Discourse Markers in TV Series: the Issue of Authenticity in Scripted Dialogue

Relatrice
Prof. Katherine Ackerley

Laureanda
Eleonora Covolan
Matr. 1163868/LTLLM

Acknowledgments

Patience is a word that perfectly describe the people around me, who constantly show me what it means to be patient and never give up. Above all, I am extremely thankful to Professor K. Ackerley, who patiently guided me through the writing of this dissertation and gave me precious advice and suggestions.

I owe a deep sense of gratitude to all my family and friends who have been there for me for all these years: to my parents R.&D. and my brother M. thank you for backing my university career; to my boyfriend N. thank you for never being sick of my whining; to my loving friends M. and A. thank you for all the karaoke; to my university partners E. and E. thank you for these amazing and terrifying years; to my friend L. thank you for all the laughter.

Now you can stop being so concerned about my graduation!

Table of content

Introduction

The entertainment system has always been fascinating to people, who perceive of it as a way to escape reality and everyday life. Cinema and Television show situations and events that one will possibly never experience. In recent years, a new wave of television products, namely TV Series, changed the way people relate to fiction. Nowadays, a multitude of series, belonging to a multitude of genres, can be watched anytime and anywhere. I decided to focus my dissertation on this topic and to analyse a specific TV Series to which I am deeply connected, and which has been there for me throughout my growth: "Friends".

This dissertation aims to investigate the frequency of specific discourse elements, called discourse markers, in scripted dialogue in order to understand the level of authenticity of language use in TV Series opposed to language use in real life. In other words, what I will do is analyse the frequency of two discourse markers (*I mean*, *in fact*) in two specific corpora (Santa Barbara Corpus of Spoken American English; Friends: the TV Series Corpus) which represent respectively authentic language and non-authentic language. the goal is to observe how real and natural language use in fiction is. Therefore, authenticity is the *fil rouge* connecting all the individual parts of this work. The procedure through which I will write my dissertation consists of three major sequences: retrieving data and then writing a literature review; selecting the specific elements to analyse and collecting the corpora; analysing those elements and discussing the results.

First, I collected all the data necessary to write an extensive literature review on authenticity in language and describe all the specific terminology to use. Materials were retrieved both from books (e.g. Bednarek, 2010; Taylor, 2004) and from articles found in "Jstor", a digital library (e.g. Schegloff, 1999). The first chapter will explain notions such as "authenticity", "fiction", and "context". Specifically, I will analyse the dichotomy between authentic language and non-authentic language. This distinction is pivotal to the investigation because it gives characteristics for both language uses, thus allowing a better understanding of differences and similarities. Another important definition is that of "context". Indeed, the process of communication is multimodal, as it includes a number of elements that are both verbal (spoken interaction) and non-verbal (body movements).

The notions of discourse analysis and discourse markers will be introduced in the second chapter and then find application in the third chapter. The framework for this part of the study includes "Conversational Routines in English" written by Karin Aijmer (1996) and "Discourse Markers" written by Deborah Schiffrin (1989). The second chapter focuses on the description of discourse analysis and the data used in the analysis. For the purposes of this dissertation, two corpora have been selected: one is the "Santa Barbara Corpus of Spoken American English", representative of authentic language; the other is the "Friends: the TV Series Corpus", representative of non-authentic language. Then, I will dig into the theory of discourse markers, which are elements typical of naturally occurring interactions, providing categorization, and grammatical and functional features.

Once I have all the theoretical part of the dissertation, I will shift to the practical side. In the third chapter I will briefly illustrate "Friends" and describe both corpora used. Then, in the actual analysis I will define the grammatical and functional features of both discourse markers (*I mean*, *in fact*) before contextualizing these elements in the results. I subdivided my analysis in two parts: one focusing on the discourse marker *I mean* and the other on the discourse marker *in fact*. The individual parts will include examples for each corpus of the context in which discourse markers occur and a subsequent explanation of the results found.

Chapter 1

Spoken language and authenticity

The first chapter of this dissertation introduces some basic notions that will be central to the general understanding of the whole work. The aim of this chapter is twofold: first, it will introduce spoken language in general, defining its main characteristics with a focus on its authenticity. Second, it will describe the logics behind language use in the media, more particularly 'the language of fictional television' (Bednarek, 2010), with an overview of the main features of the fictional sphere.

1.1 Authentic Language

In order to understand what 'authentic language' means, it is crucial to have a clear interpretation of the two concepts which define it: 'authentic' and 'language'.

As for the latter term, there are two different conceptions of language which arise from philosophical reflection on the nature of human language. One has its roots in Aristotelian thought and was central in the philosophical discussion on language from the seventeenth to the nineteenth century. It follows a function-theoretic logic, according to which language is seen 'as the full suite of abilities to map sound to meaning, including the infrastructure that supports it' (Dediu and Levinson, 2013). The second is in turn an anthropological one. It does not prioritize notions such as truth and representation, which were central characteristics linked to the previous conception, but rather those of understanding and communication. Following this method, which became the focus point of philosophical discussion on human language only in the 1960s-1970s, human language is an 'extension of human behaviour' (Hacker, 2014:1282).

The former term, the adjective 'authentic', refers to the idea of language realism. Eagleton (1983:135-6), among others, states that the idea of realism 'helps to confirm the prejudice that there is a form of ordinary language which is somehow natural' and sure enough, socially speaking, people assume that this 'natural language', as Eagleton calls it, is the reflection of reality. Therefore, 'authentic' takes on the meaning of language use in real life, as opposed to 'non-authentic', which describes language use in scripted dialogue. One clear example of authentic language is conversation, which can be defined

from different approaches. From a linguistic point of view, a conversation takes place within a social context that defines it. Speakers create spontaneous discourse from restricted characteristics depending on that context. Alternatively, the term may be used more loosely as a synonym for spoken encounter/talk. In order to avoid this latter connotation, that may be considered too trivial, authors such as Schegloff (1999) tend to use more neutral terms such as 'talk-in-interaction', which surely is an inclusive terminology, yet not all types of talk fall within this category. Included in this class are very different types of 'oral discourse', for example meetings, interviews, school presentations, and so on. Indeed, most of the communicative work of the main institutions of society (religion, family, education, economy, politics) is considered 'talk-in-interaction'. On the contrary, someone who is thinking out loud may be talking but is not participating in a 'talk-in-interaction', as he/she is not interacting with other people.

1.1.1 Standard English

Taking a closer look into English, which is the focus language of this dissertation, an important issue has to be explained. In English-speaking countries there is a specific variety of English known as 'Standard English' (SE). J.P. Gee (2011:2) states that SE is 'the variety of English that is held by many to be "correct" both in the sense that it shows no strong regional variations and that it is used widely in mainstream media and by public figures'. Dr. McDavid highlights the definition of the general term "standard variety", given by C.C. Fries: it is 'the variety used by those who hold positions of trust and respect and conduct the important affairs of a community' (Fries, in McDavid, 1968:561). In fact, Standard English is nothing more than a dialect, whose origins date back to the fourteenth century. Thanks to the growing economic clout of the merchant class in London, their East Midland dialect, which was used for public business, spread across the country. Later, it became the basis of the so-called "Received Pronunciation" or "the Queen's English" in the United Kingdom and gave rise to Standard American English – the US standard for the English language – too.

Standard English and Standard American English are two examples of how a standard variety might not be 'an ideal of which the non-standard varieties are degenerate versions' (McDavid, 1968:561). The theory advanced by McDavid is supported by two leading fundamental aspects. First, the standard and its non-standard varieties may not

share the same origins. Reporting the example given by McDavid, Northern English dialects influenced standard London English more than Cockney, the local non-standard variety of London, did. Second, the standard is not always a 'monolithic entity' (McDavid, 1968). Nations such as Italy and France recognize one upper-class variety as the standard (Florentine-roman and Parisian) whereas others include a number of non-standard varieties, an aspect that shows the division of their culture. The most obvious example is the USA. Standard American English and the non-standard varieties are one of the many manifestations of its diversity, whose origins can be found in the early history of the US. The colonies were independent foundations which were brought together into a Union of States only in 1789. This means that since the first "official" settlement, Virginia, was founded in 1607, the colonies have developed different systems, social stratifications, language differences, etc., leading to a nowadays strong local loyalty towards these conventions. Furthermore, there are many other factors that have been affecting the varieties of the standard language throughout the centuries, for example industrialization, urbanization, and immigration, All these events have operated differently, both in intensity and way, on the American communities.

It can be said that people can perceive Standard English from two points of view: one as an actual variety of language and the other as an idealized norm of English, acceptable in many social situations. SE is certainly used in most public discourse and American social institutions; the new media, the government, and a number of other figures such as teachers all consider it to be the proper mode of communication. Studies (Dose and Gross, 1994; Haas, 1982) have demonstrated that people, even those who do not speak Standard English, consider it superior to other dialects. They believe 'good/correct grammar' and 'proper mechanics' equal to language correctness, that is they claim the existence of a single standard for language. On the contrary, scholars such as Collins (1991) argue against this 'misbelief'. The truth is that people speak it in different ways, just as they do with other varieties of English. That is, a speaker brings to Standard English his/her own linguistic influences, that may come from other dialects or even languages he/she knows and speaks. Following this logic, Collins (1991), Milroy and Milroy (1991), and other authors think it is more useful to speak of a Standard Ideology instead of a standard language, which 'stresses the importance and superiority

of the standard, "literate" or "unaccented" variety of English' (Wiley and Lukes, 1991:514). Moreover, Milroy and Milroy (1991) demonstrated that through Standard Ideology, people not only elevate one variety of language to be the standard, but they concurrently stigmatize other varieties, meaning that people themselves can be affected by the variety of language they use.

1.1.2 Context

Whether it is Standard English, or a dialect spoken by a minority, language is part of a process of communication which is culturally influenced. Communication is intrinsically multimodal: there are various aspects intertwined and the verbal element is only a part of it. Indeed, both verbal and non-verbal cues allow listeners to interpret a message fully. When people communicate, they do not just exchange information but rather interact in a cultural context. According to Gee (2011:6), context 'includes the physical setting in which the communication takes place and everything in it; eyes, gaze, gestures, movements; what has previously been said and done by those involved in the communication; any shared knowledge those involved have'. Another definition of context comes from Van Dijk (2006:163), who states the following: 'contexts are not "objective" or "deterministic" constraints of society or culture at all, but subjective participant interpretations, constructions or definitions of such aspects of the social environment'. From Van Dijk's point of view, the various contexts that exist within any communicative action, and which represent relevant aspects of situations and society, function as an interference, whether positive or negative, in the mental processes that take place in both production and comprehension of communication. Van Dijk goes on stating that contextual factors, such as roles, positions, social relations of participants, goals, intentions, etc., can 'control' our mental operations only when we conceive to these factors as relevant external features. This means that it is not the 'objective' aspect that defines context but how people interpret and use such constructions.

One characteristic of context that goes by the name of 'shared cultural knowledge' is pivotal to a correct interpretation and understanding of communication. It can be defined as all the knowledge, assumptions, and inferences that an entire community take for granted. That is, people belonging to a specific community see certain aspects of the context they live in as 'normal' and 'natural'. Recalling the example given by J.P. Gee,

'shared cultural knowledge' can be explained as follows: people in a determined cultural group assume that dinner time is within a certain range of time (for example 6-9 p.m.). This range of time equals to dinner time for that specific community, but it will be different for another community, which may be used to different traditions. This means that, more generally speaking, outsiders, i.e. people who do not belong to a certain cultural group, may not understand what insiders, i.e. people belonging to the cultural group, assume is the norm and therefore take for granted. Thus, in simple communication, and more specifically in discourse analysis (which will be described more extensively in the next chapter), what speakers say and the context in which they speak should not be taken for granted, because things may be 'old' for someone but 'new' for others.

1.2 Language in Television Series

As already stated in the previous paragraph, language use in the entertainment system, particularly in fictional television and cinema, is called 'non-authentic language' and it noticeably differs from 'authentic language', term used to define spontaneous conversations. Film/television use of language is deliberately non-spontaneous, as it is made up. It is an ensemble of several factors that contribute to its non-authenticity.

The main difference between the two types of language use can be found in their level of predictability. Several studies have shown evidence that this opposition is clear though in the last years screenplay writers have been trying to recreate spontaneous speech in fictional dialogue. Taking as reference the experiments conducted by the University of Trieste (see Taylor 2004), the results of such investigations show that fictional language, especially in stylised genres (science fiction series, westerns, etc.), is used in a much more crafted and predictable way than authentic language. Moreover, and the following aspect is central in this dissertation, the same experiments show that several elements such as discourse markers, which are typical of spoken language, display different frequency levels between the two types of language use: they are much more used in authentic language than non-authentic.

One underlying factor which defines the predictability of fictional language is the fact that any product belonging to television and cinema is always a 'team effort', as Taylor (2008) defines it. Indeed, a large number of figures is included in the creation of film scripts: starting from screenplay writers, adaptators, subtitlers (working with the

script); actors, dubbing actors (with the task of portraying characters as naturally as they can); even editors, producers, cameramen are involved in this process. According to Taylor (2008:168): 'the language of film is a scripted construct created by writers, subsequently altered by directors and actors, in the creation of an "artificially produced situation" (APS)'. This 'artificially produced situation' translates into several outputs, depending on the figures involved: for the writers it is the original script of a movie/episode; for the actors it is the artificial situational context in which they find themselves, attempt to 'live', and try to create a much more spontaneous situation.

Television itself is made up by various features which characterize the language used in media. Authentic and non-authentic language certainly share a number of properties, yet they differ in the distinctive details. The main properties of non-authentic language, according to Bednarek (2010), are the following: communicative context; multimodality; code of realism; nature of characters in Television.

The communicative context of Television has been described by Spitz (2005:22) in relation to scripted dialogue. Spitz states that: 'scripted dialogue is very different from ordinary, naturally occurring conversation with respect to its communicative context. It is not spontaneous talk but pre-planned, construed dialogue that is controlled by its author(s).' In turn, Bubel (2006) chose a different approach, describing the relationship between the audience and what they are watching. She analysed 'screen-to-face discourse' using a model that considers the typical elements of Television context, which are: the multiple authorship of scripts (involving screenwriters, actors, directors, etc.); the audience; and the cognitive processes involved in the interpretation of the dialogues. This model draws both on cognitive and sociological theories, meaning that the audience are considered both participants and overhearers. In other words, as indicated in Bubel's model, the communicative context of a television product involves 'world knowledge' both on the production team's and on the audience's part. With 'world knowledge' Bubel describes the common ground of knowledge that is at the basis of any script draft that can be interpreted from the target audience.

Another important factor when describing Television Language is the code of realism. The term has been defined by Bubel (2006:43) as follows: 'The imitation of reality, which holds not only for language but for all elements of the film text, is typically

achieved with the help of specific filmic conventions'. Bubel goes on stating that although realistic dialogue characterizes modern Television products, there are still some common conventions to be followed in terms of staged dialogue, which the audience have started to recognize as clichés of all Television products. In turn, when other elements such as discourse markers or utterance prefaces, which are typical in spontaneous conversation, appear in scripted dialogue, they function as a way of making the script less simulated and more natural.

All products made for television or cinema are multimodal. This means that they involve, apart from language, other elements, called semiotic modes, which cooperate to provide a meaningful whole, specific for that situation. Communication indeed 'occurs in tightly bundled clusters of data that bombard the senses' (Esslin 2002), and only some of these data are linguistic. Examples of semiotic modalities are written and spoken words, visual images, music, colour, etc. Multimodality can be seen not only in the whole product of television, but in the single aspects as well. For example, the characters imagined in the script of a television dialogue become more and more detailed and multi-faceted as the 'body and voice [of the actor] are themselves the medium through which skill is expressed' (Mills 2005:73). In particular, characters are one of the most crucial parts of fictional stories. Their nature can be seen from two different points of view. On one side there is professional practice, on the other there is media and television studies.

1.2.1 Genre

The aforementioned aspects are doubtlessly central in the making of scripted dialogue and any other fictional product. However, one element is left out. Genre, which is shared by both authentic and non-authentic language, is pivotal to the complete definition of what can be found in the fictional sphere. As a matter of fact, people use language in particular ways, chosen according to the situation. This includes the interactions we have with other people, both in spoken and written form, and of course the content and purpose of the genre which the discourse belongs to.

The term genre has been described by different authors. Richard and Schmidt (2002:224), define it as: 'a type of discourse that occurs in a particular setting, that has distinctive and recognizable patterns and norms of organization and structure and that has particular and distinctive communicative functions'. Examples of genres from this point

of view are news reports, casual conversations, etc. Another definition of genre comes from Martin (1984:25), who defines this term as 'a stage, goal-oriented, purposeful activity in which speakers engage as members of our culture'. Martin and Rose (2003:7) add to the previous definition a more specific explanation. They state that genres are social because people participate in them with other people; they are goal-oriented because people use them to 'get things done'; they are staged because there are different steps to be taken in order to achieve our goal.

While the previous definitions take into account genre as a 'type of discourse' or 'activity', other authors define it in different terms. For example, Neale (2001) uses the word 'category' and 'class', stating that genre is 'marked by a particular set of conventions, features and norms'. In common language, people see the different genres as categories in which one can include a number of works (films, books, poems, etc.) that share the same characteristics. In Television and cinema, genres include their own style of storytelling, dialogue, humour, performance, visual effects, etc. Examples of genres distinguished by the content are sitcoms, action series, science fiction series, detective series, fantasy series, mystery drama, soap drama, etc.

Genre classification clearly delimits the limits of different types of works, though these boundaries can be very dynamic. There are, indeed, 'no uniform criteria for genre delimitation – some are defined by setting (westerns), some by actions (crime shows), some by audience effect (comedy), and some by narrative form (mysteries).' (Mittel 2004:173). This means that genre boundaries are loose, and television in particular frequently conflate different generic elements, creating hybrid genres and favouriting genre embedding, i.e. when one genre is used for a different one.

1.2.2 Uses of the Term "TV-Series"

In the world of Television production, aspects such as context or genre are not the only ones that can define the limits of what we see on the small screen. One of the most basic classifications comes from broadcasting networks – and in recent years from streaming platforms too. The networks choose the target audience they want to reach and what they want to broadcast on their channels/platform in terms of formats or medium. Film and Television genres and sub-genres can be listed under three main formats: animation (puppetry, animated series, stop motion, etc.); live-action unscripted, which

includes informative programs (documentary, cooking show, talk show, news programs, etc.); live-action scripted, which includes basically all the fictional shows and divides into drama or comedy (and then all the sub-genres).

For the purposes of this dissertation, scripted live actions are of great interest and will provide the data analysed in Chapter three. As stated before, live-action scripted is a format that includes two major narrative fictions, namely drama and comedy, which themselves are divided into numerous sub-categories, such as medical, legal, political drama; or mockumentary, romantic comedy, sitcom (short for Situational Comedy).

All the fictional products listed above are called in common language "TV Series", but in more technical terms there is a specific distinction between a television series and a television serial, although nowadays many fictional programs fall between a series and a serial. A Television series usually 'feature the same characters, theme and settings. […] Each episode is relatively self-contained although there are occasional cliff-hangers across episodes and characters have a memory of what has gone on in previous episodes.' (Bednarek, 2010). A television serial, in turn, is more meta-narrative oriented, meaning that storylines stretch over various episodes, the narrative is very open-ended, and 'will go on for as long as audience interest and advertising support endure' (Huisman 2005: 154). Following the boundaries of genre, the series can be considered continuous and never-ending and let viewers decide which episode to watch; viewers are able to follow the general story even without watching every single episode. Alternatively, the serial develops storylines that last more than one episode and therefore it involves a certain level of loyalty from the viewer. Important changes came with modern definitions; as the differences between series and serials fall, they mix up and create hybrid formats and new sub-genres.

In the last few years, streaming platforms – more than Television companies, broadcasting networks, etc. – have been producing and broadcasting thousands of TV series and serials. The result can be seen in the expansion of an already mixed-up catalogue of products, so much so that new terminology was created: terms such as "TV addict" (i.e. someone who is obsessed by Television products and watches them as much as possible) are now part of mass/pop culture. Television series are popular because they combine familiarity and novelty. As Esslin (2002) writes, they provide a 'sense of

security', that is there is no need to make 'an intellectual effort' every time a new episode comes up, because the recurring characters and their status will end up being familiar to the audience, yet they will appear each time in new situations and dynamics.

The 'sense of security', their impact on different communities thanks to the huge range of topics they deal with, together with the increasingly high number of TV series existing, all make this phenomenon subject to social and linguistic studies, which may analyse single features or whole aspects of such entertaining products. For the purposes of this dissertation, TV series will be analysed from a linguistic point of view, more in particular the analysis will be about specific elements called discourse markers, which are generally typical of authentic language rather than non-authentic. Language use in Television products is not the exact transposition of language use in real life. For example, how many times, watching one of the various TV series about a group of friends, has one thought "I would never say it like that!" or "How come they never say swear words?". Questions like these lead to the realization that a Television script is a crafted element and therefore it will never become authentic (meaning real-life language).

This chapter has covered authentic and non-authentic language, defining their characteristics. Now that these basic notions have been described, the focus will shift to discourse analysis and conversational routines, both extensively described in chapter two. Discourse analysis is a tool used by analysts and scholars to answer to a variety of questions. For example, one can be interested in conversational routines (as it is the case of this dissertation), which are fixed expressions typical of authentic language: they are the element that gives the impression of naturalness. All these aspects will then be useful in the third chapter, in which the analysis of scripted dialogue will give practical evidence of what was stated in the theorical part of the dissertation.

Chapter two

Discourse analysis and inserts in conversation

Chapter one focused on the dichotomy of language, which can be 'authentic' or 'non-authentic'. Authentic language refers to language use in 'talk-in-interactions' (Schegloff, 1999), i.e. spontaneous discourse created by speakers depending on the context. Non-authentic language, in turn, refers to a more predictable language use which is typical of products belonging to the Entertainment system (e.g. fictional television, cinema, etc.). The level of predictability is the key characteristic which can show how different the two language uses are. One reason is that the grammar of a spontaneous conversation includes a great number of real-time elements, for example ellipsis or inserts, which speakers do not plan on producing. On the contrary, scripted dialogue is a crafted production of discourse and by its nature it could not possibly include such constituents; even though in recent years the figures behind scripted productions (screenwriters, etc.) are becoming aware of the fact that their works need to be less constructed in order to appear closer to real-life conversations. One device which could lead to an almost natural discourse production is the use of real-time elements such as inserts, typical of authentic language. Biber (2002) distinguished eight major classes of inserts: 'interjections (*oh*), greetings/farewells (*hi*), discourse markers (*well*), attention-getters (*hey*), response-getters (*okay*?), response forms (*right*), polite formulas (*thank you*), and expletives – usually taboo words (*damn*).' (Biber et al., 2002: 449).

Among these classes, the focus of this dissertation will be given to discourse markers (DM). This chapter will first describe discourse analysis and then will go on to introduce conversational routines and more particularly discourse markers, listing properties and functions of such elements.

2.1 Discourse Analysis

Discourse analysis is not an easy concept to describe. It is a tool which can be used to answer a multitude of questions, depending on what analysts, students, or researchers are looking for.

From a theoretical point of view, some authors conceive of the term discourse analysis as a synonym for other approaches. For example, Tannen (1989: 6) stated that: 'the name for the field "discourse analysis" says nothing more or other than the term "linguistics": the study of language'. Certainly, at the basis of discourse analysis there is 'the study of language' – intended as talk/communication – but the various approaches that might be used to analyse language are to be distinguished. What makes discourse analysis different from other methods can be found in the words that make up the term: 'discourse' and 'analysis'.

First of all, what does 'discourse' mean? It can be referred to as 'actual instances of communication in the medium of language' (Johnstone, 2002: 2). It is important to highlight Johnstone's use of the term 'communication' instead of 'language' because there are forms of interaction other than language, called 'semiotic elements', which are interconnected to language, hence they are fundamental in the understanding of discourse. Examples of such elements are photography, gestures, music, etc. In addition to spoken communication, there are other types of language use, such as written communication or manual languages (Johnstone, 2002) such as ASL (American Sign Language). Given that, one question that may arise is "why discourse analysis and not language analysis?". Analysis about language conceive of language as an abstract system, whereas analysis about discourse 'focus on what happens when people draw on the knowledge they have about language […] to do things in the world' (Johnstone, 2002: 3). In other words, discourse analysis studies language intended as an array of rules/relationships but focuses on the actions made by users of a language, for example expressing feelings, exchanging ideas, etc., which are based on 'memories of things said, written, heard, seen' (Johnstone, 2002: 3).

Usually, "discourse" is a mass noun, but some researchers might use the term as a count noun. t is the case of the scholars influenced by Foucault (1977, 1980, 1990), which define "discourses" as 'conventional ways of talking that both create and are created by conventional ways of thinking' (Johnstone, 2002: 3). Expanding on the creation of 'ways of thinking' one can say that discourse used as a count noun (discourses) is nothing more than ideology, i.e. a system of ideas and ideals. Drawing on this meaning, discourses influence and are influenced by ideas and are used to circulate power in society.

As stated before, discourse analysis focuses more on the actions related to language than on language itself, seen as an abstract. This specific point of view is useful in order to define the term "analysis" in relation to "discourse". According to Johnstone, analysis can be done following different approaches: it could mean systematically asking a number of questions or performing a variety of tests. All forms of analysis involve scholars and analysts asking questions and then finding answers to those questions. In the case of discourse analysis, the questions asked are often interdisciplinary, meaning that they are shared by discourse analysis and other studies (e.g. studies on communication, identity, etc.). In order to answer those questions one must analyse discourse in a specific way which is by 'examining aspects of the structure and function of language in use' (Johnstone, 2002: 4). When discourse is analysed, there is usually a process of breaking it down into parts or functions, according to participants, settings, processes, etc.

2.1.1 Data used in Discourse Analysis

Now that we have a clear definition of what discourse analysis is, it is just as much important to define the data analysed. The material subject to discourse analysis is often referred to as "texts", and it includes real-life occurrences of communication. The data collected is always in the written form but can be about non-written interactions. Analysing discourse in real time is very difficult because a single viewing or listening is not enough to allow a complete and accurate analysis. That is the reason why analysts are provided with transcriptions of the interactions involved in the study. The correct term used to indicate the above-mentioned transcriptions is "records". In addition, by doing such activity, non-written discourse is provided with characteristics that are typical of written texts such as boundaries or a fixed structure, and most of all it becomes a physical object.

These so-called "texts" are the result of both inclusion and exclusion of transcriptions – better of interactions in general. For example, as Johnstone states: 'a text might be one discussion or a whole complete series of television debates' (2002: 19). Actions such as including or excluding texts are at the basis of every type of approach to social-scientific research. They allow analysts to draw boundaries and collect only the data needed, thus making the analysis more effective and clear. The subjective aspect of such actions – the fact that is the analyst who decides what to analyse from the

transcriptions – is one element which highlights the diversity in the studies. For example, the same text might be excluded from one study but included in another and the result will thus be different.

Another element which increases the variety of analysis and studies is the style with which a transcription is made. Usually, 'standardized transcription systems' are used in the field of discourse analysis. Nonetheless, no single transcription system is suitable for all purposes and therefore different styles and criteria are employed to achieve different goals. What is important is that any record must be accurate, i.e. 'include what it claims to include' (Johnstone, 2002: 21). Texts are useful to researchers and analysts only if they match with the interest point of the research/analysis and do not include any additional detail that might be a distraction in the developing of the study. The question of how much detailed a transcription needs to be affects speech as much as speakers, which may be portrayed more or less accurately. There are two main points of view about this aspect: those who state that a detailed representation is important and that analysts should take every information as new, and those who think that a literal representation might negatively affect the speaker.

After choosing the most suitable transcriptions, analysts may proceed by gathering these texts into so-called "corpora". A corpus is an organized set of texts with shared characteristics (topic, register, etc.), employed in linguistic analysis. For the purposes of this dissertation, two corpora will be used in order to analyse authentic and non-authentic language. One corpus is the "Santa Barbara Corpus of Spoken American English", and it includes a number of transcriptions of spontaneously occurred conversations across the USA. It was compiled by researchers in the Linguistics Department of the University of California, Santa Barbara. The other corpus includes transcriptions of episodes from the famous TV Series "Friends", and it is the result of personal research, as I could not find corpora on the topic.

2.1.2 Genre and Register

Texts of any kind share certain characteristics with some texts and differ from others. This aspect is called "intertextuality" and describes the connection between texts in relation to both prior works of the same kind and shared elements, for example how they were written (style, medium, etc.). The intertextuality of texts can be analysed from

different points of view: one is by focusing on repetition in discourse, i.e. the way users of a language borrow, adapt and re-use structures and forms from their previous interactions. Another way to analyse intertextuality is by focusing on register and genre. These two terms are often used as synonyms, but they do not describe the same aspect of discourse. Johnstone (2002) defined register as 'a variety of language (or "style") associated with a recurrent communicative situation or set of communicative roles' (2002: 158). When describing a register, one is automatically describing the situation in which that specific register is requested, and its linguistic properties. In turn, genre is described as 'a recurrent verbal form associated with a recurrent purpose or activity' (Johnstone, 2002: 155). Nowadays, the term 'genre' refers to the categorizations of different types of work (either written or spoken) which share the same properties. The term 'genre knowledge' describes the competence of producing a form which conventionally falls into a certain category/genre. There are specific situations which require "stabler" genre, such as writing a haiku or a sonnet.

When analysing discourse, intertextuality among texts involved in the study is a key-aspect. By its own definition, a corpus includes texts that share characteristics such as genre, register, topic. For example, the corpora analysed in this dissertation share several features: they both include transcriptions of verbal interactions; such interactions happen between friends and family, thus an informal register is in use. The same description can be done on one corpus at a time. The 'Santa Barbara Corpus of Spoken American English' includes transcriptions of naturally occurred interactions, whereas the "Friends: the TV Series Corpus" includes transcriptions of fictional conversations. A proper analysis on authentic and non-authentic language use can be done thanks to both similarities and differences between these corpora. The intertextuality of the spoken interactions analysed allows a comparison on repetition in discourse, focusing on discourse markers, elements typical to spoken language, which are described below.

2.2 Conversational Routines

Within the framework of discourse analysis, studies have shown that there is a variety of phrases, frequent in spoken language, which authors have named "Conversational Routines". Conversational routines have been defined by Leech (1983: 28) as 'phrases which, as a result of recurrence, have become specialized or "entrenched"

for a discourse function which predominates over or replaces the literal referential meaning'. Conversational routines are extremely common in spoken language, and they are best known for their central role in (first) language acquisition and in the learning of a second language.

From a grammatical point of view, a routine can be made up by a single word, several words, or even quite long phrases. In addition, according to Bahns et al. (1986: 695), they can be analysed syntactically as sentences, (*how do you do*), subordinate clauses (*if I may say so*), noun-verb structures (*you know*), non-sentences (*next please*). They can even be discontinuous with a lexical slot (*as far as /.../ concerns*). Routines do not only occur alone, but they can be juxtaposed or repeated. There are three processes from which conversational routines may be created:

1. Lexicalization: linguistic process through which single words are combined to create new wholes. The new concept then must be "culturally recognized" as a conversational routine, i.e. speakers/users need to be familiar with it.

2. Grammaticalization: process involved with "decategorialization", i.e. the creation of a subcategory such as adverbs.

3. Idiomatization: words receive a meaning which is different from the meaning of the single parts. It can be strong or weak, and not all sequences of words which are lexicalized become idomatized.

Analysed in terms of the situation in which they are used, conversational routines can be grouped into several classes. According to Alexander (1984), there can be "connectives", whose function is to contribute to the cohesion of the discourse, or "conversational gambits", which serve as openings to the conversation (*So, what do you do?*). The following are just some functions attributed to conversational routines: they facilitate the transition to something new in the discourse; signal a digression; organize different aspects of the topic; and more generally they add to the structural coherence of the discourse. Furthermore, conversational routines cannot be defined in terms of compositional rules, as there can be functions shared by conversational routines belonging to different semantic sources (*let's face it; frankly*; *etc…*). The result is that these routines have one common pragmatic function, i.e. the meaning that the speaker wants to convey, that overlaps more or less dominantly with the referential meaning, i.e. the literal meaning of the expression.

Connected to the pragmatic information of routines are "frames", regarded by Aijmer (1996: 27) as: 'hypothesis about speakers' stereotypic knowledge of a situation and how this knowledge is organized in the long-term memory'. This knowledge is organized as frames due to the fact that during a conversation many situations tend to recur, leading to the repetition of linguistic behaviour, too. As a result, when a certain situation recur the long-term memory remembers which conversational routine was used before and the speaker uses it automatically another time. This association between social situation and formulae is useful for the analysis of routines in different cultures and different context.

The focus of this dissertation is to analyse a specific type of conversational routines, namely "Discourse Markers", a term coined by Deborah Schiffrin in 1987. The discourse markers analysed in the third chapter may be defined as routinized elements that contribute to the coherence of discourse. In actual fact, they occur in spoken interactions where speakers automatically repeat a linguistic behaviour, and they function as markers of conversational flow.

According to Aijmer (1996), there are two types of cohesive devices with the function of marking the progression of discourse. One option is to use prefaces. The speaker 'tells the hearer explicitly what he is going to do or how the different parts of the text are structured in a metalinguistic statement' (Aijmer, 1996: 202). Striking evidence of the use of prefaces come from literary works. For instance, in "Adventures of Huckleberry Finn", Mark Twain writes a preface in which he tells some background information about the characters and why one should read the book. The second option is to use discourse markers, which can refer backwards and forwards in the discourse context and are oriented to the speaker and/or the hearer. There is little agreement about what to account as "discourse marker" and what the different types of discourse markers have in common.

2.3 Discourse Markers

Before Schiffrin coined the term 'discourse marker' (DM) in her 1987 book *Discourse Markers*, several labels and definitions have been given to these devices (e.g. pragmatic markers, discourse connectives, etc.), but studies have failed to give them a proper description. The work conducted by Schiffrin in 1987 can be considered one of

the first studies on the matter, which resulted in an extensive description and characterization of discourse markers. She defined DMs as 'sequentially dependent elements which bracket units of talk', claiming that they 'do not easily fit into a linguistic class.' (1987:31). Schiffrin operates at a certain degree of vagueness because discourse markers are influenced by several types of units of talk. Hence, including a large variety of DMs allows a more effective and precise definition of such elements. In everyday use, speakers express their orientation toward a proposition, interjection, etc… or they organize actions and information using elements that can become markers of other discourse components. Schiffrin (1987: 328) lists some specific conditions through which an expression becomes a discourse marker:

1. It has to be syntactically detachable from a sentence.
2. It has to be commonly used in initial position of an utterance.
3. It has to have a range of prosodic contours.
4. It has to be able to operate at both local and global levels of discourse (this aspect will be described in paragraph 2.3.2).

The approach chosen by Schiffrin (which is sociolinguistic) is not the only one applied to define and label DM: other definitions come from scholars such as Fraser, who studied DM from a grammatical-pragmatic perspective and stated that they 'signal the speaker's intended relationship between the segment and the preceding one' (Fraser, 2009: 87). Blakemore (2002), in turn, was influenced by relevance theory. The principle of relevance theory is that the context in which an utterance is found clarifies its message (Sperber, D. and D. Wilson, 1986). Analysing DM from this point of view means conceiving of them as metalinguistic elements, that is parts of the utterance whose functions are to reformulate, add to, summarize, clarify, a previous utterance. Jucker (1993) uses the term 'signpost' in order to define the main function of DM, which is to display that a speaker understood previous talk and/or discloses future talk. In other words, DM allow a clear flow to the conversation, incorporating utterances and explaining their interpretations. All these different approaches and labels may find a common ground of agreement in the definition given by Carter and McCarthy: 'DMs function to organize and monitor an ongoing discourse, between one topic and the next, by indicating topic changes or by bringing a conclusion to the discourse. They also function to mark the state of knowledge between participants.' (Quaglio, 2009: 80).

2.3.1 Linguistic and Grammatical Features

According to Fraser (2009), the majority of researchers classify DM into three main groups, though the common features of each single class are still up for debate:

1. Contrastive discourse markers (*but, however*), which show a contrast between two sentences or segments.

2. Elaborative discourse markers (*and, in addition, well*), which expand information from the first sentence/segment in the second.

3. Inferential discourse markers (*so, thus*), which reach a conclusion.

Given the fact that discourse markers generate so much debate concerning their features, even from a grammatical point of view, scholars and authors disagree. Schiffrin states that 'understanding discourse markers requires separating the contribution made by the marker itself, from the contribution made by characteristics of the discourse slot in which the marker occurs.' (Schiffrin, 1987: 73). This translates into several questions regarding discourse markers, their grammatical status, semantic meaning and how these features interfere with the functions and interpretations of DM.

Another perspective comes from Fraser (1999), who, with his grammatical-pragmatic work on discourse markers, proposes three main counterarguments to the definitions given by his fellow authors:

1. What do discourse markers connect? They may connect not only a sentence to the prior one, but also to prior and following segments. Therefore, they do not have a positional fixedness, meaning that they can occur at the beginning of the first sentence, between two sentences, or at the end of the second one.

2. What grammatical status do they have? There are three main stems which may generate a discourse marker: conjunctions, adverbs, and prepositional phrases. A stem is a collocation with a lexicalized pattern containing a grammatical slot, which can be filled with lexical material. (Aijmer, 1996).

3. What is their semantic meaning? Each discourse marker has a core meaning which is mixed with the context so that it achieves an additional interpretation.

A further significant issue in the grammatical description of DM is the 'discourse marker slot'. In 1975, Andersson provided a model, which is nowadays considered a framework in the study of discourse markers. The model, displayed in fig. 1, separates

the message itself, called propositional core, from lexical material, that consists of speech-act adverbials such as discourse markers, with the function of commenting the sentence. Andersson used the term 'lexical material' to refer to all elements in the discourse marker slot, which can be recognized on two linguistic levels: semantically, because they 'have at least a component of meaning that resists truth-conditional treatment' (Levinson, 1983); and prosodically, because between the discourse marker slot and the sentence, usually, there is a tone break. In many cases, the DM slot is filled by a so-called "cluster of markers", that is a combination of different discourse markers.

Building on Schiffrin's characterization of DMs, the model indicates that the discourse marker slot only goes in initial position. In turn, drawing on the counterarguments provided by Fraser, discourse markers may occur initially or finally, thus the figure functions to display that discourse markers are not part of the utterance, but rather an external element through which one can interpret/express information.
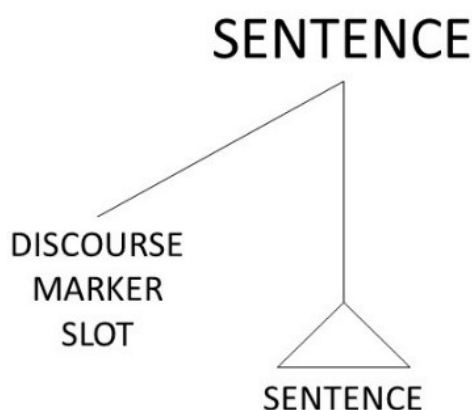
Fig. 1: the Discourse Marker Slot

2.3.2 Functional and Contextual Properties

From a functional point of view, discourse markers can be divided into two categories: global DM and local DM. Global discourse markers are used on a more general level, mainly to guide the hearer through the communication. They serve as organizers of topics, for example when the speaker makes digressions or corrections (*by the way*), adds a point (*another thing*), summarizes a discussion (*to sum up*). Another important role attributed to global markers is that of commenting on the relationship between two larger discourse units. In turn, local discourse markers are largely used to

facilitate the flowing of connections. In particular, they comment on the occurrence and importance of new information given by such connections (*actually, in fact, the point is, after all*). According to Aijmer (1996), local DMs have a pragmatic function, that is of operating on the validity of the new message, such as emphasising a certain point of view rather than another.

From a contextual point of view, Schiffrin (1987) identified five planes of talk on which DM are used: 'exchange structure; action structure; ideational structure; participation framework; information state.' (1987: 325). Drawing on this distinction, Verdonik (2008), studied the impact of context on discourse marker use. Indeed, from her analysis, several 'contextual factors' have been highlighted, such as: number of participants and their social relationships; topics of conversations/opinions on the topic; channels of communication; general goal of communication or participants in communication. The study conducted by Verdonik 'considers DMs as pragmatic elements conveying no or minimal propositional content' (2008: 760). Schiffrin (1987), in turn, states that the 'communicative effect' (1987: 317) of a discourse marker is the result of both its semantic meaning and syntactic properties. Discourse markers occur in specific positions which determine their interpretation in the discourse. They are capable of conveying any potential interpretation, it is then the context in which they occur that constrains the meaning conveyed by discourse markers. For example, if one extrapolates two utterances from their context, the relationship between the two can be interpreted differently, depending on the discourse marker that one decides to use. But when the two utterances occur within a certain context, the discourse marker used has the sole function of displaying that relationship.

Utterances, by their very nature, are context-bound: 'they are presented by a speaker to a hearer at a certain time and in a certain place' (Schiffrin, 1987: 322). Speaker, hearer, time, and space are aspects of context, i.e. situational parameters which can be often encoded through 'deictic elements' (personal pronouns, temporal and locative expressions, etc.). In the case of discourse markers, the context in which they occur is exemplified through the use of discourse parameters rather than situational. Discourse parameters refer to either participants (speaker or hearer) or text (prior or upcoming text). As for the latter, DMs functioning as deictic 'pointers' (Aijmer, 1996) may refer

23

backwards or forwards in the conversation. Depending on the context in which they occur, DMs may not fall strictly into one category. For example, "*the point is*" or *"I mean"* may refer to upcoming messages or old information respectively. DMs like *listen, look, now* point only forwards, whereas *right, okay, oh* are use more as backwards pointers. In the former setting (parameters referring to participants), DMs are used to signal the participant's attitude towards the conversation and other participants. Therefore, studies such as those of Aijmer (1996) and, before her, Schiffrin (1987) DMs expressing personal deixis have been classified into different categories, depending on the orientation they convey:

- speaker-orientation (*I mean, in my opinion*).
- Hearer-orientation (*you know*).
- Orientation to speaker and hearer (*let's face it*).
- Orientation to a third person (*as X says*) – Despite Aijmer, Schiffrin doesn't include this category.
- No speaker/hearer orientation (*the fact is*).

Grammatical and contextual aspects of discourse markers form what Aijmer (1996) has called 'cognitive frames', i.e. units of information useful for the speaker to retrieve characteristics of DMs. Concerning this dissertation, cognitive frames include all the features that will be useful for the following investigation, which will analyse the frequency and the functions of a pair of discourse markers, *I mean* and *in fact*. The two DMs will be first described through the use of cognitive frames, and then analysed inside their context.

Chapter three

Data analysis

In this dissertation, both chapter One and chapter Two function as the theoretical part of the study on language use, which will be put into practice in chapter Three. The literary review in the aforementioned chapters focused on past works in the field of language use (Schiffrin, 1989; Aijmer, 1996) and contributed to the introduction and definition of pivotal terms such as *language* and *non-authentic language*, *discourse analysis*, *discourse markers*.

Drawing from the notions defined in the previous pages, the dichotomy between authentic and non-authentic language is significant to the developing of the analysis in this chapter. It has been said that authentic language refers to language use in real-life interactions, whereas non-authentic language refers to crafted productions of conversations typical of the entertainment system (e.g. scripts of movies or TV series), and by their nature they lack spontaneity. In turn, when spontaneously communicating, people do not have the whole interaction organized yet, so they recur to specific elements such as discourse markers, which are primarily used to guide the flow of the conversation.

Recently, there has been a change in direction regarding the production of scripted dialogue. The figures behind scripts (screenwriters, producers, etc.) felt the need to use language in a much more natural way, thus including elements, e.g. discourse markers, which were once distinctive of authentic language only. This new turn in the creation of non-authentic language led to a new perception of it and is the starting point of this dissertation, whose aim is to explain the issue of authenticity in scripted dialogue. The most evident representation of authentic language in scripted dialogues comes from comedy series, where the use of informal language allows more natural elements. Among the infinite list of comedy series existing nowadays, I chose to analyse the famous American TV Series *Friends*, created by M. Kauffman and D. Crane. The Sitcom aired on NBC from 1994 to 2004, for a total of ten seasons. The main characters are a group of six friends (Monica, Chandler, Rachel, Ross, Phoebe, Joey) who live and work in New York City. Friends was nominated for and won several awards, becoming one of the most popular TV Series of all time.

In the next pages I will first proceed introducing the methodology involved in the discourse analysis, describing the corpora and the analysis tools used, and the specific elements analysed. Through discourse analysis I will answer questions such as "what is the frequency of DMs in authentic and non-authentic language?", "what meaning do they convey?", "Do they occur under the same conditions in both uses of language?". Then I will continue displaying and discussing the results of the discourse analysis.

3.1 Methodology

The methodology chosen for this study includes two corpora, namely the "Santa Barbara Corpus of Spoken American English" and the "Friends: the Tv series Corpus", whose texts are examined through the use of an analysis tool called AntConc. It is a corpus freeware with a large variety of tools for texts analysis and concordance. The subject of such investigation are discourse markers, which are conversational elements typically associated with naturally occurring interactions. Examples of DMs are *I mean*, *well*, *you know*, *so*, *like*, *anyway.* Words used as DMs acquire new meaning based on the context in which they occur. For example, the word *well* can occur as (elaborative) discourse marker, adjective, adverb (conveying different meanings) and therefore it is incredibly challenging to analyse when looking for specific aspects. As stated above, there has been a change in the way scripted dialogue is created: elements once distinctive of authentic language now increasingly occur in non-authentic language, too.

In order to study this issue and therefore show how authenticity in both language uses is generated, I will analyse the frequency of the following DMs: *I mean*, *in fact*. Drawing on the two main frameworks followed to map the theoretical part of this dissertation, Aijmer (1996) and Schiffrin (1987), I will now give a proper characterization of the discourse markers subject of my investigation. As starting point I chose the organization of DMs as cognitive frames, made by Aijmer (1996), to which I added Schiffrin's categorization of DMs use into planes of talk. Cognitive frames are 'units of information in the speaker's long-term memory' (Aijmer, 1996: 231), i.e. when a specific discourse marker occurs, the speaker recognizes it and thus remembers its features and how/when to use it. A frame includes information about formal features (grammatical and linguistic aspects) and situational features (i.e. contextual factors).

(a) *I mean*

Formal features	Situational features
Sentence external	Backward and forward orientation
Lexicalised phrase	Speaker orientation
	Local function
	Participation framework, information state, ideational structure

Table 1: cognitive frame of *I mean*

(b) *in fact*

Formal features	Situational features
Sentence external	Backward and forward orientation
Discourse marker	No speaker/hearer orientation
	Local function
	Action structure, ideational structure, information state

Table 2: cognitive frame of *in fact*

3.1.1 The Corpora Employed

This is a corpus-based study, i.e. material extracted from discourse analysis provide support, confirmation or verification to knowledge and expectations. Opposed to this method is a corpus-driven study, in which the results detected from the corpora are not based on previous information.

The first corpus is the "Santa Barbara Corpus of Spoken American English", which will be used in the investigation of discourse markers in authentic language use. It was compiled by researchers in the Linguistic Department of the University of California, Santa Barbara. I found this corpus online[1], while searching for a quite small corpus of spoken American English to analyse as an alternative to the "Corpus of Contemporary American English (COCA)", which is a one-billion-word corpus. The "Santa Barbara Corpus of Spoken American English" is made up of four parts (including both audio and transcriptions), for a total of approximately 240,000 words. It includes a large number of naturally occurring spoken interactions from all over the United States and depicts a large segment of society, representing people of different age, gender, social background, ethnicity, occupation, etc. The most represented form of interaction is face-to-face

---

[1] https://www.linguistics.ucsb.edu/research/santa-barbara-corpus

conversations between friends or family members, followed by telephone talks, on-the-job talk, and more.

The second corpus is the "Friends: the TV series Corpus", which will be used in the investigation of DMs in non-authentic language use. I collected the texts included in the corpus from a website which provided access to transcripts of all ten seasons (https://fangj.github.io/friends). I gathered those transcriptions in a Word document, I watched some episode to see if the transcriptions were correct and included every word, simultaneously checking for typos and lastly I deleted unnecessary data such as credits and copyright information. The "Friends: the TV Series Corpus" consists of transcripts from the famous TV Series "Friends", for a total of approximately 160,000 words. The spoken interactions are not spontaneous, as they are part of a script, though the actors of this particular TV Series had the opportunity to improvise. Some aspects connect this corpus to the "Santa Barbara Corpus of Spoken American English": both are based on American English; the prominent form of interaction is face-to-face conversations between friends and family members. Speaking of the people depicted in the "Friends: the TV Series Corpus", it can be said that characters are mainly young adults (in their thirties), with different occupations, quite different social backgrounds (e.g. some are poorer than others), and almost all are Americans.

Given the fact that the two corpora used are of different sizes – one including 240,000 words while the other 160,000 – the problem is how to relate the results for both corpora in order to give proper and accountable evidence of my investigation. To do so, I will normalise the results to a common base using a specific formula: divide the raw frequency by the total number of words in the corpus section and multiply the result by one thousand. For example, if the frequency of a given word is 120 in the 240,000 words corpus and 70 in the 160,000 words corpus, the normalized frequencies are:

- 120 ÷ 240,000 * 1,000 = 0,5 occurrences per thousand words
- 70 ÷ 160,000 * 1,000 = 0,45 occurrences per thousand words

3.2 Data Analysis and Results

In this paragraph I will provide results and evidence for the frequency of the three discourse markers "*I mean*" and "*in fact*", analysing them individually before reaching a

global conclusion. Table 3 shows all the normalized frequencies resulted from the investigation of the two discourse markers in both corpora:

	SANTA BARBARA CORPUS	FRIENDS CORPUS
I MEAN	**2.10** occurrences per thousand words	**1.70** occurrences per thousand words
IN FACT	**0.20** occurrences per thousand words	**0.06** occurrences per thousand words

Table 3: frequency per thousand words of the DMs *I mean* and *in fact* in the two corpora

3.2.1 *I Mean*

According to Schiffrin (1987) the semantic meaning of *I mean* highly influences its use as a discourse marker in conversations. As a discourse marker, *I* mean is used to explain what has been said before.

From a grammatical point of view, I mean is a lexicalised phrase and is sentence external: as all discourse markers it occurs in the so-called "discourse marker slot" with the function of commenting the sentence. From a contextual point of view, *I mean* is speaker-oriented, as it marks speaker orientation and focuses on the speaker's point of view. Indeed, it works on the participation framework because it involves the relationship between speaker and information. It also functions on the information state, since 'speaker orientation to ideas is related to knowledge about their content' (Schiffrin, 1987: 295). *I mean* refers backwards and forwards in the conversation: it draws on already existing information expanding its meaning. Finally, it operates locally in the discourse. Local discourse markers are closely related to pragmatic functions, i.e. they operate on the validity of messages.

(a) *I mean*

Formal features	**Situational features**
Sentence external	Backward and forward orientation
Lexicalised phrase	Speaker orientation
	Local function
	Participation framework, information state, ideational structure

Table 1: cognitive frame of *I mean*

Results show a high frequency of *I mean* in both corpora: 568 occurrences in the "Santa Barbara Corpus of Spoken American English", versus 285 occurrences in the "Friends: the TV Series Corpus".

3.2.1.1 *I Mean* in the "Santa Barbara Corpus of Spoken American English"

Out of 568 occurrences in the Santa Barbara Corpus, *I mean* is used as part of an utterance 30 times, it is not sentence external and thus cannot be analysed as discourse marker (example 1). Therefore, its frequency as a discourse marker lowers to 538, i.e. **2.10** times per thousand words, in the Santa Barbara Corpus. The investigation shows that *I mean* occurs at the beginning of an utterance in two cases: 1. It is the starting point of a second utterance, linked to the previous one, where the speaker clarifies previous information (ex. 2); 2. It is used by a speaker as a reply to another speaker's statement (ex. 3).

(1) **MARIE**: *I didn't mean* like half the things I said, just, just to like...show him. Do you know what *I mean*?

In this example, Marie uses the expression *I mean* twice. The first time in a negative form, the verb *to mean* is a synonym of the verb *to intend*. Here Marie states that the things she said have no real meaning to her. In the second case, *I mean* is part of a question asked in a rhetorical way by Marie to have some sort of feedback from her interlocutor: she wants to be sure that the hearer understood her real intentions. In both cases, I mean is part of a sentence and cooperates to its meaning. It is not sentence external, therefore not a discourse marker.

(2) **PATRICK**: we're at two totally different social levels. *I mean* it's easier for women to walk into a bar... not know anyone, and then come out knowing everybody.

In example (2), Patrick is telling his friend his struggle to meet girls at bars. He starts by saying that men and women are on different levels and then uses *I mean* to explain his idea. Here, the backward and forward orientation of *I mean* is evident, as it guides the

hearer through two sequences of discourse. The speaker develops information of a previous message in order to explain an idea.

> (3) **REED**: Alright, what are you doing over there. I thought you were doing something different. I thought you were working with languages in…
> **DARREN**: Oklahoma?
> **REED**: Huh?
> **DARREN**: Or in California.
> **REED**: *I mean*, well yeah, California, your basic project,

In example (3), two friends meet after a long time and Reed can't recall where Darren went. After Darren remembered him he was in California, Reed replies starting the conversation with *I mean*, which functions as a link to Reed's previous utterance, hence allowing the conversation to go on from where he stopped. Here, *I mean* forms a cluster of DMs with *well* and *yeah*. It focuses on the speaker's point of view on the topic and introduces new information from previous talk.

3.2.1.2 *I Mean* in the "Friends: the TV Series Corpus"

Out of 285, 10 occurrences will not be counted because *I mean* is part of an utterance and cannot be analysed as discourse marker (example 4). Hence, the normalized frequency of *I mean* is **1.70** occurrences per thousand words in the Friends Corpus. Results from the Friends Corpus show that *I mean* occurs at the beginning of an utterance, alone or part of a cluster of DMs such as *You know/Y'know*, *yeah*, *well* (example 5). In its sentence initial position, *I mean* is used by a speaker to expand on the meaning of the previous statement (example 6).

> (4) **RACHEL**: Ross, this is not how we wanted you to find out about this. You have every right to go nuts.
> **ROSS**: I'm not going nuts. Do you see me go nuts?
> **RACHEL**: No, but you know what *I mean*.
> **ROSS**: Hey, hey, hey... If you two are happy, then I'm happy for you. I'm fine!

In example (4), Ross finds out about the fling between Rachel and Joey and is talking about it with them. Rachel is worried that Ross will get mad (because once they were lovers) and is trying to make things clear. In this case, *I mean* takes part in the meaning of the utterance. It is not an external element commenting on the discourse.

> (5) **MONICA**: Well, do you love him?
> **RACHEL**: Sure.
> **MONICA**: Sure?
> **RACHEL**: Yeah, *I mean*, whatever.

Example (5) shows a cluster of DMs made up by *yeah, I mean*, and *whatever*. Here, Rachel answers Monica's questions in an uninterested way. In this example, Rachel uses *I mean* to expand the meaning of her point of view. She first says "yeah" which acts as a confirmation of the fact that she is sure about her point of view. Then, she clarifies with *I mean* (plus *whatever*) the fact that she really does not care about it.

> (6) **CHANDLER**: Yeah, yeah. Some people said some nice things about him. I think somebody should have it.
> **MONICA**: Oh, gosh, this is so weird. *I mean*, his whole life was in this apartment, and now it's gone. You know, I think it would be nice if we just took a few moments, for Mr. Heckles. *I mean*, he was kind of a pain, he was, but, he was a person.

In example (6), The gang is cleaning Mr Heckles' apartment, Monica's downstairs neighbour. Here Monica uses *I mean* twice. In the first case she says that being there is weird and then introduces with *I mean* the reason she thinks the situation is weird. Here, *I mean* is used to expand the meaning of prior information. Its orientation both backwards and forwards can be seen as the marker connects points already made with upcoming messages. Again, in the second case, *I mean* is used to develop what Monica wants to say. She wants everybody to be nice to the late neighbour and after saying so she uses *I mean* to explain why they should do so.

3.2.2 *In Fact*

In fact is a discourse marker which operates locally in the discourse, meaning that it comments on the validity of a message. Specifically, *in fact* is used 'to correct or contradict a previous speaker' (Aijmer, 1996: 223). In terms of orientation, *in fact* may be categorized by Schiffrin (1989) as oriented to both speaker and hearer. However, I believe that a much better classification comes from Aijmer (1996). Her category "no speaker/hearer orientation" is more suitable for this DM, as it includes markers such as "the fact is" and "more important", which share with *in fact* the local function of emphasizing previous topics. These discourse markers show 'no explicit reference to the speaker, hearer, or third person' (Aijmer, 1996: 221) *In fact* is oriented both backwards and forwards, as it reinforces old points with new information. It works on the action structure, as it acts as clarifier, on the information state, since it "involves the organization and management of knowledge and meta-knowledge" (Schiffrin, 1989: 28), and on the ideational structure, because of its cohesive and explanatory functions.

(b) *in fact*

Formal features	Situational features
Sentence external	Backward and forward orientation
Discourse marker	No speaker/hearer orientation
	Local function
	Action structure, ideational structure, information state

Table 2: cognitive frame of *in fact*

In fact occurs as discourse marker 44 times in the Santa Barbara Corpus and 9 times in the Friends Corpus. This discourse marker is less frequent than *I mean* but its investigation is as much important. The analysis will take into consideration the three main functions of *in fact*: clarifying/explaining, correcting, contradicting.

3.2.2.1 *In Fact* in the "Santa Barbara Corpus of Spoken American English"

In the Santa Barbara Corpus, *in fact* occurs 44 times and the normalized number is **0.20** occurrences per thousand words. Regarding its contextual functions, *in fact* is used 20 times to clarify prior points (ex. 7), 18 times to correct the message (ex. 8), and 6 times to contradict previous information (ex. 9). All occurrences show that *in fact* is used at the

beginning of an utterance, alone (ex.7, 8) or following other DMs such as *yeah*, *uhm*. In the latter case, it forms a cluster of DMs in the discourse marker slot (ex. 9).

> (7) **DANNY**: That's what the church is about, we're on a mission from God. *In fact*, let me clarify it a little more. Our mission is to participate in, and to carry out God's mission.

Example (7) clearly shows how *in fact* works both on the ideational structure and on the action structure. The speaker, Danny, explicitly says that he is going to clarify a point made in the previous utterance. He uses "in fact, let me clarify it…" as a device to manage the discourse and not only clarify prior talk.

> (8) **PAMELA**: I said… I wanna believe in Santa Claus. *In fact*, sometimes I do believe in Santa Claus. And that... that really satisfied her.

In example (8), Pamela is talking to her friend about a conversation she had with her daughter. She says that she wants to believe in Santa Claus but immediately corrects the message stating that she actually does. In this situation, *in fact* acts as a correction to a previous point made. This is typical of naturally occurring interactions, as the speaker tries to correct mistakes or wants to add points to his/her statements.

> (9) **MARILYN**: And she goes, do you…you don't mind, do you? And I said, well yeah, *in fact* I do mind. Cause I thought the lemon tree was dying.

In example (9), Marilyn is reporting a conversation she had with a neighbour, which took some lemons from Marilyn's tree without asking. The woman rhetorically asked her if she was bothered, and Marilyn replied that she "in fact" was. The DM *actually* could be used as a synonym in this situation since both act as clarifier and contradictory devices. Here, *in fact* occurs in a cluster of DMs (*well*, *yeah, in fact*) in sentence initial position. Its function is to contradict a previous statement, in this case specifically to answer a rhetorical question.

In fact occurs very few times in the Friends Corpus, only 9, marking a normalized frequency of **0.06** per thousand words. Given the extremely low number of frequencies, it is possible to give a proper analysis for each occurrence, highlighting the contextual characteristics (clarify, correct, contradict) conveyed in each example. *In fact* occurs in a sentence final position only once (ex. 10), while the remaining 8 examples show that it occurs in a sentence initial position (ex. 11).

> (10) **RACHEL**: Well, my boss was at the same restaurant where I was having my interview and he heard everything. So later he calls me to his office and he tells me that he's gonna have to let me go, because I'm not a team player. And I said "Wait a minute! Yes I am." and I had to sit there for 45 minutes while he proved that that *in fact...* was true.

In this example, Rachel is talking about how bad her career is going. Despite she believes she is a team player her boss proved her wrong. The discourse marker *in fact* occurs at the end of the conversation to introduce an opposition to Rachel's idea, hence it has a contradictory function. Still, it has no-speaker orientation and connects information backwards and forwards. As it can be seen here, *in fact* operates in the validity of discourse because it verifies the veracity of information (specifically in this example it contradicts the information).

> (11.a) **CHANDLER**: I'm Chandler. Hey, I was in the scouts too.
> **OWEN**: You were?
> **CHANDLER**: Yeah, *in fact*… my father was a den-mother.

Here, *in fact* is part of a cluster of DMs, as it occurs together with *yeah*, and is used as a reply. Owen is surprised to hear that Chandler was a scout and asks a rhetorical question – which does not imply an answer, it makes a point. Chandler then answers adding information to his previous interaction: this action highlights backward and forward orientation of *in fact*. In this example, the DM functions as a clarifier to prior information.

(11.b) **CHANDLER**: I'd like to toast, Ross and Emily. Okay. I known Ross for a
long time. *In fact*, I knew him when he was going out with his first
girlfriend.

In example (11.b) Chandler is Ross's best man and therefore has to make a toast. He decides to start off by making fun of Ross. Here, *in fact* is used as a correcting device, highlighting its backward and forward orientation: Chandler first says that he has known his friend for a long time and then he immediately corrects himself by indicating one particular period in the past. Information in the first utterance is corrected in the second one.

(11.c) **CHANDLER**: All right, fine, you know what…we'll both sit in the chair.
I'm soooo, comfortable.
JOEY: Me too. *In fact*, I think I might be a little too comfortable.

Example (11.c) depict Chandler and Joey who are arguing over a chair. Joey is commenting on the comfort of both of them sitting down in the same spot. The backward and forward orientation of *in fact* can be seen as the DM is used to connect the fact of being comfortable (old information) with the subsequent feeling of being too comfortable (new information). Here, *in fact* operates on the validity of the message clarifying the degree of comfort that Joey is experiencing. On more general terms, it expands the meaning of the message: it clarifies it but on some level it also corrects it. Joey corrects the information of being comfortable adding details (how much he is comfortable).

(11.d) **CHANDLER**: Does anyone else think David Copperfield is cute?
MONICA: No, but he told me, he thinks you're a fox.
CHANDLER: All right, Janice…likes him. *In fact* she likes him so much
she put him on her freebie list.

Like example (11.c), *in fact* clarifies the meaning of an utterance, also correcting it. In this case, *in fact* is used to describe to what extent Janice likes David Copperfield. Indeed, Chandler says that she likes him so much that, hypothetically speaking, if she had an

affair with him then Chandler, her boyfriend, has no rights to be angry. Here, *in fact* is used to expand on the fact that Janice likes Copperfield (backward information) by stating what she is willing to do (forward information).

> (11.e) **RACHEL**: What are you talking about, Ross, you just said that you read it twice! Look, y'know what, either it does or it doesn't, and if you have to even think about it...
> **ROSS**: No, Rach, no. I don't, I don't, I don't have to think about it. I*n fact*, I've decided, I've decided that, that it.......does.

In example (11.e) Rachel and Ross are discussing about a letter from Rachel to Ross. He has not read it but pretends to have done it twice. Rachel is angry that Ross cannot decide whether the letter means something to him or not so Ross corrects his previous statement saying that he has decided. This turn in his own point of view is prefaced by *in fact*, which in this case acts as a correcting device, and could be substituted with the DM *actually*. It is used by the speaker to mark a correction in his/her thoughts/ideas. The fact of having to think about something (old/backward information) changes to the fact of having decided (new/forward information).

> (11.f) **PHOEBE**: Oh, okay. Well, so tell me everything about my parents.
> **PHOEBE SR**: Ohh, well. Y'know we were always together… *In fact* they had a nickname for the three of us.

In this example, Phoebe's biological mother also called Phoebe, is talking about her relationship with Phoebe's adoptive parents. She states that they were close friends, so close that people gave them a nickname. This expansion of information is signalled by the use of *in fact*. Here, it acts as a clarifying device. The speaker gives additional information (forward orientation) to a prior message (backward orientation).

> (11.g) **PHOEBE**: I think the baby can totally hear everything. I can show you. Look, this will seem a little weird, but you put your head inside this turkey, and then we'll all talk, and you'll hear everything we say.

CHANDLER: I'd just like to say that I'm totally behind this experiment. *In fact*, I'd very much like to butter your head.

In example (11.g) Chandler shows his (fake) excitement for what Phoebe just suggested doing saying that he is "totally behind this experiment". He then adds information about the try-out he wants to do. Here, *in fact* operates on the validity of message as a clarifier introducing information which explain a prior point. Chandler's excitement (backward orientation) is explained by the activity he wants to do (forward orientation).

(11.h) **ROSS**: Hey, y'know, y'know what would make me really happy?
RACHEL: Oh yeah, no, what's that?
ROSS: If like the four of us could all y'know, hang out together. Uh, *in fact*, Emily's coming into town this weekend, why don't you say we all have dinner?

Lastly, example (11.h) shows Ross trying to tell Rachel that he is planning a double date. First, he generally talks about the possibility of hanging out together. Then he corrects what he just said by announcing that Emily is about to get to New York, and that they can all go out that weekend. In this case, *in fact* functions as a correcting device, it fixes prior information.

3.3 Discussion

Analysing the results and the contexts in which discourse markers occur, it can be said that *I mean* is one of the most used DMs both in authentic language (2.10 times per thousand words) and in non-authentic language (1.70 times). On the contrary, *In fact* can be considered an infrequent marker, as it occurs 0.20 times per thousand words in the Santa Barbara Corpus, representative for authentic language, and only 0.07 times in the Friends Corpus (non-authentic language). Moreover, The results show that both discourse markers may appear alone, preceded by other markers, or in expressions (conversational routines). Except for one occurrence of *in fact*, both discourse markers appear in sentence initial position and are used either by a speaker to go on with his/her conversation or by a speaker to reply to another speaker's utterance (whether it is a question or a statement).

I mean and *in fact* are discourse markers conveying a number of similarities. Indeed, the cognitive frames of both DMs highlight that they share contextual characteristics: they have local functions, i.e. they operate on the validity of messages and/or 'facilitate the transition from one turn to another' (Aijmer, 1996: 222); they operate on the ideational and information level of discourse since they relate and organize information, both backwards and forwards. One difference lies in the person deixis. *I mean* is speaker oriented. It operates on the participation framework, i.e. it marks the speaker orientation in the discourse. In turn, *in fact* has no speaker or hearer orientation. It works on the action level and therefore focuses on the action/information rather than on the person.

Another evidence resulted from the analysis is the fact that they convey similar meanings. *I mean* is mostly used to clarify previous points, but it can also be used by speakers to modify/correct their points of view. Likewise, *in fact* functions as a corrective device, as it modifies the meaning of prior utterances, and as a clarifying marker which introduces additional details to already known information. Besides correcting and clarifying, *in fact* can be used to contradict utterances and messages. This latter function is the least applied. In the Santa Barbara Corpus, *in fact* occurs 0.025 times/thousand words with the function of contradicting information, whereas in the Friends Corpus the frequency is 0.006. Nonetheless, the fact that they share functions does not mean that the two discourse markers are interchangeable. For instance, in example (5), *I mean* is used to clarify a point made by the speaker.

(5) **MONICA**: Well, do you love him?
　　 RACHEL: Sure.
　　 MONICA: Sure?
　　 RACHEL: Yeah, *I mean*, whatever.

Rachel's point of view is that she does not love him and clarifies that she does not care about it (message encoded in the DM *whatever*). Here, *I mean* operates on her point of view and thus on her relationship with that information. *In fact* functions as a clarifier too, but it could not substitute *I mean*, as it operates on actions and not on the participation framework. The sentence "Yeah, *in fact*, whatever" could still make sense but it seems to

lack emotions. Using *I mean* gives a much stronger sense of thought and emotions from the speaker. following the same logic, *I mean* cannot fully substitute *in fact*. Example (11.c) shows the use of *in fact* as a clarifier, function which characterizes *I mean*, too.

(11.c) **CHANDLER**: All right, fine, you know what…we'll both sit in the chair. I'm soooo, comfortable.

JOEY: Me too. *In fact*, I think I might be a little too comfortable.

Here, one could substitute in fact with I mean, since both discourse markers acts as clarifying devices. "Me too. *I mean*, I think I might be a little too comfortable" still makes sense. A hypothetical hearer understands what the speaker wants to say, i.e. the feeling of being very comfortable. In this case, the overall meaning of the message emphasizes the speaker and his/her point of view, rather than the specific action of feeling comfortable. These two interchanging experiments show that *in fact* focuses on actions whereas *I mean* focuses on points of view. It is then the context in which discourse markers occur which suggests the most suitable element to apply.

Conclusion

Findings of chapter three shed light on the research questions at the basis of my analysis: "what is the frequency of DMs in authentic and non-authentic language?", "what meaning do they convey?", "Do they occur under the same conditions in both uses of language?"

Drawing on the data collected in Table 3, an important aspect of the investigation, i.e. the frequency of the discourse markers *I mean* and *in fact*, can be discussed. It can be said that *I mean* is the most used marker of the two. Nevertheless, what is striking is the difference in frequency between the two corpora: *I mean* occurs 2.10 times/thousand words in the Santa Barbara Corpus, and 1.70 times/thousand words in the Friends Corpus. This means that this DM is **0.04%** more frequent in the former corpus, which represents authentic language. Analysing the results of *in fact*, it scored 0.20 occurrences/thousand words in the Santa Barbara Corpus, and only 0.06 occurrences/thousand words in the Friends Corpus. *In fact* is rarely used in both corpora, but it is **0.01%** more frequent in the Santa Barbara Corpus than in the Friends Corpus. These data show an interesting pattern: both discourse markers are more frequent in the corpus including authentic language use than in the corpus including non-authentic language use. Therefore, the results show that authenticity in scripted dialogue is certainly represented through the use of real-time elements, in this specific case discourse markers, but it is still less evident than naturally occurring interactions.

	SANTA BARBARA CORPUS	FRIENDS CORPUS
I MEAN	**2.10** occurrences per thousand words	**1.70** occurrences per thousand words
IN FACT	**0.20** occurrences per thousand words	**0.06** occurrences per thousand words

Table 3. Frequency per thousand words of the DMs *I mean* and *in fact* in the two corpora.

Answers to the second research question, "what meaning do they convey?", and to the third one, "do they occur under the same conditions in both language uses?" have already been given in the discussion section of chapter three. Results show that context

influences in the same way the use of the two discourse markers in authentic language use and in non-authentic language use. To sum up, three main functions have been found analysing the discourse marker *in fact*: clarifying/explaining, correcting, and contradicting. As for *I mean*, two main functions have been highlighted: it expands the meaning of prior utterances both clarifying and modifying ideas. Both discourse markers operate on the information state but focus on different elements: *I mean* highlights a speaker's relationship to information (points of view), whereas *in fact* emphasizes actions. As for the grammatical context in which the two discourse markers occur, the investigation shows that both *I mean* and *in fact* are used in sentence initial position, except for one occurrence of *in fact* in sentence final position (in the Friends Corpus). From this point of view, authenticity is well represented in scripted dialogue, which reflects functions and grammatical features of discourse markers.

The results of this analysis lead to the conclusion that authenticity is deeply attached to real life conversation. Scripted dialogue surely represents it, but the issue of portraying natural language use is still unresolved. It has to be said that my investigation has some limitations, particularly concerning the "Friends: the TV Series Corpus". First of all, it comprises a single TV Series, which represents the life of a group of friends in New York. This means that the results depict a certain language use influenced by 'shared cultural knowledge' (Gee, 2011), i.e. New Yorkers' use of language is different from other cities and therefore it may include expressions or specific elements which are characteristic of that culture but may not even exist in other cities' use of language. A second limitation lies in the corpus chosen: analysing a single TV Series gives only a partial representation of the fictional sphere. In addition, out of 236 episodes, only 100 have been collected in the corpus, restricting the analysis even more. Last but not least, the number of elements analysed does not allow to give an ultimate theory about the issue of authenticity. The investigation analysed just two discourse markers which could not possibly represent the whole category, even though one, *I mean*, scores a high number of frequencies in both corpora, suggesting that it is one of the most used DMs; whereas the other, *in fact*, ranks amongst the least used DMs in both language uses.

Altogether, the analysis of the previous chapter provides a general idea of how authenticity is represented in naturally occurring conversations, which by their nature are

authentic, and in scripted dialogue, which is a crafted product including authentic elements. But since this dissertation is limited, further studies on the analysis of authenticity in scripted dialogue may be done drawing on the analysis and the subsequent conclusion of this work. Suggestions include collecting corpora containing a wider number of TV Series of different genres and years and/or analysing different discourse markers in order to have a more in-depth analysis.

Bibliography

Aijmer, K. (1996), *Conversational routines in English: convention and creativity*. London: Longman.

Alexander, R. (1984), "Fixed expressions in English: Reference books and the teacher". In *English Language Teaching journal*, 38(2), pp. 127-34.

Andersson, L.G. (1975), "Talktsadverbial Nvsvenska studier" [in Swedish]. In Aijmer, K. 1996. *Conversational routines in English*.

Bahns, J., H. Burmeister and T. Vogel. (1986), "The pragmatics of formulas in L2 learner speech: use and development". In *Journal of Pragmatics*, 10(6), pp. 693-723.

Bednarek, M., Piazza, R., Rossi, F. (2010), *The Language of Fictional Television – Drama and Identity*. New York: Continuum International Publishing Group.

Blakemore, D. (2002), *Relevance and Linguistic Meaning: the Semantics and Pragmatics of Discourse Markers*. Cambridge University Press, Cambridge.

Biber D., Conrad S., Leech G. (2002), *Student Grammar of Spoken and Written English*. Harlow: Pearson Longman

Collins, J. (1991), "Hegemonic practice: Literacy and standard language in public education". In Mitchell, C. and Weiler, K. (Eds.) *Rewriting Literacy: Culture and Discourse of the Other*, pp. 229-254.

Dediu, D. and S. C. Levinson (2013), "On the Antiquity of Language: The Reinterpretation of Neandertal Linguistic Capacities and Its Consequences". In *Frontiers in Psychology*, IV, 397, pp. 1-17.

Esslin, M. (2002), *The Age of Television*. New Brynswick, NJ: Transaction Publishers. [originally published in 1982, San Fransisco, CA: Freeman.]

Eagleton, T. (1983), *Literary Theory*. Oxford: Blackwell.

Fraser, B. (1999), "What are Discourse Markers?". In *Journal of Pragmatics*, 31, pp. 931-952. Boston: School of Education, University of Boston.

Gee, J.P. (2011), *How to do Discourse Analysis: a toolkit*. Oxon: Routledge.

Huisman, R., et al. (2005), "Aspects of narrative in series and serials". In H. Fulton (Ed.), *Narrative and Media*, pp. 153-171. Cambridge: Cambridge University

Johnstone, B. (2002), *Discourse analysis*. Oxford: Blackwell.

Jucker, A. 1993. "The discourse marker well. A relevance-theoretical account", in *Journal of pragmatics* 19(5), pp. 435-53.

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

Levinson, S. (1983), *Pragmatics*. Cambridge: Cambridge University Press.

Martin, J. R. (1984). "Language, register and genre". In F. Christie (Ed.), *Children writing: Reader*, pp. 21–29. Geelong: Deakin University Press.

Mills, B. (2005), *Television sitcom*. London: British Film Institute.

Milroy, J., and Milroy, L. (1991), *Authority in Language: Investigating Language Prescription and Standardisation*. London: Routledge.

Mittel, J. (2004), "A cultural approach to television genre theory". In Allen, R. & Hill, A. (Ed.), *The Television Studies Reader*, pp. 171-81. London/New York: Routledge.

Neale, S. (2001), "Studying genre". In Creeber, G. (Ed.), 1-3.

Quaglio, P. (2009), *Television Dialogue*. Amsterdam: John Benjamins Publishing Company.

Redeker, G. (1990), "Discourse Markers in Attentional Clues at Discourse Transitions". In Fischer, K. (Ed.), *Approaches to Discourse Particles*, pp. 3-26.

Richards, J.C. and Schmidt, R. (2002), *Longman Dictionary of Language Teaching and Applied Linguistics* (3rd ed.). Harlow, UK: Longman.

Sperber, D. and D. Wilson. 1986. *Relevance. Communication and cognition*. Oxford: Blackwell.

Taylor, C. (2004) "The Language of Film: Corpora and Statistics in the search for Authenticity. Notting Hill (1998) – a case study". In *A Journal of English and American Studies*.

-- (2008), "Predictability in Film Language: Corpus-Assisted Research". In Taylor Torsello C., Ackerley K., Castello E. *Corpora for University Language Teachers*, pp. 167-80. Berna: Peter Lang.

Tannen, D. (1989). *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press.

Turner, G. (2001) 'The Uses and Limitations of Genre'. In G. Creeber (ed.) The Television Genre Book, pp. 4–5. London: British Film Institute.

Wiley, T. G., and Lukes, M. (1997), "English-Only and Standard English ideologies in the U.S.". In *TESOL Quarterly*, 30(3), pp. 511-534.

Sitography

Bubel, C. (2006), *The linguistic Construction of Character Relations in TV Drama: Doing Friendship in Sex and the City*. Unpublished PhD dissertation, Universität des Saarlandes, Saarbrücken (Germany). Available at: Diss_Bubel_publ.pdf (uni-saarland.de)

Hacker, P. M. S. (2014), "Two Conceptions of Language". In *Erkenntnis (1975-)*, *79*, pp. 1271–1288. From http://www.jstor.org/stable/24021394.

Max S. Kirch. (1973). "Language, Communication and Culture". In *The Modern Language Journal*, *57*(7), pp. 340–343. From https://doi.org/10.2307/324652.

McDavid, R. I. (1968), "Variations in Standard American English". In *Elementary English*, *45*(5), pp. 561–608. From http://www.jstor.org/stable/41386367.

Nordquist, R. (2020), *Standard English (SE)*. From https://www.thoughtco.com/standard-english-1692137

Schegloff, E. (1999), "Discourse, pragmatics, conversation, analysis". In *Discourse Studies*, 1(4), pp. 405-435. From http://www.jstor.org/stable/24046433.

Spitz, A. (2005), *Power Plays: the Representation of Mother-Daughter Disputes in Contemporary Plays by Women. A Study in Discourse Analysis*. Unpublished PhD dissertation, Universität des Saarlandes (Germany). Available at: dissFINALE.pdf (uni-saarland.de).

Van Dijk, T. (2006), "Discourse, context and cognition". In *Discourse Studies*, 8(1), pp. 159-177. From http://www.jstor.org/stable/24049517.

Verdonik, D. et al. "The Impact of Context on Discourse Marker Use in Two Conversational Genres". In *Discourse Studies*, vol. 10(6), pp. 759–75. From http://www.jstor.org/stable/24049381

Summary

Questa tesi analizza la rappresentazione dell'autenticità nel linguaggio, intesa come spontaneità e naturalezza con cui le persone interagiscono. Nel dettaglio, si concentra sul problema dell'autenticità nel cosiddetto linguaggio non autentico. È importante infatti distinguere tra linguaggio autentico e non autentico. Il primo termine si riferisce a conversazioni che nascono spontaneamente tra due o più partecipanti. Il secondo termine (linguaggio non autentico) si riferisce a quello che in questa tesi viene chiamato "scripted dialogue" (dialogo da copione), ossia conversazioni precostruite e organizzate. Il linguaggio non autentico caratterizza i vari prodotti d'intrattenimento per cinema e televisione, tra cui le serie TV. L'obiettivo di questa tesi è quello di valutare il grado di autenticità rappresentato dai prodotti televisivi seriali usando come campione la serie TV "Friends", una serie comedy andata in onda tra la fine degli anni '90 e l'inizio degli anni 2000.

Due fattori sono importanti nella caratterizzazione del linguaggio: il contesto e la prevedibilità. La comunicazione è un'azione multimodale: esistono diversi aspetti sia verbali che non verbali i quali contribuiscono all'interpretazione di un messaggio. Quando le persone interagiscono, lo fanno all'interno di un contesto. Questo termine può essere descritto come l'insieme degli elementi coinvolti nella comunicazione, che possono essere fisici (es. postura, sguardo) o astratti (informazioni condivise, nozioni universali conosciute da tutti). Dal punto di vista del dialogo da copione, oltre al contesto (seppur fittizio) in cui interagiscono i personaggi, è presente anche un contesto televisivo che comprende gli autori del copione, gli attori che lo interpretano e l'audience, i quali interpretano la comunicazione grazie ad una conoscenza generale.

Passando al secondo aspetto, si può affermare che la prevedibilità evidenzi profondamente la dicotomia tra linguaggio autentico e linguaggio non autentico. Il linguaggio autentico è caratterizzato dalla spontaneità con la quale nascono e si sviluppano le conversazioni, che di conseguenza includono un alto numero di elementi quali ellissi, inserzioni, marcatori del discorso. Tutti questi elementi aiutano e permettono ai partecipanti di organizzare in tempo reale le proprie idee e quello che stanno dicendo. Al contrario, il linguaggio non autentico è un prodotto creato da diverse figure

(sceneggiatori, scrittori, registi, attori) che non trovandosi in una situazione di spontaneità non hanno la necessità di usare questi elementi. La difficoltà nell'usare un linguaggio autentico è sempre più presente nelle serie TV, che nel corso degli anni si sono evolute per avvicinarsi alla realtà a cui è abituato lo spettatore. Uno degli aspetti che sono cambiati è certamente il modo con cui i personaggi interagiscono tra di loro. In passato i dialoghi erano molto più articolati, risultando però evidentemente simulati. Ultimamente si sta assistendo al fenomeno inverso: i copioni lasciano molto più spazio alla spontaneità, permettendo anche all'attore che interpreta un personaggio di avere la possibilità di improvvisare o rendere il dialogo più naturale: vengono quindi utilizzati tutti quegli elementi tipici delle conversazioni naturali come, ad esempio, i marcatori del discorso, sui quali si concentra l'analisi di questa tesi.

Il termine "marcatore del discorso" è stato coniato da Schiffrin nel 1989, il cui lavoro può essere considerato uno dei primi studi su questa materia. A lungo gli studiosi hanno cercato di definire e caratterizzare questi elementi, che ad oggi possono essere descritti come elementi con funzioni di organizzazione del flusso di informazioni nel discorso e che operano a livello del discorso. Hanno una posizione esterna all'enunciato e di conseguenza non ne modificano il significato. Le caratteristiche dei marcatori del discorso possono essere incluse nei cosiddetti "cognitive frames" (cornici cognitive), ossia unità di informazioni utili a chi parla per riconoscere un marcatore del discorso. In queste cornici sono presenti proprietà grammaticali e contestuali, tra cui la funzione deittica dei marcatori sia personale (orientamento a chi parla, chi ascolta, ecc.) che relativa al testo (orientamento indietro o avanti nel testo). Altra distinzione dei marcatori del discorso è tra marcatori locali o globali: i marcatori locali operano e commentano sulla validità e veridicità di un messaggio (quanto è importante); i marcatori globali vengono utilizzati come guida attraverso la conversazione e fungono da organizzatori di argomenti.

In questa tesi viene analizzata la frequenza di due marcatori specifici, *I mean* e *in fact*. Entrambi sono marcatori locali orientati sia indietro che avanti nel testo. Il primo è orientato a chi parla, mentre il secondo non presenta orientamento specifico da questo punto di vista. Tutti questi aspetti sono interpretabili dai risultati dell'analisi del discorso, la quale studia il linguaggio inteso come insieme di regole e si concentra sulle azioni

compiute da chi lo utilizza, per esempio esprimere emozioni, scambiare idee, ecc. I risultati dell'analisi solitamente rispondono ad una serie di domande che vengono poste precedentemente e che hanno la funzione di restringere il campo di ricerca, affinché l'analisi risulti più effettiva. Il materiale analizzato viene chiamato tecnicamente "texts" (testi) e viene raggruppato a seconda delle caratteristiche condivise in raccolte chiamate corpus. L'analisi di questa tesi riguarda due corpus specifici. Il primo è il "Santa Barbara Corpus of Spoken American English", il quale rappresenta il linguaggio autentico, composto da circa 240,000 parole. Il secondo è il "Friends: the TV Series Corpus", il quale rappresenta invece il linguaggio non autentico, composto da 160,000 parole.

Attraverso l'utilizzo di un programma di analisi chiamato AntConc, si può procedere con l'analisi dei due marcatori del discorso scelti (*I mean*, *in fact*) in entrambi i corpus. I risultati di quest'analisi permettono di rispondere alle domande poste precedentemente, che in questo caso specifico sono: "Qual è la frequenza dei marcatori del discorso nel linguaggio autentico e nel linguaggio non autentico?", "Quale significato hanno?", "vengono utilizzati allo stesso modo in entrambi gli usi del linguaggio?". Ciò che è emerso è conferma del fatto che l'autenticità è strettamente legata alla comunicazione spontanea. Ciononostante, nel linguaggio non autentico sono presenti, seppur in numero limitato rispetto al linguaggio autentico, i vari elementi che conferiscono spontaneità alla conversazione. Questi risultati sono però limitati, in quanto l'analisi si concentra solo su due marcatori del discorso, e il corpus preso in considerazione per il linguaggio non autentico comprende solo un centinaio di episodi di "Friends", che non può essere presa come unica rappresentante di tutte le serie TV per un'analisi completa. Per questo motivo, ulteriori studi potrebbero partire dai risultati di questa tesi e sviluppare ulteriori conclusioni derivanti dall'analisi di altri marcatori del discorso in un corpus più completo sia dal punto di vista del numero dei testi che del genere di serie TV.