Context and Language

Elena Manca

UNIVERSITÀ DEL SALENTO

2012

http://siba2.unisalento.it

eISBN 978-88-8305-092-3 (electronic version)

http://siba-ese.unisalento.it

CONTENTS

Introduction[1]

When the study of meaning became a linguistic discipline, the main interest of semanticists was in the technical level of language and the major focus was represented by the single word. A different view of the relationship between language and meaning was provided by Malinowski (1923) at the beginning of the 20th century. The meaning of language was interpreted in terms of context of culture and context of situation. Malinowski's theories influenced linguists such as Firth (1957) and Halliday (1985) and language became to be considered meaningful only if considered within the language events in which it is used. For this reason, language started to be analysed only in authentic contexts: the focus of interest is not the single word any more but the meaningful relations words enter into with the other words around them (Sinclair 1991; 1996).

The main theme in this book is meaning: how meaning originates and how meaningful communication can be established in texts.

The target readers of this book are those students that are approaching the study of language, particularly of the English language, from a linguistic perspective for the first time. Discovering how language works is fascinating but it may also be perceived as complex and confusing. Students may not be aware of all the constraints which limit language choices: they may be aware of that only intuitively but they cannot know how pervasive this phenomenon may be. Constraints on language are operated by the broader context of culture but also by the topic and the participants of a language event, that is to say the context of

[1] I would like to thank Teresa Carchedi for reading a draft of this book and for her precious comments.

situation. Further constraints are to be found in the linguistic environment of a word: every time we choose a w ord or a phrase we are limited by the phraseological tendency of language.
Added to this, the structure we give to our message is also a meaningful choice: what is placed in initial position of a clause or at the end of it are constrained by the final aim of our message.
Students will be guided through the five chapters constituting the book to understand the strict relationship existing between context and language and to become better users of the English language.

Outline

The book is divided into five chapters.
Chapter 1 presents the concepts of context of culture and context of situation by discussing the theories by Malinowski, Firth and Halliday. Practical examples are also provided in order to make students aware of the influence operated by the context of culture and the context of situation on the language used.
Chapter 2 introduces corpora and corpus analysis tools. Students will learn the terminology of corpus analysis and what concordances, wordlists, and keywords are. The criteria that are usually applied in the compilation of corpora will also be outlined. The chapter also presents and explains how to use some of the tools available to carry out a corpus analysis.
Chapter 3 focuses on Firth's and Sinclair's theories on the notion of collocation. Students will understand how meaning arises from the combination of words and which steps have to be taken in order to identify extended units of meaning. Some practical examples of analysis are also provided.
Chapter 4 explores issues related to the identification of equivalent units of meaning across languages. The methodology proposed by Tognini Bonelli (2001) and Tognini Bonelli and Manca (2002) on functionally complete units of meaning will be described and applied to a number of English and Italian case studies.
Finally, Chapter 5 summarizes Halliday's view on Theme and Rheme in the systemic functional tradition. Students will learn the differences between marked

and unmarked Themes and the features which contribute to create cohesion in a text. Examples of thematic progression in English will also be provided following the theories by Firbas (1964) and Danes (1974).

1 The notion of context

1.1 Context and co-text

Words do not occur in isolation. The words of a text are surrounded by their linguistic environment, called **co-text**; the text takes place in a broader environment, called **context**. Both co-text and context are of utmost importance in the identification of the meaning of a text.
For example, the word *order* in the following sentence takes its meaning from its co-text, that is to say from the words that follow and precede:

Your order will be processed within 3/4 days after your request is submitted

Items such as *processed, request,* and *submitted* contribute to the meaning of *order* that, in this case, refers to a request for a product to be delivered to you. Furthermore, the linguistic co-text of *order* allows us to understand what is going on, that is to say it helps us make inferences on the broader event where the sentence takes place. It can be clearly understood that this sentence has been uttered or written within an event whose topic is a commercial transaction, and where the participants are the customer and the seller. Furthermore, the role language is playing in this speech event can be analysed in terms of medium which, in this case, is written (probably an e-mail or a letter) and in terms of rhetorical function which is descriptive. The topic, the participants and the medium of an event constitute what has been called by Halliday (1985a) context of situation (see section 1.2.3).

A different linguistic co-text of *order* would suggest a different context of situation as in the following example:

Excuse me, can we order, please?

Excuse me, can and *please* suggest that the linguistic event is probably taking place in a restaurant where some customers (the participants) are ordering their meal (the topic) to the waiter (the participant). In this case the medium is spoken. However, there are also cases where the context helps us disambiguate the meaning of a sentence. The sentence *She is a baby* may refer either to a baby or to an adult who looks or acts as a baby. The two contexts and co-texts will help us understand the meaning of the following sentences:

1. *Introduce your child to books when **she is a baby***
2. *She shouldn't have a baby because **she is a baby** herself*

Context and co-text play a key role in the analysis of meaning as we will see later. However, before moving to a more practical description of these two concepts, the following section will provide a brief overview of some traditional theories of context.

1.2 A theoretical overview

A theory of context was first developed by the anthropologist Bronislaw Malinowski (1923; 1935). The influence of Malinowski is visible in the theories of many scholars, particularly in those of J.R. Firth, one of his colleagues at London University and of Halliday who follows both Malinowski's and Firth's ideas in the development of his theory on the context of situation.

1.2.1 Malinowski and the theory of context

Bronislaw Malinowski carried out an ethnographic field work in the Tobriand Islands (see Halliday 1985a:5ff). The inhabitants of those islands lived by fishing and gardening and spoke the Kiriwininan language. The first problem he had to face was how to translate the texts in Kiriwinian he had taken down in discussion with the Tobrianders, in order to make them accessible to the British culture. These texts were the product of a culture which was extremely different from the Western culture, for this reason a free translation would have not helped Western people to understand them. He argues (1923:301-2)

> Instead of translating, of inserting simply an English word for a native one, we are faced by a long and not altogether simple process of describing wide fields of custom, of special psychology and of tribal organisation which correspond to one term or another. We see that linguistic analysis inevitably leads us into the study of all subjects covered by Ethnographic field-work.

For this reason, he decided to add an extended commentary to the translated text, which "placed the text in its living environment" (Halliday, 1985a:6). Malinowski coined a new term (1923) which identifies the total environment, including the verbal environment and the situation in which the text was produced: the *context of situation*.

As Halliday points out (1985a: 6) Malinowski

> (…) understood that a text written by these people into this language could not be understood by any foreigners or by people living outside this society even if translated into their own languages because each message brought more meanings than those expressed through the words, meanings that could only be understood if accompanied by the situation. Thus, Malinowski introduced the notion of context of situation, meaning by this the environment of the text.

In his ethnographic treatise *Argonauts of the Western Pacific,* Malinowski identifies three main aspects of social life that he believed the ethnographer must

fit together into a unified description of a given society (1922:22; see also Langendoen, 1968:12-13):

> . . . in every act of tribal life, there is, first, the routine prescribed by custom and tradition, then there is the manner in which it is carried out, and lastly there is the commentary to it, contained in the natives' mind.

The strict relationship between language and culture is well explained by Malinowski in *'The problem of meaning in primitive languages'* (1923), where he argues that language can be explained only by considering the broader context of situation:

> Language is essentially rooted in the reality of the culture, the tribal life and customs of the people, and […] it cannot be explained without constant reference to these broader contexts of verbal utterance. […] An utterance becomes intelligible when it is placed within its context of situation.

In order to illustrate the notion of context of situation, Malinowski described a typical Tobriand fishing expedition: after the islanders had gone outside the lagoon into the open sea to fish, they had to navigate a difficult course through the reefs to get back to the lagoon. For this reason, people on the shore shouted instructions to the fishermen and the whole situation became a sort of competition between the different canoes and groups of people. The language used in such situations was full of technical terms, references to surroundings and indications of change, based on types of behaviour well-known to the participants. The linguistic material used was, therefore, inextricably dependent upon the course of the activity in which the utterances were embedded (1923:311-312). It was language in action (see Halliday 1985a: 6) and the message was clear only to those who knew what was going on. For this reason when he accounted for these situations he realised that it was necessary to provide not only a description of what was happening, the situation, but also of the total cultural background, because:

> involved in any kind of linguistic interaction, in any kind of conversational exchange, were not only the immediate sights and sounds surrounding the event but also the whole cultural history behind the participants, and behind the kind of practices they were engaging in, determining their significance for the culture, whether practical or ritual. All these played a part in the interpretation of meaning. (Halliday, 1985a:6)

Here, Halliday refers to the notion of the context of culture, which together with the context of situation is necessary for the understanding of the text.
Furthermore, Malinowski (*ibidem*) stresses the importance of the function of words, arguing that their meaning is not given by the physical properties of their referents but by the way they are used in a given situation. Langendoen (1964:22) who has provided a critical analysis of Malinowski's works puts it this way

> Malinowski had an important insight into the nature of the meaning of particular words, namely, that their meaning is not given by the physical properties of their referents but rather by their function. He insisted that all words are functionally defined, and not only all words but all possible utterances in a language, and further that the meanings are so learned only by active experience and never by explanation or paraphrase.

Malinowski also accounted for another use of language which was different from the pragmatic one: the narrative use. The islanders used to gather around and to tell stories. In this case there are two contexts of situation: the situation of the moment of narration and the situation created by the stories themselves. Langendoen (1964:23) claims that the meaning of narrative has nothing to do with the context of situation of the moment of narration. On the other hand, what Malinowski wanted to show is that these narrative texts were somehow related to the situation in which they were told. For example, stories about great famines in the past and how people reacted all together to overcome it, were told during that part of the year when food was hardly available. So, the context of situation of the moment of narration was somewhat relevant because of a direct relation between the narrative and the immediate surrounding (Malinowski, 1923:313).

1.2.2 Firth and the context of situation

The influence of Malinowski's view about the context of situation is also visible in the work of J.R. Firth. He adopted Malinowski's notion but in his linguistic theory the context of situation was the whole cultural setting in which the speech act was embedded, more than the context of human activity concurrent with, immediately preceding, and following the speech act (Langendoen, 1964:35).
In Firth's view (1957d:182) the context of situation is best used as a suitable schematic construct to apply to language events and is a group of related categories at a different level from grammatical categories but rather of the same abstract nature. The categories brought into relation by the context of situation are (*ibidem*):

A. The relevant features of participants: persons, personalities.
   (i) The verbal action of the participants.
   (ii) The non-verbal action of the participants.
B. The relevant objects.
C. The effect of the verbal action.

Firth's taxonomy has then to be applied to language events and in this way contexts of situation and types of language function can be grouped and classified. The following text taken from Dan Brown's *Digital Fortress* (1998) can be analysed according to Firth's taxonomy:

Susan's Volvo sedan rolled to a stop in the shadow of the ten-foot-high, barbed Cyclone fence. A young guard placed his hand on the roof. 'ID, please,' Susan obliged and settled in for the usual half-minute wait. The officer ran her card through a computerized scanner. Finally he looked up. ' Thank you, Ms Fletcher.' He gave an imperceptible sign, and the gate swung open. Half a mile ahead Susan repeated the entire procedure at an equally imposing

electrified fence. *Come on, guys … I've only been through here a million times.* As she approached the final checkpoint, a stocky sentry with two attack dogs and a machine gun glanced down at her license plate and waved her through. She followed Canine Road for another 250 yards and pulled into Employee Lot C. Unbelievable, she thought. *Twenty-six thousand employees and a twelve-billion-dollar budget; you'd think they could make it through the weekend without me.* Susan gunned the car into her reserved spot and killed the engine.

The relevant features of participants (persons, personalities):

1. Susan Fletcher, NSA (National Security Agency) head cryptographer
2. Young guard
3. Stocky sentry

(i) **The verbal action of the participants**: '*ID, please' … Thank you, Ms Fletcher …*

(ii) **The non-verbal action of the participants**: *Susan's Volvo sedan rolled to a stop … A young guard placed his hand on the roof … He gave an imperceptible sign … a stocky sentry with two attack dogs and a machine gun glanced down at her license plate and waved her through …*

The relevant objects: ten-foot-high, barbed Cyclone fence … her card … equally imposing electrified fence … final checkpoint … machine gun

The effect of the verbal action: *Susan obliged and settled in …*

According to Firth's theories, the meaning of what is going on is clear when all these features are analysed and considered together.

The importance of the context of situation in the identification of meaning is linked to another Firthian tenet, that is to say 'the notions of personality and language'(1957a:183) which are considered by Firth as vectors of the continuity of repetitions in the social process (ibid.). Firth stresses the importance of

studying individuals, and not speaking masses[2], in their bundles of roles and *personae*. He says (1957a:184)

> There is the element of habit, custom, tradition, the element of the past, and the element of innovation, of the moment, in which the future is being born. When you speak you fuse these elements in verbal creation, the outcome of your language and of your personality.

What Firth strongly believes in is that any social person in the multiplicity of roles he/she takes in his/her life and in the multiplicity of contexts of situation he/she finds him/herself in, is not free to say what s/he wants. We behave systematically "since experienced language is universally systemic" (1957a:187). This reinforces the relevance of the context in the identification of the meaning of an utterance: since the linguistic events and the roles we perform in given situations influence the language we use, each utterance has to be considered in the context in which it is produced.

An example of how we change our language according to the role we are performing is provided below. Let us suppose we realize that the overcrowded room we are in would need some fresh air. A friend of us is close to the window. We may verbalize our feeling this way:

- *It's so hot in here! Open the window, please.*

However, if we suppose we are in a public place (school, university, office, ...) and we want to ask for permission to someone we do not know our request would sound as follows:

- *Excuse me, would it be possible to open the window, please? The room would need some fresh air.*

[2] He dismisses the Saussurean dichotomies of *langue* and *parole.*

Similarly, if tomorrow we have to do an exam, we will probably tell a friend *Domani devo fare un esame* but to our teacher we would probably say *Professoressa, domani dovrei sostenere un esame.* The two phrases *fare un esame/sostenere un esame* have two different degrees of formality and are used in different contexts of situation. Furthermore, the use of the conditional form for the verb *sostenere* reinforces the formality of the sentence. These examples show the strong relationship between the language we use and the role we perform in context. The same difference can also be identified in English where *to do an exam*, *to take an exam*, and *to sit an exam* have different levels of formality and, as a consequence, clearly reflect the statuses and roles of the people participating in the verbal action.

This view of language also has important implications in the study of language: if language is systemic, that is to say if a specific situation attracts a range of words and phrases we are very likely to use, the linguist may identify scientifically the features of the repeated linguistic events and classify them (see chapter 3).

1.2.3 Halliday and the context of situation

As previously mentioned, Halliday (1985a) takes up both Malinowski's and Firth's ideas. In order to describe the importance of the context of situation in communication, he makes us think about the way in which we communicate. He says (1985a:9)

> What is remarkable is how often people do understand each other despite the noise with which we are continually surrounded. How do we explain the success with which people communicate? The short answer, I shall suggest, is that we know what the other person is going to say. We may be partly surprised; but the surprise will always be within the framework of something that we knew was going to happen.

Successful interactions are possible because we make predictions based on the context of situation. It provides a lot of information about the meanings that are being exchanged and that are likely to be exchanged. This perspective also

influences the definition of text, being for Halliday (1985a:10) "any instance of living language that is playing some part in a context of situation". For this reason a text is a product, an output, and a process, "a movement through the network of meaning potential" (ibid.). Halliday (1985a: 11ff) provides a taxonomy through which it is possible to define the context of situation of a text. He puts it this way

> The text, we have said, is an instance of the process and product of social meaning in a particular context of situation. Now the context of situation, the context in which the text unfolds, is encapsulated in the text, not in a kind of piecemeal fashion, not at the other extreme in any mechanical way, but through a systematic relationship between the social environment on the one hand, and the functional organisation of language on the other.

In order to show the systematic relationship existing between language and the social environment he provides his own taxonomy of the context of situation, adapted below (1985a:12):

1. **the field of discourse**, that is to say what is going on, the nature of the social action;
2. **the tenor of discourse**, which refers to the participants, their statuses and roles, their permanent and temporary relationships;
3. **the mode of discourse**, that is to say the role language is playing, the symbolic organisation of the text, its status, its function in the context, the channel and the rhetorical mode.

Halliday (*ibidem*) argues that such an analysis of the context allows us to represent the system that lies behind the unconscious process of producing and understanding texts in a context of situation.

The notion of context of situation is also strictly linked to the concept of register. It means that the features of the context of situation somehow constrain the lexis and the expressions that can be used.

In order to understand better the kind of restraints operated by the situation, we can consider the definition of register provided by Halliday (1985a:39):

> A register is a semantic concept. It can be defined as a configuration of meanings that are typically associated with a particular situational configuration of field, mode and tenor. But since it is a configuration of meanings, a register must also, of course, include the expressions, the lexico-grammatical and phonological features, that typically accompany or REALISE these meanings.

The identification of the three variables of the context of situation indicated by Halliday (see above) has different implications in language use and consequently, in register.

The variable of the field of discourse, which refers to the topic of the language event, operates a constraint on the lexis and expressions available for use. Ulrych (1992:85) provides the examples of *absolve* and *acquit*. Both mean *assolvere* in Italian. However, the first of these two verbs belongs to the set of lexis and expressions of a religious register and is likely to occur with words such as *confessional, priest, go in peace, let us pray*. Conversely, *acquit* is used in the legal register and will co-occur with items such as *charge, court, I rest my case*.

The word *depression* may co-occur with different items and may have different meanings depending on the register in which it is included. In the language of medicine it co-occurs with *suffer from* whereas in the language of economics it is included in expressions such as *periods of economic depression*, and in the language of meteorology it is qualified by *weather.*

The variable of the tenor of discourse, referring to the participants, constrains the degree of formality of a language event. Levels of formality can be formal, neutral, and informal (Ulrych, 1992:87). As already said above with the example of *to do an exam* and *to sit an exam*, formality depends on the statuses of the participants. *May I borrow your pen* and *give me your pen* clearly exemplify a different relationship between the participants of the two language events. The first example has a more formal style and presupposes that the participants do not know each other (or do not know each other well). In the second example, the use of the imperative form and the absence of expressions such as *please* reveals the high degree of familiarity between the interlocutors.

The variable of the mode of discourse refers to the medium (written, spoken, ...) and the channel (e-mail, article, essay, …) chosen for the language used. The constraints operated on language by this variable include the differences between spoken and written language and the format that different genres should have. In English, contractions should not be used in written texts and a newspaper article does not include the expression *dear readers* which, conversely, may represent the opening of a letter or of an e-mail.

1.3 Context of situation: some practical examples

In the first section of this chapter, the importance of both context of situation and co-text as a source of meaning have already been described and exemplified. In this section, further examples will be provided in order to show that the identification of the context of culture is at the basis of the comprehension of a text. Two texts will also be analysed according to the three variables of the context of situation, that is to say field, tenor, and mode.

1.3.1 Contexts across cultures and languages

The text which follows is the homepage of the Holmsdale Hotel's website (available at www. blackpoolhotel.com/holmsdalehotel.html), a hotel based in Blackpool.

British people will find reading and interpreting what is written in this text quite easy. However, Italians may get a little bit confused, particularly reading those parts of the text where the presence of children is questioned. The habit of not allowing children in some restaurants and pubs is quite common in the UK and it is culturally accepted (see Tognini Bonelli and Manca 2002). In Italy and in other European countries children are always accepted in these public places and parents do not usually go out without their children. The text can, therefore, be interpreted only if its context of culture is considered.

The Holmsdale Hotel

Sorry - for the benefit of our regular guests this hotel is a Child Free Zone - we will only accept young adults of 14 years and over.

Book Online

The Holmsdale is a select, quiet hotel privately owned and run and offering guests quality accommodation in relaxing, informal surroundings with en-suite rooms and lifts to all floors.

Every effort is made to make our guests' stay as pleasurable as possible, with friendly personal service at all times and attention to every detail. Twin, double and single rooms are available but no family rooms so the hotel cannot cater for children.

Conversely, the following Italian text[3] may sound bizarre to English people who do not understand the need to add a detailed description of how a fine may be increased if the offence is carried out in the presence of pregnant women, babies and children up to 12 years.

[3] This notice has been downloaded from the website www.usl3.toscana.it accessed in October 2010.

VIETATO FUMARE

Ai sensi della L. n. 3/2003 art. 51 e s. m. i.: ***"Tutela della salute dei non fumatori"*** e della L. R. T. n. 25/2005: ***"Norme in materia di tutela della salute contro i danni derivanti da fumo"***

i trasgressori alle predette disposizioni sono soggetti alla sanzione amministrativa di una somma:

da Euro 27,50 a Euro 275,00

La misura della sanzione è raddoppiata qualora la violazione sia commessa in presenza di una donna in evidente stato di gravidanza o in presenza di lattanti o bambini fino a dodici anni.[4]

La vigilanza sul divieto di fumo spetta ai seguenti nominativi:

……..

L'accertamento dell'infrazione spetta al personale dei Corpi di Polizia Amministrativa locale, agli Agenti ed Ufficiali di Polizia Giudiziaria

[4] My emphasis

The British equivalent of this notice does not include any detailed description of further increases in the fine due to the presence of certain categories of people, as visible in the notice provided below[5]:

Different cultures produce different texts which should be interpreted accordingly. This confirms Malinowski's theories according to which language cannot be explained without constant reference to the broader contexts of verbal utterance.

[5] This notice has been downloaded from the website of the North Lanarkshire council available at www.northlan.gov.uk, accessed in October 2010.

1.3.2 Contexts across the same language

Examples of Halliday's context of situation and the notion of register are provided in the analysis of the two texts reported below. As said above, the basic assumption is that the language used in texts is strictly correlated to the features of the three variables (field, tenor, and mode).

Both texts chosen as examples talk about fractures, that is to say they have the same field of discourse, the same topic. However, they are characterised by differences at the level of tenor and mode.

Text 1 is an abstract of a scientific article published on a specialized journal and Text 2 is a message posted to an open forum[6]. Let us have a look at the examples:

Text 1

***Subtrochanteric femur fractures* by Bedi A; Toan Le T[7]**

Abstract:

Subtrochanteric femur fractures *have demanded special consideration in* ***orthopaedic traumatology****, given the high rate of complications associated with their management. The intense concentration of* ***compression, tensile, and torsional stresses*** *and* ***decreased vascularity*** *of the region has challenged* ***orthopaedists*** *with problems of* ***malunion****,* ***delayed union****, and* ***nonunion*** *resulting from* ***loss of fixation****,* ***implant failure****, and* ***iatrogenic devascularization*** *of the* ***operative exposure****. Only recently has a better understanding of* ***fracture biology****,* ***reduction techniques****, and* ***biomechanically improved implants*** *allowed for* ***subtrochanteric fractures*** *to be addressed with consistent success.*

Text 2

Topic: Your worst injury

I've had an interesting life when it comes to ***broken bones****. It seems like I can't go 5-6 years without breaking something.*

[6]Ars Technica Openforum: http://episteme.arstechnica.com posted on July 31, 2004

[7] Published in *Orthopedic Clinics of North America* ISSN: 0030-5898, 01-OCT-2004; 35(4): 473-83

When I was in kindergarten/first grade, I was pushed off the top of a slide that was probably 10-15 feet up... when I hit the ground, my leg hit the support arm of the slide and ***it broke my thigh bone*** *in half. I was* ***in the hospital*** *for a month.*
A few years later I ***broke my left arm*** *playing tag of all things... not too bad...*
when I was in 9th grade, I wrecked my 3-wheeler and ***shattered my ankle****. I had* ***4 screws installed*** *and 2 weeks of the hospital.*
My last major injury was in Dec 1999 when I ***ruptured my Achilles tendon*** *in my right ankle playing basketball and then I* ***RERUPTURED it*** *again in March 2000 on vacation in Mexico...*
Besides those 4 major accident's, I've had stitches in various places for different accidents. I even had a skill saw accident when my parents were building their home. […]

The two texts have the same topic (they both deal with fractures) but they show a number of differences in terms of lexical density, tone, level of interaction, use of contracted forms, shared knowledge.
Text one is about fractures and, being a scientific text, the tenor is constituted by the author and his/her audience (mainly colleagues and experts). The mode, which, as already said, describes the way language is being used in the speech interaction, including the medium (spoken, written, written to be spoken, etc.), the channel (e-mail, letter, article, essay, ...) as well as the rhetorical mode (expository, instructive, persuasive, etc.) is expressed through the written channel and its rhetorical mode is mainly instructive.
Similarly, text two is about fractures, but being a message posted to an open forum, the tenor is constituted by the author and his/her audience constituted by the forum community and websurfers. The language is written and the rhetorical mode is expository.
The important difference between the two texts lies in the tenor: the different statuses and roles of the participants imply different levels of formalities and consequently a different use of language (see section 1.2.3).
Text one is highly formal: a scientific article published in a scientific journal is addressed to an expert audience. Text two is highly informal: a message posted to

an open forum addressed to a big and highly heterogeneous community. The expert-to-expert and non-expert-to-non–expert interactions are visible in the linguistic choices. A table will help us to identify the main differences in terms of lexical choice and tone:

TEXT ONE (high degree of formality)	**TEXT TWO (casual and informal tone)**
in orthopaedic traumatology	*when it comes to broken bones*
fractures	*broken bones*
femur	*thigh bone*
biomechanically improved implants	*I had 4 screws installed*
compression, tensile, and torsional stresses	• *shattered my ankle* • *I broke my left arm* • *I ruptured my Achilles tendon*

Further differences are in the mode of the two texts such as the use of contractions (*I've had an interesting life; I can't go 5-6 years)* which are absent in Text one; the use of the first person pronoun (*I was in kindergarten; I wrecked my 3-wheeler; I've had stitches)* which is very rare in scientific English where far more preferred choices are the third person pronoun, the passive form, and a construction of the sentence which depersonalize the voice of the author by positing the object of the sentence in subject position (*Subtrochanteric femur fractures have demanded; The intense concentration …has challenged; Only recently has a better understanding … allowed ...).*

1.4 Conclusion

This chapter has described how the meaning of a text arises from its context of culture, its context of situation, and from its co-text. This explains why texts, in order to be properly interpreted, should be placed in their verbal and social living environment. The crucial relationship between text and context is visible when the

notion of register is considered. Each context of situation has its own set of words and expressions available for use. These words do not acquire their meaning from the physical properties of their referents but from their function in context, that is to say from the way they are used in social and language events.

Context of culture, context of situation and co-text play a fundamental role in the process of translation. Some concepts may exist in one culture but not in another. As shown above, the *Child free zone* advertised by the British hotel would not have meaning in an Italian context of culture where the linguistic equivalent *Area libera dai bambini* would not sound so positive and meaningful. Baker (1992:33) provides the example of *Cream Tea*, which is not a *the alla crema*, as some Italian students may think, but a British traditional afternoon meal consisting of tea and scones. This concept has no equivalent in the Italian culture and translators should adapt it according to the new context of culture and the context of situation.

In Chapter 3, the focus will be on co-text, collocation and phraseology. However, before moving to the analysis of meaning by collocation, the next chapter will explore corpora (that is to say collections of texts electronically stored) and the software which allow linguists to analyse language systematically and classify repeated patterns of language.

2 Corpora

2.1 What is a corpus?

A corpus (pl. corpora) is a large collection of texts electronically stored on a computer. These texts contain authentic language used in real situations and can represent both the language used in speech and in writing.
A corpus can be used for a number of reasons. Some of them are listed here:

- to check patterns of the language and its lexico-grammatical features;
- to check the use of words;
- to compare the use of words in different varieties of the same language (for example either in the language of economics or in the language of medicine and so on …);
- to compare and contrast translation equivalents across different languages;
- to draw examples for the preparation of teaching material;
- to obtain a list of the phraseology and the terminology of a language and its varieties;
- etc. etc.… .

Technology has made the procedure of corpus assembling easier and easier both because of the large number of texts available on t he internet and of the availability of more powerful computers which can store huge amounts of bytes without slowing down the proper working of a computer.
General English corpora are made up of millions of words in order to be representative of the whole English language. Two examples are:

- The British National Corpus (BNC), a 100 million word corpus of modern British English texts, both written and spoken, made available in 1995;
- The Bank of English (BoE), a 450 million word monitor[8] corpus under continuous development at the university of Birmingham since 1980[9].

Corpora may be of different types. These types depend on the type of language we are going to analyse. As seen in the previous chapter, language is dynamic and acquires different features depending on the social event in which it is used. A list of the main categories of corpora is provided below (adapted from Bowker and Pearson 2001:11-12 and from Teubert 1996:245-246; 2000):

General reference corpus: a general reference corpus is designed to be representative of a given language as a whole and can therefore be used to obtain insights on that language. It is usually constituted of a series of text types (both spoken and written) and focuses on the language used by ordinary people in everyday situations (newspaper, fiction, radio and television broadcasts, etc.);
Special purpose corpus: a corpus which focuses on a particular aspect of language, that is to say on a particular subject field, text type or language variety. A special purpose corpus may be a corpus constituted of tourist websites, or of articles from sports newspapers. The insights we obtain from this type of corpus are only valid for the type of language contained in it.
Monolingual corpus: a monolingual corpus contains texts in only one language.
Multilingual corpus: a multilingual corpus contains texts in two or more languages and can be *comparable* or *parallel*. A multilingual corpus is comparable when it is constituted by two or more sets of texts which have similar composition. Similar composition means that all the texts contained in the corpus have the same communicative function, they all deal with the same topic, they are all of the same type of text. The texts contained in a comparable corpus are all

[8] A monitor corpus is a type of corpus which is constantly updated to monitor changes in a given language.

[9] For details of the BNC visit http:// info.ox.ac.uk/bnc Information on the BoE is available at the Collind COBUILD website, http://titania.cobuild.collins.co.uk/boe_info.html

original texts and no translations are included. For example, a comparable corpus may be constituted of British newspaper articles on the economic recession and Italian newspaper articles on the same topic; similarly, it can be composed of three subcorpora collecting the speeches delivered by Barack Obama, Gordon Brown and Mario Monti. Analysing the original texts contained in these corpora we can make observations on the features (lexical, syntactic, ...) of these languages and we can compare them in order to detect differences and similarities.
A multilingual corpus is defined parallel when it contains original texts in one language and their translations in another language. For example, a parallel corpus may be constituted by original texts of fairy tales in English and their translations into Italian (French, German, ...). Parallel corpora can provide examples of how equivalence has been established by translators and what translation strategies have been adopted at different stages. They can be identified as a big repository containing previous translators' choices (Teubert 2000).
In order to carry out a corpus-based analysis, first a corpus to analyse is needed. There are different corpora that can be used (many of them are available on the internet). Students taking their first steps in corpus analysis should use small corpora (about 80,000/100,000 words or even less depending on the purpose of the analysis) which contain texts dealing with a specific topic. Johansson (1991: 305-6) suggests that in spite of the undoubted advantages of large corpora, there is still something to be said "for smaller, carefully constructed sample corpora which can be analysed exhaustively in a variety of ways". Of course, the results of an analysis based on a specific topic will only be valid for that specific language domain and not for the language as a whole.
Among the most practical applications of corpora, particularly of special purpose corpora, there is the elaboration of wordlists and patterns of the language of a given topic. We may be requested, for example, to write or to talk about the "economic recession" but we may not be familiar with the terminology and the phraseology associated with it. We can, therefore, search the net to identify and download a number of newspaper articles which deal with global recession. A software will help us to analyse these texts and create useful *wordlists* and *concordances*.

Wordlists are a list of all the words contained in the texts chosen for analysis. These words are listed in frequency order or in alphabetical order by the software, as visible in the figure below:

to	175
of	164
in	140
a	118
and	116
BBC	96
is	82
The	78
that	66
function	55
by	55
will	55
for	53

Concordances allow the researcher to identify and analyse the linguistic co-text of a word. The format data will acquire is called KWIC which stands for Key Word In Context. The *node word* is aligned in the centre and is preceded and followed by its co-text. The words frequently preceding and following the node word in a span of five words on the left and five words on the right are called *collocates* and form with the node words repeated string of words called *patterns*.

hat does it mean? The steelworkers say it is about putting	stimulus	cash - American taxpayers' dollars - towards
y Philippa Thomas BBC News, Washington The massive economic	stimulus	package being debated in the US Senate this w
t three million jobs. President Obama has made passing the	stimulus	package his priority, saying that millions mo
te vote. "No other nation's parliament has refused a major	stimulus	package in the current environment of unprece
g to the independent Congressional Budget Office (CBO), the	stimulus	package is likely to reduce the severity of t
ama aims to create 3.5 million new jobs, but others say the	stimulus	package could create between 1.2 and 3.6 mill
me. Meanwhile, the US Senate has backed an $838bn economic	stimulus	package which will now have to be reconciled
r. But even if Mr Obama gets rapid approval for his $800bn	stimulus	plan - which has passed the House of Represen
ent said. The government says that all the measures in the	stimulus	plan are temporary and it is committed in the
disaster if radical action is not taken. Why has such a big	stimulus	plan been proposed? The US economy is enteri
n(){}, write: function(){}, show: function(){}}; Australian	stimulus	plan blocked The Australian Senate has rejec
American" provision attached to President Obama's economic	stimulus	plan clear its first hurdle in the House of R
h chambers of Congress. What is in the stimulus plan? The	stimulus	plan includes a combination of measures desig
e an impact on unemployment. How will it be paid for? The	stimulus	plan will be funded by borrowing money - push
tors in the US Congress have agreed a slimmed-down economic	stimulus	plan worth about $789bn (Â£549bn) aimed at bo
n(){}, write: function(){}, show: function(){}}; Q&A: Obama	stimulus	plan Negotiators in the US Congress have agr
e Labor government's 42bn Australian dollar ($27bn; Â£19bn)	stimulus	plan. The bill was voted down after an indep
it is passed by both chambers of Congress. What is in the	stimulus	plan? The stimulus plan includes a combinati
tion(){}, write: function(){}, show: function(){}}; Will US	stimulus	trigger a trade war? By Philippa Thomas BBC
ely. The CBO also says that although only a portion of the	stimulus	will be spent in 2009, the bulk of the money
n infrastructure projects which make up a large part of the	stimulus	package. It is also unclear how many jobs wi
k. "While much of the world is focused on bank rescues and	stimulus	packages, we should not forget that poor peop

As visible in the figure reported above, if we choose to analyse the word *stimulus* which is one of the most frequent in the language of "economic recession"[10], its collocates are immediately visible and we are provided with a pattern such as "*pass*/*refuse*/*debate* + *economic stimulus package*", which we can add to our database on global recession.

Collocates and patterns of language will be dealt extensively in chapters 3 and 4.

In order to obtain *wordlists* and *concordances,* we need a software that can be used for corpus analysis. Some of them are listed below[11]:

- **WordSmith Tools** by Mike Scott (A set of tools for text analysis. Includes a *wordlister*, a *concordancer*, a *keyword analyzer* and more. Distributed by Oxford University Press)
- **TextSTAT** - Free concordance software for Windows and Linux
- **Microconcord** (DOS version of WS Tools *concordancer*)
- **ConcApp** concordancing programs (Freeware)
- **Concordance** (Wordlists, concordances. For publishing concordances on the WEB. By R.J.C. Watt)
- **WordExpert** (A concordancer for technical translators by myteam-Software)
- **Monoconc** (*Concordancer* by M. Barlow, distributed by Athelstan)
- **Paraconc** (Concordancer for parallel texts, by M. Barlow)
- **Multiconcord** (Concordancer for parallel texts, by David Wools)
- **PWA**- The Plug Word Aligner A collection of tools for the automatic alignment of word correspondences in bilingual parallel texts
- **LEXA** *Corpus Processing Software* distributed by ICAME.
- **DBT** (*Corpus Processing Software* developed by Eugenio Picchi at the Istituto di Linguistica Computazionale del CNR in Pisa).

[10] The corpus of Economic recession is constituted of BBC newspaper articles downloaded from www.bbc.co.uk from February to March 2011.

[11] The source of this list is Federico Zanettin's website at http://www.federicozanettin.net/sslmit/cl.htm

— **TactWEB** (*Corpus Processing Software* developed by John Bradley and Lidio Presutti, University of Toronto.

Some software is available for free and can be downloaded directly from the internet.
In the following section, a corpus of newspaper articles on the economic recession will be used to describe the features of TextSTAT, a concordance software freely downloadable from the internet[12], and WordSmith Tools, one of the most widely used software for corpus analysis..

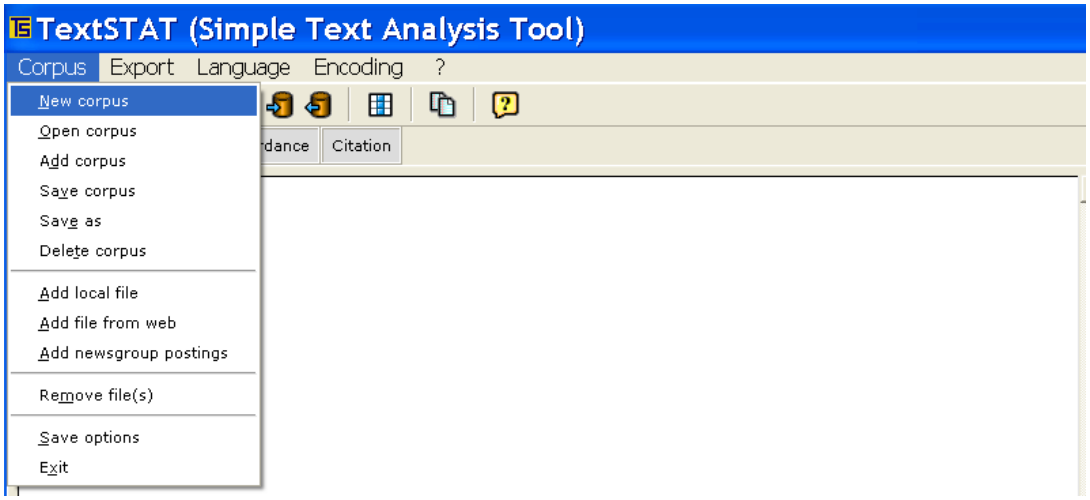
2.2 TextSTAT and WordSmith Tools

TextSTAT

TextSTAT[13] is a very simple and basic programme and only produces wordlists and concordances. However, it is ideal for those students who take their first steps into the world of corpus linguistics.
With TextSTAT it is possible to learn how often a certain word occurs or in what contexts it is used.
To use TextSTAT, first a corpus should be created by clicking on "Corpus" and then on "New corpus". (It is possible to open an existing corpus which has already been created in previous applications).

[12] http://www.niederlandistik.fu-berlin.de/textstat/software-en.html

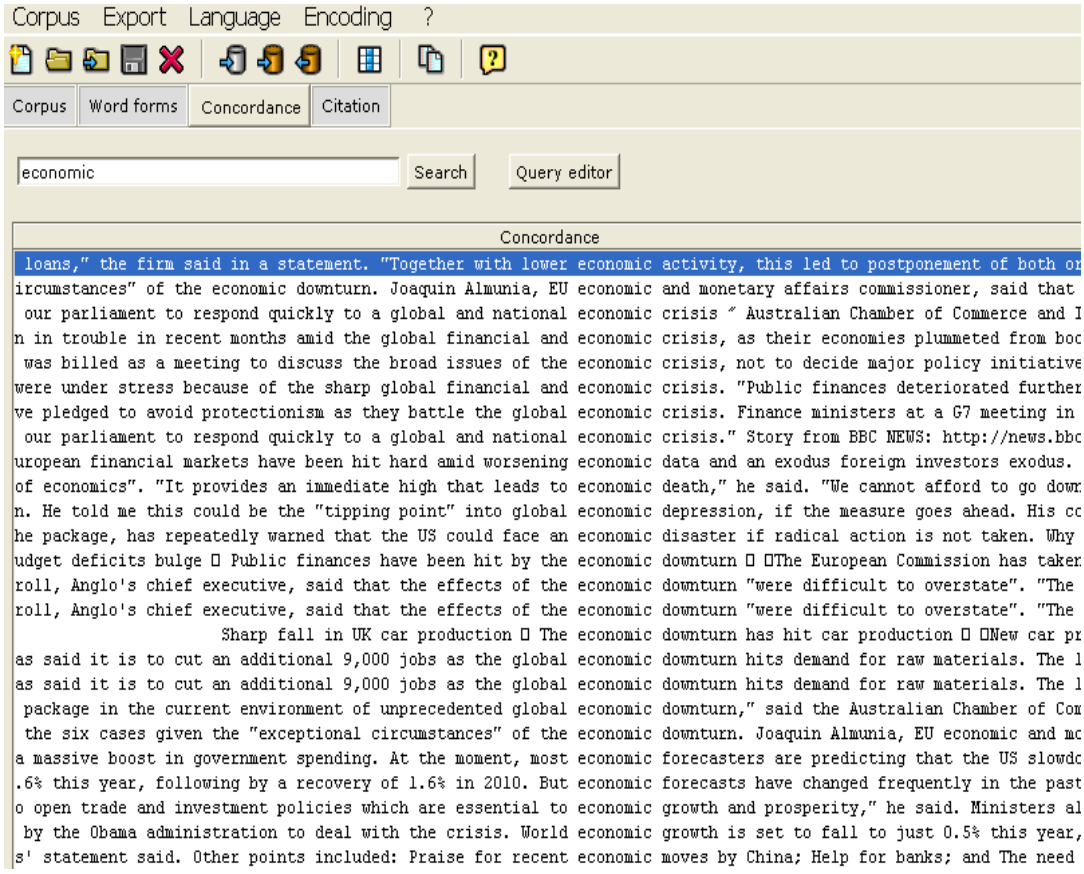
[13] A detailed description of how to use TextSTAT is provided by its creator Mattias Huening at http://www.niederlandistik.fu-berlin.de/textstat/TextSTAT-Doku-EN.html



After creating a new corpus, it is possible to add the files in your corpus by clicking on the second cylinder in the upper toolbar.

When your texts have been uploaded, click on the tab sheet "Word forms" and then on " Frequency list". A list of all the words contained in the corpus in frequency order can thus be obtained. The number of words constituting our corpus is indicated in the lower bar.

By clicking twice on each word of the wordlist, it is possible to have the concordance of the word chosen, which contains the node word and its co-text as it occurs in all the texts selected for analysis. The table below provides the concordance of the word *economic*. As can be easily seen, the words immediately following and preceding the node word can be sorted alphabetically in order to make the retrieval of recurrent patterns easier.



Furthermore, by clicking twice on each concordance line it is possible to see the original document which contains the string selected. This action corresponds to the tab sheet "citation".

WordSmith Tools[14]

WordSmith Tools has many features in common with TextSTAT but it is a more advanced tool for corpus analysis. The main tools it provides are *Concord*, *KeyWords* and *WordList* but other more specific and complex tools are also available.

Concord provides the concordance of the node word selected as well as a list of its collocates and of the recurring clusters or phrases (Scott 2001).

[14] A step-by-step guide to the use of version 5.0 of this software is provided at http://www.lexically.net/wordsmith/step_by_step/index.html.

KeyWords works with two wordlists originated by two different corpora, generally a co rpus and a b igger reference corpus. These wordlists can be compared in order to identify 'key words', that is to say those words which are more (or sometime less) frequent in the main corpus than in the reference corpus. These words are defined as keywords in that they are a key in the analysis of the language of a genre, of a domain or even of the works or speeches of a single author. Milizia and Spinzi (2008; see also Phillips 1989; Berber-Sardinha 2004; Scott 2006; Scott & Tribble 2006) have analysed the word *people* in the speeches of Tony Blair and George W. Bush. The word is heavily used by both politicians. They notice, however, that in the British list *the Iraqi people* is more frequently used than *the British people*, ranking 7th, whereas *the American people* ranks first in the American corpus. Scrolling down the list, it is clear that the main concerns of the two politicians are different.

WordList, as in TextSTAT, creates word lists, ordering them by frequency and alphabetically. However, *WordSmith* can generate wordlists that can be either one-word list or up-to-five-word lists, thus allowing the researcher to identify the most frequent patterns of a language in a few seconds. See figure below:

THE HOUSE OF REPRESENTATIVES	9	0,06
AT THE END OF	7	0,05
BY THE END OF	7	0,05
CARS REGISTERED IN THE	6	0,04
ECONOMY SHRANK BY #	6	0,04
IN THE US SENATE	6	0,04
NEW CARS REGISTERED IN	6	0,04
REGISTERED IN THE UK	6	0,04
THE END OF THE	6	0,04
HAVE BEEN HIT BY	5	0,03
IN THE FOURTH QUARTER	5	0,03
IN THE THIRD QUARTER	5	0,03
THE END OF #	5	0,03
THREE MONTHS OF #	5	0,03
A GLOBAL AND NATIONAL	4	0,03
A WIDER ECONOMIC SLOWDOWN	4	0,03
AMID A WIDER ECONOMIC	4	0,03
AND NATIONAL ECONOMIC CRISIS	4	0,03
AND REDUCE WORKERS HOURS	4	0,03
AT THE FIRM WOULD	4	0,03
AUSTRALIAN CHAMBER OF COMMERCE	4	0,03
BE OFFERED VOLUNTARY REDUNDANCY	4	0,03

2.3 Assembling corpora

The advantages of using corpora will be described in the rest of this book, particularly relating them to the fields of translation and language learning.
However, one of the main advantages of corpora lies in the authenticity of the texts contained in them which allow us to validate the theory and to have some insights into actual language that would not be possible using only our intuition of native speakers of a language.
For this reason, in order to be representative of a language and "to capture the regularities of a l anguage" (Tognini Bonelli, 2001:53), corpora should be assembled on the basis of a number of criteria. In the EAGLES recommendations on corpus typology (1996) linguistic criteria may be of two different types: 1. external, in that they concern the participants, the occasion, the social setting or the communicative function of the pieces of language; and 2. internal, in that they concern the recurrence of language patterns within the pieces of language. According to Sinclair (2005: 1) and Clear (1992) corpora should be designed and constructed exclusively on external criteria. Texts should be selected on the basis of their communicative function in the community in which they have been produced.
Tognini Bonelli (2001:54ff) analyses some authoritative definitions of corpus provided by Francis (1992:17), Sinclair (1991:171), Aarts (1991:45), the EAGLES project report (1996:2.1) and identifies three main issues which are extremely relevant to the process of text selection and corpus assembling. The issues are:

1. Authenticity of the texts included in the corpus.
All texts should represent language used in authentic language events. Invented examples or texts fabricated by the linguist cannot be the object of analysis in that they are not representative of real language (see Biber 1988; Aarts 1991; Sinclair

1991, 2005[15]; Johansson 1995; Carter and McCarthy 2003; Baroni and Bernardini 2006, to name but a few);

2. The representativeness of language included in the corpus.

Sinclair (2005: 2) says that "corpus builders should strive to make their corpus as representative as possible of the language from which it is chosen". It means that texts should be chosen according to the purpose of the analysis.

The Brown Corpus of Standard American English was the first of the modern, computer readable, general corpora[16]. The corpus contains one million words of American English texts printed in 1961. The texts for the corpus are samples of 15 different text categories. Nowadays, it is considered too small to be a good standard reference for the English language, particularly if compared to the Bank of English[17] which, at the time of writing, amounts to 400 million words. It is a collection of samples of modern English language and contains both written and spoken samples from British, US, Australian and Canadian sources. Written texts come from newspapers, magazines, fiction and non-fiction books, brochures, leaflets, reports, letters, and so on. Spoken texts are transcriptions of everyday casual conversation, radio broadcasts, meetings, interviews and discussions, etc. The corpus is constantly updated so that the corpus can be representative of the English which most people read, write, speak and hear every day of their lives. It is used by the Collins COBUILD Advanced Learner's English Dictionary as evidence of patterns of word combination, word frequencies, uses of particular words, and of meaning disambiguation.

Specialised corpora are made up of texts that belong to the same genre and deal with a specific topic. Sinclair (2001:xi) says that "if it is a general corpus, researchers expect to find in it information about the language as a whole, and if it

[15] Sinclair, J. (2005), *Corpus and Text - Basic Principles*, in Developing Linguistic Corpora: a Guide to Good Practice, Oxford, Oxbow Books, pp. 1-16.

[16] See http://www.essex.ac.uk/linguistics/clmt/w3c/corpus_ling/content/corpora/list/private/brown/brown.html

[17] See http://www.titania.bham.ac.uk/docs/about.htm

is a more specialised corpus, then the characteristics of the genre will be discoverable".

3. The sampling criteria used in the selection of texts.

According to Sinclair (2005:3) common sampling criteria used in the selection of texts for corpora include:

(a) the mode of the text; whether the language originates in speech or writing or in electronic mode;
(b) the type of text; for example if written, whether a book, a journal, a notice or a letter;
(c) the domain of the text; for example whether academic or popular;
(d) the language or languages or language varieties of the corpus;
(e) the location of the texts; for example (the English of) UK or Australia;
(f) the date of the texts.

All these criteria should be taken into account when assembling a corpus. If the purpose of analysis is the academic written language the corpus cannot include spoken texts or articles which have been published on a magazine rather than on a scientific journal. A corpus on academic written language should be constituted by scientific articles from different scientific domains (such as history, literature, language, biology, economics, mathematics, …). Furthermore, a decision should be made on whether to select only articles written by native speakers of English or to consider articles from all the scientific communities. Texts should have been published in a limited time span: selecting together scientific articles published in 1960 and scientific articles published in 2009 should be avoided; language changes over time and considering together texts in a too wide time span would make it impossible to have a clear picture of the features of modern academic written language. However, if our main aim is that of analysing this type of language as it has changed across the years, two corpora can be assembled: one made up of scientific articles published, for example, between 1960-1970 and another corpus of articles published between 1990-2000. The two wordlists

obtained could be compared in order to identify the key words which represent the changes that the academic language has undergone across the years.

2.4 Conclusion

This chapter has described the features of corpora as used in Corpus Linguistics. Corpora are large collections of texts stored in electronic format and allow both linguists and students to obtain different types of insights into a language or a variety of a language. The size of corpora may vary according to the purpose of our analysis. Observations on the language as a whole and as used in everyday contexts by ordinary people can only be obtained from big corpora constituted of millions of words. Conversely, special corpora, which although small are specialised in their content, allow us to get insight into the typical phraseology, expressions and lexico-grammatical features of a variety of a language.

The internet has many features in common with corpora. It is constituted of millions of texts belonging to different text types. However, it also has a number of shortcomings. The main problem is, obviously, constituted by authorship. Most of the times we do not have any information on who has written a text stored on the internet, we do not know their nationality, age, gender, social class, education and so on. Furthermore, information such as the date of the text and the location of the text are usually unknown.

Nevertheless, if its limits are properly taken into account, the internet may be a valid support for students when they need to check the correctness or the frequency of usage of some expressions and collocations they use in language production.

3 Collocation and units of meaning

As seen in the previous chapter, the increasing use of corpora has allowed researchers to identify systematically sets of words frequently co-occurring in language. These sets of words have been variously defined. However, according to their nature they can be put along a continuum which sees at one extreme associations of words where one item can be replaced with no change in meaning– that is to say open collocations – and at the other extreme associations of words whose components are fixed and cannot be replaced by any other elements– that is to say idiomatic sentences.

As Firth (1951:190-215) states, words do not occur in isolation and, for this reason, they should be studied in their linguistic context and their patterns of occurrence should be systematically taken into account.

This chapter provides a theoretical and practical overview of the phenomenon of collocation and of what Sinclair (1996) defines 'units of meaning'.

3.1 Firth and the notion of collocation

The British linguist J.R. Firth was the first to elaborate the concept of collocation. In his later works, Firth focused his attention mostly on the notions of prosodic analysis in phonology and in his paper "Modes of Meanings" (1951a:190-215 in 1957c) he paid much more attention to the lexical dimension (or mode) of meaning. The meaning that originated by this dimension was defined "meaning by collocation"(cf. Langendoen 1968c: 62).

Firth distinguishes between "general or usual collocations" and "more restricted technical or personal collocations". He says (1951a: 195-196)

> The commonest sentences in which the words *horse, cow, pig, swine, dog* are used with adjectives in nominal phrases, and also with verbs in the simple present, indicate characteristic distributions in collocability which may be regarded as a label of meaning in describing the English of any particular social group or indeed of one person. The word 'time' can be used in collocations with or without articles, determinatives, or pronouns. And it can be collocated with *saved, spent, wasted, frittered away,* with *presses, flies,* and with a variety of particles, even with *no.*

Firth assumes that the meaning of words is not fixed and independent but it is strictly correlated with the context it occurs within. His well-known slogan 'you shall know a word by the company it keeps' exemplifies this strong dependence of words on their use and on their possible collocations.

> The habitual collocations in which words under study appear are quite simply the mere word accompaniment, the other word-material in which they are most commonly or most characteristically embedded. It can safely be stated that part of the 'meaning' of *cows* can be indicated by such collocations as *They are milking the cows*, *Cows give milk*. The word *tigresses* or *lionesses* are not so collocated and are already clearly separated in meaning at the *collocational level*. (in Palmer 1968: 180)

Firth observes that the collocation of a word is not just a 'juxtaposition' but it is an order of *mutually expectancy*. This is why he refers to 'meaning by collocation'(1957:195-196):

> Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark,* and of *dark* of course, collocation with *night.*

Its interest in meaning by collocation is evident in its proposal for dictionary making (1957:26). He suggests collecting the collocations of the selected words into a manageable number of sets. Each group of collocations, therefore, will

suggest an arbitrary definition of the word, compound or phrase.
Firth lived in a period in which corpora had not appeared yet and for this reason he refers to 'informants' for collecting materials and definitions of words. Nonetheless his proposal to group words and their collocates sounds very modern since it represents what has been advocated for years and recently realised in modern dictionaries.

3.2 Sinclair: collocation and the principles at the basis of language

Firthian theories, briefly described above, were taken on a nd developed by Sinclair, who was one of Firth's students at Edinburgh University. As the title of his book 'Corpus Concordance Collocation' (1991) clearly shows, he considers the notion of collocation in the light of corpus evidence and defines collocation as follows (1991:170):

> Collocation is the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening. Collocations can be dramatic and interesting because unexpected, or they can be important in the lexical structure of the language because of being frequently repeated. (…) Collocation, in its purest sense, as used in this book, recognizes only the lexical co-occurrence of words.

Sinclair considers collocation as the co-occurrence of two or more words in a span of maximum for words either on the left and on the right of the word analysed. The following example is taken from a corpus of BBC articles dealing with the economic crisis and downloaded in a period from February and March 2011. The node word chosen is *downturn*:

```
onal 9,000 jobs as the global economic downturn hits demand for raw
 said that the effects of the economic downturn "were difficult to
onment of unprecedented global economic downturn," said the Australian
 finances have been hit by the economic downturn □ □The European
all in UK car production □ The economic downturn has hit car production □
 have been particularly hard hit by the downturn. They account for about
```

Frequent collocates of *downturn* are immediately visible on its left and are constituted by *economic* and *global*. However, a closer look at the concordance also reveals the presence of another collocate, the verb *hit*, which may also be found at a distance of three or four items on the left of the node word, that is to say in a 'maximum of four words intervening'. Words occurring in positions L - 5 (that is to say on the left (L) of the node word and at a distance of 5 items from it) and in position R+5 (that is to say on the right of the node word and at a distance of 5 items from it) are not usually collocates.

This example clearly shows how "words enter into meaningful relations with other words around them" (Sinclair, 1996:71) and make meanings by their combination. The word *downturn* means 'a reduction in economic or business activity' (Macmillan, 2002) but the combination with *global* and *economic* adds more details to it, thus making it acquire the meaning of global recession.

The phenomenon of collocation describes the strong attraction existing between words. According to Sinclair (1991:1996), if words attract themselves, words in texts cannot be chosen independently of one another. There exist lexical constraints which

operate at the level of word choice and since their effects are visible on repeated language patterns they can be systematically counted and analysed.

In the example below, the node word is *expected* as used in a corpus of UK weather forecasts downloaded from the website www.metoffice.gov.uk in a period ranging from February to March 2011. The use of this word in the language of weather forecasts is lexically constrained as shown by some examples reported below:

```
  Unsettled and sometimes windy weather is expected through this period,
frosts and overnight fog. Temperatures are expected to be generally near
 and central regions, the drier weather is expected to continue, with
                          Temperatures are expected to be close to or
```

The word *expected* mainly occurs with the words *weather* and *temperatures* which represent the lexical constraint in its usage. The language pattern including *expected* is

temperatures/*weather* + to be + *expected* + (to)

and it can be counted and analysed because it frequently occurs in this type of language. It should be borne in mind, however, that language patterns should occur a minimum of twice to be considered worth analysing.
The isolated meaning of *expected* would not help us understand the meaning of the pattern because it is the combination of the meanings of *temperatures* or *weather* plus the verb *expected* which explain the overall meaning.
If words attract themselves, therefore, complete freedom of word choice as well as complete determination is very rare. For this reason, Sinclair elaborates two principles which account for how language actually works and which explain the way in which meaning arises from language text: the *open-choice principle* and the *idiom principle* (1991;1996). According to Sinclair (1991:109), the *open-choice principle* is

> a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word, phrase, or clause), a large range of choice opens up and the only restraint is grammaticalness.

The open choice principle has also been called 'slot-and-filler' model, in that at each slot virtually any word may occur which would be restrained only by grammaticalness. The tendency towards the open choice principle is labelled by Sinclair 'terminological tendency' (1996), that is to say the tendency for a word to have a fixed meaning in reference to the world. But as seen above, words enter into meaningful relations with other words and tend to retain traces of these encounters. The traces of these encounters are frequently occurring language patterns. A 'slot-and-filler model' does not take into account the phraseological tendency of language, that is to say the phenomenon of lexical attraction between

words. For this reason, Sinclair (1991:110) elaborates a second principle, which he calls the idiom principle:

> The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments.

Sinclair explains that the existence of such pre-packed phrases may be due to a number of reasons but what is crucial is that the idiom principle acts massively and predominantly with respect to the open-choice which functions only alternatively.
The following phrases taken from corpora are an example of how predominant the idiom principle is in language processing. They have been chosen as examples in that they frequently occur in the corpora chosen for analysis:

- *with showers or longer spells of rain* (weather forecast language);
- *with outbreaks of rain* (weather forecast language);
- *the Austrian/Dutch/eurozone/French economy shrank by ...* (news articles on economic crisis);
- *hit by economic slowdown* (news articles on economic crisis);
- *beautiful views of the countryside* (tourist websites);
- *conveniently situated for exploring* (tourist websites);
- *a car bomb exploded* (news articles on Mahgreb crisis);
- nr. *people were wounded* (news articles on Mahgreb crisis);
- *due regard to best interest of minor children* (EU documents);
- *applications for family reunification* (EU documents);

According to Sinclair (1991:112) the principle of idiom can be elevated "from being a rather minor feature, compared with grammar, to being at least as important as grammar in the explanation of how meaning arises in text".
Following this generalization, he observes that if two words collocate significantly, they are the result of a single choice. The idiom principle suggests,

therefore, that language is not stored as individual morphemes but as chunks which are retrieved in these pre-packed sentences.

The phenomenon of collocation is at the basis of the idiom principle. However, although the concept of collocation suggests a process of crystallization of words, this fixedness is rarely absolute. Sinclair (1991:111-112) illustrates the variation within the idiom principle and points out that:

1. Many phrases have an indeterminate extent. Sinclair (*ibidem*) provides the example of *set eyes on.* This phrase attract a pronoun subject, and words such as *never*, *the moment*, *the first time*, and *has* as an auxiliary. The extent of the phrase is indeterminate as there is not a clear distinction between what is integral to the phrase and what is in the nature of the collocational attraction.
2. Many phrases allow internal lexical variation. A very frequent phrase in the Italian language of tourism is *immersione nella natura*. A frequent lexical variations of this phrase is *tuffo nella natura*.
3. Many phrases allow internal lexical syntactic variation. In the phrase *a stretto contatto con la natura*, the adjective *stretto* may be replaced by *diretto* as in *a diretto contatto con la natura.*
4. Many phrases allow some variation in word order. In the example reported above, the adjective *diretto* may occur before or after the noun *contatto* as in *a diretto contatto con la natura* or *a contatto diretto con la natura*.
5. Many uses of words and phrases attract other words in strong collocation. Sinclair (1991:112) provides the examples of *hard work*, *hard luck, hard evidence, hard facts*.
6. Many uses of words and phrases show a tendency to co-occur with certain grammatical choices. For example, the phrase *conveniently situated* is usually followed by the prepositions *to* + the infinitive form of the verb or *for* + the –ing form of the verb.
7. Many uses of words and phrases show a tendency to occur in a certain semantic environment. For example, the phrase reported above

> *conveniently situated for/to* is usually associated with verbs describing activities, such as *tour*, *explore*, *visit*.

Language is seen, therefore, as a dynamic process, where words do not remain perpetually independent in their patterning but they "begin to retain traces of repeated events in their usage, and expectations of events such as collocation arise" (1996:82).

Biber, Conrad and Leech (2002:443) support the idiom principle and maintain that the formulaic nature of speech is reflected in 'lexical bundles', that is to say, sequences of words which are frequently re-used, and therefore become 'prefabricated chunks' that speakers and writers can easily retrieve from their memory and use again and again as text building blocks. They consider conversation as being more repetitive than written registers, for this reason, lexical bundles may be more identifiable in speech. It needs to be said, however, that academic prose makes considerable use of prefabricated blocks of text as well, but it involves different linguistic features. Lexical bundles in academic prose typically involve parts of noun phrases and prepositional phrases, whereas lexical bundles in conversation typically involve the beginning of a finite clause – especially with a pronoun as subject followed by a frequent verb of saying or thinking. Some examples of lexical bundles in conversation provided by Biber, Conrad and Leech (2002: 445) are: *I don't know what, I said to him, I tell you what, I was going to, I would like to, you know what I, it's going to be, know what I mean, what do you mean ... .*
Milizia (2012) provides some examples of lexical bundles in the speeches of President Obama. The following examples are taken from a corpus constituted of speeches delivered by President Obama in a period ranging from 2009 to 2011:

- *to make sure that ...*
- *I want to thank ...*
- *thank you very much*
- *we are going to ...*

- *is going to be ...*
- *when it comes to ...*
- *one of the things ...*
- *as the result of*

These phrases occur very frequently in this type of language and their frequency of usage clearly show that they come up as single choices in the mind of the speaker who utters them.

3.3 Collocation and the phenomenon of delexicalisation

The phenomenon of collocation is linked to another important phenomenon: delexicalisation.
Dealing with this phenomenon, Sinclair (1992: 16ff) starts from the assumption that the meaning of words chosen together is different from their independent meanings. This is due to the fact that words chosen together undergo a process called delexicalisation, that is to say they lose part of their meaning. Sinclair says (1991:113):

> There is a broad general tendency for frequent words, or frequent senses of words, to have less of a clear and independent meaning than less frequent words or senses. These meanings of frequent words are difficult to identify and explain; and, with the very frequent words, we are reduced to talking about uses rather than meanings. The tendency can be seen as a progressive delexicalization, or reduction of the distinctive contribution made by that word to the meaning.

An example of delexicalisation is given by the word *welcome* when it occurs in the pattern frequently occurring in the language of tourism:

guests/visitors + *are* + *welcome* + *to* + (semantic field of activities)

In this pattern, the word *welcome* does not have the meaning indicated by dictionaries. According to the Macmillan Dictionary (2002) "if you are welcome

or a welcome visitor at a place, people are pleased that you are there". The use of *welcome* in this pattern operates a reduction of the distinctive contribution made by *welcome* to the whole meaning. When *welcome* is followed by the preposition *to* and a verb referring to an activity, the meaning of the pattern is "you may do something if you want to" (see Macmillan Dictionary of Contemporary English, 2002). This means that the word *welcome* in this pattern is delexicalized, that is to say it has lost part of its meaning.

According to Sinclair (1991:113) "normal text is largely delexicalized, and appears to be formed by exercise of the idiom principle, with occasional switching to the open-choice principle".

Sinclair demonstrates the phenomenon of delexicalization by analysing some adjectives, which are considered by grammars as elements which may add something to the noun, restrict it or add some features to it. Sinclair (1996:98) suggests that in everyday use "there is often evidence rather of co-selection and shared meaning with the noun".

He considers the adjective 'physical' and the noun it frequently occurs with:

physical assault
physical attack
physical attributes
physical bodies
physical confrontation
physical damage
physical proximity

Sinclair (*ibidem*) suggests that the adjective *physical* does not add meaning to the noun but in a way it duplicates the meaning of the noun.

The same happens with the adjective *scientific* in *scientific experiment* and *scientific analysis* for example, the adjective is delexicalised and it is used only to dignify the following word slightly. This type of adjectives is defined 'focusing' (1992:16ff) in that they underline the meaning of the following noun. On the other hand, the type of adjectives which make a selection of the meaning of the noun are defined 'selective' (*ibidem*).

In the phrase *warm welcome,* the adjective *warm* is a focusing adjective because it duplicates and emphasises part of the meaning of *welcome*. The adjective in this

co-selection is delexicalized. Conversely, in a phrase such as *Scottish welcome*, the adjective *Scottish* is defined 'selective' because it represents a selection, a part of the meaning of the noun *welcome*. In this case, the adjective is not delexicalized.

Stubbs (1996:32ff) provides the example of the lemma pair *take a* searched in a corpus of over two million words. In only 10% of a total of 400 examples the verb *take* has the literal meaning of "grasp with the hand" or "transport". In its most common use, therefore, *take* is delexicalized and in units such as *take a deep breath* the meaning is carried by the noun and not by the verb.

Stubbs (*ibidem*) suggests that the phenomenon of delexicalization is quite common and it is fairly visible with adjectives. The two types of adjectives 'selective' and 'focusing' are distinguished by Stubbs (*ibidem*) as follows:

SELECTIVE ADJECTIVES	**FOCUSING ADJECTIVES**
outward-looking	inward-looking
independent	dependent
separate choice	co-selected with noun
adds separate meaning	repeats part of meaning of noun
narrows meaning of noun	intensifies meaning of noun

The difference between selective and focusing adjectives has important implications. The more obvious is in the translation field. In fact, the two types of adjectives would require a different approach in the search for a translation equivalent. Translating a focusing adjective could be more complex than translating a selective adjective in that knowledge of the range of collocates of the noun it qualifies is crucial. Let us take the examples reported above. The combination *a Scottish welcome* can be easily and literally translated into Italian as *un benvenuto scozzese*. Conversely, the pair *a warm welcome* requires an analysis of the collocates of *benvenuto* to realize that *caloroso* more than *caldo* intensifies the meaning of the following noun.

3.4 Collocation and the definition of meaning: naked eye

The phenomenon of collocation clearly shows that the word cannot be considered as the basic unit of language and that we should rather talk about 'units of meaning' (Sinclair,1996). The units of meaning Sinclair (*ibidem*) talks about are complex units which take into account four different types of attractions between the word and its linguistic co-text. These attractions are considered by Sinclair (1996) as steps towards the definition of meaning and are: 1) collocation; 2) colligation; 3) semantic preference; 4) semantic prosody.
As already discussed above, 1) **collocation** is the lexical attraction between two or more words (as in *economic stimulus plan, global economic downturn, warm and friendly welcome, ...*).
2) The phenomenon of **colligation** represents a type of attraction at the grammar level, that is to say the frequent co-selection of a word with a grammatical category. In the language of tourism, the word *located,* when it is used to describe the convenience of the location to visit the attractions located nearby, strongly attracts the grammatical category of adverbs and that of prepositions (as in *ideally/conveniently situated to/for ...*).
3) The step defined **semantic preference** represents an attraction between a word and one or more semantic fields. Another example from the language of tourism is provided by the frequent pattern *guests are welcome to* which strongly attracts a group of verbs belonging to the semantic field of activities (as in *guests are welcome to stroll around the farm/to join the farm activities*).
4) **Semantic prosody** represents a further step into abstraction and it is used to describe the attraction between a word and a positive, negative or neutral evaluation of that word and its collocates. Sinclair (1987) provides the example of *happen* which usually gives the unit a negative connotation and Stubbs (2002:65) makes the example of *cause* which occurs with words for unpleasant events. Its main collocates are *problems, death, damage, concern, trouble, cancer, disease.*
These attractions explain why a word cannot be considered the basic unit of language. Words combine with other words, grammatical categories, and semantic fields and create what Sinclair (1996) calls 'extended units of meaning'. He says

(1996:94) that the tendencies of words to co-occur with other words (collocation), with word classes (colligation), with set of meanings (semantic preference), and attitudes (semantic prosody) are so strong that we must expect the units of meaning to be much more extensive and varied than just a single word.

To expound his theories Sinclair provides the analysis of the collocation *naked eye*.

He stresses that there is no useful interpretation for this phrase based on the core meanings of the two words although the metaphorical extension may be obvious. Some examples of the concordance to *naked eye* (Sinclair 1996) are reported below:

agents too small to see with the	the naked eye	and so they much preferred
binaries that can be seen with	the naked eye	(very few of these) or through
our galaxy that you can see with	the naked eye	Now to expand our horizons: The
is like viewing the moon with	the naked eye	. You see a disk with some
of thing you could look at with	the naked eye	"Would you like to
it is not really visible to	the naked eye	. About five years ago, a
cannot always be perceived by	the naked eye	and said, 'As I've gotten
even though nothing is visible to	the naked eye	. We should trust our patients
the opening is not visible to	the naked eye	. Typically, the closed
photoaging changes are visible to	the naked eye	. And even more disturbing
little rooftop house. Viewed with	the naked eye	, she was nothing more than a
is visible with	the naked eye	. While stroke path can be with
outlets. These could be seen with	the naked eye	from a helicopter, and the water
human ovum is barely visible to	the naked eye	. The corpus luteum forms in the
small. It can easily be seen by	the naked eye	. The time of ovulation in
is large enough to be seen by	the naked eye	. The ovary still contains the

By analysing the collocation profile of these two words he identifies in 95% of the examples the presence of the definite article *the* at position N – 1, thus establishing that *the* is an inherent component of the phrase *the naked eye*. At position N-2 he identifies two frequent prepositions *with* and *to* and other less frequent preposition for a total of 90%. The word class preposition is thus an inherent component of the phrase and we will refer to this component not in term of collocation, but of colligation, that is to say the co-occurrence of grammatical choices (Firth 1957). At position N-3, two words dominate the picture, *see* and *visible* together with verbs and adjectives which refer to the semantic field of "visibility". The attraction between a word or a set of words and a given semantic field is called semantic preference (Louw 1993; Stubbs 1996; Sinclair 1997; Hoey 1997).

The traces of repeated events retained by this unit of meaning can be summarized as follows:

"visibility" + preposition + *the* + *naked* + *eye*"

At this point, another step in abstraction should be taken, because at closer examination one more regularity attached to the node words seems to come up. Around the unit "visibility" + preposition + *the* + *naked* + *eye*" there is an aura of meaning which refers to difficulty. This is made clear by the presence of words such as *small, faint, weak, difficult* in association with *see*, or such as *barely, rarely, just* with *visible.* Furthermore, the concept of *visibility* is also frequently associated with a negative or with modal verbs such as *can* or *could*. This aura of meaning around the unit is called semantic prosody (Louw 1993) and it is said to play a leading role in the integration of an item with its surroundings. It is the pragmatic meaning of an item. Semantic prosody may be negative, as in this case, positive or neutral and it exerts a powerful influence on words like a sort of contagion that they carry with them even when associated with other words.

Going back to *naked eye*, Sinclair maintains (*ibidem*) that this word is used to express some kind of difficulty in seeing something. At this point, a difference with the Italian counterpart *a occhio nudo* should be mentioned. The aura of meaning around the Italian items is not negative at all (see Sinclair 1996), since the unit is used to stress that something is so obvious and visible that could also be seen *a occhio nudo*.

This example shows that the relationship between lexis, grammar, semantic and pragmatic choices is strong and impossible to be divided: lexis and grammar are interdependent and strictly interrelated, no di chotomies can be accepted.

The example provided below considers the node word *immigration,* as it is used in a corpus constituted of documents downloaded from the Official

Journal of the European Union[18].

The most frequent collocates of *immigration* are *illegal, policy, policies, legal, irregular.* The instances reported below exemplify the lexical attraction between *immigration* and *illegal*.

```
 Policy priorities in the fight against illegal immigration of third-country
up the fight against all these forms of illegal immigration in a number of
 he comprehensive EU approach to combat illegal immigration is guided by a set
 mediate aim of reducing and preventing illegal immigration, this also
11. A firm policy to prevent and reduce illegal immigration could strengthen
ence of such rules may in itself reduce illegal immigration by offering
it would be unrealistic to believe that illegal immigration flows can be
should be included in the fight against illegal immigration. In the context of
eral interest of the Union in combating illegal immigration shall fully respect
A central tenet of how the EU addresses illegal immigration is the removal of
ation on harmonising means of combating illegal immigration and illegal
                        □2. Addressing illegal immigration of third-country
 its intention to ´address the issue of illegal immigration with a
s responsible for combating facilitated illegal immigration and human
ect to carriers´ obligations to prevent illegal immigration, Article 26 of the
part of how the EU continues to address illegal immigration. Informal exchange
ration comes the challenge of combating illegal immigration and human
on more effective to prevent and combat illegal immigration, and to strengthen
 last but not least, the need to tackle illegal immigration.
with it new demands on policy. Fighting illegal immigration requires particular
the action of Member States in reducing illegal immigration flows. But the
ation and determined measures to combat illegal immigration are two sides of
ress has been made in the fight against illegal immigration through the use of
gration policies and measures to combat illegal immigration lose much of their
o riguarda: □Effective measures against illegal immigration will involve the
needs and provide a real alternative to illegal immigration and the informal
n measures on how to effectively tackle illegal immigration, addressing both
isation measures and measures to tackle illegal immigration;
 plan on legal migration, fight against illegal immigration, future of the
 Policy priorities in the fight against illegal immigration of third-country
riminal organisations help to encourage illegal immigration;
 policy priorities in the fight against illegal immigration of third-country
```

A closer look at the concordance shows the presence of other types of attractions apart from the lexical one. The collocation *illegal immigration* is almost always preceded by the grammatical category of verbs. This means that verbs are its colligates. These verbs are: *fight (against), combat, reduce, prevent, tackle,*

[18] available at http://eur-lex.europa.eu

address. All these verbs are semantically correlated because they describe ways of facing unpleasant situations or conditions and represent the semantic preference of the node word. The semantic prosody of the language pattern identified is obviously negative as suggested by the verbs in its co-text and by the preposition *against.*

A further example considers the word *attacks* in a corpus which collects articles from the websites of BBC News and Sky News in the period April to May 2010. The instances provided below have been grouped according to the most frequent collocates.

The first set of examples considers *attacks* collocating with different forms of the verb *claim*:

claimed responsibility for recent bombing attacks in Baghdad. The
Taliban **claims** suicide attacks on Kabul hotels
has **claimed responsibility for** any of the attacks, the North Caucasu
The Taliban **claimed responsibility for** the attacks, which President
claimed responsibility for being behind the attacks. But Russian
Taliban has **claimed responsibility for** the attacks. Zemeri Bashary,
llah Mujahid **claimed responsibility for** the attacks on behalf of the
claimed responsibility for either of this week's attacks

As visible above, the noun *attacks* is embedded in the language pattern:

name + *claimed responsibility for the attacks (in/on)*

thus providing a perfect example of the predominance of the idiom principle in text production. The collocation *claimed responsibility for* also collocates with the singular *attack* and with the synonymous word *assaults.*

Another collocate of *attacks* is the verb *blame*, as visible below:

another station, killing 10 people. Both attacks were **blamed on**
ed in a series of apparently co-ordinated attacks **blamed on** al-Qaeda
Qaeda is **blamed for** many of the deadliest attacks in Iraq in recent
ndia **blames for** the 2008 Mumbai terrorist attacks that killed 166

However, different forms of the verb *blame* are lexically attracted by *attacks* in

two different language patterns:

(1) name + (to be) + *blamed for* + *attacks*

(2) *attacks* + (to be) + *blamed on* + name

The verb *carry out* is another frequent collocate of *attacks* as exemplified in the instances reported below:

```
e Taliban said they would carry out more attacks across the country
 litants - have also carried out revenge attacks against fellow
 grenades. Persecuted minority Sectarian attacks have been carried out
ther this assault will prove larger than attacks carried out last
ers are believed to have carried out the attacks on trains that had
ikov said its investigators believed the attacks had been carried out
tacks. It is unclear who carried out the attacks, but suspicion
```

The analysis of collocates suggests the presence of two language patterns featuring the lexical attraction between *attacks* and *carry out*:

(1) (name) + *carry out* + *attacks*

(2) *attacks* + (to be) + *carried out*

Other collocates, which will not be analysed here, are *bomb, suicide, killed, wounded* (as in *killed in attacks* or *attacks that killed 85 people*), *launch* (as in *militants launched a series of attacks*), *hit* (as in *were hit by three bomb attacks*), *stepped up* (as in *militants stepped up attacks)* and other less frequent items.

To sum up the analysis of *attacks*, let us consider in detail the four steps towards the definition of meaning:

- Collocation: *claim, responsibility, blame, carry out, bomb, suicide, killed, wounded, launch, stepped up*;
- Colligation: verbs, adjectives;
- Semantic preference: 1. items referring to attribution of responsibility (*claim, blame, responsibility, ...)*, 2. items referring to damage (*killed, wounded, hit, stroke, battered, ...*), 3. items referring to actions (*carry out, launch, step up, ...*);

- Semantic prosody: mainly negative, due to the co-occurrence with negative adjectives and verbs.

Sinclair (1996) also provides the example of the collocation *true feelings* to exemplify the four steps of analysis. Some instances are reported below:

```
     incapable of experiencing true feelings . And not just as a man, but
moment ago to share his or her true feelings with a team. Courageous
    and loneliness to mask her true feelings . As the day passed she
the Princess of Wales show her true feelings . The thousands standing
  He may not want to admit his true feelings of ambivalence because he
 in his efforts to conceal his true feelings . “I’m not ill,” she said.
   Chaucer, is to disguise his true feelings : ‘And softe sighed, lest
             If they were his true feelings . Perhaps he was suffering
    and happy hero reveals his true feelings for his friend Willie
 charmer will never reveal his true feelings ; he has to appear hard,
to his audience and hiding his true feelings behind careful
  said Taylor was aware of his true feelings for Alison, but admitted
 a man who resolutely kept his true feelings under wraps, he also manages
      Now I had to confront my true feelings about my body, another
     because I had betrayed my true feelings . I picked up the glass and
   the people who mattered. My true feelings had to be buried, the
so careful about expressing my true feelings and told them things that
            you do share your true feelings . Then you can go on to make
have been unable to share your true feelings with him. As a result, it
  be time for you to show your true feelings , and stop pretending you’re
 dreams can help indicate your true feelings at the moment – take heed of
  you were forced to hide your true feelings during childhood and became
   you cannot communicate your true feelings means you put out stress
 but abruptness betrayed their true feelings . Were they disappointed to
       seething, hiding their true feelings in adolescent petulance? I
 less open about showing their true feelings and noticeably less polite
  have little regard for their true feelings about topics they know
  the lovers who conceal their true feelings behind barbed witticisms at
```

Try to identify the extended unit of meaning in which *true feelings* is embedded by looking for collocates, colligates, semantic preference and semantic prosody.

3.5 Conclusion

This chapter has described the phenomenon of collocation and the four steps towards the definition of meaning. The examples provided show that language should be interpreted as a dynamic process where words are strictly interrelated with their co-text and their context. For this reason, a single word cannot be considered as the basic unit of language because words frequently occur with

other words and meanings arise from their combination. The Italian word *soffrire* may intuitively be associated with a negative semantic prosody due to its collocation with major or minor diseases (as in *soffrire di depressione, soffrire di distubi alimentari, ...*). However, it also collocates with *solletico* which, however frustrating, cannot be considered a disease and does not carry with it a negative connotation. This further confirms what has been described in this chapter and explains how the meaning of *soffrire* both at the lexical and pragmatic level is, therefore, strictly dependent on the words it combines with.
Another example which explains how useless considering the isolated meaning of a word may be, is provided by the word *imbottigliato*. When it collocates with *vino* it usually has a positive or neutral connotation and basically refers to a liquid which has been bottled (as *in questa dicitura assicura che il vino è stato imbottigliato nello stesso luogo dove è stato vinificato*). However, when its collocate is *traffico* it obviously acquires a negative connotation and means 'stuck in a traffic jam' (as in *sono rimasto imbottigliato nel traffico*).
Implications of the strict relationship between the item and its environment (Tognini Bonelli 2001) are visible in the process of translation as described in the following chapter. Students who are taking their first steps in the translation field are usually and too often inclined to translate word for word and consider the isolated meaning of words rather than the extended unit of meaning. If we consider the examples reported above, possible results of a word for word translation would be *I remained bottled in the traffic* whose final aim or function would not be that of describing a state or a condition or a justification for being late but of making our English-speaking interlocutors laugh.
Similarly, language learners should avoid to learn lists of words both because they need details on their usage (their collocates, semantic preference, semantic prosody, ...) if they want to use them properly and in the right contexts, and because learning pre-packed phrases as single choices help improve fluency in speaking, reading, and writing.

4 Translation and functionally complete units of meaning

4.1 Meaning as function in context

As demonstrated in the previous chapters, the identification of meaning involves the analysis of the co-text of a word and the analysis of the context in which this item has been used. The interdependence between the item and its environment in the definition of meaning has important implications in the process of translation. As Sinclair et. al. argue (1996: 175)

> Translation equivalence at word level is not by any means the whole methodology. In many instances (…) there is no translation equivalent for the chosen word. Translation can only be achieved by first of all combining the word with one or more others; the whole phrase will then equate with a word or phrase in the other language.

The starting point in the translation process is, therefore, represented by the extended unit of meaning rather than by the word. However, the identification of the meaning of a unit of language is not enough when dealing with different languages. The pragmatic function performed by the unit has also to be identified. We need to understand why that string of language has been used in that co-text and in that context.

That is why Tognini Bonelli (2001), following in the Firthian tradition, sees meaning as function in context and considers the information provided by the context of fundamental importance. Her approach, both in language description and in contrastive work, postulates the existence of *functionally*

complete units of meaning (2001:131). These units are extended units of meaning which perform specific functions at the pragmatic level.

Tognini Bonelli (2001:31) develops a methodology which allows the translator or the language student to identify equivalent functionally complete units of meaning across languages. This methodology starts from the assumption that in order to identify equivalent units of meaning across languages, all the components that are necessary for the unit to function (collocation, colligation, semantic preference, semantic prosody) need to be identified. This approach (Tognini Bonelli 2001; Tognini Bonelli and Manca 2002) involves three steps. The aim of this methodology is contextualizing the unit to be translated by considering its co-text and identifying a network of possible equivalences between the unit in the source text (the text that has to be translated) and the target language (the language the text has to be translated into). In order to apply this methodology a comparable corpus is needed (see chapter 2 for definition). The three steps of the methodology are adapted below (see Tognini Bonelli 2001; Tognini Bonelli and Manca 2002):

Step 1: The initial node word in the source language is analysed in order to identify its collocation, colligation, semantic preference and semantic prosody.

1) from the **node word** ⟶ to **its unit of meaning**

Step 2: For each collocate of the node word a possible translation equivalent is posited by looking up in dictionaries. The item reported or believed as equivalent is investigated in order to identify the unit of meaning in which it is embedded.

2) from the **collocates of the node word** ⟶ to **their equivalents** ⟶ to the **collocates of the equivalents**

Step 3: Within the collocates of the equivalents we shall identify an adequate translation equivalent of the initial node word.

3) from the **collocates of the equivalents** ⟶ to the **translation equivalent of the initial node word**

Adopting this methodology the analysis of the co-text will reveal the presence or the absence of functionally similar patterns of language across different languages.

4.2 Applying the methodology: some examples

In order to illustrate this methodology we will provide below a series of examples obtained by working with comparable corpora. The first of these examples is represented by the analysis of the adverb *largely* carried out by Tognini Bonelli (2001:150ff).

Tognini Bonelli establishes a procedure that takes into account both context and function in the identification of translation equivalence.

The basic procedure she proposes consists in taking a first step 'via collocates' and a second step 'via function'. Let us consider the analysis in some detail.

The adverb *largely* frequently occurs in expressions such as *largely because*, *largely thanks to*, *largely as a result*, *largely due to*, thus displaying an overall function associated with cause or reason. We should note that, as she demonstrates, this is not the only function *largely* engages in, but here we will only concentrate on this as an example.

In order to find an adequate Italian translation equivalent for the unit *largely because* (to take the most common collocation pattern) Tognini Bonelli (*ibidem*) proceeds to posit, as a next step, a translation equivalent of *because* in Italian – that is *perché*. By scanning the collocational profile of *perchè*, Tognini Bonelli (*ibidem*) finds that *perchè* can be modified only by *soprattutto* (and not by other adverbs that are usually taken as the translation

equivalent of *largely*, that is *largamente* or *in larga misura*). This means that the Italian functional equivalent of *largely because* is not *in larga misura perchè* or *largamente perchè* but *soprattutto perchè*. The Italian unit performs the same function as the English unit and, for this reason, can be taken as equivalents.

Further original examples of how this methodology can be applied with different comparable corpora are provided and described below.

Let us start with the node word *reunification* as used in a corpus of EU documents. Bilingual dictionaries (see Il Ragazzini, 2005) provide *riunificazione* as possible translation equivalent. However, a look at the Italian corpus of EU documents reveals the absence of this word. This means that *reunification* and *riunificazione* are not equivalent in the language of EU documents. The methodology described above has to be applied in order to find a functionally similar translation equivalent in Italian. As suggested by Step 1, the node word has to be analysed in its collocational profile. Some instances of its concordance are provided below:

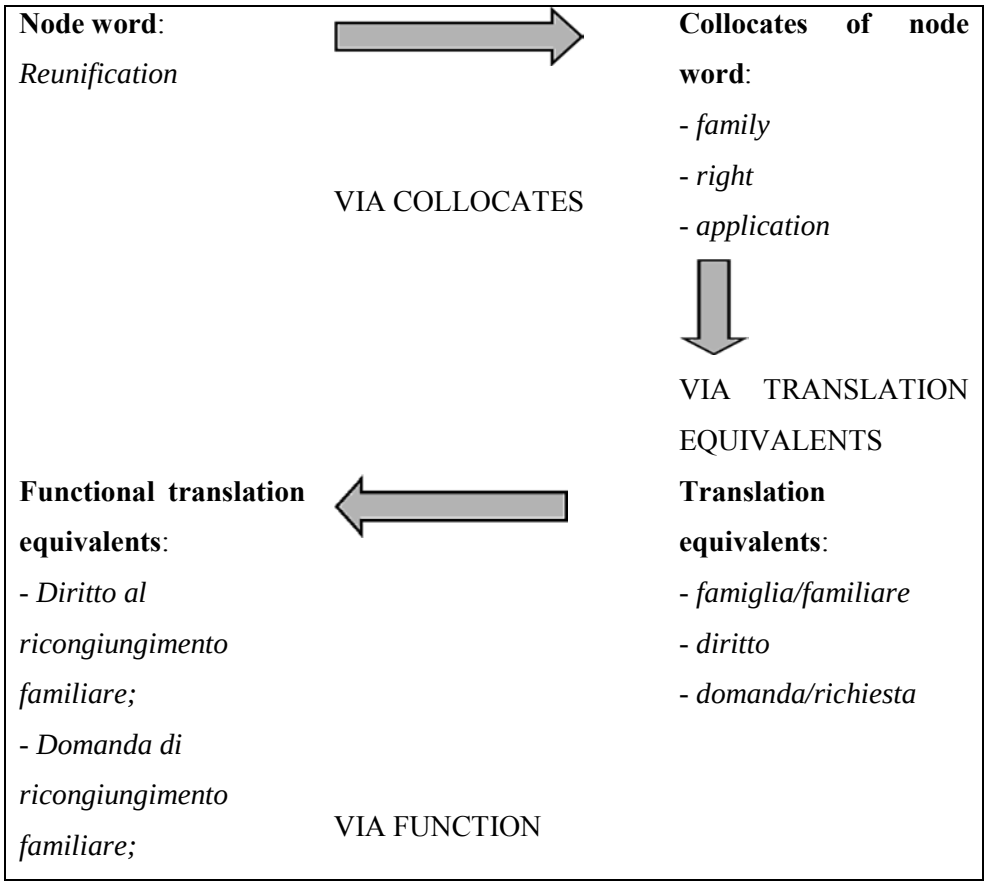
```
refused an application for family reunification has correctly
Memb er State to authorise family reunification in its territory,
y of limiting the right to family reunification of children over the
idering an application for family reunification and the Community
o require applications for family reunification of minor children to
is of ‘grounds other than’ family reunification which are not
pplying the conditions for family reunification which are prescribed
 Directive as referring to family reunification in the cases where
equivalent to a refusal of family reunification. The Council also
 for applying the right to family reunification in a harmonised
  been laid for a right of family reunification, subject to a number
 f submission of applications for reunification that are founded
 ears between the application for reunification and the issue of a
```

As visible in the instances above, the most frequent collocates of *reunification* are: *family, right, application*.

According to Step 2, for each collocate of the node word a possible equivalent has to be posited and the collocational profile of these equivalents has to be identified. *Family* can be considered equivalent both to

famiglia and *familiare*, the word *right*, in this context, is equivalent to *diritto*, and *application* has *domanda* and *richiesta* as Italian counterparts. The word *famiglia* is embedded in the patterns *indipendentemente dalla propria famiglia* and *vivere in/con la famiglia*. These units do not carry the same meaning of the English units in which *reunification* is used. Conversely, something interesting can be found in the collocational profile of *familiare.* This word is frequently preceded by *ricongiungimento* and in its co-text we also find words such as *diritto* and *domanda*. The patterns of usage are *diritto al ricongiungimento familiare* and *domanda/domande di ricongiungimento familiare.*

Starting from the collocates of the equivalents (Step 3) a translation equivalent of the initial node word has been identified. The word *reunification* in the legal context and when it co-occurs with *family, right,* and *application* can be translated into Italian with *ricongiungimento*. The presence of these patterns shows that the initial translation equivalent provided by dictionaries, that is to say *riunificazione* is not adequate in this context and similarly, that *riunificazione familiare* is not used because it has a different meaning and engages in a different function with respect to the English *family reunification.* It is likely that a more specialized bilingual legal dictionary could provide the right translation equivalent. However, the authenticity of the language contained in the corpora and the frequency of occurrence of the patterns considered for analysis represent the validity of the results obtained. To sum up the procedure:



The following example is taken from the language of tourist websites. The corpora considered for analysis are two comparable corpora constituted of websites of Italian and British farmhouses. We will start from the Italian node word *passi*, that can be literally translated into English as *steps*. Let us consider some examples from the concordance of *passi*:

```
onte al bosco della Mercanzia, a due passi da Castiglione,ecco, ben
, il paesino natio del maestro a due passi da Partigliano. Ma pochi
  nel cuore dell'Umbria verde, a due passi da Spoleto città del
 verdi colline del Montalbano, a due passi da un borgo medievale,
i queste terre basta però fare pochi passi: affacciato sullo stesso
itico mare di Capo Palinuro, a pochi passi da Paestum, dalla Certosa
 Parco naturale del Partenio a pochi passi da Roccabascerana in
orre del Sasso. Una località a pochi passi dal Centro Storico di
ico vigneto e gli olivastri, a pochi passi dal fiume. Possibilità di
la collina di Farra d'Isonzo a pochi passi dalla tenuta, dove
```

As suggested by the examples above, *passi* frequently co-occurs with *due*,

pochi, and *da* and shows a s emantic preference for geographical names (Step 1). The pattern in which it is used is

a pochi passi da + (geographical names)

Bilingual dictionaries provide *steps* as translation equivalents. However, its collocational profile in the English comparable corpus suggests a different usage of *steps*. While the Italian *passi* is used metaphorically to refer to distance, the English counterpart is used literally to refer to the steps of stairs, as in *Six shallow steps to open-plan sitting room* or *and are approached via stone steps to both front and rear.* This means that there is no literal equivalence between the two words.

In order to find an adequate translation equivalent, possible translation equivalents for each of the collocates of *passi* need to be posited and their collocational profiles have to be analysed (Step 2).

The English equivalents of *due, pochi,* and *da* are *two, few,* and *from*.

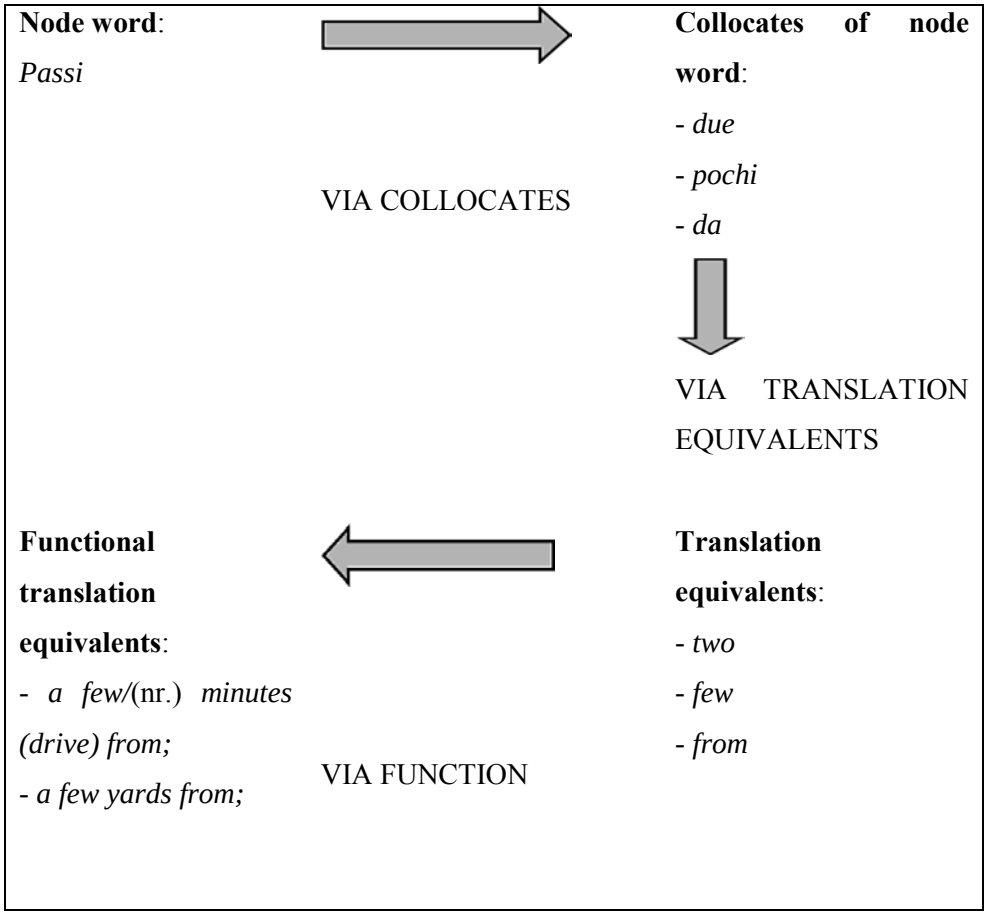
As expected, *two*, in its function as quantifier, has a varied collocational profile: it co-occurs with *bedrooms, adults, bathrooms, miles, people, single beds, twin bedded rooms*. The only item referring to distance is *miles* which is not, however, the equivalent we are looking for, in that *a pochi passi da* describes a very short distance.

The item *few* collocates with *days, miles, minutes* and *yards*. Both *minutes* and *yards* are used to describe a very short distance and are embedded in a unit of meaning very similar to the one containing *passi*. Examples are the sentences: *only a few minutes drive from the medieval town of Pembroke, …the county ground is a few minutes away, … a peaceful retreat just a few yards from your holiday home*.

The preposition *from* has, obviously, a varied range of collocations. However, when it refers to distance it ma inly collocates with *miles* and *minutes*.

In the light of the results obtained, we may take *minutes* as translation equivalent of *passi* (Step 3). In fact, its function is that of describing the

short distance existing usually between a location and a place, such as a city, a town or a nature reserve. To sum up:



Our next example will start from an Italian node word and the methodology will be applied in order to identify its English functional translation equivalent.

The node word chosen for analysis is the Italian adjective *familiare* as used in the language of tourism (see above for details on the corpora used). The English equivalents of this word are *family, domestic,* and *household* but the bilingual dictionary *Il Ragazzini* 2005 also provides in the examples the equivalents *friendly, informal, easy, well-known*, and *familiar*. A look at the collocational profile of *familiare* will help us to identify the different units of meaning in which it is used and the English equivalents.

The most frequent collocates of the Italian *familiare* are: *conduzione, atmosfera, ambiente, accoglienza, ospitalità.* Some examples are reported

below:

```
a Vigna: azienda agrituristica a conduzione familiare immersa nel
rfettamente curate da un attenta conduzione familiare,sono parte
 immerso nel verde,l'atmosfera è semplice e familiare, le camere e
etto per la pesca amatoriale e un'atmosfera familiare fatta di cose
a Il casolare offre un'inconsueta atmosfera familiare, sia per la
correre dei momenti di relax in un ambiente familiare.
a che circonda l'Agriturismo.  L'ambiente è familiare e discreto, ma
ort ed elegantemente arredati. L'ambiente è familiare e discreto e
 tranquilla e offre una cordiale ospitalità familiare. L'azienda è
lie in un'atmosfera di autentica ospitalità familiare. La tradizione
acque ioniche.L'ospite trova un'accoglienza familiare e riservata in
ico e Perna Lanera offrono una   cordiale e familiare accoglienza,
```

At this point, we should identify functional translation equivalents for the following units: *azienda agricola a conduzione familiare*, *atmosfera familiare*, *ambiente familiare*, *ospitalità familiare*, and *accoglienza familiare*. First, we need to check if there is a w ord-for-word correspondence, that is to say if the corresponding English units contain *family* or *familiar*. In case of no direct equivalence, we will proceed according to the methodology proposed.

The word *family*, in the English set of the comparable corpus used, frequently co-occurs with the following words: *run, rooms, home, bedroom, ideal for*. Only one match with the Italian units can be identified and refers to *azienda agricola a conduzione familiare* which, as suggested by the presence of *run,* can be translated as *family-run farm*. The possible equivalent *family-owned farm* suggested by bilingual dictionaries is not used in this corpus. There are no matches for the other collocations of the Italian word *familiar* both in the concordance of *family* and in the concordance of *familiar* which has no entries in the corpus used. For this reason, the analysis proceeds with Step 2 of the methodology.

The English equivalent of the collocates of *familiare* are: *atmosphere, environment/ambience, hospitality, welcome*. The collocational profiles of these equivalents need to be analysed in order to find a functionally similar units of meaning of the Italian collocations.

The word *atmosphere* collocates with *friendly* and *informal*. These collocates do not exactly refer to the family but at the functional level they refer to the in formality of the place. This is why the units *friendly atmosphere* and *informal atmosphere* can be taken as functional equivalents of *atmosfera familiare* (Step 3).

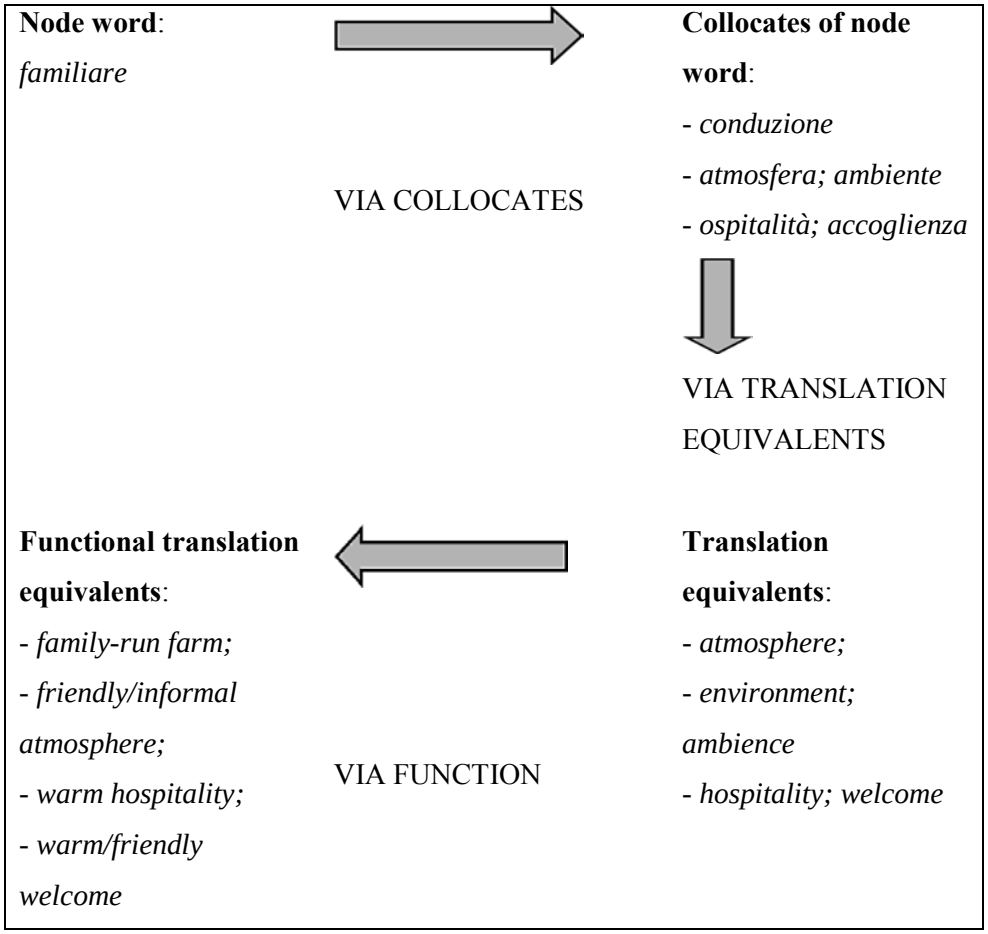
Let us proceed with *environment*. This word is used rarely in the corpus (only 4 times) and does not show any frequent co-occurrence. The same can be said for *ambience* occurring only 3 times. Both the concordances of *friendly* and *informal* do not suggest possible equivalents of *ambiente*. However, in the Italian texts and in this collocation, the function of *ambiente* is very similar to the one engaged in by *atmosfera*. It refers to the place where accommodation is offered and for this reason it may be translated with *friendly* or *informal atmosphere* with no change in meaning or function.

The other two equivalents of the Italian collocates are *hospitality* and *welcome*.

The item *hospitality* is not used frequently. However, a recurring adjective for this word is *warm* and in three instances it is also associated with the noun *warmth*. The concept of warmth is strictly linked to that of familiarity. A warm hospitality will make visitors feel at home. For this reason, the adjective *warm* can be taken as functionally equivalent to the Italian *familiare.*

As expected, the concordance of *welcome* does not show unexpected results in terms of collocations. The adjectives it frequently collocates with are *warm* and *friendly*, which, for the reasons described above, are functional equivalents of *familiare.*

The table provided below will help us to summarize the procedure followed and the results obtained.



4.3 Conclusion

The examples described in these sections have shown the importance of considering units of meaning as a starting point in the translation process. In the majority of cases, a word-for-word translation leads to mistakes and misunderstandings and does not consider the phraseological tendency of language. Starting from units of meaning means considering all the elements that a unit needs to function properly and to carry meaning. As shown above, the words *reunification, passi* and *familiare* may have different meanings depending on the contexts in which they are used and on the co-texts they are embedded in. As Tognini Bonelli (2001:150) says, if we need to establish functional equivalence across languages and cultures "it is not the word but the contextual patterning associated with it that identifies a function".

The process of translation is a complex process involving different levels of interpretation of meaning. The units of translation identified need also to be interpreted in terms of text type and its function, of register and according to the two cultures involved.

This methodology is a starting point for those students who are being trained to be translators and are taking their first steps in the difficult task of making a bridge between different languages and cultures.

In the domain of contrastive studies, comparable corpora prove to be valid tools both for the linguist and the translator in that they give insights into the patterns of a given language and allow the researcher to identify systematically the features of languages, that is to say typical and recurrent words, phrases, and expressions.

5 Textual colligation and thematic progression in English

In the previous chapters, attention has been paid particularly to lexis and to how meaning arises from word combinations. In this chapter, lexical choice will be considered from the perspective of thematic choice in the English language. According to Hoey (2003:171), "lexical choice has a major effect on features such as cohesion, Theme choice and paragraph division" and corpus analysis may also help the linguist to identify the nature of lexical choices.

As repeatedly described in the previous chapters, items should not be considered in isolation and should not be used as a slot and filler model or open choice principle would suggest (see chapter 3). Items are not arbitrarily combined but attract each other. As already seen, these attractions may be of different nature: lexical, grammatical, semantic or pragmatic. All these attractions considered together help identify units of meaning which should be interpreted as the basic units of language. Hoey (2003:174) identifies a fifth type of attraction which he defines *textual colligation*. What he firmly believes in is that lexical items are *primed* for use in textual organization, that is to say every lexical item is expected to be used in a certain way in the organization of the structure of texts. This attraction constrains us, as writers or speakers, to use words just as the other types of attractions identified by Sinclair do (see chapter 3).

According to Hoey (1985) whenever we encounter a word we note subconsciously the words it occurs with (its collocations), the meanings with which it is associated (its semantic associations), the pragmatics it is associated with (its pragmatic associations), the grammatical patterns it is associated with (its colligations), whether it is typically cohesive (its textual collocations), whether the

word is associated with a particular textual relation (its textual semantic associations), if it likes to begin sentences or paragraphs (its textual colligations), the genre, style, or social situation it is used in.

Most of the associations Hoey talks about have already been discussed and explained in the preceding chapters. In this chapter, the relation between lexical items, their function and their position in a clause will be analysed.

However, first some terminology should be made clear. The terms *lexical chains* and *cohesive chains* will be frequently encountered. A lexical chain is a sequence of related words in a text. Lexical chains are used to create cohesion in a text, that is to say to create a text which is well-organized, semantically meaningful and works as a coherent unit. Cohesive devices will be described in the following sections. However, we will provide here a couple of examples in order to make the definition of lexical chain clear. A cohesive device may be the repetition of the same item, of one of its synonyms, the use of pronouns or of co-referential items. A text dealing with Barack Obama is likely to use as cohesive devices the pronoun *he*, and the co-referential items *The President* and *The White House*. All these items form a lexical chain which gives cohesion to the text.

Texts are (or should be) cohesive and all the parts of the text have to be organized in order to have a meaningful unit as a result. Text should also be linearly developed (see Hoey 2003:172). This means that each sentence of a text should be meaningfully linked both to the sentences that follow and precede. This aspect of textual organization is represented by the Theme-Rheme structure. In order to describe this structure, the following sections will heavily draw on Halliday's works (1985a and b) who represents the Systemic-Functional tradition.

5.1 Theme and Rheme in the Systemic-Functional tradition

The clause has to be interpreted as a message. The structure which gives the clause its character as a message is defined by Halliday (1985b:38) **Thematic structure**. This type of structure is composed by two parts: the Theme and the Rheme. These two parts combined together constitute the message. Let us consider the following clause:

The hurricane Sandy strengthens into a strong category two hurricane.

It is concerned with *the hurricane Sandy*, which is the Theme of the clause. In the second part of the clause, details are added on hurricane Sandy: *strengthens into a strong category two hurricane*. This part in which the Theme is developed is called Rheme. According to Halliday (1985b:39) the Theme can be identified as that element which comes in first position in the clause. It is the starting point for the message and describes what the clause is going to be about. For this reason, part of the meaning of the clause depends on which element is chosen as its Theme.

The elements that can be selected as Theme in English are: subject, predicator, object, complement, and adjunct. The subject is the person, place, or thing that does what the verb describes. The predicator is the verbal element of the clause or of the sentence. The object is a noun, pronoun, or noun phrase representing 1) the person or thing (direct object) that something is done to, such as *book* in *We bought a book*, or 2) the person (the indirect object) who is concerned with the result of an action, such as *her* in *I gave her a book* or *I gave a book to her*, or 3) the person or thing that is joined by a preposition to another word or phrase, such as *bed* in *She was lying on the bed*. The complement is a word or phrase (esp. a noun or adjective) that follows a verb and describes the subject of a verb. The adjunct is an adverbial word or phrase that adds information to a sentence[19].

Further examples of Theme-Rheme structure are reported below (Halliday 1985b:40):

THEME	RHEME
The Queen of Hearts	she made some tarts
the man in the wilderness	said to me
for want of a nail	the shoe was lost

[19] Definitions adapted from Macmillan English Dictionary for advanced learners, 2002 and from the Longman Dictionary of Contemporary English 1990.

with sobs and tears	he sorted out those of the largest size
from house to house	I wend my way
Language – human speech -	is an inexhaustible abundance of manifold treasures
One hundred and fifty years ago, on 15 September 1830,	the world's first passenger railway … was opened …

5.1.1 Marked and unmarked themes

In declarative clauses, the Theme is usually represented by the subject. For example, in the clause *Material scientists are now actively borrowing nature's capacity for regeneration*, the theme is *material scientists* and corresponds to the subject of the clause. Putting a subject in thematic position means making an **unmarked** choice, that is to say a choice which is very likely to be made in declarative clauses. Conversely, a **marked Theme** in declarative clauses is a Theme which is not combined with the subject but with any other elements constituting the clause. Halliday (1985b:45) indicates the adjuncts as the most usual form of marked Theme in declarative clauses, such as for example *today, suddenly, at night, without much hope* whereas the most marked type of Theme in declarative clauses is a complement. He (*ibidem*) provides the example of *nature* in *nature I loved* and *this responsibility* in *this responsibility we accept wholly.*
The table reported below summarizes possible unmarked and marked choices in declarative clauses (Themes are reported in bold):

	Function	**Clause example**
UNMARKED THEME	Subject	**- engineers** have tackled the problem using a variety of strategies; - **A trans-Atlantic journey of just sixty minutes** has been promised since the dawn of supersonic flight; - **she** was given the prize.

MARKED THEME	Adjunct	- **In late September**, a secretive experimental vehicle roared into the clear blue skies; - **In the pages of popular books, magazines and newspaper comics**, the hyperfast world of airline travel was predicted to be just over the horizon;
	Complement	- **Precious** were his comments.

In interrogative clauses, verbs, auxiliaries, and Wh- elements are unmarked thematic choices. Their occurrence in first position are the regular pattern by which the interrogative is expressed. Marked themes are not usually used in interrogative clauses. However, examples of marked theme in this type of clauses are (Halliday 1985b: 48): *after tea will you tell me a story?* where the Theme is *after tea* and *in your house who does the cooking* where *in your house* represents the Theme.

Examples of unmarked and marked thematic choices in interrogative clauses are reported below (Themes are reported in bold):

	Function	**Clause example**
UNMARKED THEME	Finite verb	- **Is** this your bag? - **Can** you answer the phone, please? - **Do** you know him? - **Will** you be there? - **Should** I go?
	Wh- questions	- **Who** wrote this book? - **How** long has he been sleeping? - **How** far is the station? - **Which** house do you live in?

MARKED THEME	Adjunct	- **After dinner** will you call me? - **In the US** who will be elected?

In imperative clauses, the unmarked Theme is represented by *you, let's* and the verb. In negative imperatives, typical Theme is *don't*. Some examples are reported below (Themes are in bold):

	Function	**Clause example**
UNMARKED THEME	Finite verb (*do/don't; let's;* …)	- **Close** the door, please; - **Do** keep quiet; - **Let's** go out; - **Don't** look at me like that.
	Subject (you)	- **you** keep quiet.

As explained by Halliday (1985b: 49-52; see also Baker 1992:123), conjunctions and modal adjuncts (such as *and, or, when, even if, …;* and *probably, perhaps, usually, broadly speaking,* …) usually occur at the beginning of a clause in English. This means that these items can be considered inherently thematic and it is natural for the speaker to put them in initial position. These elements are, therefore, not considered part of the thematic structure.

5.1.2 Other types of marked Theme

The difference between marked and unmarked Themes in declarative, interrogative and imperative clauses has already been described above. In the examples provided, marked Themes were characterized by the presence of items in initial position which do not usually occupy that place by default. For this reason, the Themes combined with predicators, complements, and adjuncts are defined Fronted Themes, referring to the action of fronting, of placing in initial position. Furthermore, it has been noticed that there exist different degrees of markedness. For example, in declarative clauses the presence of an adjunct in

thematic position is marked because in the English language the initial position is occupied by the subject by default. However, thematizing place adjuncts is very common in tourist and narrative texts (see Enkvist 1987 a nd Baker 1992). Conversely, fronting a predicator is a very unusual choice in the English language. Apart from 1) Fronted Themes, there are other two types of marked structures in English: 2) Predicated Themes, and 3) Identifying Themes.

2) Predicated Themes: *it* + be + nominal/adverbial group

As already mentioned above, another type of marked Theme is the Predicated Theme (Halliday 1985b:59). This structural patterns is characterized by the pronoun *it* + the verb *be* followed by a nominal or adverbial group. Let us consider the following example:

It was John who called her yesterday night

The predicated Theme is often associated with contrast. The function of *it was* in the example above is, therefore, creating contrast. The meaning is "it was John and not someone else who called her yesterday night". However, it should be noticed that the Theme of such structures is not *it* but the element which occurs after the verb *be*. In the example above, the Theme is, therefore, represented by *John*. The personal pronoun *it* has the function of an empty subject and is used to allow a cer tain element to be placed near the beginning of a cl ause and be interpreted as a Theme (Baker 1992:135).

Another function associated with predicated Themes is linked to the interpretation of the clause as an information unit. The information unit is structured into two components: one part is the news and the other part is what is already known to the listener, the given (Halliday 1985b:59). The new usually comes at the end of the information unit and corresponds to the Rheme. Conversely, the already known, the given comes at the beginning and corresponds to the Theme. In the example above, the predicated Theme allows the speaker to present *John* as the new item even though it is in thematic position.

3) Identifying Theme

Identifying Themes differ from predicated Themes because instead of the pronoun *it* + be, a *wh*- structure is placed in initial position in the clause (Halliday 1985b: 41-43; Baker 1992:136). According to Halliday (*ibidem*), in this special thematic structure all the elements are organized into two constituents which are linked by a relationship of identity expressed by the verb *be*. Let us consider the following example (*ibidem*):

What the duke gave to my aunt was that teapot

In the clause above, what comes before the verb *be*, that is to say *what the duke gave to my aunt*, is an instance of nominalization, "whereby any element or group of elements takes on the functions of a nominal group in the clause" (*ibidem*). In this type of thematic structure, what comes before the verb *be* functions as Theme. Another example is the following:

What happened was that John called her yesterday night

The Theme of the clause reported above is *what happened* and the rest of the clause functions as Rheme.

Predicated and identifying Themes allow the speakers to structure the message in whatever way they want, that is to say overcoming restrictions on word order (see Baker 1992:136). Furthermore, they imply contrast and contribute to the meaning of exclusiveness. In the first example reported above, the meaning of exclusiveness refers to the teapot: the duke gave my aunt that teapot and nothing else.

We will briefly summarize the three types of marked Themes by considering the structure *Hurricane Sandy has killed dozens of people in the Caribbean* and by changing the position of its components. In some cases the order of the elements in the clause needs to be rearranged.

Let us start with fronted Theme. As said above, in declarative clauses a marked Theme is a T heme which is not combined with the subject but with any other

elements constituting the clause. The other elements are: subject, predicator, object, complement, and adjunct.
The clause *Hurricane Sandy has killed dozens of people in the Caribbean* is unmarked because the subject is in initial position and functions as Theme.
If the place adjunct is fronted, the structure would become marked, however not highly marked as explained above (see section 5.1.1):

Fronted place adjunct: *In the Caribbean hurricane Sandy has killed dozens of people*

The Theme is *in the Caribbean* and, as visible, the resulting structure sounds quite common.
When objects or complements are fronted the resulting structure is marked and less used. The two examples are reported below:

Fronted object: *Dozens of people hurricane Sandy has killed in the Caribbean*;
Fronted adjunct: *Killed dozens of people were*.

Dozens of people and *killed* are, respectively, the fronted object and the fronted adjunct and are placed in Theme position.
As already said, fronting a predicator is a highly marked and very unusual choice. Let us consider the following example:

Fronted predicator: *Hurricane Sandy threatened to kill dozens of people in the Caribbean, and kill them it did.*

In the above example, *kill* functions as Theme. Examples of fronted predicator in English are very rare.

Predicated Themes involve the use of the pronoun *it* + the verb *to be*. Here follow the examples:
1) *It was hurricane Sandy which killed dozens of people in the Caribbean*;

2) *It was in the Caribbean that hurricane Sandy killed dozens of people*.

As already explained above, the Themes are what comes after the *it*-structure. *Hurricane Sandy* and *in the Caribbean* are, therefore, selected as marked thematic choices.

Identifying Themes imply the use of *what* combined with the verb *to be*. Here follows the example:

What hurricane Sandy killed in the Caribbean were dozens of people

The Theme is *what hurricane Sandy killed in the Caribbean* and the Rheme, the core of the message is *dozens of people*.

The following section will describe the interpretation of the clause as an information unit, that is to say as a unit composed of what is already known to the listener or the reader and what is new.

5.2 The clause as information unit: Given and New

The clause is the nearest grammatical unit which corresponds to the information unit. According to Halliday (1985b:274):

> The information unit is what its name implies: a unit of information. Information, as this term is being used here, is a process of interaction between what is already known or predictable and what is new or unpredictable. (…) It is the interplay of new and not new that generates information in the linguistic sense.

As anticipated, in section 5.1.2 the information unit is made up of two constituents, the Given and the New. The Given is that part of the information unit

which is already known to the hearer or to the reader; the New is the remaining part of the unit which presents the new message.

Given and Theme and New and Rheme are semantically related but not equal. The Theme is the point of departure of a speaker or of a writer while the Given represents the common ground between the speaker/writer and the listener/reader. Similarly, the Rheme is what the speaker/writer says about the Theme and New is what is not already known to the listener/reader. This explains why Theme and Rheme are speaker oriented while Given and New are listener oriented. What is new and what is given, therefore, depends on the common ground, on the shared knowledge existing between the interlocutors.

The following message can be segmented in three different ways depending on the amount of shared knowledge existing between the participants to the linguistic event. In other words, the interpretation of the message depends on the context of situation:

We are meeting John and Mary tomorrow afternoon

If we are talking about what we are doing the following day, the Given is only the pronoun *we* and the rest of the sentence *are meeting John and Mary tomorrow afternoon* is New. Conversely, if the focus of our message is on who we are meeting, the Given is *We are meeting* and the Rheme is *John and Mary tomorrow afternoon.* There is also a further possibility: we may want to inform our interlocutors only about when we are meeting John and Mary. In this case, the Given is *We are meeting John and Mary* and the New is *tomorrow afternoon.*

In the English language given information is usually placed before new information in order to make the text easier to be understood.

As exemplified above, Theme and Rheme and Given and New are semantically related. The information unit typically corresponds to a clause, even though for Halliday (1985b:285) the interpretation of the Given as something occurring in initial position (in Theme position) and of the New as the element placed at the end of the clause (in Rheme position) may limit the potential of these two

systems. However, the above mentioned organization of the clause represents a valid guideline for the organization of a text.

5.3 External relationships between clauses: the concept of cohesion

Both Theme and Rheme and Given and New represent internal resources for structuring the clause as a message. However, a proper organization of a text also needs non-structural resources which are referred to by the term cohesion. Cohesion can be created in English by four different ways: reference, ellipsis, conjunction and lexical organization (Halliday 1985b:288 and ff).

5.3.1 Reference

The term **reference** broadly refers to the relationship between a word and the object it refers to. In Hallidayan terms, reference is an element introduced at one place in the text which acts as a reference point for other elements following in the text. Pronouns are the most common referential elements used in English and in many other languages. Apart from personal pronouns, demonstratives are also used in texts to refer back to something that has already been said. Let us consider the examples below:

*The White House says that **the president** was updated through the night as Hurricane Sandy carved its way up the coast - signing two declarations of disaster.*
***He** really would be in trouble if people thought **he** was ignoring a major disaster to save his political career, travelling to swing states to campaign instead of staying in the White House*[20].

[20] From BBC News bbc.co.uk *Sandy steals spotlight from Romney* by Mark Mardell, 31 October 2012.

A relationship of identity is created between the elements *the president, He,* and *he*. Both *He* and *he* refer back to *the president* and form a chain of reference which contributes to the cohesion of the text.
The following example contains a demonstrative used to refer back to something previously mentioned:

The sun shines and this delights me

Here, *this* refers to the first stretch of the message and the relationship of reference is created between *the sun shines* and *this*. In writing, demonstratives typically refer back to the preceding text (anaphoric reference). However, in some cases they may also refer to something that follows (cataphoric reference). Examples of cataphoric *that* and *those* are provided below (Biber Conrad and Leech 2002:75):

The unit of heat was defined as ***that*** *quantity [which would raise the temperature of unit mass of water …];*
We apologise to ***those*** *readers [who did not receive the Guardian on Saturday].*

5.3.2 Ellipsis and substitution

Other forms of cohesion are ellipsis and substitution. They both set up a lexicogrammatical relationship which, as Halliday (1985b:296) explains is a relationship in the wording rather than directly in the meaning.
Ellipsis refers to the omission of elements which are recoverable from the linguistic context or the situation. The examples reported below are adapted from Biber, Conrad and Leech (2002:230; 348):

- *He squeezed her hand but* (omitted element: *he*) *met with no response*;
- *He and his mate both jumped out, he* (omitted element: *jumped put*) *to go to the women, his mate* (omitted element: *jumped put*) *to stop other traffic on the bridge*;
- *He fell asleep up there – I don't know how* (omitted element: *he fell asleep up there*).

Substitution occurs when an element is replaced by another element. Some examples provided by Halliday (1985b:297ff) are reported below:

- *I've lost my voice*.
- *Get a new one*.
(*One* replaces and acts as a substitute of *voice*)

- *If you've seen them so often, of course you know what they're like.*
- *I believe so*.
(*So* is a substitute of *I believe I know what they are like*)

- *Does it hurt?*
- *Not any more. It was doing last night.*
(*Doing* replaces the verb *hurt*)

5.3.3 Conjunctions

Conjunctions are type of function words that connect clauses and sometimes, phrases or words. Baker (1992:191) summarizes the main relations expressed by conjunctions in the table reported below:

additive	*and, or, also, in addition, furthermore, besides, similarly, likewise, by contrast, for instance;*
adversative	*but, yet, however, instead, on the other hand, nevertheless, at any rate, as a matter of fact;*
casual	*so, consequently, it follows, for, because, under the circumstances, for this reason;*
temporal	*then, next, after that, on another occasion, in conclusion, an hour later, finally, at last;*
continuatives (miscellaneous)	*now, of course, well, anyway, surely, after all.*

Baker (1992:192) notices that languages vary in the type of conjunctions they prefer and in the frequency of usage of conjunctions. What students should bear in mind is that the English language has a tendency to present information in small chunks and to signal the relations between chunks in unambiguous ways. Furthermore, the frequency of use of conjunctions varies according to text types.

5.3.4 Lexical cohesion

According to Halliday (1985b:310), **lexical cohesion** occurs when items that are related in some way to those which have been previously mentioned are selected. These relationships between items may be created through repetition, synonymy, and collocation.

Repetition refers to the repetition of a lexical item. For example:

John tried to open the door. The door was locked.

The repetition of the word *door* is a form of lexical cohesion.

Repetition also refers to words which do not have the same morphological shape but are forms of the same item. Halliday (*ibidem*) provides the examples of *dine, dining, diner, dinner* and of *rational* and *rationalize.*

Synonymy is another form of lexical cohesion and involves the choice of a lexical item which is synonymous to the item previously mentioned. Examples may be *noise* and *crash* or *lightning* and *flash*.

Superordinates (general words referring to a class) and **hyponyms** (specific words referring to a class) are also included in this form of lexical cohesion. Examples are *bird* and *robin*, *car* and *sedan*, *child* and *girl*, *snake* and *phyton*, and so on.

A special case of synonymy is represented by **antonymy** which refers to lexical items which are opposite in meaning. Examples are: *warm* and *cold*, *war* and *peace*, *love* and *hate*, and so on.

We have already talked about collocation in chapter 3. Halliday (1985b:312) considers **collocation** as a form of lexical cohesion where the relationship between items is represented by their tendency to co-occur. He provides the following example (*ibidem*):

A little fat man of Bombay
Was smoking one very hot day.
But a bird called a snipe
Flew away with his pipe,
Which vexed the fat man of Bombay.

The items *smoke* and *pipe* are very strong collocates and this relation makes the occurrence of *pipe* cohesive.

5.4 Thematic progression in English

A message is textured when it has certain kinds of meaning relations. These meaning relations form the basis of cohesion between the messages of a text (Halliday and Hasan 1985:73).
In the creation of texts, the following features should be considered (Halliday 1985b:315ff):
1) Theme and focus; 2) lexical cohesion and reference; 3) ellipsis and substitution; 4) conjunctions.
We will discuss here only the first of these features, the choice of Theme.
The choice of Theme clause by clause is of utmost importance because the alternation of Theme and Rheme contributes to the development of the text. The organization of the clause in terms of Theme and Rheme is defined 'Functional sentence perspective' or FSP approach (Firbas 1964; Danes 1974). Thematic progression differs from language to language and from one text type to another.
In English there are three main thematic progressions:

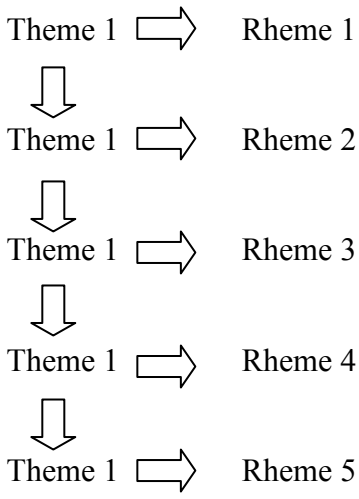
a) the main topical Theme may be repeatedly used for a certain stretch of discourse;

Let us consider the following example taken from *In Pursuit of the Proper Sinner* by E. George (1999):

David *thanked her.* ***He*** *stood in the dining room where the windows stretched from floor to ceiling and reflected all three of them in the glass.* ***He*** *admired the epergne that spilled white roses onto plaits of ivy.* ***He*** *fingered one of the thin silver forks.* ***He*** *used his thumbnail against a drip of candle wax. And* ***he*** *knew* ***he*** *wouldn't be able to force a morsel of food past the constriction in his throat.*

In *David thanked her, David* is the Theme and *thanked her* is the Rheme. The following clauses have all the same Theme and cohesion is created by the relationship of reference between *David* and the personal pronoun *he*.

The structure of the text follows the progression reported below:



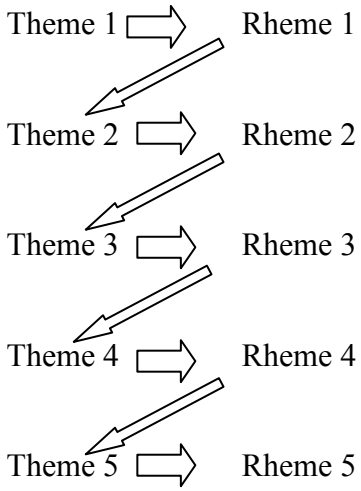
b) the Theme of one clause is selected from within the Rheme of the preceding clause;

The following example is adapted from Alexander Bain's biography[21]:

[21] available at http://inventors.about.com/od/bstartinventors/a/fax_machine.htm.

*The first fax machine was invented by Scottish mechanic and inventor **Alexander Bain** in 1843. **He** received a British patent for "improvements in producing and regulating electric currents and improvements in timepieces and in electric printing and signal telegraphs", in laymen's terms a fax machine.*

The Rheme of the first sentence contains the element Alexander Bain which is used as Theme of the following sentence.
The structure of this thematic progression is reported below:



c) a series of Themes can be developed from within a single Rheme (split Rheme).
An example of progression by means of a split Rheme is reported below[22]:

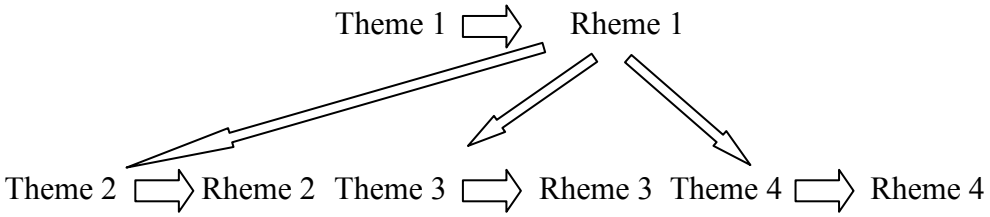
*Sedimentary rocks can be divided into **clastic rocks, chemical rocks, and organic rocks**.*
***Clastic sedimentary rocks** are accumulations of clasts: little pieces of broken up rock which have piled up and been "lithified" by compaction and cementation.*

[22] adapted from *Ask GeoMan What are the 3 basic types of rock?* available at http://jersey.uoregon.edu/~mstrick/AskGeoMan/geoQuerry13.html

Chemical rocks *form when standing water evaporates, leaving dissolved minerals behind ….*

Organic rocks *are any accumulation of sedimentary debris caused by organic processes ….*

In the above example the Rheme of the first sentence is split into three Themes. The typical structure of the split Rheme is the following:



d) Themes may derive from a hypertheme. Let us consider the example below[23]:

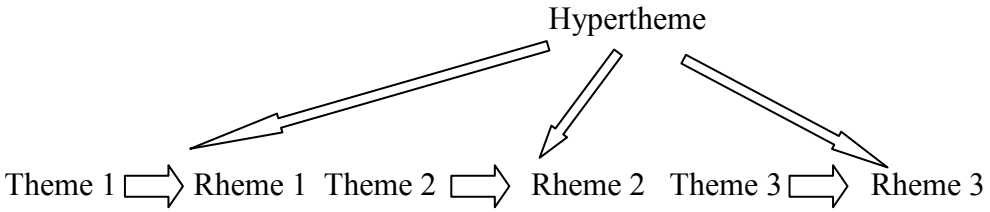
Italy is located in southern Europe and comprises the Italian Peninsula and some islands such as Sicily and Sardinia. Almost 40% of the territory is mountainous, with the Alps as the northern boundary and the Apennine Mountains forming the backbone of the peninsula.

Italian is the official language spoken by the majority of the population. The climate of Italy is highly diverse, and could be far from the stereotypical Mediterranean climate. Winters are cold and damp in the North, and milder in the South. Conditions on peninsular coastal areas can be very different from the interior's higher ground and valleys, particularly during the winter months when the higher altitudes tend to be cold, wet, and often snowy. The Alps have a mountain climate, with cool summers and very cold winters.

In the above text, *Italy* is the hypertheme and subsequent themes (*Almost 40% of the territory, Italian, The climate,* …) are derived from it.

The structure of this thematic progression is exemplified below:

[23] adapted from http://en.wikipedia.org/wiki/Geography_of_Italy and http://wikitravel.org/en/Italy



At this point, the analysis of a whole text will be carried out. The text presented below has been chosen for analysis because it contains many of the thematic progressions discussed above[24]:

At the scene *by Damian Grammaticas*

We headed towards where the tsunami hit land, close to the little village of Higashiro.
We had to pick our way through a sea of mud. What should have been a road was covered in broken branches, a squashed tractor and lots of electricity cables that had been brought down.
The destruction goes on and on.
The seashore was in the distance behind a row of trees. Here the waves toppled houses; they lie at crazy angles. Trees have been smashed into the buildings. A motorcycle lies twisted and bent.
Inside the houses, the furniture has been turned to matchsticks, possessions tossed everywhere, and on a few walls are portraits with the faces of those who once lived here, now stained by the waters which filled everything.

Analysis of thematic progression (Themes are reported in **bold** and Rhemes in *italics*) :

[24] from BBC News bbc.co.uk *At the scene* by Damian Grammaticas, 14th March 2011.

We *headed towards where the tsunami hit land, close to the little village of Higashiro.*

Theme 1 (**We**) ⟶ Rheme 1 (*headed towards where ….*)

↓

We *had to pick our way through a sea of mud.*

Theme 1 (**We**) ⟶ Rheme 2 (*had to pick our way through a sea of mud*)

In the first two sentences the Theme *we* is repeated and Rheme 1 and Rheme 2 are combined with the same Theme 1.

From within Rheme 2 (*had to pick our way through a sea of mud*), the author chooses the Theme of the following sentences:

What should have been a road *was covered in broken branches, a squashed tractor and lots of electricity cables that had been brought down.*

The destruction *goes on and on.*

Theme 2 (**what should have been a road**) ⟶ Rheme 3 (*was covered in broken …*)

Theme 3 (**that**) ⟶ Rheme 4 (*had been brought down*)

Theme 4 (**The destruction**) ⟶ Rheme 5 (*goes on and on*)

The description in Rheme 3 (*was covered in broken branches, a squashed tractor and lots of electricity cables*) is semantically similar to Theme 4 (*the destruction)* and, for this reason, they are cohesive.

Most of the Themes in the remaining part of the text are chosen from within Theme 4 which acts as a hypertheme. Theme 5, 6, 8, 9, and 10 are all descriptions of *the destruction*.

The seashore *was in the distance behind a row of trees*.

Theme 5 (**The seashore**) ⟶ R heme 6 (*was in the distance behind a row of trees*)

Here *the waves toppled houses*; **they** *lie at crazy angles*.

Theme 6 (**Here**) ⟶ Rheme 7 (*the waves toppled houses*)

Theme 7 (**they**) ⟶ Rheme 8 (*lie at crazy angles*)

Trees *have been smashed into the buildings*.

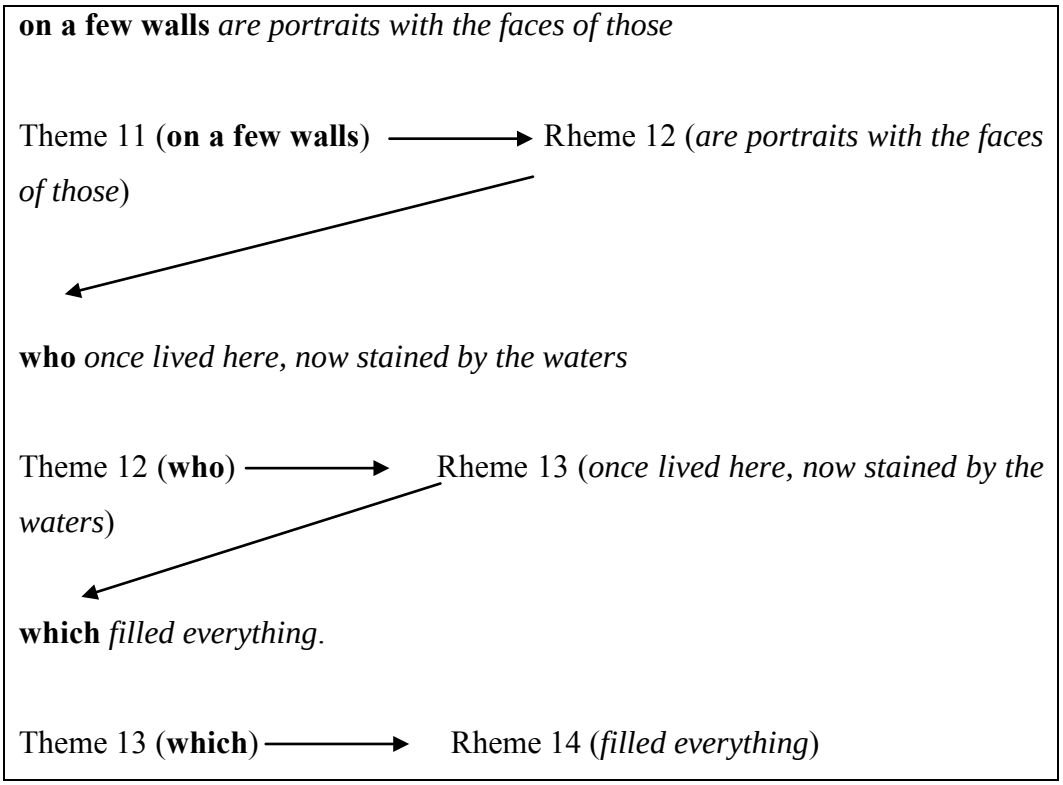
Theme 8 (**Trees**) ⟶ Rheme 9 (*have been smashed into the buildings*)

A motorcycle *lies twisted and bent*.

Theme 9 (**A motorcycle**) ⟶ Rheme 10 (*lies twisted and bent*)

Inside the houses, *the furniture has been turned to matchsticks, possessions tossed everywhere, and*

Theme 10 (**Inside the houses**) ⟶ Rheme 11 *(the furniture has been …)*



The text is highly cohesive because it has all the meaning relations which are at the basis of a well-structured text.

This chapter has briefly summarized the features of cohesion and has shown the importance of Theme-Rheme progression in the creation of texts.
It needs to be said that these theories are valid for the English language but many of them do not prove to be as valid for the Italian language, particularly because Italian is not characterized by the same rigid word order as English.
For example, fronted predicators are not a marked choice in Italian. Sentences such as *Rispondo io al telefono* or *Vado al cinema con Francesca* sound quite common in the Italian language even though they have predicators in Theme position.
Furthermore, the use of conjunctions is different in the two languages; as mentioned above, texts in English are characterized by small chunks of language. Conversely, the Italian language uses longer chunks which need to be connected by means of more conjunctions and subordination.

These differences are fundamental in the process of translation. Students trained to become translators should be acquainted with the features of the two languages which should be studied contrastively.

Meaning arises from the combination of words. These combinations constrain the way clauses are formed. However, in order for a text to be meaningful and work as a unit, structural and cohesive devices should be used and clauses should be organized according to the meaningful relations occurring between them.

References

Aarts, J., 1991. "Intuition-Based and Observation-Based Grammars". In Aijmer, K. and B. Altenberg (eds), English Corpus Linguistics. London: Longman.

Baroni M. and Bernardini S., 2006. "A new approach to the study of translationese: Machine-learning the difference between original and translated text". Literary and Linguistic Computing 21(3). 259-274.

Baker M., 1992. *In Other Words: A coursebook on translation.* London: Routledge.

Berber-Sardinha, T. 2004. *Linguistica de Corpus*. Barueri SP, Brazil: Manole.

Biber D. Conrad S. and R. Reppen, 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge Univ. Press

Biber D, Conrad S. and Leech G., 2002. *Longman Student Grammar of Spoken and Written English*. London: Longman.

Bowker L. and Pearson J., 2002. *Working with Specialized language: A practical guide to using corpora*. London and New York: Routledge.

Carter R. and McCarthy M., 1998. *Vocabulary and Language Teaching.* London, Longman.

Clear J., 1992. "Corpus sampling". In Leitner G. (ed.) *New directions in English language corpora*. Berlin: Mouton de Gruyter, pp. 21-31.

Danes F., 1974. "Functional sentence perspective and the organization of the text". In Danes F. (ed.) *Papers on Functional Sentence Perspective*. Prague: Academia /The Hague: Mouton, pp.106-128.

EAGLES, 1996. *Preliminary recommendations on corpus typology*. EAG-TCWG-CTYP/P. Pisa: Consiglio Nazionale delle Ricerche. Istituto di Linguistica Computazionale.; available on-line at http://www.ilc.cnr.it/EAGLES/corpintr/node13.html.

Enkvist N., 1987. "Text strategies: single, dual, multiple". In Steele R. and Threadgold T. (Eds.) *Language Topics: Essays in honour of Michael Halliday*. Volume II. Amsterdam: Benjamins, pp. 203-211.

Firbas J., 1976. “A study in the functional sentence perspective of the English and Slavonic interrogative sentence”. *Brno Studies in English* 12: pp.9-57.

Firbas J., 1986. “On the dynamics of written communication in the light of the theory of Functional Sentence Perspective”. In Cooper, C. and Greenbaum S. (eds.) *Studying Writing: Linguistic Approaches*. Beverley Hills: Sage, pp. 40-71.

Firth J.R., 1957a. *Papers in Linguistics*. London, Oxford University Press.

Firth J.R., 1957b. “The Technique of Semantics”, in *Transactions of the Philological Society*, 1935, reprinted in Firth J.R. (ed.), *Papers in Linguistics 1934-51.* London, Oxford University Press.

Firth J.R., 1957c. “Modes of Meaning”, 1951, reprinted in Firth J.R. (ed.), *Papers in Linguistics 1934-51.* London, Oxford University Press.

Firth J. R., 1957d. “Personality and Language in Society” in J. R. Firth, *Papers in Linguistics* 1934-1951. London, Oxford University Press.

Francis N. W., 1992. “Language corpora B.C.”. In Svartvik J. (ed.), *Directions in Corpus Linguistics.* Berlin: Mouton de Gruyter, pp.17-32.

Halliday M.A.K. and Hasan R., 1976. *Cohesion in English.* London, Longman.

Halliday M.A.K. and Hasan R., 1985a. *Language, Context and Text: a Social Semiotic Perspective*. Deakin University Press.

Halliday M.A.K., 1985b. *An Introduction to Functional Grammar.* London, Arnold.

Hoey M., 1997. "From concordance to text structure: new uses for computer corpora" in Melia, J. and Lewandoska, B. (Eds.) *PALC 97: practical applications in language corpora,* Lodz, Poland: Lodz University Press.

Hoey M., 2003. ‘Textual colligation – a special kind of lexical priming” In Aijmer K. and Altenberg B. (Eds) *Advances in Corpus Linguistics. Proceedings of ICAME 2002, Göteborg.* Amsterdam: Rodopi, pp.171-194.

Hoey M., 2005. *Lexical Priming. A new theory of words and language*. London: Routledge.

Johansson S., 1991. “Times change, and so do corpora”. In Aijmer K. and Altenberg B. (Eds.) *English corpus linguistics*. New York: Longman, pp. 305-314.

Johansson S., 1995. "The approach of the Text Encoding Initiative to the encoding of spoken discourse". In Leech G., Myers G. and J. Thomas (Eds) *Spoken English on Computer*. Harlow: Longman, pp. 82-98.

Langendoen D.T., 1968. *The London School of Linguistics: A study of the linguistic contributions of B. Malinowski and J. R. Firth*. Cambridge, Mass: MIT Press.

Louw B., 1993. "Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies". In Baker M., Francis G. and E. Tognini-Bonelli (Eds) *Text and Technology: In Honour of John Sinclair*. Amsterdam & Philadelphia: Benjamins, pp. 157-176.

Macmillan English Dictionary for Advanced Learners, 2002. Oxford, Macmillan.

Malinowski B., 1922. *Argonauts of the Western Pacific: An Account of Native Enterprise and Adventure in the Archipelagoes of Melanesian New Guinea*. New York: E. P. Dutton & Co.

Malinowski B., 1935. *Coral Garden and Their Magic*. London, Allen & Unwin.

Malinowski B., 1994. "The Problem of Meaning in Primitive Language" (1923). In J. Maybin (ed.), *Language and Literacy in Social Practice: A Reader*. Avon: The Open University Press, pp. 1-10.

Milizia D. and Spinzi C., 2008. "The terroridiom principle between spoken and written discourse". In Roemer U. and Schulze R. (Eds) *Patterns, meaningful units and specialized discourses*. Special issue of IJCL, vol. 13 (3), pp. 322-350.

Milizia D., 2012. *Phraseology in political discourse*. Milano: LED.

Palmer F.R., 1968. *Selected Papers of J.R. Firth 1952-59.* London/Harlow: Longman Linguistics Library.

Phillips M., 1989. "Lexical Structure of Text". In *Discourse Analysis Monographs*, 12 Birmingham: University of Birmingham.

Sinclair J., 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.

Sinclair J. M., 1992. "The Automatic Analysis of Corpora". In Svartvik J. (ed.), *Directions in Corpus Linguistics.* Berlin: Mouton de Gruyter.

Sinclair, J. 1996. **“** The Search for Units of Meaning”. TEXTUS IX (1). Genova: Tilgher, pp.71-106.

Sinclair J.M., Payne J. and C. Perez Hernandez (eds.), 1996. *Corpus to Corpus: a Study of Translation equivalence. IJCL* 9.3.

Sinclair J., 1997. “Corpus Evidence in Language Description”. In Wichmann, A. et al (Eds.), *Teaching and Language Corpora*. London and new York, Longman, pp. 27-39.

Sinclair J., 2001. “Preface”. In Ghadessy M., Henry A. and Roseberry R. (Eds.) *Small corpus studies and ELT*. Amsterdam/Philadelphia: John Benjamins, pp. vii-xv.

Sinclair J., 2005. “Corpus and Text - Basic Principles” in Wynne M. (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books, pp. 1-16.

Scott. M., 2009. “In Search of a Bad Reference Corpus” in Dawn Archer (ed.) *What’s in a Word-list? Investigating word frequency and keyword extraction*. Oxford: Ashgate. pp. 79-92.

Scott M. and Tribble, C., 2006. *Textual Patterns. Keywords and corpus analysis in language education.* Amsterdam: John Benjamins.

Stubbs M., 1996. *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford, Blackwell.

Stubbs M., 2002. “Two quantitative methods of studying phraseology in English”. International Journal of Corpus Linguistics 7:2, pp.215–244.

Tognini-Bonelli E., 2001. *Corpus linguistics at work.* Amsterdam: John Benjamins

Tognini Bonelli E. and Manca E., 2002. “Welcoming children, pets and guest. A problem of non-equivalence in the languages of Agriturismi and Farmhouse Holidays”. In Evangelisti P. and Ventola E. (Eds.), *English in Academic and Professional settings: techniques of Description/Pedagogical Applications. Textus XV* (2). Genova: Tilgher, pp.317-334

Teubert W., 1996. “Comparable or Parallel Corpora?” *International Journal of Lexicography*. 9(3): pp. 238–64

Teubert W., 2001. “Corpus Linguistics and Lexicography”. *International Journal of Corpus Linguistics.* 6: pp. 125–53.

Ulrych, M., 1992. *Translating Texts.* Rapallo, Cideb.

Zanichelli, 2005. Il Ragazzini 2005. Bologna: Zanichelli.

http://siba2.unisalento.it

eISBN 978-88-8305-092-3 (electronic version)

http://siba-ese.unisalento.it