch between the utterance itself (*What a genius you are!*) and the person it is said to (a stupid person). In addition, internal co-textual mismatch occurs when a polite piece of language is used side by side with an impolite piece within the same utterance resulting in possible sarcasm. This usage is termed as *verbal formula mismatches* by Culpeper (2011, p.174) and as *attitude clash* by Leech (2014, p.283). Culpeper's (2011, p. 174) paradigm example to illustrate how such a mismatch works is "Could you just fuck off?". The example starts with the conventionalized politeness formula of request "Could you just…" which may lead the sarcasm target down the garden path that a polite request is being issued. The last bit "fuck off (= go away)" is impolite in one of the rudest possible ways, and forces the target to retrospectively reinterpret the utterance as sarcasm (see Leech, 2014, p. 238). I believe that

this kind of sarcasm is much easier to recognize as it can solely be triggered by the utterance form itself, without much need to the extra-linguistic context.

As for sarcasm characteristics, the main characteristic this approach focuses on is mock politeness. Another characteristic is the involvement of a victim against whom the irony/sarcasm is directed (see 2.5).

2.5 General Pragmatic Characteristics of Sarcasm

This section lists a number of 'general pragmatic characteristics' of sarcasm which have been identified from the definitions and approaches discussed earlier. The term is adopted from Leech (1983) who applies *general pragmatics* to "the general conditions of the communicative use of language" (p. 10). In fact, several of the below-listed characteristics are, strictly speaking, sociopragmatic by nature (e.g. mock politeness and victim), whereas others are not (e.g. contradiction). Thus, I prefer to use the term 'general pragmatic characteristics' under which all these characteristics can be subsumed. Most of the general pragmatic characteristics are derived from the approaches of irony, but they are assumed to apply to sarcasm as it is a subtype of irony. The experiments conducted in this study will verify which of those characteristics apply to sarcasm and which do not, and also to what degree.

Allusion to an Antecedent

Sarcasm always refers to some antecedent. This characteristic is extracted from Sperber and Wilson's echoic mention theory. The antecedent can be specific (e.g., a previous remark of somebody else) or general (e.g., a cultural norm or a social expectation) (see 2.4.3).

Contradiction

Contradiction here simply means 'saying something and meaning the opposite'. This characteristic of sarcasm is derived from the traditional and the Gricean approaches to irony which both maintain that verbal irony communicates the opposite of what is literally said (see 2.4.1 and 2.4.2).

Insincerity

Insincerity is derived from Grice's approach of irony. It refers to triggering sarcasm by means of flouting the Quality Maxim of the CP. When doing sarcasm, the speaker, in many cases, says an untruthful remark about the victim and wants him/her to reach the intended sarcastic meaning by means of implicature. For example, when I say "What a kind person you are!" to an unkind man, he will realize that I am insincere and untruthful in what I say and will interpret the remark as sarcasm by implicature.

Flouting Quantity or Relevance

Sometimes, sarcasm is created via flouting CP maxims other than Quality, namely, flouting Quantity or Relevance. This characteristic is derived from the scholars' reactions to Grice's approach which aimed to broaden its scope (see 2.4.2 for Leech's (1983) and Myers Roy's (1978) arguments for including flouting Quantity and Relevance in irony creation respectively).

Negative Attitude

Generally speaking, sarcasm is seen by many scholars as an impoliteness device used typically for causing offence by means of showing a negative attitude indirectly. We saw in our discussion of sarcasm definitions and irony approaches how 'Negative attitude' was embodied in the definition of irony/sarcasm within these different accounts (see 2.2.2 for the scholars' definitions of sarcasm and 2.4.2, 2.4.3 and 2.4.5 for the irony approaches which include a negative attitude). Bowes and Kats (2011) claim that exacerbating rather than muting negativity has been reported more in the literature of sarcasm.

I support the view that sarcasm is prototypically used for doing impoliteness and causing offence by means of conveying an indirect negative attitude. However, I disagree with the absoluteness that sarcasm invariably does so. As was mentioned earlier, I agree with Kim (2014) that there is room for sarcasm to be used in a friendly way and convey a positive emotion or attitude (see 2.4.3, the third criticism against the echoic mention theory). The

conclusion that sarcasm can be used to convey a positive emotion has first-order evidence in this study (see 9.3.1). However, it should be explored more in future studies before this conclusion is confirmed.

Mock Politeness

Mock politeness is another general pragmatic characteristic of sarcasm which involves a double-levelled nature. It involves the use of obviously insincere politeness strategies, which remain surface realisations, to perform an FTA (Face-Threatening Act). In other words, the sarcastic polite language remains on the surface level and is not truly intended. Mock politeness is mainly derived from the (im)politeness approach of irony and the works of Leech (1983, 2014) and Culpeper (1996, 2005) (see 2.4.5).

Victim Involvement

Involving some kind of victim is a key characteristic which many scholars have embodied in their accounts of irony/sarcasm. Three kinds of victim have been identified in the literature reviewed above:

1. The person against whom the sarcasm is directed (see sarcasm definitions of Cheang and Pell (2008), McDonald (1999), Wilson (2013), Rockwell (2006), Leech (2014) in 2.2.2 and Culpeper (1996) in 2.4.5);
2. The unseeing or injudicious person the ironist is pretending to be (see Clark and Gerrig, 1984, in 2.4.4).
3. The ignorant and uncomprehending audience who take(s) the ironic utterance sincerely (see Clark and Gerrig, 1984, in 2.4.4) ; and
4. The previous remark or thought the ironist is alluding to and "those who might hold or have held it" (Wilson and Sperber, 2012, p. 125, see 2.4.3).

2.6 Pragmalinguistic Characteristics of Sarcasm

In addition to general pragmatic characteristics, a number of pragmalinguistic characteristics that are associated with sarcasm are mentioned in the literature. These characteristics have more to

do with the sarcastic utterance *per se* rather than the external context in which it occurs. Their presence can be considered, more or less, a possible indicator that sarcasm is being used. Among these characteristics are the following.

Positive wording

This pragmalinguistic characteristic is closely related to "mock politeness", i.e., it is one means whereby mock politeness is done. In most cases, sarcastic utterances are worded positively: (1) they either contain lexical items carrying positive semantic meanings (e.g. *You are a genius!*), or (2) the sarcastic utterance itself can be a formulaic expression associated with doing something polite (e.g. using *Could you……?* for mocking a polite request), (see Colston, 2002, Rockwell, 2006, Culpeper, 2011). Metaphor can also serve as a positive-wording strategy in sarcastic utterances (e.g. *You are a lion!* said sarcastically by an officer to a cowardly soldier) (see Utsumi, 2000). It is important to note here that although I said that positive wording is closely related to mock politeness, this does not necessarily mean that all cases of positive wording do mock politeness. An important use of positive wording in everyday life is to do politeness (e.g. *What a genius he is!* said genuinely in admiration to a scientist), but there are also plenty of positive evaluations that have nothing to do with (mock) politeness (e.g. *What wonderful weather today!* said in a sunny and breezy day). The point I am making in this section is that 'mock politeness is mainly done by positive wording' and not that 'positive wording is mainly doing mock politeness.'

Positive wording being an important prototypical characteristic of sarcasm in English has been revealed in a great deal of research (e.g. see Colston, 1997, 2002; Kreuz and Glucksberg,1989; Gibbs, 2000). But this does not mean that all sarcastic utterances are positively-worded. Neutrally-worded cases of sarcasm are also possible and applicable in everyday situations. For example, utterance (12) above (*You got up so early! It's still dark outside*, said by a Chinese father to his son who woke up very late) is a one case of using neutral language in sarcasm. Moreover, I also believe, based on Leech's (2014, p.233) amended Irony Principle, that negative-worded sarcasm is also possible when the ostensible sarcasm victim is the speaker him/herself. The first part of Leech's amended Irony Principle says:

In order to be ironic, S expresses or implies a meaning (let's call it Meaning I) that associates a favorable value with what pertains to O (O = other person(s), mainly the addressee) <u>or associates an unfavorable value with what pertains to S (S = self, speaker).</u> At the same time, by means of Meaning I and the context, S more indirectly implies a second, deeper meaning (Meaning II) that cancels out Meaning I by associating an unfavorable value with what pertains to O, or associating a favorable meaning with what pertains to S. (Leech, 2014, p. 233). (my emphasis underlined)

For example, consider the following situation:

Situation 1

 A is a car driver who was about to hit a pedestrian B:
 A: Sorry, mate. I didn't see you.
 B: (fuming with anger) You are right not to see me because **I am too TINY to be seen**!

By saying "*I am too tiny to be seen*", B, who is a mature adult, is using sarcasm in his reply to A. With the false self-description (*too tiny*), B apparently directs the sarcasm to himself. Nevertheless, he ultimately wants to criticize A's inability to see him by means of recognizing the falsity of that description. Describing the self as being "too tiny" is semantically negative and associates an "unfavourable value" with the speaker in Leech's terms. In other words, the sarcastic utterance is worded negatively. Accordingly, room for mock impoliteness in sarcasm is possible as well. In fact, this is another proof that sarcasm can be other than mock politeness. Leech's (2014) definition of conversational irony (conversational irony is *mock politeness*) seems to contradict what he asserts in his amended Irony Principle in the same book (i.e., the possibility of associating an unfavorable value with what pertains to the speaker). Thus, conversational irony (or alternatively 'sarcasm' in Leech's terms) could be something beyond mock politeness. This kind of contradiction attracts reveals the complex and varied nature of sarcasm.

Rockwell (2006) also speaks about negative wording in sarcasm. However, she provides irrelevant examples (such as *You look awful!* said to somebody who looks so elegant) to support the claim that it is sarcasm. It is clear that *You look awful!* in such a situation is interpreted as a case of banter (mock impoliteness) rather than sarcasm.

As was mentioned earlier in (2.4.5), sarcasm can also combine positive and negative wording together (see Culpeper 2011). Sarcastic utterances of this kind can start with a polite piece of language followed by an impolite one such as "*Could you fuck off?*" or vice versa as in "*SHUT THE FUCK UP, please!*". In support, Partington (2011) also states that irony can be triggered by juxtaposing two elements with opposing evaluative polarity. In other words, ironic force is created by means of a one element collocating with an antonym which it does not normally co-occur with. He terms this kind of irony as *evaluative oxymoron*[2]. For example,

 (13) You are the *wisest fool* I have ever met! (said by a king to his stupid advisor)

The co-occurrence of the adjective *wise* in its strongest superlative form with *fool* which carries an opposing semantic meaning creates a sense of oddness leading ultimately to interpret the utterance as sarcasm. In fact, I believe, as was mentioned earlier in (2.4.5), that the false combination of two contradictory pieces of language can help more easily trigger sarcastic interpretation, without much need for contextual factors.

 Hyperbole

   Leech (2014, p. 234) argues that exaggeration (or hyperbole in rhetoric terms) is a way of making the overt meaning of an utterance infelicitous and ultimately prompts ironic interpretation. He states that exaggeration flouts the Quality Maxim in the sense of overstating the truth. Sarcastic hyperbole refers to an exaggerated linguistic form the speaker uses to indicate that sarcasm is in operation. In the literature, researchers mention certain exaggeration-indicating forms on the word level and the sentence level which seem to have

---

[2] Oxymorons ''are traditionally defined as figures of speech that combine two seemingly contradictory elements'' (Gibbs, 1993:268).

been conventionalized for doing hyperbole. In turn, hyperbole can signal sarcasm if used in a sarcasm-triggering context.

On the word level, extreme adjectives can serve as indicators of hyperbolic sarcasm (e.g. *That's fabulous!*). In addition, superlatives can also serve the same purpose (*You've got the biggest mind ever!*) (see Kovaz et al., 2013).

On the sentence level, hyperbole can have different versions. Hyperbole can occur by repeating a word in the same utterance (e.g. *Thank you very very much!*) (see, Kreuz and Caucci, 2007). Liebrecht et al. (2013) found that sarcasm can be hyperbolic by means of using intensifiers and exclamations. Hyperbole can also arise by the co-occurrence of an adverb+(extreme)adjective (e.g. *He is really smart!, That's extremely amazing!*). Hancock (2004) terms such adverbs or adjectives used to exaggerate a statement as *verbal amplifiers*. In their study about sarcasm in the internet language, Kovaz et al. (2013) found that this adverb-adjective combination is quite prevalent in the data they investigated. They list this combination within their *Lexical Cues Hypothesis*, where they propose lexical cues that trigger sarcasm in utterances (other cues being interjections and positive emotion terms). In fact, before Kovaz et al. (2013), it was Kreuz and Roberts (1995) who attracted attention to this adverb-adjective combination as a probable indicator of irony in general. They came up with what they called the *Random Irony Generator*, which is a linguistic frame within which one can combine an adverb with an extreme positive adjective (e.g., *really fabulous*) in an utterance to generate irony (For a list of possible adverbs and adjectives, see Kreuz and Roberts, 1995, p. 25, Table1). In partial support of Kovaz's et al. (2013) finding, Partington (2007) conducted a corpus study of irony and found that "very noticeable was the number of adverbial intensifiers found in the company of 'irony' and 'ironic', including: *deeply, particularly, especially, indeed, certainly, doubly, genuinely, bitterly and supremely*", (p.1551). Seto (1998, p. 244) also argues that ironists use adverbs to intensify or exaggerate the literal meaning as a signal for the addressee to "reverse the polarity", i.e., cancel the literal meaning and recognize the ironic meaning instead.

Hyperbole can also take the form of overpoliteness represented by overindirectness. Leech (1983) contends that an increase in the level of indirectness results, in principle, in an increase in the level of politeness. Accordingly, the high level of politeness created by overindirectness

can be exploited, depending on the context, by the ironist as a hyperbolic strategy to achieve a higher degree of mock politeness when doing sarcasm. For example, "*I'm so sorry to trouble you, but could you kindly be quiet for a moment?*" said by a father to his little daughter whilst trying to speak on the phone (see Culpeper, 2011, p.166). In this example, the father uses hedging and highly-polite expressions (overindirectness) in a situation that does not sustain that (i.e., speaking to his little daughter). The purpose is to create hyperbolic mock politeness leading to a high level of sarcasm. In addition to all that, a high degree of hyperbole may arise through using expressions indicating utmostness or entirety (e.g. *You are the cleverest/most genius man in the whole world/on the planet!*).

Graphological cues of sarcasm

Sarcasm as a complex pragmatic phenomenon draws on several extra-linguistic contextual factors for triggering and grasping the sarcastic meaning in addition to the linguistic ones. Factors like tone of voice, facial expressions, and kinesics help a lot in conveying and recognizing verbal sarcasm (see Wilson, 2013). However, sarcasm becomes more difficult to understand in written language (e.g. internet language) due to the absence of those extra-linguistic factors. As a substitute, a number of graphological cues have been conventionalized and used by ironists to convey sarcasm in writing. Many of these indicators are discussed in the literature such as capitalization to foreground something or show emphasis (Culpeper, 2011), vowel elongation (e.g. *You are soooooooooo smart!*) (Shively et al., 2008), emoticons and exclamation marks (Wang 2013, Carvalho et al., 2009). As regards internet language in particular, Wang (2013) writes that it "utilize[s] visual cues as in capitalization, emoticons, punctuation, and hashtags to show the real intention of the speaker in order to achieve the effect of sarcasm and irony" (p.355-56). Likewise, Carvalho et al. (2009) identified a number of graphological cues used in internet language for denoting irony. These are: (1) emoticons, (2) onomatopoeic expression for laughter, (3) heavy punctuation marks, and (4) quotation marks. It was also noted that a number of graphological cues are used in my data for indicating sarcasm (e.g. capitalization, emoticons and laughing markers) (see 4.2.2, Table 3 for the full list of graphological cues and their frequencies).

2.7 A Prototype Definition of Sarcasm

Different scholars have provided different definitions to irony and sarcasm according to the theory and the view they adopt (e.g. see Attardo, 2000; Clark & Gerig, 1984; Culpeper, 2011; Giora, 1998; Leech, 2014; Wilson & Sperber, 2012). Some of them were mentioned in (2.2.2). As a result, no consensus is available among scholars about what sarcasm is. Some extreme views adopt a position of all-or-nothing towards sarcasm. They attempt to impose necessary and sufficient conditions for sarcasm without which it cannot come into being (e.g. see Sperber and Wilson, 1986). Those attempts cannot do justice to the varied and complex nature of this pragmatic phenomenon. As a way out of the pitfalls of these views, I prefer to adhere to prototype theory and provide a definition for sarcasm accordingly.

Prototype theory came about as a reaction to the strict and rigid Aristotelian theory of categorization (classic theory). For Aristotle, any category is distinguished from others by possessing a bunch of distinctive features. And, anything cannot be a member of a category unless it has ALL its distinctive features (see Taylor, 1995). For example, Aristotle believes that 'Man' category has the distinctive features [+two-footed] and [+animal]. These features are necessary for any entity to be categorized as 'Man'. Meanwhile, they are sufficient to classify any entity as a 'Man' (Taylor, 1995, p. 22-23). Failure to show any of these features results in excluding from the 'Man' category. That is why the classic theory is described as *all-or-nothing* theory.

On the other hand, prototype theory also relies on feature possession for categorization. Any category is distinguished by a set of features. The entity (or entities) which own all the category features is the most central and representative member of that category, i.e., the prototype of the category. Entities with less features are still members of the category but they are less central and representative, and the grading continues up to the edge of the category which contains the peripheral members with the least features. In fact, "prototype categories have a flexibility, unknown to Aristotelian categories, in being able to accommodate new, hitherto unfamiliar data" (Taylor, 1995, p.53). It is devoid of the rigidity of the classic theory and open to include new members without the need to restructure the category itself (i.e., redefining the category by modifying the criteria of inclusion).

In what follows, I provide a prototype definition of sarcasm based on the characteristics mentioned in (2.5) and (2.6). Such a definition is condition-free and more inclusive of cases of sarcasm. It tackles the characteristic-sarcasm relationship as a direct proportion (a matter of more or less): the more characteristics are available the more the sarcastic interpretation is enhanced, and vice versa. The prototype of sarcasm would be a case that shows all the characteristics mentioned in the definition. But if any characteristic is missing, this would not render the case as 'not sarcastic', only the sarcastic probability of the utterance would lessen, i.e., it would be less prototypical.

My prototype definition of sarcasm is as follows:

> Sarcasm is typically a double–levelled pragmatic phenomenon that is a subtype of verbal irony wherein the intended meaning is the opposite or different from the literal meaning and is understood by implicature. It is triggered by the flouting of the Cooperative Principle, usually the Quality Maxim, to create a sense of insincerity. It alludes to either a definite (e.g. a previous remark) or general (e.g. a social norm or expectation) antecedent. It mainly conveys a negative attitude against a target or a victim. In many cases, it utilizes positive wording and/or hyperbolic forms to do mock politeness which is exploited, in turn, as a means for conveying the negative attitude.

There are downsides to prototype theory which push it away from perfection (see Taylor, 1995 for details). A key downside relevant to my study is the overlap of features of the sarcasm category with other categories. For example, allusion to an antecedent can be a feature of parody as well. However, I believe that sarcasm cannot come into being by possessing a single feature only. There should be a minimum number of features for sarcasm to arise. Investigating what that minimum number should be and how each feature is weighted is beyond the scope of the current study, but is worth a study of its own. The second relevant downside is that sarcasm, like all prototype categories, has fuzzy boundaries, a matter acknowledged in general by prototype theorists, (see Taylor, 1995). As a result, peripheral cases are candidates to be overlapped by other similar categories such as parody or banter.

2.8. Online Sarcasm

This section reviews some of the studies that have been conducted on online irony/sarcasm. All the below-listed studies deals with online irony and sarcasm as L1 constructs. The reviewed studies were noted to mainly focus on how online irony/sarcasm is identified (with what cues and indicators) and on how to design an automatic system for detecting irony or sarcasm online. Table 1 below gives a summary of these studies.

Table 1

A Summary of Some Studies which Investigate Online Irony or Sarcasm

No.	Study	Focus	Methodology	Data	Major finding
1	Kreuz and Caucci (2007)	Sarcasm recognition in Google Books	1.Searching for and collecting instances explicitly marked as sarcastic by their authors. The prompt used is "said sarcastically". 2. Removing the word "sarcastically" from all the collected material. 2. Asking participants via a judgment task to rate sarcasm	1. One hundred excerpts from Google books containing sarcasm. 2. Fifteen control excerpts from Google books which contain no sarcasm	1. participants rated the excerpts which originally contained the word "sarcastically" as more likely to be sarcastic than the control items which did not contain that word. 2. the use of interjections within the excerpts such as *gee* and *gosh* helped the participants significantly to detect and rate sarcasm.
2	Carvalho et al. (2009)	Identifying typical indicators of irony	Investigating the comments submitted online to the website of a popular Portuguese newspaper in search for irony indicators	A set of comments submitted online to the website of a popular Portuguese newspaper	Five textual clues were found to signal the presence of irony. These are emoticons, onomatopoeic expression for laughter, heavy punctuation marks, quotation marks, and positive interjections.
3	Kovaz et al. (2013)	Identifying lexical indicators of sarcasm	Examining two kinds of online datasets which are explicitly marked as being sarcastic by their authors: (1) Twitter posts, (2) dialogues from Google Books	They collected 1222 tweets from Twitter and 100 statements from Google books.	1.Coming up with the *Lexical Cues Hypothesis* which lists some lexical indicators of sarcasm 2. Among these cues are adjective-adverb combination, terms of positive affect, and interjection
4	Liebrecht et al. (2013)	Investigating how sarcasm is detected on Twitter	Searching for and collecting tweets which are explicitly marked with the hashtag '#sarcasm'	78000 Dutch tweets	1. sarcastic utterances are characterized by using positive literal meaning (surface meaning), hyperbolic expressions (including intensifiers and exclamations) and explicit markers

					(hashtags) 2. Hyperbolic sarcastic instances are able to stand alone, whereas non-hyperbolic ones are marked with the hashtag (#sarcasm). 3. the explicit marker (hashtag) is used in sarcastic tweets to compensate for the nonverbal behaviour people use when expressing sarcasm
5	Tsur et al. (2010)	Designing an Auto-detecting system for online sarcasm	1. Collecting reviews on Amazon product which contain sarcasm 2. Asking annotators via a judgment task to rate sarcasm in the data 3. Extracting some properties of sarcasm which were later employed in the automatic algorithm they designed for recognizing sarcasm	66000 reviews of Amazon product	They claim their system to be novel and possesses a high level of accuracy (77%)
6	González-Ibánez et al. (2011)	Designing an Auto-detecting system for sarcasm on Twitter	1.Collecting tweets which were indicated to be sarcastic by their authors 2. Studying the impact of the lexical and pragmatic factors on the machine effectiveness in identifying sarcasm 3. Comparing the performance of the machine detection to human judgments	Sarcastic Tweets	The results demonstrated that the machine performance was as good as the human. But in both the accuracy was low
7	Filatova (2012)	Designing an Auto-detecting system for online sarcasm	Collecting and examining reviews on Amazon products	1000 reviews from the Amazon website	One finding was that sarcasm was mainly found in negative reviews which give low scores to the product reviewed
8	Reyes et al. (2012)	Designing an auto-detecting system for irony and humour in social media	Collecting tweets which contain the hashtags '#humour' or '#irony'	50000 tweets	The results were positive for the case of humour and encouraging for irony.

9	Ptacek et al. (2014)	Designing an Auto-detecting system for online sarcasm	Collecting tweets which contain sarcasm	7000 Czech tweets	They claim their approach to outperform the state-of-the-art methods used in English for the same purpose.
10	Farías et al. (2016)	Designing an Auto-detecting system for irony on Twitter	1. Collecting ironic tweets 2. Using a wide range of irony-related vocabulary which reflects various facets of affect	Ironic tweets	Results showed that the inclusion of affect-indicating words helps a lot in distinguishing ironic utterances from non-ironic ones.

2.9 Summary

This chapter has provided the first part of the literature review which relates to sarcasm as a pragmatic phenomenon. First, it has provided some lexical and academic definitions for sarcasm to answer the question 'What is sarcasm?', then it touched upon the terms used by researchers and laypeople to refer to sarcasm. Afterwards the chapter presented different accounts and approaches to verbal irony and sarcasm. These are traditional approach, Gricean approach, echoic-mention approach, pretence theory and (im)politeness approach. Mainly out of the discussed accounts and definitions of verbal irony, a number of general pragmatic and pragmalinguistic characteristics of sarcasm were extracted. The pragmatic characteristics are 'Allusion to an antecedent', 'Contradiction', 'Insincerity', 'Flouting Quantity or Relevance', 'Negative attitude', 'Mock politeness' and 'Victim involvement'. The pragmalinguistic characteristics are 'Positive wording', 'Hyperbole' and 'Graphological cues'. Basing on these characteristics, the researcher provided his own prototype definition of sarcasm after listing and discussing them. Finally, the chapter reviewed some of the available studies on online sarcasm and revealed what these studies have focused on. The next chapter will do a literature review for the field of L2 pragmatics within which this study is mainly conducted.

Chapter Three

Second Language Pragmatics

3.1 Introduction

   This chapter provides a literature review for the field of second language pragmatics (L2 pragmatics). It starts with presenting an overview of this field touching upon what it is concerned with, how it began, some definitions of the field and the main topics that have been researched in it (3.2). Sections (3.3) and (3.4) review some studies on two L2 pragmatics parameters that are relevant to what is done in this thesis. Section (3.5) reviews the a few previous works on L2 irony/sarcasm. Afterwards the chapter presents some of the popular instruments used in L2 pragmatic for data collection (3.6). Section (3.7) lists the research gaps that are located while reviewing the literature. Finally, the chapter re-lists the research questions of the thesis in order to remind the reader with and provides some predictive tentative answers for them according to the literature reviewed (3.8).

3.2 An Overview of Second Language Pragmatics

   Second language pragmatics– also known as L2 pragmatics – is a subfield of study which is seen as the confluence between second language acquisition (SLA) and pragmatics (see Kasper and Rose, 1999). It is concerned with examining L2 learners' pragmatic knowledge, use and development in the target language (Taguchi, 2017). The story of second language pragmatics dates back to the 1970s when Selinker (1972) coined the term *interlanguage* within the general field of SLA to refer to the special system of rules L2 learners develop for the target language they are learning. As part of SLA, the term has been applied to learning the pragmatics of the target language to form *interlanguage pragmatics* (ILP). Interlanguage pragmatics is now used interchangeably to mean second language pragmatics or L2 pragmatics (Taguchi and Roever, 2017). When first introduced, interlanguage pragmatics was originally defined as a subfield within SLA which studies the non-native speakers' (NNSs) production and comprehension of speech acts of the target language and how the knowledge related to them (speech acts) is acquired (see Kasper and Dahl, 1991). Kasper and Rose (2003) introduced a two-fold definition of interlanguage pragmatics which accommodates both the

pragmatic competence and pragmatic performance of L2 learners. As the study of the L2 pragmatic competence, ILP probes how L2 learners develop the ability to produce and comprehend pragmatic acts within the target language. And, as the study of L2 pragmatic performance, ILP investigates the use (production and comprehension) of these pragmatic acts in definite target-language contexts. These definitions were influenced by the overwhelming focus on L2 speech acts studies since the rise of the field along the 1980s and 1990s.

However, the definition of ILP was later broadened to include areas beyond speech acts in response to the various pragmatic aspects investigated. For instance, Taguchi (2017) attempts the following definition of ILP. She writes that ILP is "a branch of second language acquisition (SLA), [which] examines second language (L2) learners' knowledge, use, and development in performing sociocultural functions" (p. 153). In this definition, Taguchi uses the term 'sociocultural functions' in a broader sense to include in addition to speech acts other aspects such as the pragmatic routines, cultural norms and politeness rules. She affirms that for L2 learners to have mastery of the target language, they need to have, among other things, proper knowledge of its pragmalinguistic and sociopragmatic rules. The former represents the knowledge of the linguistic forms and rules used to realize the various pragmatic functions, whereas the latter refers to how to use and apply those linguistic forms and rules appropriately in the relevant situations (see Leech, 1983, p.10-11). A similar broad account of ILP is maintained by Bardovi-Harlig (2010, p. 1) who writes "[pragmatics] bridges the gap between the system side of language and the use side, and relates both of them at the same time. Interlanguage pragmatics brings the study of acquisition to this mix of structure and use". Taguchi and Roever (2017) point out that despite the discrepancies in the definitions provided for interlanguage pragmatics, the area of research in this field remains the same: "L2 learners' knowledge and use of language in social interaction" (p. 5).

Although interlanguage pragmatics is seen as the intersection between SLA and pragmatics, it owes most of its being to pragmatics rather than SLA. ILP has borrowed a lot from L1 empirical pragmatics, especially studies on speech acts, politeness and cross-cultural pragmatics (see Kasper and Rose, 1999, p. 81-82). Kasper and Rose (1999, p. 82) argue that the single SLA issue ILP has borrowed and addressed is pragmatic transfer. In fact, the level

of L2 proficiency and the effect of study abroad are two further SLA properties which have been prolifically researched in ILP to date (see 3.3 and 3.4).

Taguchi and Roever (2017, p. 16) assert that pragmatic studies in SLA are traced back to the late 1970s. The earliest ones were studies on speech act production by L2 learners and their use of politeness strategies (Kasper, 1979; Scarcella, 1979). But "The publication of Kasper's (1981) dissertation put interlanguage pragmatics firmly on the map in SLA research" (Taguchi and Roever, 2017, p. 16). Since the establishment of L2 pragmatics as a field of study, it has witnessed a rapid growth in the body of empirical research conducted within its framework. In the early years, cross-linguistic studies were dominant in L2 pragmatics. This is obvious through the big body of studies in the 1980s and the 1990s which addressed especially the differences in pragmatic behaviour across the investigated languages. The premise on which these studies was based is that "different cultures and languages have different ways of encoding pragmatic notions of politeness or directness into linguistic behaviors, and that these differences often serve as sources of LI transfer and areas of difficulty in learning" (Taguchi and Roever, 2017, p. 9). Taguchi (2017, p. 159) states that despite the possibility, in principle, of all aspects of pragmatics to be subject to cross-cultural pragmatic studies, L2 studies in this area concentrated, at that time, on investigating especially speech acts and politeness in the target language.

Most of the studies conducted in the early period of the ILP were cross-sectional in nature investigating mainly the use rather than the acquisition/development of speech acts (Taguchi, 2017, p. 167). The milestone and most seminal work which informed many of the studies then was the Cross-Cultural Speech Act Realisation Project (CCSARP) by Blum-Kulka, House and Kasper (1989). This project investigated and analysed requests and apologies in seven languages via a discourse completion test (DCT). A major finding is that it could reveal many culture-specific properties of the tested speech acts in those languages.

Longitudinal studies, which tracked the development of L2 learners' knowledge and use, were rare in the early period of ILP. Most developmental insights came from cross-sectional studies by means of comparing each investigated stage of learning to the other higher or lower stages. By so doing, the line of progress was tracked (see Taguchi, 2017, p. 160). The outstanding longitudinal study at that time was Schmidt (1983). It was a case study of Wes, a

Japanese artist who was naturalized in Hawaii. Schmidt traced the progress Wes was doing in learning English (target language) over a period of three years. Schmidt's study was a serious attempt to open the door wide to conduct longitudinal studies, which are true developmental studies, within this new-born field. Unfortunately, longitudinal studies remained less preferred in L2 pragmatics during that period, and it took about a decade until the second prominent longitudinal study appeared, i.e., Ellis (1992) which examined the development of two ESL learners' in initiating requests within a classroom setting (see Taguchi, 2017).

Taguchi (2017, p. 154) writes that starting from 1990s, the focus of research in L2 pragmatics shifted from cross-linguistics studies to exploring the instruction and assessment of L2 learners' pragmatic competence. In the same period, a considerable increase took place in the body of longitudinal studies addressing the acquisition and development of the target language pragmatics. The subject-matters of these studies ranged from the traditional investigation of speech acts and implicature to newly-researched constructs such as the interactional features that facilitate the learner's participation in speech events (see Taguchi and Roever, 2017; Taguchi, 2010, for a review).

The first decade and a half of the third Millennium witnessed a further growth in the ILP research which tackled the instruction, assessment and acquisition of L2 pragmatics. Researchers started applying the mainstream SLA theories to the field of interlanguage pragmatics (Taguchi, 2017, p. 154). Taguchi and Roever (2017, p.9) summarize the expansion in L2 pragmatics research since the beginning of this century as follows:

> The first decade [and a half] of this century saw further growth of instructional and acquisitional research characterized by a more explicit application of mainstream SLA theories to ILP studies, as well as technology applications to teaching and testing. At the same time, the analytical framework of pragmatic competence shifted from 'pragmatics-within-individuals' to 'pragmatics-in-interaction-in-context' by drawing on the concepts of interactional competence (Young, 2002, 2008, 2011) and discursive pragmatics (Kasper, 2006b). With present-day globalization and transnationalism, ILP has further expanded its empirical scope to the areas of intercultural competence (Spencer-Oatey & Franklin, 2009), pragmatics in lingua franca communication (House, 2010), L3 pragmatic acquisition (Alcon Soler, 2013b), and heritage learner pragmatics (Taguchi, Li, & Liu, 2013; Xiao-Desai & Wong, forthcoming).

A big body of studies is now available which probes various issues of L2 pragmatics (for a review, see Kasper and Rose, 2003; Schauer, 2010; Taguchi, 2012, 2015, 2017). In what follows, a literature review will be done on three aspects of L2 pragmatics which are relevant to the topic researched in the current study. These are the effect of L2 proficiency, the effect of study abroad, and L2 irony and sarcasm. In addition, I will also review some popular data collection instruments in L2 pragmatics.

3.3 L2 Proficiency Effect

Many L2 pragmatics studies are concerned with investigating the role of L2 proficiency in learning and developing the L2 pragmatic competence. A great number of these L2 proficiency studies are on the production of speech acts and pragmatic functions and routines (see Kasper and Rose, 1999; Taguchi 2012 and 2013). Scarcella (1979) is one of the earliest studies in L2 pragmatics. She compared the politeness strategies produced by two levels of ESL learners (beginners and advanced) to their English-native-speaker counterparts. The focus of the study was the development of L2 pragmatic competence in inviting and requesting. The author designed three videotaped open role-play contexts for this purpose. Results showed that with increased proficiency, learners approximated the level of native speakers in using the investigated speech acts.

In fact, not too many studies involved beginning learners like Scarcella (1979). Kasper and Rose (1999) state that the majority of proficiency studies compared intermediate to advanced learners and excluded beginners. This is because, generally speaking, beginning learners have not reached the required threshold of L2 pragmatic competence that enables them to comprehend and use the various pragmatic constructs correctly.

Takahashi and DuFon (1989) investigated the production of L2 requesting by Japanese learners of English. The study showed that with increasing proficiency, the learners opted for using more native-like conventions of requesting. Further and important work on requesting was undertaken by Trosborg's (1995) whose monograph "represented an important contribution to the study of acquisition in L2 pragmatics" (Barron, 2012, p. 52). The study was on the production of three L2 speech acts: requesting, complaining and apologizing. Although this was a cross-sectional study, it aimed at investigating the development of learners'

pragmatic competence of these speech acts. Participants were three groups of Danish learners of English: secondary school (grade nine), high school and university students. No proficiency test was done for the participants, but it was assumed that each educational level represents a different proficiency level. Role-play was the instrument used in this study for eliciting data. Among the study's findings was that the higher the proficiency level the more native-like the learners were as regards the investigated constructs.

Other studies dealt with the production of L2 apologies. For instance, Trosborg (1987) compared the production of apologies by Danish EFL learners at three proficiency levels to that of English and Danish native speakers. She used the role-play instrument for collecting data from participants. She noted that the learners' repertoire of pragmatic routines of apology increased with increase in proficiency. Modality markers (e.g. hedges and intensifiers) increased with high proficiency to approximately the native-level. More recently, Dalmau and Gotor (2007) also investigated the L2 apology production by 78 Catalan EFL learners at three proficiency levels. The study was centred on the learners' use of apology IFIDs (Illocutionary Force Indicating Devices), what apology strategies they employed and the type and degree of apology intensification they used. Results demonstrated that learners with high proficiency produced more native-like apologies than those with lower-proficiency.

L2 pragmatics proficiency studies on speech act comprehension seem to be less than their production counterparts. One outstanding example of such comprehension studies, which was seminal for later studies and replicated by several of them (e.g. Schauer 2006), is Bardovi-Harlig and Dörnyei (1998). It investigated the grammatical and pragmatic awareness of mainly two groups of L2 learners of English: ESL learners in the US and EFL learners in Hungary. The study also involved a secondary group of EFL learners in Italy. The total of participants in this study was 543 learners. The researchers designed their own instrument which consisted of a number of videotaped scenarios along with a two-fold judgment task containing a *yes-no* question and a rating scale. Each scenario is imagined to have a male and female students interacting with each other and the interaction ends up with a request, suggestion, apology, or refusal. All scenarios contained either a grammatical or a pragmatic error. Results showed, among other things, an obvious effect for proficiency in comprehending the speech acts in question. Within the EFL groups, the low-proficiency learners gave lower ratings for the

deliberately-incorporated grammatical and pragmatic errors vis-à-vis the high-proficiency ones. However, the high-proficiency learners demonstrated a greater awareness of the grammatical errors than the pragmatic ones. Regarding the ESL group, results also showed that the high-proficiency learners were more capable of recognizing the pragmatic errors in the scenarios than the low-proficiency learners.

Some L2 pragmatics proficiency studies combined the investigation of both the comprehension and the production of L2 speech acts. An example of such studies is Koike (1996) which examined the comprehension of L2 suggestions and the production of suggestion responses by English L2 learners of Spanish. The study involved participants from different proficiency levels (beginners, intermediate and advanced) who were all university students learning Spanish. Results showed that the performance of advanced students was significantly better than the other groups on both the comprehension and production accounts.

The current study will investigate the effect of L2 proficiency upon the Iraqi learners' recognition of online English sarcasm. They will be divided according to proficiency level and check their performances for any differences which can be attributed to proficiency.

3.4 Study-Abroad Effect

The effect of studying a target language in a native environment upon developing the L2 learners' pragmatic competence has been an area of interest since the early days of L2 pragmatics. Studies in this domain are mainly of two types: (1) cross-sectional studies which compare the performance of study-abroad L2 learners to their counterparts who study L2 at home at a certain point of time (single-moment studies), and (2) longitudinal studies which track the development of L2 learners' pragmatic competence over a period of time of studying abroad (see the reviews of Schauer, 2010 and Taguchi, 2011).

Single-moment studies seem not to be too many in the literature. One example is Takahashi and Beebe (1987), which investigated the development of L2 refusals by Japanese learners of English. The participants were 80 in total: 20 Japanese EFL learners in Japan, 20 Japanese ESL learners studying in the United States, 20 Japanese native speakers and 20 American-English native speakers. Both Japanese and English native speakers were employed as control groups. The instrument used for collecting data was a written DCT. They found,

among other things, that ESL learners did better than EFL learners as regards acquiring and using L2 English refusals.

Bardovi-Harlig and Dörnyei (1998) also investigated the effect of study abroad in addition to the effect of proficiency. They compared the ability of ESL learners to detect grammatical and pragmatic errors to that of EFL learners. They found that the ESL learners were generally better than the EFL ones in detecting pragmatic inappropriateness. Schauer (2006) replicated and extended Bardovi-Harlig and Dörnyei's (1998) study. She investigated the L2 English learners' awareness of pragmatic infelicities. Three groups of participants were involved in the study. The first comprised 16 German ESL learners studying abroad at a British university. The second included 17 German EFL learners studying in Germany, whereas the third was a control group containing 20 British-English native speakers. Results revealed that "the German EFL participants were less aware of pragmatic infelicities than the ESL group and that the ESL learners increased their pragmatic awareness significantly during their stay in Great Britain" (Schauer, 2006, p. 269-270).

Taguchi (2011) designed a cross-sectional study to test the effect of both proficiency and study abroad upon the pragmatic comprehension of L2 learners of English. She investigated the learners' conventional and non-conventional implicatures as well as their indirect refusals using a computerized pragmatic listening test. Three groups of learners were recruited in this study. They were different in terms of proficiency level and study-abroad experience. Group 1 (22 participants) was characterized with low proficiency and zero study-abroad experience. Group 2 (20 participants) had a higher proficiency level and no study-abroad experience. Group 3 (22 participants) was almost of the same proficiency level of Group 2, but had at least a one-year experience of studying in an English-speaking country. A fourth group of English native speakers (25 participants) was also employed as a control group. Taguchi found, among other things, that study-abroad has an effect upon the learners' "comprehension of nonconventional implicatures and routine expressions but not [on] indirect refusals" (Taguchi, 2011, p. 904).

Longitudinal studies that investigate the development of learners' L2 pragmatic competence while studying abroad are more pervasive in the literature. An early instance is presented by Olshtain and Blum-Kulka (1985). They examined the learners' perception of directness and positive politeness in L2 Hebrew during their stay in Israel. Request and

apology were analysed as two topics of the study-abroad experience. Rating-scale assessments demonstrated that learners initially gave low scores for the more direct strategies and positive politeness involved in L2 requests and apologies. By so doing, they were relying on their own L1 cultural and social norms (L1 negative transfer). However, after a relatively long stay in Israel, their later assessments revealed more tolerance of directness and positive politeness that coincides with the Hebrew native norms. Thus, the study attests an obvious effect of study abroad upon developing L2 pragmatic competence. It also proves that L1 negative transfer in learners' performance diminishes the longer they stay in an L2-speaking environment (see Kasper and Rose, 1999).

More recently, Schauer's (2007, 2008, 2009) studies were concerned with ESL learners' production of external modifiers in requests. Nine German ESL students were recruited for the study who were studying L2 English abroad in a British university for a full academic year. Data was elicited from those learners at three stages: at the beginning of their sojourn in October, in the middle of the year in February and shortly before the end of their stay in May. Control for the study was provided by 13 German EFL students studying in a higher educational institution in Germany and 15 British-English native speakers also studying at a British university. The data collection instrument was a computer-based multimedia elicitation task (MET). Sixteen requesting scenarios were included in the instrument which differed in terms of the interlocutor's social status (higher/equal) and the degree of imposition the request has upon the addressee (high/low). Schauer found that ESL learners developed a broader repertoire of external request modifiers after the sojourn they spent in England. In addition, the direct requesting strategies, imperatives and unhedged performatives which the ESL learners had used at the first data collection point (at the beginning of their stay) decreased significantly at the final data collection session towards the native norms. Thus, the study-abroad experience can result in gains in L2 pragmatics development. Moreover, Schauer also noticed that not all the ESL learners benefit from the study abroad experience to the same extent. They differ among each other in the level of L2 pragmatic development depending on the amount of interaction with the native speaker, how much exposure they have to the L2 input and their motivation to learn the L2 (see Schauer, 2010; Barron and Warga, 2007).

Another study that also highlighted the importance of interaction with native speakers in developing learners' pragmatic competence is Bardovi-Harlig and Bastos (2011). This study investigated the role of proficiency, length of stay and intensity of interaction in the L2 learners' recognition and production of conventional expressions (e.g. *No problem, Nice to meet you,* and *That'd be great*) in the target language. The study was conducted on 122 ESL learners studying at an American university. In addition, controls were provided by 49 American-English native speakers. Learners' proficiency was determined by an exam, the length of stay abroad was counted by month, and intensity of interaction was measured "by self-report of weekly English language use outside class with native speakers, daily use with other learners, and television viewing" (p. 347). The recognition data was collected by means of an aural task in which "the expressions were digitally recorded by a single speaker in a sound booth and were played to participants through individual headsets" (p. 360). The production data was elicited via a computer-based task involving delivering the required responses orally. Results revealed that intensity of interaction had a "significant influence" on both the recognition and production of conventional expressions, whereas proficiency was found to have an influence on the production of those expressions only. Length of stay, on the other hand, had no significant effect upon either the recognition or the production of conventional expressions.

Barron and Warga (2007) write that "the general findings from study abroad investigations … reveal that exposure to second language input triggers some important developments" (p.117-118). They add that "While many of these developments lead to an increasingly L2-like pragmatic competence, it has also been shown that some aspects of pragmatic competence do not change at all" (p. 118). An example that shows this latter fact comes from Bouton (1992, 1994) which investigated the development the ESL learners showed in comprehending L2 conversational implicatures. In these studies, the learners were asked to complete a written test containing 33 short dialogues. After each dialogue, there was a question asking about the implicature of a target utterance within the dialogue. The dialogues incorporated different types of implicatures. Results pointed out that after a 17-month sojourn in the target language community, learners still lag behind native speakers' comprehension of conversational

implicatures, particularly indirect criticism, Pope questions[3], sequence implicatures, and irony. After 4 years and a half, learners' comprehension of relevance-based implicatures (i.e., implicatures that draws on general inferencing mechanisms) reached virtually the native level. However, formulaic implicatures which are culture-specific (e.g., Pope questions and irony) were still difficult to comprehend even after this long stay in the L2-speaking community (see Taguchi, 2011).

The current study will, in addition to proficiency, also investigate the effect of studying abroad upon the Iraqi learners' recognition of online English sarcasm. The performance of the Iraqi L2 learners of English studying in the UK will be compared to that of the Iraqi learners studying at home. This is to verify whether coming to and studying in the UK would perfect the learners' recognition of sarcasm.

3.5 L2 Irony and Sarcasm

Shively et al. (2008) state that there is a scarcity of studies dealing with irony and sarcasm within L2 pragmatics despite the large body of literature written on them in general. Only a handful of studies were found that attempted to probe how irony/sarcasm is produced or recognized in the target language by L2 learners. The following is a summary of these studies.

Bouton's (1999) was another stage within his series of studies that probes the ESL learners' perception of English conversational implicatures. Two studies of this series were referred to in the previous section of study abroad (Bouton, 1992, 1994). In Bouton (1999), the researcher looked at whether learners' comprehension of conversational implicatures (including irony) can be facilitated by explicit teaching. It is a longitudinal study whereby the author tested a group of English L2 learners from a variety of L1 backgrounds at the beginning of their stay in the U.S., 17 months later, 33 months later, and 54 months later respectively. Learners were presented with written situations and were asked to signal their understandings of implicatures of single utterances via a multiple–choice instrument. Uninstructed learners

---

3 The Pope question "*Is the Pope Catholic?*" is an informal humorous or sarcastic response to another question for which the answer is obviously 'Yes' (retrieved from https:// www.urbandictionary.com; https://en.oxforddictionaries.com/. For example, if on seeing someone in the military uniform one says "Are you a soldier?" and the answer was "Is the Pope Catholic?", the answer implies "Yes, of course! There is no doubt I am a soldier".

showed improvement over time in perceiving irony and all other conversation implicatures. But especially with irony, they did not reach a native–like level of perception even after 54 months of being in an English-speaking environment. "However, Bouton discovered that explicit classroom instruction and awareness-raising about irony was effective in helping learners improve their skills in interpreting irony", (Shively et al., 2008, p. 107).

Yamanaka (2003) also examined the comprehension of conversational implicature. It investigated the effect of L2 proficiency and the length of residence in L2 environment on the pragmatic comprehension of Japanese ESL learners. The focus of the study was on comparing NS' pragmatic comprehension of implicature to that of the learners. Thirteen American English NSs as well as 43 Japanese ESL learners participated in the study. The participants were presented with 12 video clips taken from different television programs each of which was followed by Multiple-Choice options defining possible implicatures. Within the investigated implicatures, the study dealt with the comprehension of irony and sarcasm. Results show that ''the NNSs in this study found irony (and its sub-category, sarcasm) to be among the most difficult implicature types'' (p. 138-39).

Unlike Bouton's (1999) and Yamanaka (2003) which tackled conversational implicatures in general, Shively et al. (2008) conducted a study about the perception of verbal irony in specific in Spanish as a second language. Participants were 55 Spanish learners studying in a public American University in the Midwest. All participants were American-English native speakers except one student. They were divided into three proficiency groups (20 beginners, 17 intermediate students and 18 advanced students) according to the Spanish course they were in (second, fourth or sixth). The researchers used quasi-natural data in this study: a set of excerpts taken from L1 Spanish films. The researchers tested two groups of learners. The first group was presented with written scripts of the excerpts only, whereas the second one was allowed to watch the excerpts on-screen in addition to the written scripts. The objective was to "examine the interpretation of ironic utterances in Spanish-language films by L2 learners of Spanish and the impact of an audiovisual context on the ability of learners to interpret irony", (Shively et al., 2008, p. 101). The results support the previously-attested view that the recognition of irony improves as proficiency level and experience in the target language increase. In addition, there was a weak support to the hypothesis derived from Yus (1998,

2000), i.e., the more audio and visual resources are available at the hearer's disposal, the easier the irony perception will be. This hypothesis proved to be significant for only the advanced learners tested in the study.

More recently, Shively did another study (Shively 2013) in which she dealt with the production of humour in L2 Spanish. Shively (2013) is a longitudinal case study. The participant is an L2 Spanish learner (Kyle) who is spending a semester-long study-abroad stay in Toledo, Spain. Among other things, the participant used sarcasm in the target language as a one strategy to perform L2 humor. Results show that the participant, over the study-abroad course, became more proficient in using L2 humor with his NS friends. A main factor in developing this proficiency was the close friendship that he developed with his NS peers. The close relationship offered Kyle (1) the opportunity to acquire the NS-strategies of humor and (2) a high level of intimacy which enabled him to produce humor with his NS close friends.

Another quasi-natural-data study is Kim (2014) which, unlike Shively et al. (2008), focused on the perception of L2 sarcasm more specifically rather than verbal irony in general. Participants were 28 Korean EFL learners (11 males and 17 females) who all studied English in South Korea and had no study-abroad experience in any English-speaking country. Her data was a number of sarcastic video clips taken from the famous American TV sitcom "Friends" along with written scripts. Participants were asked to complete three tasks after watching each clip and reading the script: (1) sarcasm identification task, (2) speaker's intent comprehension task, and (3) potential sarcasm cue identification task. A follow-up interview with every participant was done to obtain an in-depth understanding of the answers. "Analysis revealed that learners drew upon certain features of L1 schema during the L2 comprehension process [of sarcasm]" (p. 1). Thus, this study attests an L1 negative pragmatic transfer in the Korean learners' comprehension of L2 English sarcasm.

Peters et al. (2015) investigated the role of context and prosody in comprehending sarcasm. They compared the sarcasm-comprehension of English native speakers to that of L2 learners of English (whose L1 is Arabic). The authors constructed a number of 3-sentence spoken discourses. For each discourse, the first sentence introduced an action done by someone (Person 1). The second sentence introduced a context for that action (either positive or negative). The third sentence presented another person's (Person 2) reaction towards that

action (said with either sincere or sarcastic prosody). Each discourse ends with a Yes/No comprehension question which elicits whether the participant believes that Person's 2 reaction was sincere or sarcastic. Results show that in the case of context-prosody matching, English native speakers did better in identifying sarcasm. In such cases, NSs relied on both context and prosody in identifying the sarcastic meaning. However, when context and prosody conflicted, NSs relied more on context than prosody in sarcasm comprehension. On the other hand, L2 learners appeared to rely exclusively on context in their sarcasm comprehension in almost all cases.

Finally, Togame (2016) investigated the perception of L2 irony by Japanese ESL learners adopting a relevance-theoretic account. The focus of the study was on "the extent to which non-native speakers of English understand potentially ironic utterances in a similar way to native speakers" (p. ii). For this end, the researcher designed and conducted two experiments with written and spoken nature respectively. The first experiment consisted of twenty imaginary stories each of which contained a target utterance. The participants' task was to read the story and rate how ironic the target utterance was on a scale from 'not at all' to 'very much'. The experiment was conducted online via 'Survey Monkey' website. Fifty three Japanese ESL learners participated in this experiment as well as 22 British English native speakers who provided a control. The second experiment was prosodic in nature. It also contained some devised stories (22 brief stories) which were narrated by a professional English native speaker. Each one included a target utterance which was repeated three times with different tones: " 'basic' (a kind of default, unmarked tone), 'deadpan' (with a narrower pitch range), and 'exaggerated' (with a wider pitch range)" (p. ii). The task was that the participant listens to the narrated story and identifies the ironic utterance, if any. Thirty five Japanese ESL learners took part in this experiment. In addition, a control for this experiment was provided by 30 native speakers. The first online experiment yielded "surprising results, suggesting that Japanese speakers can respond to potentially ironical utterances similarly to native speakers" (p. ii). Regarding the second aural experiment, "the results indicated that Japanese participants could perceive English prosodic structure [of irony] in similar ways to native speakers and were affected by prosodic contours in similar ways" (p. ii).

In fact, to the best of my knowledge, no previous L2 irony study has used naturally-occurring data extracted from real-life situations. Film excerpts and video clips can be considered quasi-natural or real-like data. But they cannot be as genuine as real-life situations, especially because they are controlled by their writers, producers and directors. Furthermore, natural sarcasm in online English blogs and forums has been hardly ever explored to find out how it works and how it is recognized by L2 learners of English. This kind of sarcasm almost lacks the audiovisual contextual factors which proved to facilitate sarcasm recognition in previous studies (see Kim, 2014 and Shively et al., 2008). Thus, the comprehension of such sarcasm could be problematic to L2 learners. The current study attempts to find out how online real-life sarcasm works and how it is recognized by EFL learners.

3.6 Research Methods in L2 Pragmatics

Several research and data collection methods have been used in L2 pragmatics since the inception of the field. Some of them are already used in the superordinate field of SLA (e.g. DCT), whereas others are borrowed from other disciplines (e.g. 'Response Time Measure', which is borrowed from psycholinguistics). Giving an entire overview of all research methods in L2 pragmatics is something beyond the purpose of this section. Thus, only some of the popular methods are going to be mentioned here. For a broader overview of L2 pragmatics research methods, see Kasper and Dahl (1991), Schauer (2009), Roever (2011) and Taguchi and Roever (2017).

One way of categorizing L2 pragmatics research methods is according to the kind of data to be elicited (i.e., perception or production). Perception data consists mainly in the participants of an experiment providing their judgments about a measured construct. Production data, on the other hand, are of two types: (1) non-interactive data which comes from, for example, a response to a prompt without being involved in an encounter, (2) interactive data which is collected via participants' interaction with each other (see Taguchi and Roever, 2017, chapter four). In what follows, some popular method will be mentioned for each kind of data.

3.6.1 Perception data

Judgment Task

Judgment tasks (alternatively known as 'Metapragmatic judgments' or 'Acceptability judgments') "elicit respondents' perceptions about a pragmatic feature… or their comprehension of implied meaning" (Taguchi and Roever, 2017, p. 77). L2 pragmatics researchers make use of this instrument to draw conclusions about L2 learners' awareness of the different target-language pragmatic aspects. The basic version of an item within a judgment task consists of two parts: (1) a stimulus which is usually a written scenario (with a target utterance), and (2) a (Likert) scale with even or odd number of points. The participants' task is to read the scenario and pass judgments about it on the scale (see Taguchi and Roever, 2017). Figure 1 below gives an illustrative example of a judgment task item.

Susan is leaving work and realizes that she just missed the train home. The next train is in an hour. She asks her colleague, Patrick, who lives near her, for a ride:

Susan 'Patrick, I just missed my train. Can I catch a ride with you?'

Patrick 'No, you can't.'

How appropriate is Patrick's response?

Entirely	Mostly	Somewhat	Somewhat	Mostly	Entirely
inappropriate	inappropriate	inappropriate	appropriate	appropriate	appropriate

Figure 1. Illustrative judgment task item (adopted from Taguchi and Roever, 2017)

Among the recent studies which employed a judgment task is Roever et al. (2014) which tested the learners' knowledge of L2 English sociopragmatics. They used imaginary scenarios with a target utterance in each. The scenarios exposed varied social relationships among the interlocutors in terms of power, distance and degree of imposition. By means of a five-point Likert scale, the researchers asked the participants to rate the appropriateness of the target utterances on a scale range from 'very impolite' to 'far too polite'.

Different variants of judgment tasks have been introduced and used by L2 pragmatics researchers. For instance, Matsumura (2003) employed a multiple-choice judgment task. She devised scenarios which require giving advice to an imaginary addressee. Instead of using a rating scale, the author provided the ESL learners participants with four choices in each scenario to choose from what suits the scenario best. Li and Taguchi (2014) used audio input instead of written scenarios and Bardovi-Harlig and Dörnyei (1998) used video clips as their scenarios. To the best of my knowledge, no L2 comprehension study has used naturally-occurring data as prompts for a judgment task.

Regarding the pros and cons of judgment tasks, this instrument provides insights about learners' pragmatic perception, especially their knowledge of the L2 sociopragmatic system. Judgment tasks also have the merit of being an instrument less affected by proficiency level than production tasks. Furthermore, judgment tasks are less demanding for participants as they require comprehension only on their part. However, they cannot collect data about the learners' production abilities and cannot support conclusions about learners' deployment of pragmatic knowledge. Nor can this instrument measure learners' pragmatic performance. The instrument has not also been widely-used in studies on L2 pragmalinguistics (see Taguchi and Roever, 2017, p. 77).

Another disadvantage of this instrument relates to the coding of the data obtained from the Likert scale. Usually, points on a Likert scale are assigned ascending numerical values. For example, a Likert scale with five categories ranging from 'entirely inappropriate' to 'entirely appropriate' may have the value of 1 to 'entirely inappropriate' and 5 for 'entirely appropriate'. This value assignment assumes even intervals between the values on the scale which may not be the case. Taguchi and Roever (2017) summarize this problem as follows:

> Assigning point values with equal intervals (5, 4, 3, 2, 1) to response options implies that the difference in acceptability between an option worth five points and an option worth four points is the same as between an option worth two points and an option worth one point. That may not be true, however: the four-point option may just be slightly less suitable than the five-point option, whereas the one-point option may be entirely inappropriate and cause grave offence, while the two-point option might be clumsy but inoffensive.   (Taguchi and Roever, 2017, p. 83)

Despite this limitation, judgment tasks and Likert scales continue to be used by researchers as no method or instrument is entirely perfect, limitation-free and devoid of problems.

Rank-ordering tasks

This method is among the earliest methods of measuring the pragmatic comprehension of NS and NNS participants. Rank-ordering tasks reported in the literature are mainly of two formats:

a. *Card-sorting*

This format involves providing participants with different sets of cards. Within each set, the first card describes a situation under test (e.g. requesting something). The other cards within the same set contain utterances that can be said in the described situation. Then, the participants are asked to put the utterances cards into a certain order (e.g. from the least to the most polite) (see Schauer, 2009). Card-sorting instrument was used in two of the earliest pragmatic-comprehension studies: Carrell and Konneker (1981) and Tanaka and Kawade (1982), both cited in Schauer (2009). ''Card-sorting has two main advantages: first, it is a very inexpensive method and secondly, it can be administered relatively quickly which allows researchers to collect data from a large number of participants'' (Schauer, 2009, p. 62). The major disadvantage of this instrument is that it can only provide a brief summary of the situations in which the utterances are to be used due to the limited number and size of cards (see Schauer, 2009).

b. *Filling in a questionnaire*

Rank-ordering task can take the form of filling a questionnaire (e.g. Olshtain and Blum-Kulka's, 1985 and Kitao, 1990). This instrument is similar to the judgment task in that participants are asked to rate the appropriacy of a set of utterances in a given situation on a rating scale (e.g. from 'very rude' to 'very polite'). In addition to sharing the abovementioned advantages of card-sorting, this instrument has the further advantage of ''allowing researchers to investigate a higher number of scenarios and to provide more detailed contextual information'' (Schauer, 2009, p. 63). However, the instrument, due to

being paper-based, remains far more restricted in presenting contextual information than the audiovisual methods (see Schauer, 2009).

Multiple-Choice questionnaire

Another instrument for testing pragmatic-comprehension is Multiple-Choice questionnaire. In this instrument, participants are provided with a number of scenarios each of which is followed by a set of sentences. The sentences are ''either interpretations of an utterance that is contained in the scenario's description, or possible responses to the scenario'' (Schauer, 2009, p. 63). Participants are asked to read the scenario and respond to it by typically selecting one option from the provided choices. Several studies used this instrument including Bouton (1994) and Hinkel (1997).

As for the pros and cons, this instrument, as is the case with rank-ordering tasks, has the advantage of being inexpensive, easily and quickly administered to a great number of participants, and capable of describing scenarios in details. The main limitation of the method is that participants need to rely on their imagination power to envisage the scenario context due to the absence of audiovisual elements. Thus, a disparity may occur among the participants in conceiving a scenario (e.g. whether it is friendly or hostile) and responding to it accordingly (see Schauer, 2009). To address this limitation, Schauer (2009, p. 63) suggests providing ''detailed instructions'' to the participants about the scenario context. However, this cannot eliminate the possibility of misconception completely. Other researchers made use of the modern technology to develop this instrument. For example, Taguchi (2008) used a computerized version of Multiple-choice questionnaire. She provided the participants with audio input to enable them make judgments based on the speaker's prosody. ''This is a very encouraging development that shows how existing instruments can be modified to address their limitations'' (Schauer, 2009, p. 63).

Multimedia instruments

The technology revolution in the 1990s, especially after introducing the internet, opened the door wide before researchers to exploit this technology in advancing research methods. L2 pragmatics researchers are no exception. They started to employ the multimedia technology to

collect data from participants. The abovementioned case of Taguchi (2008) is an instance of this technology-use method. Other researchers used audiovisual methods to provide participants with more contextual factors. One of these methods is the video-and-questionnaire task in which the researcher films a video containing a number of scenarios. The video is then presented to the participants along with a questionnaire about the videotaped scenarios (see Schauer, 2009). For example, Bardovi-Harlig and Dornyei's (1998) used a video-and-questionnaire task wherein they presented their participants with a number of videotaped interactions containing inappropriate utterances. Then, participants were asked to rate the severity of the inappropriate utterance on a scale.

Compared to paper-based methods, the audiovisual methods have the undeniable advantage of providing participants with ''a higher degree of carefully controlled contextual information. Thus, they considerably decrease the probability that utterances might be assessed differently based on the individual participants' imagination/perception of the scenario'' (Schauer, 2009, p. 64). Participants can make more accurate judgments based on the prosody and the body language they hear and see. However, these methods are not devoid of limitations. The main disadvantages of are: (1) the high expense in comparison to the paper-based methods (e.g. hiring actors and video recording), and (2) the high demand of logistics which makes the implementation of these methods restricted to well-equipped locations (i.e., locations equipped with the required TV screens and computers) (see Schauer, 2009).

3.6.2 Production data

3.6.2.1 Non-interactive data

Discourse Completion Task

The discourse completion task (DCT) is the most classic production task and the most widely used one in L2 pragmatics (Taguchi and Roever, 2017, p. 83). It is a kind of production questionnaire intended to elicit a response to a prompt. The basic format of a DCT comprises a written prompt describing an imaginary situation and a question asking the respondent what s/he would say in such a situation (see Taguchi and Roever, 2017, p. 83-84). Figure 2 illustrates a DCT item intended to elicit a request:

> You are at work, and writing a report that is due by the end of the day. You don't know how to create a graph for the report from a spreadsheet but your colleague Jane in the next cubicle is very good with spreadsheets. You get along well with Jane and often help each other. You decide to ask Jane to show you how to create the graph. She is at her desk reading a document. You walk up to her and say:_____

Figure 2. Illustrative DCT item aiming to elicit a request (adopted from Taguchi and Roever, 2017)

The DCT has been extensively used in the early and late research of individual speech acts and politeness (e.g. Blum-Kulka et al., 1989; Pinto, 2005). In order to be successful in obtaining the required data, the DCT should be well-designed. Taguchi and Roever (2017) assert that prompts within a DCT need to display sufficient information about the situations in question so that participants can construct a relevant and to-the-point response. The DCT designer should also attend to the kind of participants she wants to test and creates plausible scenarios that are close to the participants' real life. This is to ensure that the participants can provide realistic responses by imagining themselves in situations they are familiar with (see Tran, 2013; Taguchi and Roever, 2017). For example, if the participants were university students, the DCT scenarios should be from or close to their academic environment, not something they have not experienced before (e.g. being a custom officer).

In addition to the basic and most commonly-used format of DCT mentioned above, several variants have been used by researchers. For example, instead of a scenario, some studies used dialogues as prompts (e.g. Bardovi-Harlig & Hartford, 1993). Other studies used 'rejoinder' (=utterance following the blank) in their prompts (e.g. Rose, 1992). Multi-turn DCT used conversation prompts with several gaps within each conversation for the respondents to fill in (e.g. Cohen and Shively, 2002). Another extreme variant is interactive-like in nature. It requires respondents to write a full conversation in response to a prompt (e.g. Martinez-Flor, 2013). Li and Taguchi (2014) used an oral DCT whereby they asked their respondents to provide spoken responses to the prompts (see Taguchi and Roever, 2017).

Finally, Halenko (2016) made use of modern technology and used an innovative variant of oral DCT which she called *computer-animated production task* (CAPT). In this task, she designed some semi-interactive single-turn scenarios which contained characters animated by a computer programme. The scenarios were designed as Power Point slides. Each scenario begins with an initial slide briefing the participant with the situation. Then, after a short interval, the animated character starts talking automatically. S/he utters a sentence to which the participant should respond orally (see Halenko, 2016, Appendix 1). The author had recourse to an online movie-making site[4] to design the task. Halenko's instrument is a creative and interesting exploitation of technology in the field of L2 pragmatics. It is highly expected that it will stimulate more future studies of the kind and receive more advancement.

DCTs are characterized with great practicality "they can elicit a large amount of data under controlled conditions from a large number of respondents in a short period of time" (Taguchi and Roever, 2017, p. 85). They are also typically easier to design and administer than some other production-data instruments. However, DCTs also have some serious limitations. They elicit only "offline knowledge…In other words, they show participants' knowledge of the strategies available to produce a speech act, but they do not show what participants would actually say in real-world interaction" (Taguchi and Roever, 2017, p. 85). This is due, among other things, to the imaginary nature of prompts and the lack of interactive real-life sequences in this instrument (Taguchi and Roever, 2017). But, this problem can be partly overcome by using more sophisticated DCTs such as Halenko's one. Furthermore, Ishihara and Cohen (2010) state that several studies which used DCT have questioned the validity of the data elicited by this instrument and recommended using it with much caution (e.g. see Beebe and Cummings, 1996, Hartford and Bardovi-Harlig,1992, both cited in Ishihara and Cohen, 2010). They also argue that ''DCTs may be a valid source of pragmalinguistic (language-focused) data, but might not be reflective of the sociopragmatic (culture-focused) aspects of the speech act'' (p. 43).

---

[4] www.nawmal.com

3.6.2.2 Interactive data

Role-play

Role-play is another production-data instrument used in L2 pragmatics which is similar to DCT in having controlled and predetermined prompts and different from it in being interactive in nature. "Role plays are simulations of communicative encounters… that elicit spoken data in which two interlocutors assume roles under predefined experimental conditions" (Felix-Brasdefer, 2010, p. 47).They are often considered "a good compromise between a structured, standardized data collection procedure and natural data" (Taguchi and Roever, 2017, p. 88). Role-plays are of two types: closed and open. In both types, participants are often provided with a description for a prompt situation and asked to respond as naturally as possible as if the situation were a real-life one. In closed role-plays, participants are required to respond to a prompt role-play situation without a further reply from another interlocutor. Taguchi and Roever (2017) consider closed role-play the same as oral DCT. Open role-play, on the other hand, specifies beforehand the role for each participant to play in the role-play interaction. Then, each participant is asked to impersonate the role and interact face-to-face with somebody else who is either another participant or a trained interlocutor. The course of the encounter and its outcome are left to the interlocutors' discretion though without control from the researcher(s) (see Felix-Brasdefer, 2010 and Taguchi and Roever, 2017).

Role-plays have the advantage of eliciting interactive data which allow researchers to draw conclusions about L2 learners' prolonged use of the target language rather than the use of single-shot utterances. Another advantage is the partial control the researcher has over the prompt situations and the required roles. This enables him/her to direct the role-play interaction to the pragmatic act/phenomenon under research and obtain the required data. Such a feature is not afforded by other interactive methods such as the natural or elicited conversations. However, despite being interactive, role-plays cannot capture all the features available in naturally-occurring encounters due to dealing with simulations only. Another issue with especially open role-play is that because of the lack of control on the content of conversations per se (rather than their prompts), "no two role-play interactions, even if based on the same prompts, are likely to be identical. This raises the thorny issue of standardization vs. authenticity" (Taguchi and Roever, 2017, p. 89). The researcher needs to make a trade-off

between either obtaining authentic data from participants at the expense of standardization or vice versa. As a way out of this dilemma, Youn (2015) suggests providing more instructions for both participants and trained interlocutors specifying what speech acts to be produced at the beginning, middle and end of each conversation. By so doing, Youn could attain more comparable data with a higher level of standardization.

Recording authentic conversation

Recording what people say is another way of capturing the production of interactive data. Unlike DCT and role-play, recording speech in real-life situations yields naturally-occurring data whose results would be more accurate and more reflective to the reality (see Ishihara and Cohen, 2010). Throughout the history of (L2) pragmatics studies, different methods have been used for recording naturally-occurring data which can be summarized as follows:

Field notes is a standard old data-collection technique which involves the researcher taking notes of what people say and do while happening. Kasper (2008) states that this technique was originated in ethnography and was borrowed and adopted in some of the earlier pragmatics studies (e.g. Manes and Wolfson, 1981; Wolfson, 1983; Holmes, 1988, all cited in Kasper, 2008). The data collected via field notes are ''impressively large'' and can establish a useful basis for the realization of the construct under research (see Kasper, 2008, p. 284). However, this technique suffers from a number of downsides. First, the researcher needs to obtain permission from the subjects to take notes. This may affect the naturalness of their behaviour and make them more self-conscious to what they say. Second, taking notes relies mainly on the short-term memory of the observer which is apt to decaying rapidly. Third, the method is uncontrolled in nature which cannot be manipulated easily, very effort- and time-consuming, and having elusive data as the phenomenon under research may not be captured with ease. Finally, Field notes is ineffective in observing the prosody of speech and the body language (e.g. gestures and facial expressions) which people employ while talking (see Kasper, 2008 and Ishihara and Cohen, 2010).

Modern technology helped researchers in overcoming many of the problems of the old techniques. With *audio-recording*, researchers became able to record people's speech

permanently and capture much of its prosody. This advancement made the data obtained more controllable (i.e., almost nothing of the recorded speech can be lost) and the results more accurate and fruitful. With *video-recording*, researchers became far more able to capture the body language of the recorded interaction in addition to speech. Kasper (2008) asserts that ''Visual data afford access to the physical setting, which in turn enables inferences regarding the social, cultural and institutional organization of the setting, personal and social relationships, and participant attributes'' (p. 286). Audio- and video-recording shares with field note the first and third disadvantages mentioned above. In addition, audio/video-recording might be more expensive in terms of affording the recording devices and their accessories (see Kasper, 2008 and Ishihara and Cohen, 2010).

With the end of the discussion of data collection methods, we come to the end of the literature review. The current study will use a judgment task instrument for collecting the required data. This is because: (1) judgment tasks have been widely used in previous L2 comprehension and L2 irony studies like the current one (e.g. Togame, 2016), and (2) other popular L2 pragmatics methods (DCT, Role-play and recording techniques) are mainly production rather than comprehension data-collection instruments.

3.7 Research Gaps: A summary

The reviewed literature on online sarcasm and L2 pragmatics in this part of the thesis has a number of research gaps (listed below), some of which the current study is attempting to fill in. The research gaps listed here are used in chapter one as a rationale for conducting the current study.

1. As mentioned earlier (3.5), the number of L2 pragmatics studies dealing with irony and sarcasm is very small vis-à-vis the body of studies handling irony and sarcasm in general (Shively et al., 2008). This study investigates the comprehension of L2 sarcasm by Iraqi EFL learners and adds to the current literature.

2. Within L2 pragmatics, there are more studies on production than on comprehension (see the reviewed literature above). As the current study deals with sarcasm comprehension, it will increase the literature on comprehension and contribute to bridging the gap between the two kinds of studies. Moreover, most of the L2 pragmatics studies investigate speech

acts. By looking at sarcasm, the current study attracts attention towards investigating L2 pragmatic features and phenomena other than speech acts.

3. No study was found tackling online English sarcasm within L2 pragmatics. The studies on online sarcasm reviewed in (2.8) above are all dealing with it from L1 perspective. The current study covers this shortage by investigating the L2 learners' recognition of online English sarcasm. Furthermore, most of the L1 online sarcasm studies used data collected from Twitter and Amazon websites. No study has analysed data from special online forums as the current study does (data collected from football and parenting forums).

4. All the L2 irony and sarcasm studies reviewed above used constructed data (Bouton, 1999 and Togame, 2016) or quasi-natural data (Shively et al., 2008 and Kim, 2014). None of them have used naturally-occurring data. The current study addresses this shortage by employing naturally-occurring data which represents real life better.

5. There is a real scarcity of studies looking at Arab EFL/ESL learners within L2 pragmatics studies in general and L2 irony/sarcasm studies in particular. The shortage is more serious regarding Iraqi EFL/ESL learners within Arab learners of English. The current study sheds light on the recognition of British-English sarcasm by Iraqi EFL learners. By so doing, the study attracts attention more towards conducting studies on this category of EFL learners (Arabs including Iraqis).

3.8 Research Questions (revisited) and Hypotheses

In what follows is a reminder of the research questions of the current study along with hypotheses based on the literature reviewed in this chapter and the previous one.

1. **Can Iraqi L2 learners of English recognize written sarcasm in British English?**
'Yes, they can' is the hypothetical answer of this question. All the L2 irony/sarcasm studies reviewed in this chapter indicate that L2 learners can recognize irony/sarcasm in the target language (see 3.5).

2. **If so, how does Iraqi L2 learners' ability to recognize written sarcasm compare to that of native speakers of English?**

Some previous L2 studies (e.g. Shively et al., 2008) assert that irony/sarcasm perception in the target language improves with the increase in L2 proficiency and target language experience. However, it is expected, as Bouton (1999) points out, that learners' perception of the L2 irony/sarcasm would not reach the native-level even after spending a long time in the target language community.

3. **What factors influence Iraqi L2 learners' ability to recognize written sarcasm (age, gender, L2 proficiency, study abroad)?**

According to the literature review above, it is expected to see effect for L2 proficiency and study abroad upon the L2 learners' sarcasm recognition (see Schauer, 2009 and Shively et al., 2008). As for age and gender, their effect, if any, will be investigated in this study. But no expectations are given due to the scarcity of L2 studies focusing on these two factors.

4. **What are the general pragmatic and pragmalinguistic characteristics of sarcasm that English native speakers and Iraqi L2 learners of English draw on in the process of recognizing written sarcasm? Which characteristics are more prototypical and which are less?**

It is expected to find almost all the pragmatic and pragmalinguistic characteristics mentioned in (2.5) and (2.6) in the data of this study. It is also expected that, within pragmatic characteristics of sarcasm, 'Negative attitude' and 'Victim' to be central prototypical characteristics because several definitions and accounts of irony/sarcasm have included them as key properties (see 2.5, sections of 'Negative attitude' and 'Victim'). 'Insincerity' is also a candidate to be a prototypical characteristic as several scholars have highlighted its importance for irony/sarcasm creation and recognition (see Grice, 1975, 1989; Leech, 1983, 2014; Culpeper, 1996, 2005). Regarding pragmalinguistic characteristics, the expected candidates to be prototypical features of sarcasm are 'Hyperbole' and 'Graphological cues' according to the reviewed literature (see 2.6).

3.9 Summary

    This chapter has reviewed the literature of several L2 pragmatics aspects which relate to the current study. In the beginning, the chapter provided an overview of this field which mainly showed what L2 pragmatics is concerned with and the major topics investigated within its domain. Then, the chapter reviewed two parameters used in L2 pragmatics which have direct connection to the work of this thesis, i.e., L2 proficiency effect and study abroad effect. Some studies which dealt with these parameters were reviewed to see what impact those parameters had on the L2 constructs they investigated. The chapter also reviewed the previous studies on L2 irony and sarcasm to see what aspects were covered and specify the gaps that need to be filled. Afterwards a review was done for the popular instruments of data collection used in L2 pragmatics. This was to decide which instrument is adequate for the purpose of this study. Next to that, the chapter lists some research gaps found while doing the literature review which relate to the scarcity of L2 pragmatics studies dealing with: (1) the recognition of online L2 sarcasm, (2) naturally-occurring data, and (3) Arab EFL learners. The current study attempt and contributes to filling these gaps. Finally, the chapter rehearses the research questions of this study and provides some predicative answers for them according to the literature review. With this final step, we come to the end of the literature review. In the next chapter, we shall start the second part of this thesis which is concerned with conducting the first study in this thesis, i.e., the corpus study.

PART 2

Corpus Study

Chapter Four

Investigating Online Sarcasm

4.1 Introduction

This chapter and the next one consider the first study done in this thesis, i.e., the corpus-based study of sarcasm. As for this chapter, it consists of two main components: (1) Methods and (2) Results and Discussion. The methods component explains the process of data collection and how the data was analysed. At the beginning, it mentions which online sources were used to collect sarcasm data from, and the strategy adopted for locating and extracting that data. Then, it describes the process of analysing the data, which was mainly based on the general pragmatic and pragmalinguistic characteristics of sarcasm reviewed in the chapter two. The results and discussion component first presents the results of the general pragmatic and pragmalinguistic analyses in tables, and then comments on the implications of these results with regard to the general pragmatic and pragmalinguistic characteristics of sarcasm. The chapter will end with a short summary of the conclusions.

4.2 Data Collection from Online Forums

The current study involved obtaining naturally-occurring instances of sarcasm from online British English forums. As I mentioned in 1.1, there is a shortage of irony and sarcasm studies utilizing naturally-occurring data. Such data would yield more realistic findings that could reflect 'reality' better; in other words, they improve ecological validity. That is why it was decided to employ naturally-occurring data in this research. This section states how the real-life sarcasm corpus was collected from the internet.

A careful search was conducted in many of the online forums in order to pinpoint the ones that were rich in using sarcasm. I preferred to start searching within the sport domain for sarcasm data. Sport is a big and ramified field that involves lots of people, especially fans. If the search proved not to be fruitless, I would probe some other domains for sufficient data. Within sport, I decided to search for the required data in the most popular sports first. If the search was not successful, I move down to the less popular sports. Popular sports entail more

fans involved. This, in turn, involves a more chance for obtaining sufficient data. The search was done first within football, which is a highly popular sport in the UK. The search was applied to the websites of some well-known British football clubs (namely, Arsenal[5], Chelsea[6], Liverpool[7], Manchester City[8], and Manchester United). Except for Manchester United, all the other clubs have outside social media pages on Facebook and Twitter for their fans to chat in. The problem with Facebook and Twitter is that these sites do not have internal search engine for looking for data. On the other hand, Manchester United official website has a built-in fans' forum[9] with a search engine of its own. The forum is very sophisticated and contains hundreds of fans continually interacting with each other.

The search within the Manchester United forum resulted in a good amount of sarcasm data. But football forums are assumed to be mainly populated by men. Most of the pseudonyms used in Manchester United forum, for example, are masculine. In addition, there is an excessive use of the pronouns *he* and *him* in the users' references to each other. Therefore, some other forums mainly populated by women were also investigated in order to avoid gender bias and make the data collected as diverse as possible. In the beginning, the preference was to remain within the sport area to control the variable of the search area (sport). Thus, another search was done within the online forums of showjumping sport, which is mainly a female sport in British culture. Some evidence of the users' gender of such forums came from female pseudonyms, female pictures posted, and the overuse of *she* and *her* in chatting when referring to each other. The search aimed at obtaining an amount of sarcasm comparable to that extracted from the football forum. Unfortunately, these forums were not found to be rich enough in sarcasm. Hence, there was a need for finding a practical substitute.

Mumsnet[10] and *Netmums*[11] are two well-known British websites dealing with parenting issues. They are big sites populated by hundreds of thousands of mainly female users and contain internal chat forums. Mumsnet was founded by a sport journalist called Justine Roberts and some other co-founders in Jan 2000. Since inception, the site has hosted

---

[5] https://www.arsenal.com/
[6] https://www.chelseafc.com/
[7] http://www.liverpoolfc.com/welcome-to-liverpool-fc
[8] https://www.mancity.com/
[9] http://community.manutd.com/forums/
[10] https://www.mumsnet.com/Talk
[11] https://www.netmums.com/

webchats on different parenting issues. The site has grown rapidly and now it includes over 1.25 million registered users. The former British Prime Minister described Mumsnet as a "national institution"[12] and its founder Justine Roberts was ranked the seventh in the BBC's list of the most powerful women in the UK in 2013[13]. The other site Netmums was founded by Siobhan Freegard and other co-founders in 2000. It has also hosted online chats on parenting issues and has now over one million registered members. In 2009, the site was chosen by the *Independent* newspaper as one of the top 50 websites and blogs for parents.[14] In 2014, the Freegard and other founders of Netmums received the 'Most Excellent Order of the British Empire' reward from the Queen for their services to families.[15]

   The two sites appeared to be good sources for female dataset. Thus, a search was conducted in these sites in pursuit of sarcasm. First, the search was within the Mumsnet site and some sarcastic data was obtained. However, the amount of sarcasm was not comparable to the one obtained from the football forum. Thus, there was a need for doing a further search within the Netmums site for more data. The search was fruitful and managed to collect a good amount of extra data. The total dataset collected from the parenting sites (Mumsnet and Netmums) was eventually comparable in size to the football dataset and achieved counterbalance with it.

   The data collection procedure was as follows. The relevant websites were navigated in search for the terms *sarcasm* and *sarcastic* by the internal search engines. In addition, Google search engine was also used to do the same search, but within the wanted websites by using a special command (e.g. sarcastic or sarcasm site:mumsnet.com/talk). In fact, Google search proved to be more efficient and was used more as it could yield a great deal of results not caught by the forums' search engines. As a result, a list of links was displayed on the screen each of which led to a thread in which *sarcasm* or *sarcastic* occurred. Links were followed one by one, and threads were read thoroughly as they were potential environments for sarcasm. Any thread was considered as containing sarcasm if it included a remark that was metalinguistically judged as being 'sarcastic' by a later remark. The metalanguage could be

---

[12] http://www.bbc.co.uk/news/uk-12238447
[13] http://www.bbc.co.uk/programmes/articles/3J92brPmK0hskzhpTV3CrZ0/the-power-list-2013
[14] https://en.wikipedia.org/wiki/Netmums
[15] https://en.wikipedia.org/wiki/Netmums

made by the speaker him/herself (e.g. *I was being sarcastic*) or by somebody else within the thread (e.g. *Don't be sarcastic*). Indicating sarcasm by means of a metalinguistic judgment is termed as "metalanguage strategy". For example, consider the following thread from Manchester United football forum:

(14)

A:

Did u see Man city match yesterday? *I blame Berbatov[player name] for [Man]city defeat ..Well he was so lazy on the pitch he missed 3 clear chances one to one with a goalkeeper.he was so useless can't be bothered to run around like headless chicken.it all Berbatov fault!* (sarcasm italicized)

B:

You are being sarcastic. He didn't play! (metalanguage underlined)

In example (14), B's turn comes as a reaction to A's mention of "*Berbatov*" describing that mention as "being sarcastic". The first utterance of B's turn provides a metalinguistic judgment for the sarcasm in A's turn. All such encountered cases were extracted from the forums as possible sarcasm data.

If sarcasm was identified in any thread by the metalanguage strategy, the whole thread was extracted from the forum as it provided a context for comprehending the sarcastic turn (or turns) occurring in it. However, if the thread was long, a sufficient number of turns was extracted before and/or after the one in question to provide a proper context for grasping the sarcastic point. Sometimes, a thread may contain a sarcastic remark, but the thread is excessively long and may extend to several pages. Thus, extracting some turns around the sarcasm may not provide a sufficient context. Such a thread was dropped from consideration

for reasons of practicality, i.e., it would not make an examinable stimulus for the study's experiments discussed in Part 2 and Part 3.

In fact, not all the threads in the search results proved to include sarcasm. The occurrence of the terms *sarcasm* and *sarcastic* was, sometimes, for purposes other than indicating sarcasm such as denying sarcasm (e.g. *I'm not being sarcastic here*) and describing friends and family members as being sarcastic in their everyday life (e.g. *My husband is very sarcastic*). That was a limitation in the metalanguage strategy used for pinpointing sarcasm in this study. Another limitation was that, sometimes, the metalinguistic remark did not come immediately after the sarcastic remark, but long after it. In several cases, this resulted in elongating the thread extensively and making it impractical for being a stimulus in the study's experiments. However, apart from these two issues, the metalanguage strategy ultimately proved to be successful in collecting the required data of sarcasm from the online sources. By means of this strategy, a total of 142 threads were collected from forums that had potential sarcasm — 70 from the football forum and 72 from the parenting forums. Some threads turned out to have more than one instance of sarcasm. Hence, the total of sarcasm instances was 149, 73 within football dataset and 76 within the parenting dataset.

4.3 Data Analysis

The total of the collected sarcasm instances was analysed in this phase of the thesis to see how sarcasm was used and with what characteristics. The analysis was made in the light of the general pragmatic and pragmalinguistic characteristics reviewed in the literature (see 2.5 and 2.6). Regarding the general pragmatic characteristics, the data was investigated to see which of the following general pragmatic characteristics of sarcasm were used and how often: (1) Contradiction, (2) Insincerity, (3) Flouting Quantity (4) Flouting Relevance, (5) Mock politeness, (6) Allusion to antecedent, (7) Negative attitude, and (8) Victim. In addition, the data was also investigated for any other characteristics that were not covered in the literature. A similar procedure was also adopted with the pragmalinguistic characteristics of sarcasm. The data was investigated to find out which and how often the following characteristics were used: (1) Positive wording, (2) Hyperbole, and (3) Graphological cues (Capitalization, Emoticon, Laughing marker, Exclamation mark). The data was also inspected for any other

pragmalinguistic characteristics that were not or hardly ever discussed in the literature. Both football and parenting datasets were separately investigated for both kinds of characteristics in order to highlight any gender differences. The next section presents the results of the analysis conducted.

4.4 Results and Discussion

4.4.1 General Pragmatic Characteristics

Results of data analysis reveal differences in the frequencies of occurrence of general pragmatic characteristics. In addition, they also manifest discrepancies between the football dataset and the parenting dataset in employing those characteristics in the online sarcasm they used. Table 2 displays the distribution of general pragmatic characteristics within the collected sarcasm data and reveals these differences and discrepancies.

Table 2

Sarcasm within the Collected Excerpts from Online Forums: Frequencies of General Pragmatic Characteristics

Forum	General Pragmatic Characteristics									
	Allusion to Antecedent	Contradiction	Insincerity (Flouting Quality)	Flouting Quantity	Flouting Relevance	Mock Politeness	Negative Attitude	Victim	Other	
									Characteristic	No.
Football	38	41	67	2	2	25	34	52	Elaboration	3
									Mocking a previous utterance	2
									Simile	2
									Metaphor	1
Parenting	51	30	52	0	0	15	42	55	Elaboration	1
									Metaphor	2
									Sincerity	13
Total	89	71	119	2	2	40	76	107		24

Note. 'Other' category comprises Characteristics encountered in the data which are never or hardly ever talked about in the literature.

Table 2 reveals that 'Insincerity' comes top in the list of characteristics frequencies with 119 times of use in total (67 within football dataset and 52 within parenting dataset). It seems that 'Insincerity' has the greatest weight in creating sarcasm. In fact, some samples have 'Insincerity' as the only characteristic used. This strengthens the hypothesis that 'Insincerity' is so basic for creating sarcasm in English. However, it is not claimed here that 'Insincerity' is a necessary condition of sarcasm. Rather, It could be a central prototypical characteristic, and, perhaps, the most central one according to the results of this study. The excessive appearance and importance of 'Insincerity' in my data is a finding favours Grice's account of irony/sarcasm. It also runs contrary to that of Campbell & Katz (2012) (i.e., insincerity is

relatively weak in creating sarcasm) and that of Colston (2000) (i.e., insincerity is less important for doing sarcasm). Unlike these two studies, one advantage of the current study is that it involves naturally–occurring data. I have not come across any previous study that considers 'Insincerity' as the most important and frequent characteristic of sarcasm in English.

The existence of a victim also proves to be an important characteristic of sarcasm. According to the results shown in the Table 2 (107 in total: 52 within football dataset and 55 within parenting dataset), it comes second in the list after 'Insincerity'. This high score qualifies 'Victim' to be another highly prototypical feature of sarcasm. These results support any account that involves victim existence in its very definition of sarcasm (e.g. Attardo, 2000 and Toplak and Katz, 2000).

Regarding the other characteristics, 'Allusion to an antecedent' also scored high in the results (89 in total: 38 within football dataset and 51within parenting dataset). These results favour Sperber and Wilson's (2012) account of irony. Parenting dataset, for some reason, seem to apply this feature somewhat more than football dataset when using sarcasm. 'Negative attitude' scored 76 in total (34 within football dataset and 42 within parenting dataset), which is rather high as well, and this supports all scholars' works which include 'Negative attitude' in their irony/sarcasm definitions (e.g. Culpeper, 1996; Colston, 1997; Toplak and Katz, 2000; and Kim, 2014). Again, parenting dataset also use this feature somewhat more than football dataset for some reason. 'Contradiction', which is the traditional Aristotelian characteristic associated with sarcasm, was found in nearly half of the cases (71 in total out of 149 samples— 41 within football dataset and 30 within parenting dataset). This time, it is football dataset which seems to use this feature somewhat more frequently than the parenting dataset. Results qualify these three features to be candidates for being prototypical features of first-order sarcasm since they were used in the majority of cases (i.e., around or more than half of the cases).

Regarding 'Mock politeness', it scores 40 only in the total of 149 sarcastic instances. This low result cannot make mock politeness eligible for being a central prototypical feature of sarcasm. However, one should not deny the possibility of mock politeness for being such. Perhaps, by investigating other kinds of data (e.g. sarcasm in oral conversations), different results may appear. But with the case of the current data investigated (i.e., written sarcasm on

the internet), evidence does not support the eligibility of mock politeness as a central feature of written online sarcasm. As for the disparity between the football dataset and parenting dataset in using mock politeness for sarcasm, results show a rather big difference in the number of times it was used by both datasets (25 within football dataset and 15 within parenting dataset). It is not clear why this difference is, but one might speculate that mock politeness may enhance the offensive power of sarcasm, and it was noticed in the current investigated data that the assumed males (football dataset) were more aggressive in their sarcasm than assumed females (parenting dataset).

All the instances of mock politeness were of the type of polite language that remains on the surface level. Only one exceptional single case was in the parenting dataset that belongs to Culpeper's (2011) "verbal formula mismatches" (polite language+ impolite language in the same remark). Again, it is not clear why this is the case.

(15)

A:

Our roof needs a few new slates and as a busy Mother of three I thought I'd ask if anyone could recommend a local roofer…The admin posted up a reply very quickly saying that I should only post about parent related stuff. WTF? AIBU to think that the power has gone to her head a little?!

B:

But don't stoop to sarcasm. Just say **"**oops, my mistake, I will take my post elsewhere. <u>Thanks you cunt!"</u>.

(sarcasm underlined)

In (15) above, verbal formula mismatch occurs in "Thanks you cunt". In this expression, the highly polite word "Thanks" is forced to collocate with "you cunt" which is a very offensive taboo phrase.

Example (16) below is one of the best illustrative examples in the analysed data that includes most of the general pragmatic characteristics, namely, Allusion to antecedent, Insincerity, Mock politeness, Negative attitude and Victim.

(16)

(Oshea [player name] Out For Two Months)

Gabranth:

I'm not happy that he's injured, I'm just happy for all of us that he won't be playing.

RedYankee:

For all those who think Gabranth is a United fan

He is happy Oshea[player name] is hurt. What a great United fan he is, huh?

(Sarcasm underlined)

In 16, the ironist when saying "What a great United fan he is, huh?", he alludes to Gabranth's previous remark (antecedent). He is also being insincere in saying this exclamatory sentence in which he describes Gabranth as being a "great United fan". In addition, exclamatory sentence typically shows admiration towards somebody/something and this implies politeness towards the addressee. Using this exclamatory sentence sarcastically is an obvious example of mock politeness. Finally, the sarcasm in this example is used to express a negative attitude against Gabranth who is happy that Oshea (Man United player) will not play again and it is clear that Gabranth is the victim of this sarcasm.

'Flouting Quantity' occurred twice in the whole total as well as 'Flouting Relevance', and in football instances exclusively. This very low score may indicate how rare the flouting of these two CP maxims could be in the performance of sarcasm online.

The last column in the table presents the 'Other' characteristics of sarcasm. These were some extra characteristics encountered while analyzing sarcasm in the data. To my

knowledge, only metaphor was touched upon in the literature (see Kim 2014), whereas the rest do not seem to have been investigated. What attracted my attention in particular was the 13 times 'sincerity' was used in comparison to the other characteristics in the 'Other' category and within parenting dataset only. I believe areas such as this may benefit from future in-depth qualitative research exploring the complicated nature of sarcasm which seems to draw on features other than the well-known ones enclosed in the literature. In what follows are some illustrative examples of those 'Other' characteristics.

(17) Elaboration (from football data)

A:

Would you swap Nani [player name] for Bale [player name]?

B:

i would swap Obertan [player name] or Bebe [player name] for Bale but not Nani. i have always been a big fan of bale however i do think Nani is better so i would choose him over Bale.

C:

Yes in a heartbeat, I'd also throw Giggs [player name], Rooney [player name], Scholes [player name], Evra [player name], Vidic [player name], and Anderson [player name] in as part of the deal.

As you can guess I'm being sarcastic.

NO WAY !!!!!!!!!!!!!!!!!

(Sarcasm underlined)

(18) Metaphor (from parenting data)

A:

.... Maybe I have "dodged a bullet" here...I have felt for a while that there is a slight resentment towards my son for various reasons and the last thing I want is him to grow up

feeling like a second class citizen in our house.....I told my partner a while ago that my son will always come first, I will always choose him if forced to make a choice...and what mother wouldn't do that? These words obviously stuck in my partner's head as he said to me, in one of his texts, that maybe my "little prince will be happy now he has me all to himself" (in the most sarcastic tone)

(Sarcasm underlined)

(19) Mocking a previous utterance (from football data)

A:

If we could manage to get Robinho [player name] for between £5-£8 million, or Silva [player name]for around £10 million, that would be very interesting. Anything more, I'd say no ways jose.

B:

hahahahahahahahaha....................Silva there NEW signing being instantly sold for £10Millions.................jeez some fans are such IDIOTS

yes, I'm being sarcastic

(Sarcasm underlined)

 (20) Simile (from football data)

A:

I recommend Rooney [player name] on the bench against Fulham!

B:

I recommend you get some sense, and fast. Dropping Rooney would be incredibly stupid as he NEEDS games.

Wouldn't be surprised if it was just another person wanting Berbatov [player name]to be in the starting line up, at whatever cost.

C:

Yeah, that's right. It's just another Berbatov's fan, <u>they're like plague</u> (sarcastic).

(Sarcasm underlined)

(21) Sincerity (from parenting data)

A:

My 8 year old is starting to develop an attitude!! Feel like im at my wits end with her. Everything i say she has some sarcastic comment to reply to, for example if i ask what she would like for dinner the reply is usually, "<u>you r the mom!</u>" I love her so much but im finding it hard to enjoy her at the moment.

(Sarcasm underlined)

4.4.2 Pragmalinguistic characteristics

All the pragmalinguistic characteristics of sarcasm discussed in the literature review (positive wording, hyperbole and graphological cues, see 2.6) were found within the sarcasm of the collected data. Table 3 shows the frequencies of those characteristics within the total of the collected data. Additionally, it also shows their frequencies within the football and parenting datasets.

Table 3

Sarcasm within the Collected Excerpts from Online Forums: Frequencies of Pragmalinguistic Characteristics

Forum	Pragmalinguistic Characteristics							
	Positive Wording	Hyperbole	Graphological Cues				Other	
			Capitalization	Emoticon	Laughing Marker	Exclamation Mark	Characteristic	No.
Football	25	31	17	1	5	15	Interjection and exclamation	10
							Attention getter	1
							Structure Repetition	2
							Vowel elongation	2
Parenting	15	13	3	21	1	23	Interjection and exclamation	14
							Italics	2
							Quotation marks (for stress)	1
							Structure repetition	1
							Vowel elongation	1
							Boldface	1
Total	40	44	20	22	6	38		35

Note. 'Other' category comprises features encountered in the data which are never or hardly ever talked about in the literature.

'Hyperbole' had the highest score among the pragmalinguistic characteristics with 44 times in total. But there was a noticeable discrepancy between football and parenting datasets in how often they applied this feature in their sarcasm (31 within the football dataset and 13 within the parenting dataset). It could be the case that hyperbole may result in enhancing the force of sarcasm used and that the assumed British males prefer to use exaggerated forceful sarcasm

more than the assumed females. To illustrate how 'Hyperbole' works in sarcasm, consider the following example from the football dataset:

(22)

bruk misghina:

hello united. This is the time to build new united.let our renascence begin 86hichi86w. We need THIAGO SILVA,specially the build up of our MF I think the right man for scholsy is BASTIAN more than WESLY,let our formation be like this.

NEUER,RAFAEL,THIAGO SILVA(MERTESACAR),VIDA,ASAWA KOUTTO,RIBERY,BASTIAN,KAKA'(DE ROSSI),DAVID SILVA,ROO & THE BOY WONDER 86hichi! OUR NUMBER 7 MUST BE GIVEN TO EITHER KAKA OR RIBERY

lorik:

Wow I made a post yesterday saying people should be allowed to post what they want. But your on some next *** there. You been smoking rocks or something?

mazhar08:

(Sarcastic tone)

Please, support him. He is a genius. He knows more than everyone, including God. He must be a FOOTBALL GOD who everyone must worship. He will lead us to glory for all time. LOL!!! 😬

(Sarcasm underlined)

In this example, '**mazhar08**' describes the first poster (**bruk misghina**) as ''He knows more than everyone, including God'' and as ''a FOOTBALL GOD who everyone must worship''. These descriptions are obviously hyperbolic in nature and are used to enhance the force of the sarcasm expressed.

'Positive wording' achieved the second highest score among these characteristics with a total of 40 (25 within the football dataset and 15 within the parenting dataset). Positive wording is

the pragmalinguistic tool by which mock politeness is realised. That is why its score coincides with that of mock politeness mentioned above (see Table 2). Example (16) above is an illustrative case of 'Positive wording'.

Regarding 'Graphological cues' (Capitalization, Emoticon, Laughing marker and Exclamation mark), they vary in the total scores they achieved between the highest 'Exclamation Mark' (38 in total) and the lowest 'Laughing Marker' (6 in total). Example (22) above also illustrates how 'Graphological cues' are used to convey sarcasm. As for the 'Other' category, it comprises some pragmalinguistic features found in the data about which no or little literature has, to my knowledge, been written. Within this category, interjections (e.g. *Oh, Ah, Wow*) were used much more than the rest. This finding comes in support of Kovaz's et al. (2013) note that interjections highly occur with sarcasm and can function as an indicator to it as well. Regarding 'structure repetition' and 'vowel elongation' within this category, nothing was found in the literature about the former, whereas the latter was incidentally touched upon in Shively et al. (2008) and Kim (2014). How to classify these two features was also not stated in the literature. However, I class them within the framework of hyperbole as they were used in the data to give a sense of exaggeration. Examples (23) and (24) illustrate how 'structure repetition' and 'vowel elongation' were used in the data to express sarcasm respectively.

(23)

Gary Mcleod

If we could manage to get Robinho for between £5-£8 million, or Silva for around £10 million, that would be very interesting. Anything more, I'd say no ways jose.

lewis.No.9

hahahahahahahahaha....................Silva there NEW signing being instantly sold for £10Mill.................jeez some fans are such IDIOTS

(Sarcasm underlined)

In this example, the second poster (**lewis.No.9**) when saying ''Silva there NEW signing being instantly sold for £10Mill'', he almost repeats the same structure used by the first poster (**Gary Mcleod**) ''Silva for around £10 million''. '**lewis.No.9**' used this structure repetition to mock the other poster's remark sarcastically. '**lewis.No.9**' also used the interjection ''jeez'' and capitalization to emphasize his expression of sarcasm against the other forum user.

(24)

garp01us:

once I get home I'll check my dvr and then give you the exact time in the match it happened because it did happen. Eitherway besides the one goal that was all down to Berba, Hernandez does nothing positive for us when he plays simple as that, like it or not it is the truth!

chicles:

<u>Finally a voice of reason. You're sooooo right!!! Scoring 4 goals in 4 straight games isn't positive enough for ManUntd, lets sell the boy now!!!</u>

andrea71:

i think he was being sarcastic lol

(Sarcasm underlined)

In this example, the poster '**chicles**' elongated the vowel in *so* when saying ''You're sooooo right!!!'' in order to indicate and intensify the sarcasm he is directing against the first poster ''**garp01us**''.

4.5 Conclusion

This chapter, firstly, elucidated the process of collecting naturally-occurring sarcasm from online sources which are rich in it (football and parenting online forums). The collection was by means of the metalanguage strategy which proved to be efficient for this purpose. Secondly, the

collected data was analysed to find out which general pragmatic and pragmalinguistic characteristics of sarcasm were there.

Results revealed that 'Insincerity' is the candidate to be the most central and prototypical general pragmatic characteristic of sarcasm due to the highest frequency it achieved in the data (119 out of 149). This outcome favours Grice's account of irony/ sarcasm which rests mainly on flouting the Quality Maxim. This finding is also contrary to those of Colson (2000) and Campbell & Katz (2012) who all underestimate 'Insincerity' in performing sarcasm. The second prototypical characteristic which comes next to 'Insincerity' is 'Victim' (107 out of 149).This result supports all the accounts and definitions of irony/sarcasm that involve victim (see 2.5, victim involvement). Less prototypically come 'Allusion to antecedent' (89 out of 149), 'Negative attitude' (76 out of 149) and 'Contradiction' (71 out of 149). These results support all the accounts of irony/sarcasm that include these general pragmatic characteristics (see 2.5). The low score of 'Mock politeness' (40 out of 149) does not qualify it to be a prototypical characteristic of sarcasm. This finding does not consolidate both Leech's (1983, 2014) and Culpeper's (1996, 2005) view of sarcasm (i.e., that sarcasm is purely mock politeness). Similarly, with the very low scores of 'Flouting Quantity' and 'Flouting Relevance' (2 each out of 149), both these characteristics lag far behind the others in being prototypical characteristics of sarcasm.

As for pragmalinguistic characteristics, 'Hyperbole' scored the highest frequency of occurrence (44 out of 149). This high score qualifies 'Hyperbole' to be the most prototypical pragmalinguistic characteristics of sarcasm. This finding partly coincides with the finding of Kovaz et al. (2013) and Partington (2007) who found that hyperbolic expressions are quite prevalent in ironic utterances. The high scores of 'Positive wording' (40 out of 149) and 'Exclamation mark' (38 out of 149) also entitle them to be further prototypical pragmalinguistic characteristics of sarcasm. Results also show several discrepancies between the football dataset and the parenting dataset in applying both kinds of characteristics. This observation may suggest the need for conducting more in-depth studies to uncover the reasons behind such differences.

Finally, it is worth mentioning that general pragmatic characteristics seem to bear most of the burden of creating and conveying sarcasm. Supporting evidence for this claim comes from

the fact that many of the investigated 149 instances of sarcasm were devoid of any pragmalinguistic characteristic (positive wording, hyperbole and graphological cues). Nevertheless, they were successful in conveying sarcasm to the reader by means of only the general pragmatic characteristics they possessed (e.g. contradiction, insincerity, allusion to antecedent, etc.). Consider this illustrative example from the football dataset (sarcasm underlined).

(25)

(Edinson Cavani [player name]to Chelsea[football club] done deal)

A:

Chelsea will be waiting on the Rooney [player name]saga

B:

Rooney [player name]must be their 2nd or 3rd choice then. Since they were in for Falcao [player name]and Cavani [player name]first.

C:

Remember how many players we have been linked with already, not everything is necessarily true. Papers seem to write any old rubbish now to sell, next it'll be we've submitted a £15m bid and a chunky kitkat for Messi[player name]

The sarcasm used in this example is devoid of any pragmalinguistic characteristic discussed above. Nevertheless, the speaker was successful in creating and conveying sarcasm by means of employing general pragmatic characteristics only (e.g. 'Insincerity').

Chapter Five

Preparing Stimuli for the L2 Pragmatics Study

5.1 Introduction

The second purpose of collecting online sarcasm data was to use it as an item pool for the L2 pragmatics study (see part 3). This chapter will outline the processes undertaken to prepare stimuli for that study. It will describe the two filters used in trimming the raw material into stimuli. First, it will explain how the length filter was systematically operated to trim the data by means of setting upper and lower word count limits for the accepted stimuli. Secondly, it will also demonstrate how the data which survived the first filter was further trimmed by a judgment task designed for this purpose. The chapter will cover how the judgment task was designed, piloted for the first and second times, and how it was ultimately conducted. Afterwards, it will spell out the procedure of selecting the final stimuli for the L2 pragmatics experiment depending on the results of the stimuli selection task. The chapter will end with a short summary of what is discussed in it.

5.2 Trimming the Corpus

For matters of practicality, it was not possible to use all the total of 142 threads into the L2 pragmatics study. More than one filter were applied for reducing that total and selecting varied testable stimuli from it for the L2 study. The first filter was length. The length of test items has to be reasonable for the testees. Lengthy items might be boring to read, difficult to control, cognitively stressing and time consuming. On the other hand, very short items may not provide a sufficient and a clear-cut context for grasping the sarcasm in them, especially if there is a need to cut something out of them as part of the experiment. Hence, there was a need to set upper and lower limits for item length. Ranges of word counts were calculated for all the excerpts. Table 4 below shows the word count ranges of the whole corpus as well as the ranges of its divisions (the football and parenting datasets). It also shows the means and standard deviations for all the ranges.

Table 4

Corpus Word Count: Ranges, Means and Standard Deviations

	Word Count Range	Mean	SD
Whole Corpus	19-400	105.31	62.36
Football Dataset	19-291	91.84	51.31
Parenting Dataset	30-400	118.41	69.36

According to Table 4, the defined upper and lower limits of the includable word count were 40-143 words for the football dataset and 49-188 words for the parenting dataset. The procedure of defining the upper limit was by adding one standard deviation to the mean of the dataset, whereas the lower limit was determined by subtracting one standard deviation from the mean. This procedure was applied to both datasets. Using more than one standard deviation would have resulted in including most of the 142 threads again within the L2 tested data and rendered the length filter worthless in that case.

The total of 142 threads was reduced to 71 items only according to the length filter. Although the total was reduced by half, the remaining 71 threads were still too many for an L2 experiment, especially if that experiment was a double-tasked one (see chapter seven). Hence, the 71 threads needed to be filtered more and reduced to a testable number. Furthermore, the sarcasm in these threads also needed to be validated by a number of British English (BrE) native speakers in order to be more reliable for the final L2 experiment. The further filtering and sarcasm validation was the job of the stimuli selection task which is discussed next.

5.3 Stimuli Selection Task (SST)

After reducing the total of excerpts to 71, this total was involved in a judgment task. As was mentioned earlier in 3.6, judgment task is the suitable instrument for this purpose because the experiment is a comprehension task, not a production one. The judgment task is designed to serve the following purposes:

1. Filtering and reducing the 71 excerpts more for the final L2 experiment.

2. Being a first pilot for the L2 experiment as well.

3. Validating the sarcasm in the excerpts by exposing them to a group of English native speakers.

The method of collecting sarcastic samples from online forums (metalanguage strategy) was systematic and left no room, roughly speaking, for the researcher's subjectivity. The samples were judged as being sarcastic by BrE native speakers within the forums. Each sarcasm sample was judged to be so by only one native speaker (either the speaker him/herself or some other forum user). In addition, the forum judgments were validated further by means of the validation task in question (SST) which involved native-speaker informants. The extra validation of the SST also provided a remedy to one limitation of the current study, i.e., the anonymity of the forum users. This anonymity made it difficult to tell which user was an English native speaker and which was not. If a sarcasm sample was judged by a non-native speaker of English on the forum, the judgments of the English native speakers in the validation task would dis/approve the forum-user judgment. The results of this task will render the original forums' judgments more reliable and the selected excerpts will be used in the final L2 experiment with more confidence. In fact, it is the judgments of the SST that will be relied on for the final selection of the L2 experiment stimuli.

The SST involved giving the 71 excerpts to a small group of BrE native speakers. The informants were asked to rate sarcasm in the excerpts on a 7-point Likert scale ranging from *Strongly disagree* to *Strongly agree*. Actually, when rating, informants were doing two tasks simultaneously. First, they were passing judgments whether or not there was sarcasm in each excerpt and by doing so they were in/validating the original forums' judgments. Secondly, if sarcasm was available, they were rating its degree according to three points on the scale (i.e., *Slightly agree, Agree, Strongly agree*). By doing the latter task, BrE NS informants judged the range of strength of the sarcasm used in the excerpts. The items which got high scores on the scale were considered to contain more prototypical (central) sarcasm, whereas those with low scores contain marginal sarcasm. Knowing that would help considerably in designing a solid main L2 experiment containing a wide range of sarcasm from borderline up to very strong. The Likert scale used in the SST was as follows:

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

The Likert scale was the best scaling technique to suit the purpose of this task as it was mainly invented to elicit opinions and judgments. It has been extensively used in studies since it was first introduced in 1932 "due to the fact that the method [Likert scale] is simple, versatile and reliable" (Dörnyei, 2010, p. 27). The original Likert scale was made of five points (*Strongly disagree*, *Disagree*, *Neither agree nor disagree*, *Agree*, *Strongly agree*). However, it was modified in later studies to include six or even seven points (Dörnyei, 2010, p. 28). I preferred to use the 7-point scale in the experiments of this thesis in order to capture more nuances of judgments among participants. As for the middle option *Neither agree nor disagree*, some researchers retain it in the Likert scale they use, whereas other exclude it because it is an indeterminate category. The middle option was retained in this study as it allows for rating the indeterminacy of sarcasm as a pragmatic phenomenon. Pragmatic phenomena can be indeterminate or *ambivalent*. That is to say that the intended force of an utterance can be negotiable (meaning one thing or the other) (see Leech, 1977, p. 99; Thomas, 1995, p. 195).

5.3.1 Piloting the stimuli selection task

Piloting is a crucial step for developing a solid test. It would highlight the weak and strong points of the test so that any necessary adjustments are done before the main test is carried out. Piloting would also give the experimenter an insight into how the main experiment would take place (see Mackey and Gass, 2016, p. 52). This section will detail the process of piloting the stimuli selection task.

5.3.1.1 *Stimuli selection task: First pilot task*

Participants

As a first attempt, the stimuli selection task was piloted with a small group of BrE native speakers. Three people (two males and one female) were recruited for this purpose who were

all British nationals and students at different UK universities. Their ages were 19, 27 and 27, and the UK regions they grew up in were Oxfordshire (2) and East Midlands (1) respectively. Two of them were approached via email and one was approached in person. All of them were asked to do an online task and give feedback about it afterwards.

Before doing the task, participants were provided with an information sheet which briefed them with the aims of the study and the kind of data used in the task (see appendix A). The information sheet also assured the participants about the anonymity of participation, the confidentiality of their data and emphasized the option of withdrawal. After reading the information sheet, they were asked to provide their consent for participation by signing a special form prepared for this purpose. The task, information sheet and consent form were all approved by the Research Ethics Committee/Lancaster University beforehand (see appendices A and B).

Materials

The materials of this pilot were the 71 excerpts of sarcasm data that passed through the length filter (see appendix B).

Instrument

The 71 excerpts were involved in a judgment task containing the 7-point Likert scale mentioned above. The task was designed to be done online via 'Qualtrics Survey Software' (http://www.qualtrics.com/). The judgment task consisted of two parts: Part 1 contained 40 items taken from the football forum of Manchester United club, and Part 2 comprised 31 items taken from Mumsnet and Netmums websites. The material of part 1 contained some special terms, names and acronyms. The meanings of these were explained in red font within square brackets (see Figure 3). Within each part, all the questions were randomized and each participant answered a different version of the task. Question numbering was hidden in order to avoid any numbering confusion caused by randomization. Figure 3 presents a screenshot of what the online SST looked like in the first pilot test.

Figure 3. A Screenshot showing a sample of the online SST on Qualtrics Website: First pilot test

After doing the task online, participants' responses were saved on the Qualtrics website under pseudonyms and converted automatically into PDF files. The PDFs were later downloaded and analysed.

Procedure

A covering message was emailed to the participants briefing them with the experiment and requesting them to participate in it. Attached with it were the information sheet and the consent form. The participants were asked to read the sheet and sign the form before doing the

experiment. Afterwards they were asked to follow the experiment link provided at the end of the covering message and do the task in one session. The pilot test was done online and was self-administered by the participants themselves. The task completion time ranged from 49 to 56 minutes. All the participants were paid 5 GBP for their participation.

Results and Discussion

Regarding the pilot test results, all the participants answered the 71 items without skipping any. The following table shows the distribution of responses to the scale categories along with the means and standard deviations (SD) for all the participants.

Table 5

Distribution of Participants' Responses to Scale Categories: First Pilot Test of the Stimuli Selection Task

	Scale Categories							Mean	SD
	Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree		
Participant1	1	2	2	5	7	37	17	10.14	13.04
Participant2	7	6	5	6	19	12	16	10.14	5.58
Participant3	2	7	5	4	22	17	14	10.14	7.55
Total	10	15	12	15	48	66	47	30.42	22.66

On balance, responses have normal distributions around their means. The table reveals that the participants favour answering within the right side of the scale (*Slightly agree, Agree, Strongly agree*) in which scores ranges from 7 to 37, whereas the scores of the left side (*Strongly disagree, Disagree, Slightly disagree*) ranges between 1 to 7 only. This means that there was a general tendency amongst the participants to respond with agreement to the

sarcasm in the excerpts. Thus, in most cases, the task items were realized as sarcastic as was expected. The middle option "*Neither agree nor disagree*" was used 15 times only out of 213 responses. This indicates the scarcity of indeterminate sarcasm cases in the tested data.

Calculating average scores and average scale categories

In this pilot study, the participants varied more or less sharply in their responses for each sarcasm item in the test. It is only in a couple of cases that the responses were identical. Therefore, there was a need to calculate the mean of responses (average score) of each item. Then, average score will assign the item to a specific category within the Likert scale (average category). To calculate the average score and average scale category for each item, the following steps were followed:

1. **Pre-calculation numbered scale**. Before doing the calculations, the categories of the Likert scale were assigned values from 1 to 7 as follows:

Strongly disagree=1, Disagree=2, Slightly disagree=3, Neither agree nor disagree=4, Slightly agree=5, Agree=6, Strongly agree=7.

Juffs (2001, cited in Mackey and Gass 2016, p. 64) advocates using positive values always for scoring Likert scale (e.g. 1 to 7) rather than negative and positive values with 'zero' as the middle option (e.g. -2,-1, 0, 1, 2). He argues that the latter case makes it difficult to interpret the 'zero' as the *don't know* or *not sure* midpoint.

2. **Mean calculation**. For each item, the summation of the scores (responses) was calculated according to step 1 and divided by the number of the participants (3) to get the mean.

3. **Post-calculation numbered scale.** The resultant mean of each task item assigned that item to an average category on the Likert scale according to the following numerical ranges:

Strongly disagree=1-1.4, Disagree=1.5-2.4, Slightly disagree=2.5-3.4, Neither agree nor disagree=3.5-4.4, Slightly agree=4.5-5.4, Agree=5.5-6.4, Strongly agree=6.5-7.

This scale is similar to the pre-calculation numbered scale (step 1). The only difference is that, instead of an integer, each category on the scale is assigned a numerical range setting its upper and lower limits with the relevant integer based in the centre of the range. To take '*Disagree*' as an example, the integer of this category is (2) which is based right in the middle of the numerical range of this category (1.5-2.4). Using numerical ranges will help deal with the fractions that may result from the calculations done in step 2. The fractions are rounded to the nearest integer. By doing that, the fraction is assigned to specific scale category.

To give an example of calculating an average category, suppose in item (1), the responses were as follows: participant 1 (*Slightly disagree*), participant 2 (*Slightly agree*), participant 3 (*Agree*). According to step 1, the responses would be assigned the following numbers: (*Slightly disagree*= 3), (*Slightly agree*= 5), (*Agree*= 6). Afterwards the mean (average score) would be calculated according to step 2 (3+5+6= 14 → 14/3 = 4.6). according to step 3, 4.6 falls within the '*Slightly agree*' numerical range (by rounding it to 5). Consequently, item (1) would be assigned to the '*Slightly agree*' average category on the post-calculation scale. Table 6 shows how many items were assigned to each average category according to the calculation process above.

Table 6

Items Distribution to Average Categories: First Pilot Test of the Stimuli Selection Task

Average Categories						
Strongly disagree (1-1.4)	Disagree (1.5-2.4)	Slightly disagree (2.5-3.4)	Neither agree nor disagree (3.5-4.4)	Slightly agree (4.5-5.4)	Agree (5.5-6.4)	Strongly agree (6.5-7)
1	1	6	7	23	26	7

Total: 71 items

As shown in table 6, the results skew towards the categories on the right (*Slightly agree, Agree, Strongly agree*). In fact, this was quite expected and justified as all the data was collected after being judged as 'sarcastic' in the metalanguage. In addition, the participants' evaluations were likely influenced by that metalanguage as well, which was exposed in all the tested items, and skewed towards priming the recognition of sarcasm accordingly. The results as they were did not provide balance between the 'sarcastic' right side of the scale and the 'not sarcastic' left side. There was a severe shortage in the 'not sarcastic' items. Such a balance was required for providing an even item pool for the L2 experiment. Investigating sarcasm as a prototype notion requires testing a range of items starting from 'not sarcastic' items going through items with borderline sarcasm and ending with 'very sarcastic' items. In addition, 'not sarcastic' items will serve as control items (distractors) in the main L2 experiment. They will disillusion the participants from the impression they may get that all the items in the L2 experiment have sarcasm. Hence, in order to create a balanced item pool, a second pilot test was required in which more 'not sarcastic' samples were to be used. This is discussed in the next section.

A final word is about the participants' feedback in this pilot study. All participants said that the judgment task was manageable. One noteworthy point was that all participants found part 1 more difficult to respond to than part 2. They said that the contexts of part 1 were hardly enough for understanding and responding to the items. They also mentioned that they did not have enough background knowledge about the topics chatted about on the football forum. The overall attitude of participants about the task as a whole was that it had been well-designed. The participants proposed a couple of suggestions: (1) changing the "*Neither agree nor disagree*" option into "*don't know/not sure*" and (2) using a numerical scale to measure the degree of sarcasm in the task items. In fact, they may sound good suggestions, but they cannot be considered for the following reasons. As for the first suggestion, the original scale wording has been abundantly used in studies and tested by questionnaire designers for validity and reliability (e.g. see Dörnyei, 2010). Therefore, replacing it with another formula needs to be tested by specialists before it can be approved. On the other hand, for a numerical scale to be efficient, it needs two opposite adjectives to be placed on its extremes in order to guarantee polarity evenness on the scale (e.g. *Impolite<-1-2-3-4-5-6-7->Polite*). Such polarity

evenness cannot be guaranteed in the case of sarcasm as the adjective "*Sarcastic*" has no opposite adjective.

5.3.1.2 *Stimuli selection task: Second pilot task*

The second pilot test followed a similar path to that of the first one. The details are as follows:

Participants

Four BrE native speakers participated in this test. The participants were one male and three females aged 19 all. They were all students at different UK universities. The participants were all from Lancashire and were approached via a circulated email from the department of Linguistics and English language/Lancaster University.

Materials

In order to obtain some 'not sarcastic' material, a new search was done within the same online forums (Manchester united club and Mumsnet & Netmums) to collect sufficient data that was judged as 'not sarcastic' in the metalanguage. The search used the prompts *not sarcastic* and *No sarcasm* to search for the required material. The search succeeded in locating 11 new 'not sarcastic' excerpts which were extracted and added to the original material used in the first pilot task (71 excerpts). An example of the 'not sarcastic' data is the following excerpt from Manchester United forum (original pseudonyms retained):

(26)

numChUk NoRis

I live in the United States and I'm a DIE HARD Manchester United fan who doesn't miss a single game.

No Mercenaries:

You are proof that being a true fan is not limited to people whom live in the Manchester postal code.

nige49

Legend mate, I stand up and clap for you. <u>And I am not being sarcastic</u>.

(metalanguage underlined)

The new 82 total of excerpts was processed to hide the metalanguage used in it. This was to avoid the probable influence the metalanguage may have upon the participants' evaluations. After hiding the metalanguage, it turned out that the contexts of some of those excerpts were defected, and this made the affected excerpts unable to convey the not/sarcastic points they had. Accordingly, 12 excerpts were excluded from the total due to insufficiency of context after hiding the metalanguage. As a result, the final total of excerpts was 70. This total was the material of this second pilot test (see appendix C).

Instrument

As was the case with the first pilot test, the 70 excerpts were put into a judgment task with the same 7-point Likert scale. The task was also designed online via 'Qualtrics Survey Software'. The judgment task consisted of two parts: Part 1 contained 35 items taken from the football forum of Manchester United club, and Part 2 included 35 items taken from Mumsnet and Netmums websites. All the questions in the task were randomized and each participant answered a different version. Like the first pilot task, question numbering was also hidden in order to avoid any numbering confusion caused by randomization. Figure 4 shows a sample screenshot of how an item in the SST/second pilot task appeared on the Qualtrics website.

Figure 4. A Screenshot showing a sample item of the online SST on Qualtrics Website: Second pilot test

As was the case with the first pilot task, participants' responses were also saved on the Qualtrics website under pseudonyms and converted automatically into PDF files. The PDF documents were later downloaded and analysed.

Procedure

Similar to the first pilot, the second pilot test was also done online in one session and was self-administered by the participants themselves. The participants received the same covering message used in the first pilot attached with it the same information sheet and consent form as they experienced no changes. They were asked to read the information sheet, sign the consent form and follow the link in the message to do the experiment. The time of completing the task

ranged from 55-62 minutes. Participants in this second pilot task were also paid 5 GBP for their participation.

Results and Discussion

Table 7 shows the results of the second pilot test. It presents the distribution of the participants' choices to the scale categories, i.e., how many each participant chose each scale category in his/her responses of the 70 excerpts. The table also shows the mean and the standard deviation of responses of each participant.

Table 7

Distribution of Participants' Responses to Scale Categories: Second Pilot Test of the SST

	Scale Categories							Mean	SD
	Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree		
Participant 1	5	14	1	0	17	6	27	10	9.79
Participant 2	6	10	1	4	16	21	12	10	7
Participant 3	0	14	10	6	5	23	12	10	7.41
Participant 4	8	14	10	3	10	14	11	10	3.78
Total	19	52	22	13	48	64	62	40	21.45

As it was the case with the first pilot, most of the participants' responses fall within the right-hand categories (*Slightly agree, Agree, Strongly agree*), especially the '*Agree*' category which three participants chose more than others. This indicates that in most cases the metalanguage was not necessary for comprehending sarcasm. On the other hand, what is noticeable was the increase in the 'not sarcastic' (left-hand) responses in comparison to the first pilot. This is

likely to be due to the new strategy adopted in this pilot (i.e., hiding the metalanguage and adding extra non-sarcastic items).

Average categories for the second pilot test were calculated in the same way mentioned in the first pilot task (i.e., according to the same 3-step procedure mentioned in "*Calculating average scores and average scale categories*" above). Table 8 shows the distributions of the 70 items to the average scale categories after calculating the items' average scores.

Table 8

Items Distribution to Average Categories: Second Pilot Test of the Stimuli Selection Task

Average Categories						
Strongly disagree (1-1.4)	Disagree (1.5-2.4)	Slightly disagree (2.5-3.4)	Neither agree nor disagree (3.5-4.4)	Slightly agree (4.5-5.4)	Agree (5.5-6.4)	Strongly agree (6.5-7)
0	7	10	14	11	18	10

Total: 70 items

As it is shown in the Table 8, 17 items were recognized as 'not sarcastic' in comparison to 8 only in the first pilot which represented a considerable increase. Were it the main stimuli selection test, these results would provide a rather balanced item pool for the final L2 experiment— Bearing in mind that we only need few 'not sarcastic' items to function as control items in the final L2 experiment. The biggest concern in these results was the increase in the middle indeterminate category '*Neither agree nor disagree*' from 7 in the first pilot task to 14 in this pilot task. This could be attributed to the absence of the metalanguage that made the participants hesitant in evaluating some of the items. However, 14 items with indeterminate judgments out of 70 items is still an acceptable proportion. The results of this

pilot test proved that the test design was OK as it was and that the way was paved before conducting the main SST experiment.

5.3.2 Conducting the main stimuli selection task (SST)

After considering the results of the first and the second pilot tests, the main SST was ready to be conducted. The details of this experiment are as follows:

Participants

Fourteen BrE native speakers took part in this experiment (six males and eight females). Their ages ranged from 18 to 39 years old ($M = 23.35$ years, $SD = 5.95$ years). They grew up in different UK regions (Lancashire 3, South East England 3, Mid Wales 1, North west England 1, Worcestershire 2, Yorkshire 4). All participants were approached via an email circulated by the postgraduate co-ordinator at the Department of linguistics and English Language/Lancaster University. After receiving requests for participation from willing people, I replied to them using the same covering message of the previous pilot studies. The information sheet and the consent form received no negative notes from participants in both pilot studies. Thus, they were kept the same and used in the main SST without changes.
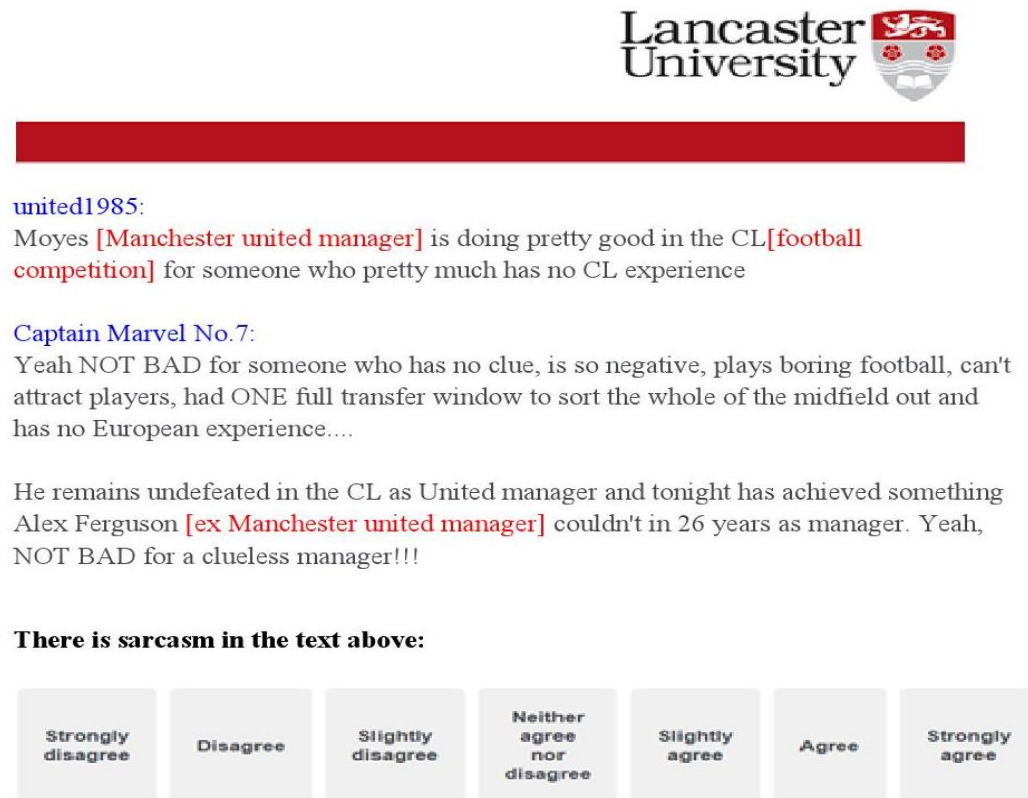
Materials

The materials of the main SST were the same 70 excerpts used in the second pilot test (see appendix D).

Instrument

As was the case with the first and second pilot studies, the main SST was also designed and conducted online via 'Qualtrics Survey Software'. The 70 excerpts were put into a judgment task with the same 7-point Likert scale used in the first and second pilot studies. Similar to the second pilot test, the judgment task consisted of two parts: Part 1 contained 35 items taken from football dataset (Manchester United forum), and Part 2 included 35 items taken from parenting dataset

(Mumsnet and Netmums forums). All the questions in the task were randomized and each participant answered a different version. Question numbering was hidden in order to avoid any numbering confusion caused by randomization. Figure 5 gives a sample screenshot of the main SST on the Qualtrics website.

Figure 5. A Screenshot showing a sample of the Main SST on Qualtrics website

The same procedure of data saving used in the pilot studies was also applied here. Participants' responses were saved on the Qualtrics website under pseudonyms and converted automatically into PDF files. They were later downloaded for analysis and obtaining results.

Procedure

Like the first and second pilot tests, the task was done online, in one session, and was self-administered by the participants. The same covering message used in the two previous pilot tests was also used here. Attached with it were the information sheet and consent form. The

participants were asked to read the information sheet and sign the consent form before doing the experiment. Afterwards they were asked to follow the link enclosed in the message to do the main SST (see appendix D). The time range of completing the task was between 52 to 68 minutes. Participants received 5 GBP each for their participation in this experiment.

Results and Discussion

Table 9 shows the results of the main SST. Like Tables 5 and 7, it presents the distribution of the participants' responses to the scale categories, i.e., how many times each participant chose each scale category in his/her responses of the 70 excerpts. The table also gives the mean and the standard deviation of the responses of each participant.

Table 9

Distribution of Participants' Responses to Scale Categories: Main SST

	Scale Categories							Mean	SD
	Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree		
Participant 1	40	15	3	1	6	3	2	10	14.05
Participant 2	23	18	7	2	7	9	4	10	7.66
Participant 3	17	15	6	3	13	12	4	10	5.60
Participant 4	12	15	5	2	18	13	5	10	6.00
Participant 5	8	16	3	3	12	19	9	10	6.11
Participant 6	7	15	2	3	13	19	11	10	6.30
Participant 7	5	13	4	0	14	19	15	10	6.98
Participant 8	3	12	4	2	8	25	16	10	8.35
Participant 9	0	13	4	2	7	26	18	10	9.47
Participant 10	0	12	5	1	7	22	23	10	9.42
Participant 11	0	6	9	1	4	25	25	10	10.68
Participant 12	0	5	5	3	3	17	37	10	13.08
Participant 13	0	3	4	3	3	8	49	10	17.36
Participant 14	0	2	1	1	5	7	54	10	19.56
Subtotal	115	160	62	27	120	224	272	140	86.36
Total		337		27		616			

Table 9 reveals that participants, generally speaking, favoured the right-side categories of the scale (*Slightly agree, Agree, Strongly agree*) in most of the cases (616 in total). This indicates that the majority of the sarcasm used in 70-excerpt material was validated by the NS participants. On the other hand, the total of responses to the left of the scale (*Strongly disagree, Disagree, Slightly disagree*) was also significant (337 in total). The middle option '*Neither agree nor disagree*' was chosen only 27 times in total and this denotes the low cases of sarcasm indeterminacy within the data. The results of individual participants anticipate a good item pool for the final selection of stimuli to the L2 experiment.

To calculate average scores and average categories, scores of the individual participants were processed according to the same 3-steps procedure used in the first and second pilot studies (see 5.3.1.1). Table 10 below presents the distribution of the 70 items of the SST to the average categories.

Table 10

SST Items Distribution to Average Categories: Main SST

	Average Categories						
	Strongly disagree (1-1.4)	Disagree (1.5-2.4)	Slightly disagree (2.5-3.4)	Neither agree nor disagree (3.5-4.4)	Slightly agree (4.5-5.4)	Agree (5.5-6.4)	Strongly agree (6.5-7)
	1	10	7	6	19	18	9
Subtotal		18		6		46	
Total				Total:70			

As shown in Table 10, 46 items out of 70 fall within the right/agreement side of the scale. This means the sarcasm in these items is ultimately validated. On the other hand, only 18 items are within the left/disagreement side. It is a rather low number, but seems to be enough for providing sufficient control items (distractors) to the L2 experiment. Only 6 items fall within the middle average category '*Neither agree nor disagree*' which indicates that the level of sarcasm indeterminacy was low in the total of the tested items. The distribution shown in Table 10 offers a

good item pool for the final L2 experiment. But average categories differ among each other in how many items they include and these different proportions need to be taken into account in the final selection of L2 experiment stimuli. This is because these different proportions represents different degrees of sarcasm judgments in reality and any well-made experiment should reflects the reality as it is. Overinclusion of one average category or more in the L2 experiment at the expense of other categories may seriously affect the experiment and distort its results and its findings ultimately. Thus, the final stimuli of the L2 experiment had to be carefully selected. The final selection is discussed next.

5.4 Stimuli of the L2 Pragmatics Study: Final Selection

For reasons of practicality, it was not possible to include all the 70 excerpts used in the main SST into the L2 pragmatics experiment. The main SST was a single-task experiment involving rating sarcasm only and was done by solely English native speakers. Despite that fact, the time range of completing that task exceeded one full hour. On the other hand, the planned L2 experiment was intended to be a double-task experiment involving rating sarcasm and identifying it within excerpts (see chapter seven). It was also intended to be done by one group of BrE native speakers and two groups of Iraqi EFL learners who were expected to take longer than NSs to finish the experiment. The L2 experiment was also planned to take no more than one hour to finish. For the above reasons, only some of the 70 excerpts can be ultimately selected as stimuli for the L2 experiment, not all of them.

Regarding the completion time of the L2 experiment, one hour was the maximum time with which one could make the most of the informants' participation. Exceeding that time limit would, possibly, have resulted in the *fatigue effect*[16]. Dörnyei (2010, p.12) writes that in the L2 field " Most researchers agree that anything that is more than four to six pages long and requires over half an hour to complete may be considered too much of an imposition". However, as the L2 experiment was complex in being double-tasked and containing different degrees of sarcasm, I needed to exceed Dörnyei's limit and have the L2 experiment been done within a full hour.

---

[16] Fatigue effect occurs when respondents feel tired or bored after doing a long or monotonous experiment. Respondents may give inaccurate answers towards the end of the experiment due to this effect (see Dörnyei, 2010, p. 9).

Coming to the procedure of selecting the final stimuli, it was not that complicated. First, I decided to choose 30 items out of 70 to be the material of the L2 study. Thirty was a reasonable number of items to be answered within one hour. But it was the job of the pilot test of the L2 experiment to tell us whether 30 items can be done within one hour or not, i.e., whether they need more trimming or adding (see chapter six). Secondly, as was mentioned earlier, if the L2 experiment is to be well-designed and reflecting what happens in real life, all the average categories of the SST need to be represented in the final selection of stimuli.

The representation of each average category relied on its score in the main SST (see Table 10). Table 10 presents the scores or the proportions of the SST average categories out of 70. Similar proportions should appear in the final stimuli out of 30. By using this simple equation ($x/70 = y/30$), I could calculate how many items of each average category would go in the final stimuli (with the fractions rounded to the nearest integer). The results were as follows: (*Strongly disagree*, 1 item), (*Disagree*, 4 items), (*Slightly disagree*, 3 items), (*Neither agree nor disagree*, 2 items), (*Slightly agree*, 8 items), (*Agree*, 8 items), (*Strongly agree*, 4 items), (total, 30 items). Regarding which items to be selected from each average category, this was also systematic. Each average category had a numerical range (e.g. *Slightly agree* = 4.5-5.4) along which the items of that category spread. The selection of items was from the beginning, the middle and the end of each category range respectively.

5.5 Summary

This chapter has expounded the process of preparing stimuli for the coming L2 pragmatics study. First, it stated how the collected 142-excerpt data was initially trimmed into 71 excerpts by applying the length filter. It was systematically applied by means of setting upper and lower word count limits for the excerpts to be selected as stimuli. The limits were calculated according to the word counts of the football and parenting datasets along with their means and standard deviations. Second, the chapter also stated how the 71 excerpts, which passed through the length filter, were further trimmed by means of the SST. The SST was designed to filter the data more as well as to validate the sarcasm in it. The SST was piloted twice and after doing the necessary changes, the main SST was conducted. According to the results of the main SST, 30 items were

ultimately selected to be the final stimuli of the L2 experiment. The selection was via a systematic procedure (i.e., basing on the items scores within average categories).

By selecting 30 items form the data to be the final stimuli of the L2 pragmatics study, we come to the end of part 2 of this thesis. In part 3, we shall start detailing what happened in conducting the L2 experiment as well as its results and their discussions.

PART 3

Measuring Sarcasm Recognition

(L2 Pragmatics Study)

Chapter Six

Piloting the L2 Pragmatics Study

6.1 Introduction

At the end of chapter five, it is mentioned that 30 items were ultimately selected from the item pool as the final stimuli for the L2 pragmatics study which would measure sarcasm recognition. Before conducting the main L2 study, these items were put into a pilot study which served a number of purposes. First, it was intended to test whether the selected items were appropriate stimuli for sarcasm recognition in terms of item length, item reading time, and sufficiency of sarcasm context. Second, the pilot study also aimed at testing whether the judgment task designed was capable of measuring sarcasm recognition. The sarcasm recognition task used in this pilot test was different from the stimuli selection task in being a two-fold instrument (see 5.3.2). The third purpose of the test was to uncover how long the experiment would take. Fourth, the test was also made to highlight and fix any problems that relate to the design and content of the experiment. A further fifth purpose of this pilot study was to give us an insight of what kind of results we can expect in the main L2 study. This chapter is devoted to discussing the piloting of the L2 pragmatics study in detail.

6.2 Research Questions

The pilot study attempts to answer the following questions:

1. Are the 30 items used appropriate stimuli for sarcasm recognition in terms of length, sarcasm context sufficiency and time spent for reading them?

2. Can sarcasm recognition be measured by means of a judgment task?

3. Are the instructions given at the beginning of the task informative enough to recognize what is required?

4. Are the task format and design participant-friendly?

5. How long will the judgment task take?

6. Are the examined groups significantly different from each other as regards sarcasm

recognition?

6.3 Methods

6.3.1 Participants

As the main study was to be applied to British English native speakers and Iraqi EFL learners, it was necessary to choose a sample from all those populations for the pilot study. Nine participants were recruited for piloting the L2 pragmatics study (four males and five females; age range 18-45 years, $M = 26.55$, $SD = 8.10$ years). They were divided into three groups. First, "English-NS" group which contained three British-English native speakers who were all from Lancashire. Two of them were undergraduate students and one was working in the private sector. Second, "Iraqi EFL-UK" group which involved three Iraqi EFL learners who had been studying in the UK for 2-3 years. They finished their BA and MA degrees in English in Iraq and were all PhD students at different UK universities. As for English language proficiency, their IELTS scores were 6, 6.5 and 6.5. Based on these scores, two of them were advanced (C1 level) and the third participant was upper intermediate (B2 level) in English according to the *Common European Framework of Reference for Languages* (CEFR) (see appendix G). Third, "Iraqi EFL-home" group which comprised three Iraqi EFL learners studying in Iraq who had never been to any English-speaking country. Two of them were PhD students and one was an MA student. Their IELTS score were 5.5, 5.5 and 6. They were all of upper intermediate level (B2) as regards English language proficiency. All the Iraqi learners were Arabic native speakers. Three of them grew up in the south of Iraq, two in the west of Iraq and one grew up in the Iraqi capital Baghdad. All participants were invited to participate via email. Some of them were friends and others were colleagues. A snowballing element was involved in recruitment as some participants invited others for the experiment. The same information sheet and consent form used earlier were used in this pilot study as they received no changes.

6.3.2 Materials

The material of this pilot study was the 30 items which were finally selected from the item pool according to the stimuli selection task (SST) (see 5.4 and appendix E).

6.3.3 Instrument

This pilot study was designed as a judgment task experiment similar to the SST with the same 7-points Likert scale (see 5.3.2 and appendix D). Regarding content, the judgment task consisted of two parts: Part 1 which contained 15 items from the football forum of Manchester United club, and Part 2 which also contained 15 items from the parenting forums of Mumsnet and Netmums websites. Regarding what was required, the judgment task was a two-fold instrument. First, participants were asked to read the items and rate the degree of sarcasm in each one on the 7-points Likert scale. This was the 'Sarcasm Rating' part of the task which was similar to the task done in the SST. Second, participants were also asked to highlight/underline the part(s) of the item text which they believed to be sarcastic. This was the 'Sarcasm Identification' part of the task.

As for the design of the pilot judgment task, the initial attempt was to design the experiment online on Qualtrics website as was done with the SST. Unfortunately, Qualtrics did not allow a text-editing function (e.g. highlighting or underlining) and this made it impossible to do part 2 of the task (Sarcasm Identification). Thus, this option was excluded. Then, Google docs was used to do the online design instead. Google docs afforded a text-editing function. However, it was not a practical option due to some failures. Five of the EFL learners reported that they faced difficulty in doing part 2 of the task as they were not familiar with Google docs and how to edit texts on it. Some of them suggested designing the task with some other format. Hence, online design was abandoned altogether and manual design was adopted alternatively. As for manual design, PDF format also proved not to be practical. Although it allows minor editing to texts such as highlighting or underlining, some participants failed to do especially part 2 of the task on it. Perhaps, the adobe reader version they had was not up-to-date. Ultimately speaking, the judgment task was designed as a Word file because it allows editing texts easily and people are more familiar with this electronic file format. Participants were asked to use either the highlighting or underlining functions built-in

in Word to do both parts of the judgment task. For randomization purpose, three versions of the experiment Word file were produced each of which had a different order of the items. Figure 6 shows a sample of how an item within the pilot judgment task was answered on Word.

Figure 6. A Screenshot showing a sample of how the pilot L2 pragmatics study was answered on Word. The purple highlight within the Likert scale represents the participant's sarcasm rating, whereas the highlight within the text indicates where the participant believes the sarcasm to be (Sarcasm Identification).

This pilot experiment was submitted to the Research Ethics Committee at Lancaster University and was approved by it prior to putting it into practice.

6.3.4 Procedure

The participants were approached via email. They were sent a covering message similar to the one used in the SST. Attached with it were the judgment task, an information sheet and a

consent form. Participants were asked to read the information sheet and sign the consent form and then they were asked to do the task in one session and time themselves while doing it. They were allowed to have a short break if they needed to so as to avoid any potential fatigue effect. Participants were also requested to take extra care not to change the original wording of items while providing answers. To maintain randomization, the participants received different versions of the experiment file which contained different item orders. After doing the task, every participant saved the changes on his/her Word file and emailed it back to me along with the signed consent form. They were also asked to provide feedbacks about the design and content of the experiment. The experiment was self-administered by the participants. The task completion time ranged from 50 to 60 minutes. The participation of the Iraqi EFL learners was free due to being friends of the researcher, whereas the English native speakers were paid 5 GBP each for their participation.

6.3.5 Data Analysis

Each part of the judgment task was analysed separately. First, the data of the 'Sarcasm Rating' part of all the three groups, which were the participants' choices on the Likert scale, were converted into values according to the following numerical range:

Strongly disagree=1, *Disagree*=2, *Slightly disagree*=3, *Neither agree nor disagree*=4,

Slightly agree=5, *Agree*=6, *Strongly agree*=7.

Then, within each group, the "average score" was calculated for every participant. That was the mean of the participant's answers to all the 30 items. For example, suppose participant 'A' had the following ratings to the 30 items in the experiment after converting them into values according to the numerical range above:

3+2+5+6+6+6+7+3+4+7+7+5+5+6+4+3+1+4+2+6+7+6+4+5+5+4+6+7+5+6

The average score of participant 'A' would be the mean of these individual score (147/30 = 4.9). In this way, the average score was calculated for all participants. Afterwards, "overall score" was calculated for each examined group. That was the mean of the average scores of the three

participants in each group. For example suppose Group 1 had the following average scores (Participant A = 4.5, Participant B = 5.2, Participant C = 6.4), the overall score of Group 1 would be (4.5+5.2+6.4= 16.1 → 16.1/3 = 5.36). The next step was checking groups overall scores for the assumptions of normality of distribution and homogeneity of variances. The former was verified by the Shapiro-Wilk test and the latter by the Levene's test using SPSS. Then, One-Way ANOVA was applied to the data to obtain results (see 6.4).

As for the 'Sarcasm Identification' part of the task, the data was compared across the examined groups. This was to reveal any matches and mismatches among participants in the parts they highlighted within items texts (i.e., where they believed sarcasm to be). The procedure of analysing the data obtained from this part of the task was a two-step procedure. The first step was that within each examined group, every item was inspected to find out which of its parts was highlighted by participants and how often. On the group level, any highlighted part was considered a 'sarcastic part' if it achieved consensus within the group or was highlighted by the majority of that group members (2 members out of 3). For example, in Figure 6 above, there is one highlighted part within the item text "*What a surprise there…*". Suppose that this part of the text was highlighted by all the three participants in the English-NS group, by two participants out of three in the Iraqi EFL-UK group, and by one participant out of three in the Iraqi EFL-home group. In this case, it would be considered as a sarcastic part for the English-NS group because it achieved consensus (highlighted by all participants in the group). It would also be considered as a sarcastic part for the Iraqi EFL-UK group because it was highlighted by the majority of this group (2 out of 3). However, it would not be considered as a sarcastic part for the Iraqi EFL-home group as it was highlighted by only the minority of this group (1 out of 3).

The threshold of consideration is that any highlighted part was considered as a sarcastic part provided it is chosen by the majority of at least one examined group. In this way, the sarcastic parts within all the tested items were identified. This procedure was applied to the data of all the three examined groups. After identifying sarcastic parts comes the second step of the analysis procedure. The identified sarcastic parts were compared across the groups to find out sarcasm matches and mismatches. Any sarcastic part was considered: (1) a 'Full match' if it was highlighted by all the examined groups, (2) a 'Partial match' if it was chosen

by two groups only, and (3) a 'No match' (Peculiar sarcasm) if chosen by one group only. The results of the pilot analysis are presented in the next section.

6.4 Results

Regarding sarcasm rating, results show that the mean rating was 4.13 (*SD* = 1.66) in the English-NS group, 5.24 (*SD* = 1.30) in the Iraqi-EFL group and 5.16 (*SD* = 1.14) in the Iraqi EFL-home group. The assumptions of distribution normality and homogeneity of variances were examined for all the groups and were met (the Shapiro-Wilk test: English-NS *p* = .40, Iraqi EFL-UK *p* = .10, Iraqi EFL-home *p* = .16; Levene's test: $F(2, 6) = 1.42$, *p* = .21). One-way ANOVA was run to test whether any statistically significant difference was available among the three groups. ANOVA results reveal a significant difference found among the examined groups [$F(2, 6) = 5.98$, *p* = .004]. The effect size of that difference is moderate (η^2 = .12). Results of the post-hoc Tukey HSD test show two significant differences. The first is between the 'English-NS' group and the 'Iraqi EFL-UK' group (*p* = .007), whereas the second is between the 'English-NS' group and the 'Iraqi EFL-home' group (*p* = .01). No significant difference is found between the 'Iraqi EFL-UK' group and the 'Iraqi EFL-home' group.

The above results show that both learners groups performed differently from the English native speakers. Learners were more likely to place their choices within the right side of the Likert scale (*Slightly agree, Agree, Strongly agree*) than the native speakers. On the other hand, the learners groups were not significantly different from each other in their performance. It seems that studying abroad has no effect in perfecting the performance of Iraqi EFL learners towards the native standard as regards sarcasm recognition. If the main L2 pragmatics study would yield similar results, an explanation could be provided for why this is the case after analysing the results.

Regarding 'Sarcasm identification' part of the task, Table 11 presents the results of this part. It shows which items contain sarcasm and which do not. In addition, it also shows the cases of full match, partial match and peculiar sarcasm across the tested items.

Table 11

Sarcasm Identification Results: Sarcasm Availability and Sarcasm-Matching Status across the Tested Items

Item	Sarcasm Available?	Sarcastic Part		Sarcasm-Matching Status		
		Subtotal	No.	Full Match	Partial Match	No Match (Peculiar Sarcasm)
1	Yes	1	1			Iraqi EFL-UK
2	Yes	1	1			English-NS
3	No					
4	No					
5	Yes	1	1	Yes		
6	Yes	2	1	Yes		
			2			Iraqi EFL-home
7	Yes	1	1			Iraqi EFL-home
8	Yes	1	1		Iraqi EFL-UK and Iraqi EFL-home	
9	Yes	1	1		Iraqi EFL-UK and Iraqi EFL-home	
10	Yes	1	1		English-NS and Iraqi EFL-home	
11	Yes	2	1			Iraqi EFL-home
			2			Iraqi EFL-home
12	Yes	1	1		Iraqi EFL-UK and Iraqi EFL-home	
13	Yes	1	1	Yes		
14	Yes	2	1			Iraqi EFL-home
			2			Iraqi EFL-UK
15	Yes	2	1			Iraqi EFL-home
			2		Iraqi EFL-UK and Iraqi EFL-home	
16	Yes	3	1			English-NS
			2	Yes		
			3	Yes		
17	No					

Item	Sarcasm Available?	Sarcastic Part		Sarcasm-Matching Status		
				Full Match	Partial Match	No Match (Peculiar Sarcasm)
		Subtotal	No.			
18	No					
19	Yes	1	1	Yes		
20	Yes	1	1	Yes		
21	Yes	1	1	Yes		
22	Yes	1	1	Yes		
23	Yes	1	1		English-NS and Iraqi EFL-UK	
24	Yes	2	1			Iraqi EFL-UK
			2			Iraqi EFL-UK
25	Yes	2	1			Iraqi EFL-UK
			2	Yes		
26	Yes	1	1			English-NS
27	Yes	1	1	Yes		
28	Yes	1	1	Yes		
29	Yes	1	1			Iraqi EFL-home
30	Yes	1	1			Iraqi EFL-home
		Total: 34				

Table 11 shows that only 4 items out of 30 were not recognized as having sarcasm. In fact, this number is only half of 8 which is the number of the disagreement items (i.e., items judged with *Strongly disagree*, *Disagree* or *Slightly disagree*) included in this test according to the SST. This means that participants in this pilot study have different judgments from those provided by the SST participants. The table also shows that there are 34 sarcastic parts in total and that some items have 2 or even 3 sarcastic parts within them. Only 12 sarcastic parts achieve a full match which is a rather low proportion. In addition, only 6 sarcastic parts are partially matched and in most cases the partial match is between learners' groups. Sixteen sarcastic parts were 'peculiar sarcasm' to only one of the examined groups or the other. Most of this number goes to the learners' groups as well. The above results indicate a deviation by

the Iraqi EFL learners from the native standard of sarcasm perception. But this assumption needs to be confirmed by the main L2 study before it can be attested as a finding.

6.5 Discussion

After conducting the pilot study and obtaining the results, it is time to answer the research questions raised at the beginning of the chapter to see whether the objectives of the pilot study have been achieved.

1. Are the 30 items used appropriate stimuli for sarcasm recognition in terms length, sarcasm context sufficiency and time spent for reading them?

Yes, they are. No serious notes were received from participants as regards the length and content of items as well as the time spent in reading them. However, the only note I received in this regard was that some participants said they were not familiar with the material of the football forum. But they all said that the explanations provided between square brackets were very helpful in clarifying things.

2. Can sarcasm recognition be measured by means of a judgment task?

Judgment task in this pilot study was successful in obtaining judgments of sarcasm recognition on the 'Sarcasm Rating' and 'Sarcasm Identification' levels. Hence, the answer is yes. This instrument can be used to measure sarcasm recognition and will be retained and used in the main L2 study. Sarcasm rating task and sarcasm identification task were separately and successfully used by Togame (2016) and Kim (2014) respectively. The current pilot study (and the main L2 pragmatics study as well) combines both of these tasks into a single two-fold judgment task. This is to see whether this combination would be good enough innovative improvement that can yield more interesting and interrelated results. The pilot study results are encouraging and predict that this combination would make a better instrument.

3. Are the instructions given at the beginning of the task informative enough to

   recognize what is required?

Yes, they are informative enough as most participants were able to recognize what was required in both parts of the judgment task. In addition, I received no negative feedback about these instructions. However, they will be amended in the main L2 study to avoid any possible failures similar to the ones occurred in this pilot study (see appendix E).

4. Are the task format and design participant-friendly?

In their feedback, participants said that Word format was good and easy to manage although some suggested it had been better if the experiment was designed online. In fact, this was already in mind, but online design was not possible due to the problems mentioned in 6.3.3 above. Furthermore, the Likert scale used received no complaints or notes from the participants. Thus, it will be retained and used as it is in the main L2 study.

5. How long will the judgment task take?

In this pilot study, participants accomplished the task within 50 to 60 minutes. No negative feedback was received regarding the time of the experiment. This means the time span of finishing the task is reasonable. As the study was done within around an hour, which is the planned completion time (see 5.4), the same 30 item used in this pilot study will be retained for the main L2 study without reduction or addition from the item pool.

6. Are the examined groups significantly different from each other as regards

   sarcasm recognition?

Regarding 'Sarcasm Rating' part of the judgment task, ANOVA and Tukey results show statistically significant differences between learners' groups and native speakers group. On the other hand, learners' groups are not significantly different from each other. As for 'Sarcasm

Identification' part of the task, results show rather few cases of full match (only 12 out of 34 sarcastic parts) among the examined groups and even fewer ones of partial match (6 only). Around half of the sarcastic parts (16) are 'peculiar sarcasm' cases belonging to different groups. These results indicate differences among the examined groups in sarcasm identification.

6.5.1 Limitations

Apart from the problems relating to design discussed earlier (see 6.3.3), the major setback encountered in this pilot study was the misuse of part 2 of the task (sarcasm identification). Some participants answered part 1 (Sarcasm Rating) of some items with disagreement (i.e., they chose *Strongly disagree, Disagree,* or *Slightly disagree*) or chose the middle option *Neither agree nor disagree*. This indicated that there was no sarcasm within the given texts or the existence of sarcasm in those texts was questionable. Despite that, they confusingly highlighted parts of those texts in indication of where sarcasm was. Consequently, the instructions of the main judgment task were amended to address this possible failure in the main L2 study. The participants were contacted and asked to correct their answers. Another failure in this pilot study was that two of the participants did only highlight choices on the Likert scale and forgot to highlight sarcastic parts within texts. They were approached again and were asked to do the 'sarcasm identification' part of the task. Finally, a minor failure was that one of the participants did not use the underlining or the highlighting tools to provide answers. Instead, he changed the original font colour to indicate his choices on the scale and the sarcasm area in texts. This was rather confusing, but responses were eventually located successfully.

6.6 Conclusion

In this chapter, the experiment of piloting the L2 pragmatics study was detailed. First, the chapter presented the research questions the pilot study attempted to answer. Second, it detailed the methods of performing the experiment. Nine participants took part in this pilot study. They were three British-English native speakers, three Iraqi EFL learners studying in the UK, and three Iraqi EFL learners studying at home (Iraq). The material of the study was

the 30 items that were finally selected according to the SST results. The instrument used was a judgment task which was ultimately designed as a Word file. The judgment task was a two-fold instrument. The first part was to measure the degree of sarcasm in texts on a 7-point Likert scale (Sarcasm rating), whereas the second part was designed to determine where sarcasm to be within texts according to the participants' beliefs (Sarcasm identification). As for the procedure, the experiment was self-administered by the participants. The experiment file was sent to them by email and after doing the experiment they saved the changes and emailed the file back to me.

ANOVA and Tukey results of the 'Sarcasm rating' data showed a statistically significant difference between learners' groups and the English natives' group. In contrast, no significant difference was found between the learners' groups themselves. The results of the 'Sarcasm identification' data revealed few cases of full match among the groups and even fewer cases of partial match. Most of the partial match cases were between the learners' groups. About half of the sarcastic parts highlighted in this experiment (16 out of 34) are cases of 'peculiar sarcasm' belonging to one examined group or the other.

Towards the end of the chapter, answers are provided to the research questions. They are based on what happened in this pilot experiment. The answers of these questions show that the objectives of this pilot study, which were mentioned at the beginning of the chapter in 6.1, have been generally achieved. By considering the results of this pilot study and answering the research questions, the way is paved for conducting the main L2 pragmatics study. Chapter seven will give a full account of the methods of this main study.

Chapter Seven

L2 Pragmatics Study: Methods

7.1 Introduction

This chapter describes the main L2 pragmatics study which aims at measuring the sarcasm recognition of Iraqi EFL learners and comparing it to that of the English native speakers. The study measures two aspects of sarcasm recognition: rating sarcasm and identifying it within the given texts. The chapter is devoted to detailing the methods of conducting the main L2 study and the methods of analysing the data obtained from it. The methods are mainly based on the results of the pilot L2 study which provided useful insights of how to design the main study and analyse the data.

7.2 Participants

Ninety participants were recruited to do the main L2 pragmatics study. They were 47 females and 43 males. Their ages ranged from 18 to 47 years ($M = 33.1$, $SD = 9.22$ years). Thirty of them were monolingual British-English native speakers, 58 were Arabic-English emerging bilingual (EFL learners, Arabic L1), and only two were Arabic-Turkish and Arabic-Kurdish balanced bilinguals (EFL learners as well). All participants were students studying at different British and Iraqi universities (23 undergraduates and 67 postgraduates). Their ethnic groups were white British (29 participants), British Asian (1 participant), Arab (58 participants), Iraqi-Kurdish (1 participant), and Iraqi-Turkmen (1 participant).

Participants were evenly divided into three groups as follows:

1. **English-NS**. This group consisted of 30 British-English native speakers participants who grew up in different UK regions [19 females/11 males, age range 18-47 years ($M = 26.53$, $SD = 11.55$ years), 23 undergraduates and 7 postgraduates]. They were studying different majors at Lancaster University. Participants of this group were paid 5 GBP each for their participation. The group was used as a benchmark group to which the performances of the other two experimental groups were compared.

2. **Iraqi EFL-UK**. This group comprised 30 Iraqi EFL learners studying at different UK universities [14 females/16 males, age range 27-43 years ($M = 36.96$, $SD = 4.45$ years)]. Their length of stay in the UK ranged from 1 to 4 years. They were all postgraduates students (MA and PhD students). In fact, I attempted to make this group a mixture of undergraduates and postgraduates. Unfortunately, no Iraqi undergraduate EFL learners were studying in the UK at the time of conducting the experiment. As for English language proficiency, the levels of the 'Common European Framework of Reference for Languages' (CEFR) were adopted in this study for determining L2 proficiency (see appendix G for CEFR levels). Most of the participants in this group (28 participants) scored 6.5-7 in IELTS. This range of score classifies them as having advanced English language proficiency (C1 level). Only two participants in this group were of upper intermediate proficiency (B2 level) as their IELTS scores were 6 for both. The L1 of this group was Arabic except one participant who was a Kurdish-Arabic balanced bilingual. Participation of this group was voluntary and without payment in return.

3. **Iraqi EFL-home.** This group also had 30 EFL Iraqi learners studying at different Iraqi universities [14 females/16 males, age range 22-45 years ($M = 35.8$, $SD = 6.33$ years)]. They were all MA and PhD postgraduates students. This was to control the 'Study Level' variable and create balance with 'Iraqi EFL-UK' group in this regard. Regarding English language proficiency, 20 participants scored 6.5 and were classified as advanced in using English (C1 level). The remaining 10 all scored 6 in IELTS and were of upper intermediate proficiency (B2 level). As for L1, they were all Arabic native speakers except one who was a Turkish-Arabic balanced bilingual. Participants of this group have never been to any English–speaking country before. As was the case with Iraqi EFL-UK group, participation in this group was voluntary and without payment in return as well.

An attempt was made to achieve balance among the groups with regard to gender, age, and English language proficiency variables. However, the voluntary nature of participation yielded the current numbers mentioned above, some of which were not fully balanced.

The examined groups were statistically tested to uncover any significant differences among them in terms of gender and age. Chi-square results indicated no significant difference among the groups in terms of gender [$\chi^2(2) = 2.23$, $p = .33$]. As for English language proficiency, results revealed a significant difference between the learners' groups [$\chi^2(1) = 6.67$, $p = .01$]. The One-Way ANOVA showed a statistically significant difference among the groups as regards age, $F(2, 87) = 15.20$, $p < 0.001$, $\eta^2 = 0.26$. In order to pinpoint where exactly was the significant difference, the Games-Howell[1] post hoc test was run. It has revealed that the significant differences was between English-NS group and Iraqi EFL-UK group ($p < .001$), and between English-NS group and Iraqi EFL-home group ($p = .001$). On the other hand, there was no significant difference between Iraqi EFL-UK and Iraqi EFL-home groups ($p = 0.68$) in terms of age.

Regarding the sampling of participants, it was not possible to adopt the *probability sampling*[2] (which involves that every member in the population has an equal chance to be selected for the sample, see Rea and Parker, 1992) in the current study as is the case with almost all researches in social sciences. Non-probability sampling[3] was used instead in recruiting participants for the study which is quite common in second language research (see Mackey and Gass, 2016). Among the several non-probability sampling techniques, two were operated in the current study: *convenience sampling* and *snowballing sampling*. Convenience sampling is "the most common non-probability sampling type in L2 research" (Dörnyei, 2010, p. 61). It involves that "members of the target population are selected for the purpose of the study if they meet certain practical criteria, such as geographical proximity, availability at a certain time, or easy accessibility" (p. 61). In other words, the researcher in this technique approaches and samples only the members of the population that are convenient and easy to reach for him/her. Snowballing sampling involves the researcher invites few people to do the study. Then, they are further asked to invite other members of the same population to participate in the study and so on. By continually adding new people, the number of participants will grow like a snowball (see Dörnyei, 2010 and Griffee, 2012).

The convenience and snowballing techniques in this study were online in nature. Participants were studying at different places and countries and it was almost impossible to gather them physically to do the study. Convenience technique was mainly used to recruit

English native speakers (by means of an email circulated to Lancaster University students) and Iraqi EFL learners studying in the UK (by means of an announcement posted on Facebook). On the other hand, snowballing technique was used with Iraqi EFL learners studying in Iraq. Facebook was also employed for this purpose. Some Facebook friends were approached and asked to do the study as being Iraqi EFL learners. Afterwards they were asked to approach friends and colleagues of theirs and ask them to participate in the study as well. All participants were sent the experiment file as an attachment via email or Facebook along with a covering message. They were also sent the same information sheet and consent form used in the pilot L2 study as they received no changes.

Finally, we turn to address why the study have chosen this sample size (90 participants in total, 30 participants in each group). Dörnyei (2010, p. 62) states that there is no hard-and-fast rule for how many to sample in second language research. The general rule in scientific research is that the larger the sample the more it approaches the population size and the more it is representative of it (see Brown, 1988). Sample size is a key issue for the generalizabilty of results to the population as very small samples might not be representative of their populations. In sum, the larger the sample the better and sample size must be deteremined with care.

When it comes to my sample size, I can assume that a total N of 90 participants is a good sample size. Dörnyei (2010, p. 62-63) argues, from the statistical significance perspective, that in order to reach statistical significance in L2 studies, we need a sample size of 50 participants and more. Most of the reviewed studies on L2 irony/sarcasm used near this number of participants or less (see 3.5).

Regarding the sample sizes of individual groups (n-size), 30 participants each looks a rather small number, but it is not. Plonsky (2013) assessed 606 published SLA studies and found that the median sample size in these studies is 19 participants. Furthermore, most of the L2 pragmatics studies in general and L2 irony/sarcasm studies in particular reviewed in this study used similar n-sizes for their individual groups (see 3.3, 3.4 and 3.5). In fact, the 30-participant n-size for all the groups was mainly determined by the Iraqi EFL-UK group. The population of Iraqi EFL learners in the UK was rather small (73 students only according to a statistic by the Iraqi embassy in London). Due to voluntary nature of participation, only 30 people from this population were responsive when they were invited to do the experiment. Therefore, I decided

to take the same number of participants from the other two populations (British-English native speakers and Iraqi EFL learners studying in Iraq). This was to avoid the statistical problems caused by unequal sample sizes[4].

Supportive evidence comes from the literature that validates the capacity of the current *n*-size (30 participants) to yield generalizable results. For example, Dörnyei (2010) states that "a basic requirement is that the sample should have a *normal distribution*, and a rule of thumb to achieve this, offered by Hatch and Lazaraton (1991), is that the sample should include 30 or more people" (p. 62). Similarly, Brown (2001) argues that when the sample size is within the domain between 25 to 30 participants, it is likely to achieve distribution normality which is, in turn, a requirement for accurate statistics and generalizable results. Finally, Fraenkel and Wallen (2003) propose minimum *n*-sizes as guidelines for different kinds of studies. They propose that 15-30 participants per group is the minimum *n*-size for experimental studies.

7.3 Materials

The materials of the main L2 pragmatics study were the 30 stimuli which were carefully selected by means of the SST (see appendix E). All the average categories of the SST were represented with different proportions in this study according to their scores in the SST (see 5.4). Half of the stimuli was football data, whereas the other half was parenting data. As detailed in chapter six, the stimuli were piloted beforehand in order to highlight and fix any problems before involving them into the main L2 study.

7.4 Instrument

The selected stimuli were put into a sarcasm judgment task similar in design to the one used in the pilot L2 pragmatics study. Taking the pilot study results into account, the main task was designed as a Word file which contained the following (for details, see appendix E):

1. **Introduction.** It included four distinct things: (1) a starting sentence meant to arouse the participant's interest, (2) the participant was briefed with the experiment topic, (3) the participant was ensured about the data confidentiality and anonymity of his/her identity, and (4) a final thanking sentence. Two versions of the introduction were made with a slight

difference between them. That was to address whether the participant was an English native speaker or an Iraqi EFL learner.

2. **Instructions**. After the introduction, the participant was briefed with the number of items s/he would answer and the structure of the judgment task. S/he was also instructed of how to do the two required parts of the task followed by an illustrative example.

3. **Main body**. The main body of the judgment task presented the tested stimuli. It contained 30 items. Each item contained one excerpt (stimulus) quoted from the source online forum (football or parenting). The original wording, layout and even the pseudonyms were retained in all. So, participants were as if reading from the online forum directly. The main body was divided into two parts: Football part and Parenting part. Each part contained 15 items. All items ended with the exact 7-point Likert scale used in the earlier experiments.

4. **Information questions**. The last section of the judgment task comprised some factual questions about the participants. Questions about gender, age, L1, L2 proficiency, length of stay in the UK, etc. were all incorporated in this section. I followed Dörnyei's (2010) advice to include such personal questions towards the end of the instrument.

The judgment task in the main L2 study was a two-fold task. Participants were asked to read the text (prompt dialogue) in each item and respond to the two parts of the judgment task simultaneously:

1. **Part 1(Sarcasm Rating)**: passing judgments whether or not the text had sarcasm. Judgments were given on the provided Likert scale.

2. **Part 2 (Sarcasm Identification)**: If the text contained sarcasm (i.e., the answer of part 1 was *Slightly agree, Agree,* or *Strongly agree*), participants were asked to identify where sarcasm lied within the text. This part of the task was inoperative if no sarcasm was judged in part 1 (i.e., the answer of part 1 was *Strongly disagree, Disagree, Slightly disagree,* or the middle option *Neither agree or disagree*). In this case, only part 1 was done.

For answering both parts of the judgment task, participants were instructed to use the underlining or highlighting tools built-in in Word. Figure 7 presents an illustrative example of how the main L2 pragmatics study was answered on Word.

Figure 7. A Screenshot showing a sample of how the main L2 pragmatics study was answered on Word. The purple highlight within the Likert scale represents the participant's sarcasm rating, whereas the highlight within the text indicates where the participant believes the sarcasm to be (Sarcasm Identification).

Finally, in order for the judgment task to yield trusted results, it should prove to have reliability. Reliability will be verified by means of testing the *internal consistency*[5] of the task's items. If the items prove to have internal consistency, this would mean that the instrument used to collect the data is reliable. *Cronbach's alpha*[6] will be used to test the internal consistency of the task's items (see chapter eight for the results).

7.5 Ethical Procedure

In order to meet the ethical standards of research of Lancaster University, the main L2 pragmatics study was submitted to the Research Ethics Committee at Lancaster University prior to practice. The committee reviewed the document and gave their approval to it. Regarding the information sheet and consent form, they were already submitted to and approved by the same committee in the previous experiments.

7.6 Procedure

As most participants were geographically scattered, I chose to administer the main L2 pragmatics study online. Online administration ensures easy access to participants who are just one click away from the researcher whatever physically distant they are. In addition, online-run experiments are both cost-friendly and time-friendly activities (see Dörnyei, 2010, p. 69-70). As was mentioned earlier, the judgment task, the information sheet and the consent form were sent as attachments to all participants either via email or Facebook along with a covering message. They were asked to read the information sheet and give their approval to participate by signing the consent form prior to doing the experiment. They were asked to do the task in one session but were allowed to have a short break if needed. This was to avoid, at least partly, the fatigue effect (see Dörnyei, 2010, p. 9). Participants received randomized versions of the task and were asked to time themselves while doing it. Participants were not monitored while doing the task as it was of the self-administered kind. Perhaps, this could be one disadvantage of online administration as the researcher could not run and monitor the experiment. However, it could also be an advantage as participants were doing the judgment task while being at ease and without the pressure of researcher's presence.

All participants did both parts of the task concurrently by means of the highlighting or underlining tools. Some participants used one tool for one part and the other for the other part. No problems were reported with regard to using these tools. Participants were instructed not to do part 2 (sarcasm identification) unless they respond positively to part 1 (i.e., choosing *Slightly agree, Agree,* or *Strongly agree*). This instruction was added to the general instructions at the beginning of the judgment task due to some problem emerged in the pilot study (see 6.5.1)— Some participants in the pilot L2 study contradictorily highlighted parts in some texts while

they rated sarcasm with disagreement (i.e., *Strongly disagree, Disagree, Slightly disagree*) or chose the middle option *Neither agree nor disagree*.

As the judgment task was a writable Word file, participants were asked to take extra care not to change the original wording of items. After finishing the task, the Word file was saved and emailed back to me/sent back via Facebook along with the signed consent form. Participations were completely anonymous and answers were confidential, in accordance with Lancaster University ethical standards.

Other issues to speak about here are the time needed to complete the judgment task and the problems emerged due to that some participants did not do the task properly. As for time, the three examined groups varied in the time range of completion. The English-NS group had a time range between 45-58 minutes to finish the task ($M = 52$, $SD = 4.19$ minutes). Participants of the Iraqi EFL-UK group needed between 51-66 minutes to do the task ($M = 59$, $SD = 4.6$ minutes). The time range of the Iraqi EFL-home group was 56-68 minutes ($M = 62$, $SD = 3.85$ minutes). Although the time range exceeded one hour (the planned time of task completion, see 5.4) in the learners groups, the range was still acceptable especially with having in mind that they were non-native speakers of English. Thus, they normally need extra time to read and understand in English. In fact, the amount of time exceeding one hour (6-8 minutes) was rather insignificant and does not bear very negative effect upon their cognitive ability, especially because they were able to have a short break if they needed to.

Regarding problems, four participants (1 within the English-NS group, 1 within Iraqi EFL-UK group, and 2 within Iraqi EFL-home group) provided incomplete answers. They all answered part 1 of the task and missed part 2. All those participants were approached again and asked them to do part 2. Two of them responded and did it, whereas the other two did not and were eventually replaced by new participants. Another problem was similar to the one happened in the pilot L2 study, i.e., misuse of part 2 of the task. Although instructed beforehand, two participants (1 in the Iraqi EFL-UK group and 1 in Iraqi EFL-home group) highlighted some parts within texts in an indication of where sarcasm is, whereas they had disagreed to sarcasm existence within the same texts. Their answers were excluded altogether and they were replaced by new participants as well. These are not considered big problems and challenges to the experiment as they constituted only a small group of the participants'

total (6 out of 90) and were remedied rather easily. Apart from that, the conducting of the main L2 pragmatics study went on smoothly and effectively without reporting major problems. This was a proof that the instrument (judgment task) used and the procedure followed in data collection were efficient and appropriate.

7.7 Data Analysis

This section presents the methods followed in analysing the data obtained from the main judgment task. First of all, it details what relates to part 1of the task, i.e., how the Likert scale was coded and how its scores were analysed. Afterwards it turns to discussing how the data obtained from part 2 of the task (sarcasm identification) was analysed.

7.7.1 Sarcasm rating analysis (judgment task/part 1)

The procedure followed in analysing 'Sarcasm Rating' data was similar to the one used in the pilot L2 experiment discussed earlier (see chapter six). First, Likert scale categories were coded numerically from 1 to 7 as follows:

Strongly disagree=1, Disagree=2, Slightly disagree=3, Neither agree nor disagree=4, Slightly agree=5, Agree=6, Strongly agree=7

Second, within each examined group, the responses of every participant were converted into scores according to the coding system above. Third, the *average score* was calculated for every participant within each examined group. That was the mean of the participant's individual scores in all the 30 items of the judgment task. Fourth, after calculating the average scores for all the participants within the same group, *overall score* was calculated for that group. Overall Score was the mean of all participants' average scores within the same group. In addition, standard deviation was also calculated for the average scores of each examined group to see how they were distributed around the mean (overall score). Tables 12 and 13 illustrate how average scores and overall scores were calculated.

Table 12

Calculating Average Score for Individual Participants (Fictitious Data)

Participant No.	Judgment Task Items					Mean (Average Score)
1	1	2	3	4	5	7+7+6+5+5+5+6+4+
	Strongly agree=7	Strongly agree=7	Agree= 6	Slightly agree= 5	Slightly agree= 5	6+7+6+6+7+2+1+7+
	6	7	8	9	10	6+7+7+7+3+7+6+2+
	Slightly agree= 5	Agree= 6	Neither agree nor disagree= 4	Agree= 6	Strongly agree=7	6+7+7+7+6+7= 172/30= 5.7
	11	12	13	14	15	
	Agree= 6	Agree= 6	Strongly agree=7	Disagree= 2	Strongly disagree= 1	
	16	17	18	19	20	
	Strongly agree=7	Agree= 6	Strongly agree=7	Strongly agree=7	Strongly agree=7	
	21	22	23	24	25	
	Slightly disagree = 3	Strongly agree=7	Agree= 6	Disagree= 2	Agree= 6	
	26	27	28	29	30	
	Strongly agree=7	Strongly agree=7	Strongly agree=7	Agree= 6	Strongly agree=7	

Table 13

Calculating Overall Score (Fictitious Data)

Group	Participants' Average Scores					Mean (Overall Score)	SD
1	P1*	P2	P3	P4	P5	7+7+6.2+5.7+5.2+5 +6+4.2+6.1+7+6+6+ 7+2.3+1.8+7+6+7+7 +7+3.1+7+6+2.6+6+ 7+7+7+6.1+7= 175.3/30= 5.84	1.53
	7	7	6.2	5.7	5.2		
	P6	P7	P8	P9	P10		
	5	6	4.2	6.1	7		
	P11	P12	P13	P14	P15		
	6	6	7	2.3	1.8		
	P16	P17	P18	P19	P20		
	7	6	7	7	7		
	P21	P22	P23	P24	P25		
	3.1	7	6	2.6	6		
	P26	P27	P28	P29	P30		
	7	7	7	6.1	7		

Note. P = Participant

It is worth mentioning here that Taguchi and Roever (2017) question analysing Likert-scale data on the level of individual items. On the other hand, they (2017, p.83) report J.D. Brown's (2011) position that it is only legitimate to statistically analyse the data obtained from a Likert scale on the total level. That is, the researcher can add up the rating scores of all the items in the instrument and then apply a statistical procedure to the total. Alternatively, s/he can group items according to (in)dependent variables within the instrument and apply a statistical analysis to the resultant totals. And, this is what was done in this experiment (see below).

Calculating average and overall scores as well as standard deviations for the examined groups was part of doing descriptive statistics to the data which is a necessary step for calculating inferential statistics. In addition, other calculated descriptive statistics were N (total number of participants), groups n-sizes, N's overall response and N's overall standard

deviation. Furthermore, descriptive statistics (e.g. *M, SD,* etc.) were also calculated for all the subgroups analysed. All calculations were done with SPSS.

Before applying inferential statistics to the data, the reliability of the instrument (judgment task) was tested. The internal consistency of the task items was measured with Cronbach's alpha using SPSS. The results are shown in chapter eight below.

Regarding inferential statistics, the 'Sarcasm rating' data was statistically tested on two levels to uncover any significant differences among the participants and groups. These levels were overall level and group level. On the overall level, the analysis was done for the data of all participants as a whole regardless of their examined groups. This analysis was to find out any significant differences among the 90 participants in terms of age and gender. On the group level, the analysis was applied to the data of participants as members of their examined groups. First, statistical analysis was applied to highlight any significant differences among the three examined groups in general as regards sarcasm rating. Second, the analysis was also applied to reveal any differences within and across the examined groups in terms the independent variables they had (age, gender, English language proficiency,…etc.). The latter analysis entailed dividing the examined groups into subgroups according to those independent variables.

One-way ANOVA was used to statistically test the mean-differences among the groups on both levels. In general, ANOVA is mainly used to test significant differences among three groups and more, but it can also be used to test the mean difference of two groups only instead of *t*-test and produces the same result (see Larson-Hall, 2010, p. 274). Hence, One-Way ANOVA was also used with the case of testing the mean difference of two (sub)groups as well. Furthermore, the Pearson correlation coefficient was applied to examine the relationship between any two continuous variables in the data. It was applied to the case of age and sarcasm rating on both the overall and group levels, and the case of 'length of stay in the UK' and sarcasm rating on the group level.

Being parametric tests, ANOVA and Pearson correlation tests work according to two underlying assumptions: (1) *normality of distribution* (i.e., the data is, at least roughly, normally distributed), and (2) *homogeneity of variances* (i.e., the variances of the tested

groups are equal– applying to ANOVA only) (see Larson-Hall, 2010). Hence, it was crucial to test both of these assumptions prior to applying these tests to the data. To test the normal distribution of the data, the *Shapiro-Wilk* [7] test was used which is provided by the SPSS software. In addition, I used histograms and Q-Q plots to verify the normality of distribution visually.

As for homogeneity of variances, the *Levene's test* [8] was applied to check the homogeneity of the groups' variances. In case of violating this assumption, I used *Welch's* robust *F*-statistic which is a rectified ratio of ANOVA' *F* and can deal with the problem of heterogeneity of variances (see Field, 2013, p.491-93). It was used to dis/confirming ANOVA's *F*-result.

A Significance level (α) of 0.05 was used in all the inferential statistics in this study. If ANOVA tested three groups and its *F*-statistic was significant ($p < 0.05$), a post hoc test was used to determine where exactly the significant difference was (between which groups). I used the *Tukey HSD*[9] and *Games-Howell* tests in the situations of meeting and violating the 'homogeneity of variances' assumption respectively. Finally, *effect size* was also reported in case of a significant ANOVA' statistic to assess the magnitude of the significant difference found. For ANOVA, the most frequently-used measure of effect size is *eta squared* (η^2) (see Dörnyei, 2007). The current study used Cohen's (1988, cited in Larson-Hall, 2010, p. 119) guidelines for interpreting eta squared values. These are 0.01 = *small effect*, 0.06 = *moderate effect*, and 0.14 = *large effect*.

7.7.2 Sarcasm identification analysis (judgment task/part 2)

Regarding 'Sarcasm identification', the data obtained from this part of the judgment task was analysed according to a procedure similar to the one followed in the pilot L2 study (see chapter six). The data consisted of highlighting parts within the items texts where participants believed sarcasm to be. The analysis was a two-fold procedure. First, the highlighted parts (after achieving the status of being 'sarcastic parts', see below) were compared across the examined groups to reveal matches and mismatches among the groups as regards sarcasm identification. Second, the sarcastic parts were also inspected to see which general pragmatic and pragmalinguistic characteristics were used in them and how often they are used.

The procedure of finding out mis/matches across the groups was as follows: First, within each examined group, participants responses of every item were inspected to find out how many highlighted parts were there in that item. Any highlighted part was considered a 'sarcastic part' if, at least, it was highlighted by the majority of any examined group (16 participants and more out of 30). If highlighted by less than that, the highlighted part was dropped from consideration. In this way, the sarcastic parts in all the tested items were identified. This procedure was applied to the data of all the three groups of participants. Second, sarcastic parts were compared across the groups to find out sarcasm matches and mismatches. Any sarcastic part was considered: (1) a 'Full match' if it was highlighted by all the groups, (2) a 'Partial match' if it was chosen by two groups only, and (3) a 'No match' (peculiar sarcasm) if chosen by one group only.

The general pragmatic characteristics analysed were the following: (1) Contradiction, (2) Insincerity, (3) Flouting quantity, (4) Flouting relevance, (5) Mock politeness, (6) Allusion to antecedent, (7) Negative attitude, (8) Victim, and (9) Other (i.e., the characteristics that may emerge in the data which are not or hardly ever touched upon in the literature).The verified pragmalinguistic characteristics were: (1) Positive wording (2) Hyperbole, (3) Graphological Cues, and (4) Other. The sarcastic parts in all the tested items were examined for the existence of both kinds of characteristics and their frequencies in the data were counted as well. Checking the frequencies of sarcasm characteristics could determine which ones are more prototypical to online English sarcasm and which are less. For the same purpose, characteristics were also checked in both datasets used (football dataset and parenting dataset) to find out which characteristics were used more in either dataset. All the results are shown in the next chapter.

7.8 Summary

This chapter presented the methods of conducting the main L2 pragmatics study as well as the methods of analysing the data obtained from it. Regarding the methods of conducting the study, it gave details of the ninety participants who took part in this study and how they were recruited. Then, the chapter summarizes how participants were statistically tested to reveal any significant differences among them in terms of the independent variables of age, gender and

English language proficiency and presents the results. Afterwards the chapter gives details of the materials and instrument used in the main L2 pragmatics study followed by a summary of the procedure of carrying it out.

Regarding the methods of data analysis, the 'Sarcasm rating' data and the 'Sarcasm identification' data were analysed separately. The former is analysed by means of converting the Likert scale ratings into numbers and then calculating the average scores (i.e., the mean of the participant's responses) and overall scores (i.e., the mean of the average scores of each group) for all the examined groups. The analysis of the latter depended on determining the sarcastic parts within the data and then looking at the sarcasm matching status among the groups as well as investigating the general pragmatic and pragmalinguistic characteristics in these sarcastic parts. Next chapter will presents the results of the main L2 pragmatics study.

Chapter Eight

L2 Pragmatics Study: Results and Discussions

8.1 Introduction

This chapter will report the results of the main L2 pragmatics study and provide discussions for them. First of all, it will report the descriptive and the inferential statistics of the statistical tests done for the data collected from the judgment task/part 1 (sarcasm rating) (8.2). In this regard, the chapter will present, in the beginning, what was found regarding the participants as a whole regardless of being native speakers or Iraqi EFL learners (overall results) (8.2.1). The 'Overall Results' section will report the results of gender and age as these are the only common variables the participants have regarding the current experiment. Afterwards the chapter will mention the 'Sarcasm rating' results of the three main examined groups (English-NS, Iraqi EFL-UK, Iraqi EFL-home) (8.2.2). It will mention any significant differences found among them in general (8.2.2.1). Then, the chapter will present the results of the subgroups to which the control and the experimental groups are divided (8.2.2.2). The main groups will be divided into subgroups in term of gender and L2 proficiency, and data will be compared on the within-group and between-groups levels. In addition, the result of the correlation between age and sarcasm rating will be reported for all the groups as well as the correlation between the length of stay in the UK and sarcasm rating (for the Iraqi EFL-UK group only).

Second, this chapter will also summarize the results of the second part of the judgment task, i.e., 'Sarcasm identification' (8.3). In this regard, the chapter will report, in the form of tables, the results of the following: (1) the availability of sarcasm within the tested items and its matching status among the three examined groups, (2) the availability and frequency of the general pragmatic characteristics of sarcasm found in the identified sarcastic parts, and (3) the availability and frequency of the pragmalinguistic characteristics of sarcasm within the same sarcastic parts. Finally, the chapter will provide discussions for the results of both the sarcasm-rating results (8.4.1) and the sarcasm-identification results (8.4.2). The chapter ends with a summary of the major conclusions that arise out of the results (8.5).

It is worth mentioning before reporting the results that Cronbach's alpha test was run using SPSS in order to check the internal consistency of the tested items. Results showed that the level of internal consistency is .794 (approximately 80%). This level of reliability is adequately high and validates all the inferential and descriptive statistics of the judgment task/part 1.

8.2 Sarcasm Rating

8.2.1 Overall results

As mentioned earlier, gender and age are the only common variables among all participants as regards this experiment. The sarcasm data obtained from the total of participants ($M = 4.77$, $SD = 0.72$) was statistically investigated to find out any significant differences in terms of those common variables. For this purpose, One-Way ANOVA was run to check gender differences. In addition, Pearson correlation coefficient was operated to test the relationship between age and sarcasm rating.

Gender Results

According to gender, the total of participants was divided into 'Male group' ($n = 43$, $M = 4.8$, $SD = 0.79$) and 'Female group' ($n = 47$, $M = 4.7$, $SD = 0.65$). The underlying assumptions of the One-Way ANOVA (i.e., normality of distribution and homogeneity of variances) were examined first and were met[Shapiro-Wilk test: (Male group $p = .47$, Female group $p = .62$); Levene's test $F(1, 88) = 0.98$, $p = .32$]. The null hypothesis (H_0) is that there exists no significant difference between the gender groups as regards sarcasm rating, whereas the alternative hypothesis (H_1) is that a significant difference does exist between the two groups in rating sarcasm. The results of the One-Way ANOVA are not significant and in favour of the null hypothesis [$F(1, 88) = 0.27$, $p = .61$]. Thus, we fail to reject the null hypothesis. There seems to be no significant difference between males and females in sarcasm rating.

Age and sarcasm rating (correlation results)

Unlike gender, age is a continuous variable which has no levels (groups). Thus, Pearson correlation coefficient was used to test its correlation with sarcasm rating which is another continuous variable. The null hypothesis (H_0) is that age and sarcasm rating have no relationship between one another, whereas the alternative hypothesis (H_1) is that a positive or negative relationship is available between the two variables. The underlying assumption of normality of distribution was tested beforehand and was met [Shapiro-Wilk test: Overall data (Age: $p = .08$, Sarcasm rating: $p = .40$)]. Pearson test results fail to reject the null hypothesis. They reveal no significant (positive or negative) correlation between age and sarcasm rating [$r(88) = .19$, $p = .27$].

8.2.2 Sarcasm rating among the English-NS, Iraqi EFL-UK and Iraqi EFL-home groups

In this section, we turn to mentioning the results of sarcasm rating among the three examined groups (English-NS, Iraqi EFL-UK, Iraqi EFL-home). By means of One-Way ANOVA, these main groups were tested to see whether any significant difference was available among them as regards sarcasm rating. Furthermore, differences among subgroups, whether within the main groups or across them, were also checked. Regarding age and length of stay and their relationship with sarcasm rating, Pearson correlation test was applied to examine the significance of this relationship. Prior to running the statistical tests, their underlying assumptions were tested first. Both assumptions were met [Shapiro-Wilk test: English-NS $p = .78$, Iraqi EFL-UK $p = .91$, Iraqi EFL-home $p = .34$; Levene's test: $F(2, 87) = 0.48$, $p = .62$]. Means and standard deviations were calculated to all the groups. They are shown in Table 14 below.

Table 14

Sarcasm Rating: n-size, Mean and Standard Deviation of the Main Examined Groups

Examined Group	Sarcasm Rating		
	n	*M*	*SD*
English-NS	30	4.52	0.60
Iraqi EFL-UK	30	5.08	0.71
Iraqi EFL-home	30	4.97	0.63

8.2.2.1 *Sarcasm rating across the main groups*

The null hypothesis (H_0) states that there is no statistically significant difference among the three examined groups as regards sarcasm rating. Conversely, the alternative hypothesis (H_1) proposes that at least one group is significantly different from either or both of the other groups in rating sarcasm. In fact, the null hypothesis is rejected as the One-Way ANOVA results are significant and in favour of the alternative hypothesis [$F(2, 87) = 6.19$, $p = .003$, $\eta^2 = 0.12$]. Effect size is of a moderate level according to Cohen's (1988, cited in Larson-Hall, 2010) guidelines.

In order to specify where exactly lie(s) the significant difference(s) (between which groups), a Tukey HSD multiple-comparison test (post hoc test) was performed. The Tukey test was chosen over the Games-Howell test because the 'homogeneity of variances' assumption was satisfied. Tukey's multiple-comparisons are presented in Table 15.

Table 15

Sarcasm Rating: Tukey's HSD Multiple-Comparisons of the Main Examined Groups

Tukey's Comparisons	*p*-value	95% Confidence Interval	
		Lower Bound	Upper Bound
English-NS – Iraqi EFL-UK	.004**	-0.9571	-0.1552
English-NS – Iraqi EFL-home	.023*	-0.8510	0.0501
Iraqi EFL-UK – Iraqi EFL-home	.80	-0.2953	0.5061

Note. (*) indicates a significant result at (0.05) significance level.

(**) indicates a significant result at (0.01) significance level.

Tukey's results show that both the Iraqi EFL-UK and Iraqi EFL-home were significantly different from the English-NS group. We are 95% confident that the actual difference between the means of the English-NS population and Iraqi EFL-UK population lies within this interval CI = [-0.9571, -0.1552], and the actual difference between the English-NS population and Iraqi EFL-home population lies within this interval CI = [-0.8510, 0.0501]. In fact, the difference between the English-NS group and Iraqi EFL-home group is rather expected as the latter group has no contact with English native speakers. However, the result of the English-NS group and Iraqi EFL-UK group is rather surprising. It is expected that the sarcasm rating of the Iraqi EFL-UK group to be close to that of the English-NS group due to the expected high level of interaction between the EFL learners in this group and English Native speakers. Both of the learners groups gave higher sarcasm ratings than the English-NS group. This means that the Iraqi EFL learners recognized sarcasm (or gave higher positive judgment to it) in contexts where English native speakers did not. Perhaps, part 2 of the judgment task (sarcasm identification) can uncover some rationale for why this is the case. Part 2 will provide an analysis for the general pragmatic and pragmalinguistic factors that helped participants to judge the existence of sarcasm within the given excerpts. On the other hand, Tukey results reveal no significant difference between the Iraqi EFL-UK and Iraqi EFL-home groups (CI = [-0.295, 0.506]). Thus, for the subsequent analyses, these two groups will be integrated into one single group called "Iraqi EFL learners".

8.2.2.2 *English-NS and Iraqi EFL learners: Comparing the data of gender, age,*
   L2 proficiency and length of stay in the UK

The English-NS group was divided into subgroups in terms of gender. In addition, the Iraqi EFL learners group was divided into subgroups according to gender and English language proficiency. As was the case with the main groups, subgroups were also tested for significant differences in sarcasm rating. Effects of gender upon participants' sarcasm rating were tested on the within-group level for both the English-NS group and the Iraqi EFL learners group. In addition, the effect of English language proficiency upon Iraqi EFL learners' performance was also tested on the within-group level. Afterwards the subgroups of the Iraqi EFL learners were compared to the English-NS group to find out which subgroup is closer in performance

to English native speakers. In this regard, the "Male" and "Female" subgroups of learners as well as the "Advanced" and "Intermediate" subgroups were all compared via One-Way ANOVA to the English native speakers to check differences. Results are displayed below.

For all subgroups analyses, the null hypothesis (H_0) was that subgroups were not significantly different from each other/English native speakers as regards sarcasm rating. By contrast, the alternative hypothesis (H_1) was that at least one subgroup was significantly different in rating sarcasm from the other tested subgroup/ English native speakers.

As for age and length of stay in the UK, their relationship with sarcasm rating was tested by means of Pearson correlation test. This is because they are all continuous variables. In what follows, the descriptive and inferential statistics of sarcasm rating are reported for the tested subgroups in turn.

Gender

According to gender, both the English-NS group and the Iraqi EFL learners group were divided into "Males" and "Females" subgroups. On the within-group level, differences between males and females in sarcasm rating were tested for significance by means of One-Way ANOVA. The underlying assumptions of ANOVA were checked and met for all the subgroups [Shapiro-Wilk test: English-NS-Males $p = .83$, English-NS-Females $p = .96$/ Iraqi EFL learners-Males $p = .46$, Iraqi EFL learners-Females $p = .42$; Levene's test: English-NS subgroups $F(1, 28) = 0.01$, $p = .76$, Iraqi EFL learners subgroups $F(1, 58) = 0.011$, $p = .91$]. The ANOVA results are shown in Table 16 below.

Table 16

Sarcasm Rating of Gender Subgroups (within-group Level)

Examined Group	Gender		ANOVA	
	Males subgroup (*n, M, SD*)	Females subgroup (*n, M, SD*)	*F*-value	*p*-value
English-NS	(11, 4.30, 0.62)	(19, 4.64, 0.57)	$F(1, 28) = 2.28$	.14
Iraqi EFL Learners	(32, 4.98, 0.78)	(28, 4.79, 0.72)	$F(1, 58) = 0.99$	.32

Table 16 shows no significant differences between males and females in rating sarcasm within both the English-NS group and Iraqi EFL learners group.

The Iraqi EFL learners' subgroups of gender were compared to the English-NS group to find out which subgroup is more native-like in rating sarcasm, if any. One-Way ANOVA was conducted to do this comparison between the means of the groups. ANOVA underlying assumptions were examined beforehand and were satisfied [Shapiro-Wilk test: English-NS *p* = .78, Iraqi EFL learners-Males *p* = .46, Iraqi EFL learners-Females *p* = .42; Levene's test: $F(2, 87) = 0.36$, *p* = .70]. ANOVA results indicate a significant difference between the tested groups [$F(2, 87) = 3.32, p = .04$] with a moderate effect size ($\eta^2 = 0.07$). Hence, a post hoc test was conducted to expose between which groups occurs that significant difference. As the assumption of 'homogeneity of variances' was met, the Tukey HSD test was used. Table 17 presents the results of the Tukey post hoc test.

Table 17

Sarcasm Rating: Comparing Iraqi EFL Learners' Subgroups of Gender to each other as well as to English Native Speakers

Tukey's Comparisons			*p*-value	95% Confidence Interval	
				Lower Bound	Upper Bound
English-NS (*n* = 30, *M* = 4.52, *SD* = 0.60)	–	Iraqi EFL Learners-Males (*n* = 32, *M* = 4.98, *SD* = 0.78)	.03*	-.8886	-0.0334
English-NS (*n* = 30, *M* = 4.52, *SD* = 0.60)	–	Iraqi EFL Learners-Females (*n* = 28, *M* = 4.79, *SD* = 0.72)	.32	-0.7092	0.1750
Iraqi EFL Learners-Males (*n* = 32, *M* = 4.98, *SD* = 0.78)	–	Iraqi EFL Learners-Females (*n* = 28, *M* = 4.79, *SD* = 0.72)	.54	-0.2415	0.6293

Note. (*) indicates a significant result at (0.05) significance level.

Tukey's test results capture a significant difference between male subgroup of Iraqi EFL learners and English native speakers. This indicates that male Iraqi EFL learners are far from being native-like in their rating of sarcasm. On the other hand, female Iraqi EFL learners seem to be closer to native speakers in sarcasm rating as the difference between the two groups is statistically insignificant.

Age

In 8.2.1 above, the Pearson correlation coefficient was used to test the correlation between the age of the total of participants and sarcasm rating. Similarly, the Pearson correlation coefficient was also used to test the correlation between age and sarcasm rating within each of the two main groups: English-NS group and Iraqi EFL learners group. The underlying assumption of 'normality of distribution' was examined beforehand for the Pearson correlation test and was satisfied [Shapiro-Wilk test: English-NS (Age: $p = .11$, Sarcasm rating: $p = .78$), Iraqi EFL learners (Age: $p = .25$, Sarcasm rating: $p = .30$)]. Table 18 presents the correlation test results for both groups.

Table 18

Age and Sarcasm Rating: Correlation Results within English-NS Group and Iraqi EFL Learners Group

Examined Group	Pearson Correlation Result (Age and Sarcasm Rating)	*p*-value
English-NS	$r(28) = -.15$	.44
Iraqi EFL Learners	$r(58) = .08$	.54

As shown in table 18, no significant (positive or negative) association was detected between age and sarcasm rating within the tested groups.

English language proficiency

To test the effect of English language proficiency upon sarcasm rating, the Iraqi EFL learners group was divided into 'Advanced' and 'Intermediate' subgroups. The division was according to the 'Common European Framework of Reference for Languages' system (CEFR) which, in turn, depends on the participants' IELTS scores (see appendix G). One-Way ANOVA was applied which served two purposes: (1) investigating any significant difference between the advanced and intermediate subgroups in sarcasm rating, and (2) comparing their sarcasm ratings to that of the English-NS group to find out which subgroup has a more native-like performance. ANOVA underlying assumptions were tested and met [Shapiro-Wilk test: English-NS $p = .78$, Iraqi EFL learners-Advanced $p = .09$, Iraqi EFL learners-Intermediate $p = .22$; Levene's test: $F(2, 87) = 1.12$, $p = .33$]. ANOVA results indicate a highly significant difference among the tested groups with a moderate effect size [$F(2, 87) = 5.99$, $p = .004$, $\eta^2 = 0.07$]. Tukey HSD post hoc test was performed to reveal where exactly lies this difference. Table 19 displays the results of the Tukey HSD multi-comparisons.

Table 19

Sarcasm Rating: Comparing Iraqi EFL Learners' Subgroups of English Language Proficiency to each other as well as to English Native Speakers

Tukey's Comparisons			*p*-value	95% Confidence Interval	
				Lower Bound	Upper Bound
English-NS (*n* = 30, *M* = 4.52, *SD* = 0.60)	–	Iraqi EFL Learners-Advanced (*n* = 48, *M* = 5.02, *SD* = 0.66)	.004**	-0.8574	-0.1333
English-NS (*n* = 30, *M* = 4.52, *SD* = 0.60)	–	Iraqi EFL Learners-Intermediate (*n* = 12, *M* = 5.06, *SD* = 0.75)	.047*	-1.0691	-0.0064
Iraqi EFL Learners-Advanced (*n* = 48, *M* = 5.02, *SD* = 0.66)	–	Iraqi EFL Learners-Intermediate (*n* = 12, *M* = 5.06, *SD* = 0.75)	.98	-0.5444	0.4597

Note. (*) indicates a significant result at (0.05) significance level.

 (**) indicates a significant result at (0.01) significance level.

As shown in the table above, both intermediate and advanced Iraqi EFL learners are significantly different in sarcasm rating from English native speakers. It seems that English language proficiency has no role to play in perfecting the sarcasm recognition of Iraqi EFL learners towards the native level. This hypothesis is further supported by the insignificant difference found between the intermediate and advanced learners in sarcasm rating. The advanced Iraqi EFL learners did not do any better in rating sarcasm than the intermediate ones.

Length of stay in the UK

 The length of stay of the Iraqi EFL learners in the UK was also tested for significance. The purpose was to find out whether the length of stay in the UK has any effect upon the learners' recognition of British-English sarcasm. As length of stay was a continuous variable with no levels (subgroups), Pearson correlation coefficient was applied to test the correlation between length of stay in the UK and sarcasm rating within Iraqi EFL-UK group only. The underlying

assumption of normality of distribution was tested and met [Shapiro-Wilk test: Iraqi EFL-UK (Length of stay: $p = .13$, Sarcasm rating: $p = .91$)]. Results show no significant (positive or negative) relationship between the length of stay in the UK and sarcasm rating [$r(28) = -.12$, $p = .53$]. Accordingly, it appears that the length of stay in the UK has no effect upon advancing the sarcasm rating of Iraqi EFL learners.

8.3 Sarcasm Identification

The data obtained from the judgment task/part 2 (sarcasm identification) was analysed to reveal two different things: (1) the matches and mismatches among the three examined groups as regards the sarcastic parts within the tested texts, and (2) the general pragmatic and pragmalinguistic characteristics of sarcasm which are found in those sarcastic parts and their frequencies. As for the latter, an inter-rating reliability test was conducted to verify the researcher's own judgments regarding the existence of the general pragmatic characteristics within the obtained data. Inter-raters were two British-English native speakers who are knowledgeable about pragmatics. They were given the participants' answers (sarcastic parts within texts) and were asked to identify which general pragmatic characteristics are found in each case (researcher's judgments were not shown to them) (see appendix F). Out of 85 sarcasm- general-pragmatic-characteristic judgments made by the researcher and the inter-raters, 70 cases of judgment-match were available, whereas the judgment-mismatch cases were 15 only. In the case of a judgment-mismatch, the judgment of the majority was adopted (2 out of 3). Regarding pragmalinguistic characteristics, no such a test was needed for them as they are material properties within texts which can be located without contention.

Table 20 below presents the sarcasm availability in the tested items and the matching status of the sarcastic parts among the three examined groups. Table 21 sums up the information detailed in table 20. Table 22 presents the frequencies of the general pragmatic characteristics of sarcasm found in the participants' data. It is supplemented by table 23 which shows the 'Other' general pragmatic characteristics (i.e., the characteristics which are not or hardly ever tackled in the literature) and their frequencies. Likewise, tables 24 and 25 presents the pragmalinguistic characteristics, including the 'Other' category, and their frequencies in the data (see appendix H for the source of information of tables 20-25).

Table 20

Sarcasm Identification Results: Sarcasm Availability and Sarcasm-Matching Status across the
Tested Items

Item	Sarcasm Available?	Sarcastic Part		Sarcasm-Matching Status		
				Full Match	Partial Match	No Match (Peculiar Sarcasm)
		Total	No.			
1	Yes	2	1			English-NS
			2	Yes		
2	Yes	1	1	Yes		
3	Yes	1	1		Iraqi EFL-UK and Iraqi EFL-home	
4	No					
5	Yes	1	1	Yes		
6	Yes	1	1	Yes		
7	No					
8	Yes	1	1		Iraqi EFL-UK and Iraqi EFL-home	
9	Yes	1	1	Yes		
10	Yes	1	1	Yes		
11	Yes	1	1			Iraqi EFL-home
12	Yes	1	1		Iraqi EFL-UK and Iraqi EFL-home	
13	Yes	1	1		English-NS and Iraqi EFL-UK	
14	No					
15	Yes	1	1			English-NS
16	Yes	3	1			English-NS
17	Yes	1	1	Yes		
18	No					
19	Yes	1	1	Yes		
20	Yes	1	1	Yes		
21	Yes	1	1	Yes		
22	Yes	1	1	Yes		
23	Yes	1	1	Yes		

Item	Sarcasm Available?	Sarcastic Part		Sarcasm-Matching Status		
		Total	No.	Full Match	Partial Match	No Match (Peculiar Sarcasm)
24	No					
25	Yes	1	1	Yes		
26	Yes	1	1			English-NS
27	Yes	1	1	Yes		
28	Yes	1	1	Yes		
29	Yes	1	1			Iraqi EFL-UK
30	No					

Table 21

Sarcasm Identification Results: Summing up the Information Detailed in Table 20

	Tested Items		Sarcastic Parts within the Tested Items	Matching Status		
	Items with Sarcasm	Items with No Sarcasm		Full Match	Partial Match	No Match (Peculiar Sarcasm)
Total	24	6	27	17	4	6

Table 22

Sarcasm Identification Results: Frequencies of the General Pragmatic Characteristics in the Sarcastic Parts

Forum	General Pragmatic Characteristics							
	Allusion to antecedent	Contradiction	Flouting Quantity	Flouting Relevance	Insincerity (flouting Quality)	Mock politeness	Negative attitude	Victim
Football	13	6	1	0	11	1	10	11
Parenting	13	6	0	2	12	3	10	12
Total	26	12	1	2	23	4	20	23

Table 23

Sarcasm Identification Results: Frequencies of the Other General Pragmatic Characteristics in the Sarcastic Parts (Supplement to Table 22)

Forum	Other General Pragmatic Characteristics[1]			
	Elaboration on previous remark	Friendliness	Formula Mismatch	Metaphor
Football	1	3	1	1
Parenting	0	2	0	0
Total	1	5	1	1

1. These are the characteristics found in the data which are hardly tackled in the literature.

Table 24

Sarcasm Identification Results: Frequencies of the Pragmalinguistic Characteristics in the Sarcastic Parts

Forum	Pragmalinguistic characteristics					
	Positive wording	Hyperbole	Graphological cues			
			Capitalization	Emoticon	Laughing marker	Exclamation mark
Football	3	3	8	0	3	4
Parenting	3	4	1	4	0	3
Total	6	7	9	4	3	7

Table 25

Sarcasm Identification Results: Frequencies of the Other Pragmalinguistic Characteristics in the Sarcastic Parts (Supplement to Table 24)

Forum	Other Pragmalinguistic Characteristics[1]						
	Attention getter	Boldface	Interjection	Italics	Question mark	Quotation marks	Rhetorical question
Football	1	0	4	0	0	0	0
Parenting	0	1	2	1	2	4	6
Total	1	1	6	1	2	4	6

1. These are the characteristics found in the data which are hardly tackled in the literature.

8.4 Discussion

The results of the judgment task summarized above indicated some similarities and differences between English native speakers and Iraqi EFL learners. In this section, we shall discuss the results of each part of the judgment task in turn.

8.4.1 Sarcasm rating

On the overall level, no statistically significant difference was found between the gender or age of participants and sarcasm rating. This means that both gender and age have no effect on how participants rate online sarcasm. On the group level, results show that both Iraqi EFL-UK and Iraqi EFL-home groups can recognize written English sarcasm, but they are significantly different from native speakers in this regard. This indicates no effect for studying abroad on perfecting the Iraqi learners' recognition of online English sarcasm. It could be, as Bouton (1999) believes, that sarcasm is a highly culture-specific pragmatic phenomenon which learners cannot have a perfect command on even after spending a long period of time in the target language community. Another reason for this difference is that English sarcasm may not be used online as prolifically as speech acts (e.g. stating and requesting), thus Iraqi learners are not quite familiar with it. A further reason relates to the intensity of interaction. In their feedback of the experiment, most of the Iraqi learners in the UK mentioned that they did not have intense contact with English native speakers outside their academic environment. Most of them reported that they live within Arab and Muslim communities and they use their L1 in everyday life most of the time. In fact, one limitation of the current study is the lack of measuring the intensity of communication with native speakers.

As for English language proficiency, both intermediate and advanced learners were significantly different in their sarcasm recognition from native speakers. This also denotes no effect for L2 proficiency upon sarcasm recognition. The advanced learners did not do any better than the intermediate in rating sarcasm. It seems that sarcasm is a complex phenomenon the recognition of which is not dependent on proficiency level. This result comes contrary to what Shively et al. (2008) found, i.e., that irony recognition improves as the level of proficiency gets higher. In fact, I believe that the result I arrived at is somewhat justified because most of the measures the researchers adopt for gauging language proficiency are

language tests (e.g. IELTS and TOEFL). Such tests measure mainly the grammatical and phonological abilities of examinees. Although they touch upon the examinee's L2 pragmatic comprehension and production in their listening and speaking parts respectively, the span given to L2 pragmatics is still not that much in these tests. I believe it is high time pragmaticists developed a separate, efficient and recognized proficiency test for L2 pragmatics. Regarding sarcasm recognition, perhaps what matters more in polishing it is the kind and/or amount of L2 input the learner is exposed to.

Regarding age, it showed no significant influence on sarcasm rating in all the three examined groups. There seems to be no relationship between how old the native speaker/Iraqi learner is and how s/he would rate sarcasm online. No previous study on L2 irony or sarcasm has, to my knowledge, investigated the effect of age on sarcasm recognition so that we can compare the current results to it. And, most of the other L2 pragmatics studies have not, generally speaking, concentrated on the effect of age on the constructs they investigated.

Length of stay in the UK, which is examined within the Iraqi EFL-UK group only, seems to have no impact on the learners' rating of online English sarcasm. No significant correlation was detected between sarcasm rating and the Iraqi learners' length of stay in the UK. This latter result contradicts with what Bouton (1999) found, i.e., that study-abroad L2 learners showed improvement in irony perception the longer they stayed abroad. However, it could be that the smaller sample size used in my study (30 participants in comparison with 375 L2 learners in Bouton, 1999) or its cross-sectional nature (in comparison to the longitudinal nature of Bouton, 1999) is what yielded this result.

As far as gender is concerned, no significant difference was observed between Iraqi male and female learners. However, male Iraqi learners were found to rate online sarcasm significantly differently from English native speakers. No such a difference was detected with female Iraqi learners. This result indicates that Iraqi female learners are closer in their sarcasm ratings to native speakers than male learners are. It could be that the female learners who participated in this study have more intense interaction with native speaker or are more familiar with English sarcasm than Iraqi male learners.

8.4.2 Sarcasm identification

Tables 20 and 21 show that out of the 30 items of the judgment task, sarcasm was identified in 24 items and not identified in 6 only. In the 24 sarcasm-containing items, 27 instances of sarcasm were found. Seventeen sarcasm instances achieved consensus (i.e., they were identified by the En-NS group, Iraqi EFL-UK group, and Iraqi EFL-home group). This means that the Iraqi learners can recognize written English sarcasm and that they have agreed in more than half of the cases with the native speakers on identifying sarcasm within the items. This is a rather good proportion of agreement, but it does not indicate that learners have reached the native-level of sarcasm comprehension, even those who are studying abroad (the latter group has agreed in one further item with the native speakers making the total of agreement 18 only). This outcome favours Bouton's (1999) conclusion that even after spending a long time in the target language community, L2 learners cannot reach the native-speaker level of irony comprehension. In addition to the 17 consensus instances, the two learners' groups agreed on 3 further instances. This makes the total of the agreed upon instances of sarcasm between the learners' groups 20 out of 27 instances. This indicates a rather small difference between the learners' groups in sarcasm recognition (7 instances only out of 27) which, in turn, supports the finding that study-abroad learners have no better sarcasm recognition than the home learners. As for the 'peculiar sarcasm' instances, they are only 6 in number (4 for the En-NS group, 1 for the Iraqi EFL-UK group, 1 for the Iraqi EFL-home group). This is rather encouraging as it indicates the scarcity of peculiarity cases of sarcasm recognition in all the examined groups. Regarding L2 proficiency, the sarcastic parts of the intermediate and advanced groups were compared to each other and to those of the native speakers. The results were close to those in table 21 (Full match= 15, Partial match= 5, No match (peculiar sarcasm) = 7) (see appendix H). Likewise, the matching status of the sarcastic parts was also checked for the gender groups in comparison with the native speakers. Similar results were found as well (Full match= 16, Partial match= 5, No match (peculiar sarcasm) = 6) (see appendix H).

General pragmatic characteristics seem to be more important for sarcasm recognition than the pragmalinguistic ones. This conclusion is sustained by two observations: (1) the frequency of general pragmatic characteristics in the data is generally higher than that of pragmalinguistic characteristics (see tables 22 and 24), (2) sarcasm was identified in some items of the judgment

task which contain general pragmatic characteristics and are devoid of or containing few pragmalinguistic characteristics only (see items 5, 6, 9, 15, 17, 26, 27 and 28 in appendix H). Furthermore, another observation the results provide is that the more general pragmatic and pragmalinguistic characteristics are available the more comprehensible the sarcasm is by both native speakers and Iraqi learners and vice versa [e.g. see items 1, 5, 16 (for more characteristics) and 3&29 (for less characteristics) in appendix H].

Within general pragmatic characteristics, table 22 shows that 'Allusion to antecedent' is the most frequent sarcasm characteristics employed in the data (26 times in a total of 27 sarcastic parts). This result denotes the prototypicality of this characteristic in online English sarcasm and comes in favour of Wilson and Sperber' (1992, 2012) account of irony which concentrates on 'Allusion to antecedent' as the major property of irony (see 2.4.3). 'Insincerity' and 'Victim' also prove to be further prototypical characteristics of online English sarcasm relying on their high frequency of occurrence (23 times each). The results support the Gricean account of irony, which focuses on insincerity in irony (see 2.4.2), and any other victim-involving account of sarcasm [e.g. Leech (1983) and Culpeper (1996) (see 2.4.5 and 2.5, victim involvement)]. Negative attitude can also be another prototypical characteristic of online English sarcasm as it has achieved 20 times of occurrence in the data. This also sustains any account of irony/sarcasm involving negative attitude [e.g. Colston (1997) and Toplak and Katz (2000) (see 2.5 above, Negative attitude)]. The low occurrence of contradiction does not qualify it to be a central prototypical feature of online English sarcasm. This result supports the criticism made by many researchers (e.g. Wilson and Sperber, 1992 and 2012) against the traditional account of irony which considers the contradiction between the literal meaning and the intended meaning as the main property of irony. However, the current study does not sustain the extreme version of this criticism as the frequency of contradiction is not that bad (12 out of 27). Contradiction was used in assertive sarcasm only. It was not used in interrogative sarcasm at all. As for 'Mock politeness', the very low score of occurrence (4 out of 27) does relegate it to a peripheral rather than a prototypical characteristic of online English sarcasm. This result does not come in favour of Leech's (1983, 2014) and Culpeper's (1996) accounts of sarcasm which both focus on mock politeness as the principal property of sarcasm. Flouting quantity and flouting relevance seem to be two further peripheral features of online sarcasm due to their scarcity in the data (see table 22). Table 22 also reveals no noteworthy difference between the assumed males (football

dataset) and the assumed females (parenting dataset) in employing any of the general pragmatic characteristics. This observation indicates the generality and non-gender-specificity of these characteristics.

As for pragmalinguistic characteristics, tables 24 and 25 show that 'Capitalization' is the most frequent characteristics (9 out of 30). Although the number is not very big, it still indicates that capitalization is more prototypical than all the other pragmalinguistic characteristics of sarcasm as regards the modality tested (i.e., written online English sarcasm). 'Hyperbole' and 'Exclamation mark' (scored 7 each out of 30) also seem to be further two prototypical pragmalinguistic characteristic. Less prototypical pragmalinguistic characteristics are 'Positive wording', 'Interjection', and 'Rhetorical question' which all scored 6 out of 30. The rest seem to be peripheral rather than central prototypical pragmalinguistic characteristics due to the low frequencies they have scored.

The results of 'Sarcasm identification' yielded three further noteworthy observations with regard to pragmalinguistic characteristics of sarcasm. First, Iraqi EFL learners seem to be more sensitive to graphological cues than native speakers. They identified sarcasm where native speakers did not on seeing these cues. For example, on seeing capitalization and exclamation marks, Iraqi learners identified sarcasm in items (3, 8 and 11, see appendix H) wherein native speakers did not detect any. A similar case occurs in item (12) on seeing capitalization and a laughing marker and a further case is in item (29) on seeing an emoticon. This observation accounts for, at least in part, the difference found between Iraqi learners' and native speakers in rating sarcasm (see the end of section 8.2.2.1). Second, a further observation about graphological cues is that the assumed males (football dataset) use more capitalization and laughing markers in their online sarcasm than the assumed females (parenting dataset) and that females use more emoticons than males (see table 24 above). It is not clear why this is the case and I found no previous literature which can provide an explanation for this observation. Third, it was also noticed that some identified sarcasm took the form of rhetorical questions (see items 17, 19, 21, 22, 25, 27 in appendix H).

Finally, it is also worth noting that the identified sarcasm in this study was not always nasty and conveying negative attitude. Sometimes, sarcasm was used for fun in a friendly way or to convey a positive emotion. The friendliness was indicated by using a laughing marker such as

'LOL' (e.g. see item 2 in appendix H) or a smiley (e.g. see item 20 in appendix H). This finding, which is based on first-order observation, is in contradiction to any account of irony/sarcasm that necessitates inevitable conveyance of negative attitude by sarcasm (e.g. Leech, 2014; Culpeper (1996); Colston,1997; Toplak and Katz, 2000). On the other hand, the finding coincides with Kim's (2014) finding that "positive emotions can trigger a speaker to yield light-hearted sarcasm in a friendly way" (p. 1).

8.5 Conclusion

In this chapter, the results of the main L2 pragmatics study was reported and discussed. First, it reported the results of the first part of the judgment task (Sarcasm Rating). In this regard, the chapter reported the overall results of the whole participants as regards gender and age vis-à-vis sarcasm rating. The results indicate no significant difference or correlation between the compared variables. The chapter also reported the sarcasm-rating results of the three examined groups. Both Iraqi EFL-UK and Iraqi EFL-home groups were found to be significantly different from the native speakers, the matter which uncovers no effect for studying abroad on perfecting the Iraqi learners' recognition of online English sarcasm. The results of the main examined groups also revealed no effect for L2 proficiency, age and length of stay in the UK upon sarcasm rating. The only significant difference found is related to gender: Male Iraqi EFL learners were different from native speakers in rating sarcasm, whereas female learners were not.

Second, this chapter also reported the results of the second part of the judgment task (sarcasm identification). Sarcasm was identified in 24 items out of 30. A total of 27 sarcastic parts was identified in these 24 items. Seventeen sarcastic parts from this total achieved consensus by all the examined groups which does not indicate that the Iraqi learners have reached the native level in comprehending L2 online sarcasm (close results were found regarding the L2 proficiency groups and gender groups). Regarding sarcasm characteristics, the chapter reported the frequencies of all general pragmatic and pragmalinguistic characteristics found in the data. According to the results, general pragmatic characteristics seem to be more important than pragmalinguistic ones for sarcasm comprehension for all participants (learners and native speakers). Within general pragmatic characteristics, 'Allusion to antecedent',

'Insincerity', 'Victim' and 'Negative attitude' are all more frequent than others and appear to be candidates for being central prototypical features of online English sarcasm. As for pragmalinguistic characteristics, 'Capitalization' seems to be the most central prototypical characteristic followed by 'Hyperbole' and 'Exclamation mark'. Less prototypical characteristics are 'Positive wording', 'Interjection', and 'Rhetorical question'. Iraqi learners seem to be more sensitive to those pragmalinguistic characteristics (namely to graphological cues) than native speakers, i.e., they tend to identify sarcasm more than native speakers on noticing these cues within texts. The final noteworthy conclusion is that some of the identified sarcasm within the data was used in a friendly way to convey a positive emotion rather than a negative attitude.

Chapter Nine

Conclusions

9.1 Introduction

The first section in this chapter is devoted to answering the research questions in the light of the results of the two studies done in this thesis (corpus study and L2 pragmatics study) and the literature review done in chapters two and three. Some of the findings of this thesis are integrated with the answers of research questions, whereas the rest are listed under 'Other observations' at the end of the section. The next section (9.3) outlines the theoretical and methodological contributions the thesis has made to the field of L2 pragmatics in general and L2 sarcasm in particular. This section also summarizes the pedagogical implications the study has for teaching L2 sarcasm. The limitations of the study are listed in (9.4) and some directions for future research are provided in (9.5). The chapter ends with concluding remarks which contain pieces of advice to L2 researchers (9.6).

9.2 Research Questions: Answers and Other Observations

1. Can Iraqi L2 learners of English recognize written sarcasm in British English?

Results of the L2 pragmatics study in this thesis provide evidence for a positive answer to this question from the performance of the sampled Iraqi EFL learners. They were able to identify and rate sarcasm in most of the cases (see the results in 8.2 and 8.3). This confirms the hypothesis given for this research question in 3.8. This positive answer is also in line with the results of L2 irony/sarcasm studies reviewed in 3.5. All those studies indicate that L2 learners can recognize irony/sarcasm in the target language.

2. If so, how does Iraqi L2 learners' ability to recognize written sarcasm compare to that of native speakers of English?

Results of the L2 pragmatics study reveal a significant difference between Iraqi EFL learners and English native speakers in rating sarcasm (see 8.2). In addition, the 'Sarcasm

identification' results show that the Iraqi EFL learners achieved consensus (full match) with the native speakers in only 17 sarcastic parts out of a total of 27 (see 8.3). All these results affirm that Iraqi EFL learners have not reached the native level in sarcasm recognition, even those who have been studying abroad in the UK for a long period of time (length of stay results show no significant difference between learners who has spent one year of study abroad and those who have spent four years, see 8.2.2.2). This finding confirms what was expected in 3.8 which is based on Bouton's (1999) finding, i.e., that L2 learners of English could not reach the native level of irony perception even after spending 54 months in an English-speaking environment. It could be the case that English sarcasm is more culture-specific than other pragmatic aspects. Thus, learners need to have a very good mastery of L2 English cultural norms before they can comprehend all the sarcasm they experience in the target language. Taguchi (2011) refers to this likelihood when she says "some formulaic implicatures that draw on culture-specific knowledge (e.g., Pope question and irony) remain difficult to comprehend" (p. 916).

3. What factors influence Iraqi L2 learners' ability to recognize written sarcasm (age, gender, L2 proficiency, study abroad)?

It was expected that study abroad experience and L2 proficiency would have an effect upon the sarcasm recognition of Iraqi EFL learners (see 3.8). However, the 'Sarcasm rating' results affirm a significant difference between the Iraqi EFL-UK group and EN-NS group, whereas no such a difference is found between the Iraqi EFL-UK group and the Iraqi EFL-home group. Moreover, the results of 'Sarcasm identification' also count slightly more matches between the Iraqi EFL-UK group and the Iraqi EFL-home group (20 matches) than between the Iraqi EFL-UK group and native speakers (18 matches) (see 8.3). All these results suggest no effect of study abroad on the sarcasm recognition of Iraqi learners. This finding runs contrary to the findings of many study-abroad studies (e.g. Schauer, 2009 and Taguchi, 2011), which indicate an advantage of studying abroad on developing learners' L2 pragmatic competence. Perhaps, what matters most is the intensity of L2 interaction and the amount and kind of L2 input the learners are exposed to, as Schauer (2010) and Bardovi-Harlig and Bastos (2011) have pointed out. In fact, as was mentioned in 8.4.1, most of the Iraqi EFL-group reported that during their stay in the UK they lived within Arab and Muslim

communities. They use their L1 in everyday life most of the time, and reserve using L2 English for their academic environment only. This reality can account, at least in part, for the passive effect of their study abroad experience on sarcasm recognition.

Regarding L2 proficiency, results of sarcasm rating reveal a significant difference between both intermediate and advanced learners and native speakers. In addition, the same results reveal no significant difference between the intermediate and advanced learners. That is to say, the advanced learners were not better than the intermediate in sarcasm recognition. Thus, it seems that L2 proficiency has no effective role in advancing the sarcasm recognition of Iraqi EFL learners. This finding counters what was expected in 3.8, and counters the finding of Shively et al. (2008) which attests that irony recognition improves as the level of L2 proficiency increases. It could be the case that the advanced learners in this study were not 'advanced' enough in the pragmatic sense of the word. Perhaps, to become pragmatically proficient, the learner needs to spend a considerable time in the L2 community and gets immersed in its culture. The measure used for deciding the level of L2 proficiency in this study was the IELT or TOEFL score. These and similar English tests measure the four skills (listening, speaking, reading and writing) as well as the grammar and vocabulary of English as a target language. They do not measure learners' L2 pragmatic competence as they are not designed for this purpose. Thus, using the scores of such language tests to determine L2proficiency status is very tricky in L2 pragmatics studies. As I stressed earlier in 8.4.1, it is high time for pragmaticists to develop a separate, efficient and recognized proficiency test for L2 pragmatics.

As for age, it also seems to have no influence on the sarcasm recognition of the learners in question as no significant correlation is found between age and sarcasm rating. Finally, gender results of sarcasm rating uncover a significant difference between male learners and native speakers, whereas no such a difference is found between female learners and native speakers. This means that female Iraqi EFL learners, for some reason, recognize written English sarcasm as well as the native speakers, whereas male Iraqi EFL learners are far from the target-level of sarcasm recognition. This finding is not found in any of the previous L2 irony/sarcasm studies (see 3.5).

4. **What are the general pragmatic and pragmalinguistic characteristics of sarcasm that English native speakers and Iraqi L2 learners of English draw on in the process of recognizing written sarcasm? Which characteristics are more prototypical and which are less?**

All the general pragmatic characteristics that are extracted from irony/sarcasm definitions and accounts (discussed in 2.5) are found in written online English sarcasm. These are 'Allusion to antecedent', 'Contradiction', 'Flouting Quantity', 'Flouting Relevance' 'Insincerity', 'Mock politeness', 'Negative attitude' and 'Victim'. In addition to these, other general pragmatic features are found in the analysed data of sarcasm which are hardly ever touched upon in the literature. They also seem to be less typical for English sarcasm as they scored very low in terms of frequency of occurrence (see Tables 2and 23). These are 'Elaboration on previous remark', 'Friendliness', 'Formula Mismatch', 'Metaphor', 'Mocking a previous remark', 'Simile' and 'Sincerity'. As for pragmalinguistic characteristics, all the pragmalinguistic characteristics derived from irony/sarcasm accounts and discussed in the 2.6 are found in the collected data. These are 'Hyperbole', 'Positive wording' and 'Graphological cues' (Capitalization, Emoticon, Exclamation mark, Laughing marker). There are other pragmalinguistic features which are encountered in the data analysed that are not or hardly dealt with in the literature. These are 'Attention getter', 'Interjection', 'Rhetorical question', 'Structure repetition' and 'Other graphological cues' (Boldface, Italics, Question mark, Quotation mark, Vowel elongation). The native speakers and the L2 learners draw on the above general pragmatic and pragmalinguistic characteristics to comprehend online English sarcasm.

Regarding which characteristics are more prototypical and which are less, the results of the corpus study and the L2 pragmatics study indicate that 'Insincerity' is the most prototypical (most central) general pragmatic characteristic of written English sarcasm followed by 'Victim' and 'Allusion to antecedent'. What supports the centrality of 'Insincerity' to sarcasm recognition appeared in the corpus study results (see 4.4.1), namely, that some instances of sarcasm within the collected data have 'Insincerity' as the only general pragmatic characteristic available. This finding favours Grice's (1975, 1989) account of irony, Wilson and Sperber's (1992, 2012) account of irony and any account or definition of irony/sarcasm that involves victim (see 2.5, Victim involvement). In a recent work dealing with flouting

Quality in interaction, Vergis (2017) found out that flouting Quality (i.e., insincerity) is crucial for crystallizing banter which is akin pragmatic phenomenon to sarcasm. This may provide partial support for the 'Insincerity' finding of this study. Results also show that 'Negative attitude' is a less prototypical (less central) general pragmatic characteristic of sarcasm and 'Contradiction' and ' Mock politeness' are even lesser. This latter finding does not provide strong support to the traditional account and (mock)politeness account of irony/sarcasm. The rest of the general pragmatic characteristics are more peripheral characteristics or the least prototypical characteristics of sarcasm. The above finding is not entirely in line with what was expected (see 3.8), i.e., that 'Negative attitude' and 'Victim' would be the most prototypical characteristics of sarcasm.

As for pragmalinguistic characteristics, the results of the corpus study show that 'Hyperbole' is the most prototypical pragmalinguistic characteristics followed by 'Positive wording' and 'Exclamation mark' respectively. Less prototypical ones are 'interjection', 'Emoticon' and 'Capitalization'. However, the results of the L2 pragmatics study show that 'Capitalization' is the most prototypical pragmalinguistic characteristics followed by 'Hyperbole' and 'Exclamation mark' (both have the same frequency). Less prototypical characteristics are 'Positive wording', 'Rhetorical question' and 'Interjection' (All have the same frequency). By combining the outcomes of both studies, 'Hyperbole' could be the most prototypical pragmalinguistic characteristic of written English sarcasm followed by 'Positive wording' and 'Exclamation mark'. Less prototypical characteristics are 'Capitalization', 'Interjection' and 'Emoticon'. The rest seem to be peripheral rather than central characteristics. This finding confirms what was expected in 3.8 with regard to hyperbole. It supports all the accounts that highlight the role of hyperbole in expressing irony/sarcasm (see 2.6, Hyperbole). With regard to graphological cues, the expectation was to find that most of the graphological cues are among the prototypical characteristics, whereas only 'Exclamation mark' is found to be a rather central prototypical characteristic. Some other graphological cues are less prototypical characteristics ('Capitalization' and 'Emoticon'), whereas the rest of them are more peripheral ones.

Other observations

In addition to the findings enclosed in the answers of the research questions above, what follows are some other findings arrived at through the results of the corpus study and the L2 pragmatics study.

1. No significant (positive or negative) correlation was found between age and sarcasm recognition of English native speakers.

2. The prototypical general pragmatic and pragmalinguistic characteristics the results have yielded validate the prototypical definition proposed for sarcasm in 2.7.

3. General pragmatic characteristics of sarcasm have a more fundamental role in the recognition of written English sarcasm than pragmalinguistic characteristics. Support for this finding comes from cases in the data which contained no or few pragmalinguistic characteristics. Nevertheless, they were recognized as sarcastic relying on the general pragmatic characteristics they have only (e.g. see example 21 in 4.5).

4. There are discrepancies between the assumed English males and females in employing general pragmatic and pragmalinguistic characteristics in their written sarcasm (see Tables 2 and 3).

5. Generally speaking, the more general pragmatic and pragmalinguistic characteristics are available in any sarcastic part the more perceptible it would be and the higher rating it scores among participants (native speakers and learners) (see appendix H). In other words, sarcasm becomes easier to comprehend when it contains more general pragmatic and pragmalinguistic characteristics.

6. Iraqi EFL learners seem to be more sensitive to graphological cues than native speakers. They identified sarcasm where native speakers did not on seeing these cues. It seems that the learners developed associative indication between these cues and sarcasm.

7. The identified sarcasm in this study was not always nasty and conveying negative attitude. Sometimes, sarcasm was used for fun in a friendly way or to convey a positive emotion (e.g. see items 2 and 20 in appendix H).

8. Some first-order sarcasm in this study takes the form of a rhetorical question (e.g. see items 27 in appendix H). This finding is not apparent in the literature.

9.3 Contributions and Implications of the Study

The current study contributes to and extends the current knowledge and theory of (L2) pragmatics in several ways. Furthermore, the study also has some contributions to the current methods used in L2 pragmatics studies. These contributions are summarized below.

Theoretical contribution

The current study contributes to making up for the research shortage in L2 impoliteness in general and L2 sarcasm in particular by investigating the comprehension of English sarcasm by L2 learners. It is one of a few L2 pragmatics studies dealing with the comprehension rather than production of an L2 construct. The study contributes to increasing the body of literature on L2 comprehension and redressing the balance with L2 production literature. This study is also, to my knowledge, the first to investigate online English sarcasm as an L2 pragmatic phenomenon, adopts a gender balance in the data collected from the internet and analyse gender difference in the investigated sarcasm. All the previous studies have tackled online English sarcasm within the domain of L1 and the L2 irony/sarcasm studies did not do any analysis for gender difference (see 2.8 and 3.5).

This study is also innovative in investigating English sarcasm as it is naturally used by native speakers and see how L2 learners of English recognize it with all its complexities. In this sense, it is the first to study L2 English sarcasm as a first-order construct wherein the focus is on how the laypersons native speakers use sarcasm. The study judges the different scholarly accounts of irony/sarcasm according to this use. The study provides empirical support for some of these accounts and invalidate others. This is also the first L2 sarcasm study that adopts a prototypical view rather than a sufficient-and-necessary-condition view

when dealing with sarcasm. It is the first to extract the general pragmatic and pragmalinguistic characteristics of sarcasm from the different accounts and definitions of irony/sarcasm and investigate which are more prototypical and which are less.

It is also, to my knowledge, the first study to find out that 'Insincerity' is the most prototypical general pragmatic characteristics of sarcasm and favours Grice's account of irony based on evidence from naturally-occurring data. Perhaps, the only parallel finding comes from Vergis (2017) who discovered that flouting Quality, which results in insincerity, is a crucial act for materializing banter which is akin pragmatic phenomenon to sarcasm.

Finally, the current study is one of the few L2 pragmatics studies that involves Arab L2 learners of English and the first, to my knowledge, to investigate Iraqi learners. By so doing, it attempts to turn the spotlight rather away from the Japanese, Chinese and European L2 learners of English towards Arab learners who are highly under-researched in L2 pragmatics studies.

Methodological contribution

The current study seems to be the first to introduce a metalanguage strategy as a data collection procedure to L2 irony/sarcasm studies. The strategy proved to be effective in locating the target data within the investigated corpora without much effort. The two-fold judgment task used in this study with 'Sarcasm rating' part and 'Sarcasm identification' part is also something new to L2 irony/sarcasm studies. Combining these two parts together in a single judgment task adds more precision to the measuring capability of that task. Two participants may give the same rating to the sarcasm used in the same tested item, but they may differ in identifying where sarcasm is within the text of that item. This reveals that although they give the same rating in the same item, they rate different things. In addition, being a two-fold task enabled it to be versatile and serve different purposes at the same time. The 'Sarcasm rating' part of this judgment task served to reveal the degree of rating, whereas the 'Sarcasm identification' part helped us know the matching status of the recognized sarcasm among the examined groups and revealed which general pragmatic and pragmalinguistic characteristics are used and their frequencies.

The stimuli selection task, which is used for selecting final stimuli for the L2 pragmatics study, is another contribution for this study. According to its results, the task systematically selected the required number of stimuli for the L2 experiment depending on the score of every average category (see table 10). In other words, all the average categories are represented in the final selection of stimuli according to their proportions or scores in the stimuli selection task. This is to prevent the overrepresentation of some average categories at the expense of others which may, in turn, distort the results of the L2 study. This technique was not used before, to my knowledge, in any L2 irony/sarcasm study or even in any L2 pragmatics study.

The main L2 pragmatics study was conducted online and self-administered by the participants themselves. This procedure proved to be effective in the case of distant participants. Participants who are scattered over different regions or different countries cannot be reached physically easily. Thus, conducting the study online is an ideal solution in such a case. This is the first L2 sarcasm study to be conducted online using Facebook and email communication. This procedure saves time, effort and money if participants are not easy to reach.

Pedagogical implications

The difference found in sarcasm recognition (sarcasm rating and sarcasm identification) between Iraqi learners and native speakers attracts attention to a gap in teaching L2 irony and sarcasm. Teachers of pragmatics can utilize the discussed general pragmatic and pragmalinguistic characteristics of sarcasm especially the more prototypical ones when giving explicit instructions to learners regarding L2 irony/sarcasm. Furthermore, they can make use of the judgment task used in the current L2 pragmatics study for some pedagogical purposes (e.g. designing a 'conscious raising task'). For example, teachers can prepare some texts and ask some native speakers to identify where sarcasm is within them (e.g. via highlighting). Then, they give the same texts to their L2 learners and ask them to identify sarcasm as well. Afterwards the teachers disclose the native speakers' highlights to the learners for comparison. Finally, they comment on the learners' errors in the light of mainly the characteristics of sarcasm.

9.4 Limitations of the study

The major limitation is that the study did not measure the 'intensity of interaction' of the learners with the native speakers. It did not seek to know what kind of L2 input the learners are exposed to and for how long. Measuring that could have yielded some more interesting results. The second limitation relates to the human source of the online data. With this kind of data, one cannot tell for sure that all the bloggers who interact with each other on the forums are native speakers of English due to the anonymity of their identities. Furthermore, it is also not known whether the forums' users are from all the regions of the UK so that we can assume that the sarcasm produced provides a true picture for the 'British-English' sarcasm. The third limitation is that the collected data could not cover all the possible cases of sarcasm in British English due to it being controlled by website moderators. The last limitation relates to the metalanguage strategy used for collecting the required data. Although this strategy was successful in locating and extracting enough sarcastic instances for the study, it misfired the required purpose sometimes. Not all the returned hits (threads) proved to include sarcasm, meaning that manual screening of the hits was necessary. The occurrence of the search terms *sarcasm* and *sarcastic* was, sometimes, for purposes other than indicating sarcasm such as denying sarcasm (e.g. *I'm not being sarcastic here*) or describing friends and family members as being sarcastic in their everyday life (e.g. *My husband is very sarcastic*).

9.5 Directions for future research

L2 impoliteness is still a rather virgin area of research which needs further exploration. Researchers are highly recommended to investigate impoliteness aspects such as rudeness, patronizing, pejorative expressions and impoliteness speech acts (such as insulting) within the domain of L2 learning. L2 irony/sarcasm also needs more attention as the number of studies dealing with it is still far too few. It would be interesting to investigate L2 irony/sarcasm in the spoken modality and consider prosody and kinesics in the investigation by means of employing some audiovisual instruments. The production of irony/sarcasm by L2 learners would also be a further interesting subject. Furthermore, more in-depth studies are needed to inspect the general pragmatic and pragmalinguistic characteristics of sarcasm in terms of their number, availability and prototypicality. Gender differences found in this study which relate

to the employment of higher densities of characteristics by one gender or another are also worthy of further investigation.

9.6 Concluding Remarks

The current study has expanded the existing knowledge of L2 sarcasm which is already underresearched in L2 pragmatics. Researchers are encouraged to complete the way this study has started through scrutinizing the L2 sarcasm phenomenon more, zooming in on the current sarcasm characteristics and shedding light on any extra properties that may come out. They are also encouraged to adopt the properties-investigation procedure when tackling any pragmatic phenomenon. This dismantling procedure would help a lot in providing a better and comprehensive understanding of the construct under investigation. The study also highlights the importance of naturally-occurring data in L2 pragmatics research as it reflects the reality better, and asserts that this kind of data should be the main source of any new theorizing or any updating of an existing theory.

References

Attardo, S. (2000). Irony as relevant inappropriateness. *Journal of Pragmatics*, *32*, 793-826.

Ball, D. W. (1965). Sarcasm as sociation: The rhetoric of interaction. *Canadian Review of Sociology and Anthropology*, *2*,190-98.

Barbe, K. (1995). *Irony in context*. Amsterdam. John Benjamins.

Bardovi-Harlig, K. (2010) 'Exploring the pragmatics of interlanguage pragmatics: Definition by design'. In A. Trosborg (ed.) *Handbook of Pragmatics: Pragmatics across Languages and* Cultures (pp.219-260).Berlin: Mouton de Gruyter.

Bardovi-Harlig, K., & Bastos, M. T. (2011). Proficiency, length of stay, and intensity of interaction and the acquisition of conventional expressions in L2 pragmatics. *Intercultural Pragmatics*, *8*(3), 347-384.

Bardovi-Harlig, K., & Dörnyei, Z. (1998). Do language learners recognize pragmatic violations? Pragmatic versus grammatical awareness in instructed L2 learning. *Tesol Quarterly*, *32*(2), 233-259.

Bardovi-Harligr, K. & Hartford, B.S. (1993). Refining the DCT: Comparing open questionnaires and dialogue completion tasks. In L.F. Bouton (Ed.), *Pragmatics and Language Learning*, *4* (pp.143-165). Urbana, Ill.: DEIL.

Barron, A. (2012). Interlanguage pragmatics: From use to acquisition to second language pedagogy. Language Teaching, 45(1), 44-63.

Barron, A., & Warga, M. (2007). Acquisitional pragmatics: Focus on foreign language learners. *Intercultural Pragmatics*, *4*(2), 113-127.

Blum-Kulka, S., House, J., & Kasper, G. (1989). *Cross-cultural Pragmatics: Requests and apologies*. Norwood, NJ: Ablex Pub. Corp.

Booth, W. (1974). A rhetoric of irony. Chicago: University of Chicago Press.

Bouton, L. (1992). The interpretation of implicature in English by NNS: Does it come automatically without being explicitly taught? In L. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning monograph series, vol. 3* (pp.66–80). Urbana-Champaign: University of Illinois.

Bouton, L. (1994). Can NNS skill in interpreting implicature in American English be improved through explicit instruction?: A pilot study. In L. Bouton & Y. Kachru (Eds.), *Pragmatics and language learning monograph series vol. 5* (pp. 88–108). Urbana-Champaign: University of Illinois at Urbana-Champaign.

Bouton, L. (1999). Developing nonnative speaker skills in interpreting conversational implicatures in English. In E. Hinkel (Ed.) *Culture in second language teaching and learning* (pp.47-70). New York: Cambridge University Press.

Bowes, A. & Katz, A. (2011). When sarcasm stings. *Discourse Processes*, *48*(4), 215-236.

Brown, J. D. (1988). *Understanding research in second language learning: A teacher's guide to statistics and research design*. Cambridge: Cambridge University Press.

Brown, J. D. (2001). *Using surveys in language programs*. Cambridge: Cambridge University Press.

Brown, J.D. (2011). Likert items and scales of measurement. *SHIKEN: JALT Testing & Evaluation SIG Newsletter, 15*(1), 10-14.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language use*. Cambridge: Cambridge University Press.

Burgers, C. (2010). *Verbal irony: Use and effects in written discourse*. Unpublished PhD Thesis, University of Nijmegen.

Campbell, J. D. & Katz, Albert N. (2012). Are there necessary conditions for inducing a

      sense of sarcastic irony? *Discourse Processes, 49*, 459–480.

Capelli, Carol A., Nakagawa N., & C. A. Madden (1990). How children understand

      sarcasm: The role of context and intonation. *Child Development, 61*, 1824-184.

Carvalho, P., Sarmento, L., Silva, M. & Oliveira, E. (2009). Clues for detecting irony

    in user-generated contents: Oh...!! It's "so easy" ;-). In *Proceedings of the 1st*

    International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion

    (TSA'09) (pp. 53-56). New York, NY: ACM.

Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.

Clark, H. & Gerrig, R. (1984). On the pretence theory of irony. *Journal of Experimental*

      Psychology: General, 113, 121-126.

Cheang, H. & Pell, M (2008). The sound of sarcasm. *Speech Communication, 50*, 366–381.

Cohen, A.D., & Shively, R. L. (2002). Measuring speech acts with multiple rejoinder

     DCTs. *Language Testing Update*, *32*, 39-42.

Colston, H. L. (1997). Salting a wound or sugaring a pill: The pragmatic functions

      of ironic criticism. *Discourse Processes, 23*, 25–46.

Colston, H. L. (2000). On the necessary conditions for verbal irony comprehension.

     Pragmatics and Cognition, 8, 277–324.

Colston, H. L. (2002). Contrast and assimilation in verbal irony. *Journal of Pragmatics, 34,*

    111–142.

Culpeper, J. (1996). Towards an anatomy of politeness. *Journal of Pragmatics, 25*, 349-367.

Culpeper, J. (2005). Impoliteness and *The Weakest Link*. *Journal of Politeness Research*, *1*(1),35–72.

Culpeper, J. (2011). *Impoliteness: Using language to cause offence.* Cambridge: Cambridge University Press.

Culpeper, J., Haugh, M. & Kadar, D. (2017). *The Palgrave Handbook of Linguistic (im)politeness.* London: Palgrave Macmillan.

Cutrone, P. (2011). Politeness and Face Theory: Implications for the Backchannel Style of Japanese L1/L2 Speakers. *Language Studies Working Papers*, *3*, 51-57.

Dalmau, M. S., & Gotor, C. H. (2007). From 'sorry very much' to 'I'm ever so sorry': Acquisitional patterns in L2 apologies by Catalan learners of English. *Intercultural Pragmatics*, *4*(2), 287-315.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.

Dörnyei, Z. (2010).*Questionnaires in Second Language Research: Construction, Administration, and Processing*. New York: Routledge.

Ellis, R. (1992). Learning to communicate in the classroom. *Studies in Second Language Acquisition,14*, 1-23.

Farias, D., Patti, V., Rosso, P. (2016). Irony detection in Twitter: The role of affective content. *ACM Transactions on Internet Technology*, *16*(3), 59-83.

Felix-Brasdefer, J. C. (2010). Data collection methods in speech act performance. In A. Martinez-Flor & E. Uso-Juan (Eds.), *Speech Act Performance: Theoretical, empirical and methodological issues* (pp.41-56). Amsterdam: John Benjamins.

Felix-Brasdefer, J. & McKinnon, S. (2016). Perceptions of impolite behavior in study abroad contexts and the teaching of impoliteness in L2 Spanish. *Journal of Spanish Language Teaching*, *3*(2), 99-113.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: SAGE Publications Ltd.

Filatova, E. (2012). Irony and sarcasm: Corpus generation and analysis using crowdsourcing. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12)* (392–398). European Language Resources Association (ELRA), Istanbul.

Fraenkel, J., & Wallen, N. (2003). *How to design and evaluate a research in education* (5th ed.). New York: McGraw-Hill.

Gibbs, R. (1993). Process and products in making sense of tropes. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 252–276). Cambridge: Cambridge University Press.

Gibbs, R. (2000). Irony in talk among friends. *Metaphor and Symbol*, *15*, 5-27.

Giora, R. (1998). Irony: Graded salience and indirect negation. *Metaphor and Symbol*, *13*(2) 83-101.

González-Ibanez, R., Muresan, S. & Wacholder, N. (2011). Identifying sarcasm in Twitter: A closer look. In *The 49th annual meeting of the association for computational linguistics: human language technologies ACL HLT* (pp. 581-586). Portland: Omnipress, Inc.

Grice, H. P., 1975. Logic and conversation. In: P. Cole and J. Morgan (Eds.), *Syntax and Semantics* (pp. 41-58). New York, NY: Academic Press.

Grice, H. P. (1978). Further Notes on Logic and Conversation. In P. Cole (Ed.), *Pragmatics: Syntax and Semantics*, Vol. 9 (pp. 113-127).New York, NY: Academic Press.

Grice, H. P. (1989). *Studies in the way of words*. Cambridge: Harvard University Press

Griffee, D. (2012). *An introduction to second language research methods: Design and data*.
    California: TESL-EJ Publications.

Halenko, N. (2016). *Evaluating the explicit pragmatic instruction of requests and*
    apologies in a study abroad setting: The case of Chinese ESL learners at a UK
    Higher Education Institution. Unpublished PhD Thesis, Lancaster university.

Hamamoto, H. (1998). Irony from a cognitive perspective. In R. Carston & S. Uchida (Eds.),
    Relevance theory: Applications and implications (pp. 257–270). Amsterdam: John Benjamin.

Hancock, J. T. (2004). Verbal irony use in face-to-face and computer-mediated conversations.
    Journal of Language and Social Psychology, 23, 447–463.

Haugh, M. (2014). *Im/politeness Implicatures*. Berlin: Mouton de Gruyter.

Hinkel, E. (1997). Appropriateness of advice: DCT and multiple choice data. *Applied*
    Linguistics, *18*(1), 1–26.

Ishihara, N., & Cohen, A.D. (2010). *Teaching and Learning Pragmatics: Where Language*
and Culture meet. Harlow: Longman.

Iwasaki, N. (2011) . Learning L2 Japanese 'Politeness' and ' Impoliteness' : Young American
    Men's Dilemmas During Study Abroad. *Japanese Language and Literature*, *45*, 67-106.

Jorgensen, J. (1996). The functions of sarcastic irony in speech. *Journal of Pragmatics,*
    26(5), 613-634.

Kasper, G. (1979). Errors in speech act realization and use of gambits. *Canadian Modern*
    Language Review, 35(3), 395-406.

Kasper, G. (1981). *Pragmatische Aspekte in der Interimsprache*. Tuebingen: Narr.

Kasper, G. (2008). Data collection in pragmatics research. In H. Spencer-Oatey (Ed.)

    Culturally speaking: Culture, communication and politeness theory (pp. 279-303).

    London: Continuum International Publishing Group.

Kasper, G. & Dahl, M. (1991). Research methods in interlanguage pragmatics. *Studies in*

    Second Language Acquisition, 13, 215-247.

Kasper, G. & Rose, R. (1999). Pragmatics and SLA. *Annual Review of Applied*

    Linguistics, 19, 81-104.

Kasper, G & Rose, K. (2003). *Pragmatic development in a second language.*

    (*Language Learning Monograph Series).*Oxford: Wiley-Blackwell.

Katz, A. (2000). Introduction to the special issue: The uses and processing of irony and

    sarcasm. *Metaphor and Symbol, 15* (1), 1-3.

Kaufer, D. (1981). Understanding ironic communication. *Journal of Pragmatics*, *5*(6), 495-510.

Kitao, K. (1990). A study of Japanese and American perceptions of politeness in requests.

    Doshida Studies in English, *50*, 178–210.

Kim, J. (2014). How Korean EFL learners understand sarcasm in L2 English. *Journal of*

    *Pragmatics,60,*193-206.

Koike, D. A. 1996. Transfer of pragmatic competence and suggestions in Spanish foreign

    language learning. In S. M. Gass and J. Neu (Eds.) *Speech acts across cultures* (pp.257-

    281)*.* Berlin: Mouton de Gruyter.

Kovaz, D., Kreuz, R. J., & Riordan, M. A. (2013). Distinguishing sarcasm from literal

    language: Evidence from books and blogging. *Discourse Processes, 50*(8), 598 – 615.

Kreuz, R. J., & Caucci, G. M.(2007). Lexical Influences on the Perception of Sarcasm. *FigLanguages*

    07 Proceedings of the Workshop on Computational Approaches to Figurative Language, 1-4.

Kreuz, R. J., & Glucksberg, S. (1989). How to be sarcastic: The echoic reminder theory of verbal irony. *Journal of Experimental Psychology, 118*(4), 347-386.

Kreuz, R. J., & Roberts, R. M. (1995). Two cues for verbal irony: Hyperbole and ironic tone of voice. *Metaphor And Symbolic Activity, 10*(1), 21-31.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS.* New York: Routledge.

Leech, G. (1977). *Language and tact*. Linguistic Agency, University of Trier.

Leech, G. (1983). *Principles of pragmatics*. London: Longman.

Leech, G. (2014).*The pragmatics of politeness.* Oxford: Oxford University Press.

Li, S., & Taguchi, N. (2014). The effects of practice modality on the development of pragmatic performance in L2 Chinese. *Modern Language Journal*, *98*, 794-812.

Liebrecht, C., Kunneman, F. & Bosch, A. (2013). The perfect solution for detecting sarcasm in tweets #not. In *Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp.29-37). Association for Computational Linguistics, Atlanta, Georgia.

Mackey, A. & Gass, M. (2016). *Second Language Research*: *Methodology and Design*. New York: Routledge.

Macmillan English Dictionary, retrieved from http://www.macmillandictionary.com/

Martin, R. (1992). Irony and universe of belief. *Lingua, 87*, 77–90.

Martinez-Flor, A. (2013). Learners' Production of Refusals: Interactive Written DCT versus oral Role-Play. *Utrecht Studies in Language & Communication*, *25*, 175-211.

Matsumura, S. (2003). Modelling the relationships among interlanguage pragmatic development, L2 proficiency, and exposure to L2. *Applied Linguistics*, *24*(4), 465-491.

McDonald, S. (1999). Exploring the Process of Interference Generation in Sarcasm: A
    Review of Normal and Clinical Studies. *Brain and Language*, *68*, 486–506.

Mugford, G. (2012). 'I Wouldn't Say that if I Were You': Face to Face with Foreign-Language
    Impoliteness. *Journal of Politeness Research*, *8*,195-221.

Myers Roy, A. (1978). *Irony in conversation.* Ann Arbor, MI: University Microfilms
    International.

Olshtain, E. and Blum-Kulka, S. (1985). Degree of approximation: Normative reactions to
    native speech act behavior. In S. Gass and C. Madden (Eds.) *Input in second language*
    acquisition (pp.303-325)*.* Rowley, MA: Newbury House.

Oxford English Dictionary, retrieved from http://www.oed.com/

Partington, A. (2007). Irony and reversal of evaluation. *Journal of Pragmatics*, *39*,1547–1569.

Partington, A. (2011). Phrasal irony: Its form, function and exploitation. *Journal of Pragmatics*,
    43,1786–1800.

Peters, S., Wilson, K. , Boiteau, T., Gelormini-Lezama, C., Almor, A. (2016). Do you hear it now? A
native advantage for sarcasm processing. *Bilingualism: Language and Cognition*, *19* (2), 400-414.

Pinto, D. (2005). The acquisition of requests by second language learners of Spanish. *Spanish*
    in context, *2*(1), 1-27.

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices
    in quantitative L2 research. *Studies in Second Language Acquisition*, *35*, 655–687.

Ptacek, T., Habernal, I. & Hong, J. (2014). Sarcasm detection on Czech and English
    Twitter. In *Proceedings of COLING, the 25th International Conference on*
    Computational Linguistics: Technical Papers (pp.213–223). Coling, Dublin.

Rea, L. M., & Parker, R. A. (1992). *Designing and conducting survey research: A*
    comprehensive guide. San Francisco, CA: Jossey-Bass.

Reyes, A., Rosso, P. & Buscaldi, D. (2012). From humor recognition to irony detection: The figurative language of social media. *Data Knowl. Eng*., *74*, 1–12.

Reyes, A., Rosso, P. & Veale, T. (2013). A multidimensional approach for detecting irony in Twitter. *Language Resources and Evaluation*, *47*(1), 239–268.

Rockwell, P. (2000). Lower, slower, louder: vocal cues of sarcasm. *Journal of Psycholinguistics Research*, *29*, 483-495.

Rockwell, P. (2006). *"Yeah, right!" A linguistic analysis of self- reported sarcastic messages and their contexts.* Los Angeles, LA: University of Louisiana at Lafayette.

Roever, C. (2011). Testing of second language pragmatics: Past and future. *Language Testing*, *28*(4), 463-481.

Roever, C., Fraser, C. & Elder, C. (2014). *Testing ESL sociopragmatics: Development and validation of a web-based test battery.* Frankfurt: Peter Lang.

Rose, K. (1992). Speech acts and questionnaires: The effect of hearer response. *Journal of Pragmatics*, *17*, 49-62.

Scarcella, R. (1979). On speaking politely in a second language. In C. A. Yorio, K. Perkins & J. Schachter (eds.) *On TESOL '79* (pp.275-287). Washington, DC: TESOL.

Schauer, G. A. (2006). Pragmatic awareness in ESL and EFL contexts: Contrast and development. *Language Learning*, *56*(2), 269-318.

Schauer, G. A. (2007). Finding the right words in the study abroad context: The development of German learners' use of external modifiers in English. *Intercultural* Pragmatics, *4*(2), 193–220.

Schauer, G. A. (2008). Getting better in getting what you want: Language learners' pragmatic development in requests during study abroad sojourns. In M. Puetz & J. Neff van Aertselaer (Eds.), *Developing Contrastive Pragmatics: Interlanguage and Cross-cultural perspectives* (pp.399–426). Berlin: Mouton de Gruyter.

Schauer, G. A. (2009). *Interlanguage pragmatic development: The study abroad context*. London: Continuum International Publishing Group.

Schauer, G. A. (2010). Study abroad and its effect on speech act performance. In A. Martinez-Flor & E. Uso-Juan (Eds.), *Speech Act Performance: Theoretical, empirical and methodological issues* (pp.91-108). Amsterdam: John Benjamins.

Schmidt, R. (1983). Interaction, acculturation and the acquisition of communicative competence. In N. Wolfson and E. Judd (Eds.) *Sociolinguistics and second language acquisition* (pp. 137-174). Rowley, MA: Newbury House.

Selinker, L. (1972). Interlanguage. *IRAL-International Review of Applied Linguistics in Language Teaching*, *10*, 209-232.

Seto, K. (1998). On non-echoic irony. In R. Carston & S. Uchida (Eds.), *Relevance theory: applications and implications* (pp.239–255). Amsterdam: John Benjamins.

Shively, R.L. (2013). Learning to be funny in Spanish during study abroad: L2 humour development. *The Modern Language Journal*, *97*(4), 930-946.

Shively, R., Menke, M.& Manzón-Omundson, S. (2008). Perception of Irony by L2 Learners of Spanish. *Issues in Applied Linguistics*, *16* (2), 101-132.

Sperber, D. (1984). Verbal irony: Pretense or echoic mention? *Journal of Experimental Psychology, General*, *113*(1), 130-136.

Sperber, D., & Wilson, D. (1981). Irony and the use-mention distinction. In P. Cole (Ed.), *Radial pragmatics* (pp.295-318). New York, NY: Academic Press.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Oxford: Blackwell.

Taguchi, N. (2008). Cognition, language contact, and the development of pragmatic comprehension in a study-abroad context. *Language Learning*, *58*(1), 33–71.

Taguchi, N. (2010) 'Longitudinal studies in interlanguage pragmatics'. In A. Trosborg (ed.) *Handbook of Pragmatics: Pragmatics across Languages and Cultures* (pp.333-361). Berlin: Mouton de Gruyter. 333-361.

Taguchi, N. (2011). The Effect of L2 Proficiency and Study-Abroad Experience on Pragmatic Comprehension. *Language Learning*, *61*(3), 904-939.

Taguchi, N. (2012). *Context, Individual Differences, and Pragmatic Competence*. New York: Multilingual Matters.

Taguchi, N. (2013). Production of routines in L2 English: Effect of proficiency and study-abroad experience. *System*, *41*, 109-121.

Taguchi. N. (2015) 'Instructed pragmatics at a glance: Where instructional studies were, are, andshould be going'. *Language Teaching, 48*,1-48.

Taguchi, N. (2017). Interlanguage pragmatics. In A. Barron, P. Grundy, & G. Yueguo (Eds.), *The Routledge Handbook of Pragmatics* (pp.153–167). Oxford/New York: Routledge.

Taguchi, N. & Roever, C. (2017). *Second Language Pragmatics*. Oxford: Oxford university press.

Takahashi, T., & Beebe, L. M. (1987). The development of pragmatic competence by Japanese learners of English. *JALT journal*, *8*(2),131-155.

Takahashi, S. & DuFon, M. (1989). *Cross-linguistic influence in indirectness: The case of English directives performed by native Japanese speakers.*[ERIC Document ED 370-439].

Taylor, C. (2015). Beyond sarcasm: The metalanguage and structures of mock politeness. *Journal of Pragmatics*, *87*,127-141.

Taylor, R. J. (1995). *Linguistic categorization: Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Thomas, J. A. (1995). *Meaning in interaction: An introduction to pragmatics*. London: Longman.

Togame, N. (2016). *Irony in a Second Language: exploring the comprehension of Japanese speakers of English*. Unpublished PhD thesis, Middlesex University.

Toplak, M., & Katz, A. (2000). On the Uses of Sarcastic Irony. *Journal of Pragmatics, 32*, 1467-1488.

Trosborg, A. (1987). Apology strategies in natives/non-natives. *Journal of Pragmatics,11*, 147-167.

Trosborg, A. (1995). *Interlanguage pragmatics.* Berlin: Mouton de Gruyter.

Tsur, O., Davidov, D. & Rappoport, A. (2010). ICWSM – A great catchy name: Semi-supervised recognition of sarcastic sentences in online product reviews. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social* Media (pp.162-169). Association for the Advancement of Artificial Intelligence, Jerusalem.

Utsumi, A., (2000). Verbal irony as implicit display of ironic environment: distinguishing ironic utterances from nonirony. *Journal of Pragmatics, 32*, 1777–1806.

Vergis, N. (2017). The interaction of the Maxim of Quality and face concerns: An experimental approach using the vignette technique. *Journal of Pragmatics*, *118*, 38-50.

Wang, P. (2013). Irony or Sarcasm—A Quantitative and Qualitative Study Based on Twitter. PACLIC, 27*, 349-356.*

Watts, R., Ide, S., and Ehlich, K., (Eds.) (1992). *Politeness in language: Studies in its history, theory and practice*. Berlin: Mouton de Gruyter.

Wilson, D. (2006). The pragmatics of verbal irony: Echo or pretence? *Lingua Journal, 116* (10), 1722-1743.

Wilson, D. (2013). Irony comprehension: A developmental perspective. *Journal of Pragmatics*, *59*, 40-56.

Wilson, D. & Sperber, D. (1992). On verbal irony. *Lingua, 87*, 53-76.

Wilson, D., & Sperber, D. (2002). Relevance theory. *UCL Working Papers in Linguistics, 14*, 249-287.

Wilson, D.& Sperber, D. (2012). Explaining irony. In D. Wilson, & D. Sperber (Eds.), *Meaning and Relevance* (pp.123-145). Cambridge: Cambridge University Press.

Yamanaka, J. (2003). Effects of Proficiency and Length of Residence on the Pragmatic Comprehension of Japanese ESL Learners. *Second Language Studies*, *22*(1), 107-175.

Youn, S. J. (2015). Validity argument for assessing L2 pragmatics in interaction using mixed methods. *Language Testing*, *32*(2), 199-225.

Yus, F. (1998). Irony: Context accessibility and processing effort. *Pragmalingüística,* 5-6, 391-410.

Yus, F. (2000). On reaching the intended ironic interpretation. *International Journal of Communication,10*(1), 27-78.

Notes

[1] Field (2013) argues that Games-Howell is the most powerful test within the *equal-variances-not-assumed* list provided by SPSS. He also contends that this test is accurate in the case of inequality of sample sizes.

[2] Rea and Parker (1992) define probability sampling as "the probability of any member of the working population being selected to be a part of the eventual sample is known" (p. 147). By definition, it requires an access to all the members of the investigated populations. This kind of sampling is not common or even feasible in second language research due to being expensive, time-consuming and involving complex procedures (see Dörnyei, 2010, p. 60).

[3] Non-probability sampling is biased by nature as not all the members of the population are given even chances to be selected. Despite this limitation, this kind is more realistic for scientific research which is generally characterized by limitedness in terms of time, recourses and effort (see Griffee, 2012).

[4] Unequal sample sizes may result in heterogeneity of variances, and the ANOVA F–statistic might not be robust enough in the case of unequal sample sizes if the data was non-normally distributed as well (see Wilcox, 2005, p. 9-10).

[5] Internal consistency is a psychometric term which refers to a kind of reliability. It relates to multi-item scales (e.g. Likert scale). On the one hand, it tests whether each item on a scale correlates with the other items. On the other hand, it also tests whether each item correlates with the total score of the scale. If attested, this means that the items work homogenously and the scale effectively measures the construct in question (see Dörnyei, 2010, p. 94).

[6] Cronbach's alpha (or simply *Coefficient alpha*) is a measure for internal consistency. It is "a figure ranging between zero and +1…and if it proves to be very low, either the particular scale is too short or the items have very little in common" (Dörnyei, 2010, p. 94). Cronbach's alpha of 0.80 and above indicates a good level of internal consistency. However, in second language

research and due to the complexities it has, a Cronbach's alpha of 0.70 is considered acceptable as well (Dörnyei, 2010, 94-95).

[7] Shapiro-Wilk test is a significance test which assumes the null hypothesis that "the data are sampled from a normal distribution. When the p-value is greater than the predetermined critical value ($\alpha=0.05$), the null hypothesis is not rejected and thus we conclude that the data is normally distributed" (Ahad et al., 2011, p. 637). Several studies have shown that this test is very powerful for testing the normality of data and would highly recommend it for researchers (see Ahad et al., 2011; Mendes and Pala, 2003; Razali and Wah, 2011; Ricci, 2005).

[8] Levene's test is also a significance test. It verifies the null hypothesis that the variances of the tested groups are homogenous. If the p-value it yields is over 0.05, then we fail to reject that null hypothesis and assume the groups variances to be equal (see Larson-Hall, 2010, p. 88).

[9] Tukey HSD test is amongst the most commonly-used post hoc tests and is claimed to be more powerful than some other tests such as *Dunn* and *Scheffé* tests (see Field, 2013).

Appendices

Appendix A

Ethics Documents

Date:

INFORMATION SHEET

As part of my PhD study in the Department of Linguistics and English Language/Lancaster University, I am carrying out a study on how Iraqi EFL Learners recognize and understand sarcasm in British English. The study involves completing a judgment task based on extracts taken from online English blogs.

Participation in this study is entirely voluntary: You are free to withdraw from the study at any time before submitting your data. At every stage, your name will remain confidential; all data is completely anonymized. The data will be kept securely and will be used for academic purposes only.

If you have any queries about the study, please feel free to contact me or the project supervisor, Dr. Patrick Rebuschat, p.rebuschat@lancaster.ac.uk, phone: 01524 - 592433.

Signed

Dheyaa Jasim Issa Al-Fatlawi

PhD Student of Linguistics

Lancaster University

Tel: 07459010065

Email: d.al-fatlawi@lancaster.ac.uk

Lancaster University

Lancaster LA1 4YL

United Kingdom

Tel: +44 (0)1524 593045

Fax: +44 (0)1524 843085

http://www.ling.lancs.ac.uk

Lancaster University

Department of Linguistics and English Language

<u>Consent Form</u>

Project title: **Online Sarcasm and its Perception by Second Language Learners: The Case of Iraqi EFL Learners in Iraq and the UK**

1. I have read and had explained to me by Dheyaa Al-Fatlawi the Information Sheet relating to this project.

2. It has been explained to me the purposes of the project and what will be required of me, and any questions have been answered to my satisfaction. I agree to the arrangements described in the Information Sheet in so far as they relate to my participation.

3. I understand that my participation is entirely voluntary and anonymous, and that I have the right to withdraw from the project any time before or during the test. But, I also understand that I can ask for cancelling and destroying my data within two weeks' time after doing the test. Otherwise, the data will be included in the study analysis.

4. I have received a copy of this Consent Form and of the accompanying Information Sheet.

Name:

Signed:

Date:

Appendix B

Stimuli Selection Task: First Pilot Test

(Supplemental Material)

This appendix presents the material used in the first pilot test of the stimuli selection task (SST). It consist of 71 sarcasm-containing threads from Manchester united forum and Mumsnet and Netmums forums. The original metalinguistic remarks are retained in the threads. To upload the appendix, click on the following link.

File: Appendix B

Appendix C

Stimuli Selection Task: Second Pilot Test

(Supplemental Material)

This appendix presents the material used in the second pilot test of the stimuli selection task (SST). It consist of 70 threads from Manchester united forum and Mumsnet and Netmums forums. The majority of them contain sarcasm, whereas the minority do not involve sarcasm and used as control items. The original metalinguistic remarks are deleted in this pilot study. To upload the appendix, click on the following link.

File: [Appendix C](#)

Appendix D

Stimuli Selection Task: Main Task

(Supplemental Material)

This appendix presents the material used in the main stimuli selection task (SST). It consist of 70 threads from Manchester united forum and Mumsnet and Netmums forums. The majority of them contain sarcasm, whereas the minority do not involve sarcasm and used as control items. The original metalinguistic remarks are deleted in this main study. To upload the appendix, click on the following link.

File: Appendix D

Appendix E

Pilot and Main L2 Pragmatics Study

Sarcasm Judgment task

Introduction & Instructions

Introduction:

What if you say something sarcastic to your international colleague and s/he does not get your point! You need to say it plainly again and explain that you were being sarcastic! As part of my PhD study at the Linguistics department / Lancaster University, I have designed the current judgment task to elicit judgments from a group of English native speakers/a group of Iraqi EFL learners about sarcasm in British English. This is not a test and there are no ''right'' or ''wrong'' answers. You do not need to mention your name and all the data you will provide will be kept anonymous and used for academic research only. Please, consider each item carefully. Thank you in advance for your help!

Instructions: (Between parentheses was added to the instructions of the main study only)

The following is 30 short excerpts taken from online British English forums. They are distributed to two parts: part 1 contains excerpts from a forum of football fans (Manchester United forum), and part 2 comprises excerpts from two forums of parenting affairs (Mumsnet and Netmums).

Please, read every excerpt carefully. Then, you have two tasks to do:

(1) Provide your judgment whether or not there is sarcasm in each excerpt on a scale from "strongly disagree" to "strongly agree".

(2) (ONLY if you agree there is sarcasm in the excerpt), underline or highlight the specific part or parts which led you to think that there is sarcasm.

An example of how you might perform these tasks is provided immediately below:

MexUnited:

Moyes is a GREAT manager

RedYankee:

Write your sh*t away from our forum

MexUnited:

THANK YOU for being nice!

RedYankee:

It's a PLEASURE!

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Notes:

1. The judgment task uses the actual pseudonyms that occur in the forums.

2. Anything embedded between two square brackets [ ] is the researcher's addition

   to the actual wording. This is for clarifying some vague words and phrases.

Part 1

Football Forum

(Manchester United Club)

united1985:

Moyes [Manchester united manager] is doing pretty good in the CL[football

competition] for someone who pretty much has no CL experience

Captain Marvel No.7:

Yeah NOT BAD for someone who has no clue, is so negative, plays boring football,

can't attract players, had ONE full transfer window to sort the whole of the midfield

out and has no European experience....

He remains undefeated in the CL as United manager and tonight has achieved

something Alex Ferguson [ex Manchester united manager] couldn't in 26 years as

manager. Yeah, NOT BAD for a clueless manager!!!

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

zimpapa:

UNITED, I THINK SHOULD CONSIDER GETTING ONE OR TWO MORE

PLAYERS WITH SPECIAL ABILITIES TO HELP RAISE OUR EVER

FORMIDABLE STATE.

robi prosser:

i agree i think we need to sign two one with the ability to fly no one would ever

tackle him and one who can turn to stone no one would ever get past him in goal..lol

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q3

*(in a debate about buying new players)

Le king cantona:

money wise were not going to spend 40 to 45 million on one player . so I'd go for

Nuri Sahin [player name] fee 16 million, Bruno Soriano[player name] good

holding cm[centre middle] …these be great in the centre of our team.

diwan alhaji:

there's another post on Sahin.. and not we don't need a good player we need a

WORLD CLASS PLAYER!

ChrisGManUfan:

Hey why spend money at all on young players, just get a bunch on frees[unpaid

players] in the summer and keep the ones that work out and sell the rest, this way

we can keep our cash to bring Rooney[former player] back.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q4

 *Ashley Young! [young player name]

deadmanavir

What the F**k happened to him? Judging ONLY on the games he played under

LVG[Manchester United coach], what do you think of him? Are we seeing a new Young?

Michael Ballard

A small sample size but so far he looks like a new signing. Hope he keeps it up. You

gotta tip your cap the guy has been class.

Ariana99

Nope. Looks like the same Ashley Young I've always watched just in a different position than
he is used to.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q5

(Oshea [player name] Out For Two Months)

Gabranth:

I'm not happy that he's injured, I'm just happy for all of us that he won't be playing.

RedYankee:

For all those who think Gabranth is a United fan

He is happy Oshea[player name] is hurt. What a great United fan he is, huh?

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q6

Yorick:

I hope we never go for Silva[player name]

rooneygeniusno.8/no.9bigfan:

What a surprise there. We just forgot you know everything that isn't proven, Yorick.

Yorick:

did I ever say I know everything.no but you know nothing. it's true.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q7

denismf:

Why would Rooney [player name]absence be crucial? We have got Chicharito[player name],

he's gonna win it for us. We don't even need

Berbatov[player name] let him stay on the bench, Chicha [player name]

will do the job alone.

ManUTDecade:

i don't remember you being a Chicha [player name]fanboy. Anyway, are you guys

forgetting the Berbatov[player name] and Hernandez [player name] partnership

when Rooney [player name]was out ? I thought it worked out well.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q8

anibalin:

What were the fans chanting during the second half?

I couldn't figure out what were the fans chanting, I know it was something like "city,

la la la la" , lol, I'm just curious.

xxRedDevilxx (2):

It made me laugh when the [Manchester]United fans took the piss out of the City

fans after they wanted a penalty for a hand ball. The United fans chanted "HAND

BALL!" every time a United player got the ball.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q9

(Edinson Cavani [player name]to Chelsea[football club] done deal)

J1210:

Chelsea will be waiting on the Rooney [player name]saga

fuldagap:

Rooney [player name]must be their 2nd or 3rd choice then. Since they were in for

Falcao [player name]and Cavani [player name]first.

J1210:

Remember how many players we have been linked with already, not everything is

necessarily true. Papers seem to write any old rubbish now to sell, next it'll be we've

submitted a £15m bid and a chunky kitkat for Messi[player name]

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q10

Beirut1987

Man Utd [Manchester united]must move quickly to land Benfica [Portugal football

club]wingback

Coentrao [player name]

ManUTDecade:

I think IF we bought him, it will be a replacement for Giggs[player name]..

mancrob:

but we're signing Sanchez[player name], goal.com said so.

ManUTDecade:

Oh, Wow so we must be really in need for another RIGHT footed winger ?? And

goal.com is not really reliable right?

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q11

vivek bheeroo

Who do you'll think will win the CHAMPIONS LEAGUE[football competition]?

Chris87

CHELSEA[football club]........

just put a grand on it!!!!!

garp01us

that was a bad bet, sorry dude but Chelsea never had a chance.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q12

(MAN SHITY'S MEGA BARGAIN SALES!)

Gary Mcleod

If we could manage to get Robinho[player name] for between £5-£8 million, or Silva[player name] for around £10 million, that would be very interesting. Anything more, I'd say no way jose.

lewis.No.9

hahahahahahahahaha....................Silva there NEW signing being instantly sold for £10Million.................jeez some fans are such IDIOTS.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q13

(Fergie[manager name] on the situation with Ravel Morrison[player name])

PaddyDevil

His agent has been working hard to get him another club. We've offered him terms

which he has refused. We've rejected an offer from Newcastle[football club]. We'll

see how that progresses. His demands are unrealistic as far as we're concerned.

Carlsberg King:

OMG WE HAVE NO MONEY AND WE CAN'T OFFER HIM 100k A WEEK.

GLAZIERS[player name] OUT SAF [Manchester united manager]OUT

MOURINHO[football club manger] IN.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Nash33

Alexander Büttner[player name]: Dutch LB on his way to [Manchester] United

(Reports/Rumors)

Reports in Netherland say we are close to signing him or is it merely a rumor

bw.k:

lol Lucas saga [player name] took ages...and this transfer done in apparently less

than 1 hour

therealry:

How do you know how long it took? Just curious

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q15

I'm the man:

If utd [Manchester united] is to sell Nani [player name] i like Ben Arfa [player name] to replace him. Not bale[player name] or Rodriguez[player name]. Ben Arfa is the man you want

we are the team to beat:

Oh look, *I'm the man* finally making some sort of sense.

Brittain 10:

No YOU'RE not, why would we replace a winger with an even less consistent one who's not as good? Will say this makes no less sense than the blasphemy you spew up the rest of the time.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Part 2

Parenting Forums

Mumsnet & Netmums

(I'd really like to thank the woman)

BellaBear:

I'd really like to thank the woman who followed me (waddling at 30 weeks) towards

the first capital connect train at Kentish Town this morning and at the last minute

pushed past me and took the last spot on the train. No, really, thanks. For the push

AND for the twenty minutes I had to wait for the next train.

Obviously, you may not have realised I was pregnant... it's obviously fine to push

past normal commuters and take their space. ???????????!

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

*Islam is anyone interested in it? No preaching, just asking...

PasseBlanc

Just wondering if anyone here is or has ever been interested in coming to Islam,

despite all the bad press/media hype etc?

I'm not here to quote and preach, I just want to maybe answer a few questions from

my own perspective as a convert muslim.

CanISawItOff

I'm interested from the point of view that I am comfortable in my own faith and

have no desire to convert.

PasseBlanc

I personally made the choice to convert because of the amount of scientific evidence

I came across.

CanISawItOff

Science convinced you to convert to a religion founded a few thousand years ago? 😵 really?

PasseBlanc

CanIsawItOff, do you have a few minutes where I can give you some examples

before you give me that comment and 😵 face?

CanISawItOff

it's genuine confusion. Not sure if you saw my thread from the other day regards

JEW's?

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q18

*Am I taking what dp [partner] said the wrong way?

Phylis81

Bit pissed off. Most weekends dp is grumpy, complaining about being tired and

stressed out and generally moangy. Normally we spend weekends together (not 24/7

but most of it) This is the first weekend where I have had to work from 9am until

10pm yesterday and from 8am until 9pm tonight … we've hardly seen each other.

I'm stuck at work now (obviously) and he's just texted saying he's had a really great

weekend thanks to me and the kids. I've not been there! I feel like he's thanking me

for not being around! Am I overreacting?

Why can't he be happy when I'm there?

He's also constantly asking me what time I'm finishing work etc and seems happier

when it's later. Feeling like he prefers it when I'm not around.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q19

Dull conversation

Tolalola:

Random shop assistant in the States – "And where are you visiting from?"

Me – "The Virgin Islands"

Shop assistant – "Oh, what do you do there, are you retired?"

Me – "Erm...no. I'm 38."

My stock response to situations like these consists of a withering look and remark

like "Gosh, are you *quite sure* you want to be a diplomat when you grow up?"

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

(To wonder why the head of the "Mothers' Union" is a man?)

Meditrina:

I've just seen him on the news. Reg Bailey. I've no reason to think he's doing

anything other than an excellent job and that he was the right person for the job.

But, in representational terms, it seems really, really odd. What do you think?

ResurrectionByChocolate:

Yep I thought it was odd too.

TheOriginalFAB:

I heard him mentioned on the radio this morning and was surprised too.

DontCallMePeanut:

It's because us women need a man to speak up for us...

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

(Cherry for a girl? HELP!)

Camilla C(30):

Hello all!

I am less one week away from giving birth and we can't agree on a name for our

baby girl!

I really like the name Cherry (like Cherry Healey[British television presenter] bbc3)

and my husband doesn't mind it either. I am just wondering what people think. The

reaction most people give is "really?! why not call her tomato" etc. I like unusual,

uncommon names so any other suggestions would be fab as well!

I'm not sure of middle names to go with Cherry either so any suggestions for that

would be brilliant!!

Thanks for reading!

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

JURGITA J(3):

BBC NEWS | England | Lancashire | Baby died after 'mother's kiss'

ok there is the link for mums that think it just a tale. yes i kiss all my daughter all

over, as a mum you can't help yourself. but i won't let anyone else except her dad to

do that.

Sara D(11):

As awful as it is it's kind of just one of those unfortunate things that happen

sometimes. Something like 10 people die a year while brushing their teeth, shall we

stop our kids doing that to avoid any danger?

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

(How to deal with friend who brags?)

Lauren D(63):

I love my friend i really do, but she is driving me insane. She's now working and

has spent everything on her child, which i totally understand. Except she sends me

pictures like "wow she's a spoilt child, bet you can't wait to do this", "I'm just glad i don't have

to get a loan", so on and so forth. It's just driving me crazy how she's

rubbing it in my face and she got offended when i told her it was way too much

Natasha H(169):

I'd just reply "that's nice!"

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q24

*To think that being passive is as bad as being controlling

trampstamp

I really feel being a passive parent is just as damaging as being a controlling one.

And sadly I think that this type of passive parenting is growing …There even been a

few threads on here about letting children **regulate their own behaviour**

sadbodyblue

if you lay down good foundations in the early years and be strict but loving you can

slacken off as they get older as they have learned boundaries and respect. I am

friends with my grown up lads and am becoming that relationship with my teen

daughters.

sadbodyblue

I think if you are 'passive' with toddlers and young children you have already lost

the teen.

trampstamp

That's very good saying ..wow

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q25

*What age do you expect to start paying for children's meals out?

Katie (2)

Just wondering really because we have recently stopped going to our Chinese buffet

restaurant because they were charging us for 3 children & 2 adults meals.

I thought this was unfair because DD3 [child]was only 19months the last time we

went & I always feed her something before we go to eat so she doesn't create a fuss

if we have to wait for food & her appetite isn't huge so usually ate a small bowl of

rice & a prawn cracker or two.

Kristine G(3)

When we go to our local Chinese buffet, my nan (who has the appetite of a sparrow)

only eats a tiny bowl of soup and one or two prawn crackers... So should we not

have to pay for that because the amount she's eaten doesn't justify the £8.95 per

head cost? 🤩

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q26

(kitchen roll!)

louise t(360):

ok bit of a nosey question but for you guys that use kitchen roll

how much a week do u go through and what do you use it for?

Kelly R(341):

Sorry to be really preachy but that's a lot of paper to be throwing away. What's

wrong with using a washable cloth? Not only is it better for the environment but it's

far cheaper. Of course I'm sure you all buy recycled kitchen towel...

lottie h(2):

i buy the shops own or what's on offer, not recycled.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q27

*whooping cough vaccine

Rm D

What are your reasons for not having whooping cough just out of interest?

Donna C(503)

2 main reasons 1) nobody knows the long term side effects

And 2) I don't feel it's necessary for me or my unborn baby to have it.

Why do you not feel it's necessary for your baby to have it? Is your baby naturally

immune to whooping cough?

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q28

Not currently a smoker...

but I have a baby due this August who we need to buy a lot of stuff for... so I was

thinking of taking up smoking, you know then maybe my doctor will give me £400

to quit so I can buy the baby some stuff?

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q29

Emma C(680):

My son was 15 months when he had chicken pox. He had some really itchy spots so

I phoned the doctor to ask for some advice. When I told her she replied "chicken

pox, well what do you want me to do about it?"!!!!! Don't get me wrong, it was

positively mild compared to this poor boy, but the point is, she refused to see him

because it was 'only' chicken pox 😎 So I can well believe that the drs were rubbish.

Poor little boy, a very sad story 🙁

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Q30

Clair D(33):

Reluctant to tell friends and family about my fourth pregnancy

lynsey c(94):

Hi i felt the same when i caught for my fourth, and dreaded telling people, like

others have said i got the comments "you not got a tele" and "you must be mad" but

since she's been here (now 4 months) she's the best thing I've ever done and has

made our little, well slightly large family even closer.

There is sarcasm in the text above.

Strongly disagree	Disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Agree	Strongly agree

Respondent information (English native speakers):

Please, provide the following information for statistical purposes:

Gender:

Male☐   Female☐

Nationality:

Age:

UK region you grew up in:

==

Respondent information (Iraqi EFL learners):

Please, provide the following information for statistical purposes:

1. Gender:

Male☐   Female☐

2. Age:

3. Nationality:

4. Native language(s):

5. Level of study

MA☐   PhD☐   Other (Please, specify)☐:

6. Place of Study:

Country ┌─────────┐   University ┌─────────┐

7. Major:

a. Linguistics ☐   b. English Literature ☐   c. Applied Linguistics or TEFL ☒

d. Translation ☐   e. Other (Please, specify) ☐:

8. English Language Proficiency

    a. Have you done any of the following exams?

    1. IELTS ☐
    2. TOEFL ☐
    3. PTE ☐

4. Other (Please, specify) □:

Overall grade (IELTS, TOEFL, PTE, Or other) or the equivalent after completing the pre/in

sessional course:

9. How long have you been studying in the UK (for students studying in the UK only)?

10. Have you ever been to an English-speaking country for more than 6 months (for

students not currently studying in the UK)?

No□   Yes (Please, specify for how long) □:

Appendix F

Categorization Test (Inter-rating Test): A Sample Item

Dear Sir/Madam,

   Thank you for your participation in this 'categorization test'. The following is a sarcasm judgement task that was done by some former participants. They were asked to read the excerpt in each item, which had been taken from an online forum, and pass judgments where sarcasm was within each excerpt. They were asked to highlight or underline the sarcastic part(s) within the texts they read.

   I have highlighted the choices of those participants (where sarcasm is) in all the task items. Please, your job in this test is to categorize the sarcasm in each item pragmatically. That is, you indicate which pragmatic characteristics are available in each sarcastic case.

Q5

(Oshea [player name] Out For Two Months)

Gabranth:

I'm not happy that he's injured, I'm just happy for all of us that he won't be playing.

RedYankee:
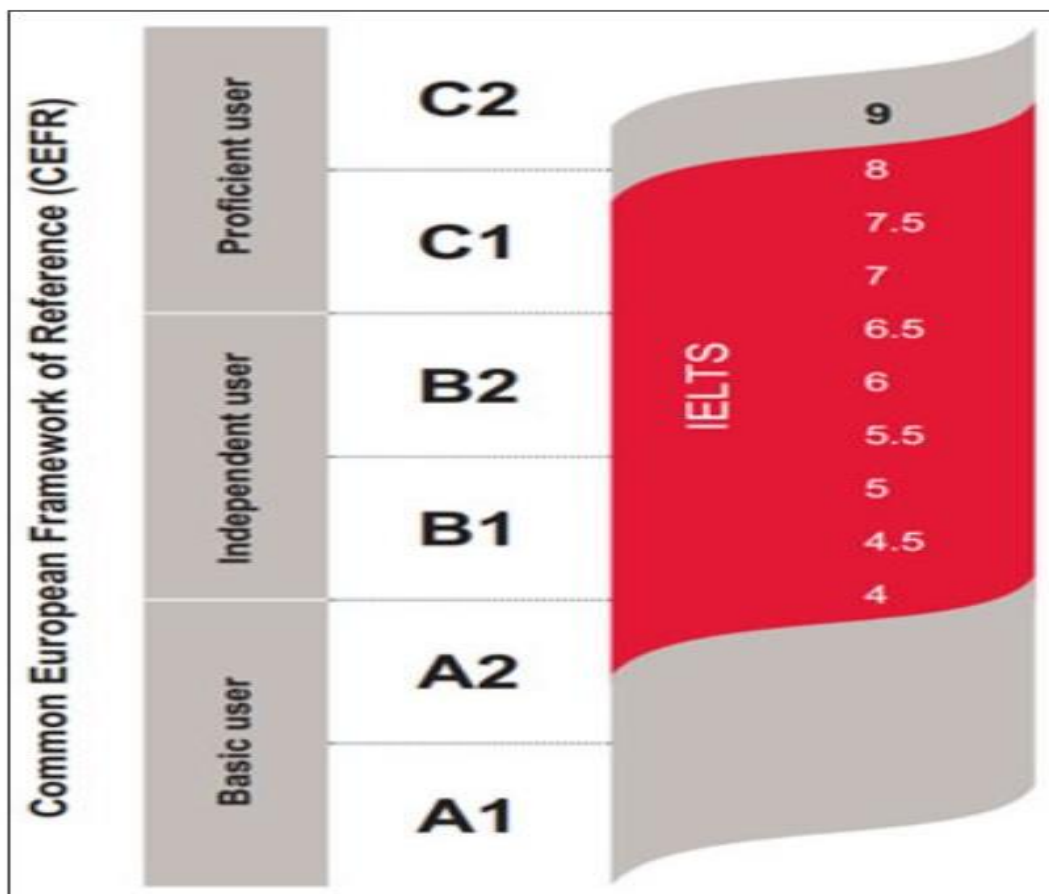
For all those who think Gabranth is a United fan

He is happy Oshea[player name] is hurt. What a great United fan he is, huh?

General Pragmatic Characteristics								
Contradiction	Insincerity (flouting Quality)	Flouting Quantity	Flouting Relevance	Mock politeness	Allusion to antecedent	Negative attitude	Victim	Other Characteristics
x	x				x	x	x	

Appendix G

Assessment of L2 Proficiency

IELTS Scores in comparison to the levels of 'Common European Framework of Reference for Languages' (CEFR)

Retrieved from: https://www.ielts.org/ielts-for-organisations/common-european-framework.

PROFICIENT USER	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.
INDEPENDENT USER	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes & ambitions and briefly give reasons and explanations for opinions and plans.
BASIC USER	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.

Retrieved from: http://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale

Appendix H

Detailed Statistics of the Main L2 pragmatics Study

(Supplemental Material)

This appendix provides detailed statistics for all the experimental items used in the main L2 pragmatics study. For every item, it provides the following:

1. How many scores each point on the Likert scale has got from the three examined groups.

2. The sarcastic part(s) which was/were highlighted by some of/all the three examined groups along with how many scores each sarcastic part has got (i.e., how many participants in each examined group has chosen the sarcastic part in question).

3. How many intermediate and advanced Iraqi EFL learners have chosen the available sarcastic part(s).

4. How many male and female Iraqi EFL learners have chosen the available sarcastic part(s).

5. A Statistic of the available general pragmatic characteristics of sarcasm.

6. A Statistic of the available pragmalinguistic characteristics of sarcasm.

To upload the appendix, click on the following link:

File: Appendix H