

© copyright Edizioni Università di Trieste, Trieste 2008

Proprietà letteraria riservata.

I diritti di traduzione, memorizzazione elettronica, di riproduzione e di adattamento totale e parziale di questa pubblicazione, con qualsiasi mezzo (compresi i microfilm, le fotocopie e altro) sono riservati per tutti i paesi.

ISBN 978-88-8303-252-3

EUT - Edizioni Università di Trieste
p.zza Europa, 1 - 34127 Trieste
<http://eut.units.it>

Ecolingua
The Role of E-corpora
in Translation and
Language Learning

edited by
Christopher Taylor

Table of contents

- | | | | |
|----|------------------------------------------------------------------------------------------------------------------------------------------------------|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 7 | <i>Christopher Taylor</i>
Introduction | 113 | <i>Maria Grazia Busà XXX</i>
Teaching Prosody to Italian Learners of English: Working towards a New Approach |
| 11 | <i>Silvia Bruti</i>
<i>Elisa Perego</i>
Vocatives in Subtitles: a Survey across Genres | 127 | <i>Sara Gesuato</i>
Encoding of Information in Titles: Academic Practices across Four Genres in Linguistics |
| 52 | <i>Maria Freddi</i>
Continuity and Variation across Translations: Phraseology in the Pavia Corpus of Film Dialogue | 158 | <i>Vanessa Leonardi</i>
Increasing or Decreasing the Sense of "Otherness": the Role of Audiovisual Translation in the Process of Social Integration |
| 71 | <i>Francesca Bianchi</i>
The Distribution of Authorial Presence in Experimental Psychology Articles | 173 | <i>Amanda C. Murphy</i>
Mediated Language in Non-native Speaker Texts in the European Commission |
| 93 | <i>Francesca Bianchi</i>
<i>Elena Manca</i>
Discovering Language through Corpora: the Skills Learners Need and the Difficulties they Encounter | 185 | <i>Pierfranca Forchini</i>
The get-unit in Corpora of Spontaneous and Non-spontaneous Mediated Language: from Syntactic Versatility to Semantic and Pragmatic Similarity |

Introduction

CHRISTOPHER TAYLOR
University of Trieste

The current volume, as the final act of the PRIN research project *Ecolingua: the role of e-corpora in translation*, in language learning and in testing, is largely based on the presentations made during ECOLINGUA DAY, an event organised at the University of Trieste in order to hear papers illustrating the results of a number of the various research sub-projects brought to conclusion by the five university units involved (The Catholic University of Milan, the University of Padua, the University of Pavia 1 & 2, the University of Trieste).

The ECOLINGUA project gave the opportunity to many of those involved in the previous DIDACTAS project, also financed by the Italian Ministry of Education, to continue and refine their various research programmes and produce useful materials for research in linguistics, in translation practice and teaching, in language teaching and in testing language competence. The scientific background to the project was more or less the same for all the local units even though each concentrated on a number of specific tasks. This background included systemic linguistics, information technology, text analysis, multimodality, translation (particularly film translation), language and translation teaching/learning and testing, and the glue that held all these strands together was the 'corpus' and corpus-based research. This common denominator of the whole project has for a long time formed the basis for a great deal of research in the field of both theoretical and applied linguistics. Much of the linguistic research carried out has been based on systematic observation of authentic communicative events, namely

texts (oral, written and multimodal) which are representative of the cultural and situational contexts in which they are produced and exchanged.

Following in these footsteps, the various units looked to computers and corpora to carry out their ongoing research projects. Members of both Pavia units were concerned with multimodal text analysis and the translation of multimodal products, particularly with a didactic objective. The articles by Elisa Perego and Silvia Bruti and by Maria Freddi in this volume are testimony to this line of research. Francesca Bianchi and Elena Manca provide other corpus-based research articles dealing respectively with academic language and student reaction to corpora. The Padua unit, represented here by Maria Grazia Busà and Sara Gesuato produced a great deal of valuable work on linguistic and pedagogical subjects, all the result of painstaking corpus-based research. The Trieste unit was mainly involved in the analysis and translation of screen material, pursuing new angles such as identifying the phenomenon of predictability in film and television products, and in the teaching of film translation. Vanessa Leonardi discusses here the positive and negative effects that dubbing can have on social integration. Finally the Milan group's research was again totally corpus-based with Amanda Murphy and Pierfranca Forchini providing contributions for this volume.

The joint paper by Silvia Bruti and Elisa Perego (*Vocatives in Subtitles*) continues their research into the function of vocatives and their translation in interlingual subtitles. A small corpus of films of various genres forms the raw material for the project. Summarising the results of the research, the authors' survey points up differences in the type of vocatives chosen in each genre of film, but also shows that the translation of these terms does not always reflect those differences. The rigorous analyses of the nine video products revealed the use of eighteen types of vocative ranging from proper names to insults and their relative frequency of use within the various genres. A quick glance at one of the tables provided by the authors shows how such vocatives are at times not translated or are translated in different ways e.g. first names instead of last names, diminutives instead of kinship terms, etc.

Maria Freddi (*Continuity and Variation across Translations*) analyses a small corpus of British and American films dubbed into Italian in order to check for variation in translated products. The corpus methodologies adopted are principally those of Baker (the identification of translation specificities and tendencies) and Diaz-Cintas (audiovisual translation). Using text alignment technology Freddi discovers that while continuity patterns can be observed as translators find many common solutions, individual choices are not uncommon and can be attributed to a number of factors.

Francesca Bianchi (*The Distribution of Authorial Presence in Experimental Psychology Articles*) investigates to what extent scientific writing adheres to an impersonal style. Her corpus of tagged experimental psychology papers is used to show not only the level of overt authorial presence, largely revealed through

pronouns and determiners, but also how certain key words, analysed quantitatively, highlight various distribution patterns and indicate the most frequent collocates. The results show that all the articles (43 papers) contained some form of overt authorial presence but it was also revealed that the Results and Discussion sections proved to be the most personal, and the Abstract, Method and Conclusions sections the least. As the author claims, this study of a single discipline provides useful insights into the use of a little analysed rhetorical practice.

Francesca Bianchi's second contribution, in collaboration with Elena Manca (Discovering Language through Corpora), goes to the heart of the question of the use of corpora in a university environment. The objective of the experiment conducted by the authors was to assess the level of intrinsic difficulty encountered by students in performing corpus-based tasks. Two different groups of students of different levels of competence and experience, and their responses to a variety of such tasks, led to the creation of a General Difficulty List of Corpus Analysis Tasks. The authors are at pains to point out that they were measuring intrinsic difficulty and not external or environmental factors such as previously acquired skills and courses attended. The results, as the authors claim, should be very useful in the designing of corpus analysis tools for students.

Maria Grazia Busà's article 'Teaching Prosody to Italian Learners of English' concerns the implications for English language pedagogy of the expanding role of English in the world. She concentrates on the question of prosody and the shift towards a greater recognition of this aspect of language learning in terms of *mutual intelligibility*. Accordingly Busà discusses aspects of Italian pronunciation in English which may affect intelligibility. Reviewing some of the technological advances in the field and the instruments now available to teach prosodic features, the author points out that prosody is now more accessible to the non-expert with concrete benefits for the learner in terms of sounding natural and communicating successfully.

Sara Gesuato, in her article 'Encoding of information in titles', likens the titles of academic publications to a business card, and describes these 'mini-texts' as useful, logical and reader-friendly, though to varying degrees. In her research the author finds much variation as she analyses a hefty corpus of English language titles culled from four distinct yet connected academic genres - books, dissertations, journal articles and proceedings papers on the subject of linguistics. She shows how the titles differ along several dimensions such as length in words, richness in technical vocabulary and content, denotational precision, etc., and how these differences can be attributed to context of situation, communicative goals and target readerships. Gesuato's minute analysis shows that there are similarities across the genres though also differences, particularly in terms of expansion, pre-modification strategies, post-modification resources and so on.

Vanessa Leonardi (Increasing or Decreasing the Sense of Otherness) also uses a small corpus of Walt Disney films to analyse the impact and potential of

audiovisual translation in the process of social integration. Starting from the premise that language has always been associated with power, the author examines the strategies of foreignisation and localisation in terms of an increasing or decreasing sense of 'otherness'. In particular she analyses dubbing and how it can be used as an instrument of linguistic nationalism but also as a means of strengthening social integration. The Italian versions of films that in the original contained potentially racist elements (accentuated 'foreign' accents, stereotyped behaviour and language patterns, etc.) consciously or unconsciously erased these features thereby contributing to portraying diversity in a positive light.

Amanda Murphy (Mediated Language in non-native speaker texts in the European Commission) looks at examples of mediated language in non-native speaker texts in European Commission documents. She compares edited with non-edited texts in order to see whether editing can be considered a form of language mediation. These texts are subjected to further comparison with general reference material from the British National Corpus. Some preliminary conclusions show firstly that most editing concerns objective criteria such as grammar and house style, though personal subjective modifications are also detected, highlighting the seemingly conflicting strategies of concision and explicitation.

In her paper on the 'get-unit' Pierfranca Forchini (Milan) makes an exhaustive analysis of the use of the ubiquitous English verb 'get', making a new contribution to the many studies already conducted on this aspect of verbal grammar. The author shows, through rigorous exemplification based on the US spoken sub-corpus of the *Bank of English* and a corpus of films, that the verb 'get' can be depicted semantically and pragmatically as a result marker. She also makes a convincing case for 'get' displaying prevalent negative semantic prosody. Thus, in spite of its syntactic versatility, the use of 'get' can be seen to imply some sort of transformation, often of a negative nature, and across a range of registers and text types.

Vocatives in Subtitles: a Survey across Genres

SILVIA BRUTI (University of Pisa)
ELISA PEREGO (University of Trieste)¹

0. OVERVIEW

The purpose of this contribution is to investigate the function of vocatives and their translation in interlinguistic subtitles over different film genres. Our previous investigation (Bruti, Perego 2005) was based on a small corpus that included 2 British and 2 American films, belonging approximately to the same genre, i.e. COMEDY (with the exception of *The Talented Mr Ripley*, which begins as a COMEDY but turns into a DRAMATIC MYSTERY STORY). This project aims to investigate the various roles vocatives play in the construction of the narrative according to the different needs that different film genres aim to fulfil. The corpus has therefore been extended to include: a full-length animated feature from Walt Disney Pictures (*Bambi*, D. Hand, 1942), an action film (*Lethal Weapon 4*, R. Donner, 1998), an adaptation from a literary masterpiece (*Sense and Sensibility*, A. Lee, 1996), a popular comedy series (two episodes of *Sex and the City*, Season 4, “The Agony and the Ecstasy”, M.P. King, 2001 and “I heart NY”, M.P. King, 2002) and an animated series (two episodes of *The Simpsons*, “Homer in the night”, R. Moore, 1989-90; “Homer the Moe”, J. Kamerman, 2001-02).

1. VOCATIVES: FUNCTIONS AND POSITIONS IN DISCOURSE

Vocatives² can be distinguished on the basis of the different functions they perform in verbal exchanges. Zwicky (1974) assigns them two roles and names them

calls and *addresses*. The former are used to catch the addressee's attention, the latter to maintain or reinforce the contact between the speaker and the addressee. Quite interestingly, Zwicky recognises the "extraordinary idiomaticity" (1974: 788) of vocatives, given the very many combinations in which they can appear.

The distinction between the two functions is however fuzzy, as Zwicky notes that some vocatives used as calls cannot be used as addresses (e.g. *cabby*, used to refer to a taxi driver), including the item *you*. Davies (1986) disagrees and claims that *you* can instead be employed to refer to somebody who is already paying attention to what the speaker is saying. She therefore proposes a different categorisation of the functions of vocatives: an "identifying" function, that either selects an addressee, or – especially if the vocative is repeated – works as a reminder within a conversation, lending it a personal tone; and an "expressive" function, when it is not necessary to select an addressee but the speaker's attitude towards the addressee is specified. She then discusses the power of different vocatives, i.e. proper names alone or accompanied by titles vs. definite NPs with some modification, as identifying labels. As she shows, if a vocative is used to identify somebody, the information it contains needs to unambiguously specify who this person is; if, on the contrary, it is used mainly to express the speaker's attitude towards the addressee there are fewer constraints. A rather large group of nouns that can be used as vocatives (e.g. *boy, girl, man, lady, child, brother, sister*, but also terms of endearment like *honey, treasure, pet*, and derogatory items like *rascal, idiot, nigger, Communist, Nazi*) are reminiscent of nicknames and sometimes in fact do turn into real nicknames (Davies 1986: 97 mentions the case of *baby* > *Babe*).

Biber *et al.* (1999 and previously Leech 1986), on the basis of extensive corpus investigation, identify three different uses and relate them to discourse position: getting someone's attention, identifying the addressee, creating, maintaining or reinforcing the social bond between collocutors. The identifying function – a way of recognising or sanctioning the addressee – is the most typical feature of an act of address and can occur along with the other two. The attention-getting function is usually linked to an act that aims to address the interlocutor who has not been referred to before the use of the vocative. This aim is often intertwined with the pragmatic functions that are linked to perlocutionary aims (e.g. convince someone). In addition, pragmatic functions alone may be implied by the use of the vocative: one might for instance think of the stereotyped answering formulae in which vocatives are directly linked to politeness requirements but are not at all essential to identify speech participants. Vocatives that appear at the beginning of an utterance usually fulfil an attention-getting function, possibly combined with that of identifying the addressee; those that are placed at the end most often contribute to expressing socio-pragmatic meanings which can also accommodate an identification of the addressee.

Huddleston and Pullum (2000: 523, their *italic*) also remark that "vocative terms generally convey a considerable amount about the speaker's social relations or emotive attitude towards the addressee, and their primary or sole purpose is often to give expression to this kind of meaning, as in *Yes, sir!* or *I agree, my dear, that it's quite a bargain*". This is especially true for those languages that like English no longer have the distinction between T and V forms and have to resort

to the “parafeatures that congregate around the address pronouns” (Hickey 2003: 402), among which notably nominal address, in order to express more subtle sociolinguistic nuances of meaning.

On the basis of these distinctions some working hypotheses may be put forward. When vocatives serve the purposes of either selecting an addressee or catching someone’s attention, they can more easily be omitted in interlinguistic subtitles for two main reasons. Firstly, since subtitling is not a substitutive form of translation like dubbing but provides a graphic addition that complements the original soundtrack, many vocatives can be omitted because they are clearly perceptible through the auditive channel. The majority of them are proper nouns, kinship terms or titles, designators that can also be aurally recognised by a foreign audience. Secondly, then, in film texts the selection of the next speaker can be achieved through the visual channel, for example with some movements of the head, the eyes, the hands (i.e. pointing) or with posture. There can be cases in which images do not show the selection of the interlocutor, for example when the characters appear at a distance or are with their back to the camera. In such circumstances, if there are a number of compatible addressees, and it is therefore the verbal code that disambiguates among them, in order to avoid vagueness, the subtitle should include the vocative. This is especially the case when it is a descriptor such as a term of endearment or of abuse. When instead vocatives are embedded in the exchange as supporting elements that build up or reinforce a social relationship between co-speakers (Martiny 1996: 767), they are clearly loaded with a socio-pragmatic meaning. Interlinguistic subtitles should therefore try to convey this meaning potential or at least to compensate its loss by reproducing the illocutive value of the utterance.

2. VOCATIVES AND SUBTITLES

Subtitling is known to be a reduced form of audiovisual translation. As reductions are often inevitable, information that might be relevant is bound to be deleted. Apparently, subtitlers follow precise criteria in order not to let subtitles remain “frustratingly ‘incomplete’” (Brondeel 1994: 28). The selection of superfluous information, however, does not necessarily correspond to an unquestionable choice. At times, the total or partial deletion of some elements can spoil the detailed pragmatic picture that the film director has carefully drawn, although it can be argued that the expressive meaning conveyed by vocatives is not always necessarily lost in the subtitled version (Bruti, Perego 2005: 46). When subtitling, whatever pertains to the phatic and expressive function of language is omitted in order to favour the factual, narrative and referential function. As a consequence, markers that linguistically emphasise the interpersonal relationships between the characters (e.g. appellatives, vocatives and phatic expressions) are often expunged (Kovačič 1994: 250, 1996: 108; Becquemont 1996: 152-153; Assis Rosa 2001: 216) with the result that “subtitling may create a substantially different interpersonal dynamics from that intended” (Hatim, Mason 2000: 438)³. This has patent repercussions on the pragmatic dimension. The conciseness of subtitles may entail a loss in meaning as well as in the expression of emotion, and

often jeopardises the transmission of connotative and pragmatic meanings (see Hatim, Mason 2000 as far as politeness is concerned).

In our previous study, we observed that the appropriate or inappropriate use and transposition of terms of address could adequately reflect or distort the dynamics of interactions. The analysis of a first sample of English films with Italian subtitles revealed that, whenever possible, the function of vocatives and terms of address tends to be respected with reference to the overall texture of the film at stake, and that subtitles manage to capture some of the pragmatic force of the original version and make the dialogue sound as authentic and effortless as possible (cf. results for English subtitles of Polish soap operas in Szarkowska *forthc.*). Of course, technical constraints do not always allow the subtitler to be fully faithful to the original, even though a tendency exists not to remove vocatives artlessly. In order for us to see whether this is somehow linked to film genre⁴, we started working towards defining the genre of a wider sample of films⁵.

3. FILM GENRE AND FILM TRANSLATION

Genre is a ubiquitous phenomenon common to all instances of discourse (Neale 2000: 2), which justifies the interest in this topic at different levels. Film and genre have recently been analyzed by various authors and from different perspectives, as the recent wide-ranging literature in the field demonstrates (Altman 1999; Kozloff 2000; Neale 2000; Frezza 2001; Aimeri, Frasca 2002; Campari 2002; Eugeni, Farinotti 2002).

In linguistics, genre has been studied in relation to language in order to explore, for both theoretical and applied purposes, the range of predictable patterns and processes used to produce texts that reflect an intended purpose for an intended audience (Bhatia 1993; Swales 1990). Genre analysis, therefore, comprises the searching for the rationale behind specific linguistic choices.

In this paper, we aim to relate the issues of film genre and audiovisual translation, thus merging two distinct research areas and possibly understanding their mutual influences. This approach appears to be a productive means of reflecting upon decision-making in film translation. In particular, we set out to determine to what extent genre diversity has a bearing on the translation or non-translation of vocatives in interlinguistic subtitles. Indeed, the genre of a film appears to determine linguistic choices in the first place, and translational choices in the second place due to its vast influence on rhetorical and pragmatic strategies (Malvasi 2002: 283). However, the impact of film genre specifically on the translation of vocatives is not always easy to pinpoint. This difficulty arises from the more general and serious difficulty of defining film genre.

Tim Dirks (2006, <http://www.filmiste.org>) opens up his thorough description of film genres by providing the following definition:

Film genres are various forms or identifiable types, categories, classifications or groups of films that are recurring and have similar, familiar or instantly-recognizable patterns, syntax, filmic techniques or conventions - that include one or more of the following: settings (and props), content and subject matter, themes, period, plot, central narrative events, motifs, styles, structures, situations, recurring icons (e.g. six-guns

and ten-gallon hats in Westerns), stock characters (or characterizations), and stars⁶. Many films straddle several film genres.

He goes on by dividing film genres into three main categories⁷, each comprising a range of labels as shown in Table 1.

FILM GENRES		
NON-GENRE FILM CATEGORIES	MAIN FILM GENRES	FILM SUB-GENRES
Animated films British films Children/kids/family films Classic films Cult films Documentary films Serial films Sexual/erotic films Silent films	Action Adventure Comedy Crime/gangster Drama Epics/historical Horror Musicals Science fiction War Westerns	Biographical films (Biopics) 'Chick' flicks (or Gal films) Detective/mystery films Disaster films Fantasy films Film noire 'Guy' films Melodramas or women's "weepers" Road films Romance films Sports films Supernatural films Thrillers/suspense films

Table 1 Genre classifications (Dirks 2006)

Determining the exact genre to which each film belongs is deceptively simple. As Neale (2000; 2, Eugeni, Farinotti 2002) points out, genre is a multi-dimensional, heterogeneous phenomenon. The problem of defining film genre is most of the times created by the fact that some basic features⁸ combine with a range of secondary but still recognizable and important ones which contribute to the nature of the audiovisual product. For this reason we have been induced to use compound labels to refer to the genre of a given film.

The generic diversity of our corpus, made up of an old (Bruti, Perego 2005) and a new sample of films, is represented in Table 2, where films are ordered alphabetically. Films have been divided, where possible, according to the major traits identified by critics, despite the fact that in all of them there are components overlapping with a different genre, sub-genre or non-genre. This lack of a precise internal hierarchy of traits in which dominant features are recognisable is at the basis of any film genre study (Eugeni, Farinotti 2002: 140). This eventually leads to presenting each film as a unique network of traits. In our categorization, we tried to label each film according to its dominant generic and sub-/non- generic traits as for Table 1. As a consequence, when we use several partitions we also in-

clude “mega-genre” (e.g. ANIMATED FILMS, BRITISH FILMS, CHILDREN FILMS) and sub-genre (e.g. ROMANCE, FANTASY, MYSTERY, etc.), as this kind of information seems to be vital in distinguishing each corpus component.

The most problematic genres in terms of definition and identification are COMEDY and DRAMA. COMEDY, in its various guises, appears to be the least rigid and the most pervasive film genre due to the diversity of topics it deals with. Brancato *et al.* (2001: 71-72) maintain that COMEDY cannot be referred to as a “pure” genre (in fact they use the label “transversal ‘super-genre’”, our translation). COMEDY merges together a series of narrative functions linked to humorous effects (the Italian film *comico*) and it is liable to literally cross the borders towards more articulated expressive registers. Another fuzzy and normally over-used label is that of DRAMA, again a structurally complex genre-type. The formal and historical complexity of this genre escapes any easy classification and leaves room for different interpretations (Amendola, Frezza 2001: 166). In other words, DRAMA crosses the frontier and supports other film genres, imposing itself on other traits (Amendola, Frezza 2001: 166). Hence the need to further specify the sub-genre for each film of our corpus. Being specific on the definition of film genre by using compound labels is here necessary to correlate with precision the original choices to the ones made in the subtitled version. In Table 2, main genres are in bold and may occur in combination; when several labels are present, the first one, even though it does not always correspond to a main genre, is the one that best describes the film.

Old sample	<i>East is East</i>	COMEDY / DRAMA > FAMILY
	<i>Shallow Hal</i>	COMEDY
	<i>Sliding Doors</i>	COMEDY / DRAMA
	<i>The Talented Mr Ripley</i>	DRAMA > MYSTERY & THRILLER
New sample	<i>Bambi</i>	FULL LENGTH ANIMATION / CHILDREN > ADVENTURE > FANTASY
	<i>Lethal Weapon 4</i>	ACTION / ADVENTURE
	<i>Sense and Sensibility</i>	LITERARY ADAPTATION > COMEDY / DRAMA > ROMANCE
	<i>Sex and the City</i>	SERIAL > COMEDY
	<i>The Simpsons</i>	SERIAL ANIMATION / AMERICAN PRODUCTION > COMEDY

Table 2 Genre of the films analysed

3.1. FILM GENRE AND DVD SUBTITLES

The choice to subtitle a film or a TV programme attempts to fulfil different functions, i.e. to make a film product accessible to an international audience, or to make reception easier for an audience with hearing problems (respectively interlingual and intralingual subtitling). In addition, there is a third possible use, that in which various combinations of subtitles and soundtracks can be used in language teaching (cf. Mariotti 2002, Caimi 2008).

The development of recent technologies, e.g. satellite television and DVDs, make subtitles much more pervasive than they were only a decade ago. DVDs feature in fact numerous soundtracks and subtitles in many different languages. As Mary Carroll claims, “the quantity of DVD subtitling has boomed to such an extent in the past few years that quite different work processes have emerged. Unlike the small-scale DVD subtitling of up to approximately six languages that is common for corporate DVDs and European films, a Los Angeles-London axis has evolved to coordinate the localization of subtitles into 40 or more global languages for Hollywood releases on DVD. The price wars are fierce, the time-to-market short, the fears of piracy rampant. The aim of the subtitling companies is to deliver the best multilingual subtitles possible under the given circumstances” (Carroll 2004).

But crucial and critical questions arise: who are the addressees of the subtitles, for which ‘ideal’ audience have they been prepared and for what purpose? If one can easily understand that a foreign viewer of a famous literary adaptation like *Sense and Sensibility* wants to have access to the original version but needs some linguistic help and therefore draws on subtitles (either in English or in the viewer’s own language) to enhance his/her comprehension, the choice to use subtitles when watching a cartoon like *Bambi* (cf. Salaets 2004) or an action movie like *Lethal Weapon* is certainly harder to explain. *Bambi* is essentially destined for an audience of children, who – if already literate – certainly do not read subtitles. Likewise, the viewers of a very fast action film like *Lethal Weapon* are presumably too absorbed by the frantic rhythm of the action to concentrate on reading the subtitles.

DVD subtitling is different in nature, because it has grown to meet the needs of a booming market, where multinational companies compete with multiple language releases (Georgakopoulou 2004). Productivity procedures have therefore been speeded up, and the subtitles in the different languages are not translated from the original but from a “template” (usually in English) with obvious consequences for the quality of the translation. As a result, both the audience of audiovisual productions and their genre have a direct bearing on the quality of the subtitles that are formulated.

4. ANALYSIS OF THE CORPUS

In what follows, we propose an account of our findings, based on a comparison of old and new data. For the films in the old sample we present some summarising remarks and refer the reader to Bruti, Perego 2005 for a more detailed account.

4.1 VOCATIVES IN EAST IS EAST

East is East can be duly labelled as a COMEDY, given its pervasive “light-hearted plot consistently and deliberately designed to amuse and provoke laughter [...] by exaggerating the situation, the language, action, relationships and characters” (Dirks 2006). Terms of address here appear to perfectly embody the language-exaggeration trait emphasised by Dirks’s definition of COMEDY. Beside using and preserving an overall substantial number of terms of address (324 in the ST and

232 in the TT), exaggeration is clear in the number (60) and the nature of insults chosen in the original version. They are particularly varied and seem to be markers of both the characters' age and social class. Insults epitomise juvenile language targeting a young audience and contribute to amusing the viewer. However, some laughter-provoking descriptors do not easily find a translation in Italian. So, *freak*, *fucking gunnet*, *gob-shite*, *Jimmy little bleeder*, *you big daft get*, *you cheeky little bleeder*, *you daft tute*, *you pucker/puckers*, *you whiffer* do not have any counterpart in the subtitled version (but cf. also Table 12 and 4.6, where the extent of the problem is discussed in detail). The multi-ethnicity of the family (the Pakistani Muslim father George Khan, the British mother and children) is only rarely, but effectively, reflected in vocatives through culture-specific nicknames (cf. *Ghandi*) or titles (cf. *Sahib*, used formerly as a form of respectful address for a European man in colonial India and the Pakistani names of the children or of some of the characters. Names are at times combined with titles, which are dropped when they do not have a correspondent in the target language (*Popah Khalid* > Khalid; *Mrs Khan* > Signora Khan; *Mrs Shah* > Signora Shah; *Mr Shah* > Signor Shah). The ethnic diversity of different characters, with special reference to the "Pakistani-ness" of George Khan, is made clear through a few specific expressions clearly referring to a Pakistani habit (see the insult *you cow worshipping bastards* > *bastardi adoratori di vacche*) and through other frequent references to the Muslim culture (e.g. to George Khan's first wife residing in Pakistan, to the mosque, etc.) delivered throughout the film.

These remarks should remind us of the underlying DRAMA-like nature of *East is East*, where political, social and racial issues are dealt with, though light-heartedly, and where the development of the life conditions of both a family and of its individual members are central (Amendola, Fezza 2001: 165). The sense of family is rendered through a specific range of terms of address which highlight the relevance of family ties but also of cultural clashes within the family itself. Among the terms belonging to the semantic area of family, *kid(s)* ('fratellone', 'fratello', 'fratellino', 'ragazzi') is constantly used but not consistently translated (4 instances in the subtitles stand for the 7 in the Source Text¹⁰), but in general, kinship terms are a large number (50 > 32). The mother's affection for her children, as well as her social status, are delineated by the use of endearment descriptors such as *love/luv* and *cock* (12 overall instances), mainly addressed to the youngest and not always translated (7). The cultural clash and the constant father-and-son arguments are rendered through the use of all possible variants of the English *bastard*, duly maintained in the subtitles as well.

4.2 VOCATIVES IN *SHALLOW HAL*

Shallow Hal is a prototypical instance of COMEDY, as the plot is light-hearted and carefully contrived to provoke laughter. Hal promises to his dying father to date only women who are physically beautiful. One day, however, he runs into self-help guru Tony Robbins, who hypnotises him into recognising only inner beauty. Hal then meets Rosemary, an obese young woman whom only he can see as beautiful. In the end, although Hal's equally shallow friend attempts to undo the hypnosis, their relationship will survive. The film is on the whole hilarious and, despite a few moments of bitter laughter involving some of the physically

impaired characters, closes off on a happy note. The easy going, off-hand relationships that are depicted in the film are responsible for the high number of descriptors (42 in the ST and 25 in the TT), especially generic names, but also insults and endearments (descriptors: *kid(s)/kiddo, my friend, buddy, man, fellas*, etc.; endearments: *sweetie, darling, beautiful*; insults: *jackass, banana hands, mudwhistle, wise guy, you self righteous little prick*). While generic terms are drastically reduced in the subtitles, both endearments and insults, consequential emotional indices of the relationship between the collocutors, are always transposed (Bruti, Perego 2005: 39). Other types of vocatives featuring in *Shallow Hal* are kinship terms, titles and vocational titles. All of them usually appear in the subtitles, with a major change triggered by the item *sir*, which is appropriately translated only when applied to job-related speech situations (9 occurrences in the ST and 3 in the TT).

4.3 VOCATIVES IN *SLIDING DOORS*

Sliding doors can best be described as a mixture of COMEDY and DRAMA whose plot strand is devoted to several intersecting love stories. Young Helen is fired from her job at a PR company, and when the sliding doors of the tube close on her, we start to see what would have happened if she had taken the train and if she hadn't. The two realities move forward in tandem.

The heavy reduction in the number of vocatives from the ST to the TT (150 > 66) largely depends on the drop in proper names (122 > 54). Apart from names – the widest category of vocatives in all genres – insulting terms play an important role and are triggered by the several skirmishes of the two couples of lovers, Helen and Jerry on the one hand and Lydia and Jerry on the other. Interestingly, as the story is based on the love triangle between two women – Helen and Lydia – and worthless Jerry, cheating on both, insults are used from women to men and, since they contribute a great deal to depict the ongoing interaction, they are reproduced in the subtitles half of the times.

Generic descriptors have also undergone a drastic cut (from 9 to 0), for two main reasons: firstly they are used in exchanges between friends or intimates and thus do not establish address; secondly there are few effective translations in Italian for *man* (1 occurrence) and *mate* (6) (Bruti, Perego 2005: 32).

4.4 VOCATIVES IN *THE TALENTED MR RIPLEY*

Apparently, *The Talented Mr Ripley* is the only film where vocatives are used more neutrally, and where there is no direct connection between the use and the type of vocatives and the film genre. This might be due to the nature of the film itself. Its intricately plotted narrative makes it difficult to classify. We decided to label it as a MYSTERY AND THRILLER DRAMA to merge the traits of the major genre this film belongs to, i.e. DRAMA, with those of the two sub-genres, i.e. MYSTERY and THRILLER which float up in the plot at different times. As in a pure THRILLER, the low pace of the plot (vs. the dynamicity typical of chasings as in *Lethal Weapon 4*) generates suspense and involves the viewer emotionally (Menelao 2001: 227). As DRAMAS typically are, *The Talented Mr Ripley* is a serious, plot-driven presentation of reality where “intense character development and interaction” (Dirks 2006) are portrayed. Throughout the film, all characters go through a drastic change which is clear from the way they live, dress, move and talk. In general terms, language

expresses both emotions and social relationships. Here, terms of address are particularly wittingly used for the latter purpose. As a consequence, those vocatives which best determine the level of closeness and affection (e.g. endearments) or, vice versa, remoteness and coldness among characters, i.e. the parameters of distance and proximity in terms of social relations as in Brown and Gilman (1960), are maintained in the subtitled version. Proper last names (1), full names (2), title and last names (9) and terms of endearment (4) are not dropped at all, nor is, to a certain extent, the title *sir*, used 10 times in the original version and transposed in the subtitled version 8 times out of 10, without being translated. Proper names, which are used as an unmarked mode of address and are not relevant in establishing the type of relationship between interactants, are reduced roughly to a half (from 110 to 63) without impairing the communication of pragmatic meaning.

Apart from these few devices, the elaborate duality of Tom Ripley – the protagonist who wants to acquire an unaccounted identity (cf. Derry 1988: 175) – which is at the centre of the plot, is not delineated through the use of specific terms of address. The major traits of both the *MYSTERY* and the *THRILLER* sub-genres, i.e. the focus on an unsolved crime (more specifically, on the disappearance of one of the characters) and the mixture of tension, uncertainty and anxiety, cannot be delivered through vocatives but are made available through other predominantly non-linguistic but semiotic devices. The use of vocatives is limited to delineating the social relationships and the roles of the characters. Thus, it is not genre-defining or genre-defined, but social-situation driven.

4.5 VOCATIVES IN *BAMBI*

*Bambi*¹¹ is a (FULL LENGTH) ANIMATED (as opposed to LIVE-ACTION films) CHILDREN FILM. ANIMATIONS are not a strictly-defined genre category (Dirks 2006), but rather a film technique in which two-dimensional drawings, pictures and illustrations are combined with movements. Together with fairy tales and stop-motion films animations often appeal to children. The label “CHILDREN FILM” as well is not a genre category, but it certainly applies to *Bambi*, as it is a non-offensive, proper, and entertaining product that excludes violence, foul language and other profanity and is enjoyable without being excessively sentimental. In general, Disney cartoons, although they may be suitable for all age groups, are destined primarily for young children (cf. Zipes’s view, according to which they are “meant to captivate the ‘child’ in all the viewers”, 1997: 94), whereas *The Simpsons* (cf. 4.9) are watched by a more heterogeneous public of viewers. Here the dialogues, with the humour, the wisecracks and the many cultural references they contain, can in fact best be appreciated by adult viewers.

In the case of *Bambi* the genre proper is difficult to pin down because it contains some elements that are typical of *ADVENTURE* but others that fit the type *FANTASY* better. *FANTASY FILMS* often have an element of magic, myth, wonder, escapism, and the extraordinary. In *Bambi* there are in fact talking animals, i.e. woodland creatures and a young deer in particular, that behave like human beings and undergo a process of growth at the end of which they have learned the necessary skills to face life.

The storyline is quite simple and much of the narration is entrusted to images, songs and music. The soundtrack (nominated in 1943 for the Oscar award

in the category Best Music, Scoring of a Dramatic or Comedy Picture) is in fact essential in structuring the plot and defining it. In an interview, Walt Disney pointed out that there are only 950 words of dialogue in the entire movie (http://soundtracks.monstersandcritics.com/reviews/article__5809.php). When asked why, he answered that he “wanted the action and the music to carry” the movie. This is exactly the impression one gets when watching *Bambi*: the soundtrack actually tells part of the story.

The dialogues are therefore limited to the essential and are often reduced to very short exchanges between the characters, where cues rarely exceed the length of six/seven words. Yet vocatives, especially names and kinship terms, are extensively used in a not much talked film, not only as identifying labels that strengthen the characters’ identification for the benefit of the audience, but as substitutes for complete and articulated utterances whose meaning is conveyed visually much more than verbally. This happens especially in some of the most dramatic scenes, where children can interpret the seriousness of the situation through the images and the paralinguistic signals, such as the tone, the volume, the rhythm of the voice, as well as from the score.

VOCATIVES	ST (Tokens)	ST (Types)	ST (Tokens)	TT (Types)
Proper names (first)	47	31 Bambi 6 Thumper 3 Flower (1 Pretty Flower) 7 Faline	34	21 Bambi 6 Tippet 3 (1 Fiore carino) 4 Occhidolci
Kinship terms	20	15 Mother 3 Mama 1 Papa 1 My son	15	10 Mamma 3 Mamma 1 Papà 1 Figlio mio
Titles	4	3 Little prince 1 Young prince	5	3 Principino 2 Principino (*)
Titles + names	2	1 Mrs Quail 1 Prince Bambi	1	1 Signora Quaglia 0 (*translated as Principino)
Descriptors: generic names	5	4 Friend Owl 1 Fellas	5	4 Amico Gufo 1 Amici
TOTAL	78		60	

Table 3 *Bambi*

The evocative power of nouns is to be ascribed to their function of carriers of identity, even more so when they are meaningful¹²: e.g. *Thumper* > *Tippete*, a de-verbal noun is turned into an onomatopoeic expression; *Flower* > *Fiore*; *Faline*¹³ > *Occhidolci*.

As vocatives are a high percentage out of the total number of words in the screenplay, they are most often translated in the subtitles (78 in the original, 60 in the subtitles), in many cases replacing complete utterances. Here the meaning potential is attained by the various intersecting semiotic planes: images, sounds, score and often, lastly, language.

Kinship terms are second in rank, quite expectedly, as importance is attached to strong familiar bonds in the animal world and are omitted only when they are repeated to emphasise the tension or the climactic moment in the plot. For example, in the heart-rending scene in which Bambi's mother is shot, the little deer calls his mother repeatedly, with increased pitch, in a crescendo of anguish. The intensity of the feeling is entrusted to the images, to the music and to the pitch of the voice. The subtitles use in fact only one vocative but the cumulative effect of its repetition is achieved through the soundtrack. Both titles and descriptors have rarely been used but have always been translated.

4.6 VOCATIVES IN *LETHAL WEAPON 4*

Lethal Weapon 4 (R. Donner, 1998) perfectly fits Neale's definition of ACTION/ADVENTURE film (Neale 2000: 52-60; cf. "action movie" in Maiello 2001 and "police/detective film" in Menelao 2001: 224, cf. Italian "poliziesco"¹⁴). As Officer Riggs ironically puts it, what it is all about is usually "gunfights, explosions, sharks".

Indeed, this major genre is characterised by "a propensity for spectacular physical action, a narrative structure involving fights, chases and explosions, and in addition to the deployment of state-of-the-art special effects, an emphasis in performance on athletic feats and stunts" (Neale 2000: 52). Dirks (2006) uses the two labels separately as indicators of two different main genres, admitting the similarity of adventure to action films. In this respect, in *Lethal Weapon 4* action traits are predominant and include features such as "high energy, big-budget physical stunts and chases, [...] with rescues, battles, fights, escapes, destructive crises [fire], non-stop motion, spectacular rhythm and pacing, and adventurous, often two-dimensional 'good-guy' heroes [...] battling 'bad guys'". Nevertheless, a typically adventurous feature is the exciting story of the plot (Dirks 2006).

VOCATIVES	ST (Tokens)	ST (Types)	TT (Tokens)	TT (Types)
Proper names (first)	34	15 Leo 5 Trish 4 Ping 3 Lorna 3 Roger 2 Rianne 1 Bruce 1 Stephanie	20	10 Leo 1 Trish 0 4 Lorna 3 Roger 0 1 Bruce Lee 1 Stephanie
Proper names (last)	75	57 Riggs 4 Butters 3 Murtaugh 1 Biscuit 1 Boogers 1 Burger 1 Burton 1 Butler 1 Buttkiss 1 Hong 1 Putter 1 Summers	33	19 Riggs 4 Butters 0 1 Biscuit 1 Boogers 1 Burger 1 Burton 1 Butler 1 Buttkiss 1 Hong 1 Putter 1 Summers
Familiar forms	19	18 Rog 1 Vicki	5	4 Rog 1 Vicki
Kinship terms	8	7 Uncle Benny 1 Dad	3	3 Zio Benny 0
Titles	4	4 Sir	2	1 Signore, 1 sissignore
Titles + last names	1	1 Mr Proody	1	1 Sig. Proody
Vocational titles	15	9 Captain 1 Captains 1 Coast Guard 1 Doc 1 Doctor 1 Lieutenant 1 Officer	13	8 Capitano 1 Capitani 1 Guardia costiera 1 Doc 0 1 Tenente 1 Agente
Vocational titles + last names	12	4 Captain Murtaugh 4 Captain Riggs 2 Sergent Murtaugh 1 Doctor Woods 1 Sergent Riggs	12	4 Capitano Murtaugh 4 Capitano Riggs 2 Sergente Murtaugh 1 Dott. Woods 1 Sergente Riggs

Descriptors: generic names	32	5 Man 3 Boy 3 Guys 3 Kid 3 You guys 2 Oh, brother! 2 Everybody 1 Babe 1 Baby 1 Boys 1 Buddy 1 Falks 1 Gang 1 Girls 1 Grandfather 1 Oh man 1 You 1 You two	20	1 Cazzo 1 Cucciolo, 1 mamma mia 2 Ragazzi, 1 gente 2 Ragazzo, 1 ragazzino 1 Ragazzi 1 Mamma mia! 2 Gente 0 0 1 Ragazzi 1 Amico 1 Ragazzi 1 Gente 1 Donne 1 Nonno 0 0 1 Voi due
Descriptors: modifier(s) + generic names	2	1 Telephone tough guy 1 You little phoner	1	1 Il duro del telefono 0
Descriptors: endearments	10	8 Honey 1 Love 1 Tiger	6	5 Tesoro, 1 cara 0 1 Tigre
Descriptors: insults	22	7 Son of a bitch 2 "you plik" 1 Asshole 1 Bitch 1 Fucker 1 Fuckface 1 Fucking asshole 1 You fucking leprechaun 1 Jerk 1 Mr Big Shot 1 Naughty 1 Pigs 1 You little shit 1 You son of a bitch 1 You stupid shit	19	6 Figlio di puttana 1 "blutto stlonzo" 1 Stronzo 1 Deficiente 1 Stronzo 1 Stronzo 1 Testa di cazzo 1 Folletto 1 Stronzo 1 Grand'uomo 1 Cattivello 0 1 Pezzo di merda 1 Piccolo figlio di puttana 1 Merdaccia
TOTAL	234		135	

Table 4 *Lethal Weapon 4*

The centrality of aesthetics (both in the human hyperbolic bodies and in the hyperbolicity of the actions), the speeded-up rhythm and the relevance of both the main characters' sexual and professional identity conditions the use of vocatives to a great extent. The proliferation of vocational titles, at times combined with last names, seems to reinforce both the sexual and the professional identity of male characters, especially of those who represent *the hero*¹⁵. Vocational titles are

mainly used in situations where the addressee's authority and role are to be emphasised either positively or negatively, but they may also serve to give precise information on the professional profile of a specific character if this is important in a given context (cf. *coast guard* > *guardia costiera*; *lieutenant* > *tenente*; *officer* > *agente*). Functional labels in the form of titles + last name (*Captain Murtaugh*, *Sergent Murtaugh*, *Captain Riggs*, *Sergent Riggs*) are not necessarily used at workplace and in asymmetrical relationships, but may be used by colleagues to give an impression of intimacy. An extreme example is provided by several vocational titles in a row, used tongue-in-cheek by Riggs and Murtaugh when they get a promotion as Captains, thus becoming even more virtuous heroes by raising their professional status. The function of these specific vocatives appears to be intentionally retained in most subtitles (Figure 1).



Figure 1

In so much as heroic characters are proclaimed through language, secondary and less heroic characters are attributed this status in the same way. The wide range of mocking and incorrect surnames chosen or used inadvertently mainly by Sergeant Murtaugh addressing Detective Lee Butters appears to be used to discredit the young man and to threaten his masculinity and professional identity, which are central hero-like features in the texture of this genre, and by the director to amuse the audience. The incorrect surnames, i.e. Biscuit, Boogers, Burger, Burton, Butler, Buttkiss, Putter, Summers, have always been kept in the subtitles, but they have not been translated, so the puns deriving from their semantics are lost¹⁶.

ACTION-ADVENTURE films normally comprise the presence of single groups of people that will be eventually defeated by the hero(s). In *Lethal Weapon 4*, the friends and LAPD officers Martin Riggs (Mel Gibson) and Roger Murtaugh (Danny Glover) try to stop the Chinese Triads from doing their illegal business in the U.S.. The screenplay is therefore plain and dry (Maiello 2001: 341), though the story is exciting and meant to be an energetic experience for the viewer. Action and violence are predominant elements, emphasised by a quick editing, fast shots and taboo language. As for the latter, insults have well-established address-

ees: criminals, gangsters and underworld figures who operate outside the law. In other words: the enemy. In terms of quantity, insults are not many (22 instances in the ST and 19 in the TT), even if the film genre would allow for more, and the dark cynical way of using language, a language of conflict, is overwhelming in the original and partly maintained in the subtitled version. Its genre-defining nature does not allow the subtitler to remove or modulate it without compromising the original communicative intentions and the genre requirements. Proper names (first and last) are not so relevant as in other genres (e.g. COMEDY), which is clear from their overall number¹⁷. Generic names are restricted to a few, neutral situations, as insults are preferred as genre-markers. Finally, endearments strongly clash with insults in terms of addressees and situational context. Their overall number is low (10 occurrences), and it mainly encompasses very common ones like *honey* and *love*, regularly translated, if at all, with “tesoro” or “caro/a” (cf. § 4.7). Endearments are used in familiar settings to underline the moral side of the hero, or rather the non-hero and human nature of the protagonists. Hence, the endearments-insults opposition is here functional as it traces the line between two opposite and opposing realms, i.e. family (as a representation of affection, warmth, protection, tenderness and relaxed atmosphere) and work (as a representation of hate, lack of feeling, violence, brutality and on-the-alert atmosphere).

These two categories of terms of address also differ formally: insults are normally complex noun groups, heavily modified and situation-based (Table 12) as opposed to endearments, which appear to have a simpler internal structure and to occur in more standardised patterns. This does not allow for the same extent of creativity in terms of language, but facilitates translation. An instance of an unusual endearment, thus labelled because of the context in which it occurs (cf. the same remarks for insults in *Sex and the City*), is the common name *tiger* used by Riggs when he refers to his to-be-born baby while embracing his partner’s big belly (Example 1).



RIGGS: Easy **tiger!** Here he goes again
Example 1

Interestingly, no sharp distinction between these specific terms of address exists in terms of referents in films such as *East is East*, *Shallow Hal*, *Sex and the city* or *The Simpsons*. Here, endearments and insults are used both in everyday life and at work, and insults can be used playfully within family or among friends. The use and function of insults in *Lethal Weapon 4* therefore shows that the same category of terms of address has a different meaning and a different communicative purpose in different film genres.

By way of conclusion, we should point out that proper names, first and last, are chiefly used with the function of attention-getters, and are not regularly present in the subtitles. Most of the exchanges between Martin Riggs and Roger Murtaugh, who work together and constantly interact to defeat the enemy, occur in frantic and excited situations. Here the highest concentration of omissions can be noticed. Verbal exchanges (like shots and editing) are rapid, and the most important thing to pass is the message, often in the form of directions or requests for approval. Since style, spectacle, visual excesses, atmosphere and tone are equally important in most action pictures (Neale 2000: 54), non-particularly-relevant vocatives are typically omitted (Examples 2 and 3).

ORIGINAL	SUBTITLES
RIGGS: Don't turn around. C'mon. Will him with me, Rog .	Non girarti. Ordinaglielo con me.

Example 2

ORIGINAL	SUBTITLES
MURTAUGH: Hei Riggs , hei. Do you think the bird thing helped eh? You think it helped?	Credi che la cosa dell'uccello, sbattere le braccia ¹⁸ , abbia aiutato?

Example 3

The use of vocatives is then restored when conversation gets quiet again. This explains the gap between the occurrences of *Riggs* and *Rog* used frantically in the original dialogue (respectively 57 and 18) which drop off to 19 and 4 in the subtitles¹⁹.

4.7 VOCATIVES IN *SENSE AND SENSIBILITY*

The 1995 version of *Sense and Sensibility* directed by Ang Lee can be regarded as a delicate adaptation of Austen's novel, which deals with British social customs through the story of two very different sisters and their quest for the ideal marriage. The definition of LITERARY ADAPTATION (cf. Dirks 2006, among the various types of "mega genre"), although it does not specify the genre proper, is essential when categorising films like this. What is instead more difficult to pin down is the genre, as it is actually located at the intersection of COMEDY and DRAMA²⁰ in that it partakes of some features of both. In addition, the thematic content suggests

that the parameter of the sub-genre *ROMANCE* is especially relevant in describing this film. *ROMANCE FILMS* consist of love stories that centre on passion, emotion, and the romantic, affectionate involvement of the main characters (usually a leading man and woman) and they in fact make the love story the main plot focus. In most cases in *SCREEN ROMANCES*, and *Sense and Sensibility* follows this trend, lovers face obstacles and the dangers of hardship, finances, illness, racial or social class status, occupation, psychological restraints, or family that menace their union and attainment of love. As in all love relationships, tensions of day-to-day life, temptations (of infidelity), and differences in temper enter into the plots of romantic films.

As conversation is so central in the film, verbal action almost completely encroaches on physical action and vocatives appear with remarkable frequency. The setting of the story, i.e. 18th century England, determines the choice of types: in fact names, either first or last, and titles are the most frequent form of address. Even though emotions are part and parcel of the plot of this film, the social conventions of the period that the film aims to depict impose restraint and decorum in expressing them. Consequently, generic names and offensive terms are very few, and terms of endearment, although quite frequent, are quite formulaic and stereotypical in nature. This is perfectly in keeping with the rules of behaviour of British society at the time Austen wrote: social ranks were quite rigid and forms of address necessarily reflected status configurations.

VOCATIVES	ST (Tokens)	ST (Types)	TT (Tokens)	TT (Types)
Names (first)	92	26 Marianne 22 Elinor 13 Edward 13 Margaret 6 Lucy 5 Thomas 3 John 2 Charlotte 2 Fanny	40	12 (1 translates Miss Marianne) 7 Elinor 8 Edward 4 Margaret 2 Lucy 2 Thomas 2 John 2 Charlotte 1 Fanny
Names (last)	18	13 Willoughby 5 Brandon	11	8 Willoughby 3 Brandon
Common nouns		1 Virtue	1	1 Virtù
Nicknames		4 Pigeon 1 Pooter		2 Pigeon 1 Pooter
Kinship terms	16	13 (12 Mamma + 1 Mother) 3 Father	9	7 Mamma 2 Padre
Titles	15	12 Madam 3 Sir	4	4 Signora 0

Titles + last names	97	49 Miss Dashwood (1 fictional) 10 Mr Ferrars 8 Mrs Dashwood 8 Mrs Jennings 8 Mr Willoughby 5 Miss Steele 5 Mr Palmer 3 Mrs Bunting 1 Mrs Palmer	23	12 Signorina Dashwood 2 Signor Ferrars 0 1 Signora Jennings 4 Signor Willoughby 1 Signorina Steele 1 Signor Palmer 2 Signora Bunting 0
Titles + first names	29	19 Miss Marianne (1 fictional) 9 Sir John 1 Mr Robert	9	7 Signorina Marianne 2 Sir John 0
Titles + names	1	1 Mr Impudence	1	1 Signora Impudenza
Vocational titles	22	22 Colonel	10	9 Colonnello
Vocational titles + last names	2	2 Colonel Brandon		1 Colonnello Brandon
Descriptors: generic names	1	1 My boy	0	0
Descriptors: modifier(s) generic names	3	1 Delightful creatures 1 Poor thing 1 You poor souls	1	1 Adorabili creature 0 0
Descriptors: endearments	42	11 Dearest 9 My dear (7 + 2 My dear (interjection)) 6 (My) dear + title - 2 My dear Miss Dashwood - 1 My dear Mrs Dashwood - 1 Dear Mrs Jennings - 1 My dear Sir John - 1 My dear Madam (fictional in Willoughby's letter) 4 Dear 3 My dear + first name 3 (My) dear ladies 3 My love 2 My darling 1 Beloved Marianne 1 My dears 1 My Marianne	19	6 (2 Cara, 2 Mia cara, 2 Tesoro) 1 Mia cara 4 - 1 Mia cara Signorina Dashwood (+1 Signorina Dashwood) - 0 - 0 - 1 Caro Sir John - 1 Cara Signora 1 1 2 (1 + 1) 2 (Tesoro, Amore mio) 0 1 (Mia amata Marianne) 1 0
Descriptors: insults	2	1 Viper in my bosom 1 You sly thing	0	0 (Ho cresciuto una vipera in seno) 0 (Che furbetta!)
TOTAL	344		131	

Table 5 *Sense and Sensibility*²¹

In the original version, vocatives are 344, of which only 126 have been retained in the subtitles. Titles are the most frequent option with 166 occurrences in various combinations (title alone e.g. *Madam*, title + last name e.g. *Miss Dashwood*, title + first name e.g. *Miss Marianne*, or vocational title e.g. *Colonel*) followed by names with 116 occurrences. The percentage of translated vocatives in subtitles is instead reversed, with 55 translated names and 47 titles. The discrepancy in number between vocatives in the original and in the translation is due to the fact that both names and titles can be accessed quite easily through the original soundtrack.

Titles are used either alone or in combination with names. A remarkable case in point is provided by the two address forms used for the Dashwood sisters: Elinor (starring Emma Thompson) is always referred to as *Miss Dashwood*, as a form of respect because she is the eldest sister, whereas her younger sister Marianne (starring Kate Winslet) is usually called *Miss Marianne*, a usage that is marked for temporal dialect in English (Quirk *et al.* 1985: 1320), but, interestingly, not in Italian.

Generic names are very few (4) and have been translated only once. The reason why they are so poorly represented is that they presuppose a certain degree of informality and familiarity between collocutors. In the 4 cases identified they are used downwards, i.e. by social superiors to inferiors (e.g. *poor thing*) or by older to younger people (e.g. *you poor souls*, *delightful creatures*, *my boy*).

There are only two examples of derogatory terms (e.g. *you sly thing*, *viper in my bosom*) that have been omitted in the translation, which is a consequence of the fact that in very few relationships a certain degree of intimacy could be reached.

Finally, terms of endearment (42 in the original and 19 in the subtitles) are most times routine expressions, whose meaning does not presuppose a strong, heartfelt emotional involvement (e.g. *dear*, *dearest*, *my dear(s)/my dear + first name*, *my dear + title*)²². Conversely, in different genres, both endearment and offensive terms abound in number and are characterised by lexical creativity and inventiveness (e.g. the many examples in *East is East*, *Lethal Weapon 4*, *Sex and the City*, *Shallow Hal*, *The Simpsons*; cf. insults in Table 12; cf. also the function of *Marilyn Monroe* as an insult and *Prince Charming* as a form of endearment in *Sex and the City 4.8*).

On the whole, it can be noticed that when vocatives are translated the same type of vocative is almost always used: there is in fact just one exception in the whole film, a case where *Miss Marianne* is rendered with the first name *Marianne* in the subtitles. The use of the combination title + first name is quite significant as it is used only to address Marianne, whereas Elinor is always referred to as *Miss Dashwood* (cf. above). Changing the type of vocative is not a serious fault in itself, but it weakens the conventional pattern of address used in the original and consequently does not reproduce the frozen quality of social behaviour.

The only instance of creativity in the use of terms of address can be observed in Willoughby's parody of Mrs Jennings, where he mocks the old lady and employs some 'fictional' vocatives, two of which are common modes of address (*Miss Dashwood* and *Miss Marianne*), but one is instead rather inventive, as it is directed to an imaginary character, *Mr Impudence* (and is translated into Italian with a necessary change of grammatical gender, "Signora Impudenza")²³.

4.8 VOCATIVES IN *SEX AND THE CITY*

Sex and the City is a SERIAL COMEDY that was broadcast for six years running on American TV (and afterwards in Europe as well) with huge success. Set in New York, the show focuses on the sex lives of four female best friends, three of whom are in their mid-to-late thirties, and one of whom, Samantha, is in her forties. The series became famous for shooting scenes on the streets and in the bars, restaurants and clubs of New York City while pushing the envelope of fashion and shattering sexual taboos. Even though it was based on the book that was compiled from the New York Observer column *Sex and the City* by Candace Bushnell, only the first season can be defined as a free adaptation of its source material, but from the second season on, it took on a life of its own. The language used is very colloquial as in *Sex and the City* the four women friends talk with candour and humour about men, sex and relationships, inspiring a new language and providing a new model of the ways women of all ages all over the world talk about men, dating and sex.

The hilarious quality of the series depends on both comic situations and verbal humour. This is reflected in the use of puns, allusions, and inventive and piquant language, a feature that is also suggested in the type of vocatives that have been used. As most of the time is spent chatting or discussing (apart from some linking narrative bits where Carrie – Sarah Jessica Parker – reads aloud what she has written on her laptop for the column *Sex and the City*, cf. above), vocatives are often employed, most of the time with a pragmatic function. In fact, in face-to-face conversation, both the topic of the exchange and the various cinesic and proxemic signals are enough to select addressees. Carrie, as the main character in the series, is both the sender and addressee of most of the vocatives in the episodes that have been analysed, i.e. “The Agony and the Ex-tasy” (from now on referred to as episode one) and “I heart NY” (referred to as episode two), both belonging to the fourth series.

The overall number of vocatives in the two episodes is respectively 29 and 32, which have been translated in both cases 19 times.

VOCATIVES	ST (Tokens)	ST (Types)	ST (Tokens)	TT (Types)
Names (first)	16	4 Carrie 4 Charlotte 2 Miranda 2 Trey 1 Hailey 1 Phil 1 Samantha 1 Sheila	12	2 Carrie 4 Charlotte 1 Miranda 2 Trey 1 Hailey 1 Phil 0 1 Sheila
Nicknames	1	1 Shecky	1	1 Spiritosona
Vocational titles	2	1 Friar Fuck 1 Sister	2	1 Frate Scopata 1 Sorella
Descriptors: generic names	4	1 Lady 1 Old chap 1 You guys 1 World	2	0 1 Vecchio mio 0 1 Gente
Descriptors: Endearments	6	3 Baby 3 Honey	2	1 Tesoro 1 Tesoro
Descriptors: insults	1	1 Marilyn Monroe	1	1 Marilyn Monroe
TOTAL	30		20	

Table 6 *Sex & the City*: The Agony and the Ex-tasy
VOCATIVES IN SUBTITLES

VOCATIVES	ST (Tokens)	ST (Types)	ST (Tokens)	TT (Types)
Names (first)	16	4 Carrie 3 Bobbo 2 Danny (1 <u>fictional</u>) ²⁴ 2 Miranda 2 Steve 1 Eric 1 Richard 1 Samantha	7	0 2 Bobbo 1 Danny 1 Miranda 1 Steve 1 Eric 1 Richard 0
First + last	1	1 Richard Wright	0	0
Nicknames	2	1 Mummy biggest 1 Prince Charming	2	1 Mammina 1 Principe azzurro
Kinship terms	1	1 Pops	1	1 Babbo
Titles	1	1 Ma'am	1	1 Signora
Vocational titles	2	2 Nurse	2	2 Infermiera
Descriptors: endearments	8	2 Kid 2 Gorgeous 1 Baby 1 Lover (<u>fictional</u>) 1 Sweet 1 Sweetie	5	2 Piccola 1 Bella 1 Tesoro 1 Amore mio 0 0
Descriptors: Insults	1	1 You bastard	1	1 Bastardo
TOTAL	32		19	

Table 7 *Sex & the City*: I heart NY

First names are here as well the most frequent vocative, but as was hinted at before, they seldom perform an identifying function, apart from the obvious cases in which they are used either in telephone conversations, for instance when a message is recorded on the addressee's answering machine (cf. Example 4 taken from episode 1).

ORIGINAL	SUBTITLES
CHARLOTTE: Carrie , I'm in traffic. They're paving Fifth Avenue. I can't get through. Don't wait for me, happy birthday.	Carrie , sono nel traffico. Stanno pavimentando la Quinta strada./ Non aspettarmi, buon compleanno.

Example 4

More often than they somehow contribute to increase or diminish the force of a speech act (cf. Example 5 taken from episode 2)

ORIGINAL	SUBTITLES
SAMANTHA: If he's cheating on me I've got to know now. <i>Carrie</i> : Samantha , if you love him don't you think there's even the slightest possibility that he loves you?	Se mi tradisce, devo saperlo./ Tu lo ami, non credi sia possibile che anche lui ti ami?

Example 5

In Example 5, for instance, Carrie tries to convince Samantha that Richard, the man she's going out with, might be sincerely in love with her. The vocative in initial position lends strength to her remark by involving Samantha and getting her to take her hypothesis into account. Despite its pragmatic function as an illocutionary force indicating device, the vocative is not retained in the subtitles, nor is its loss made up for.

Another instance is provided by Example 6 (episode 1), where Charlotte is told off for her fussy attitude and detachment. The vocative is exploited to reinforce the imperative that follows and is effectively retained in the subtitles.

ORIGINAL	SUBTITLES
SAMANTHA: Oh, come on Charlotte , get that judgemental puss off and join in.	Charlotte , non fare la critica e partecipa.

Example 6

Endearments are the second frequent form of vocative. It is not always easy to draw a clear boundary between endearments and nicknames on the one hand and insults on the other. If one takes vocatives at their face value confining oneself to their literal meaning, *Marilyn Monroe* (episode 1) should be considered a positive evaluation, but its situational context (Carrie steps into fresh tarmac in front of a group of annoyed workers) makes it clear that it is an attack and is therefore loaded with irony and sarcasm (Figure 2 and 3).



Figure 2



Figure 3

In episode 2, Carrie calls Big – the man with whom she has a long on-and-off relationship – *Prince Charming* after he has behaved nicely to her and driven her in a carriage to Mount Sinai Hospital to assist Miranda in her labour. It is certainly meant to flatter him and also to thank him for being such a gentleman, but there is a more hidden implication, evoked by the frame of the fairy tale, for Carrie would like him to commit and settle down with her. So in both cases well-known referential expressions are used to exploit both the encyclopaedic meaning they evoke and also the more specific meaning they represent on a precise occasion. For the purposes of the present classification, nicknames have been distinguished from endearments on the basis of one main parameter, i.e. whether they are sometimes innovative expressions whose interpretation is heavily context-dependent (nicknames) or if they unambiguously convey a positive meaning (endearments). By way of summarising, both nicknames and insults are always retained in the subtitles; endearments are sometimes omitted due to their inconsiderable semantic load. Many of them are in fact conventional expressions that do not presuppose much emotional involvement. Some peculiarities that deserve notice are the ironic use of titles in episode 1 (e.g. *Friar Fuck*, *Sister*) and of kinship terms in episode 2 (e.g. *Pops*), both of which have been translated (e.g. *Frate Scopata*, *Sorella*, *Babbo*, the former, however, quite disappointingly, as the intertextual allusion to Robin Hood's friend is lost).

Furthermore, in episode 2 there are two instances of “fictional” vocatives²⁵, i.e. forms of address that are directed at someone who is not present or who does not exist in the real world (cf. similar instances in *Sense and Sensibility*, 4.7). In Example 7 Steve is choosing a name for his future son and tries to address him to test how the name ‘Danny’ sounds in a piece of hypothetical dialogue. The second part of his turn signals the boundary of the fictional frame and his coming back to reality.

ORIGINAL	SUBTITLES
STEVE: Hey Danny you want to go shoot some hoops? It sounds right.	" Danny ti va di tirare a canestro?" Suona bene.

Example 7

In Example 8 (episode 2) Carrie is in front of a shop window and gazes at a pair of sandals that she would love to buy and wear on her date with Big. So she talks to the shoes as if they were a person. Then she immediately resumes normal speech, as she explains her reaction, most of all to the advantage of the audience.



ORIGINAL	SUBTITLES
CARRIE [to a pair of shoes in a shop window]: Hello, lover! I'm needing those for my last Big night on the town.	Ciao, amore mio! Mi servono per la mia ultima notte con Big.

Example 8

In both cases, since they express unexpected, unconventional meanings, the vocatives have been kept in the subtitles.

4.9 VOCATIVES IN *THE SIMPSONS*

In terms of genre, *The Simpsons* is a peculiar audiovisual product. We labelled it as SERIAL ANIMATION, which allows us to highlight and merge the predominant features of the audiovisual product at stake. Typically, animations are considered kids or family oriented films (Dirks 2006; cf. 4.5 on *Bambi*). *The Simpsons* is mainly targeted at a teenage or adult public, and not necessarily at children, as this segment of audience does not have the necessary background knowledge to grasp the references in the plot.

VOCATIVES	ST (Tokens)	ST (Types)	TT (Tokens)	TT (Types)
Proper names (first)	45	19 Homer 12 Bart 9 Marge 2 Eugene 1 Barney 1 Mike 1 Moe	32	13 Homer 8 Bart 6 Marge 2 Eugene 1 Barney 1 Mike 1 Moe
Proper names (last)	11	8 Simpson 3 Smithers	11	8 Simpson 3 Smithers
Familiar forms	1	1 Barn	1	1 Barn
Full names	2	2 Homer Simpson	1	1 Homer Simpson
Kinship terms	10	5 Dad 2 Mom 2 Son 1 Daddy	9	4 Papà 2 Mamma 2 Figliolo 1 Papà
Titles	11	7 Sir 1 Ladies 1 Lady 1 Madam 1 Mister	6	4 Signore 1 Ragazze 0 0 1 Signore
Titles + last names	1	1 Mr Burns	1	1 Signor Burns
Vocational titles + titles	1	1 Mr Maestro	1	1 Maestro
Vocational titles + last names	1	1 Reverend Lovejoy	1	1 Reverendo Lovejoy
Descriptors: generic names	17	4 Man 2 Buddy 2 Folks 2 Guys 2 People 1 Boy 1 Fat boy 1 My boy 1 Pal 1 You two	10	1 Signora 1 Bello 1 Gente 1 Ragazzi 2 Gente 1 Figliolo 1 Grassone 0 1 Amico mio 1 Voi due
Descriptor: modifier(s) + generic names	1	1 Liberty bell	1	1 Campanella
Descriptors: endearments	7	2 Honey 1 Baby 1 Dear 1 Doll 1 Me little bucko 1 Princess	7	2 Tesoro 1 Piccola 1 Caro 1 Bambolina 1 Mio piccolo amico 1 Principessa
Descriptors: insults	7	1 Full moon! 1 Pint-size 1 That purple fruit thing! 1 You big b... 1 You gross! 1 You big -- 1 You little --	7	1 La luna piena 1 Tappetto 1 Dannato frutto viola! 1 Grosso ... 1 Che schifo! 1 Grosso disg... 1 Piccolo disg...
TOTAL	116		89	

Table 8 *The Simpsons*: Homer in the night

Proper names (first)	32	9 Moe 8 Homer 3 Bart 2 Formico 2 Lenny 2 Michael 1 Cecil 1 Dagmar 1 Julian 1 Lisa 1 Marge 1 Shaggy*	28	7 Moe 7 Homer 3 Bart 2 Formico 2 Lenny 1 Michael 1 Cecil 1 Dagmar 1 Julian 1 Lisa 1 Marge 1 Shaggy
Diminutives	2	1 Bob 1 Homie	2	1 Bob 1 Homie
Full names	1	1 Moe Syzslak	1	1 Moe Syzslak
Common nouns	8	7 Turkey 1 Cougar	8	7 Tacchino 1 Puma
Kinship terms	5	4 Dad 1 Father	4	3 Papà 1 Padre
Vocational titles	5	4 Professor 1 Barkeep	4	4 Professore 1 Barista
Descriptors: generic names	10	3 Pal 1 Baby 1 (Oh), boy! 1 Buddy 1 Guys 1 Kids 1 Man 1 Young man	9	3 Vecchio mio 1 Baby 1 Bene, bene 1 Amico 1 Ragazzi 1 Ragazzi 0 1 Giovanotto
Descriptors: modifier(s) + generic names	2	1 Coffee boy 1 You glass-wipe	1	0 1 Vecchio canovaccio
Descriptors: endearments	1	1 Doll	1	1 Bambola
Descriptors: insults	5	2 Suckers 1 Smart guy 1 Son of a ... 1 You dirty teens	5	1 Fresconi! 1 Che polli! 1 Intelligentone 1 Figlio di ... 1 Sporchi ragazzacci
TOTAL	71		63	

Table 9 *The Simpsons*: Homer the Moe

The episodic form, the 15-to-20-minute length of each episode, the presence of permanent hero-like characters that present themselves again from week to week are all typical features of SERIAL FILMS (Cati 2000; Dirks 2006).

The Simpsons is a particularly versatile audiovisual product, liable to adjust well to different sub-genres depending on the main theme of each episode. It exploits adult-oriented topics – a typical feature of teenpics (Neale 2000: 121) – and intertextual references, and besides engaging both the spectator’s generic and specific knowledge to a great extent, it tends to borrow devices from other established genres and to either foreground or background them conveniently (Altman 1996: 279). Language adapts from episode to episode. Nevertheless, the fixed expository pattern of the series seems to generate a general trend whereby the most widely used and preserved terms of address are proper names that identify the main characters, those who are at the centre of the attention, those who are the stars of the series itself or, at times, the stars of a specific episode (Table 10).

Proper names			
<i>Homer in the night</i>		<i>Homer the Moe</i>	
ST	TT	ST	TT
19 Homer 12 Bart 9 Marge 2 Eugene 1 Barney 1 Mike 1 Moe	13 Homer 8 Bart 6 Marge 2 Eugene 1 Barney 1 Mike 1 Moe	9 Moe 8 Homer 3 Bart 2 Formico 2 Lenny 2 Michael 1 Cecil 1 Dagmar 1 Julian 1 Lisa 1 Marge 1 Shaggy	7 Moe 7 Homer 3 Bart 2 Formico 2 Lenny 1 Michael 1 Cecil 1 Dagmar 1 Julian 1 Lisa 1 Marge 1 Shaggy

Table 10 Distribution of proper names in *The Simpsons*

Although proper names dominate in term of number, when reference is not made to the stars of the series, language is used creatively. The creative use of language is a double edged weapon: it enriches the original but has the potential of triggering inadequate translations in the subtitled version and, in worst cases, unwanted shifts in register and pragmatic meaning, as in Example 9, where the slang expression used by a young guy, dressed-up in pirate-clothes and with a bandanna, gains a totally different illocutionary force:



ORIGINAL

WAITER: Ahoy! I spy the children's menu!
 BART: Ahoy! This place bites.
 MARGE: Bart!
 WAITER: So, what's it gonna be, **me little bucko**?

SUBTITLES

Ehi! Vedo un menu per bambini!//
 - Ehi! Questo posto fa pena!
 - Bart!//
 Allora cosa scegli,
mio piccolo amico?

Example 9

Chiefly Irish for young man or lad, *bucko*, which also marks pirate language, seems here to be used with the meaning of blustering or bossy person, a person who is domineering and bullying. It is in fact addressed to Bart, who is actually conducting himself in an arrogant and superciliously pompous manner, with an air of overbearing self-confidence when reading the menu aloud and deciding what to eat. The reference to Bart's behaviour, no matter how mocking, is not only lost but also completely distorted by the Italian version *mio piccolo amico* 'my little friend'.

The sense of continuity and of closeness among characters in the series is made explicit by the choice of the generic descriptor *pal*, consistently translated as "vecchio mio" (but see "amico mio" in "Homer in the night"), as if it worked as an in-group making device.

As previously mentioned, language adapts from episode to episode according to the plot and to the most outstanding generic trait. In "Homer the Moe", the plot has a profound effect on the selection of vocatives, which requires the ability of the subtitler to maintain the original intentions whenever possible. Homer, temporarily substituting Moe in his bar and therefore working as a bartender,

uses language accordingly and adapts his register to the social circumstances even when at home, where he addresses his kids with unusual terms of address and resorts to bar-linked gambits²⁶.

Overall, we observed that a substantial number of terms of address is concentrated in relatively short episodes, which makes us wonder whether the number (and not only the type) of vocatives may depend on genre. Secondly, the apparent overuse of vocatives, as if there was no sufficient context to rely on in the identification of the addressees, seems to be dictated by the search for continuity or as a cohesive devices used to interconnect episodes. Thirdly, both episodes comprise a scant number of insults, although all very inventive and context-based, but as a general trend offensive descriptors are avoided and broken off both in the original and in the subtitles, as in the case of *son of a ... , you little b..., you big b...*. This is in opposition to what happens in films which are differently targeted. In *The Simpsons* insults are meant to be hilarious and enjoyable, neither abusive nor provoking. Finally, findings are in line with the general and cross-genre trend whereby vocatives are dropped when the referent is visible (cf. MOE: Wait a minute **Homer** => Un attimo) and they are preserved when they are functional, i.e. in the case of attention-getters (MARGE: **Bart?** This is Doctor Kaufman => **Bart?** Questo è il dottor Kaufman).

Descriptors in the form of generic names are the second most frequent type of vocatives used (17 > 10 “Homer in the night”; 10 > 9 “Homer the Moe”), in line with the American English usage (Biber *et al.* 1999: 1109). The same applies to the use of the typically American generic name *man*, only used by Bart in the episodes analysed. After these considerations, and crossing-over the traditional film genre types, it is possible to tag *The Simpsons* with the non-genre foil category label AMERICAN PRODUCTION (Dirks 2006). Indeed, culturally American features appear all through the series on different planes, that of language and that of content.

4.10 DESCRIPTORS: A FEW REMARKS

Descriptors are a very versatile instrument to describe and evaluate the addressee. As Allerton (1996: 621) puts it:

When speakers wish to refer to an individual concrete entity – a person, a creature, a thing or a place – they find that their language offers them a range of possible linguistic units for doing so; most of these are noun phrases. The variety of structural possibilities for noun phrases, together with the range of vocabulary means that even when it comes to picking out a particular item in a given context, a language provides its speakers with a choice of linguistic expressions.

Thus speakers select a descriptor on a relevance basis (Allerton 1996: 622), i.e. of the addressee’s most outstanding features, on his/her role in society or in the situation.

As already observed (cf. Bruti, Perego 2005), the creativity of the English language often fails to have a proper counterpart, or a counterpart at all, in Italian, where, for example, the same lexeme translates a variety of ingenious English expressions (e.g. *stronzo* for *fuckface*, *fucker*, *jerk*, *asshole* and *prick* in *Lethal Weapon 4* and for *slug* and *you cheeky twat* in *East is East*). Furthermore, English shows an

inclination to choose and eventually make up terms of address in terms of their functionality in a given context (cf. Allerton 1996). In other words, English can easily exploit the current situation and the speaker's standpoint in the selection of an appropriate referring expression, which may cause misunderstandings and meaningless translations.

The variety of structural possibilities for noun phrases used in English is shown in Table 11 as far as evaluative descriptors are concerned, and Table 12 as far as insults are concerned.

SOOURCE TEXT	TARGET TEXT
Telephone tough guy (<i>Lethal Weapon 4</i>)	Il duro del telefono
You little phoner (<i>Lethal Weapon 4</i>)	o
Barkeeper (<i>The Simpsons</i>)	Barista
Coffee boy (<i>The Simpsons</i>)	o
Liberty bell (<i>The Simpsons</i>)	Campanella
You glass-wipe (<i>The Simpsons</i>)	Vecchio canovaccio

Table 11 Evaluative descriptors (modifier + generic name)

SOOURCE TEXT	TARGET TEXT
You big daft get (<i>East is East</i>)	o
You cheeky twat (<i>East is East</i>)	Stronzo
You cow worshipping bastards (<i>East is East</i>)	Bastardi adoratori di vacche
You mard arse (<i>East is East</i>)	o (Quanto sei cretina)
Viper in my bosom (<i>Sense and Sensibility</i>)	o (Ho cresciuto una vipera in seno)
You sly thing (<i>Sense and Sensibility</i>)	o (Che furbetta!)
Pint-size (<i>The Simpsons</i>)	Tappetto
Mudwhistle (<i>Shallow Hal</i>)	Ciccio
Unibrow (<i>Shallow Hal</i>)	Scimmione
You little warthog (<i>Shallow Hal</i>)	Schifoso maiale
You self-righteous little shit (<i>Shallow Hal</i>)	Arrogante testa di cazzo
You drunken eejit (<i>Sliding Doors</i>)	Ubriaca e suonata
You lazy git (<i>Sliding Doors</i>)	Pigrone
You sad, sad wanker (<i>Sliding Doors</i>)	o (Sei un patetico, triste segaiolo)
You useless shagging bastard (<i>Sliding Doors</i>)	o (Sei un farabutto, scopatore, buono a nulla!)
That purple fruit thing! (<i>The Simpsons</i>)	Dannato frutto viola!
You dirty teens (<i>The Simpsons</i>)	Sporchi ragazzacci

Table 12 Insults

By way of illustration, let us consider evaluative descriptors in *The Simpsons*. They are mostly situation-based and cannot work without the visuals. We can imply the meaning of *liberty bell* only when this term is fully contextualised, i.e. when we see that Homer is scratching a scratch'n'win card, and finds 2 bells and a piece of fruit (cf. Ex. 10 and Figure 4).

ORIGINAL	SUBTITLES
<p>HOMER: Oh, liberty bell! Eh? Another liberty bell! Another and I am millionaire! C'mon, liberty bell, win, win, win! Oh! That purple fruit thing! Where were you yesterday?</p>	<p>Una campanella. // Forza campanella, ti prego, ti prego, ti prego. // Un'altra campanella! // Un'altra e sono ricco! // Dannato frutto viola! / Dov'eri ieri?</p>

Example 10



Figure 4

4.1.1 VOCATIVES AS SECONDARY INTEJECTIONS

An interesting remark concerns the intersection area between vocatives and interjections. Vocatives can sometimes be employed in conversation as secondary interjections, that is expressions that still have autonomous meaning but can be used to express a sudden feeling or reaction. The employment of vocatives as secondary interjections in conversation has been observed in different forms in line with different film genres. Overall, *man* is the most widely used form in situations of surprise or disappointment²⁷ and in genres which comprise informal or humorous situations along with an American background (cf. *Shallow Hal* (Bruti, Perego 2005: 39-40), *Lethal Weapon 4* (§ 4.6) and *The Simpsons* (§ 4.9)).

When the function of *man* is clearly that of an interjection indicating intense feeling, or if its role is that of a generic name (cf. Example 11, *Lethal Weapon 4*), it is usually not rendered.

ORIGINAL	SUBTITLES
RIGGS: Sorry, Rog, about your loss. MURTAUGH : Thank you, man , thank you.	-Mi dispiace per i tuoi danni. -Grazie.

Example 11

Nevertheless, in such a case as Example 12, it is not clear whether the Italian taboo word substitutes for the interjection *man* or for the English colloquial intensifier *fucking*.

ORIGINAL	SUBTITLES
LEE BUTTERS: He's dead, eh? He's fucking dead, man	È morto. È morto, cazzo .

Example 12

The same trend has been observed in *The Simpsons*, and, more generally, in some of the films previously examined (*Shallow Hal*, *Sliding Doors*). In *Shallow Hal* (§ 4.2) it appears 6 times and is never translated in the Italian subtitles; in *The Simpsons* (§ 4.9), for instance, of the 4 occurrences of the episode “Homer in the night” only one is rendered as “signora” as its function is verbalised (Table 13); in *Sliding Doors* it appears once as *old man* and is omitted in the subtitles. It is therefore clear that genre does not have an effect on the way it is rendered nor has it on the preferred choice of omitting it. The use of *man*, though, is a marker of the language variety of the film, which is known to be typically lost in subtitles. Although it usually functions as an interjection, we decided to consider it a vocative which tends to be rendered according to its pragmatic meaning, with the subtitler opting for functionally equivalent solutions, where the original interjection disappears (Table 13, Ex. 13)²⁸.

ST	TT	CONTEXT
Wow! Cool, man!	Che forza!	Bart is looking at some ads on the newspaper and cries
Oh, thanks, man!	Ah! Grazie, signora.	Bart is excited because he eventually gets a spy-camera by mail. He thanks the female carrier animatedly
Whoa, man!	Wow!	Bart is admiring his new spy-camera
Good luck, man!	Buona fortuna	Bart wishes good luck to his father, who has to redeem for having danced with a hot lady in a club

Table 13 Interjections



BART: Whoa, **man!** Look at the size of this thing!
(Bart is admiring his new spy-camera)

Example 13

5. CONCLUSIONS

On the whole, the survey has shown that there is a certain difference in the type of vocatives chosen in each film depending on the genre and in some cases on the country of production. As far as translation is concerned, similar trends in the choice of what and how to translate have been observed in spite of genre. That is, a larger corpus has allowed us to observe that the tendency to discriminate between more or less relevant/meaningful vocatives is maintained when possible. In other words, vocatives with a relevant pragmatic meaning are usually retained. Some problems that are shared by all film genres arise when the subtitler has to deal with English informative and creative labels (i.e. descriptors,

either insults or endearments), which may be problematic in terms of meaning and length.

The class of nouns deserves a special mention: it is in fact the widest category in all genres, with some remarkable differences. They take on special significance in both *Lethal Weapon 4* and *Bambi*, and consequently in the subtitles in which they are most often translated, although for different reasons: in the former Murtagh intentionally manipulates Detective Lee Butters's name to create puns and thus to pour ridicule over him; these names are all meaningful and convey Murtagh's low opinion of the young detective. In *Bambi* nouns not only have great evocative power, but are meaningful and contribute therefore to personifying the characters. They are therefore used and translated even when it would have been possible to avoid them. This also happens in *The Simpsons*, where the density of names seems to be particularly high both in source and target text on the grounds that there are a group of stable characters in the series but other that change from episode to episode. A certain degree of continuity needs to be preserved and this is achieved through names as identifying referential labels selecting the most important characters. Repetition of names is in other words a signal of importance, and helps the audience to memorise and prioritise them, according to the role the referents play in the plot.

Another interesting finding concerns the different functions that the same category of vocatives have across genres: insults, for example, seem to be the category that is most strongly influenced by genre, which in fact determines their number, type, function, and translation. In *Lethal Weapon 4* they belong to the language of violence and conflict that is connected to the professional role of the two protagonists, whereas in comedies (*East is East*, *Shallow Hal* and also in *The Simpsons*) they are non offensive and amusing. Endearments, which are formally simpler and less situation-based than insults, acquire a genre-specific status only when they enter into opposition with insults (cf. *Lethal Weapon 4*). On the other hand, if their semantic load is inconsiderable or they are used as routine expressions (*Sex & the City* and *Sense and Sensibility*), they tend to be removed. Titles, conjoined or disjoined from first and last names, are another category that is tightly linked to (sub-)genre: literary adaptation > romance are defined by the frequency of titles, which are necessary markers of social customs, rigid ranks and status configuration. On the other hand, they are used in other film genres only in specific social situations. Generic names are preferably used in American productions (*The Simpsons*, *Shallow Hal* and *Lethal Weapon 4*) or as social-class markers (*East is East*), in the presence of young protagonists (*Sex and the City*, *East is East*, *Shallow Hal* and *Lethal Weapon 4*) but they are poorly represented in those cases where the film genre implies a high degree of formality and colloquial language (*Sense and Sensibility*).

	<i>Bambi</i>		<i>East is East</i>		<i>Lethal Weapon 4</i>		<i>Sense and Sensibility</i>		<i>Sex and the City</i>		<i>The Simpsons</i>		<i>Shallow Hall</i>		<i>Sliding Doors</i>		<i>The Talented Mr Ripley</i>	
	ST	TT	ST	TT	ST	TT	ST	TT	ST	TT	ST	TT	ST	TT	ST	TT	ST	TT
Proper names (first)	47	34	166	116	34	20	92	40	32	19	60	77	116	87	122	54	110	63
Proper names (last)	0	0	1	1	75	33	18	11	0	0	11	11	0	0	0	0	1	1
Familiar forms	0	0	8	6	19	5	0	0	0	0	1	1	0	0	1	1	1	0
Diminutives/Nicknames	0	0	5	5	0	0	5	3	3	3	2	2	8	8	0	0	11	1
Full names	0	0	0	0	0	0	0	0	1	0	3	2	1	1	1	1	2	2
Common nouns	0	0	0	0	0	0	1	1	0	0	8	8	0	0	0	0	0	0
Kinship terms	20	15	52	32	8	3	16	9	1	1	15	13	9	10	1	1	1	1
Titles	4	5	10	5	4	2	15	4	1	1	11	6	11	5	2	0	10	8
Title + first name	2	1	0	0	0	0	29	9	0	0	0	0	0	0	0	0	0	0
Title + last name	0	0	11	7	1	1	97	23	0	0	1	1	6	4	0	0	9	9
Title + common noun	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
Vocational titles	0	0	4	2	15	13	22	9	4	4	5	4	5	5	1	0	0	0
Vocational title + last name	0	0	0	0	12	12	2	1	0	0	1	1	0	0	0	0	0	0
Vocational title + title	0	0	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0	0
Descriptors: generic names	5	5	10	4	32	20	4	1	4	2	27	19	25	8	6	0	0	0
Descriptors: generic + modifier	0	0	0	0	2	1	3	1	0	0	3	2	0	0	1	0	0	0
Descriptors: endearments	0	0	15	9	10	6	42	19	14	7	8	8	9	9	6	5	4	4
Descriptors: insults	0	0	60	45	22	19	2	0	2	2	12	12	8	8	9	4	0	0
TOTAL	78	60	343	233	234	135	349	132	62	39	186	151	198	145	150	66	149	89

- 1 While both authors are responsible for the introduction and the conclusions (par. 0 and 5) along with par. 4.10 and 4.11, Silvia Bruti is the author of par. 1, 3.1, 4.2, 4.3, 4.5, 4.7, 4.8 and Elisa Perego is the author of par. 2, 3, 4.1, 4.4, 4.6, 4.9.
- 2 For a descriptive account of vocatives see both Davies 1986 and Gramley and Pätzold 1999.
- 3 From a sociolinguistic point of view, the way individual speakers or group of speakers use the repertoire of address variants available to them is extremely meaningful, as address behaviour reflects the speaker's (and, more generally, the dyad's) social and linguistic background (Braun 1988: 13), by exposing his/her status or illustrating what relationship (e.g. close, intimate, strained, troubled, etc.) is established between the collocutors.
- 4 On the connection between genre and dialogue pattern see Kozloff 2000: 137 and ff.
- 5 We are most grateful to Yves Gambier for providing us with the stimulus to work in this direction. His comments on our presentation in Alicante in May 2004 were challenging and thought-provoking.
- 6 Along with specific use of specific linguistic features, as this work highlights.
- 7 I.e. main film genres, sub-genres and non-genres. Main film genres are the most common and identifiable ones; film sub-genres are identifiable sub-classes within the larger film genre container; non-genre films cross-over many traditional film types.
- 8 Both attributes or intentions may be referred to in order to define film genres (Tudor 1986: 4). Altman (1986: 31) proposes instead a distinction between semantic and syntactic approaches to genres, the former based on the genre's building blocks, the latter on the structural sequence.
- 9 Cf. Carroll "Such a template can make sense if it is thoroughly researched and well-timed, especially if subtitlers are free to use it as an aid but are not compelled to force their translation, regardless of its structure, into its mold. However, the rigidity of such files can result in poor subtitling with little adherence to now common standards of good subtitling practice".
- 10 From now on the abbreviations ST and TT will be used referring respectively to Source- and Target Text.
- 11 The story is based on a novel published in 1926 by the Viennese author Felix Salten (cf. <http://disney.go.com/vault/archives/characters/bambi/bambi.html>).
- 12 Quite interestingly, the names in the Italian subtitled and dubbed version differ: in dubbing, in fact, *Tippete* becomes *Tamburino* and *Occhidolci* becomes *Feline*. The second choice seems to privilege a shorter term, but the first one uses a longer word, with the same initial sound as the one used in the subtitles.
- 13 A search on the web gave the following result: *Faline* is a name of Latin origin, meaning "like a cat". Possible variants are *Faeleen*, *Fayline*, *Felina*, *Feline*. As for its popularity, the name *Faline* was not ranked among 4275 first names for females of all ages in the 1990 U.S. Census (<http://www.thinkbabynames.com/name/o/Faline>).
- 14 Menelao (2001) highlights a further interesting trait of this genre, not mentioned elsewhere, whereby the crescendo of suspense is tightly connected with the speed of the actions.
- 15 I.e. the central character, one around which the plot is built, a person involved in a dangerous experience either by chance or for duty (Maiello 2001: 340).
- 16 Most of these fake surnames have a subtly negative connotation

or are meant to be funny. *Biscuit* might be considered as having a clear feminine connotation; *Booger* is a US slang expression for a piece of dried mucus from inside the nose; *Burger* refers to food; *Burton* has a possibly endless number of referents (<http://www.answers.com/burton>); *Butler* is a male servant; *Buttkiss* is a slang expression (or a possible variant of *bupkis*) for 'nothing'. In any case, no matter what they exactly refer to, they are used in line with the speaker's (or the director's) intention to create a funny situation.

17 34 vs. 166 in *East is East*, 122 in *Sliding Doors*, 116 in *Shallow Hall*, 110 in *The Talented Mr Ripley*, and in the new sample 92 in *Sense and Sensibility*, 77 in *The Simpsons*, 47 in *Bambi* and 32 in *Sex and the City*.

18 Interestingly, a non-verbal message conveyed by the speaker imitating the gesture, is encoded linguistically.

19 It is interesting to make mention of three instances (which have not been tabled) where a vocative is present in the Italian version only. Apparently, inserting a vocative in the written text compensates for repetitions in the spoken text, contributes to give a self-contained appearance to the subtitle and allows the subtitler to avoid unnatural formulae (cf. LEO to ROG: Don't leave me! Don't leave me! > Non lasciarmi **Rog**; OFFICER: You're all mine! All mine! Sei tutto mio, **bello**./Tutto mio.; RIGGS to LORNA: Bye bye! > Ciao, **tesoro**.)

20 Cf. definitions for COMEDIES and DRAMAS in Dirks 2006, <http://www.filmsite.org/comedyfilms.html> e <http://www.filmsite.org/drama-films.html>.

21 Both Pigeon and Pooter have been classified as nicknames. The former refers to Mrs Jennings's parrot and the second to her butler. A search in dictionaries of English names/surnames gave no results. Interest-

ingly, both names start with a /p/ sound and are, at least partially, sound-symbolic (conveying the idea of clumsiness; S. George, personal communication).

22 They do in *Lethal Weapon* 4.

23 The same type of 'fictional' address, where a vocative is employed to refer to someone who is either not physically present in the situation or is a totally invented character, is quite common in the series *Sex and the City*.

24 Cf. further on in the text for an explanation.

25 This kind of vocative seems to be quite frequently resorted to in the series, as it also appears in other episodes of other seasons.

26 Cf. HOMER to BART: Freshen your drink, **pal!** => Ti riempio il bicchiere, **vecchio mio**; HOMER to LISA: There you go, **doll** (*offering her a cigarette*) => Ecco, **bambola**; HOMER to MAGGIE: Look, **buddy**. I don't know where you go but you can't sleep here => Senti, **amico**. Non mi importa dove vai, // ma non puoi dormire qui.

27 "[I]n the place of 'damn' or even 'whoa'. May be considered regional and more so used in the Texas or Louisiana and surrounding areas. Can be cross used with 'meng' esp. when stretched in its use" (<http://www.urbandictionary.com/define.php?term=man&defid=1385405>).

28 The same phenomenon concerns the form of endearment *my dear* in *Sense and Sensibility* (§ 4.7): MRS JENNINGS: **My dear**. Well, you do not waste time, Miss Marianne => Non perdetevi tempo con // la vostra lettera; MRS JENNINGS: Oh **my dear**, he's not the only young man worth having => Non è l'unico giovane // che valga la pena avere.

- Aimeri, Luca / Frasca, Giampiero (2002) *Manuale dei generi cinematografici. Hollywood: dalle origini a oggi*, Torino, UTET.
- Allerton, David J. (1996) Proper names and definite descriptions with the same reference: A pragmatic choice for language users, *Journal of Pragmatics*, 25, pp. 621-633.
- Altman, Rick (1986) "A semantic/syntactic approach to film genre" in Barry Keith Grant (ed.), *Film Genre Reader III*, Austin: University of Texas Press, pp. 27-41.
- Altman, Rick (1996) "Cinema and Genre" in G. Nowell-Smith (ed.), *The Oxford History of World Cinema*, Oxford; Oxford University Press, pp. 294-303.
- Altman, Rick (1999), *Film/Genre*, Milano, Vita e Pensiero (tit. or. *Film/Genre*, first published in 1999 by the British film institute).
- Amendola, Alfonso / Frezza, Gino (2001), "Dramma. Forza e squilibri di un trans-genere", in Gino Frezza (a cura di), pp. 165-195.
- Assis Rosa, Alexandra (2001), "Features of oral and written communication in subtitling", in Yves Gambie, Henrik Gottlieb (eds), (*Multi media translation: concepts, practices and research*, John Benjamins, Amsterdam/Philadelphia, pp. 213-222.
- Bargiela, Francesca / Boz, Corinne / Gokzadze, Lyly / Hamza, Abdurrahman / Mills, Sara / Rukhadze, Nino (2002), "Ethnocentrism, politeness, naming strategies", *Working Papers on the Web - Linguistic Politeness and Context*, vol. 3, <http://www.shu.ac.uk/wpw/politeness/>.
- Bathia, Vijay Kumar (1993), *Analysing Genre. Language Use in professional Settings*, London, Longman.
- Becquemont, Daniel (1996), "Le sous-titrage cinématographique: contraintes, sens, servitudes", in Yves Gambier (ed), *Les transferts linguistiques dans les médias audiovisuels*, Presses Universitaires du Septentrion, Villeneuve d'Ascq, pp. 145-155.
- Biber Douglas et al., 1999, *The Longman Grammar of Spoken and Written English*, London, Longman.
- Brancato, Sergio / Denunzio, Fabrizio / Frezza, Gino (2001), "Comico e commedia. Corpi e mutamenti", in Gino Frezza (a cura di), pp. 65-109.
- Brondeel, Herman (1994), "Teaching subtitling routines", *Meta*, 39, 1, pp. 26-33.
- Brown, Roger / Gilman, Albert (1960), "The pronouns of power and solidarity", in Thomas Sebeok (ed.) *Style in Language*, Cambridge, Mass., MIT Press, pp. 253-276.
- Bruti, Silvia / Perego, Elisa (2005), "Translating the expressive function in subtitles: the case of vocatives", in J.D. Sanderson (ed.), *Research on Translation for Subtitling in Spain and Italy*, Alicante, Publicaciones de la Universidad de Alicante, pp. 27-48.
- Caimi, Annamaria, (2008). "Pedagogical insights for an experimental English language learning course based on subtitling", in Chris Taylor (ed.).
- Campari, Roberto (2002), *Cinema. Generi tecniche autori*, Città di Castello (PG), Mondadori.
- Carroll, Mary (2004), "Subtitling: Changing Standards for New Media?", *The Globalization Insider*, XIII, 3, 3, <http://www.lisa.org/archive/newsletters/2004/3.3/carroll.html>
- Cati, Alice (2002), "Serial di genere? Come la serialità interpreta il genere nel cinema delle origini", in Ruggero Eugeni e Luisella Farinotti (a cura di), pp. 267-271.
- Davies, Eirlys E. (1986), "English Vocatives: A Look into their Function and Form", *Studia Anglica Posnaniensia*, 19, pp. 91-106.
- Dickey, Eleanor (1997), "Forms of address and terms of reference", *Journal of Linguistics*, 33, pp. 255-274.

- Dirks, Tim (2006), *Film Genres*, <http://www.filmsite.org/genres.html>.
- Eugeni, Ruggero / Farinotti, Luisella (2002), "Introduzione", *Comunicazioni Sociali. Rivista di media, spettacolo e studi culturali* [Special Issue: Territori di confine. Contributi per una cartografia dei generi cinematografici, a cura di Eugeni, Ruggero e Farinotti, Luisella], Anno XXIV Num 2, pp. 137-152.
- Frezza, Gino (a cura di) (2001), *Fino all'ultimo film. L'evoluzione dei generi nel cinema*, Roma, Editori Riuniti.
- Georgakopoulou, Yota (2004), "DVD subtitling: a search for the Holy Grail?", paper presented at the International Conference *In So Many Words*, London 6-7 February 2004.
- Grant, Barry K. (ed.) (1986), *Film genre reader III*, Austin, University of Texas Press.
- Halupka-Reštar, Sabina / Radiš Biljana (2003), "Animal names used in addressing people in Serbian", *Journal of Pragmatics*, 35, pp. 1891-1902.
- Hatim, Basil / Mason, Ian (1997), "Politeness in screen translating", in Lawrence Venuti, Mona Baker (eds), *The translation studies reader*, London, Routledge, pp. 430-445.
- Hickey, Raymond (2003), "The German address system: Binary and scalar at once", in A.H. Jucker, I. Taavitsainen (eds), *Diachronic Perspectives on Address Term Systems*, John Benjamins, Amsterdam, pp. 401-425.
- Huddleston, Rodney / Pullum, Geoffrey K. (2002), *The Cambridge Grammar of the English language*, Cambridge, Cambridge University Press.
- Jucker Andreas H. / Taavitsainen, Irma (2003), "Diachronic perspectives on address term systems", in Andreas H. Jucker, Irma Taavitsainen (eds), *Diachronic Perspectives on Address Term Systems*, John Benjamins, Amsterdam, pp. 1-25.
- Kozloff, Sarah (2000), *Overhearing Film Dialogue*, Berkeley, University of California Press.
- Kovačič, Irena (1994), "Relevance as a factor in subtitling reductions", in Cay Dollerup, Anne Loddegaard (eds), *Teaching Translation and Interpreting 2. Insights, Aims, Visions. Papers from the Second Language International Conference, Elsinore, Denmark, 4-6 June 1993*, John Benjamins, Amsterdam/Philadelphia, pp. 245-251.
- Leech, Geoffrey (1999), "The Distribution and Function of Vocatives in American and British English Conversation", in Hilde Hasselgård, Signe Oksefjell (eds), *Out of Corpora: Studies in Honour of Stig Johansson*, Amsterdam/Atlanta, GA: Rodopi, pp. 107-118.
- Maiello, Fabio (2001), "Action Movie. Un genere vissuto pericolosamente" in Gino Frezza (a cura di), pp. 337-348.
- Malavasi, Luca (2002), "Passioni di genere", in Ruggero Eugeni, Luisella Farinotti (a cura di), pp. 283-287.
- Mariotti, Cristina (2002), "Dall'immagine al sottotitolo: strategie d'uso dei materiali audiovisivi nell'insegnamento dell'inglese L2", in Annamaria Caimi (a cura di), *Cinema: Paradiso delle Lingue. I sottotitoli nell'apprendimento linguistico*, (= Rassegna Italiana di Linguistica Applicata, XXXIV, 1-2), pp. 207-213.
- Martiny, Thierry (1996), "Forms of address in French and Dutch: a socio-pragmatic approach", *Language Science*, 18, 3-4, pp. 765-775.
- Menelao, Veronica (2001), "Thriller, poliziesco, spy story. Quando la visione si fa complice", in Gino Frezza (a cura di), pp. 221-233.
- Neale, Stephen (2000), *Genre and Hollywood*, London, Routledge.
- Quirk, Randolph / Greenbaum, Sidney / Leech, Geoffrey / Svartvik, Jan (1985), *A Comprehensive Grammar of the English Language*, London and New York, Longman.
- Salaets, Heidi (2004), "Multilingual DVD cartoons", paper presented at the International Conference *In So Many Words*, London 6-7 February 2004.
- Szarkowska, Agnieszka forthc., "'Why are some forms of address not omitted in subtitling? A study based on three selected soaps broadcast on TV Polonia", in Łukasz Bogucki, Krzysztof Kredens (eds), *Proceedings of Intermedia 2007 Conference* (Lodz, 13-14 April 2007), Bern, Peter Lang.
- Swales, John (1990), *Genre Analysis. English in Academic and Research Settings*, Cambridge, Cambridge University Press.
- Tudor, Andrew (1986), "Genre", in Barry Keith Grant (ed.), *Film Genre Reader III*, Austin, University of Texas Press, pp. 3-11.
- Zipes, Robert (1997), *Happily Ever After. Fairy Tales, Children, and the Culture Industry*, Routledge, New York.
- Zwicky, Arnold M. (1974), "Hey, whatsyourname!", *Papers from the tenth regional meeting of the Chicago Linguistic Society*, pp. 787-801.

WEB REFERENCES

- Disney characters, <http://disney.go.com/vault/archives/characters/bambi/bambi.html>
- Interview with Walt Disney, http://soundtracks.monstersandcritics.com/reviews/article__5809.php
- Names for females in the 1990 U.S. Census, <http://www.thinkbaby-names.com/name/o/Faline>
- <http://www.answers.com/>
- <http://www.urbandictionary.com/>

FILMOGRAPHY

- Bambi, D. Hand, 1942, USA.
- East is East, D. O'Donnel, 1999, UK.

Lethal Weapon 4, R. Donner, 1998,
USA.

Sense and Sensibility, A. Lee, 1996,
USA/UK.

Sex and the City, Season 4, “The Ago-
ny and the Ex-tasy”, M.P. King, 2001
and “I heart NY”, M.P. King 2002,
USA.

Shallow Hal, Farrelly Brothers, 2002,
USA.

Sliding Doors, P. Howitt, 1998, UK.

The Simpsons, “Homer in the night”,
R. Moore, 1989-90; “Homer the
Moe”, J. Kameron, 2001-02, USA.

The Talented Mr Ripley, A. Minghella,
1999, USA-Italy.

Continuity and Variation across Translations: Phraseology in the *Pavia* *Corpus of Film Dialogue*

MARIA FREDDI (University of Pavia)

1. INTRODUCTION

The present paper describes the phraseology of original American and British filmic speech by relating it to issues of translation – dubbing – into Italian. In so doing, it brings together stances from audiovisual translation research, theory of phraseology, as developed by corpus linguists both with reference to monolingual analysis and in relation to the translation process, and corpus-based translation studies.

Recently, stress has been placed on sociocultural contextual factors affecting the translation process and products (see in particular Baker 2004; Laviosa 2004) and especially research from within corpus-based translation studies (hereafter CTS) has advocated that empirical studies of translators' variability should be carried out. This is in order to incorporate contextual parameters such as the actual agents of the translation process, professional and commitments constraints, etc. into the analysis of translated texts, something corpus methodology allows for. Strong criticism of a logocentric view of text corpora stems from translation scholars such as Mona Baker, who want to value translation as a variety in its own right, showing specificities and tendencies which are not present in other varieties of original, non-translated texts (cf. Baker 1995; Olohan 2004). Such specificities, often classified as 'translation universals', are related to contextual, non-linguistic factors. So, for example, the ob-

served tendency of translations towards explicitation practices is interpreted as a result of the translator's need to spell out things more explicitly for the target community.

The same invite to study the policy that regulates the whole translation process and a similar call for the incorporation of context into text corpora also comes from audiovisual translation (hereafter AVT), see in particular Diaz Cintas (2004: 25-29), and seems to share with CTS scholars many concerns, including the very notion of 'translation norms', although the two groups rarely speak to each other. The present paper uses the annotated corpus as a tool to study similarities and patterns of translation behaviour in film translation from this context-sensitive perspective.

2. AIMS AND METHODOLOGY

Within AVT, work on translated film language has shown how routinized translations are so common in film dubbing that it is possible to advance the hypothesis of a 'third norm', which is neither the source nor the target language norm (Pavesi 1994; Herbst 1996; Malinverno 1999). This has also been termed 'dubese', stressing both its autonomy and internal consistency (Freddi 1999). From a target-oriented and translational perspective, dubbese shows significant differences from non-translated language (Alfieri et al. 2003; Bucaria 2008; Bruti and Pavesi 2008), while from a source-oriented perspective, it is described as a series of semantic and structural calques which are found to occur repeatedly, hence the term 'translational routines' (Pavesi 2005: 48), such as, for example, the pairs *l'hai detto-you said it*, *la sai una cosa?-you know something?*, *scordatelo-forget about it*, etc. (for the full list of examples, see Pavesi 2005: 49).

A more recent angle on translated film language, particularly from English into Italian, but also from English into Spanish, has shown how distribution of certain expressions varies when dubbed film and television dialogues are compared with corpora of natural spoken discourse chosen as reference corpora (Romero Fresco 2006, 2009; Pavesi 2008; Bruti and Pavesi 2008), or with data found in reference grammars and dictionaries (Alfieri et al. 2003). This occurs at a time when translation scholars in general, not just re audiovisual translation, tend to focus on translated language as a variety in its own right to be compared with non-translated language, though not necessarily thought of as ancillary to it or related to the source language (Kenny 2005). This perspective is paralleled by the comparisons of original film and television dialogue with spontaneous conversation, again highlighting different distributional profiles (Quaglio 2008; Romero Fresco 2009) with divergence explained by Quaglio 2008 in terms of functional (re, for example, the role of vagueness and emotional language) and situational differences (limited range of conversation topics and settings).

This AVT perspective on the formulaic nature of film translation matches well with theory of phraseology, particularly with regard to Sinclair's 'idiom principle', or the 'single choices' the language user makes when producing and processing a text (Sinclair 1991: 110).

Within monolingual corpus linguistics, various scholars have shown patterns that come out as a result of repeated usage. In particular, lexical bundles, as defined by Biber and Conrad (1999: 183), are recurrent lexical sequences 'of three or more words that show a statistical tendency to co-occur', e.g., in conversation, *I don't know what, I don't know how, well I don't know, I don't want to, I don't think so*, etc. As is clear from the examples, lexical bundles do not require structural completion and become a reliable indicator of genre variation (Biber and Conrad 1999; Biber et al. 1999) as different registers or genres are characterised by different distributions of clusters. The extreme example of this tendency words have to cluster together is idiom in the traditional sense where frequency is accompanied by lexico-grammatical fixedness and semantic non-compositionality (Moon 1998: 7-8). A broad concept of idiomaticity as a continuum together with a pragmatic classification of idioms allow us to move from frequent co-occurrence to pure idioms and to study phraseology in film dialogues as strategies for fostering interaction as well as markers of textual organization (Moon 1998: 17).

Also of help is the notion of 'functional equivalence' coming from bilingual contrastive corpus analyses (Tognini Bonelli 2001, 2002), whereby two forms chosen as translational equivalents are such if their respective environments are considered and found to be equivalent as well, ensuing the recognition that the phraseological unit reunites lexical and grammatical patterns with a specific semantic dimension and a specific pragmatic function.

Therefore, this kind of analysis can be carried out by means of the *Pavia Corpus of Film Dialogue* (see Freddi and Pavesi in press), an annotated corpus that not only aligns the texts sampled, but also combines them with the contextual information that has been encoded in it and consequently establishes multi-level relations between data which are apparently distant, in an attempt to provide the 'situated explanation' Baker (2004: 183) calls for.

3. THE PHRASEOLOGY OF FILMIC SPEECH

The perspective of the present study is both monolingual – is there a phraseology of filmic speech? – and contrastive, that specific to AVT – is the repetitiveness and fixedness of certain translation solutions the direct consequence of the formulaicity of original film dialogue? The approach therefore exploits the parallel corpus in both directions, shunting between original texts and their translations examined in parallel and translated texts as such (the translational component of the parallel film corpus) in the framework of CTS.

In order to identify phrases that might be typical of film dialogue, both original and translated, the interrogation of the raw corpus has been combined with the annotated corpus search. More in detail, to approach the first perspective, the overall corpus frequency wordlist is looked at to hypothesise frequent words which might belong in frequent phrases. Then, concordances are generated through the database (the annotated corpus) query and analysed, allowing for further distinctions to be made between the American and British sub-corpora and for searches of single films as well as across films.

Once a phraseology of original film dialogue is identified, the corresponding translations are examined to see whether an equivalent set of recurring patterns are observable, what their specific pragmatic functions are and how the same functions are expressed in original and translated texts. Finally, these findings are compared with the instances of dubbese found in the literature (cf. Pavesi 2005; Taylor 2000, 2006, his ‘framing moves’ and predictability of filmic speech, i.e. ‘predictability of textual occurrences and frequencies as associated with particular scenes and scene types’), which are the result of a translational perspective. Analysis of data at this stage is aimed at checking whether the same kinds of patterns are present in the corpus. In the following paragraphs the findings are discussed in relation to the methodology followed.

3.1 FROM FREQUENT (LEXICAL) WORDS TO LEXICAL BUNDLES TO PRAGMATIC IDIOMS

With more than a thousand occurrences, *what* is a high frequency word in the overall corpus. Coming after the two top frequencies corresponding to the second and first person singular pronouns *you* and *I*, it ranks twelfth right after other usually very frequent grammatical words. It is evenly distributed across the British and the American film sub-corpora and found to occur in each film in the corpus.

Starting from frequency counts of individual words such as this, information on frequent collocates can also be retrieved through a software for automatic text processing (*Wordsmith Tools*). An interesting kind of output is the list of 3-word clusters (bundles) calculated within a span of 5 words to the left to 5 to the right from node word *what*, with minimum frequency set at 15, of which the top three 3-word clusters are as follows:

Cluster	Freq.
WHAT DO YOU	93
WHAT ARE YOU	71
YOU KNOW WHAT	48

Table 1 – *Wordsmith Tools’ list of 3-word clusters (both left and right horizons)*

The first two both present subject verb inversion typical of the structure of questions, with *what* functioning as interrogative pronoun.

The next most frequent one, *you know what*, is interesting as it includes the fixed phrase *You know what?* functioning as turn launcher. This is also listed by Romero Fresco 2006 among the set phrases typifying the original dialogue of the television series *Friends* and matches Pavesi’s *You know something?* whose translation she records among the calques typical of dubbed Italian (Pavesi 2005: 49). In the whole corpus the string *You know what?* occurs 15 times, almost all of which distributed across 2 films, both American ones, namely *Erin Brockovich* and *Crash*. There is one single occurrence in one British English film,

Mike Leigh's *Secrets and Lies*, where the sequence combines with *Listen* in initial position, used as attention-getting device, and is translated as *Te lo dico io, sai cosa facciamo?*

The other occurrences in the corpus are variously translated as *Smettila* (perhaps a more apt equivalent for *stop it than you know what?*) / *Sai che ti dico* (3 times) / *Facciamo così / Sa(i) una cosa?* and *La sai una cosa?* (3 times). Among the renditions there are also 2 zero correspondences and hapaxes such as *Hey senti okay / Senta / Ecco, visto / Beh* and *lo sai?* pointing backwards, thus annulling the projecting cataphoric function of *you know what?*, by reason of which it has been labelled 'utterance launcher' (Biber et al. 1999).

It is interesting to notice that the variety of translation solutions includes both that which according to Pavesi 2005 sounds like natural Italian, namely *Sai che ti dico?*, and its more literal counterpart *La sai una cosa?*

A similar query was generated with min. frequency set at 5 and horizon 0 to the left to 5 to the right from node word *what*, thus considering only the clustering taking place in the right co-text:

Cluster	Freq.
WHAT DO YOU	91
WHAT ARE YOU	70
ARE YOU DOING	30
WHAT IS IT	25
DO YOU THINK	24
WHAT THE FUCK	21
WHAT DID YOU	21
DO YOU MEAN	21
WHAT THE HELL	16
DO YOU DO	12
WHAT KIND OF	11
DID YOU SAY	10
DO YOU WANT	10
WHAT WAS THAT	9
WHAT IT IS	8
WHAT WOULD YOU	7
WHAT I WAS	7
WHAT HAPPENED TO	7
WHAT AM I	7
ARE YOU TRYING	6
THE FUCK ARE	6
WHAT IT WAS	6
ARE YOU TALKING	6
WHAT YOU WANT	6
WHAT YOU THINK	6

ARE YOU GONNA	6
WHAT CAN I	5
WHAT I WOULD	5
WHAT THE HELL'S	5
WHAT DID HE	5
WHAT I SAID	5
WHAT YOU MEAN	5
WHAT IS THAT	5
WHAT DO I	5
WHAT IS THE	5
WHAT IF I	5
WHAT ARE THEY	5
WHAT ABOUT YOU	5

Table 2 – *Wordsmith Tools'* list of 3-word clusters (right horizon only)

All the clusters above are characterised by structural incompleteness, that is, they do not fulfil a syntactic function, rather they are simply the result of statistical association and even distribution across the texts in the corpus.

However, a first look at the list shows that the overall number of occurrences of *what do you* (91) partially overlaps with occurrences of *do you think*, *do you mean*, *do you do* and *do you want* (highlighted in bold). Moreover, as a consequence of the transcription process, orthographic variants such as *d'you* / *do you* should also be taken into account when searching the corpus, so numbers might become even bigger.

In the same way, *what are you* (70) is to be seen in combination with *are you doing* (30), *are you trying* and *are you talking*. We thus get a series of *What*-questions *What are you doing?*, *What are you trying to say? / to do?* and *What are you talking about?* as frequent patterns of film dialogue.

On the basis of these results, some database queries were set up in order to be able to relate the occurrence of these phrases to the parameters encoded in the corpus (e.g. film, line number, or cue, character, language, translator-adaptor). Results for the phrase *what do you mean* are shown below in Table 3:

Table 3 – *what do you mean*

FILM	ID_CUES	LANGUAGE	CHARACTER	TEXT	ADAPTER
Ae Fond Kiss	661	English	ROISIN	What do you mean? Why not? Did you talk to your folks?	
Ae Fond Kiss	661	Italian	ROISIN	<u>Cosa? Perché no ? Hai parlato con i tuoi?</u>	Depaolis Federica
Billy Elliot	441	English	BILLY	What do you mean?	
Billy Elliot	441	Italian	BILLY	<u>E perché ?</u>	Cosolo Carlo
Crash	493	English	CAMERON	What do you mean?	
Crash	493	Italian	CAMERON	<u>Che vuoi dire?</u>	Ottoni Filippo
Dead Man Walking	58	English	SISTER HELEN	What do you mean?	
Dead Man Walking	58	Italian	SISTER HELEN	<u>Che vuoi dire?</u>	Bertini Lorena
Dead Man Walking	626	English	SISTER HELEN	A lot of commotion for nothing. I... What do you mean you didn't know?	
Dead Man Walking	626	Italian	SISTER HELEN	<u>Vuoi dire che non lo sapevi?</u>	Bertini Lorena
Dead Man Walking	798	English	MATTHEW PONCELET	What do you mean?	
Dead Man Walking	798	Italian	MATTHEW PONCELET	<u>Che vuoi dire?</u>	Bertini Lorena
Erin Brockovich	228	English	GEORGE	What do you mean I cheat?	
Erin Brockovich	228	Italian	GEORGE	Eh eh... <u>Come sarebbe imbroglio?</u>	Mete Marco
Erin Brockovich	396	English	GEORGE	What? What do you mean you got fired? You're working so hard.	
Erin Brockovich	396	Italian	GEORGE	<u>Cosa? Come sarebbe licenziato? Non fai altro che lavorare.</u>	Mete Marco
Erin Brockovic	736	English	ERIN	What do you mean?	
Erin Brockovich	736	Italian	ERIN	<u>Che vuoi dire?</u>	Mete Marco

FILM	ID_CUES	LANGUAGE	CHARACTER	TEXT	ADAPTER
Finding Forrester	95	English	TERREL	What do you mean “nothing”? This getting in the way of your plans or something?	
Finding Forrester	95	Italian	TERREL	Come sarebbe niente? È, è una cosa che rovina i tuoi piani per caso?	Caporello Elettra
Finding Forrester	138	English	JAMAL'S MOTHER	Don't know? What do you mean you don't know?	
Finding Forrester	138	Italian	JAMAL'S MOTHER	Non lo so? <u>Cosa vorrebbe dire</u> questo non lo so?	Caporello Elettra
Finding Forrester	412	English	JAMAL	What do you mean “that's a foul”? I had the spot.	
Finding Forrester	412	Italian	JAMAL	<u>Ma quale</u> fallo? Il posto era mio.	Caporello Elettra
Notting Hill	91	English	ANNA	Alright. Well, what do you mean “just over the street”? Give it to me in yards.	
Notting Hill	91	Italian	ANNA	D'accordo, bé, <u>cosa intende per</u> “altro lato della strada”? Quanti metri dista?	Vairano Francesco
Ocean's Eleven	1080	English	MR BENEDICT	What do you mean there's no money in the bags?	
Ocean's Eleven	1080	Italian	MR BENEDICT	<u>Come sarebbe</u> non c'erano soldi nelle borse?	Mete Marco
One Hour Photo	147	English	NINA YORKIN	What do you mean you feel bad?	
One Hour Photo	147	Italian	NINA YORKIN	<u>In che senso</u> sei preoccupato?	Valli Carlo

FILM	ID_CUES	LANGUAGE	CHARACTER	TEXT	ADAPTER
Secrets & Lies	572	English	MONICA	What do you mean ?	
Secrets & Lies	572	Italian	MONICA	<u>In che senso?</u>	Bucciarelli Elisabetta
Secrets & Lies	647	English	ROXANNE	What do you mean?	
Secrets & Lies	647	Italian	ROXANNE	<u>Che vuoi dire?</u>	Bucciarelli Elisabetta
Secrets & Lies	1140	English	CYNTHIA	What do you mean?	
Secrets & Lies	1140	Italian	CYNTHIA	<u>Beh che dici?</u>	Bucciarelli Elisabetta
Sliding Doors	172	English	HELEN 2	Anna! I'm over him. What do you mean I'm not? How do you know I'm not?	
Sliding Doors	172	Italian	HELEN 2	Anna, con lui ho chiuso. <u>Perché dici</u> il contrario? Come fai a saperlo?	Vairano Francesco
Sliding Doors	365	English	HELEN 2	What do you mean you want to see me? Jesus Christ ,Gerry! What.. How did you get in?	
Sliding Doors	365	Italian	HELEN 2	<u>Come sarebbe a dire</u> che volevi vedermi? Santo Dio! Gerry, cosa... come sei entrato?	Vairano Francesco
Sliding Doors	418	English	LYDIA	What do you mean "what am I trying to do"? I think I've broken my toe.	
Sliding Doors	418	Italian	LYDIA	<u>Come sarebbe</u> "qual è il mio scopo"? Si sarà fratturato l'alluce.	Vairano Francesco

The following observations can be made: the string *what do you mean* occurs in 11 out of the 12 films in the corpus, unevenly distributed across the films but always filling the same function of taking up the interlocutor's words to rebuke them. As far as the Italian version is concerned, there seems to be no preferred rendering, translations oscillating between *che vuoi dire?* and *come sarebbe?* also depending on the translator-adaptor: the same translator, Marco Mete, chooses both *come sarebbe* and *che vuoi dire* within the same film; *come sarebbe* is chosen by another translator, Francesco Vairano in another film. More peripheral instances such as *in che senso?* together with *cosa intende per...* make it difficult to see any regularity in the translation, thus we can only tentatively hypothesise a tendency towards functional equivalence between *what do you mean?* and *che vuoi dire? / come sarebbe?*

What is preferably translated as *che* than *cosa* (cf. also the fixed equivalence of the first most frequent phrase *What do you think? - Che ne pensi?*). Rather than a feature of dubbese, *cosa* being the unmarked expression in natural spoken Italian in the north, *che* is unmarked in Southern Italy (see also Pavesi 2005: 39 on colloquial and regional forms). Indeed, most translator-adaptors are based in Rome. In another query, analysis of *What are you doing?* gave *Che stai facendo?* far outnumbering other less frequent solutions like *Cosa fai? / Che fai / etc.* without the continuous aspect. Interestingly, when the expression is followed by the place adverbial as in *What are you doing here?* (6 occ.), the continuous aspect disappears and becomes *Che ci fai qui?* (3) / *Che ci fai tu qui?* / *Come mai sei qui?* / *Che cosa ci fai qui?*

Some of the occurrences contain a swearword stressing the action-quality of the situation, *What the fuck are you doing?*, or another mark of the colloquial register as in *What are you doin' man?* with the contracted form and the Vocative *man*, which in the Italian dubbing are rendered by the patterned formulations *Che credi di fare?*, the allocution having totally disappeared, or *Dove accidenti vai?* where the loss of the Vocative *man* is compensated for by the expletive *accidenti*. Sometimes, the tension accrued by the use of the Vocative in the source-text, *Tony, what are you doing?*, is made explicit in the target-text by means of the weak connective together with the verb of volition, as in *Tony, ma che vuoi fare?*

Analogously, the expression *What are you talking about?* (12 occ. of which 4 are transcribed as *What're you talking about?*) regularly occurs in confrontational scenes in each film, as is emphasised by the one occurrence, from *Erin Brockovich*, *What the fuck are you talking about?* with the swearword in between the interrogative pronoun and the finite element. Its function is to carry on the plot by forcing the interlocutor to produce the missing bit of information. The following translation solutions were found for this set phrase: *Ma che cosa stai dicendo? / (Adesso) di che cosa stai parlando? / Di che stai parlando? / Ma di che cazzo stai parlando? / Che cosa stai dicendo? / Ø*, while the 4 instances of the contracted form are translated as *Che stai dicendo? / Ma cosa stai dicendo?* Again, the continuous aspect together with the weak connective *ma* seem to be recurrent features (cf. Pavesi 2005: 49).

A similar fixed discursal function can be seen in the occurrences of *What are you trying to do?* (4) and *What are you trying to say?* (2). They are both stock expressions in that they help carry the action forward, with the interlocutor usually being prompted an explanation or summary of some events that will be seen later on in another scene, or else just talked about in the next turn.

As far as their repetitiveness in the translation goes, the literal **che stai cercando di fare?* is never used. Rather, two quite unusual renderings *mi dici qual è il tuo scopo?* and *E allora, qual è il tuo scopo, eh?* come from the same film and turn. Their fixedness being internal to a single speaker's turn, they can hardly be significant in quantitative terms and only indicative of a very specific stylistic choice (of translator Francesco Vairano). The remaining two occurrences are instances of a specific construct, whereby a bare infinitive lexicalizes the action verb *do*: *What are you trying to do, just rub it in my face?*-*Così pensi di sbattermi in faccia la realtà?*; *What are you trying to do, make him a fucking scab for the rest of his life?*-*Stai cercando di farne un crumiro del cazzo per tutta la vita?*

On the other hand, there is a perfect correspondence between the repetition of *What are you trying to say?* in the source-text and its translation as *che stai cercando di dire?* in the target-text. The Italian corpus also has an instance of *è questo che stai cercando di dirmi?* resulting from the pseudo-cleft structure *Is that what you're trying to say?* More data is needed, however, to prove the hypothesis of the expression being fixed and a calque of the original.

If we go back to the 3-word clusters list above, the fourth most frequent one is *what is it* for which another corpus query was made. The results are best displayed in Table 4 where the English version has been kept separate from the Italian to help visualize any regularities in the translation.

Table 4 – *what is it*

FILM	ID_CUES	LANGUAGE	CHARACTER	TEXT	ADAPTER
Ae Fond Kiss	135	English	MR KHAN	What is it?	
Ae Fond Kiss	634	English	CASIM	What is it?	
Ae Fond Kiss	676	English	MR KHAN	What is it?	
Ae Fond Kiss	808	English	ROISIN	Jamin, thanks very much. What is it?	
Ae Fond Kiss	986	English	ROISIN	What is it?	
Billy Elliot	304	English	MRS WILKINSON	What is it?	
Dead Man Walking	46	English	FARLEY	Do you know what you're getting into? So what is it , Sister? Morbid fascination? Bleeding heart sympathy?	
Dead Man Walking	600	English	UNIDENTIFIED	What is it? What happened?	
Dead Man Walking	850	English	SISTER HELEN	What, Mat? What is it?	

FILM	ID_CUES	LANGUAGE	CHARACTER	TEXT	ADAPTER
Erin Brockovich	2	English	ERIN	No. I have kids. Learned a lot right there. Yeah, I mean, I've seen nurses give my son throat cultures, and, what is it?	
Erin Brockovich	831	English	ERIN	There's no baby in here. What is it?	
Notting Hill	741	English	ANNA	What is it about men and nudity, huh?	
Notting Hill	757	English	ANNA	What? What is it?	
Notting Hill	807	English	WILLIAM	Blimey, what is it?	
Ocean's Eleven	358	English	LINUS	What is it?	
One Hour Photo	495	English	BILL OWENS	What is it , Yoshi? You can't just leave your station unmanned down there.	
Secrets & Lies	5	English	MONICA	What is it?	
Secrets & Lies	314	English	HORTENSE	What is it?	
Secrets & Lies	689	English	CYNTHIA	Yes, what is it you want, darling? Hello? Did you want Roxanne? She's gone out.	
Secrets & Lies	704	English	CYNTHIA	Listen, darling, what is it you want?	
Secrets & Lies	770	English	CYNTHIA	What is it , sweetheart?	
Secrets & Lies	788	English	CYNTHIA	Why? What is it?	
Sliding Doors	80	English	ANNA	What is it you've done, anyway? It can't be that nice if she's walked out on you.	

FILM	ID_CUES	LANGUAGE	CHARACTER	TEXT	ADAPTER
Sliding Doors	547	English	ANNA	She saw you, James, at the hospital. With your wife. You know, the one with the wedding ring. What is it with you bloody men?	
Ae Fond Kiss	135	Italian	MR KHAN	<u>Che c'è?</u>	Depaolis Federica
Ae Fond Kiss	634	Italian	CASIM	<u>Ma cosa?</u>	Depaolis Federica
Ae Fond Kiss	676	Italian	MR KHAN	<u>Che c'è?</u>	Depaolis Federica
Ae Fond Kiss	808	Italian	ROISIN	Grazie, <u>ma che cos'è?</u>	Depaolis Federica
Ae Fond Kiss	986	Italian	ROISIN	<u>Di che si tratta?</u>	Depaolis Federica
Billy Elliot	304	Italian	MRS WILKINSON	<u>Che cos'è?</u>	Cosolo Carlo
Dead Man Walking	46	Italian	FARLEY	Lo sa a cosa sta andando incontro? <u>Che cosa la spinge sorella?</u> Un'attrazione morbosa? Pietà per le anime dannate?	Bertini Lorena
Dead Man Walking	600	Italian	UNIDENTIFIED	<u>Che c'è, che è successo?</u>	Bertini Lorena
Dead Man Walking	850	Italian	SISTER HELEN	Vuoi parlare. <u>Che cosa c'è?</u>	Bertini Lorena
Erin Brockovich	2	Italian	ERIN	No. Ma ho dei figli. E ho imparato parecchio. Sì, cioè, ho visto le infermiere fare i tamponi per la gola a mio figlio e, <u>che ci vuole...</u>	Mete Marco
Erin Brockovich	831	Italian	ERIN	Non c'è un bambino qui dentro. <u>Che cos'è?</u>	Mete Marco
Notting Hill	741	Italian	ANNA	Perché gli uomini sono fissati con il nudo, huh?	Vairano Francesco
Notting Hill	757	Italian	ANNA	Chi è? <u>Cosa c'è?</u>	Vairano Francesco
Notting Hill	807	Italian	WILLIAM	Perbacco! <u>Cos'è?</u>	Vairano Francesco

FILM	ID_CUES	LANGUAGE	CHARACTER	TEXT	ADAPTER
Ocean's Eleven	358	Italian	LINUS	<u>Che cos'è?</u>	Mete Marco
One Hour Photo	495	Italian	BILL OWENS	Che c'è, Yoshi? Lo sai che non puoi lasciare il laboratorio.	Valli Carlo
Secrets & Lies	5	Italian	MONICA	<u>Che c'è?</u>	Bucciarrelli Elisabetta
Secrets & Lies	314	Italian	HORTENSE	<u>Cos'è?</u>	Bucciarrelli Elisabetta
Secrets & Lies	689	Italian	CYNTHIA	Sì, <u>che c'è gioia?..</u> Pronto?..Volevi Roxanne? È uscita.	Bucciarrelli Elisabetta
Secrets & Lies	704	Italian	CYNTHIA	<u>Che cosa vuoi?</u>	Bucciarrelli Elisabetta
Secrets & Lies	770	Italian	CYNTHIA	Sì, <u>che c'è?</u> Dica.	Bucciarrelli Elisabetta
Secrets & Lies	788	Italian	CYNTHIA	Perché, <u>che c'è?</u>	Bucciarrelli Elisabetta
Sliding Doors	80	Italian	ANNA	<u>Che cosa ha fatto</u> si può sapere? Niente di carino se ti ha piantato.	Vairano Francesco
Sliding Doors	547	Italian	ANNA	Ti ha visto, James, all'ospedale con tua moglie. Sai, quella con la fede al dito. <u>Ma che vi prende</u> a voi uomini. Non siamo qui per...	Vairano Francesco

This cluster contains the question *What is it?* meaning 'is there any problem?' or simply replacing the whole-turn *What?* translated as *Che c'è?*. Sometimes it is followed by the preposition *with* as in *What is it with you?* with a function akin to *What (i)'s the matter with you?* and *What (i)'s wrong with you?*

These expressions, in their turn, are present in the corpus and translated as *Ma che cosa hai?* / *Cosa ti succede?* / *Che succede?* / *Ma che ti/vi prende?*, occasionally with an intervening swearword as in *What the hell's wrong with you?* translated as

Ma che diavolo ti prende?. Notice that *what the hell* also appears among the next most frequent 3-word clusters with 16 occurrences in the overall corpus (although not in each film, and yet in all American ones), see Table 2 above.

The Italian string *che ti prende* can also be the result of the English *What is / What's going on?* (and the cursing *What the fuck is going on? / What the hell's going on?*). These are more commonly rendered as *Che (ti) succede* than *che sta succedendo* (apart from one *Che diavolo sta succedendo?*), with the plain present tense in lieu of the progressive seen before. However, the frequent use of the weak connector *ma* in initial position as in *What's going on with you? - Ma che ti prende adesso?* – a typical feature of dubbed Italian (Malinverno 1999; Pavesi 2005, 2008) – adds to the register-specificity of a phrase characterised by the carry-the-action-forward function which has already been observed with reference to many of these set phrases.

Finally, it is interesting to notice that Table 4 also exemplifies instances of the split structure *what is it you've done?* and *what is it you want?*. The translation, however, neutralizes the emphasis of the English construct by choosing the unmarked word order *che cosa hai fatto? / che c'è? / che cosa vuoi?* in a trend contrary to what has been observed for marked word orders in dubbed texts (Pavesi 2005, 2008).

It is clear how analysis of the clusters in the list and their collocates can proceed along the same lines to get the complete mapping of the phraseology of filmic speech. What emerges is that the idiomaticity of these expressions does not entail semantic opacity, rather it is pragmatic and register-specific to the extent that it works at the communicative and narrative levels.

3.2 TOP-DOWN INVESTIGATION OF PHRASES IN FILM DIALOGUE

Another way of looking at phraseology in film dialogue is by resorting to the inventory of fixed or semi-fixed phrases that are known in the literature to be used in natural oral conversation to see whether they are present in the corpus and to what extent. This approach is closer to a top-down rather than the bottom-up approach adopted so far.

Thus, the following expressions from Pavesi (2005: 49) were searched for: *Forget it!* (4 occ.) translated as *scordatelo* (once), *non ci raccontiamo scemenze / non importa / neanche per sogno* together with *(you can) forget about... ti puoi scordare / ti scordi pure / lascia perdere*. Again, it is interesting that the double clitic construction, said to be calqued from English, only once is the result of the expression *Forget it! Scordatelo*, however, counts another 2 occurrences in the corpus which are resorted to to translate the two elliptical answers *We're not* and *I can't*.

Conversely, search for *lascia perdere* gave 9 occurrences which, if cross-checked with the instances of *forget about*, only once translates it, while it is rather the result of various formulas in the original text, particularly *leave it (alone)* (4), *doesn't matter, it's fine, I don't care what* and is once inserted anew by the translator in lieu of a turn uttered in Punjabi in the original version. In the same vein, the Italian string *lascia stare* was found to occur 8 times translating very different realizations including *It's okay, My sister's okay! - Lascia stare mia sorella!*, *Leave her alone - Lascia stare tua sorella*, *Never mind tea - Lascia stare il tè adesso* and *don't touch your nose - lascia stare il naso*.

Also from Pavesi 2005, node word *kidding* was searched for to get all patterns of *Are you kidding (me)?* (4), *You're kidding?!* (4), *You're fucking kidding me?* (1) and *You've got to be kidding me* (1). The 10 instances of this expression come from just 7 out of the 12 films in the corpus, translated as *sta(i) scherzando?* (5 times), twice as *scherzi?* and once as *vuoi scherzare?*. The remaining two instances *You're fucking kidding me?* and *Are you kidding me?* both come from the same quarrel scene and are rendered as *Mi stai prendendo per il culo?* and *Mi stai prendendo in giro?*, uttered by the same character addressing the ambiguous main protagonist of *One Hour Photo* in both cases.

Other expressions known to characterise spontaneous conversation (Taylor 2000; Quaglio and Biber 2006: 704) were found, among which the utterance launchers and attention getting devices *I'm telling you*, *Tell you what*, *Guess what*, and *Thing is*.

The expression *I'm telling you* is not very frequent (4 occ. in overall corpus) and gets even more diluted in the translation, cf. *E adesso ti dico che / Ma te lo ripeto / Dammi retta / Vedrai* (notice, however, the reiterated turn-initial use of the pragmatic connectives *e* and *ma*).

The cognate expression *Tell you what* together with its variants *I'll tell you what* and *I tell you what*, also not very frequent (5 occ.), is translated as *Sai che facciamo? / Facciamo così* (2) / *Sai che ti dico?* and *Va bene*. A similar function is shared by *Guess what?* (3 occ.) invariably translated as *indovina un po'?* Similarly, *Thing is* (1) plus *The thing is* (5) is almost invariably translated as *E' che / La cosa è ... che / Il fatto è che*.

Unquestionably contributing to the simulation of interactivity (cf. Quaglio and Biber 2006: 716), the sparse occurrence of these phrases throughout the corpus triggers the usage of fixed forms in the target-texts, particularly pragmatic connectives and split structures, which have been shown to be a translation strategy typical of dubbed texts (Pavesi 2005, 2008; Pavesi and Perego 2008: 127). Once more is the idiomaticity of these expressions linked to their pragmatic function in the original product.

4. CONCLUSIONS

The analysis thus carried out allows for some conclusions to be drawn. First, a frequency-based description of contemporary American and British film dialogue brings to the fore holistic chunks of speech with specific functions. The functions are primarily pragmatic in that the phrases thus identified help achieve a move in dialogue. This move can be glossed as speaker A challenging speaker B usually by means of a question whose tenor is confrontational. This is why routines can be said to also have a fundamental diegetic function, in that they help advance the situations and events occurring within the fictional world of films. In fact, unlike natural conversation, film dialogues are not only linked to the situational context and the moment of uttering, they also link up to the narrative dimension of the diegesis, so that formulas like *What do you mean?*, *What're you trying to say?*, etc. have a function similar to that of a narrator whispering to the audience "hey audience, watch out as something is going to happen and character X is going

to tell us about it". The phraseology identified is therefore part of the narrative space of film dialogue.

Because of this combination of pragmatic – internal to the dialogue as anchored to the communicative situation – and diegetic – internal to the fictional world narrated by film – functions, phrases affect the kinds of interactions taking place on screen and acquire conventional significance for the type of text examined, i.e. film dialogue, thus contributing to the making of register-specificity. The second question of what the implication of this clustering in the original texts is for film translation, dubbing in particular, has been answered, showing, however, that in some instances translational routines leave place for variation and creativity on the part of the translators recreating a similar discourse function and register characterisation of the original. When translating-adapting, some of these formulas retain their formulaicity and communicative function, e.g. through repeated usage of weak connectives, progressive tense and split structures, others are diluted through variation across translations. We could then ask if such fixedness is a feature of 'filmese', as in Taylor 2006, thus reflected in dubbese, or else it is mirrored in original Italian film dialogue. To this purpose, a comparable corpus of original Italian films sampled according to the same criteria should be compiled to allow for quantitative comparisons to be made.

From a translational viewpoint, the paper has tried to show how phraseology, by reuniting the lexical and the syntactic levels, is the preferred carrier of the register-specificity of the original product and therefore also the place where translators' variability can be best observed. This is why the annotated parallel corpus and the comparative approach allow to better investigate the relation between original and translated film dialogue thought of as a particular text-type, thus showing a possible outcome of the fruitful cross-fertilization of translation studies and corpus linguistics.

- Alfieri, G., S. Contarino, D. Motta 2003, "Interferenze fraseologiche nel doppiaggio televisivo: l'italiano di E.R. e di Beautiful" in A. Sullam Calimani (a cura di) *Italiano e inglese a confronto. Atti del convegno "Italiano e inglese a confronto: problemi di interferenza linguistica"*, Venezia, 12-13 aprile 2002, 127-149
- Baker, M. 1995, "Corpora in translation studies: An overview and some suggestions for future research", *Target*, 7 (2), 223-243.
- Baker, M. 2004 "A corpus-based view of similarity and difference in translation" *International Journal of Corpus Linguistics*, 9:2, 167-193.
- Biber, D. and S. Conrad 1999, "Lexical bundles in conversation and academic prose", in H. Hasselgard and S. Oksefjell (eds.), *Out of Corpora. Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, 181-190.
- Biber, D., S. Johansson, G. Leech, S. Conrad and E. Finegan 1999, *Longman Grammar of Spoken and Written English*. London: Longman.
- Bruti, S. and M. Pavesi 2008, "Interjections in translated Italian: Looking for traces of dubbed language", in A. Martelli and V. Pulcini (eds.), *Investigating English with Corpora. Studies in Honour of Maria Teresa Prat*. Monza: Polimetrica, 207-221.
- Bucaria, C. 2008, "Acceptance of the norm or suspension of disbelief? The case of formulaic language in dubbese", in D. Chiaro, C. Heiss and C. Bucaria (eds.), *Between Text and Image. Updating Research in Screen Translation*. Amsterdam/Philadelphia: John Benjamins, 149-163.
- Diaz Cintas, J. 2004, "In search of a theoretical framework for the study of audiovisual translation" in Orero (ed.), *Topics in Audiovisual Translation*, Amsterdam/Philadelphia: John Benjamins, 21-34.
- Freddi, M. 1999 "Dialogue analysis and multimedia translation" in B. Naumann (ed.), *Dialogue Analysis and the Mass Media. Beiträge zur Dialogforschung*, 20 *Proceedings of the I.A.D.A International Conference*. Tübingen: Niemeyer, 149-158.
- Freddi, M. and M. Pavesi (eds.) in press, *Analysing Audiovisual Dialogue. Linguistic and Translational Insights*. Bologna: Clueb.
- Herbst, T. 1996 "Why dubbing is impossible", in C. Heiss, R.M. Bollettieri Bosinelli (a cura di), *Traduzione multimediale per il cinema, la televisione, la scena*. Bologna: Clueb, 97-115.
- Kenny, D. 2005, "Parallel corpora and translation studies: old questions, new perspectives? Reporting that in GEPOLC: a case study" in G. Barnbrook, P. Danielsson, M. Mahlberg (eds.), *Meaningful Texts: The extraction of semantic information from monolingual and multilingual corpora*. London: Continuum, 155-165.
- Laviosa, S. 2004, "Corpus-based translation studies: Where does it come from? Where is it going?", *Language Matters*, 35 (1), 6-23.
- Malinverno, A. 1999, "La resa delle varietà non-standard dal film americano all'italiano del doppiaggio", in M. Vedovelli (a cura di), *Indagini sociolinguistiche nella scuola e nella società italiana in evoluzione*, Milano: Franco Angeli, 51-72.
- Moon, R. 1998, *Fixed Expressions and Idioms in English. A corpus-based approach*. Oxford: Clarendon.
- Olohan, M. 2004, *Introducing Corpora in Translation Studies*. London: Routledge.
- Pavesi, M. 1994, "Osservazioni sulla (socio)linguistica del doppiaggio", in R. Baccolini, R.M. Bollettieri Bosinelli, L. Gavioli (a cura di), *Il doppiaggio. Trasposizioni linguistiche e culturali*. Bologna: Clueb, 129-142.
- Pavesi, M. 2005, *La traduzione filmica. Aspetti del parlato doppiato dall'inglese all'italiano*. Roma: Carocci.
- Pavesi, M. 2008, "Spoken language in film dubbing: Target language norms, interference and transla-

- tional routines”, in D. Chiaro, C. Heiss and C. Bucaria (eds.), *Between Text and Image. Updating Research in Screen Translation*. Amsterdam/Philadelphia: John Benjamins, 79-99.
- Pavesi, M. and E. Perego 2008, “Il dialoghista in Italia: indagine sociologica e norme linguistiche”, in A. Baldry, M. Pavesi, C. Taylor Torsello and C. Taylor (eds.), *From Didactas to Ecolingua. An Ongoing Research Project on Translation and Corpus Linguistics*, Trieste: EUT, 111-139.
- Quaglio, P. and D. Biber 2006, “The grammar of conversation” in B. Aarts and A. McMahan (eds.), *The Handbook of English Linguistics*. Oxford: Blackwells, 692-723.
- Quaglio, P. 2008, “Television dialogue and natural conversation: Linguistic similarities and functional differences”, in A. Ädel and R. Reppen, (eds.), *Corpora and Discourse: The challenges of different settings*, Amsterdam/Philadelphia: John Benjamins, 198-210.
- Romero Fresco, P. 2006, “The Spanish dubbese: a case of (un)idiomatic Friends”, *Journal of Specialized Translation* 6.
- Romero Fresco, P. 2009, “Naturalness in the Spanish dubbing language: A case of not-so-close Friends” *Meta*.
- Sinclair, J. 1991, *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Taylor, C. 2000, “Look Who’s Talking. An analysis of film dialogue as a variety of spoken discourse”, in L. Lombardo, L. Haarenan and J. Morley (eds.), *Massed Medias: Linguistic tools for interpreting media discourse*, Milano: LED, 247-278.
- Taylor, C. 2006, “‘I knew he’d say that!’ A consideration of the predictability of language use in film”, in *MuTra 2006 – Audiovisual Translation Scenarios*, EU-High-Level Scientific Conference Series, 1-11.
- Tognini Bonelli, E. 2001, *Corpus Linguistics at Work*. Amsterdam/Philadelphia: John Benjamins.
- Tognini Bonelli, E. 2002, “Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach”, in B. Altenberg and S. Granger (eds.), *Lexis in Contrast. Corpus-based approaches*. Amsterdam/Philadelphia: John Benjamins, 73-95.

The Distribution of Authorial Presence in Experimental Psychology Articles*

FRANCESCA BIANCHI
Università del Salento

ABSTRACT

Traditionally, teachers, prescriptive grammars, and writing guides have imposed the use of impersonal style in scientific writing. And indeed, in scientific papers, authorial presence is frequently hidden behind passive forms or the personification of the text or experiment. Other times, however, the author surfaces primarily by means of singular or, more frequently, plural first person pronouns or determiners. A few quantitative contributions dealing with overt authorial presence in experimental papers exist, but none of them focuses specifically on psychology. Therefore, the current study aims at expanding the existing literature by analysing the distribution of overt authorial presence in a corpus of experimental psychology articles. The corpus consists of 43 articles (298,332 running words), divided by move; each sentence in the corpus was manually tagged to identify the step it performs. Hypothetically relevant key words (*I, me, my, myself, we, us, our, ourselves, author, authors, author's*) were used as starting points for the identification of distributional patterns. The data were analysed quantitatively, in order to highlight: the distribution across files of each of the selected key words; their general distribution in the corpus; their distribution across moves and steps; the most frequent relevant moves in which they appeared per section; and their most frequent collocates per section.

Keywords: psychology; corpus; communicative function.

1. INTRODUCTION

The experimental research article is probably one of the most frequently studied academic text types. It has been analysed by a huge number of authors, from several different perspectives (see for example Bazerman 1988; Berkenkotter & Huckin 1995; Bhatia 1993; Dudley-Evans 1998; Gross *et al.* 2002; Paltridge 1997; Swales 1990). From a structural and rhetorical viewpoint, research articles tend to show a fairly limited number of recursive structural elements, often organised according to conventional textual schemata. Rhetorical strategies are acquired and perpetuated within a community by means of readings of previously published papers, imposition from reviewers (who work on the basis of their acquired knowledge), and writing guides. Traditionally, writing guides, as well as teachers, and prescriptive grammars, have imposed the use of impersonal style in scientific writing. And indeed, in scientific papers, authorial presence is frequently hidden behind passive forms or the personification of the text or experiment. On other occasions, however, the author surfaces primarily by means of singular or, more frequently, plural first person pronouns or determiners. A few quantitative contributions dealing with overt authorial presence in experimental papers exist, but none of them focuses specifically on psychology. Therefore, the current study aims at expanding the existing literature by analysing the distribution of overt authorial presence in a corpus of experimental psychology articles. The analytical perspective adopted in the current study is in line with the British tradition established by Swales (1990) and analyses the rhetorical structure of the article in terms of moves and steps, i.e. two hierarchical communicative levels in which a move (frequently but not necessarily corresponding to a section or paragraph of the article) may include one or more steps. Both moves and steps correspond to communicative functions. For a review of this tradition compared to two other major traditions of genre analysis, see Hyon (1996).

1.1 EXPERIMENTAL PSYCHOLOGY AS RHETORICAL DISCOURSE

Experimental psychology 'was the first human science to establish a specialised discourse distinguished from traditional philosophical discourse' (Bazerman 1988: 259) and this community adopted 'what they perceived to be the methods of the physical and biological sciences' and imitated 'the forms of argument developed within the natural sciences' (Bazerman 1988: 257). Their rhetorical devices are summarised in the *Publication Manual* (2001) of the American Psychological Association, which is now in its 5th edition and to which prospective authors are referred by the vast majority of psychology journals.¹ The *APA Publication Manual* is composed of 'approximately two hundred oversized pages of rules, ranging from such mechanics as spelling and punctuation through substantive issues of content and organization' (Bazerman 1988: 259). Indeed, the *APA Publication Manual* provides details of the structure the articles should have, how to name each section, what to include in each section, how to quote previous literature, and much more. This manual, however, was not meant to be a linguistic description, but rather a series of suggestions for prospective authors made by psychologists and editors.

The current study focuses on one rhetorical strategy of academic writing, namely overt authorial self-reference. The following paragraphs offer an overview of the indications provided by reading and writing guides to psychology researchers.

1.2 AUTHORIAL PRESENCE: REVIEW OF GUIDELINES IN READING AND WRITING MANUALS

A brief description of the possible ways in which the author or experimenter is expressed in scientific experimental articles is offered by Rossini Favretti while introducing students to the reading practice relating to psychology articles (Rossini Favretti & Bondi Paganelli 1988). She lists three strategies: explicit mention through personal pronouns (first person singular, or more frequently plural); agent hidden behind passive forms; or agent hidden behind personification structures (e.g. 'This chapter starts with...').

The APA *Publication Manual* mentioned above deals with many topics, but does not provide guidelines for the use of personal pronouns, or passive forms. Some writing guides, such as the *Guide to Grammar and Style* by Jack Lynch² – suggested by the online Encyclopedia of Psychology – state that

In scientific writing [...] sentences are routinely written in the passive voice; the authors are therefore given less importance, and the facts are made to speak for themselves.

On the other hand, the writing guidelines provided by the Psychology Writing Centre of the University of Washington read as follows:

Active voice is usually clearer and more to the point. Replace “nests are built” with “birds build nests”. Write “I found” instead of “it was discovered”. Don’t be afraid to use the passive voice if it really is clearer (as we do in this paper), but favor using subject, verb, object.³

Finally, some other guides to style, such as the document *How to write a paper in scientific journal style and format*⁴ by the Department of Biology of Bates College, Lewiston, take into consideration the fact that the use of one or another strategy may depend on the discipline and on the different sections of a scientific article and declare that

Some disciplines and their journals (e.g., organismal biology and ecology) have moved away from a very strict adherence to the third person construction, and permit limited use of the first person in published papers. Other disciplines, especially the biomedical fields, still prefer the third person construction. Limit your use of first person construction (i.e., “I (or we) undertook this study ...”): usually it is most acceptable in the Introduction and Discussion sections, and then only to a limited extent. Use first person in the methods sparingly if at all, and avoid its use in the results.

It emerges that guidelines are rather vague and sometimes even contradictory as far as this particular topic is concerned. This may depend on the fact that actual practice reveals a mixture of techniques, that preferred trends are not easily visible through qualitative analysis, or that analysis should be highly discipline-specific. In the literature, we found three authors providing quantitative⁵ contributions on overt authorial presence in experimental papers; of these, however, only one considered psychology, and analysed abstracts only. These contributions are summarised in the following paragraphs.

1.3 OVERT AUTHORIAL PRESENCE: REVIEW OF QUANTITATIVE STUDIES

Kuo (1998) analysed all personal pronouns in 36 scientific journal articles from three hard science fields, namely computer science, electronic engineering, and physics. With regard to pronouns that might indicate authorial presence, this author found that first-person plural pronouns and the corresponding determiner (*we/us/our*) were the most frequent ones in the corpus. On the other hand, first-person singular pronouns (*I/me/my*) did not occur at all; this absence was at least partly explained by the fact that even in single-authored papers, the writer referred to him/herself as *we*. Kuo also analysed distribution of *we* across discourse functions, but this was done on a sample of 9 research articles only. The analysis showed that exclusive *we/us/our* appeared in sentences performing one of the following functions: explaining what was done (the vast majority of cases for *we* and *us*, and highly frequent for *our*); showing results or findings (highly frequent for both *we* and *our*); showing commitment or contribution to research (the majority of cases for *our*; rather frequent also for *we*); proposing a theory or approach; stating a goal or purpose; justifying a proposition; hedging a proposition or claim; giving a reason or indicating necessity; comparing approaches or viewpoints; expressing wish or expectation. Some of these functions, such as explaining what was done, seemed to be distributed across all the sections of a journal article. Distribution of functions across sections, however, was not analysed systematically by this author.

Hyland (2001; 2003) focused on the use of self-citation and exclusive first person pronouns in two subsequent papers, both based on a corpus of 240 research articles in eight disciplines from both hard and soft fields (mechanical engineering, electrical engineering, marketing, philosophy, sociology, applied linguistics, physics, and microbiology). Unfortunately, psychology did not appear in this corpus. The 2001 paper focuses primarily on the key words *I, me, my, we, us, our, this writer, the research team* (and other similar expressions) used for self-mention, and on their discourse functions. In Hyland's data the vast majority of occurrences were pronouns (81%). As far as frequency of self-mention forms per discipline is concerned, each discipline seemed to prefer one or the other type of pronoun, regardless of the traditional distinction between hard and soft sciences. However, one general trend is visible in Hyland's tables, in almost all disciplines and with both pronouns/determiners: the subject pronoun is generally more frequent than the determiner, which in its turn is more frequent than the object pronoun. The author concluded saying that authorial presence surfaces at those points in the article where the authors are best able to promote themselves and their individual contributions; their intrusion is aimed at strengthening their credibility and their role in the research and helps them gain credit for their claims. Furthermore, in different disciplines, different degrees of authorial presence are allowed. In the 2003 paper, Hyland counted and analysed frequency and forms of self-reference, making a distinction between articles and abstracts. Furthermore, he analysed the distribution of authorial presence across discourse functions, on a limited sample of 10 research articles from each discipline. Self-reference was generally lower in abstracts than in papers, in almost all disciplines. Singular first person pronouns were virtually inexistent in all the hard sciences considered,

while in the soft sciences both singular and plural pronouns were frequent. A few occurrences of other forms of reference were found in the papers, scattered across all disciplines except philosophy, but none was found in the abstracts. Finally, four main purposes were identified for self-reference: stating a goal or outlining the structure of the paper; explaining a procedure; stating results or making a claim; elaborating an argument. The frequency 'hierarchy' of these functions depended on the discipline and was different in papers and abstracts.

Lastly, Martín Martín (2003) analysed the distribution and frequency of occurrence of first person pronouns in phonetics and psychology. His attention was focused on abstracts only, and his aim was to compare English and Spanish rhetorical practice. His data showed that first person pronouns, though present in the corpus, are not a relevant feature of abstracts in these disciplines. Almost 30% of the occurrences in English indicated 'the author as describer of the research', while 16.6% of occurrences were of 'the author as fully committed claim maker'. The other identified functions for exclusive first person pronouns were: 'the author as experiment conductor'; 'the author as opinion holder', and 'the author as cautious claim maker'. Interestingly, several of the occurrences of the author as fully-committed claim maker appeared in the Results and Discussion sections of the abstract.

The aim of the current study, therefore, was to expand the existing literature by analysing the distribution of overt authorial presence in a corpus of experimental psychology articles and taking advantage of corpus concordancing tools⁶ and quantitative methods of analysis.

2. MATERIALS AND METHODS

The current study is based on an annotated corpus of 40 experimental psychology articles taken from 15 different journals, for a total of 232,244 running words. This is a subset of the *Psychology Corpus* described in Bianchi & Pazzaglia (2007, subsection 1.4.1)⁷. In the corpus considered, slightly more than 80% of the journals are publications of the American Psychological Association (APA). The other articles all come from psychology journals published by Hogrefe & Huber.

Following Gledhill's and Bowker and Pearson's approaches (Bowker & Pearson 2002; Gledhill 1995, 1996, 2000), this corpus is divided into folders, each folder corresponding to a different move typical of experimental articles. Folders do not strictly correspond to the different sections of articles (structural units), but rather to moves (primary-level functional or communicative units) which may or may not overlap with the actual sections of articles. Moves were selected *a priori* following the work of various authors including Gopnik (1972) and Gläser (1995) and the preliminary analysis of a limited number of randomly selected articles. The moves appearing in the corpus are the following: Titles; Notes; Thanks; Abstract; Introduction; Literature Review; Method; Results; Discussion; Conclusion. For the sake of comparison with similar analyses carried out in the literature, only the following moves were considered and discussed in the current study: Abstract; Introduction; Literature Review; Method; Results; Discussion; Conclusion⁸.

Furthermore, steps, or communicative functions, were manually annotated within each move. The steps identified during the annotation phase are listed in

Table 1, along with the tags used. Any of the steps listed in the table could theoretically be found in any of the moves listed above.

For the purpose of the current study, the analyses were carried out starting from concordances of key words that might be indicative of explicit authorial presence. The following key words were taken into consideration: *I, me, my, myself, we, us, our, ourselves, author, authors, author's*. Concordance lines were retrieved with *Wordsmith Tools* for each key word and for each of the moves. The data retrieved were tabulated in Excel and integrated with information about the file from which the concordance line was taken, the step/communicative function in which the key word appears, the syntactic role of the key word (subject; direct object; other), and, when necessary, other information (such as inclusive or exclusive function, in the case of key words *we/us/our*).

Name of step	Tag	Description
Background	<backg> </backg>	reporting past research
Conclusions	<conc> </conc>	final remarks – conclusions
Discussion	<discussion> </discussion>	author's comments about findings, material, procedure, etc.
Findings	<findings> </findings>	description of results
General statement	<gen> </gen>	author's general considerations
Limits	<limits> </limits>	limits of current experiment
Material	<mat> </mat>	description of material
Method	<method> </method>	in the case of descriptions of subjects, material and procedure are not distinct
Need for extension	<further> </further>	need for further data, analysis, etc.
Object	<obj> </obj>	aim of the study
Procedure	<proc> </proc>	description of procedure
Rationale	<rationale> </rationale>	when reference is made to a theory and not to empirical past research
Relevance	<relevance> </relevance>	why are these previous studies relevant to the current experiment?
Subjects	<subj> </subj>	description of subjects
Niche	<niche> </niche>	describing/explaining one's research niche
Metatextual reference	<metatextual> </metatextual>	Metatextual reference
Hypothesis	<hypothesis> </hypothesis>	Statement of hypothesis
Quote	<quote> </quote>	Citation of other authors or texts (eg. questionnaire items)
Criteria	<criteria> </criteria>	Description of criteria/logic used to create material or analyse data
Acknowledgments	<genack> </genack>	Acknowledgments
Note to the text	<textnote> </textnote>	Note to the text

Table 1. Steps: Annotation scheme

The data were analysed quantitatively, in order to highlight: a) the distribution of each of the selected key words across files; b) their general distribution in the corpus; c) their distribution across moves and steps; d) the most frequent relevant steps in which they appeared per move; and e) their most frequent collocates per move. Some of the analyses were carried out considering all of the selected keywords, while others were carried out only on first person plural pronouns and determiners, as this was the only category of authorial self-reference with a statistically significant number of occurrences. Any irrelevant occurrence of a keyword, i.e. occurrences that were not indicative of authorial presence, were excluded from our calculations. Among the concordance lines excluded there were, for example, all those cases in which the word *author/s* referred to the author of a test or paper quoted for reference, rather than the author of the paper under analysis (e.g. *According to the test authors, the MC scale has satisfactory internal consistency*), or cases where the key word appeared within a quote (e.g. *“The teacher wants us to try new things”*: quote from an experimental subject’s response). The following sections illustrate and discuss the results of the analyses we performed.

3. RESULTS AND DISCUSSION

3.1 DISTRIBUTION OF AUTHORIAL PRESENCE ACROSS FILES

Analysis of the distribution of the relevant instances of the selected key words provided interesting insights into rhetorical standards of the psychology community and the (lack of) impact of traditional writing guidelines. The key word *author* appeared in 10 different articles, mostly in the abstract; *I/my* appeared in 3 different articles; and *we/us/our* appeared in all the remaining articles. Interestingly, those articles where the word *author* was used as self reference in the Abstract then shifted to using *we/our* in the sections that followed. All the articles in the corpus contain some form of overt authorial presence, despite traditional writing guidelines, including those of APA, which advocate impersonal style. It would be interesting to see whether any connection exists between these authorial choices and the two publishing houses.

3.2 GENERAL DISTRIBUTION

First of all we analysed the distribution of overt authorial presence in the whole corpus. To this end, wordlists were generated for each folder/move using *WordSmith Tools*, concordance lines were run for each of the selected key words, and irrelevant occurrences were deleted. This provided a first means of comparison with the previous studies.

The results of this general-level analysis are summarised in Table 2: the first column indicates whether the author resorted to first person singular, first person plural, or impersonal type of explicit reference; the second column shows the type of word representing overt authorial presence; the third column provides the raw occurrences of that word; the fourth column shows the relative percentage with respect to the given type of reference; the fifth column offers the percentage with respect to the entire set of overt references in the corpus; finally, the last column shows the syntactic position in which the key word appeared (subject/object/other).

Type of reference	Key word	Hits	Relative % (type of ref.)	General % (overt references)	S / O / Other position
1st person singular pron.	I	5	83	0.58	S (100%)
	Me	0	0	0	
	My	1	17	0.12	S (100%)
	Myself	0	0	0	
1st person plural pron.	We	599	69	68.77	S (100%)
	Us	17	2	1.95	O (99%); Other (1%)
	Our	240	29	27.55	S (50%); O (6%); Other (44%)
	Ourselves	0	0	0	
Impersonal ref.	Author	0	0	0	
	Authors	9	100	1.03	S (67%); O (11%); Other (22%)
	Author's	0	0	0	
	Authors'	0	0	0	

Table 2: General distribution of overt referential key words

As the number of hits shows, first person plural pronoun reference clearly dominated. Even in cases of impersonal overt reference the plural form (*authors*) was the only one used. Indeed, dominance of plural reference was expected, given that psychology articles are typically characterized by joined authorship. Furthermore also Kuo (1998) and Hyland (2003) have shown neat prevalence of plural reference in most academic fields. However, a few examples of reference by singular pronoun also occurred in the Psychology Corpus. This is in keeping with Hyland's (2003) study, where a high number of instances of singular personal pronouns were found in soft science papers.

Abstracts almost exclusively included references by means of the key word *author* (9 instances) and one case only of reference by personal pronoun (*we*). This finding is in keeping with Martín Martín (2003), who found that the use of personal pronouns was not a widespread practice in psychology abstracts.

Let us now take into consideration personal reference through pronouns only. If we explore the details of their syntactic roles, subject roles (*we/I*) outnumbered determiners (*our/my*), which in turn outnumbered object roles (*us/we*). This is true for both first person singular and first person plural reference. This type of distribution ($S > DET > O$) appeared in four moves, namely: Literature review, Introduction, Method, and Results. So far, the results match those in Hyland (2003). However, in our corpus, the Discussion and Conclusion moves behaved differently, showing a virtually equal number of subject and object pronouns. Abstracts could not be considered in this analysis, as they included only one instance of overt authorial presence through personal pronoun, in the form of the key word *we*.

Furthermore, if we analyse the noun phrases where referential determiners appear, authorial presence surfaces primarily in thematic/subject position. In fact, in this corpus, the number of *we* plus the number of *our* in subject position amounted to 83.5% of total cases of first person plural reference; and the only instance of *my* appeared in a subject noun phrase. In line with this, 67% of the hits of *author* were subjects.

To complete this series of analyses, we looked at the distribution of *our* in terms of Subject (S), Direct Object (DO) or Other Indirect (OI) role, across sections (Table 3). Results confirmed similarity between the Literature Review, Introduction, Method, and Results moves, where Other Indirect role preceded Direct Object, which in turn preceded Subject (OI > DO > S). Once more, the Discussion and Conclusion moves behaved differently, Subjects outnumbering Direct Objects, outnumbering in turn Other Indirect roles (S > DO > OI)⁹. Finally, no instances of *our* appeared in the Abstract section.

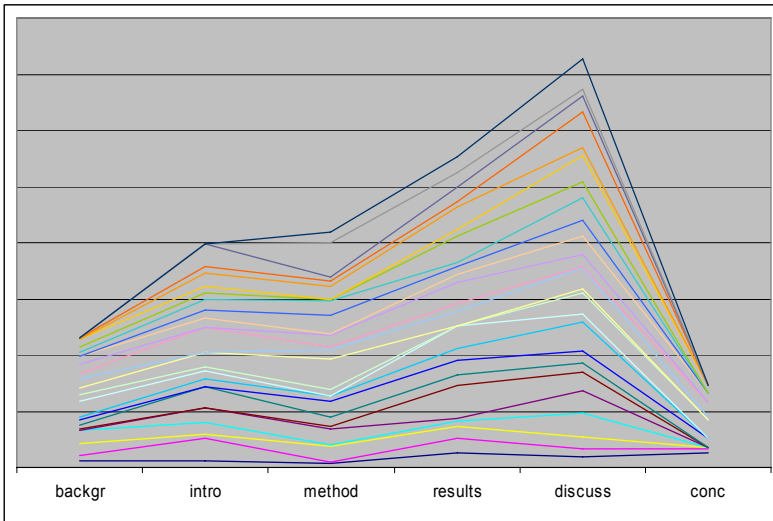
OUR	Abstr.	Lit. rev.	Intro.	Method	Results	Discuss.	Conc.	Notes	Tot.	%
Subj.		3	7	4	14	72	20	1	121	50.42%
Obj.		1	1	2	5	3	1		13	5.42%
Ind.		7	11	9	25	47	7		106	44.16%
TOT	0	11	19	15	44	122	28	1	240	100%

Table 3: Distribution of key word *our* in terms of S/DO/OI role, across sections

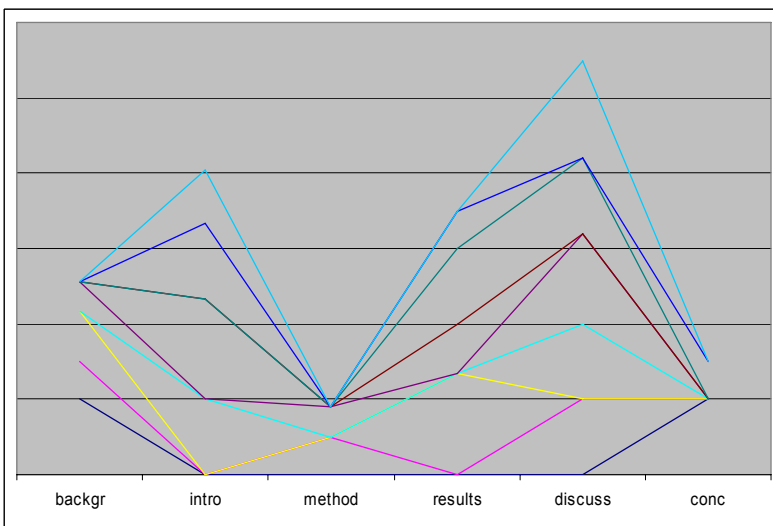
3.3 DISTRIBUTION OF AUTHORIAL PRESENCE ACROSS SECTIONS

To establish the incidence of authorial presence across sections, we decided to concentrate on the key words *we/us/our* only, as these represent the overwhelming majority of occurrences of overt reference. Only cases of *we/our/us* of the exclusive type were considered.

Graphs 1 and 2 show the distribution of overt authorial presence by first-person plural pronoun/determiner across moves. To produce these graphs, percentage values were used, calculated on the basis of the total number of occurrences of the three key words. In the graphs, each line corresponds to a different file in the corpus. Moves are reported on the x axis. The Abstract move is not present, because it never included occurrences of the key words taken into consideration. We must recall, however, that this move includes a few instances of reference using the key word *authors*. For a clearer graphical representation, two separate graphs were produced: one for those files where instances of the key words appeared in all sections, and one for those files where presence of the key words was discontinuous across sections.



Graph 1. Distribution of key words *we/us/our* across sections: files with constant presence of these key words



Graph 2. Distribution of key words *we/us/our* across sections: files with discontinuous presence of these key words

In both graphs, the Discussion section clearly appears as the top rising point of inflection, i.e. the section with the highest number (%) of the key words considered. This is true for all files, except 5 (85% of the files). In the vast majority of the files, the Introduction and Results sections are also points of inflection, though lower than the Discussion section. On the other hand, the Literature Review, Method and Conclusions sections are falling points of inflection, i.e. sections with the lowest number (%) of the key words taken into consideration.

These results may tentatively be explained by the content of each move: in the Literature Review move, focus is on previous literature rather than on the current study and its author/s; authorial presence, then, increases in the Introduction, as this move presents the current study and/or the article and is the place where authors frequently present their working hypotheses; in the Method move focus shifts to participants, materials and procedures, which are supposed to be described in an objective, detached type of language; attention returns to the current experiment in the Results section, where data are presented, and above all in the Discussion section, where the author/s can openly comment the results and their hypotheses; finally, the Conclusion section is probably poor in authorial presence markers because of the need to try and generalize results. Confirmation of these preliminary explanations will be searched for in an analysis of the distribution of steps across the different sections.

3.4 DISTRIBUTION OF AUTHORIAL PRESENCE ACROSS STEPS

Distribution of authorial presence across steps was assessed considering all types of authorial reference. Table 4 shows the distribution of authorial presence across steps, in decreasing order of frequency, regardless of move. In the third column, percentages are reported, shortened to the second decimal place. In the last column, examples from the corpus are provided (in shortened form, because of space limitations), with the indication of the move in which they were found. Here, the key word has been capitalized.

Step	Hits	%	Example from the corpus
Discussion	203	23.23	<discussion>OUR effects were quite consistent for estimates of the causes ...</discussion> [Conclusions]
Method	157	17.96	<method>The picture-recall test allowed US to analyze effects of valence and arousal separately;</method> [Discussion]
Object	107	12.24	<obj>OUR questions of interest related to the variables added after the academic skills' variable.</obj> [Results]
Hypothesis	101	11.56	<hypothesis>OUR hypothesis was that sad subjects would perceive situationally caused events as more likely... </hypothesis> [Results]
Findings	72	8.24	<findings>For m. corrugator activity, WE found a significant Defensiveness x Emotion effect:</findings> [Results]
Procedure	68	7.78	<proc>In Session 2, WE assessed behaviors over the preceding 2 weeks.</proc> [Method]
Limits	44	5.03	<limits> Another shortcoming of the study involves OUR use of measures for which only limited validity and reliability...</limits> [Discussion]

Step	Hits	%	Example from the corpus
Material	25	2.86	<mat>OUR protocol, however, diverged in two ways from the CTS.</mat> [Method]
Niche	23	2.63	<niche>What is different about OUR study when compared with others is that we focus on signif...</niche> [Introduction]
Metatextual reference	17	1.95	<metatextual>Before turning to the details of the longitudinal study, WE will briefly summarize the literature that explores our main...</metatextual> [Introduction]
Background	14	1.6	<backg>Our predictions were based on appraisal theories of emotion,...</backg> [Conclusion]
Subjects	7	0.8	<subj>We recruited 305 participants between the ages of 13 and 17</subj> [Method]
Conclusion	7	0.8	<conc>WE suggest that the study of LGB identity may enrich pres...</conc> [Conclusion]
General	6	0.69	<gen>These more directive methods, which WE refer to as curriculum centered, typically involve structured...</gen> [Introduction]
Note to the text	6	0.69	<textnote> ...10 periods) were not chosen as baseline in this report because WE decided to contrast a rest period with the induction period, ...</textnote> [Notes]
Criteria	5	0.57	<criteria>Specifically, six or more symptoms of either inattention or hyperactivity/impulsivity had to be endorsed (as "pretty much present" or "very much present") on the DBD Rating Scales to meet OUR criteria for ADHD.</criteria> [Method]
General acknowledgment	4	0.46	<genak>WE thank Maria Boardman, Susan Bunton, Phillis George, ...</genak> [Notes]
Relevance	4	0.46	<relevance>This study builds on previous work by the first AUTHOR (Levendosky & Graham-Bermann, 2001) in which an ecol...</relevance> [Literature review]
Need for extension	3	0.34	<further>OUR assessment of peer relationships could be improved by ...</further> [Discussion]
Rationale	1	0.11	<rationale>Recently, WE reported data demonstrating that young infants both discriminate among faces on...</rationale> [Literature review]

Table 4: Steps including authorial self-reference, in order of frequency

Overt authorial presence was found in all steps, except Quote which, as expected, never included instances of authorial self-reference. However, in order to compare the steps considered in this analysis to those that emerged in the previous studies, 'a naming conversion table' is needed, as the names and descriptions used by the different authors cannot be easily and univocally matched (Table 5). Indeed some doubts still remain, expressed by a question mark next to this matching. Martin's labels were not considered, as they were really difficult to compare.

Kuo's	Hyland's	Current
Explaining what was done	Explaining a procedure	Material / Method / Subjects / Procedure
Proposing a theory, approach, etc.		Hypothesis (?)
Stating a goal or purpose	Stating a goal or purpose	Object
Showing results or findings	Stating results or claim	Findings
Justifying a position	Elaborating an argument	Discussion / relevance
Hedging a proposition or claim	Elaborating an argument	Discussion
Showing commitment or contribution to research	Elaborating an argument	Discussion / conclusion / relevance
Comparing approaches, viewpoints, etc	Elaborating an argument	Discussion
Giving a reason or indicating necessity	Elaborating an argument	Discussion / limits
Expressing wish or expectations		Hypothesis (?)
	Expressing self-benefits	Discussion

Table 5: Steps: naming conversion table

The picture emerging from this table led us to group some steps together: the Material, Procedure, Method, and Subjects steps were grouped into 'Procedure+'; the Discussion, Conclusion, and Relevance steps were grouped into 'Discussion+'. This, in turn, led us to a general review of the initial steps and to a grouping of the Background, Rationale, and Criteria steps as 'Background+'. Recalculation of the results of the current analysis provided the ranking in Table 6, in which only the most frequent functions are reported.

Step	%
Discussion +	29.8
Procedure+	29.4
Object	12.2
Hypothesis	11.6
Findings	8.2

Table 6: General distribution of overt authorial presence: most frequent steps, after grouping

Interestingly, all the functions in which authorial self-reference emerged in the previous literature are among the top five in our corpus. However, in terms of ranking no similarity seems to exist between the data in our corpus and those considered by the other authors. This confirms Hyland's (2003) result that the frequency 'hierarchy' of these functions depends on the discipline.

If we look at overt authorial distribution within the different moves, the pictures in Tables 7a-7f emerge. Moves are listed in decreasing order of frequency. Abstracts have not been included in these tables, given the few occurrences of the authorial key words used in this study; however, this move will be briefly discussed after the comments on the other moves.

Literature Review	%
Object+	31.4
Hypothesis	26.5
Procedure+	18.0
Discuss+	15.7
Background+	6.0
Niche	2.4

Table 7a. Steps with overt authorial presence: Literature review move

Introduction	%
Object+	29.7
Hypothesis	26.9
Procedure+	25.5
Niche	4.9
Background+	4.1
Discuss+	4.1
Metatextual reference	3.4
Gen	0.7
Findings	0.7

Table 7b. Steps with overt authorial presence: Introduction move

Method	%
Procedure+	82.1
Discuss+	14.2
Hypothesis	2.8
Background+	0.9

Table 7c. Steps with overt authorial presence: Method move

Discussion	%
discussion+	42.8
Limits	14.8
Findings	14.0
Hypothesis	6.2
Niche	5.4
procedure+	6.0
Object	6.0
background+	2.5
General	1.4
Further	0.9

Table 7e. Steps with overt authorial presence: Discussion move

Results	%
Procedure+	46.6
Discuss+	16.2
Findings	13.9
Object	9.5
Hypothesis	7.2
Metatextual reference	4.9
Background+	0.9
General	0.4
Limits	0.4

Table 7d. Steps with overt authorial presence: Results move

Conclusion	%
discussion+	51.9
Limits	13.5
Findings	11.8
Hypothesis	9.5
Object	5.7
Further	1.9
General	1.9
procedure+	1.9
Niche	1.9

Table 7f. Steps with overt authorial presence: Conclusion move

Each move is characterized by a specific ranking of steps, but similarities can be seen between groups of moves. Furthermore, as we will see, overt authorial presence tends to appear in steps that seem to be particularly significant for and expected in the given move.

The Literature review move shows several similarities with the Introduction move. In particular, all the steps present in the Background move authorial presence list are also present in the Introduction move, the first three (Object+; Hypothesis; Procedure+) exactly in the same order, the other three (Discussion+;

Background+; Niche) in reverse order. The Introduction list also includes three other steps with overt authorial presence, namely Metatextual reference, General and Findings. These, like the ones higher up in the ranking, are typical steps for an Introduction, where one would expect to see a brief description of the contents and structure of the whole paper (Metatextual reference) and maybe also a brief summary of the results of the study (Findings).

The Method move shows several similarities with the Results move. As before, all the steps in the Method move (Procedure+; Discussion+; Hypothesis; Background+) are also present in the Results one. Furthermore, the Results move includes several other steps with overt authorial presence. This mirrors the wide variety of communicative functions that characterize the Result section of an article. Here, too, authorial presence tends to appear in steps that seem to be particularly significant for and expected in the given move, namely description of results (Findings) and description of tables and graphs (Metatextual reference).

Finally, the Discussion and Conclusion moves are among the richest in terms of variety of steps with authorial presence. Their lists share almost all the same steps, the top four (Discussion+; Limits; Findings; Hypothesis) even in the same order. A major difference is the presence of step Background+ in the Discussion move. Indeed, reference to the literature can be expected when discussing results.

The Abstract section represents a move on its own. Indeed, we have already mentioned that very few instances of overt authorial presence were found in this section, and nearly all of them were characterized by the key word *authors*, rather than *we/us/our*. These few occurrences were scattered through five steps: object, procedure+; discussion; findings; and hypothesis. Given the low number of hits, it is clearly pointless to rank steps in order of frequency. However, as was the case with the other moves, most of these steps (object, hypothesis, procedure, and findings) seem to be highly typical of an abstract.

These results are only partially in keeping with our tentative explanation of general distribution across moves (Section 3.3). On the one hand, the hypothesis that low presence of authorial reference in the Literature review move might be due to focus on previous literature rather than on the current study and its author/s is supported by the very few occurrences of self reference in the Background step. Furthermore, an increase in authorial presence in the Introduction due to topic shift to the current study and the author's working hypotheses is indeed supported by the wide number of steps and their ranking within this move. Finally, the wide difference in authorial presence that characterizes the Discussion and Conclusion moves (the highest number vs. one of the lowest numbers) can only be explained in terms of topic shift, as our analysis of steps shows very little difference between these two sections. On the other hand, the hypothesis that participants, materials and procedures – which are the main topics of the Methods move – are described in an objective, detached type of language, is contrasted by the fact that a high number of occurrences of authorial self-reference were found in the step Procedure+ (top of the list in the Method and Results moves, and among the top three in the Literature review and Introduction moves). Furthermore, increased presence in the Results move seem connected to the wide variety of steps that characterize this move, which accompanies topic shift to the presentation of results.

3.5 COLLOCATES PER SECTION

Finally, we decided to attempt an analysis of collocates, to see whether this would give us some further clues to the discussion of our quantitative findings. Collocates (N+1/2/3) were assessed only for the key words *we* and *our* – the most frequent key words of all – and are summarized in Table 8. In each column, the first row of each key word shows the most frequent collocate. The Conclusion move does not appear in this list because it did not include relevant occurrences of either key word.

Key word	Abstract	Literature review	Introduction	Method	Results	Discussion
WE	studied	expected	hypothesized	used	found	found
		hypothesized	examined	assessed	conducted	expected
		examined	believed	had	used	hypothesized
			expected	tested	entered	may
					performed	
					calculated	
					computed	
					ran	
					predicted	
OUR		data	hypothesis	analyses	hypothesis	findings
		study	study	criteria	analysis	predictions
		research	aim	hypothesis	data	study
		work	groups	study	research questions	model
		view	sample	goal	expectations	account
		knowledge	procedure	raters	model	data
			intent	results	criteria	version
			expectations	laboratory	focus	effects
			prediction	variable	design	results
			knowledge	protocol	sample	
			model		participants	
			belief		predictions	
			focus		tests	
			attempts		measures	
		research				

Table 8: Collocates (N+1) of key words *we* and *our*

In the Psychology Corpus, *we* collocated in N+1/2/3 position with verbs only, as expected, all of them in the active form. Interestingly, these verbs are connected to a limited number of communicative functions: Hypothesis; Procedure; Object; and Findings. Collocations with words indicating Hypothesis are scattered in the following moves: Literature review (collocates: *expected*; *hypothesized*); Introduction (collocates: *hypothesized*; *believed*; *expected*); Results (only collocate: *predicted*); and Discussion (collocates: *expected*; *hypothesized*; *may*). *We* collocates with the idea of Procedure primarily in the Results move (collocates: *conducted*; *used*; *entered*; *performed*; *calculated*; *computed*; *ran*), but also distributed in the following moves: Literature review and Introduction (only collocate: *examined*); and Method (only collocate: *used*). Collocates referring to function Object are found in the Method (collocates: *assessed*; *had*; *tested*), and Abstract (only collocate: *studied*) moves. Finally, *we* collocated with the idea of Findings in the Results and Discussion moves (only collocate: *found*).

On the other hand, N+1 collocates of the key word *our* are associated to a wider number of communicative functions, and are sometimes difficult to classify. As expected, they are all nouns. The key word *our* frequently collocates with words indicating Hypothesis. This is true in particular for the Introduction (collocates: *hypothesis*; *intent*; *expectations*; *prediction*), and Results (collocates: *hypothesis*; *research questions*; *expectations*; *predictions*) moves. However, one collocate was also found in the Method (only collocate: *hypothesis*), and Discussion (only collocate: *predictions*) moves. Other collocates lead us back to the Procedure + function, with the majority of words indicating either Subjects or Method. These collocates were found in the following moves: Introduction (collocates: *groups*; *sample*; *procedure*; *attempts*); Method (collocates: *analyses*; *raters*; *laboratory*; *variable*; *protocol*); Results (collocates: *analysis*; *model*; *design*; *sample*; *participants*; *tests*; *measures*); and Discussion (only collocate: *model*). Collocates of *our* also frequently express the function Findings; these are scattered primarily in the Discussion move (collocates: *findings*; *account*; *data*; *effects*; *results*), but also in the Method move (only collocate: *results*), and in the Introduction and Results moves (only collocate: *data*). Less frequently collocates refer to the Object, Discussion, or Criteria functions. Collocates indicating the Object of the study were found in the following moves: Introduction (collocates: *aim*; *focus*); Method (only collocate: *goal*); and Results (only collocate: *focus*). Collocates indicating Discussion were found in the Literature review (collocates: *view*; *knowledge*), and Introduction (collocates: *knowledge*; *belief*) moves. Finally, one collocate only was found referring to Criteria (only collocate: *criteria*), in the Method and Results moves. The remaining collocates (*study*; *research*; *work*; *version*) are too general to be matched to a specific function.

Partial coincidence can then be seen between the functions highlighted by the collocates of the two key words. In fact, the Hypothesis, Findings and Object functions are common to both, while the Procedure function of collocates of *we* does not coincide with the Procedure+ function of collocates of *our*, as the former refers to the description of the procedure itself, while the latter includes mainly description of subjects or of method. Analysis of syntactic roles, however, only partially confirmed the hypothesis that the wider number of functions connected to collocates of *our* might be due to those cases when the noun phrase did

not appear in subject position. In fact, most, but not all, instances of *our+collocate* in subject position refer to the same function as collocates of *we*; others, instead, refer to other functions.

Furthermore, we checked whether direct correspondence existed between the step performed by a sentence including authorial reference and the communicative function suggested by the collocate. In several cases correspondence existed, as in the following examples: <findings>*Furthermore, WE found patterns of relationship qualities that distinguished...*</findings>; <hypothesis>*In addition, WE expected a positive relationship between defensiveness...*</hypothesis>. However many instances were also found where sentence and collocate inspired different communicative functions. In the following sentence <conc>*In summary, WE replicated the pattern of results of Study 2...*</conc>, for example, collocate *replicate* describes a Method or Procedure, and the communicative function of the whole sentence (Conclusion) is performed by phrase *In summary*, in initial position. In the sentence <limits>*WE also attempted to call the siblings of each subject but...*</limits>, instead, collocate *attempt* suggests a Discussion, or Method, or Procedure function, and the Limits step was derived from the whole sentence, and marked by the conjunction *but*. These instances also rule out the existence of a direct connection between step/collocate-function, and subject role of the key word.

4. CONCLUSIONS

This study – based on a corpus of 43 articles (298,332 running words), divided by move (each move roughly corresponding to a different section of the article) and tagged for steps – has shown that, despite the indications in the APA manual and the general indications of style guides, the authors frequently surface in experimental psychology articles. Indeed, every article in the corpus contained some form of overt authorial presence. The preferred referential expression was the use of first person plural reference (*we/our/us*), although a few instances of reference by the word *authors* (always in the plural) or *I/my* were also found. Furthermore, authorial references mostly appeared in subject position.

Overt authorial presence was seen in all moves. The highest number of hits was found in the Results and Discussion moves, while the lowest was in the Abstract, Method and Conclusions moves. In terms of steps, key words indicating authorial presence were found in all steps except Quote. The highest number of instances belonged to the Procedure+, Discussion+, Method, Object, Hypothesis and Findings steps, in that order. Within the different moves, analysis of the distribution of authorial presence across steps showed a different ranking of steps in each move. Similarities were found between the Literature review and Introduction moves, the Method and Results moves, and the Discussion and Conclusions moves, while Abstract emerged as a move with its own characteristics. Furthermore, each move also included steps directly connected to the contents of the move.

Analysis of collocates was also carried out. Collocates replicated the most frequent steps (Hypothesis, Findings+, Procedure, Object) plus a few other functions. However, no direct correspondence was found between collocate and sen-

tence in terms of communicative function, not even when the key word was in subject position.

Comparison of results with previous findings was also attempted, despite methodological differences in the definition of steps. Generally speaking, the findings of this study are in keeping with those described in the literature, even though the other studies analysed articles from fields other than psychology. Comparison seems to confirm that, for several disciplines including psychology, the rhetorical habits observable in experimental papers tend to depart from the strategies suggested by writing guides and publication manuals, at least as far as authorial presence is concerned. In particular, our data seem perfectly in keeping with considerations by Bazerman (1988: 275), who declared that although in the 'last twenty years, a major style change in the [APA] psychological journal has [...] started to take place [as a] result of the rising influence of a cognitive psychology based on the computer model', the 'new style has not yet affected the *Publication Manual* in any significant way'. Comparison between the analysis of authorial presence across moves and steps have shown a highly articulated picture which could benefit from further analysis.

To conclude, despite any shortcomings this study might have, we believe that this paper may be of interest for the linguistic community, as it provides insight into a little analysed rhetorical practice of psychology researchers. By focusing on one single discipline, by considering all the main sections of psychology research articles, by performing detailed quantitative analysis on a full tagged corpus, and by analysing the distribution of overt authorial presence across both moves and sections, we hope to have covered a niche and provided hints for future research in the field.

NOTES

* A preliminary version of this study was presented at the ICAME 2006 Conference, 24-28 May 2006, Helsinki.

1 Interestingly, APA norms have been adopted as standards for publication also by journals of other disciplines, including linguistics.

2 <http://andromeda.rutgers.edu/~jlynch/Writing/p.html>

3 <http://web.psych.washington.edu/writingcenter/writingguides/pdf/style.pdf>

4 <http://abacus.bates.edu/~ganderso/biology/resources/writing/HTWgeneral.html>

5 Qualitative contributions on personal reference in academic writings are not taken into consideration here, though some interesting papers exist in the literature; among them: Harwood (2005), who proposes a functional taxonomy of self-promotional *I* and *we* and shows that scientific articles in the hard sciences carry a self-promotional flavour with the help of personal pronouns; Vladimirou (2007) who argues that expert writers in the ‘academic discipline of linguistics’ are surprisingly inconsistent in their use of personal reference.

6 For a discussion of the benefits of an integration of ESP and corpus approaches, see Flowerdew 2005.

7 The *Psychology Corpus* includes 67 empirical studies in psychology taken from 20 different international journals, with a total amount of 462,772 words. It exists in two parallel versions: an annotated one and a non-annotated one. When this study was carried out, the annotated version was undergoing revisions by the author; therefore only the part of the corpus that had been completely reviewed was used.

8 In the analyses carried out on the most general level, also the Note move was considered in so far as it contains further information about the experiment described in the paper. However, when getting to more specific types of analyses, this move was ignored, given the specificity of its steps and the low number of authorial references.

9 This trend was confirmed by the fact the only *my* appears in the discussion section and is in subject position.

10 This trend was confirmed by the fact the only *my* appears in the discussion section and is in subject position.

REFERENCES

Bazerman, Charles (1988), *Shaping Written Knowledge. The genre and activity of the experimental article in science*. London: The University of Wisconsin Press.

Berkenkotter, Carol & Huckin, Thomas N. (1995), *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*, Hillsdale - New Jersey: Lawrence Erlbaum Associates.

Bhatia, Vijay K. (1993), *Analysing Genre. Language Use in Professional Settings*, London & New York: Longman.

Bianchi, Francesca & Pazzaglia, Roberto (2007), “Student writing of research articles in a foreign lan-

guage: metacognition and corpora”. In: Facchinetti, Roberta (ed.), *Corpus Linguistics 25 Years on*. Amsterdam: Rodopi, pp. 261-289.

Bowker, Lynne & Pearson, Jennifer (2002), *Working with specialized language. A practical guide to using corpora*, London & New York: Routledge.

Dudley-Evans, Tony & St John, Maggie Jo (1998), *Developments in English for Specific Purposes. A multi-disciplinary approach*. Cambridge Language Teaching Library, Cambridge: Cambridge University Press.

Flowerdew, Lynne (2005), “An integration of corpus-based and genre-

- based approaches to text analysis in EAP/ESP: countering criticisms against corpus-based methodologies", *English for Specific Purposes*, 24: 321-332.
- Gläser, Rosemarie (1995), *Linguistic features and genre profiles of scientific English*, Frankfurt-am-Main: Peter Lang GmbH.
- Gledhill, Christopher (1995), "Collocation and genre analysis: the phraseology of grammatical items in cancer research abstracts and articles", *Zeitschrift für Anglistik und Amerikanistik*, 43: 11-29.
- Gledhill, Christopher (1996), "Science as collocation: phraseology in cancer research articles". In: Botley, Simon, Glass, Julia, McEnery, Tony & Wilson, Andrew (eds.) *Proceedings of the teaching and language corpora 1996*. Lancaster: UCREL: 108-126.
- Gledhill, Christopher (2000), "The discourse function of collocation in research article introductions", *English for Specific Purposes*, 19 (2): 115-135.
- Gopnik, Myrna (1972), *Linguistic structures in scientific texts*, The Hague: Mouton.
- Gross, Alan G., Harmon, Joseph E., Reidy, Michael (2002), *Communicating Science. The Scientific Article from the 17th Century to the Present*, Oxford: Oxford University Press.
- Harwood, Nigel (2005), "Nowhere has anyone attempted... In this article I am to do just that. A corpus based study of self-promotional I and we in academic writing across four disciplines", *Journal of Pragmatics*, 37 (8): 1207-1231.
- Hyland, Kan (2001), "Humble servants of the discipline? Self-mention in research articles", *English for Specific Purposes*, 20: 207-226.
- Hyland, Ken (2003), "Self-citation and self-reference: Credibility and promotion in academic publication", *Journal of the American Society for Information Science and Technology*, 54 (3): 251-259.
- Hyon, Sunny (1996), "Genre in three traditions: implications for ESL", *TESOL Quarterly*, 30 (4): 693-722.
- Kuo, Chich-Hua (1999), "The use of personal pronouns: Role relationships in scientific journal articles", *English for Specific Purposes* 18: 121-138.
- Martin Martín, Pedro (2003), "Personal attribution in English and Spanish scientific texts", *Bells*, 12, retrieved on 10 Feb. 2006 from www.publicacions.ub.es/revistes/bells12/PDF/arto9.pdf
- Paltridge, Brian (1997), *Genre, Frames and Writing in Research Settings*, Amsterdam & Philadelphia: John Benjamins' Publishing Company.
- Publication Manual of the American Psychological Association* (2001), Washington, D.C.: American Psychological Association.
- Rossini Faretta, Rema & Bondi Paganelli, Marina. (1988), *Il testo psicologico. Aspetti della traduzione e della lettura in lingua inglese*, Bologna: Pitagora Editrice.
- Swales, John (1990), *Genre analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- Vladimirou, Dimitra (2007), "I suggest that we need more research'. Personal reference in linguistics journal articles". In: Gabrielatos, Costas, Slessor, Richard & Unger J.W. (eds), *Papers from the Lancaster University Postgraduate Conference in Linguistics and Language Teaching, 2006, Vol. 1: Papers from LAEL PG 2006*, Lancaster: Lancaster University.

Discovering Language through Corpora:*

The Skills Learners Need and the Difficulties they Encounter

FRANCESCA BIANCHI, ELENA MANCA
Università del Salento

ABSTRACT

Most scholars agree on considering corpora as a valuable source of linguistic information for native and non-native speakers alike. Few researchers, however, have dealt with and systematically analysed the objective difficulties encountered by students while trying to exploit corpus data. The current paper describes a quantitative study of corpus consultation by learners and aims to establish whether different corpus analysis tasks can be considered to have different degrees of intrinsic difficulty. To this end, 26 corpus project work assignments produced by two different groups of students were assessed and tagged according to specific parameters that reflect the skills needed in corpus analysis. The data were analysed applying both parametric (ANOVA) and non parametric tests (Mann-Whitney U-test), which showed that, despite clear individual and teaching/learning environment differences between the two groups of students, the students' results in most of the tasks were due to different levels of intrinsic difficulty. This led to the creation of a General Difficulty List of Corpus Analysis Tasks.

Keywords: corpus analysis, skills, student difficulties, analysis of project works, teaching planning

1. INTRODUCTION AND RESEARCH QUESTION

Most scholars agree on considering corpora as a valuable source of linguistic information for native and non-native speakers alike. For this reason, many linguists have been increasingly advocating the use of corpora in language learning/teaching (Aston 2001; Cobb 1997; Flowerdew 1993; Levy 1990, 1997; Owen 1996; Sinclair 2003, 2004; Steven 1991; Tribble & Johns 1997). The possible uses of corpora in language learning and in translation have been widely discussed (Frankenberg-Garcia 2005b; Gavioli 2005; Gavioli & Zanettin 1997; Granger & Tribble 1998; Sharoff 2004; Tognini Bonelli 2001; Tribble & Jones 1990, 1997; Zanettin 2002; Zanettin *et al.* 2003), although some authors have illustrated the need for corpora specifically created for pedagogic purposes (Braun 2005). Furthermore, some researchers have suggested direct student access to corpora (Gavioli & Aston 2001), and others have described the serendipitous discoveries that students have made while directly accessing corpora (Bernardini 2000a, 2004; Bernardini & Zanettin 1997). Few researchers, however, have dealt with and systematically analysed the difficulties encountered by students while trying to exploit corpus data. A brief review of the major papers on this issue is provided below.

One of the first authors to deal with the processes and results of students' corpus exploration is Bernardini (2000b). Her paper focuses on students using the *British National Corpus*. Her observation of how the students approach corpus investigation reveals some problematic tendencies, including the fact that the students often ignore variants, do not look for alternative and more successful approaches and tend to make only summary analyses.

Kennedy and Miceli (2001) provide a fairly detailed qualitative analysis of the way students proceeded in using a corpus as a reference for writing in a foreign language (Italian). They consider four steps in corpus investigation: formulating the question; devising a search strategy; observing the examples and selecting relevant ones; drawing conclusions. Their recordings and interviews show that students have problems with all the steps considered, which led the authors to devise some tips for each step, so as to guide the students towards more precise and fruitful research practices.

Sun (2003) analyses the learning process and the strategies used by three undergraduate English FL learners when accessing corpus data to proofread texts with grammar mistakes. She also examines the factors that impact on the students' behaviour. The students received a relatively quick introduction to concordance analysis, and their problem-solving strategies were collected using a think-aloud protocol. This author classifies four cognitive skills required in the analysis of concordance lines, namely: comparing; grouping; differentiating; and making inferences. From Sun's description, it seems that the three students went constantly through all the phases mentioned, even though the teacher's help was at times needed for the students' correct progress. The author concludes by stating that four factors influenced the learners' investigations and the strategies they used: prior knowledge; cognitive skills; teacher intervention; and skills in using the concordance software.

Another paper mentioning and analysing student difficulties in corpus use is by Yoon and Hirvela (2004). The major focus in their study, however, is on student responses to corpus use so that the analysis of problems/difficulties is carried out with the goal of providing evidence for student likes and dislikes. The types of difficulties they take into consideration revolve around what the students feel as problems in accessing the corpus and include matters such as: data analysis is time consuming; concordance output provided too many or too few sentences; texts or chunks were difficult to read or included unknown vocabulary; Internet connection was too slow or not available. Only one item in their list generally refers to difficulty in 'concordance output analysis'.

Chambers (2005) examines the strategies generally used by her students in accessing corpora, and their efficacy or otherwise. This was part of a study designed to 'examine a number of aspects of course design in corpora and language learning involving direct access by learners' (Chambers 2005: 112) and to 'draw some conclusions concerning the factors that favour the integration of corpora and concordancing into the language-learning environment and the obstacles which remain to be surmounted' (Chambers 2005: 112). Her discussion is based on qualitative analysis of 11 end-of-course essays. Her data highlighted

a considerable amount of variation in the students' ability to explore the corpus (Chambers 2005: 119), which led her to conclude that "differences in motivation or learning styles may explain the considerable variation in the success of the activity. In addition to the variation in analytical ability, there was also considerable variation in the students' ability to reflect on the nature and limitations of the corpus, an ability which came easily to some students, but was totally lacking in others (Chambers 2005:119).

Finally, Frankenberg-Garcia has dedicated more than one study to this issue. Her 2005 paper focuses on translation students and how they combine the use of corpora, termbanks, the Web and printed references. Her plenary speech in 2006 at the 7th TALC Conference (Paris)¹, provided a detailed description of novice users' problems in accessing corpus data and presented task-based, non-corpus-specific

conscious-raising exercises aimed at helping [novice users] gauge different corpora and discern which ones are best suited to their purposes, develop basic corpus-searching strategies, and get used to interpreting corpus data (Frankenberg-Garcia 2006: 5).

Her list was inspired by a general review of the literature as well as by personal observation of the way students used the *COMPARA* corpus. Her comments and exercises focused on issues such as problems in choosing a suitable type of corpus or sub-corpus, formulating corpus queries and follow-up queries, and interpreting corpus data.

The above-mentioned studies are substantially different in terms of focus of interest, the way they were conducted, the types of students involved, the teaching objectives of each course/module, and the way corpora were introduced to the students. Furthermore, their results are frequently rather contex-

tualized. However – quoting from Frankenberg-Garcia (2006: 5) – ‘they all converge to suggest that corpus skills which come as second nature to experts are not obvious at all to the untrained’. This was previously pointed out by Sinclair (2004: 2) when he stated that corpora are not a simple object and that lack of training and experience in retrieving data may lead students to consider nonsensical conclusions as insightful ones. Thus, the teachers who decide to adopt a corpus approach to language teaching/learning should be aware of the difficulties that this applied discipline involves and pace the training according to the skills one might expect from students. Meaningful corpus analysis requires not only good knowledge of the basic theoretical concepts of the subject, but also practical experience, as well as skill in using concordancers and in observing, identifying, classifying, and generalizing data.

The current paper attempts a systematic analysis of the difficulties encountered by students in approaching language through concordancing. Attention is given to the phases that follow concordance line retrieval and which include tasks such as selecting concordance lines, categorizing collocates, analysing collocation and colligation, and using the data retrieved to make generalizations about language or to find a suitable translation equivalent.

As a general hypothesis we may presume that the performance of a task depends on: 1. the difficulty of the task itself (intrinsic difficulty); 2. individual factors, i.e. individual abilities and background knowledge; and 3. environmental factors, such as course and exam focus. So far, corpus linguists do not seem to have analysed intrinsic difficulties in corpus analysis tasks. Starting from empirical observations, we developed the following working hypothesis: if two different groups of students show similar difficulty in performing specific tasks, the influence of individual and teaching/learning environment factors can be considered less relevant than task-intrinsic difficulty. The following sections describe how this hypothesis was tested using two randomly selected groups of students.

2. DESIGN OF THE STUDY

Two separate groups of foreign language students specializing in translation studies participated in this study: 40 bachelor students from the University of Lecce, and 10 MA students from the University of Genoa. Both groups were introduced to corpus consultation and analysis and were asked to complete an end-of-course corpus research assignment. The assignment papers were analysed using a specially developed taxonomy of twelve corpus analysis tasks. Analyses were carried out at individual, group, and general levels.

2.1 PARTICIPANTS

Two groups of Italian students participated in this study: 40 undergraduate students enrolled at the Faculty of Foreign Languages of the University of Lecce, and 10 MA students enrolled at the Faculty of Foreign Languages of the University of Genoa. The two groups attended separate courses on how to use corpora for analysing language and finding translation equivalents: the courses were held by the authors of this paper (hereafter, researchers or we), one in Lecce and one in Genoa. None of the students had ever heard of corpora or corpus analysis before,

except Student 102 who had very basic knowledge in the field. The students differed in terms of foreign language background, general academic background, and familiarity with assignment writing. Moreover, they were exposed to different teaching methods. However, both groups of students were introduced to corpus analysis tools and methods and were asked to submit a similar corpus research assignment at the end of the course, which represents the rationale for the comparison and contrast of their results. A schematic summary of the similarities and differences between the two groups is provided in Table 1.

Feature	Lecce	Genoa
Native tongue	Italian	Italian
Course level	Bachelor	MA
Year	2 nd	1 st and 2 nd
Number of students	40	10 (5+5)
Number of hours of lessons (including practice)	60 hours	20 hours
Language in which the course was taught	English	Italian
Language in which the project work was carried out	English	Foreign language of student's choice
Languages of the comparable corpus used	English - Italian	Language of student's choice - Italian
Level of proficiency in the FL of the project work	B1/B2	B2/C1
Assignment	Pair work	Individual work

Table 1: The two groups participating in the study

As table 1 shows, the students were all Italian native speakers. Lecce students were all specializing in English, their course was taught in English and the assignment papers all analysed an English-Italian comparable corpus. On the other hand, the Genoa group included students specializing in a range of different European languages; for this reason the course was taught in Italian and the students analysed comparable corpora in Italian and a foreign language (FL) of their choice. The two groups also differed in terms of proficiency level in the foreign language: B1/B2 in the European Framework of Reference for Lecce students, and the higher B2/C1 for the students in Genoa.

The following section provides a brief description of the contents and teaching methods of the two courses. The description attempts to highlight similarities and differences between the two courses with respect to the tasks considered in the current study. Contents unrelated to the tasks considered have been omitted for the sake of clarity and focus.

2.2 COURSE CONTENTS

Both courses illustrated the following basic corpus linguistic concepts: corpora; word lists; running and sorting concordances; collocation; colligation; phraseology; and semantic prosody. However, each of the researchers adopted an individual approach, partly due to the different number of hours and students characterising each course.

In Lecce, the course included two parallel modules: a 40-hour theoretical classroom module, and a 20-hour practical lab module. Following the British tradition of Firth (1957), Halliday (1985), Sinclair (1996), Stubbs (1996), and Tognini-Bonelli (2001), the theoretical module introduced the students to the basic corpus linguistics concepts mentioned above, plus the other relevant concepts of context, meaning in context, and semantic preference. Furthermore, it explained how to find translation equivalents using comparable corpora (Tognini Bonelli & Manca 2002). The practical module, which took place in a computer lab, taught the students how to assemble their own corpora, use *Wordsmith Tools* (a corpus concordancer), and retrieve and analyze data. When the students seemed to be ready to work on their own, they were put in pairs, so that they could help each other out, and tutored in performing a given series of tasks required for autonomous use of corpora for linguistic analysis and translation. In this phase of the course, the students were asked to run the wordlists of the two comparable corpora they had created, search for the most frequent words in each wordlist, compare the two wordlists, and look for mismatch in frequency between items in the two wordlists. They were then encouraged to choose one or two English content words, run their concordances, sort the concordance lines, and find immediate collocates and colligates. As a further step, they were asked to enlarge the linguistic co-text in order to find collocates in $N-2/3/4$ and $N+2/3/4$. Once they had identified the most frequent collocates occurring with the node word, the students were asked to group the collocates into semantic fields, and identify the recurrent phraseology of the node word and its patterns of use. As a last step, they were invited to find Italian translation equivalents for each of the senses identified for the node word.

At the end of the course, the students were asked to hand in a paper with the following assignment (pair work): *Choose 1 or 2 words among the most frequent in your English corpus. For each word identify collocation, colligation, semantic preference, and semantic prosody. Identify the phraseology around the node word. Identify possible translation equivalents of the node word in your Italian comparable corpus using the methodology seen in class.*

In Genoa, all lessons were carried out in a computer lab provided with two concordancing programs: *Wordsmith Tools* and *ConcApp*. Each lesson included both theory and practice, for a total amount of 20 hours. The course focused on the same basic concepts as the Lecce course, except semantic preference. The students were also shown some 'automatic' retrieval features in *Wordsmith Tools* not presented to the Lecce students: keyword lists; the Cluster feature (which retrieves n-grams); and the Collocate feature. The topics and order in which concepts were presented was loosely inspired by Sinclair (1991), Partington (1996), and Bowker and Pearson (2002). Theoretical concepts were explained to the stu-

dents using a seminar-like approach and every topic was immediately followed by hands-on exercises. Examples and concordances to work on were given in Italian. When necessary, comparable corpora in other languages (either freely available on the Internet or provided during the course) were used to look for translation equivalents. When the students were considered sufficiently acquainted with basic corpus analysis, concepts and techniques, attention shifted to translation problems and solutions. After a brief review of the issues of polysemy/homonymy, suggestions were given about how to use comparable corpora to find translation equivalents, based on Tognini Bonelli (2001) and a simplified version of Sharoff (2004).

Finally, the students spent some hours on guided review exercises aimed at raising autonomy in the use of corpus tools. This work was carried out individually.

At the end of the course, the students were asked to carry out individual project work and hand in a paper. The following instructions were given: *Choose 3 or 4 words in one of the languages you study. For example, choose words that gave you problems in your last translation, synonyms provided by a dictionary, near synonyms whose subtle semantic differences seem difficult to distinguish, or simply 3 terms that tend to show up in the same semantic context. Analyse each word, along with their translation equivalents. Compare the information provided by your corpus/corpora with that provided by dictionaries.*

2.3. MATERIALS AND METHODS

The Genoa course produced 10 assignment papers, each one written by a different student. The Lecce course, on the other hand, produced 21 assignment papers, as most of the students worked in pairs. However, in the current study only 16 assignments from the Lecce group were considered, since at the time when the study was being carried out 5 papers were no longer available.

The student's assignments were manually marked up using a tagging scheme that was specifically and jointly developed by the two researchers. The tagging scheme focused on 12 tasks that the researchers considered of primary importance in the given assignments; theoretical and practical explanations about how to perform each of these tasks were given during the two courses. Table 2 provides a list of the tasks considered, along with the corresponding tags.

Task	Tag
Identifying collocations and collocates	<COLLOC>
Identifying colligations	<COLLIG>
Identifying the meaning of a word	<MEANING>
Disambiguating word meanings	<DIS>
Identifying semantic prosody	<PROS>
Selecting concordance lines for analysis	<LINES>
Grouping and classifying collocates or concordance lines according to semantic fields	<SEMFIELD>
Identifying phraseology	<PHRAS>
Searching for translation equivalents	<TRANSL>
Questioning the information provided by dictionaries	<DICS>
Making generalizations on the language system and using the results for stylistic or cross-cultural considerations	<GEN>
Creating or choosing a suitable corpus. (This tag was applied only in assessing the Genoa group.)	<CORPUS>

Table 2. Assessment scheme

Furthermore, the performance of each task was rated according to a scale ranging from 1 to 4 (Table 3), where: 1 indicates that the phenomenon was wrongly described or identified; 2 that the phenomenon was poorly described; 3 that the description provided was acceptable; and 4 that the description was excellent. The researchers' judgments took into consideration the focus of the assignment and how the project work was carried out as a whole.

Skill rating scale	Meaning
1	Wrong
2	Poor
3	Acceptable
4	Excellent

Table 3. Skill rating scale

The students' assignments were analysed and tagged by both researchers. This was done in order to avoid subjective marking and bias due to the fact that some students were well known to one of the two researchers. Furthermore, given that the two project work assignments were slightly different and required the students to focus on different linguistic aspects, the two researchers also considered project work specificity. Selection or creation of a suitable corpus was assessed only in the assignments of the Genoa group, since Lecce students had all worked on the same comparable corpora.

Figure 1 illustrates an example of tagging, taken from the assignment of a Genoa student: the tags <COLLIG 3> and <COLLOC3> indicate that colligation and collocation were acceptably described; tag <PHRAS4> indicates that phraseology was identified in a very clear and correct way.

At the end of the tagging process the two researchers went through all the assignments together and discussed the tasks that they had tagged differently. There were very few differences, and agreement was soon reached². Furthermore, where differences existed, they were related to the rating scale rather than to the tags. This is further proof that while checking the assignments the same criteria and assessment scheme were adopted by the two researchers.

<COLLIG3><COLLOC3><PRAS4>

	por el			
	al	Impuesto		
	el			
	en el		sobre la renta	(de)
reforma				
imponible	del			
contribuyente				
	de los	Impuestos		
	la	Imposición		

tarifa		
cuenta	del	
ley		impuesto sobre la renta
reforma		
crea	el	

(en) la declaración de la renta (de)

Figure 1. Extract from an assignment, after tagging.

Individual results were tabulated and statistical analyses were carried out on both group and collective values, using SPSS. Analyses, which are presented and discussed in Section 3, included distribution, calculation of mean and median values, ANOVA and Mann-Whitney U test³.

3. RESULTS AND DISCUSSION

The students' results are reported in the Appendix (Table A). Genoa students are numbered 101-110, while Lecce students are numbered 201-216. For each student, and for each task the following data are shown: mean value (Mean); number of observations (N); standard deviation (Std. Dev.). Furthermore, the last column shows each student's overall mean result, considering all the tasks in the assignment. Finally, the last three rows in the table refer to the whole group of students participating in the study. Table A in the Appendix shows great individual variation, both 'among students' and 'within students'. Overall individual results range from as low as 1.90 (Student 206) to as high as 3.40 (Student 101). Within-student variability is usually very high, with just one student (Student 107) showing a consistent mean in all tasks.

Group results in the different tasks are summarised in Table 4; tasks are listed in decreasing order.

Genoa		Higher results ↓ Lower results	Lecce	
Line selection	3.5		Colligation	3.1
Meaning	3.3		Translation equivalent	2.8
Question dictionaries	3.0		Collocation	2.7
Corpus	2.9		Meaning disambiguation	2.7
Meaning disambiguation	2.8		Meaning	2.5
Translation equivalent	2.8		Question dictionaries	2.5
Collocation	2.6		Semantic prosody	2.3
Phraseology	2.5		Semantic field	2.2
Generalization	2.4		Phraseology	2.2
Colligation	2.2		Generalization	2.2
Semantic prosody	2.0		Line selection	1.9
Semantic field	2.0			

Table 4: Group results

Before attempting to comment on the differences between the results of the two groups, we decided to carry out a statistical comparison, to see whether differences between the two groups could be considered significant. To this end, after assessing the distribution of both group and whole-group results⁴, we decided to apply both a parametric test (ANOVA) and a non parametric one (Mann-Whitney U test), for greater certainty.

Measure	Colloc	collig.	mean.	meaning disamb.	pros.	l. sel.	sem. field	phras.	transl. equiv.	quest. dics	gen.
Mann-Whitney U	1129	410.5	79	44	24	80.5	419	915	918	47	438
P	0.64	0.000	0.06	0.90	0.65	0.000	0.45	0.07	0.84	0.29	0.44
ANOVA F	0.16	29.53	3.95	0.06	0.44	40.78	0.42	3.77	0.00	1.75	0.58
P	0.68	0.00	0.05	0.79	0.51	0.00	0.51	0.05	0.92	0.19	0.44

Table 5: Results to the second decimal place of Mann-Whitney U and ANOVA tests

As Table 5 shows, the two tests gave the same type of results: in the vast majority of cases (eight tasks out of eleven: Collocation, Meaning Disambiguation, Semantic Prosody, Semantic Field, Phraseology, Translation equivalent, Question Dictionaries and Corpus) the difference between the two groups was not significant; this is tantamount to saying that, as far as these tasks are concerned, the two groups can be thought of as belonging to one and the same population. Thus, the results of the two groups in these tasks could probably be considered primarily due to intrinsic difficulty rather than individual and environmental factors (see Section 2.1 for a description of individual and environmental differences between the two groups). The remaining three tasks (Line Selection, Meaning, and Colligation), on the other hand, showed significant differences between the two groups ($P < 0.05$), which suggests that the difference in results is hardly due to chance. This, however, does not rule out the existence of intrinsic difficulty in these three tasks, but simply suggests that intrinsic difficulty did not emerge at group level in the current experiment.

Let us now consider whole-group mean results. Following the hypothesis that different tasks are characterized by different intrinsic difficulty levels, we can assume correspondence between lower mean results and greater difficulty of the task. The ranking in Table 6 was obtained by listing whole-group mean results in decreasing order.

	Mean	
Question dictionaries	3.0	<div style="text-align: center;"> Less difficult ↓ More difficult </div>
Colligation	2.9	
Line selection	2.9	
Corpus	2.9	
Meaning disambiguation	2.8	
Translation equivalent	2.8	
Meaning	2.7	
Collocation	2.6	
Semantic prosody	2.3	
Phraseology	2.3	
Semantic field	2.2	
Generalization	2.2	

Table 6. General Difficulty List for Corpus Analysis Task

This list can tentatively be considered a General Difficulty List for Corpus Analysis Tasks. However, it should only be taken as a preliminary hypothesis of ranking of the tasks considered, as we believe ranking should be verified in further studies on a wider population and a higher number of observations.

4. CONCLUSIONS

The current study sprang from the general observation that the results of a student in performing a corpus investigation task depend partly on the difficulty of the task itself (intrinsic difficulty) and partly on external factors, such as the student's cognitive skills, and environmental factors, including course and exam focus. The working hypothesis we formulated was that if two different groups of students showed similar difficulty in performing some analytical tasks using corpora, then, when it comes to those particular tasks, intrinsic difficulty could be considered more relevant than the influence of external and environmental factors.

Statistical analyses, which included calculation of mean results, normality test, ANOVA, and Mann-Whitney, were performed on data from 26 participants belonging to two different groups. Analyses showed that, despite the known differences between the two groups of students (environmental factors) and the existence of individual differences among the participants, Genoa and Lecce students could be considered as a single population with normal distribution, in almost all tasks. The statistical analyses also suggested that, in most of the tasks, the students' higher or lower results were probably not to be considered dependent on environmental factors, but rather on the different intrinsic difficulty of each task.

Consequently a General Difficulty List for Corpus Analysis Tasks was created using whole-group mean results. This list takes into account the difficulties encountered by the students of both groups, who were exposed roughly to the same course content, but differed in terms of level of studies, previously acquired analytical and research skills, course attended, teaching methods they were exposed to, and assignment given. Although the General Difficulty List that emerged in this study needs further verification on a wider population and a higher number of observations, we believe that such a list could be of great significance when designing courses that include the use of corpus analysis tools.

As a final rejoinder, this study leads us to suggest an analytical, rather than a holistic approach to project work assessment. In fact, while tagging our students' assignments, we noticed that our previous holistic assessments had, at times, been influenced by factors such as each student's fluency in expressing concepts, the general level of presentation of project work, and the order in which assignments were assessed. Finally, an analytical approach when assessing students' work may help avoid possible bias towards individual students based on their previous results.

Appendix

Table A: Individual mean results

student	measure	colloc.	collig.	mean.	meaning disamb.	pros.	line sel.	sem. field	phras.	transl. equiv.	quest. dics	gen.	corpus	mean per student
101	Mean	2.8	4.0	3.7	3.8		4.0		2.4	3.0	3.4	3.7	3	3.40
	N	4	1	3	5		8		5	3	10	3	1	
	Std. dev.	0.5	.	0.6	0.4		0.0		0.5	1.0	0.7	0.6	.	
102	Mean	3.0	2.0	4.0		3.0	3.5	2.0	2.8	3.4	3.0		4	3.14
	N	3	1	2		1	2	1	4	5	1		1	
	Std. dev.	0.0	.	0.0		.	0.7	.	0.5	0.9	.		.	
103	Mean	3.0	2.0	3.0	3.0		3.5	2.0	3.0	3.1	3.7	2.0	4	2.97
	N	4	3	1	3		2	2	3	7	3	1	1	
	Std. dev.	0.0	0.0	.	0.0		0.7	0.0	0.0	0.7	0.6	.	.	
104	Mean	2.9	2.0		2.0		2.7		2.7	2.8		2.0	3	2.61
	N	7.0	3.0		1.0		3.0		6.0	5.0		2.0	1	
	Std. dev.	0.4	0.0		.		1.2		0.8	0.4		0.0	.	
105	Mean	1.7	2.3	2.0	2.0	1.0	4.0		2.4	2.8	1.5	1.0	3	2.45
	N	3.0	7.0	1.0	1.0	1.0	5.0		5.0	4.0	2.0	1.0	1	
	Std. dev.	0.6	1.0	.	.	.	0.0		0.5	0.5	0.7	.	.	
106	Mean	2.3	2.0	3.0	1.0		4.0	2.0	2.3	3.0	2.0	2.0	1	2.58
	N	3.0	1.0	1.0	1.0		4.0	3.0	3.0	6.0	1.0	2.0	1	
	Std. dev.	0.6	.	.	.		0.0	1.0	0.6	0.0	.	1.4	.	
107	Mean	3.0	3.0				3.0		3.0	3.0	3.0	3.0	3	3.00
	N	1.0	1.0				1.0		1.0	1.0	2.0	2.0	1	
	Std. dev.	0.0	0.0	.	

Table A: Individual mean results (continued)

student	measure	colloc.	collig.	mean.	meaning disamb.	pros.	line sel.	sem. field	phras.	transl. equiv.	quest. dics	gen.	corpus	mean per student
108	Mean	2.0	2.0		2.0		2.0		2.0	2.0	3.0		4	2.21
	N	4.0	4.0		2.0		1.0		3.0	2.0	2.0		1	
	Std. dev.	0.0	0.0		0.0		.		0.0	0.0	0.0		.	
109	Mean	2.8	2.0	3.0			3.0	2.0	2.4	2.0	2.9	1.0	3	2.55
	N	9.0	5.0	1.0			3.0	2.0	7.0	4.0	11.0	1.0	1	
	Std. dev.	0.4	0.0	.			1.0	0.0	0.8	1.2	0.5	.	.	
110	Mean	2.3	2.7				2.0		3.0	2.3	3.0		2	2.43
	N	3.0	3.0				1.0		1.0	3.0	1.0		2	
	Std. dev.	0.6	0.6				.		.	0.6	.		0	
201	Mean		2.5	2.5		2.0	1.3	1.9				2.0		2.00
	N		4.0	2.0		1.0	3.0	7.0				1.0		
	Std. dev.		1.0	2.1		.	0.6	0.9				.		
202	Mean	3.0	2.0			2.5		1.0	3.0	2.3	1.0	2.0		2.25
	N	2.0	2.0			2.0		1.0	2.0	4.0	1.0	2.0		
	Std. dev.	1.4	0.0			0.7		.	0.0	1.0	.	0.0		
203	Mean	3.6	3.3		4.0		4.0	2.9	3.0	2.9		3.0		3.19
	N	5.0	4.0		1.0		3.0	7.0	2.0	7.0		3.0		
	Std. dev.	0.9	1.0		.		0.0	0.7	1.4	0.7		1.0		
204	Mean	3.0	3.0	1.0		2.3		1.5		3.0		1.5		1.95
	N	1.0	3.0	3.0		3.0		8.0		2.0		2.0		
	Std. dev.	.	0.0	0.0		0.6		0.5		0.0		0.7		

Table A: Individual mean results (continued)

student	measure	colloc.	collig.	mean.	meaning disamb.	pros.	line sel.	sem. field	phras.	transl. equiv.	quest. dics	gen.	corpus	mean per student
205	Mean	1.8	2.9	2.0		2.7	1.2	1.8	1.3	3.3		1.5		1.98
	N	6.0	7.0	9.0		3.0	5.0	6.0	3.0	3.0		11.0		
	Std. dev.	1.0	0.4	1.0		0.6	0.4	1.0	0.6	1.2		0.7		
206	Mean	2.5	2.8	2.0	2.3	1.7	1.8	1.6	1.3	2.0		1.4		1.90
	N	13.0	13.0	4.0	3.0	7.0	5.0	13.0	12.0	1.0		18.0		
	Std. dev.	0.7	1.0	0.8	1.2	0.5	0.8	0.7	0.5			0.6		
207	Mean	2.6	3.3	3.3		3.0		2.8	2.4			3.2		2.80
	N	7.0	10.0	4.0		2.0		17.0	16.0			5.0		
	Std. dev.	1.1	0.8	1.0		0.0		0.7	0.9			0.4		
208	Mean		3.7	4.0			2.0	2.3	3.3	3.0		3.0		2.97
	N		3.0	3.0			1.0	9.0	3.0	3.0		9.0		
	Std. dev.		0.6	0.0				1.0	0.6	1.0		0.9		
209	Mean		3.0				1.0	3.0	2.0	2.3				2.47
	N		3.0				1.0	3.0	2.0	6.0				
	Std. dev.		0.0					0.0	0.0	0.5				
210	Mean	2.4	3.0	4.0	4.0		1.0	2.2	2.5	2.7		2.0		2.36
	N	5.0	5.0	1.0	1.0		1.0	11.0	8.0	6.0		18.0		
	Std. dev.	1.1	0.0					1.0	0.8	0.5		0.9		
211	Mean	3.2	4.0			2.0		1.7	3.0	4.0	3.0	3.8		2.96
	N	5.0	2.0			1.0		6.0	4.0	2.0	1.0	4.0		
	Std. dev.	0.4	0.0					0.8	0.0	0.0		0.5		

Table A: Individual mean results (continued)

student	measure	colloc.	collig.	mean.	meaning disamb.	pros.	line sel.	sem. field	phras.	transl. equiv.	quest. dics	gen.	corpus	mean per student
212	Mean	2.5	3.0		1.0	3.0	2.0		1.8	2.3				2.28
	N	4.0	3.0		1.0	1.0	1.0		4.0	4.0				
	Std. dev.	0.6	1.0		.	.	.		0.5	1.3				
213	Mean	3.0	4.0			3.0		2.7	1.0	3.0		3.0		2.89
	N	5.0	4.0			4.0		19.0	1.0	2.0		1.0		
	Std. dev.	0.0	0.0			0.0		0.7	.	1.4		.		
214	Mean	2.5	3.2	2.0		3.0		1.8				1.5		2.23
	N	2.0	5.0	1.0		1.0		11.0				2.0		
	Std. dev.	0.7	0.4	.		.		1.0				0.7		
215	Mean	3.5	3.7	4.0	3.0			3.0	3.0	3.6	3.0	3.3		3.36
	N	2.0	3.0	1.0	1.0			1.0	3.0	5.0	2.0	4.0		
	Std. dev.	0.7	0.6	.	.			.	0.0	0.5	0.0	0.5		
216	Mean	2.0	3.0	3.0		2.0		2.3		2.5		2.3		2.29
	N	1.0	1.0	1.0		4.0		4.0		2.0		4.0		
	Std. dev.	.	.	.		0.8		1.0		0.7		0.5		
Total	Mean	2.6	2.9	2.7	2.8	2.3	2.9	2.2	2.3	2.8	3.0	2.2	2.9	2.55
	N	99.0	101.0	38.0	20.0	31.0	50.0	131.0	98.0	87.0	37.0	96.0	11	
	Std. dev.	0.8	0.8	1.2	1.1	0.7	1.2	0.9	0.8	0.8	0.8	1.0	0.94388	

Table B: Test of normality: whole-group results to the second decimal place.

group	measure	colloc.	collig.	mean.	meaning disamb.	pros.	line sel.	sem. field	phras.	transl. equiv.	quest. dics	gen.	Corpus
Whole group	Asymmetry	-0.29	-0.28	-0.20	-0.45	-0.55	-0.40	-0.002	-0.22	-0.21	-0.74	0.33	-0.66
	Std. asymmetry error	0.24	0.24	0.38	0.51	0.42	0.33	0.211	0.24	0.25	0.38	0.24	0.66

NOTES

* This study, in the form of preliminary analyses, was presented at the 7th Conference on Teaching and Learning with Corpora, TALC 2006, 1-4 July 2006, Paris.

1 The PowerPoint presentation Frankenberg-Garcia gave at the 2007 TALC Conference is available on her website, at the following address: <http://www.linguatca.pt/documentos/Frankenberg-GarciaTaLC7PowerPoint.pdf>.

2 Although there are statistical ways of measuring inter-rater agreement, we do not consider it necessary to apply them here, given the very low number of cases of initial inter-rater disagreement.

3 We thank Prof. Carla Ge (University of Pavia) for her help in performing the statistical analyses and in interpreting results.

4 Distribution was checked by means of a test of normality. Dis-

tribution asymmetry and asymmetry standard error were calculated; when the rate between these two values is 1, distribution is normal; a rate higher than 2 (in absolute value) compels us to reject the normal distribution hypothesis. Although at group level (Lecce vs. Genoa) distribution of results within individual tasks was normal for some, but not all tasks, due to the low number of observations and students (10 vs. 16), when the whole group of 26 students was considered, distribution was within the normal range for almost all tasks, except Semantic Field and Question Dictionaries. It must be said, however, that the rate between asymmetry and standard asymmetry error of Question Dictionaries is very close to the normal range limits. The results of the normality test on the whole group of students (N = 26) are reported in Table B in the Appendix.

REFERENCES

- Aston, Guy (ed.) (2001), *Learning with corpora*, Houston: Athelstan.
- Bernardini, Silvia (2000a), "Systematising serendipity: Proposals for large-corpora concordancing with language learning". In: Burnard, Lou & McEnery, Tony (eds.), *Rethinking language pedagogy from a corpus perspective*. Frankfurt: Peter Lang, 225-234.
- Bernardini, Silvia (2000b), *Competence, Capacity, Corpora*, Bologna: CLUEB.
- Bernardini, Silvia (2004), "Corpora in the classroom: An overview and some reflections on future developments". In: Sinclair, John (ed.), *How to Use Corpora in Language Teaching*, Amsterdam/Philadelphia: Benjamins, 15-36.
- Bernardini, Silvia & Zanettin, Federico (eds.) (1997), *I corpora nella didattica della traduzione*, Bologna: CLUEB.
- Bowker, Lynne & Pearson, Jennifer (2002), *Working with Specialized Language. A practical guide to using corpora*, London & New York: Routledge.
- Braun, Sabine (2005), "From pedagogically relevant corpora to authentic language learning contents", *ReCALL*, 17 (1): 47-64.
- Chambers, Angela (2005), "Integrating corpus consultation in language studies", *Language Learning & Technology*, 9 (2): 111-125.
- Cobb, Tom (1997), "Is there any measurable learning from hands-on concordancing?", *System*, 25 (3): 301-315.
- Firth, John Rupert (1957), "A Synopsis of Linguistic Theory 1930-1955", *Studies in Linguistic Analysis*, Special volume: 1-32.

- Flowerdew, John (1993), "Concordancing as a tool in course design", *System*, 21 (2): 231-244.
- Frankenberg-Garcia, Ana (2005a), "A peek into What Today's Language Learner as Researchers Actually do", *International Journal of Lexicography*, 18 (3): 335-355.
- Frankenberg-Garcia, Ana (2005b), "Pedagogical uses of monolingual and parallel concordances", *ELT Journal*, 59 (3): 189-198.
- Frankenberg-Garcia, Ana (2006), "Raising teachers' awareness to corpora, paper presented at the 7th Conference on Teaching and Language Corpora", 1st-4th July 2006, Paris.
- Gavioli, Laura & Aston, Guy (2001), "Enriching reality: Language corpora in language pedagogy", *ELT Journal*, 55 (3): 238-246.
- Gavioli, Laura & Zanettin, Federico (1997), "I corpora bilingui nell'apprendimento della traduzione". In: Bernardini, Silvia & Zanettin, Federico (eds.), *I corpora nella didattica della traduzione*. Bologna: CLUEB, 61-80.
- Gavioli, Laura (2005), *Exploring Corpora for ESP Learning*. Amsterdam: Benjamins.
- Granger, Sylviane & Tribble, Chris (1998), "Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning". In: Granger, Sylviane (ed.), *Learner English on Computer*. London: Longman.
- Halliday, Michael (1985), *An introduction to functional grammar*. London; Baltimore, Md., USA: Edward Arnold.
- Kennedy, Claire & Miceli, Tiziana (2001), "An evaluation of inter-Meante students' approaches to corpus investigation", *Language Learning & Technology*, 5 (3): 77-90.
- Levy, Mike (1990), "Concordances and their integration into a word-processing environment for language learners", *System*, 18 (2): 177-188.
- Levy, Mike (1997), *Computer-assisted language learning: context and conceptualization*. New York: Oxford University Press.
- Owen, Charles (1996), "Do concordances require to be consulted?", *ELT Journal*, 50 (3): 219-224.
- Partington, Alan (1996), *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Studies in Corpus Linguistics. Amsterdam-Philadelphia: John Benjamin Publishing Company.
- Sharoff, Serge (2004), "Harnessing the lawless: using comparable corpora to find translation equivalents", *Journal of Applied Linguistics*, 1 (3): 333-350.
- Sinclair, John (1991), *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, John (1996), "The Search for Units of Meaning". *Textus*, 9 (1): 75-106.
- Sinclair, John (2003), *Reading Concordances*. Harlow: Longman.
- Sinclair, John (ed.) (2004), *How to Use Corpora in Language Teaching*. Amsterdam/Philadelphia: Benjamins.
- Stevens, Vance (1991), "Classroom concordancing: vocabulary materials derived from relevant authentic text", *English for Specific Purposes*, 10 (1): 35-46.
- Stubbs, Michael (1996), *Text and Corpus Analysis*. Oxford: Blackwell.
- Sun, Yu-Chin (2003), "Learning process, strategies and web-based concordancers: a case study", *British Journal of Educational Technology*, 34 (5): 601-613.
- Tognini Bonelli, Elena & Manca, Elena (2002), "Welcoming Children, Pets and Guests. A Problem of non-equivalence in the languages of 'Agriturismo' and 'Farmhouse Holidays'", *Textus*, 15 (2): 317-334.
- Tognini Bonelli, Elena (2001), *Corpus Linguistics at Work*. Amsterdam/Philadelphia: Benjamins.
- Tribble, Chris & Jones, Glyn (1990), *Concordances in the Classroom*. London: Longman.
- Yoon, Hyunsook & Hirvela, Alan (2004), "ESL student attitudes toward corpus use in L2 writing", *Journal of Second Language Writing*, 13: 257-283.
- Zanettin, Federico (2002), "Corpora for translation practice". In: Yuste-Rodrigo Elia (ed.) *Language Resources for Translation Work and Research, LREC 2002 Workshop Proceedings*. Las Palmas de Gran Canaria, 10-14.
- Zanettin, Federico, Bernardini, Silvia & Stewart Dominic (eds.) (2003), *Corpora in Translator Education*. Manchester-Northampton: St Jerome.

Teaching Prosody to Italian Learners of English: Working towards a New Approach

MARIA GRAZIA BUSÀ
Università di Padova

1. INTRODUCTION: ENGLISH LANGUAGE TEACHING IN A GLOBAL WORLD

The expanding role of English as the language of international communication has important implications for English language pedagogy. Statistics show that speakers of English as a second and foreign language worldwide largely outnumber the speakers of English as a first language, so that International English can no longer be considered the 'exclusive property' of its native speakers (Ferguson, 2006; Kachru, 1985; Trudgill, 1998). At the theoretical level, the need is felt for revisiting traditional definitions of standards, norms and models to use in language descriptions and teaching, so as to integrate varieties of English spoken by non-native speakers, and developed both within well-established traditions, e.g., Indian English, and new traditions, e.g. European English (also called EuroEnglish, see Jenkins, 2006; Seidlehofer, 2004). At the applied level, new teaching practices are emerging, which represent "a 'paradigm shift' away from conventional EFL models" (Graddol, 2006: 15), and emphasize skills and abilities that can meet the evolving needs of International English users. For example, because of the large number of English non-native speakers worldwide, the view is being accredited that learners should be exposed to different varieties of English, including non-native ones.

The current debate on English as an International Language is placing pronunciation in a new position with regard to language instruction. This is due to the fact that interlanguage communication rests on the concept of mutual *intelligi-*

bility, and pronunciation is one of the main factors contributing to it (Jenkins, 2000; Seidelhofer, 2004; Munro and Derwing, 1995a; 1995b; Munro et al., 2006; Patil, 2006; Pickering, 2006). In other words, instruction should emphasize pronunciation to ensure that L2 learners' accent in English does not interfere with their ability to make themselves understood in interactions with speakers from different linguistic backgrounds. To this purpose, studies are being carried out internationally to investigate characteristics and dynamics of interlanguage communication in relation to foreign accent and intelligibility (Jenkins, 2000; Seidelhofer, 2004; Munro and Derwing, 1995a; 1995b; Munro et al., 2006; Patil, 2006; Pickering, 2006).

This paper discusses aspects of Italian pronunciation in English which may affect intelligibility, and reviews a method for the teaching of prosody which was implemented experimentally at the University of Padova, showing great potential for improving in-class and at-home pronunciation practice.

2. TRENDS IN PRONUNCIATION TEACHING

Attitudes and approaches to English pronunciation teaching have changed considerably over the past fifty years. Since the advent of the communicative approach, pronunciation has been recognized a key role in improving the learner's oral skills, and contributing to the success of oral communication (Anderson-Hsieh, 1989; Celce-Murcia, 1987). In fact, by improving the learner's oral skills, an accurate pronunciation is believed to help the learner increase self-confidence, promote social interactions outside the classroom (Cunningham Florez, 1998; Morley, 1991), and contribute to clarity and efficiency in professional exchanges (Neri, Cucchiaroni, Strik, and Boves, 2002). Also, because non-native pronunciation may be socially stigmatized and contribute to the negative stereotyping of some second-language learners, pronunciation teaching is viewed as a means to improve the learner's personal social acceptance, decreasing the odds of social or professional discrimination (Derwing, Rossiter, and Munro, 2002; Munro, 2008).

In the last decade, with the change of the role of English in the world, the objectives of pronunciation teaching have changed too. Traditionally, the aim of pronunciation teaching was the eradication of any trace of 'foreign accent' and the model varieties of English to imitate were British or American English. Today, in a world characterized by an infinite number of English accents, promoting a 'perfect English accent' appears unsustainable. Rather, the emphasis has shifted on pronunciation that is *intelligible*, and the focus on instruction is a list of features which are deemed particularly detrimental to intelligibility if non corrected (e.g., Brown, 1989; Derwing and Munro, 2005; Jenkins, 2000; Richards and Renandya, 2002; Tarone, 2005).

However, the notion of intelligibility in L2 speech is far from being well defined (Pickering, 2006), as it is highly dependent on non-linguistic factors such as degree of exposure or familiarity with a certain variety of English, listeners' attitudes towards second language speakers, etc. Research has shown that L2 learners are often unable to produce the characteristics of all individual L2 sounds and their allophonic variations, and this leads to the perception of 'foreign accent'.

But it is not clear whether it is the segmental vs. suprasegmental aspects of L2 speech which are more likely to affect L2 speakers' intelligibility.

In fact, studies in second language phonology have lagged behind studies of other aspects of learners' language, such as morphology, syntax, discourse or pragmatics (Tarone, 2005), and research has particularly neglected the phonology of second language suprasegmentals, though there is an increasing awareness that they play an important role in second-language acquisition processing, and in the perception and judgment of L2 speech (for a review, see Trofimovich and Baker, 2006). More systematic investigations are needed to examine how L2 suprasegmentals are learned, what factors influence their learning, and how they affect classroom-based and/or naturalistic interactions. This kind of research will provide essential data for language teachers to improve their teaching methods and materials, and to enhance learners' communication skills.

3. ITALIAN AND ENGLISH PHONOLOGICAL SYSTEMS: A COMPARISON

A good starting point for teaching pronunciation is to single out L2 learners' most recurrent problems and work from there.

Notoriously, one of the big problems for Italian learners of English is the production of vowels. This shows up in both the Italians' inability to differentiate words distinguished in English solely by the vowel (e.g., *sheep* – *ship*, *beg* – *bag*, etc.), which may lead to intelligibility problems, and their tendency to have a little epenthetic vowel at the end of English words, especially those ending with a stop consonant (e.g., in *did*, *big*, etc.), a feature which is often emphasized in stereotypes of the Italian accent in English. In fact, it has been shown that the production of English vowels by Italians largely correlates with Italian speakers' perceived degree of accent in English (Busà, 1995; Flege et al. 1999; Flege et al. 2003; MacKay et al. 2001; Piske et al. 2002). It would seem obvious, then, that teaching English pronunciation to Italian native speakers should start from the English vowels.

But what exactly is involved in the production of English vowels, and how can Italians be taught to produce vowel differences not existing in Italian?

Italian speakers' difficulties in producing English vowels stem from differences existing between the Italian and English phonological systems, both at the segmental and suprasegmental levels. At the segmental level, English has 11-13 vowels in its inventory, depending on the variety of English under consideration, while Italian has only seven vowels. The fewer vowels in the Italian system condition the Italian speakers' production and perception of English vowels, and lead to frequent hypo-differentiations of vowel contrasts. In addition, while in English vowels may span from full to reduced, in both quality and duration, and even disappear, depending on the degree of stress they receive in the utterance, in Italian, vowel quality tends to remain quite stable, regardless of the degree of stress on the vowel or any other phonological condition of the utterance. Thus, in English, phonological rules operating at the level of suprasegmentals (i.e., syllable structures, rhythmic tendencies, stress assignment rules, and intonation) trigger vowel reduction processes and create distinctions between vowels in

'strong' and 'weak' syllables. In Italian, these rules do not operate: syllables tend to have the same 'weight', and vowels are always fully pronounced.

The pronunciation of English vowels is not the only major problem Italian learners of English are faced with. Preliminary studies (Busà, 2008 and forthcoming) suggest that Italian learners of English may be unable to convey the appropriate English prosodic information, with a possible effect on the outcomes of their communication in English. That inappropriate prosody may affect overall communication is not surprising, given the fact that prosodic features have been found to play a major role in the production and perception of foreign accent (Boula de Mareüil, Vieru-Dimulescu, 2006; Munro, Derwing, 1995a). Also, prosody is a very important part of speech, as it conveys linguistic and pragmatic meaning. In both Italian and English, though with different modalities, prosody is used in the disambiguation of structurally ambiguous sentences, to signal the information status of an utterance (i.e., given vs. new information, emphasis or contrast, etc.), and to define speech functions (for example by differentiating between statements, questions, requests, etc.). Prosody may also convey paralinguistic information, for example with regard to the emotional state of the speaker (e.g., anger, happiness, love, etc.), the truth value of the proposition (e.g., certainty vs. uncertainty) or the level of the speakers' engagement (i.e., when the speaker is seeking support, responding to something, anticipating possible responses and objections, etc.).

A comparison of English and Italian prosodic features reveals what could be the possible pronunciation issues for an Italian speaker of English. The major differences in the Italian and English suprasegmental features are summarized below.

At the syllabic level, English has mainly CVC-type syllable structures, and allows complex consonant clusters both in syllable initial and final position; Italian has mainly CV-type syllable structures, with a distribution of long vowels in open syllables and short vowels in closed syllables, does not allow complex consonant groups in syllable-initial or final position, and only a limited set of consonants in word-final position. As far as rhythm is concerned, English has been referred to as a stress-timed language, and Italian as a syllable-timed language (for a review, see Busà 1995). This means that English will show a tendency to keep intervals between stresses equal, independently of the number of intervening unstressed syllables, by compressing sequences of unstressed syllables; Italian, on the other hand, will show a tendency to keep syllables at about constant duration, and more syllables (stressed or unstressed) will proportionally increase the duration of the sentence (Schlüter, 2005). Thus, whereas English is characterized by full vowels in stressed position and (highly) reduced vowels in unstressed position, Italian has no vowel reduction at the phonological level and limited vowel reduction at the phonetic level (Farnetani and Busà, 1999).

English and Italian also differ markedly in the way they use intonation to signal discourse information structure and focus, as well as in the intonation patterns used linguistically. In the first place, this difference concerns the relation between word order and intonation. Word order and intonation are the two most commonly used focus marking devices, and languages differ in the preference for one over the other, and in the ways in which the position of the fo-

cal constituent in the sentence may affect the intonational realization of focus (Chen et al. 2007). English has few inflections and a relatively fixed word order, and it relies heavily on intonation to convey grammatical information or focus elements in the sentence. Italian, on the other hand, has more inflections and a more flexible word order than English, and so provides its speakers with the option of giving prominence to some information by rearranging words in the sentence. In addition, English uses intonational accent (or extra stress) to mark grammatically salient elements (for example new or emphatic information) as prominent, while given or old information is de-accented; typically, focus accent in English is found on the last major word of the sentence, but can come earlier to emphasize one of the earlier words or to contrast it with something else. In Italian, prosody is not used to distinguish between new and given information, that is, givenness is not prosodically marked by deaccenting elements carrying given information or by using a particular type of pitch accent; rather, prominence is given to elements that are in focus (Avesani and Vayra, 2005; Bocci and Avesani, 2008). Thus, following the strategies of their native language, Italian speakers of English will be unable to mark salient discourse information through intonation, and show instead a tendency to either move syntactic elements around in the sentence, or use other linguistic devices (for example lexical items) to mark discourse focus.

4. TEACHING PRONUNCIATION THE TRADITIONAL WAY

The traditional methods for teaching pronunciation to L2 learners treat segments and suprasegmentals as separate component parts of pronunciation. Lessons focusing on segments typically involve simplified explanations of the articulation of individual sounds, as well as refinement of the learners' perception of the new sounds through aural discrimination exercises and minimal pair drills, that is, exercises where the learner is presented with the sounds in contrast. However, different sounds pose different types of difficulties for learners, though they are often treated in the same way in many pronunciations books. For example, articulatory diagrams are used to explain the position of the articulators during the production of a sound. Such diagrams may be a source of accurate information for the production of consonants, which require an obstruction or approximation of the articulators in specific places of the vocal tract, but they certainly do not provide much help for the production of vowels, which require the mobile articulators only to approach the fixed articulators. Unlike consonants, vowels are hard to teach because learners cannot 'feel' how to position the tongue correctly to create vowel distinctions, and instruction can only describe vowel production in relative terms. Learning the L2 vowels rests largely on perception. In this sense, minimal pair drills are highly effective in raising the learners' awareness of the differences between the L1 and the L2 vowel phonemic inventories, while diagrams are more conducive to favoring an enhancement of the learners' production for consonant sounds.

In general, even when learners do come to some understanding of the segmental differences existing between the native language and the L2, producing

them accurately in their spoken language is quite a different story. In fact, in order for perception exercises (like minimal pair drills) to be effective for improving the learners' articulation of new sounds, learners should get individual feedback on their productions, as their self-correcting ability is limited by the influence of their native language (Carey, 2005). But this is a highly unlikely situation in classrooms with only one language teacher. Generally, rather than creating new features, learners will try to revert to the features which are found in their native language. For example, Italians who are trying to create a distinction between pairs like 'bead' and 'bid' may either end up exaggerating the duration of the vowel in 'bead' to produce oppositions like [bi::d] and [bid], or, in the attempt to produce a distinction between the tense and lax vowels in the target words, revert to the closest contrast existing in Italian, that is [i] and [e], and thus produce vowels like the one in 'bid' as [bed], or sometimes even [bɛd]. Another likely outcome of pronunciation exercises focusing on segmentals is hypercorrection mechanisms, that is the over articulation of the target sounds. These may occur at the expenses of other processes which are important for successful communication in English. For example, a problem arising when the production of full English vowels is emphasized is that, anxious to produce vowel distinctions not existing in their native language, learners may pronounce all spoken vowels in their full, citation forms, rather than aiming at differentiating between full and reduced vowels. This contributes to increasing a well-know difficulty (i.e., vowel reduction processes) for learners of syllable-timed languages like Italians, who are particularly prone to pronounce all English vowels as full vowels.

It is unquestionable that a knowledge of L2 word stress and intonation patterns will help English learners to pronounce both full and reduced vowels, as well as assign a more stress-timed rhythm to sentences, giving prominence to semantically loaded words. As for sound perception, knowing that English vowels may present themselves in highly variable forms --from full to reduced, by effect of word and sentence stress and intonation, will help learners to associate actual spoken words to their citation forms, and thus will help them understand the spoken language.

Thus, because in speech, segmental and suprasegmental aspects of speech overlap and contribute to each other in many important ways, in pronunciation classes they should be taught together rather than separately. Focusing on stress, rhythm and intonation helps learners to improve their overall pronunciation, and to sound more natural, and can lead to more comprehensible speech as well as better understanding of others' speech. In fact, it has been found (Derwing, Munro, and Wieber, 1998) that learners are more likely to learn to produce segmentals effectively in spontaneous speech if they had instruction emphasizing suprasegmental features than if they only received instruction on the L2 segmentals (i.e., vowels and consonants).

However, teachers report a general difficulty teaching some features of prosody, mostly because they feel they lack the amount of competence necessary to embark in a such a cumbersome task as is required for describing all the variations in pitch, stress and rhythm part of the English prosodic system. As with all aspects of pronunciation, teaching L2 suprasegmentals requires some knowl-

edge of the L1 and the L2 prosodic systems. In addition, to make teaching more fruitful, it is important to distinguish between “what English speakers do” and “what learners of English need to learn” (Roach, 1996: 47), that is what L2 features should be taught and corrected and which do not need to be because they are not relevant for effective communication. This requires a good knowledge of how intonation in discourse functions in the L1 and L2.

Unfortunately, no systematic contrastive studies of Italian and English prosody have been carried out, and we are largely unaware of how Italians transfer their prosodic patterns into their L2 and how this affects their communication in English. Currently, some studies are being conducted at the University of Padova, which seem to be yielding promising results. The following section illustrates how the method used can both give indications of pronunciation problems for the learners of English and be used effectively in language pronunciation classes.

5. TOWARDS A NEW METHOD FOR TEACHING PROSODY TO ITALIAN LEARNERS OF ENGLISH

As seen in the section above, one of the problems with many pronunciation drills in the language classroom is that learners have no precise feedback they can rely on to compare and contrast their productions in L2 with the native speakers’ target productions. However, improvement in the L2 is strongly connected to the amount of feedback learners can get, since influences from L1 prevent them from discriminating or reproducing L2 sounds and phonological processes. In addition, as regards intonation properly, it appears difficult for learners (and for native speakers too) to gain a ‘conscious awareness’ of pitch movements, both in production and perception. This may affect their ability to hear and produce native-like intonation patterns, though they may have no difficulty producing appropriate pitch fall or rise in statements and questions, since all languages use falling and raising pitch linguistically (Chela-Flores, 2003).

To provide learners with pronunciation feedback, today’s tendency in pronunciation pedagogy is to expose learners to a wide variety of techniques, also involving multisensory models of presentation (e.g., visual, auditory, tactile, physical and affective senses, see Underhill, 2005). Some of these techniques are aimed at letting learners compare the physical characteristics of the L2 target with the closest native L1 phoneme, so that learners can get a real measure of the distance of their own productions from the target phones. By allowing learners to acquire a conscious awareness of the differences between the L1 and the L2, these techniques stimulate improvement through self-monitoring and self-correction.

A technique that is becoming increasingly popular among pronunciation teachers is the use of speech analysis software in L2 pronunciation classes as a source of audio-visual feedback for students’ productions. The use of speech analysis software allows learners to record and visualize their speech output on their computer monitors to obtain real-time information about the acoustic properties of this output. These visualizations can be used by both learners and teachers to compare and evaluate learners’ productions with those of native speakers, and learners can get an awareness of what details of their productions they

need to change to approach the native speaker's target. In addition, by recording and visualizing their pronunciation over time, learners can get a concrete idea of their progress. This method has been considered highly effective, and particularly for speech prosody, because, through it, learners can easily learn about L1 and L2 segments durations, rhythmic tendencies, reduction processes, pitch and intonation contours, since learners easily learn to associate the patterns on the display with the sounds, having only minimal knowledge of acoustic phonetics (e.g., Carey, 2005; M. Chun, 1998; De Bot, 1983; Eskenazi, 1999; Lambacher, 1996a, 1996b; Spaai and Hermes, 1993; Stibbard, 1996; Wennerstrom, 2000).

This section will show an example of how this method has been used and what results can be achieved with it.

The experiment

In some of the English language courses taught at the University of Padova, experiments are being carried out on the use of the speech analysis software *Praat* (freely downloadable from <http://www.fon.hum.uva.nl/praat/>) as a tool to aid pronunciation teaching/learning (see also Busà, 2008; and forthcoming), with particular regard to prosodic features. The aim of the experiments is to investigate the short-term and long-term benefits of the use of audio-visual feedback in pronunciation learning; the final purpose of the study is to implement a system that students can use in-class or to study autonomously. The study is also aimed at investigating how non-native, Italian intonation in English affects intelligibility.

For the experiment reported in this paper, 2 native (NS) English speakers (from Great Britain) and 8 native (NNS) Italian speakers (from the North-East of Italy) served as subjects. The subjects were asked to read aloud short dialogues which were recorded and digitized using the speech analysis software *Praat*. Some phrases were extracted for comparison and analysis (see below). The study was designed to obtain preliminary data on the differences in intonation patterns of three sentence types (open questions, yes-no questions, and salutations) by English and Italian speakers. Through the comparison of the native and non native sentences, the study aimed at getting evidence of how differences in intonation patterns may affect the intelligibility of the Italian-accented English speech.

Comparison of the intonation pattern in NS and NNS productions and improvement after audio-visual feedback

Figures 1-9 compare the intonation patterns in one of the NS' productions of an open question ('What are you doing this evening?', Fig. 1), a yes-no question ('Are you going?', Fig. 4), and a salutation ('Bye!', Fig. 7) with corresponding sentences produced by the NNs, *before* and *after* audio-visual feedback (Figs 2-3; 5-6; 8-9). The figures reproduce the type of visualizations of the speakers' utterances pitch patterns and sound waves that were obtained with *Praat*. In each figure, the upper box shows the speech sound wave, and the lower box the corresponding pitch contour.

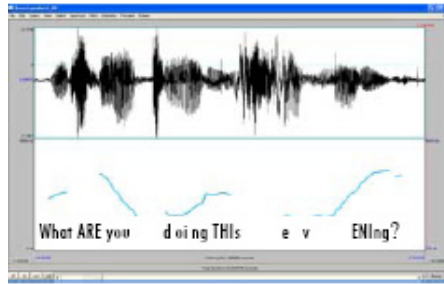


Fig. 2.: NNS' production of the sentence 'What are you doing this evening?' before audio-visual feedback

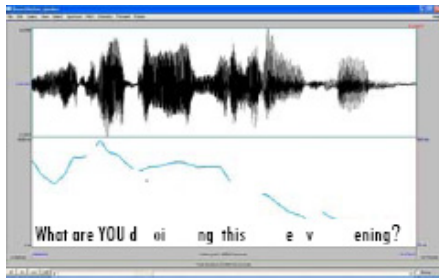


Fig. 1.: NS' production of the sentence 'What are you doing this evening?'

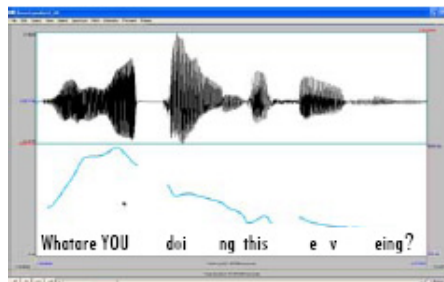


Fig. 3.: NNS' production of the sentence 'What are you doing this evening?' after audio-visual feedback

Figures 1-3 show a comparison of a NS' production of the open question 'What are you doing this evening?' (Fig. 1) and a NNS' production *before* (Fig. 2) and *after* (Fig. 3) audio-visual feedback. It can be observed that the NS has a clearly falling intonation, with a prominence peak around the word 'you'. By contrast, in Fig. 2, the NNS shows a pitch contour with three prominence peaks and a final rising intonation. After the feedback, the NNS' production closely resembles the NS', with a prominence peak on the word you and a falling intonation.

Figures 4-6 show the productions of the yes-no question 'Are you going?' by a NS (Fig. 4) and a NNS *before* (Fig. 5) and *after* (Fig. 6) audio-visual feedback. Here, again there is a clear difference in the intonation patterns produced by the NS' and the NNS' before feedback: while the NS (Fig. 4) shows a marked rising-falling contour, with a pitch peak on the word 'gOing', the NNS' utterance (Fig. 5) shows an intonation pattern which is characterized by a relatively level contour, a less prominent pitch peak than the NS, and placed in a different position in the utterances (at the onset of the vowel in gOing). On the other hand, the NNS production shows a noticeable improvement after audio-visual feedback (Fig. 6), with the production of a native-like rising-falling intonation contour, and a pitch peak at the end of the word 'you', following the prominence assignment rule of the NS.

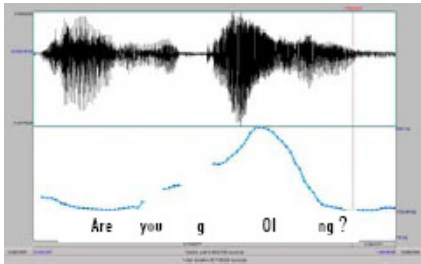


Fig. 4.: NS' production of the sentence 'Are you going?'

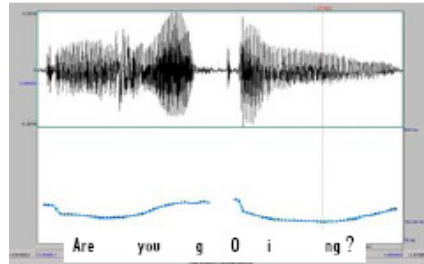


Fig. 5.: NNS' production of the sentence 'Are you going?' *before* audio-visual feedback

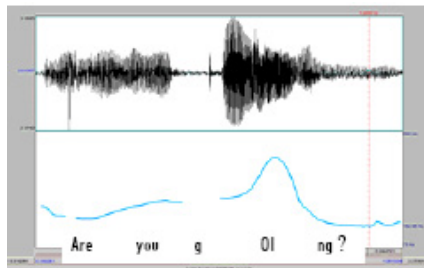


Fig. 6.: NNS' production of the sentence 'Are you going?' *after* audio-visual feedback

Finally, Figures 7-9 show the visualizations of the utterance 'Bye!'. The NS' intonation contour (Fig. 7) is first rising and then level on a vowel segment that is 704 ms long. In comparison, the Italian's utterance represented in Fig. 8 has an intonation pattern that is much more 'flat', i.e., with no clear contour or pitch peak. In addition, the Italian speaker's diphthong in 'bye' is much shorter than the NS', with a duration of 250 ms. After the audio-visual feedback (Fig. 9), the NNS' utterance shows a clear improvement in both pitch contour and vowel duration, and it resembles much more closely that of the NS.

All the examples above prove that using audio-visual feedback helps learners to improve their L2 productions and get closer to the target utterance. Working with these visualizations provides learners and teachers with an immediate and easy-to-read image of the differences existing between the L1 and the L2, and does not necessarily require much knowledge of the phonological systems of the L1 and L2. In addition, most students enjoy the hands-on experience of working with their own language and discovering facts about it.

More investigations are needed to study the effects of this method in the long term.

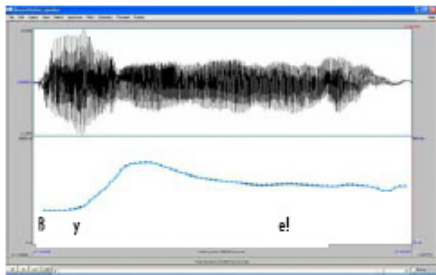


Fig. 7.: NS' production of the utterance 'Bye!'

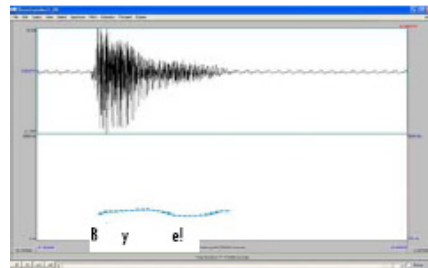


Fig. 8.: NNS' production of the utterance 'Bye!' before audio-visual feedback

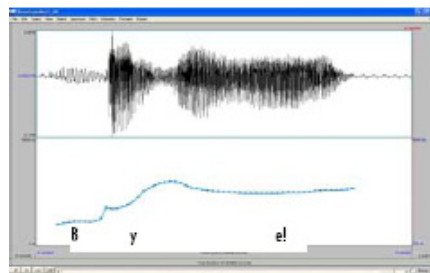


Fig. 9.: NNS' production of the utterance 'Bye!' after audio-visual feedback

6. CONCLUSIONS

The status of English as an international language calls attention to interlanguage communication, and to the dynamics which may affect interactions between speakers from widely diverging linguistic backgrounds. An unclear pronunciation may be the cause of unsuccessful communication –by being detrimental to intelligibility, or be a likely source for discrimination and prejudice, and affect the L2-speaker's social relations.

In the search for the linguistic features which affect L2 speech intelligibility, prosody has been indicated as one of the main factors, since it is used to signal information status in discourse, and to provide an interpretation to pragmatic meaning. Thus, language teachers are called upon to provide learners with practice on English prosody, to help them communicate successfully. But if prosodic features are very important in discourse, they are also very hard to teach. Perhaps this is the reason why prosody is still largely underemphasized in English language programs in spite of its recognized role in the perception and production of L2.

A promising way to address the teaching of intonation has been through the use of speech visualizing technology, which has recently become widely available. Indeed, software like Praat allows users to record, listen and visualize their own speech, and compare it with native speakers' productions. This allows learners to grow an awareness of their 'distance' from the target language and stimulates self monitoring and autocorrection.

This study has shown the kind of improvements that can be obtained with such speech-visualizing technology. The comparison of prosodic patterns in English as L1 and L2 has shown that the Italian speakers' utterances differ from the native speakers' in a number of significant features, and that audio-visual feedback helps the NNSs to improve their English prosodic patterns considerably, to the point that the latter come to resemble closely those of the NSs.

At this stage, more research is needed to investigate the long-term effects of the improvements obtained with the use of speech-visualizing technology. However, this study, in line with similar studies which are being carried out on a variety of languages, opens exciting perspectives for pronunciation teaching and learning. Prosody is becoming more accessible to the non expert, with concrete benefits for the learners. With an increased understanding of how prosody works for the L1 and the L2 speaker, teachers can help students to sound natural and be successful in their communication in English.

- Anderson-Hsieh, J. 1989. Approaches Toward Teaching Pronunciation: a Brief History, *Cross Currents*, 2, 73-78.
- Avesani, C., Vayra, M. 2005. Accenting deaccenting and information structure in Italian dialogue, *Proceedings 6th SIGdial Workshop on Discourse and Dialogue*, Lisbon, Portugal, September 2-3, 2005, 19-24.
- Bocci, G., Avesani, C. 2008. Deaccent Given or Define Focus? Where Italian does not sound like English, paper presented at the 34° *Incontro di Grammatica Generativa*, Padova, 21-23 Febbraio, 2008.
- Boula de Mareüil, P.; Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent, *Phonetica*, 63, 4, 247-267.
- Brown, A. 1989. Models, standards, targets/goals and norms in pronunciation teaching, *World Englishes*, 8, 2, 193-200.
- Busà, M.G. 1995. *L'Inglese degli Italiani. L'acquisizione delle Vocali*, Padova: Unipress.
- Busà, M.G. 2008. New Perspectives in Teaching Pronunciation, in in A. Baldry, M. Pavesi, C. Taylor Torsello, C. Taylor (eds), *From DIDACTAS to eCoLingua: an ongoing research project on translation and corpus linguistics*, Trieste: EUT, 171-188.
- Busà, M.G. forthcoming. Effects of L1 on L2 pronunciation: Italian prosody in English, *Proceedings GlobEng: International Conference on Global English*, 14-16 February 2008, University of Verona, Italy.
- Carey, M. 2005. Pronunciation Pedagogy: Historical Development and Traditional Classroom Practice. Retrieved from <http://www.ling.mq.edu.au/speech/phonetics/phonology/interlanguage/pronpedagogy.html>.
- Chela-Flores, B. 2003. Optimizing the Teaching of English Suprasegmentals, BELLS (Barcelona English Language and Literature studies, 12. <http://www.publicacions.ub.es/revistes/bells12/>
- Chun, D. 2002. *Discourse intonation in L2: from theory and research to practice*, John Benjamins, Philadelphia.
- Crystal, D. 2000. The future of English', in D. Lynch & A. Pilbeam (eds), *Heritage and Progress*, Proceedings of the SIETAR Europa Congress 1998 (Bath: LTS Training and Consulting, 2000), 6-16.
- Cunningham Florez, M.A. 1998. Improving Adult ESL Learners' Pronunciation Skills. *Center for Adult English Language Acquisition*, http://www.cal.org/caela/esl_resources/digests/Pronun.html.
- De Bot, K. 1983. Visual Feedback of Intonation 1: Effectiveness and Induced Practice Behavior. *Language and Speech* 2, 4, 331-350.
- Derwing, T.M., Munro, M.J. and Wiebe, G. 1998. Evidence in Favour of a Broad Framework for pronunciation Instruction, *Language Learning*, 48, 3, 393-410.
- Derwing, T.M., Munro, M.J. 2005. Second Language Accent and Pronunciation Teaching: A Research-Based Approach, *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect*, 39, 3, 379-397.
- Derwing, T., Rossiter, M., Munro, M.J. 2002. Teaching native speakers to listen to foreign-accented speech, *Journal of Multilingual and Multicultural Development*, 23(4), 245-259.
- Eskenazi, M. 1999. Using a Computer in Foreign Language Pronunciation Training: What Advantages? *Calico Journal* 16, 3, 447-469.
- Farnetani, E., Busà, M.G. 1999. Quantifying the range of vowel reduction in Italian, in *Proceeding of the 14th International Congress of Phonetic Sciences*, San Francisco, USA, 1-7 August 1999, 1, 491-494.
- Ferguson, G. 2006. *Language Planning And Education: Edinburgh textbooks in applied linguistics*, Edinburgh: Edinburgh University Press.

- Graddol, D. 2006. *English Next*, British Council.
- Jenkins, J. 1998. Which pronunciation norms and models for English as an International Language?, *ELT Journal*, 52(2), 119-126.
- Jenkins, J. 2000. *The Phonology of English as an International Language*, Oxford: Oxford University Press.
- Jenkins, J. 2006. 'Points of view and blind spots: ELF and SLA', *International Journal of Applied Linguistics*, 16, 137-162.
- Kachru B. 1985. Standards, codification and sociolinguistic realism: The English language in the outer circle, in R. Quirk. & H.G. Widdowson (eds), *English in the world: Teaching and learning the language and the literature*, Cambridge: Cambridge University Press/ British Council, 11-30.
- Lambacher, S. 1996a. Teaching English Pronunciation Using a Computer Visual Display. Paper presented at the IATEFL 29th International Annual Conference, York, England, 1995. Retrieved at: <http://www.u-aizu.ac.jp/~steve/york95.html>.
- Lambacher, S. 1996b. Spectrogram analysis as a tool in developing L2 pronunciation skills. *Speak Out!* London, IATEFL, 32-35.
- Morley, J. 1991. The Pronunciation Component in Teaching English to Speakers of Other Languages. *TESOL Quarterly*, 25(3), 491-520.
- Munro, M.J. 2008. Foreign accent and speech intelligibility, in J. G. Hansen & M. L. Zampini (eds) *Phonology and Second language Acquisition*, Amsterdam: John Benjamins, 193-218.
- Munro, M.J., Derwing, T.M. 1995a. Foreign Accent, Comprehensibility and Intelligibility in the Speech of Second Language Learners, *Language Learning*, 45, 73-97.
- Munro, M.J., Derwing, T.M. 1995b. Processing time, accent, and comprehensibility in the perception of foreign-accented speech, *Language and Speech*, 38, 289-306.
- Munro, M.J., Derwing, T., Morton, S.L., 2006, *Studies in Second Language Acquisition*, 28, 111-131.
- Neri, A., Cucchiari, C., Strik, H., Boves, L. 2002. The Pedagogy-Technology Interface in Computer Assisted Pronunciation Training. *CALL Journal*, 15(5), 441-467.
- Patil, Z.N. 2006. On the Nature and Role of English in Asia, *The Linguistics Journal*, 1, 2, 88-131.
- Pickering, L. 2006. Current Research on Intelligibility in English as a Lingua Franca, *Annual Review of Applied Linguistics*, 26, 219-233.
- Rajadurai, J. 2007. Intelligibility studies: a consideration of empirical and ideological issues, *World Englishes*, 26, 1, 87-98.
- Richards, J.C., Renandya, W.A. (eds). 2002. *Methodology in Language Teaching. An Anthology of Current Practice*. Cambridge, Cambridge University Press.
- Roach, P. 1996. Emotions, attitudes and the English speaker. *Speak Out!* 18, 45-58.
- Seidelhofer, B. 2001. Pronunciation, in R. Carter and D. Nunan (eds), *The Cambridge Guide to Teaching English to Speakers of Other Languages*, Cambridge: Cambridge University Press, 56-65.
- Seidelhofer, B. 2004. 'Research perspectives on teaching English as a lingua franca', *Annual Review of Applied Linguistics*, 24, 209-39.
- Schlüter, J. 2005. *Rhythmic Grammar: The Influence of Rhythm on Grammatical Variation and Change in English*, Berlin - New York: Walter de Gruyter.
- Spaai, G. and Hermes, D. 1993. A visual display for the teaching of intonation, *CALICO Journal* 10, 19-30.
- Stibbard, R. 1996. Teaching English Intonation with a Visual Display of Fundamental Frequency. *The Internet TESL Journal* 2/8, August 1996, <http://iteslj.org/Articles/Stibbard-Intonation/>.
- Tarone, E. 2005. Speaking in a Second Language, in E. Hinkel (ed.), *Handbook of Research in Second Language Teaching and Learning*, London: Routledge, 485-527.
- Trudgill, P. 1998. World Englishes: convergence or divergence?, in H. Lindquist, S. Klintonberg, M. Levin and M. Estling (eds), *The Major Varieties of English: Papers from MAVEN '97*.
- Underhill, A. 2005. *Sound Foundations*, Oxford, UK: MacMillan.
- van den Doel, R. 2007. International Intelligibility in EIL, *The Asian EFL Journal Quarterly*, 9, 4, Art. 2.
- Wennerstrom, A. 2000. The Role of Intonation in Second Language Fluency. In H. Riggensbach (ed.) *Perspectives on Fluency*, Ann Arbor, MI: University of Michigan Press.

Encoding of Information in Titles: Academic Practices across Four Genres in Linguistics

SARA GESUATO
University of Padua

1. INTRODUCTION

The title of an academic publication names, introduces, metonymically represents and advertises the content it labels, circulating from one text to the next. It is meant to be useful (informative and precise), logical (relevant to the work it names, easily classifiable and storable in databases) and reader-friendly (concise, understandable and appealing; Huth 1987; Swales and Feak 1994; Yitzhaki 1994; Busch-Lauer 2000; Dressler and Eckkramer 2001; Goodman et al. 2001; Yakhontova 2002; Hartley 2005; Lewinson and Hartley 2005; Hartley 2008). Yet, its encoding varies depending on editorial policies, individual authors' stylistic preferences, and awareness of the role and expected impact of the title on the communicative situation.

The following titles, which identify a book, a dissertation, a journal article and a proceedings paper, respectively, illustrate this variation:

- (1) "The languages of a bilingual community"
- (2) "The Acoustic-Phonetic Characteristics of Infant-Directed Speech in Mandarin Chinese and Their Relation to Infant Speech Perception in the First Year of Life"
- (3) "On the Status of Infixation and Circumfixation in English Morphology"
- (4) "Is This My Position? Teenagers' Response to Mass Media Discourse".

Titles (1), (3) and (4) are shorter than (2); (2) and (3) are richer in vocabulary and content than (1) and (4); (2) and (3) contain more denotationally precise terms than (1) and (4); (1) and (4) are easy to process, consisting of simple syntactic constituents, while (2) and (3) are more syntactically elaborate; finally, only (4) is partly formulated so as to directly involve the reader in the interaction. This linguistic heterogeneity can be attributed to the different content addressed, communicative goals and target readerships of the genres the titles are relevant to (Soler 2007: 91-92).

Books need to attract prospective readers and buyers; clear and appealing titles serve this marketing goal. Dissertations are expected to contain detailed, focused content about research carried out by fledgling academics, whose adequacy is evaluated by specially appointed examiners; their titles have to make a good first impression on these gate-keepers. Journal articles are meant to convey accurate information to interested expert peer readers, who are likely to use titles as guidelines when looking for data relevant to their work; their titles will be of help if they are non-misleading, and precise. Finally, proceedings papers typically originate as conference presentations meant to attract a large, interested expert audience; their titles contribute to this purpose by arousing the intended addressees' attention, and later on, they may retain the original formulation (cf. Busch-Lauer 2000). The different contextual expectations may call for different communicative strategies, or linguistic-textual realisations of functionally equivalent, communicative products.

More specifically, book titles that are short, divided into syntactically simple, easy-to-decode information units may be perceived as reader-friendly and thus appropriately label publications seeking commercial success. Titles of dissertations may achieve descriptive adequacy if they are exhaustive and to the point, although not necessarily appealing, easy to remember or easy to understand for the layperson: syntactic complexity, high level of information packaging and the use of jargon are appropriate choices. Journal article titles can be of use to the readers if they are highly informative and non-misleading; mild elaborateness can be tolerated for the sake of accuracy, and is unlikely to cause decoding problems to the expert, self-motivated readership. And titles of conference papers can arouse interest if they are structurally simple (easy to process), stylistically catchy (entertaining), but also lexically focused (informative).

If such linguistic-textual strategies actually recur in titles and lead to systematic differences across genres, experienced readers of academic discourse may become sensitive to them, and learn to recognize which genres given titles are associated with. In this paper I explore which systematic differences in structure, content and wording, if any, can be identified in titles of academic works in linguistics representing different genres. To this end, I overview the literature on titles of academic publications. Next I illustrate the data used for this study, and describe the analytical approach adopted. I then detail the findings. Finally, I summarise and comment on the results obtained.

2. LITERATURE REVIEW

There is a vast literature on titlelogy, including titlelogy in academic discourse. Here I overview findings from recent linguistic and, partly, scientometric publications in English on titles of scientific publications.¹

Buxton and Meadows (1977) measured the information content of research paper titles by looking at content words. They examined hundreds of journal article titles from English, French and German periodicals in several disciplines. The authors found that titles in natural sciences (especially chemistry and botany) had higher information content values than those in the social sciences (with philosophy having the lowest values). They also found that in most disciplines there had been an increase in the number of content words from 1947 to 1973, including in the English translations of the German and French titles, and that in chemistry these content words mostly identified new techniques and aspects studied. The authors attributed this finding to a widespread attributive use of nouns, but also to readers' need for easy retrievability of information as a result of ever-increasing numbers of papers being published. Finally, the authors pointed out that, independently of their information content, titles in the social sciences were less suitable to retrieval due to a lack of semi-systematic nomenclature, which is typical of the natural/hard sciences.

Dillon (1981) compared the use of colon in the titles of journal articles differing in degree of scholarship (defined in terms of dimensions abstraction of thought, protraction of endeavour and relation of entailment). His examination of 804 titles from journals in education revealed that the presence of colons in titles was the strongest in theoretical research journals (34%), less marked in empirical research journals (18%) and quite low in pedagogical journals (10%).

Dillon (1982) explored the use of colons in titles of 1,150 journal articles in education, psychology and literary criticism over a 100-year period (roughly, 1880-1980). The author noticed a steady increase in the use of colons across the disciplines, recording its first occurrence in a literary criticism journal.

Michelson (1994) examined over 2,000 titles in industrial relations journal articles. About 38% contained colons, and the presence of colons was inversely proportional to the status of the journals, as determined by their age.

Yitzhaki (1994) analysed the titles of research articles from 14 sciences, social sciences and humanities journals over a 60-year period. The author measured title informativeness as defined on the basis of number of content words, and its possible correlation with the number of authors. In the scientific fields there was a moderate positive correlation between number of authors and number of content words for most of the periods considered. However, in the social sciences the correlation was low and relevant to a minority of the titles considered. Finally, the humanities mainly displayed a negative correlation. The author attributed the positive correlation identified in the sciences to the high rate of multiple authorship in scientific papers.

Berkenkotter and Huckin (1995) examined 350 journal biology articles published between 1944 and 1989, and showed how, over time, the titles of these works had become more informative, syntactically more complex, semantically richer and more promotional in style.

Nord (1995) examined 12,000 titles and headings of various publications – including scholarly journals articles – in English, French, German and Spanish in order to describe their communicative functions, generic features, cultural conventions, structural patterns, and similarities and differences in source vs target texts in translation. She identified three essential functions of titles – relevant to the general communicative situation – and three optional functional of titles – determined by specific interactional circumstances. The first set was found to include the distinctive function, which ensures the unmistakable distinctiveness of the publication the title names, the metatextual function, meant to ensure the recognisability of the genre the publication exemplifies, and the phatic function, which involves arousing and maintaining attention in the culture-specific target audience. The second set was said to comprise the referential function, which corresponds to the understandability of the information conveyed, the expressive function, which has to do with the author's attitude toward the text labelled by the title, and the appellative function, which means the potential, envisaged attractiveness of the title for the prospective readers. Nord found that the four language- and culture-specific corpus components considered displayed the same frequency hierarchy of optional functions – in an average relationship of 100 : 30 : 6 for the referential, appellative and expressive function, respectively. But she also noticed intra-corpus differences, for instance, the highest and lowest degree of expressivity, respectively, in the German and the Spanish titles. The author also pointed out the lack of culture-specific differentiations in genres like poems and scholarly articles, the latter displaying a low proportion of appellative and expressive elements. Moreover, Nord exemplified the macrostructural types titles realise: simple title, title-subtitle combinations, duplex titles with “or” and series titles (sequences of titles belonging to different texts), and their syntactic forms: nominal, verbal, sentence, adverbial, attributive and interjection titles. Finally, the author discussed the implications of her findings for translation practice, especially the need to be loyal both to the authors' intentions in the source text and the recipients' expectations in the target culture.

Fortanet et al. (1997) analysed the structure and content of 200 titles research articles in computer science, applied linguistics, business and economics, and chemistry. The chemistry and the linguistics titles turned out to contain, respectively, the highest and the lowest number of words. Three punctuation marks (colon, semicolon, and full stop) were the most frequent in business and economics titles, and the least common in the computer science titles. Their occurrence marked the presence of two information units expressing, respectively, the general framework of the article and the specific topic, or alternatively, the topic and the method. As for content, most titles (especially in chemistry, and least of all in linguistics) conveyed the general topic and specific focus of the studies, while one third mentioned the nature of the research conducted.

Fortanet et al. (1998) examined specific linguistic features of the above-mentioned corpus of titles. The most common syntactic structure was ‘premodifier + head + postmodifier’, although combinations of heads were more frequent in linguistics and business and economics, while combinations of pre- and postmodifiers were more frequent in chemistry and computer science. Both linguistics

and business and economics displayed a majority of *-ing* forms – functioning as either nouns or verbs – while chemistry and computer science showed a balanced distribution between *-ing* and *-ed* forms – functioning as either verbs or adjectives. Finally, the linguistics and the economics and business titles favoured the use of definite articles, while the chemistry and computer science titles showed an even distribution of definite and indefinite articles.

Whissell (1999) examined over 3,000 abstracts of psychology articles from highly cited journals. Their titles turned out to be, on average, 12 words long, and to contain punctuation marks marking strong conceptual pauses about 18% of the time, which contributed to the complexity of the abstracts, together with other textual strategies.

Busch-Lauer (2000) explored the appropriateness (i.e. length, structure and communicative effectiveness) of 150 German and English titles in linguistics and medicine collected from journals articles and conference papers, and 25 English titles written by German researchers. The linguistics titles were shorter than the medicine titles (8.4 vs 9.9 words), and the German titles were shorter than the English ones. Also, the medicine titles in English (whether L1 or L2) preferred a mono-structure format, while the titles in German and in linguistics preferred a title-subtitle structure. The sequencing of syntactic constructions in title-subtitle structures was mostly nominal-nominal, although the linguistics titles more frequently instantiated verbal and clausal constructions. The most common semantic relationship between the titles and relevant subtitles was ‘general-to-specific topic’ across disciplines and languages; moreover, the linguistics subtitles were often the only comprehensible and informative component of the title-subtitle combinations. As for content, the medicine titles turned out to be precise and informative about the sub-genre of the relevant papers (e.g. case study), and about the purpose and/or results of the studies. The linguistics titles, instead, often mentioned the process of the research carried out rather than the findings, were vague and unspecified, but also creative, richer in the use of rhetorical devices, and reflective of the writers’ stylistic preferences. Overall, the medicine titles were long, precise, informative and helpful to researchers and bibliographers for their research and documentation purposes, respectively. The linguistics titles, instead, were short, vague, abstract, catchy and stylistically individualised, but less effective in helping readers trace recent research developments.

Anthony (2001) examined the length, word frequency, and preposition and punctuation usage of 600 titles representing various computer science sub-disciplines. The average title length was 9 words, most titles clustering around 6-to-12 words. On average, two-unit titles, with the colon separating them, made up about 13% of the data. The two most frequent semantic relationships holding between the two parts of a title were ‘name : description’ and ‘topic : scope’, but with considerable variation across journals. Other semantic categories identified for the title units of the corpus were ‘name of approach/algorithm/application’, ‘description of approach/algorithm/application’, ‘topic of research’, ‘scope of research’ and ‘method of research’. Statistically significant high-frequency words varied from journal to journal, reflecting their sub-disciplinary content specificity.

Goodman et al. (2001) examined the content of the titles of 420 peer-reviewed medical journal articles and queried the journals' editors on title-specific editorial policies and practices. About 40% of the titles provided information only on the topics discussed, 33% on the studies' topics and methods, 18% on their topics and results, and 2% on their topics and conclusions, while others were ambiguous. The journal editors occasionally modified titles to increase their clarity and informativity, and the only journal having a policy on titles merely addressed the issue of title length.

Yakhontova (2002) compared structural, syntactic, semantic and pragmatic characteristics of 400 titles of conference presentation abstracts representative of the disciplines of linguistics and applied mathematics in English vs Ukrainian and Russian. Various realisation patterns were instantiated in the two language-specific sub-corpora, but to different degrees. Nominative constructions dominated in the corpus, especially in the Slavic component and in linguistics; titles consisting of two parts separated by a colon were more frequent in the English component and in mathematics; and titles realised as incomplete sentences were more common in the Slavic sub-corpus, especially in linguistics. The English titles were more self-promotional from a structural, lexical and rhetorical point of view. Also, the English and Slavic titles in mathematics were more similar to each other than those in linguistics. Finally, the Slavic titles more often than the English ones highlighted the aspects of the research themes investigated and focused on the theoretical aspects of the investigations carried out.

Haggan (2004) compared the syntactic encoding, structure and content of over 700 journal article titles in literature, linguistics and education. In the three disciplines considered, similar syntactic and structural choices were made, but to different degrees: titles could consist of single information units (especially in science and linguistics) or compound ones (especially in literature), and could be formulated as complete sentences (especially in science) or phrases (especially nominal), the latter being characterised by the coordination of heads (especially in literature and linguistics) and/or their post-modification (especially in science). However, Haggan also identified important disciplinary differences in terms of content and rhetorical effects, which were independent of the structural-syntactic encoding of the titles: the science titles presented straightforward information on the findings or topics of the papers; this was probably meant to quickly and efficiently orientate the reader. The literature titles, instead, offered titillating and enigmatic hints of the content to follow, which revealed an attempt to seduce and attract the reader. The linguistics titles fell in between, showing formal features typical of science titles, but content features more similar to literature titles.

Hartley (2005, 2007a, 2007b, 2007c, 2008) and Lewinson and Hartley (2005) conducted a series of studies on thousands of titles of academic publications in various disciplines. According to their findings, scholars and students preferred titles with colons; preference for colonic titles was strong among the Arts and Social Sciences, single-authored papers and keynote address speeches; the number of titles with colons had been increasing over time; titles with colons were longer and more informative than those without; there was an even distribution be-

tween colonic titles with a longer first part (opening statement) and those with a longer second part (qualification); preference for colonic titles was similar among both highly-cited and infrequently cited papers; titles differed widely across disciplines in length (8-15 words), structure (with or without colons), and content (e.g. they could introduce a general subject, specify a precise theme, indicate a question, express the author's argument, emphasise the methodology, suggest guidelines, and attract readers through allusions, alliteration or vagueness); the nationality of the authors had little effect on title length or structure; finally, titles differed across genres, book titles being shorter and more to the point than article titles, and to prefer full sentences over sentences split by colons.

Soler (2007) examined the structure, lexicon and syntax of the titles of 480 journal review papers and 90 journal research papers in biological and social sciences. The author identified four main structural constructions, which, however, reveal a conflation of syntactic, semantic and textual classificatory parameters: nominal group, compound, full-sentence and question titles. The most common construction across disciplines and genres was the nominal group, consisting of nominal or verbal heads, possibly with pre- and/or post-modifiers. This structure was used to name, classify and describe the phenomena studied. The full-sentence construction, instead, was a generic and disciplinary peculiarity of biology research papers, and was used to present findings of experiments in a conclusive and synthetic way. The compound construction, common in the research papers and the social sciences, was used to focus attention on specific aspects of the objects of study, the first part introducing a general topic, and the second a specific one. The question construction was infrequently used, but more common among the review papers. The length of the titles varied across disciplines: the average number of words was 10.89 in the social sciences (and 7.98 in linguistics), and 14.98 in the natural sciences (and 15.48 in medicine).

Wang and Bai (2007) analysed the structure and encoding of 417 titles in medical research articles. The average length of the titles was 10.9 words, most of them being realised as nominal groups (99%), with no subtitles (98%), and characterised by the presence of single heads (75%) accompanied by post-modifying prepositional phrases (68%).

Mungra (2007) looked at the use of metaphors in the titles of the 1,426 articles published in one medical journal over a one-year period. Only 62 titles (4%) were metaphorical, and these mostly labelled editorials or opinion articles. Two main types of metaphors occurred in the titles and the body of the texts: primary ones, having a straightforward reference to the tenor or vehicle, and complex ones, formed by the blending of two input domains.

Hyland (2002) and Campagna (2008) also briefly addressed the topic of titles in academic language. In his 1.8 million word corpus of research articles, textbooks and students' essays, Hyland noticed that questions in titles only occurred in research articles in the soft fields. Campagna, who examined conference hand-outs reproducing PowerPoint slides, found that while the typographic layout of titles tended to become more varied, its content and structure was similar to that of article titles, whose information flow moved from the generic to the specific.

In conclusion, research has revealed that titles in academic discourse vary

across disciplines, sub-disciplines, languages and cultures along several dimensions: length in words, punctuation strategies, structural organisation, syntactic encoding, lexical choices, content conveyed and the semantic relationships between structural components. Titles of scholarly publications have also been found to change over time, becoming more complex syntactically, more informative semantically, and more promotional stylistically. The focus of the research on journal article titles – except for a few studies examining conference papers and conference presentation abstracts – has left unaddressed the question of whether titles of academic works differ across the genres of the publications they name. Marginal exceptions to this trend are Hartley's, Dillon's and Soler's studies, reviewed above. Hartley noticed differences in length and content between book titles and article titles. Dillon noticed that the use of colons in titles correlates with degrees of publication (i.e. colons are more common in the titles of published rather than unpublished works, and in the titles of books and articles rather than dissertation abstracts). He also observed that the presence of colons correlates with degrees of scholarship among published works (i.e. it is the strongest in the titles of theoretical publications). Soler noticed generic preferences in the structure of titles, the full-sentence and compound titles being common among research papers, and the question construction among review papers. However, no study has investigated whether titles of publications representing different genres inherit the constraints and options of those genres, that is, whether the different communicative goals that specific genres satisfy call for different communicative strategies also in the titles metonymically representing them (see section 1.). This paper addresses the issue of the possible inter-generic differences among titles within the discipline of linguistics.

3. DATA AND APPROACH

I describe 1,000 English titles of publications in linguistics, dated between 1970 and 2004, grouped into four 250-title sets. Each set exemplifies one publication type (i.e. books (BOOK), dissertations (DISS), journal articles (JOURN) and proceedings papers (PROC)) and 10 keywords (i.e. *bilingual(s); discourse; learning; morphology; phonetic/phonological; pragmatic(s); semantic; sociolinguistic; speech act(s)/lexical; syntactic/syntax*)² representing different areas and topics of investigation in linguistics.

I collected the titles from the MLA bibliography by using on-line queries. The query interface allowed me to specify the keywords in the *Title* box, the publication years in the *Years* box, and the publication type in the *Publication* box, except for *proceedings paper*. Therefore, to retrieve proceedings paper titles, I typed the search word *proceedings* in the *Anywhere* box. From each query output, I selected the first 25 relevant records, but excluding six types of titles:

(i) titles not (completely) in English; e.g.

(5) "Svenska partikelverb med in, ut, upp och ner: En semantisk "Svenska partikelverb med in, ut, upp och ner: En semantisk studie ur kognitivt

perspektiv/Swedish Phrasal Verbs with in, ut, upp och ner: A Semantic Study from a Cognitive Perspective”,

unless the non-English words or expressions identified the object of study; e.g.:

(6) “Declaring Speech Act in Conversation: A Study of Japanese Connective Datte” (2.sp-act/lex-proc);

(ii) titles not unambiguously about linguistics; e.g.:

(7) “Plastic Glasses and Church Fathers: Semantic Extension from the Ethnoscience Tradition”;

(iii) titles containing a segment not unambiguously classifiable as a subtitle; e.g.:

(8) “Knowledge of Meaning: An Introduction to Semantic Theory; Bradford Book”;

(iv) titles of atlases, bibliographies, reviews, dictionaries, anthologies, readers, encyclopaedias and multi-volume works; e.g.:

(9) “Teaching and Learning a Second Language: A Review of Recent Research” (review)

(10) “Ultra Lingua auf Deutsch: German-English Bilingual Dictionary” (dictionary)

(11) “Routledge Encyclopedia of Language Teaching and Learning” (encyclopedia);

(v) titles whose keyword was joined to another word by means of a hyphen or slash; e.g.:

(12) “Morphology-Driven Syntax: A Theory of V to I Raising and Pro-Drop” (keyword: *morphology*),

unless the query keyword itself consisted of two graphic units, possibly occurring hyphenated; e.g.:

(13) “Evidence for the Imperative as a Speech-Act Category” (16.sp-act/lex-journ; keyword: *speech act*);

and (vi) titles already included in another data set; e.g.:

- (14) “Sociolinguistic Constructs of Ethnic Identity: The Syntactic Delineation of an American Indian English” (2.socio-book; possible keyword: *syntactic*, but title already chosen for the socio-book corpus component).

I examined the length, textual organization, syntactic realizations and content of the titles. To measure length, I considered the number of words and syntactic constituents, both phrasal and clausal, making up the titles. To describe the textual organization of the titles, I examined their constituent information units, as identifiable by the presence of specific punctuation marks. The analysis of the syntactic realization of titles comprised, at a macro level, the classification of information units into phrasal vs clausal structures, and at a micro level, the identification of expansion strategies (i.e. embedding, complementation, apposition, and pre- and post-modification structures) in phrasal constituents. The examination of content involved the measurement of lexical density, and the classification of the semantic relationships between the main title and the subtitle in two-unit titles.

4. FINDINGS

4.1. LENGTH IN WORDS

I measured the length of titles in number of words. I defined words typographically, as strings of letters preceded and/or followed by spaces or punctuation marks. I thus regarded non-hyphenated compounds, capitalised abbreviations and numerical sequences as single words. Moreover, I resorted to syntactic-semantic criteria in case of hyphenated words: I regarded hyphenated sequences of strings of letters as instances of multiple words, if their constituents could function as independent units within a clause, and as instances of single words if their constituents functioned as bound morphemes; e.g.:

- (15) “You Know My Steez: An Ethnographic and Sociolinguistic Study of Styleshifting in a Black American Speech Community” (3.socio-diss; 17 words; *styleshifting*: 1 word)

- (16) “Implicatures in Discourse: The Case of Spanish NP Anaphora” (14.disc-book; 9 words; *NP*: 1 word)

- (17) “Acquisition of Spanish Verb Morphology by Bilingual Children: A Longitudinal Study between the Ages of 2;9 and 3;3” (18.bil-diss; 18 words; 2;9 and 3;3: 3 words)

- (18) “English-Learning Toddlers’ Sensitivity to Agreement Morphology in Receptive Grammar” (2.morph-proc; 9 words; *English-learning*: 2 words)

- (19) “The Effect of Character Structure on Children’s Learning of Chinese Pseudo-Characters” (2.learn-journ; 11 words; *pseudo-characters*: 1 word).

Table 1 shows the total number of title words in the title-genres considered, and the average number of words per title across title-genres. The longest titles occur in DISS (about 13 words each, on average), followed by JOURN (about 11 words), PROC (about 10 words) and BOOK (about 9 words).³

Title-genre	Total words	Average words
BOOK	2,307	9.2
DISS	3,232	12.9
JOURN	2,701	10.8
PROC	2,496	9.9
Global	10,736	10.7

Table 1: Total words in sub-corpora and average words per title

4.2. STRUCTURAL ORGANISATION

To examine the global structure of the titles, I distinguished between single-unit titles, consisting of one information unit, and multi-unit titles, comprising two or more. The presence of structural units within titles was determined by the occurrence of specific punctuation marks (i.e. colons, semicolons, full stops, question marks or dashes) marking strong internal pauses (cf. Anthony 2001: 189). However, I did not count other types of punctuation marks (e.g. commas, parentheses), or punctuation marks inside other punctuation marks, or linkers and prepositions as markers of title-internal boundaries;⁴ e.g.:

- (20) “Teaching Language, Learning Culture” (22.learn-book; 1 unit)
- (21) “The Sociolinguistic Situation of the Polish Language of the Slavic-Lithuanian Borderlands (the Region of the Present-Day Countries: Byelorussia, Lithuania and Latvia)” (24.socio-journ; 1 unit)
- (22) “Genuine Training in Academic Discourse or an Artificial Construct? Reconsidering the Past, Present, and Future of the College Research Paper” (11.disc-diss; 2 units)
- (23) “Vocal Communication in the Small-Eared Bushbaby (*Otolemur Garnettii*): Morphology, Sound Structure, and Social Context” (20.morph-diss; 2 units)
- (24) “Pragmatic Particles – Polite but Powerless? Tone-Group Terminal *hein* and *quoi* in Contemporary Spoken French” (3.pragm-journ; 3 units)
- (25) “‘I Lost the Bus: Can You Give Me a Ride Home?’: Native and Nonnative English Speakers’ Speech Act Production and Metapragmatic Judgments: A Study of Apologies, Complaints and Requests” (19.sp-act-diss; 3 units)
- (26) “A War of Words: From Lod to Twin Towers: Defining Terrorism in Arab and Israeli Newspapers 1972-1996 (2001): A Study in Propaganda, Semantics and Pragmatics” (23.pragm-book; 4 units).

Title-genre	1-unit titles	2-unit titles	3-unit titles	4-unit titles	Total units
BOOK	96 (38.4%)	151 (60.4%)	2 (0.8%)	1 (0.4%)	408
DISS	128 (51.2%)	120 (48.0%)	2 (0.8%)	0 (0.0%)	374
JOURN	144 (57.6%)	104 (41.6%)	2 (0.8%)	0 (0.0%)	358
PROC	134 (53.6%)	116 (46.4%)	0 (0.0%)	0 (0.0%)	366
Total titles	502 (50.2%)	491 (49.1%)	6 (0.6%)	1 (0.1%)	1,506

Table 2: Structure of titles

Table 2 shows the frequency and distribution of one-, two-, three- and four-unit titles in the corpus, and the total number of units per corpus component. One- and two-unit titles account for most the data. They are fairly equally distributed across the sub-corpora. BOOK prefers one-unit titles (about 60%), while the other three sub-corpora slightly favour two-unit titles. Table 3 shows that the average number of words per unit varies from a minimum of 5.6 in BOOK to a maximum of 8.6 in DISS, while JOURN and PROC have intermediate values (7.5 and 6.7, respectively).

Title-genre	Total words	Total units	Average words per unit
BOOK	2,307	408	5.6
DISS	3,232	374	8.6
JOURN	2,701	358	7.5
PROC	2,496	366	6.7
Global	10,736	1,506	7.1

Table 3: Total words, total units and average words per unit across sub-corpora

4.3. SYNTACTIC ENCODING

Title units in the corpus are realised as noun phrases, prepositional phrases, adjectival phrases, verb phrases, clauses and combinations of the above.⁵ Additionally, these basic structures can be combined through coordination, or alternatively, expanded through embedding and/or pre- or post-modification of the heads of phrasal units.

The title units realised as noun phrases mostly have nominal, but occasionally verbal or even adverbial heads (cf. Wang and Bai 2007). They can consist of one or a series of coordinated phrases,⁶ each of which can be enriched by modification or other forms of expansion; e.g.:

- (27) “But Still a Yet” (12.sem-diss; single NP with an adverbial head)
- (28) “Bilingual conversation” (10.bil-book; single NP with adjectival pre-modification)
- (29) “The Filipino Bilingual’s Competence” (15.bil-book; single NP with multiple pre-modification)
- (30) “The Development of Past Tense Morphology in L2 Spanish” (13.morph-book; single NP with post-modification)
- (31) “Rhetoric as Discourse” (19.disc-diss; single NP with complement expansion)
- (32) “The Reasons We Speak” (10.disc-book; single NP with relative clause expansion)
- (33) “A Pragmatic Logic for Commands” (20.pragm-book; single NP with pre- and post-modification)
- (34) “Hellenistic and Roman Greece as a Sociolinguistic Area” (18.socio-book; single NP with pre-modification and complement expansion)
- (35) “The Semantic Structure of Roget’s, a Whole-Language Thesaurus” (14.sem-diss; single NP with pre- and post-modification and appositional expansion)
- (36) “Control vs. Cooperation” (12.bil-diss; coordinated, unmodified NPs)
- (37) “One System or Two?” (15.bil-proc; coordinated NPs, of which one with pre-modification and the other with ellipsis of the head)
- (38) “Semitic and Indo-European, Volume II” (4.morph-book; coordinated NPs with appositional expansion)
- (39) “Emotive Signs in Language and Semantic Functioning of Derived Nouns in Russian” (20.sem-book; coordinated NPs with nominal and verbal heads, the first NP with pre-modification and the second with pre- and post-modification)
- (40) “A Pragmatic Logic for Commands” (20.pragm-book; NP with adjectival pre-modification and prepositional post-modification).

Other title units are realised as prepositional phrases, either single or coordinated; the latter are optionally characterised by the deletion of the prepositional head; e.g.:

- (41) “On Predicting Pragmatic Relations” (18.pragm-proc; single PP)
- (42) “Not by Perception Alone” (2.sem-proc; single PP)
- (43) “From the Viewpoint of ‘Perception’ and ‘Cognition’” (22.sem-journ; single PP)
- (44) “From UG to Universals” (3.learn-journ; coordinated PPs)
- (45) “On the Placement and Morphology of Clitics” (10.morph-book; coordinated PPs)
- (46) “On Learning and Teaching a Second Language” (20.learn-book; coordinated PPs).

Title units are occasionally realized as adjectival phrases; these can be single or coordinated, and may be accompanied by pre- and/or post-modification; e.g.

- (47) “Utterly Content in Each Other’s Company” (19.sem-journ; single AP with adverbial pre-modification and embedded post-modifying PP)
- (48) “Phonetic or phonological?” (8.phon-proc; coordinated APs)
- (49) “Semantic and Syntactic” (2.sem-book; coordinated APs).

Title units are often realized as verb phrases. Single verb phrases always come with some form of nominal, prepositional and/or verbal expansion; e.g.

- (50) “Making Semantic Interpretation Parser-Independent” (13.sem-proc; single VP with object NP and object complement)
- (51) “Exploring the Role of Morphology in the Evolution of Spanish” (11.morph-book; single VP with post-modified object NP and an adverbial)
- (52) “Based on Phonological and Morphological Principles” (17.phon-book; single VP with object PPs)
- (53) “Learning to read” (3.learn-diss; single VP with object VP).

In the case of coordinated verb phrases, instead, post-verbal expansion or complementation is an option; e.g.:

- (54) “Pretending and Meaning” (7.pragm-book; coordinated VPs with no expansion)
- (55) “Teaching Language, Learning Culture” (22.learn-book; coordinated VPs with object NPs)
- (56) “To Be an Actor or to Be an Observer?” (14.disc-journ; coordinated VPs with NP complementation).

The presence of a *V-ing* form as the head of a phrase required part-of-speech classification based on co-textual information. When a *V-ing* form was associated with the syntax typical of nouns (e.g. preceded by a determiner and/or adjective, and/or followed by an embedded prepositional phrase), I regarded it as the head of a noun phrase. When it co-occurred with the arguments typically required by the verb used as a predicate, I counted it as the head of a verb phrase; e.g.:

- (57) “Meaning and Time” (12.disc-diss; *V-ing* coordinated with a noun: NP)
- (58) “Speech-Language Pathologists’ Training and Confidence in Serving Spanish-English Bilingual Children” (20.bil-journ; *V-ing* with nominal pre-modification: NP)
- (59) “The Representation and Processing of Verbal Morphology in the First and Second Language” (14.morph-diss; *V-ing* with the syntax typical of nouns and coordinated with a noun: NP)
- (60) “Choosing the Right Spelling in Greek” (2.morph-journ; *V-ing* + arguments: VP).

I found only a few examples of clausal title units; e.g.

- (61) “Where Lexicon and Syntax Meet” (21.syn-book; declarative clause)
 (62) “Does Latent Semantic Analysis Actually Have a Latent Structure?”
 (18.sem-diss; interrogative clause).

Certain syntactically complex title units required more elaborate classification procedures. I classified units displaying two or more different syntactic encoding strategies as a combination of syntactic constituents; instead, I classified units compatible with more than one syntactic interpretation as ambiguous; e.g.

- (63) “Metaphoring as One Kind of Speech Act” (6.sp-act/lex-proc; double coding: VP + NP)
 (64) “Changing Economy, Changing Markets” (16.socio-diss; NPs or VPs?: ambiguous).

Syntax of title units	BOOK (%)	DISS (%)	JOURN (%)	PROC (%)	Average %
NP	357 (87.5)	325 (86.9)	299 (83.5)	300 (82.0)	84.9
VP	27 (6.7)	32 (8.6)	18 (5.0)	31 (8.5)	7.2
PP	13 (3.2)	5 (1.3)	13 (3.6)	11 (3.0)	2.7
AP	1 (0.2)	0 (0.0)	2 (0.6)	1 (0.3)	0.2
Clause	5 (1.2)	9 (2.4)	20 (5.6)	21 (5.7)	3.7
Unclear	5 (1.2)	3 (0.8)	6 (1.7)	2 (0.5)	1.0
Global	408 (100)	374 (100)	358 (100)	476 (100)	

Table 4: Frequency and distribution of units across syntactic types

Table 4 shows the frequency and distribution of units over their syntactic realizations. NP title units account for about 85% of the data, the other encoding options being less frequent. This corroborates previous findings (e.g. Haggan 2004; Soler 2007; Wang and Bai 2007).

The frequency hierarchy of the syntactic options for encoding title units is NP > VP > Clause > PP both for the corpus as a whole and for the BOOK, DISS and PROC sub-corpora. The frequency hierarchy in JOURN is slightly different: NP > Clause > VP > PP. These main syntactic structures show different distributional preferences: noun phrases are the most common in BOOK and the least in PROC; verb phrases are the most frequent in DISS and the least in JOURN; prepositional phrases are the most common in JOURN and the least in DISS; and clauses are mostly found in PROC and the least frequently in BOOK.

In two-unit titles, various combinations are attested: AP/NP, NP/NP, NP/AP, NP/PP, NP/VP, NP/Clause, PP/PP, PP/NP, PP/VP, VP/VP, VP/NP, VP/PP, VP/Clause, Clause/Clause, Clause/NP and Clause/VP. However, only NP/NP accounts for most of the data, as shown in the following frequency hierarchy:

NP/NP (71.8%) > VP/NP (9.1%) > Clause/NP (4.6%) > NP/VP (3.6%).

Table 5 shows that the distribution preferences of these combinations are heterogeneous across the sub-corpora: NP/NP is the most common in BOOK (76.8%) and the least common in PROC (64.6%); PP/NP is the most common in PROC (3.4%) and totally absent from DISS (0%); and both VP/NP and Clause/NP are the most common in PROC (12% and 6.2%, respectively) and the least common in BOOK (6.2% and 2.6%, respectively).

Syntax of 2-title units	BOOK (%)	DISS (%)	JOURN (%)	PROC (%)	Average (%)
AP/NP	0 (0.00)	0 (0.00)	1 (0.96)	0 (0.00)	1 (0.20)
NP/NP	116 (76.82)	88 (73.33)	74 (71.15)	75 (64.65)	353 (71.89)
NP/AP	1 (0.66)	0 (0.00)	0 (0.00)	1 (0.86)	2 (0.40)
NP/PP	4 (2.64)	3 (2.50)	3 (2.88)	2 (1.72)	12 (2.44)
NP/VP	6 (3.97)	7 (5.83)	2 (1.92)	3 (2.58)	18 (3.66)
NP/Clause	0 (0.00)	2 (1.66)	3 (2.88)	2 (1.72)	7 (1.42)
NP/Unclear	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/PP	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/NP	1 (0.66)	0 (0.00)	2 (1.92)	4 (3.44)	7 (1.42)
PP/VP	1 (0.66)	1 (0.83)	0 (0.00)	0 (0.00)	2 (0.40)
NP/Unclear	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/PP	1 (0.66)	0 (0.00)	1 (0.96)	0 (0.00)	2 (0.40)
PP/NP	1 (0.66)	0 (0.00)	2 (1.92)	4 (3.44)	7 (1.42)
PP/VP	1 (0.66)	1 (0.83)	0 (0.00)	0 (0.00)	2 (0.40)
VP/VP	1 (0.66)	0 (0.00)	0 (0.00)	1 (0.86)	2 (0.40)
VP/NP	10 (6.62)	13 (10.83)	8 (7.69)	14 (12.06)	45 (9.16)
VP/PP	1 (0.66)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.20)
VP/Clause	0 (0.00)	0 (0.00)	1 (0.96)	2 (1.72)	3 (0.61)
Clause/Clause	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.86)	1 (0.20)
Clause/NP	4 (2.64)	4 (3.33)	6 (5.76)	9 (7.75)	23 (4.68)
Clause/VP	0 (0.00)	0 (0.00)	0 (0.00)	2 (1.72)	2 (0.40)
Unclear/NP	3 (1.98)	1 (0.83)	2 (1.92)	0 (0.00)	6 (1.22)
Unclear/PP	1 (0.66)	0 (0.00)	0 (0.00)	0 (0.00)	1 (0.20)
Unclear/VP	0 (0.00)	1 (0.83)	0 (0.00)	0 (0.00)	1 (0.20)
Total	151 (100)	120 (100)	104 (100)	116 (100)	491 (100)

Table 5: Syntactic encoding of two-unit titles

4.4. SUB-PHRASAL SYNTAX

An examination of phrase-internal syntax in the title units reveals that pre-modification is a common strategy. Given the prominence of noun phrases, pre-modification is often applied to nominal heads, which can be qualified by nouns, verbs and/or adjectives in various combinations.⁷ The frequency hierarchy of nominal pre-modification is DISS (611) > JOURN (525) > PROC (505) > BOOK (426); e.g.:

- (65) “Karelian-Russian Language Alternation” (4.bil-book; pre-modifier: compound adjective)
- (66) “Semantically Ambiguous Words” (19.sem-diss; pre-modifiers: adverb + adjective)
- (67) “Arab and American Panel News Interviews” (25.disc-diss; pre-modifiers: coordinated adjectives and compound noun)
- (68) “Chinese Indonesian Mother-Daughter Pairs” (24.socio-diss; pre-modifiers: adjectives + compound noun)
- (69) “Web-Based Language Learning System” (1.learn-proc; pre-modifiers: noun + verb form (adjectival compound))
- (70) “ESL Korean Learners’ Decision-Making Processes” (17.sp-act/lex-journ; combination of pre-modifiers)
- (71) “The Changing Sociolinguistic Status” (18.socio-journ; combination of pre-modifiers)
- (72) “Mixed Language Varieties” (25.disc-journ; combination of pre-modifiers).

Similar forms of pre-modification apply to the verbal heads of noun phrases. Their frequency hierarchy is DISS (27) > JOURN/PROC (26) > BOOK (18); e.g.:

- (73) “Saudi Arabic-English Intrasentential Codeswitching” (17.syn-diss; pre-modifier: adjective)
- (74) “Language Switching” (22.bil-journ; pre-modifier: noun)
- (75) “Foreign Language Learning” (11.learn-book; pre-modifiers: adjective + noun)
- (76) “Data-Driven Learning” (3.learn-proc; pre-modifiers: noun + verb form).

The most frequent pre-modification types are ‘single adjective’ (55.5%), ‘single noun’ (18.8%), ‘sequence of adjectives’ (11.8%), ‘single adjective and single noun’ (8.2%) and ‘sequence of nouns’ (5.7%). Their distribution varies across the sub-corpora: BOOK displays the most limited occurrences of all pre-modification types; DISS has the highest number of three types, while JOURN and PROC have intermediate frequency values for two pre-modification types (i.e. ‘single adjective’ and ‘single adjective + single noun’). Table 6 visually summarises these findings.

Pre-modification	BOOK	DISS	JOURN	PROC	Global (%)
1Adj+ <u>N</u>	239	262	245	242	988 (55.5)
≥ 2Adj+ <u>N</u>	44	61	48	57	210 (11.8)
1Adj+1N+ <u>N</u>	28	50	39	29	146 (8.2)
1N+ <u>N</u>	51	97	82	105	335 (18.8)
≥ 2N+ <u>N</u>	19	27	30	25	101 (5.7)
Total	381	497	444	458	1780 (100)
Percentage	21.4%	27.9%	24.9%	25.8%	100%

Table 6: Most frequent types of pre-modification (heads are underlined)

Like pre-modification, post-modification typically qualifies the nominal and verbal heads of noun phrase title units. Post-nominal modification comprises the use of single or coordinated nouns, prepositional phrases or equivalents, and reduced relative clauses headed by *V-ing* or *V-ed* forms.⁸ Moreover, pre- and post-modification can co-occur; e.g.:

(77) “The Pronouns *rsw* and *rsaccaw*” (4.socio-proc; post-modifier: coordinated nouns)

(78) “the Understanding of Verbal Irony” (15.pragm-journ; post-modifier: PP)

(79) “Language Learning as Social Modeling in the Northwest Amazon” (4.learn-journ; post-modifier: PP equivalent)

(80) “A Case Study Comparing Quebecois in Montreal and Texas Spanish in San Antonio” (17.bil-diss; post-modifier: *V-ing*-headed reduced relative clause)

(81) “The Modifying Strategies Used by Deaf Students in the Speech Act of Apologizing” (23.sp-act/lex-diss; post-modifier: *V-ed*-headed reduced relative clause)

(82) “English Spatial Prepositions by, on and into” (22.bil-diss; pre- and post-modification).

Title units with *V-ing* heads having a predicative function can be accompanied by nominal or clausal complementation. There are 165 occurrences of such complementation; they are twice as frequent in DISS (54) as in BOOK (23), while JOURN and PROC display intermediate frequency values (47 and 41, respectively); e.g.:

(83) “Understanding What Is Said and What Is Implicated” (24.pragm-diss; post-verbal clausal complementation)

(84) “Positioning Gender in Discourse” (7.disc-book; post-verbal nominal complementation)

(85) “Learning How to Do Things with Words in a Study Abroad Context” (5.learn-book; post-verbal clausal complementation).

Another form of expansion of phrasal title units is coordination. This usually combines phrases with heads of the same word class; e.g.:

- (86) “Polite but Powerless?” (3.pragm-journ; coordinated APs)
- (87) “Syntax, Information Structure and Intonation” (7.syn-book; coordinated NPs)
- (88) “From the Japanese Case to a General Sociolinguistic Perspective” (17.socio-book; coordinated PPs)
- (89) “Emblematizing or Stereotyping?” (25.disc-proc; coordinated VPs)
- (90) “Phonological Phrasing and Syntactic Derivation” (19.syn-diss; coordinated NPs, of which one with a *V-ing* head).

Post-modification and coordination	BOOK	DISS	JOURN	PROC	Total
<u>NP</u> +1PP	129	90	116	112	447
<u>NP</u> + ≥ 2PP	64	137	84	89	374
<u>NP</u> <u>1NP</u>	42	36	27	42	147
<u>NP</u> <u>1NP</u> +1PP	12	10	16	16	54

Table 7: Most frequent types of post-modification and coordination (heads are underlined)

Table 7 shows that the most common forms of post-modification are single prepositional phrases or sequences of prepositional phrases (cf. Wang and Bai 2007: 395), and that the most common forms of co-ordination are sequences of two noun phrases, optionally accompanied by a prepositional phrase. The former are much more frequent than the latter (821 vs 201 occurrences, respectively). Their distribution patterns vary, with one form (i.e. multiple prepositional post-modification) being the most and the least frequent, respectively, in DISS and BOOK.

Expansion form / Head	BOOK	DISS	JOURN	PROC	Total
Pre-modification / Adj	0	0	0	1	1
Pre-modification / N	426	611	529	505	2,071
Pre-modification / V-ing	18	27	26	26	97
Pre-modification / Other	0	1	0	0	1
Post-modification / Adj	0	0	2	2	4
Post-modification / N	220	267	242	233	962
Post-modification / O	7	5	16	6	34
Complementation / V-ing	23	54	47	41	165
Coordination / Adjs	1	0	1	1	3
Coordination / Ns	70	75	64	68	277
Coordination / PPs	7	1	4	1	0
Coordination / V-ings	4	3	0	1	8
Coordination / Other	4	7	16	3	30
Other	2	2	7	0	11
Total	782	1,053	954	888	3,677
Average per title	3.0	4.0	3.6	3.4	

Table 8: Distribution of expansion strategies

Table 8 summarises the details of the distribution of expansion forms (i.e. pre-modification, post-modification, coordination and complementation strategies) over the main word classes. Pre-modification and post-modification of nominal heads are the most frequent and the second most frequent expansion strategies, respectively, in the sub-corpora. The frequency hierarchy of four of the most frequent expansion strategies (i.e. pre-modification of nouns and *V-ings*, post-modification of nouns, and complementation of *V-ings*) is DISS > JOURN > PROC > BOOK. However, the frequency hierarchy of coordinated nouns is DISS > BOOK > PROC > JOURN.

4.5. LEXICAL DENSITY

Lexical density is the amount of information conveyed as a function of the number of content words employed; the assumption is that the higher the density of a discourse excerpt, the higher its informativity. Lexical density can be measured as the overall ratio of content words to function words, or as the frequency of content words per ranking (i.e. non-embedded) clause. In a corpus of titles, however, which mostly consists of sub-clausal units, it makes more sense to consider title units as the genre-specific equivalent of ranking clauses.

Assuming that, as written texts, the titles would be richer in content words than function words, I counted the latter. I regarded as function words occurrences of articles, conjunctions, prepositions, pronouns, possessives, *wh*-ques-

tion words, demonstratives and auxiliaries, unless these were mentioned as topics of research. I applied the same principle when similar reference was made to bound morphemes and phonemes; e.g.:

(91) “Universal Semantic Primes, and Their Application to French Monolingual and English-French Bilingual Lexicography: English Spatial Prepositions by, on and into in French Translations of ‘Alice’s Adventures in Wonderland’” (22.bil-diss; 8 function words).⁹

Word counts	BOOK	DISS	JOURN	PROC	Global
Content words	1,574	2,207	1,852	1,731	7,364
Function words	733	1,025	849	765	3,372
Percentage of content words	68.2%	68.3%	68.6%	69.4%	68.6%
Average content words per title	6.3	8.8	7.4	6.9	7.4
Average function words per title	2.9	4.1	3.4	3.1	3.4
Average content words per title unit	3.9	5.9	5.2	3.6	4.9
Average function words per title unit	1.8	2.7	2.4	1.6	2.2
Total words / content words	1.5	1.5	1.5	1.4	1.5
Total words / function words	3.1	3.2	3.2	3.3	3.2
Content words / function words	2.1	2.2	2.2	2.3	2.2

Table 9: Measures of lexical density

Table 9 shows the total number of content and function words and various measures of lexical density. Content words outnumber function words both in the whole corpus and the sub-corpora. On average, there are about 7.3 content words per title and 4.8 content words per title unit. DISS and JOURN, however, have higher than average lexical density values per title and title unit.

4.6. CONTENT ANALYSIS

To approach a description of the content of the titles, I used an intuitive, bottom-up approach. That is, I repeatedly read the titles until I started to notice a general pattern, namely that the information conveyed was relevant to two elements of the studies identified by the titles: the context of the research, including the phenomena investigated, and the research itself, or the way it had been carried out. After classifying the content of the titles into these two broad categories, I noticed that contextual information comprised reference to the spatio-temporal

setting of a linguistic phenomenon, the relevant language, the texts (or other data sources) examined, and the speakers involved, including their characteristics (e.g. problems, developmental stages, interactional scenarios and communities of practice). In coding the data, I classified each contextual element separately, while I grouped together the above-mentioned speaker-specific characteristics under the label ‘context: other’; e.g.:

- (92) “Punjabi, Urdu, English in Pakistan: A Sociolinguistic Survey” (12. socio-book; context: spatial setting)
- (93) “Triglossia and Pragmatic Variety Choice in Nineteenth-Century Bruges: A Case Study in Historical Sociolinguistics” (11. pragm-journ; context: temporal setting)
- (94) “Aspects of the Syntax, the Pragmatics, and the Production of Code-Switching Cantonese and English” (24. pragm-book; context: languages)
- (95) “Cross-Language Blending of /l/ Gestures by Korean-English Bilingual Children” (3. bil-proc; context: speakers)
- (96) “Genuine Training in Academic Discourse or an Artificial Construct? Reconsidering the Past, Present, and Future of the College Research Paper” (11. disc-diss; context: text)
- (97) “A Comparison of Three Interventions for Children with Co-Occurring Pragmatic Language and Behavior Problems” (2. pragm-diss; context: other: speakers’ problems)
- (98) “Imitation as a Basis for Phonetic Learning after the Critical Period” (9. phon-proc; context: other: speakers’ developmental stages)
- (99) “Exploring Linguistic Differences in Academic Cultures” (22. disc-proc; context: other: communities of practice)
- (100) “Discourse Analysis at the Interface of Politics and the Media: Election Night Coverage” (24. disc-proc; context: other: interactional scenario).

Table 10 shows the frequency of occurrence of the most common context-relevant information units across corpus components and category types. ‘Language’ and ‘speaker’ are the notions most frequently referred to in the titles, followed by ‘text’, and then ‘place’. The distribution hierarchy of information units about context is DISS > JOURN > PROC > BOOK.

Information unit	BOOK	DISS	JOURN	PROC	Global
Place	30	39	32	15	116
Language	91	107	97	82	377
Speaker	29	77	62	64	232
Text	40	56	36	28	160
Other	4	58	27	20	109
Total	194	337	254	209	994

Table 10: Information units about context with > 10 occurrences in one or more sub-corpora

When the title units contain information about the studies they name, the content expressed can have an appealing or promotional function or be clouded in vagueness (cf. Busch-Lauer 2000; Yakhontova 2002; Haggan 2004); e.g.:

(101) “Give Syntax a Chance” (18.syn-journ; study: alluring/advertising function)

(102) “Negotiating Bilingual and Bicultural Identities: Japanese Returnees betwixt Two Worlds” (1.bil-book; study: vagueness)

(103) “Complaining and Commiserating: A Speech Act View of Solidarity in Spoken American English” (4.sp-act/lex-book; study: vagueness).

Much more often, however, research-focused titles do orientate the reader, by providing information about the linguistic fields relevant to the studies carried out, the domains investigated, specific facets of the linguistic phenomenon examined, the methods adopted in carrying out the studies (e.g. type of investigation, description of the project, the process of research, approach, implications or results), the theories used as background frameworks for the studies, and/or the work of individual scholars (cf. Busch-Lauer 2000; Yakhontova 2002; Haggan 2004); e.g.:

(104) “Creolistics and Sociolinguistic Theories” (19.socio-journ; study: field)

(105) “Expressivity and a Pragmatic Constraint on Object Reduplication in Bulgarian” (21.pragm-journ; study: domain)

(106) “Sex Differences in Voice Onset Time: A Developmental Study of Phonetic Context Effects in British English” (7.phon-journ; study: facet)

(107) “Blue in Old English: An Interdisciplinary Semantic Study” (7.sem-book; study: method)

(108) “Speech Act Taxonomy as a Tool for Ethnographic Description: An Analysis Based on Videotapes of Continuous Behavior in Two New York Households” (7.sp-act/lex-book; study: method)

(109) “Derivations: Exploring the Dynamics of Syntax” (11.syn-book; study: method)

(110) “Phonetic Features in Language Production: An Experimental Examination of Phonetic Feature Errors” (13.phon-diss; study: method)

(111) “Investigating Semantic Inhibition Using a Modified Independent Cue Task” (8.sem-diss; study: method)

(112) “Connectionism, Language Production and Adult Aphasia: Elaboration of a Connectionist Framework for Lexical Processing and a Hypothesis of Agrammatic Aphasia” (22.sp-act/lex-book; study: theory)

(113) “A Short Introduction to X-Bar Syntax and Transformations” (4.syn-book; study: theory)

(114) “Brilliance, Energy and Size in Vowels: A Cross-Linguistic Study of Phonetic Symbolism” (15.phon-diss; study: theory)

(115) “Investigating the Neuropsychological Bases of Script Knowledge: Differential Effects of Executive Dysfunction and Semantic Impairment”

- in Dementia” (1.sem-diss; study: theory)
 (116) “Pace Panini: Towards a Word-Based Theory of Morphology” (19.morph-book; study: scholar)
 (117) “Syntactic Theories and Syntactic Methodology: A Reply to Seuren” (2.syn-journ; study: scholar).

A specific type of information about the studies consists in reference to their technical aspects, or to components of the phenomena studied or of the relevant investigations; these technical elements are identified by means of nominal or adjectival expressions; e.g.

- nominal: *alignment, analogy, causative, ergative, evidentiality, factivity, grammaticalisation, grapheme, implicature, infixation, logophority, minimal pair, move-structure, onset, phonotactics, reduplication, register, schemata, semantic field, stem, transformation, umlaut, variable*;

- adjectival: *constraint-ranking, corpus-driven, derivational, diachronic, electrophysiological, paleographic, pragmalinguistic, productive, right-branching, scalar, typological*); e.g.

- (118) “Speech Act Schemata and Discourse Type” (3.sp-act/lex-proc; technical term: nominal)
 (119) “Modeling Syntactic Constraints on Anaphoric Binding” (17.syn-proc; technical terms: adjectival and nominal)
 (120) “Designing a Corpus-Based Grammar for Pragmatic Terminographic Definitions” (9.pragm-journ; technical terms: adjectival).

Content units about the studies are fairly equally distributed across the sub-corpora (see Table 11). They are the most common in PROC (550 occurrences) and DISS (549), less so in BOOK (522) and the least in JOURN (483). The most frequent content unit is ‘Technical’. The second and third most frequent ones are ‘Method’ and ‘Facet’ in BOOK and DISS, and ‘Facet’ and ‘Domain’ in JOURN and PROC. The overall frequency hierarchy of information units about research is given below:

‘Technical’ (793) > ‘Facet’ (384) > ‘Method’ (346) > ‘Domain’ (279) > ‘Theory’ (177) > ‘Vague’ (86) > ‘Field’ (20) > ‘Advertising’ (19).

Information unit	BOOK	DISS	JOURN	PROC	Global
Advertising	3	3	10	3	19
Vague	20	21	27	18	86
Field	2	24	1	2	20
Domain	44	51	84	100	279
Facet	71	100	108	105	384
Method	115	104	57	70	346
Theory	39	39	42	57	177
Technical	228	229	151	185	793
Total	522	549	583	550	2,104

Table 11: Information units about studies with >10 occurrences in one or more sub-corpora

4.6.1. INFORMATION SEQUENCING

I examined the sequencing of information in two-unit titles, that is, the types of semantic relationships holding between the first and second part of such titles, by drawing on my own classification of content units reported above (see section 4.6.) and on findings from previous studies, especially Swales and Feak (1994) and Busch-Lauer (2000). I identified four main types of information suitable for classifying the global content of title units: on the one hand, 'topic', 'method' and 'theory', relevant to the content of the studies, and on the other, 'context', relevant to the context of the studies. I also found it necessary to resort to three more content categories to deal with titles conveying the same type of information in both their first and second part. I classified the relevant information units as 'paraphrase' (if the same content was conveyed through alternative expressions), and 'general topic' or 'specific topic' (if the same type of content was being presented in broader vs more specific terms, respectively).¹⁰ In the following examples, the hyphen (-) signals 'sequencing of information units in the first and second part of the title', while the plus sign (+) signals 'combination of content units in the same title unit':

(121) "Japanese Pidgin English in Hawaii: A Bilingual Description" (25.bil-book; context - method)

(122) "Literacy Learning in a Bilingual Classroom for Deaf Students: Negotiating between New Zealand Sign Language and English" (20.bil-diss; topic - context)

(123) "Symbolic Values of Foreign Language Use: From the Japanese Case to a General Sociolinguistic Perspective" (17.sem-book; topic - context+method)

(124) "Bilingual Instruction of Immigrant Children: A Theoretical Overview and Results from Empirical Research" (8.bil-book; topic - method)

(125) "Language Choice, Language Attitudes and Ethnic Identity in Bilingual Speakers: A Case Study Comparing Quebecois in Montreal and Texas Spanish in San Antonio" (17.bil-diss; topic - method+context)

(126) "Semantic Structure of Spanish: Meaning and Grammatical Form" (13.sem-book; general topic - specific topic)

(127) "Bilingual Education: Theories and Issues" (14.bil-book; topic - theory)

(128) "The Pragmatic Nature of Theatrical Discourse: The Performance-Text as a Macro Speech Act" (23.pragm-journ; topic - paraphrase)

(129) "Sociolinguistic Analysis of the Language Preferences of Adolescent-Bilinguals: Shifting Allegiances and Developing Identities" (8.socio-journ; specific topic - general topic).

Table 12 shows the frequency and distribution of the five most frequent sequencing patterns relevant to two-unit titles, accounting for 79% of the data. The frequency hierarchy of such sequences in two-unit titles reveals a high incidence of ‘general-specific’ semantic relationships (cf. Busch-Lauer 2000; Haggan 2004), followed by ‘topic – method’ in BOOK (26), DISS (16) and JOURN (13), and ‘topic - method+context’ in PROC (13):

‘general-specific’ (212) > ‘topic-method’ (63) > ‘topic-context / topic-method+context’ (40) > ‘specific-general’ (33).

Information unit	BOOK	DISS	JOURN	PROC	Global
Topic – context	91	2	11	8	40
Topic – method	26	16	13	8	63
Topic – method+context	7	16	4	13	40
General – specific	57	57	51	47	212
Specific – general	8	4	10	11	33
Total	107	105	89	87	388

Table 12: Five most frequent sequencing patterns in two-unit titles

5. DISCUSSION AND CONCLUSION

Titles of academic publications count as *texts* (Nord 1995: 280; Haggan 2004: 312): they are cohesive and coherent self-contained discourse units, which are meant to satisfy a specific communicative purpose (i.e. to label and succinctly describe publications), and which can be removed from their context of production to be used in a different context of reception (e.g. for citation purposes). Despite being short, titles are *important* texts content-wise: in summarising the content of the publications they name, they affect the reader’s first understanding of those publications. Also, titles are *consequential* texts at the level of writer-reader interaction: being the first, and visually prominent, components of larger texts, they determine the reader’s first impressions of those texts, and thus influence the reader’s decision of whether to read more of them (Bazerman 1985; Busch-Lauer 2000: 77; Day 1994: 15; Swales 2003: 179; Wang and Bai 2007: 389; Hartley 2007c: 554). Moreover, from the perspective of documentation and information science, titles are the *major* means of cataloguing and retrieving scientific literature (Busch-Lauer 2000: 90; Soler 2007: 91; Wang and Bai 2007: 389). Finally, titles are *demanding* texts for their writers: they have to be concise (due to space limitations), informative and/or appealing (to orientate and attract the reader) and acceptable (i.e. to conform to cultural, linguistic, generic, disciplinary and audience expectations).

The situational options and constraints of titles determine their shared characteristics such as economy (necessary to summarise and label other texts), nominal encoding (fit for naming entities) and informativity (use of content words, and of pre- and post-modification; Haggan 2004; Soler 2007: 97-98; Wang and Bai 2007), which are instantiated independently of the titles' relevance to specific languages, cultures, disciplines or genres. Yet, as contextually situated texts, titles also reflect the communicative needs, goals and peculiarities of the disciplines, sub-disciplinary fields, studies and texts they are relevant to (Anthony 2001; Fortanet et al. 1998), and thus differ across languages, cultures, disciplines and genres (see section 2.).

The above analysis has revealed similarities and differences among titles in linguistics across publication categories. The similarities outweigh the differences, and include: high lexical density; preference for one or two information units; sequencing of information in two-unit titles as a transition from a general topic to a specific one; frequent syntactic encoding of title units as noun phrases; high frequency of coordinated noun phrases, of adjectival pre-modification of nominal heads, and of prepositional post-modification of nominal heads; and frequent reference to the languages and technical aspects of the studies named by the titles. Such similarities corroborate previous findings, and are attributable to the titles' common genre and discipline membership, linguistic encoding, and communicative goals (i.e. to inform and guide, clearly, precisely and economically).

Among the differences that have been detected, some are systematic, that is, characterised by consistently ordered frequency hierarchies (i.e. DISS > JOURN > PROC > BOOK) in the distribution of certain features: the length of the titles in number of words (per sub-corpus, title and title unit), and the occurrence of total expansions, pre-modification strategies, post-modification resources, function words and information units about context. Other features, however, are non-systematic, revealing variable cross-genre distributional preferences, as is the case with number of total units, frequency of one-title units, occurrence of coordination or post-modification strategies, and sequencing of information units in two-unit titles. The following frequency hierarchies succinctly illustrate these divergent patterns:

- Title units:	BOOK > DISS > PROC > JOURN
- One-unit titles:	JOURN > PROC > DISS > BOOK
- Syntactic encoding of title units as NPs:	BOOK > DISS > JOURN > PROC
- Syntactic encoding of title units as VPs:	DISS > PROC > BOOK > JOURN
- Syntactic encoding of title units as Clauses:	PROC > JOURN > DISS > BOOK
- Post-modification strategies alone:	DISS > PROC > BOOK > JOURN
- Coordination strategies alone:	PROC > BOOK > DISS > JOURN
- NP/NP sequence:	BOOK (76.8%) > DISS (73.3%) > JOURN (71.1%) > PROC (64.6%)
- VP/NP sequence:	PROC (12.0%) > DISS (10.8%) > JOURN (7.6%) > BOOK (6.6%)
- Clause/NP sequence:	PROC (7.7%) > JOURN (5.7%) > DISS (3.3%) > BOOK (2.6%)
- NP/VP sequence:	DISS (5.8%) > BOOK (3.9%) > PROC (2.5%) > JOURN (1.9%)

It appears that the sub-corpora display fairly distinctive traits, but not totally clear-cut differences. BOOK prefers relatively short titles, often divided into two parts, each encoded as a noun phrase, and characterised by limited pre- and post-modification of head words in title units. These titles convey short, easy-to-decode and reader-friendly chunks of information units.

At the other end of the continuum, DISS prefers relatively long titles, with either one or two precise, detailed, elaborate information units. DISS shows the highest frequency of title units encoded as verb phrases, a strong preference for multiple post-modification of head words and high frequency of function words.

In between BOOK and DISS are the JOURN and PROC titles. The JOURN titles are shorter than the DISS ones, but longer than the PROC ones. Like those in DISS, they tend to consist of single units. As for their content and sequencing of information units, the JOURN titles are similar to the PROC ones, but their expansions (i.e. internal syntactic elaboration) are less frequent than in PROC. Overall, they are highly informative and moderately elaborate.

Finally, PROC comprises titles that are shorter than in JOURN, but longer than in BOOK. They display a slight preference for one-unit titles, in this more closely resembling DISS than BOOK. They are similar to the JOURN titles regarding types of information units. Surprisingly, their expansions are more frequent than in JOURN, and more generally, they are not as informal as expected (cf. Busch-Lauer 2000; Yakhontova 2002; Haggan 2004, reviewed in section 2.). A reason for this may be that the titles in this group came from published proceedings, and not from conference oral presentations; it may be that they were modified from their original formulation (contra Busch-Lauer 2000). Overall, the PROC titles are structurally simple, but moderately elaborate from a syntactic point of view, that is, easy to process and informative.

This study has explored intra-disciplinary and cross-generic similarities and differences in the encoding of titles of academic publications. Part of the findings – shared characteristics and distributional differences of certain features – are in line with previous findings, and can be plausibly explained with reference to the common communicative goal and different audience expectations of the genres the titles are relevant to. Other findings, revealing non-systematic differences, may tentatively be accounted for with reference to the corpus compilation procedure adopted. The collection of data was not carried out randomly. It involved the choice of keywords for the on-line queries identifying sub-disciplinary linguistic fields. The technical nature of these terms may have contributed to the collection of *similar* titles, that is, focused on the studies named rather than representative of varied content (including the encoding of evocative or vague notions). Given that, in general, lexicon is inextricably tied to phraseology, the a-priori selection of the same lexemes across title-genres may have triggered similar co-textual associations among the titles collected, which may have obscured or distorted some inter-genre differences.

The study can therefore be enhanced in two ways. First, further aspects of the titles collected can be investigated, such as the semantic relationships holding between coordinated noun phrases within the same title unit; the frequency of

occurrence of terms denoting abstract vs concrete entities; the occurrence of rare words; the semantic fields represented by keywords identified by comparing the title corpus with a larger, general corpus; the use of rhetorical effects; and distributional preferences of linguistic features already identified across sub-disciplinary linguistic fields (i.e. phonology vs discourse analysis; cf. Anthony 2001). In addition, the same analytical procedure followed here can be applied to sets of titles selected through different, non-lexical means (e.g. the titles of linguistics dissertations produced in a given year at a specific institution; the titles of the linguistics books published within a given series; the titles of the articles published by given leading journals over a specific period; and the titles of the presentations given at a series of conferences held at various institutions). The two complementary directions of investigation, oriented towards breadth and depth of analysis, are likely to shed more light on the degree and nature of intra- and inter-generic variation among titles.

NOTES

- 1 For studies addressing issues relevant to translation practice, the pedagogy of academic writing, and information comprehension, recall, selection and appeal, see the references in the literature reviewed below.
- 2 Parentheses and slashes signal alternative forms of a keyword that were typed together as variants of a given query by using the Boolean operator OR (e.g. *bilingual* OR *bilinguals*; *speech act* OR *speech acts* OR *lexical*).
- 3 Nord (1995: 274), instead, identified optimum book title length across cultures in the range of 3-4 words. Also, Soler (2007) found that the linguistics research papers in her corpus were 7.98 words long, on average.
- 4 Nord (1995: 280), however, identified slightly different macrostructural types of titles: single title, title + subtitle, title + *or* + title, and sequence of titles (see section 2.).
- 5 This is in line with the findings reported in other studies, such as Nord (1995) and Yakhontova (2002). However, Nord (1995) also retrieved adverbial and interjection titles.
- 6 Wang and Bai (2007) also reported similar variation in the realization of NP heads, and found instances of bi-head titles realized as text units coordinated by the function words *and*, *vs* and *or*.
- 7 Wang and Bai (2007) also found that pre-modifiers included mostly nominal or adjectival classifiers and deictics, but did not provide any frequency data.
- 8 Wang and Bai (2007) also found only embedded constructions as post-modifiers in their corpus, including one type not attested in my data, namely *to*-infinitive clauses.
- 9 Here and elsewhere, the words underlined are those relevant to the specific linguistic aspect being examined.
- 10 Hamp-Lyons (1987) suggested an alternative categorisation of titles – in terms of topic, focus, comment and viewpoint – which Anthony (2001) was able to adapt to the computer science paper titles he analysed.

REFERENCES

- Anthony L. (2001) "Characteristic Features of Research Article Titles in Computer Science", *IEEE Transactions on Professional Communication*, 44(3): 187-194.
- Bazerman C. (1985) "Physicists Reading Physics. Schema-Laden Purposes and Purpose-Laden Schema", *Written Communication*, 2(1): 3-23.
- Berkenkotter C., Huckin T. (1995) *Genre Knowledge in Disciplinary Communication: Cognition/Culture/Power*, Hillsdale, NJ: Lawrence Erlbaum.
- Busch-Lauer I.-A. (2000) "Titles of English and German Research Papers in Medicine and Linguistics Theses and Research Articles", in Trosborg A. (ed.) *Analysing Professional Genres*, Amsterdam/Philadelphia: John Benjamins, 77-94.
- Buxton A. B., Meadows A. J. (1977) "The Variation in the Information Content of Titles of Research Papers with Time and Discipline", *Journal of Documentation*, 33(1): 46-52.
- Campagna S. (2008) "Projecting Visual Reasoning", paper presented at the *CERLIS Conference "TRADING IDENTITIES: Commonality and Individuality in English Academic Discourse"*, University of Bergamo, 19-21 June, 2008.
- Day R. A. (1994) *How to Write and Publish a Scientific Paper*, Cambridge: CUP.

- Dillon J. T. (1981) "Functions of the Colon: An Empirical Test of Scholarly Character", *Educational Research Quarterly*, 5: 71-75.
- Dillon J. T. (1982) "In Pursuit of the Colon: A Century of Scholarly Progress: 1880-1980", *Journal of Higher Education*, 53(1): 93-99.
- Dressler W. U., Eckkramer E. M. (2001) "Functional Explanation in Contrastive Textology", *Logos and Language*, 2(1): 25-43.
- Fortanet I., Coll J. F., Palmer J. C., Posteguillo S. (1997) "The Writing of Titles in Academic Research Articles", in Chamorro R. M., Navarrete A. R. (eds.) *Lenguas aplicadas a las ciencias y la tecnología. Aproximaciones*, Cáceres, Spain: Universidad de Extremadura, Servicio de Publicaciones, 155-158.
- Fortanet I., Posteguillo S., Coll J. F., Palmer J. C. (1998) "Linguistic Analysis of Research Article Titles: Disciplinary Variations", in Vazquez I., Camilleu I. (eds.) *Perspectivas Pragmáticas en Lingüística Aplicada*, Zaragoza, Spain: Anubar, 443-447.
- Goodman R. A., Thacker S. B., Siegel P. Z. (2001) "What's in a Title? A Descriptive Study of Article Titles in Peer-Reviewed Medical Journals", *Science Editor*, 24(3): 75-78.
- Haggan M. (2004) "Research Paper Titles in Literature, Linguistics and Science: Dimensions of Attraction", *Journal of Pragmatics*, 36: 293-317.
- Hamp-Lyons L. (1987) *Study Writing: A Course in Written English for Academic and Professional Purposes*, Cambridge: CUP.
- Hartley J. (2005) "To Attract or to Inform: What are Titles for?" *Journal of Technical Writing and Communication*, 35: 203-213.
- Hartley J. (2007a) "Colonic titles!", *The Write Stuff: The Journal for European Medical Writers*, 16(4): 147-149.
- Hartley J. (2007b) "There's More to the Title than Meets the Eye: Exploring the Possibilities", *Journal of Technical Writing and Communication*, 37: 95-101.
- Hartley J. (2007c) "Planning that Title: Practices and Preferences for Titles with Colons in Academic Articles", *Library and Information Science Research*, 29: 553-568.
- Hartley J. (2008) *Academic Writing and Publishing: A Practical Handbook*, Abingdon: Routledge.
- Huth E. (1987) *Medical Style & Format: An International Manual for Authors, Editors, and Publishers*, Baltimore: Williams & Wilkins.
- Hyland K. (2002) "What do they Mean? Questions in Academic Writing", *Text*, 22(4): 529-557.
- Lewinson G., Hartley J. (2005) "What's in a Title? Number of Words and the Presence of Colons", *Scientometrics*, 63(2): 341-356.
- Michelson G. (1994) "Use of Colons in Titles and Journal Status in Industrial Relations Journals", *Psychological Reports*, 74: 657-658.
- Mungra P. (2007) "Metaphors among Titles of Medical Publications: An Observational Study", *Ibérica*, 14: 99-122.
- Nord C. (1995) "Text-Functions in Translation: Titles and Headings as a Case in Point", *Target*, 7(2): 261-284.
- Soler V. (2007) "Writing Titles in Science: An Exploratory Study", *English for Specific Purposes*, 26: 90-102.
- Swales J. (2003) *Genre analysis. English in Academic and Research Settings*, Shanghai: Shanghai Foreign Language Education Press.
- Swales J., Feak C. B. (1994) *Academic Writing for Graduate Students. A Course for Nonnative Speakers of English*. Ann Arbor: The University of Michigan Press.
- Wang Y., Bai Y. (2007) "A Corpus-based Syntactic Study of Medical Research Article Titles", *System*, 35: 388-399.
- Whissell C. (1999) "Linguistic Complexity of Abstracts and Titles in Highly Cited Journals", *Perceptual and Motor Skills*, 88: 76-86.
- Yakhontova T. (2002) "Titles of Conference Presentation Abstracts: A Cross-Cultural Perspective", in Ventola E., Shalom C., Thompson S. (eds.) *The Language of Conferencing*, Frankfurt: Peter Lang, 277-300.
- Yitzhaki M. (1994) "Relation of Title Length of Journal Articles to Number of Authors", *Scientometrics*, 30: 321-332.

Increasing or Decreasing the Sense of “Otherness”: the Role of Audiovisual Translation in the Process of Social Integration

VANESSA LEONARDI
University of Ferrara

Whether domesticating or foreignizing in its approach, any form of audiovisual translation [...] ultimately plays a unique role in developing both national identities and national stereotypes. The transmission of cultural values in screen translation has received very little attention in the literature and remains one of the most pressing areas of research in translation studies. (Baker and Hocheil, 2001: 76)

ABSTRACT

This paper is aimed at analyzing the impact and potential of audiovisual translation (AVT) in the process of social integration. AVT is mainly characterized by the use of language which, far from being neutral, can be used in many different ways in order to manipulate meanings and exert a strong influence on society as a whole. Language has always been associated with power, manipulation and ideology and as such is at the core of any study carried out in the field of AVT. The type of language transfer chosen in any country depends on many factors, such as historical, political, cultural, religious and even economic preferences. However, it has become more and more common to refer to ‘foreignizing’ and ‘domesticating’ (translation) strategies when deciding upon and judging the type of language transfer to be used in particular contexts. It is not surprising that these strategies are used in translation in order to decrease or increase the sense of ‘otherness’. Subtitling is an example of foreignizing translation where the ‘otherness’ is given freedom to express itself because it is the target culture that moves towards the source culture. Despite being viewed initially as a lower quality product as compared to dubbing, subtitling is now much more valued than it was in the past thanks to its ‘respect’ for the culture and language of other

countries. Dubbing, on the other hand, is a classical example of linguistic nationalism aimed at domesticating the 'otherness' in order to minimize the sense of 'foreignness' for the target culture. There is nowadays an increasing awareness of the importance and the need to strengthen language learning and to promote a new model of cultural pluralism which breaks away from xenophobic and racist attitudes and empowers minority groups in various countries. AVT is an important field of interdisciplinary research which should be deeply analyzed in order to better value its potential of promoting multilingualism and multiculturalism in the world. Drawing upon examples of dubbed Italian versions of some Walt Disney movies, this paper aims at re-evaluating dubbing in a much more positive light and showing how it can help strengthen social integration in an increasingly globalized world.

1. INTRODUCTION

Translation studies is a very large academic field which over the years has included a large variety of studies and has combined many different fields together. This is why it is not surprising that many translation scholars are undertaking research in the field of Film Studies focusing especially on the phenomena of dubbing and subtitling. Many translation theorists in the past have recognized the power exerted by translation in terms of an ideological tool for manipulating language. In the field of Film Studies as well there has been a considerable interest in considering television and cinema as two major channels of power distribution. This is because both fields are characterized by the use of language which, far from being neutral, can be used in many different ways in order to manipulate meanings and exert a strong influence on society as a whole.

Nevertheless, it is fundamental to recognize the importance played by television and/or cinema and translation in general in terms of breaking down language barriers and reaching the whole world through communication. It is thanks to translation that many different cultures and peoples in the world have had the chance to communicate with one another; and it is also thanks to the mass media that these peoples have enjoyed, and still do, a different and much more international environment. This interest of translation scholars in the field of Film Studies has created a typology of translation called 'screen translation' or 'film translation' or even 'audiovisual translation'.

2. DEFINITIONS

Audiovisual translation (AVT) is a term used to refer to any language and cultural transfer aimed at translating original dialogues coming from any acoustic or visual product. AVT is also named film translation, screen translation, cinema translation and multimedia translation. It seems, however, that the term AVT is more appropriate in that it includes all these definitions within it. AVT is not a new subject at all, although it has been only recently developed. Research and seminars started to become more and more frequent after 1995, the year in

which the European Council celebrated the anniversary of the birth of the cinema and, on that special occasion, a forum on audiovisual communication and language transfer was organised (Perego, 2005:7). AVT is very much criticised in all its forms by many people who either do not consider it as a proper form of translation but rather as an adaptation or who regard it as a linguistic game used to 'spot the error' as also acknowledged by Shochat and Stam (1985:46). There are, indeed, many websites in which dubbing or subtitling gaffes are carefully reported. Nevertheless, few people really understand the importance of this extremely challenging and complex form of translation. It involves, indeed, a number of technical, linguistic and cultural constraints, not to mention social and economic constraints.

There are many types of audiovisual translation, although subtitling and dubbing are among the most commonly used and mentioned in most parts of the world. AVT can, indeed, be divided into two main groups, that is, intra-lingual (or monolingual) subtitling for the 1) hard-of-hearing and the deaf, 2) audio descriptions for blind people, 3) subtitling for the theatre or the opera and 4) live subtitling (generally used in the news broadcasts); and inter-lingual translation which includes 1) dubbing, 2) subtitling and 3) voice-over. Broadly speaking, AVT can also be simply divided into oral (dubbing) and written (subtitling) forms of language transfer.

Traditional dubbing countries include the so-called FIGS group, that is, France, Italy, Germany and Spain and the reasons for their choice is mostly linked to history and to their past political regimes. Subtitling countries are the Netherlands, Greece, Portugal, Slovenia, Croatia, Denmark, Norway and Sweden among others. Economic and social reasons partly explained their preference for subtitling which is generally cheaper than dubbing and allows for bilingualism. According to O'Connell (2007) the target audience is also an important factor which exerts some kind of influence on the selection of dubbing or subtitling. Factors such 'as the age, sex, educational background and social class of the audience' tend to be taken into consideration in the selection process (ibid.:128). Delabastita also acknowledges that film translations do not only pose linguistic problems, but they are 'conditioned to a large extent by the functional needs of the receiving culture and not, or not just, by the demands made by the source films' (1990:99). It is not, however, the aim of this paper to look in detail into the reasons that motivate the choice of a particular AVT technique but rather to analyse what impact a particular choice can have from the point of view of increasing or decreasing the sense of 'otherness'.

3. GLOBALISATION AND LINGUISTIC PLURALISM

In an increasingly globalised world, there seems to be the need to preserve identity and promote cultural pluralism and the use of AVT could either increase or decrease the sense of 'otherness'. Globalisation has had significant impact not only on countries' economic and political systems but also on the language itself which is used as a means of communication. Globalisation has greatly contributed to the homologation and linguistic standardization which are evident in the

translation strategies used both in subtitling and dubbing. This would, in turn, contribute to the gradual disappearance of linguistic and cultural heritages, especially for those minority or endangered languages all around the world. Dubbing, in particular, seems to be the major communication means which tends to standardise the language towards the 'Self' community and damaging or making the 'Other' community disappear. Initially adopted as a way of simplifying and enhancing the comprehension of foreign films, its use in some countries, such as Italy for instance, has been the symbolic expression of a xenophobic purism aimed at avoiding contacts or contaminations with different foreign 'Other' realities. It is not surprising, indeed, that dubbing has been chosen in the past in order to hide the foreignness of films coming from different countries. Strong nationalistic countries, such as Germany, Italy and Spain for instance, had adopted dubbing for real ideological reasons or, as Danan (1991:612) puts it, as 'an assertion of the supremacy of the national language and its unchallenged political, economic and cultural power within the nation boundaries [...]']

There are also, however, several significant technical reasons which explain the adoption of dubbing in some countries, such as the impossibility or complexity of rendering specific non-standard linguistic varieties. Among them, there are all those regional varieties, slang, accents and colloquial expressions which are extremely difficult to render through dubbing. Nevertheless, it is worth noting that these non-standard linguistic varieties represent an expression of identity and belong to the linguistic community of the country of origin. Suppressing and avoiding these varieties means suppressing all those cultural and linguistic diversities which belong to the 'Others' around us thus threatening the survival of some minority languages and dialects. Many people argue that this does not normally happen in subtitled films or programmes, but this is not entirely true. In the case of subtitling in minority language settings, such as Wales or Ireland for instance, there seems to emerge a tendency to reinforce the major language through subtitles instead of promoting the minority ones. O'Connell mentions the study carried out by d'Ydewalle et al. in Belgium in 1987 which proved the impossibility of avoiding reading subtitles. As a consequence, since reading requires more concentration than listening, one could argue that subtitling tends to favour and strengthen the major rather than the minority languages (2007:131).

Dubbing, on the other hand, should not always be regarded as something negative aimed at domesticating the Others' language and culture in favour of the target audience. The merit of reducing, in some cases, racist stereotypes in all those movies where particular linguistic and cultural communities are targeted should be recognised. In order to prove this point, this paper will explore how a specific category of films, Walt Disney feature films, can contribute to reinforcing stereotypes and how the dubbed (Italian) versions maintain or eliminate them altogether.

4. SUBTITLING

Subtitling is 'the process of providing synchronized captions for film and television dialogue' (Shuttleworth and Cowie, 1997:161). From a linguistic point of view,

subtitles can be *intralingual* (within one language) and *interlingual* (between two languages), based on the distinction made by Jakobson in 1959; however, from a more technical point of view, they are divided into *open* (not optional, i.e. shown with the film) and *closed* (optional, i.e. shown via teletext) (*ibid.*).

There are many obvious advantages related to the use of subtitling which is generally regarded as a cheaper and faster modality of audiovisual translation. Thanks to subtitling the original soundtrack can be preserved and it can also play an important role in language acquisition. In this respect, Danan (1991:613) asserts that interlingual subtitling, in particular, 'indirectly promotes the use of a foreign language as an everyday function in addition to creating an interest in a foreign culture'.

Nevertheless, there are also several constraints which should not be disregarded. In his article on subtitling Gottlieb (1992) discusses all those *formal (quantitative)* and *textual (qualitative)* constraints which characterise television subtitling. *Formal* constraints include all those issues related to the *space factor* and the *time factor*. On the other hand, *textual* constraints derive from the visual context of the film. Among the most influential theorists who have discussed subtitling constraints extensively, there is also Delabastita (1989) who considers, among other issues, the topic of reduction as being one of the most problematic aspects of this form of translation. Constraints involve selection on the behalf of the translator who has thus a very complex task and responsibility. From a pragmatic perspective, some scholars have applied notions derived from relevance theory to subtitling and claim that 'decisions about deletions are context-dependent' (Kovačič, 1994:250).

Subtitling can be seen as a form of foreignization where the stress is laid upon the identity of the foreign text leaving the sense of 'otherness' unchanged.

5. DUBBING

The technique of dubbing was introduced further to the introduction of sound in films and this explains why this term can be assigned more than one definition. For the purpose of this study this term refers to lip-sync dubbing, that is 'where the foreign dialogue is adjusted to the mouth movements of the actor in the film' (Dries, 1995:9).

It is interesting to note that when watching a dubbed film, not many people would realise that they are presented with a translated version of the original and therefore, this technique tends to be less criticised in terms of 'error spots' (Shochat and Stam, 1985:49). This, however, does not mean that dubbing is not subject to any constraints. Many scholars seem to agree upon the fact that one of the major constraints of dubbing is that of synchrony defined as 'the agreement between the articulatory movements seen and the sounds heard' (Barbe 1996:259). Script editors make, from time to time, changes in the translated versions to make sure that each utterance matches the lip movements of people acting in the film.

One of the merits of dubbing is that it does not distract people's attention from the image and is therefore suitable for people or children who do not have

particularly fast or good reading skills. It also involves less reduction of the original dialogue as compared to subtitling and, generally speaking, it contributes to a more relaxed viewing. However, despite these important positive elements, there are also several constraints related to the use of dubbing. First of all, it is very expensive and takes much more time and effort to realise as compared to subtitling. Secondly, it involves the inevitable loss of the original soundtrack. Thirdly, viewers are not given the opportunity to listen to the foreign language and this has a very significant impact on language learning issues. Dubbed material is, to a certain extent, limited to its country of origin and as such 'exclude(s) certain categories such as tourists and other visitors who may not speak the local language' (Baker and Hochel, 2001:75). Dubbed material tends to be also associated with ideology and censorship as well as with other kinds of 'undetected textual manipulation' (O'Connell, 2007:126). A very interesting example of this can be seen in the Italian version of the American series the *Nanny* that was aired on CBS from 1993 to 1999. The Italian version has been adapted in such a clever way as to reproduce the same cultural impact on the target audience. In both versions there are examples of stereotypes and comic elements which, although culturally different, seem to be very successful. In the US version the protagonist, Fran, is a young Jewish woman coming from Flushing, Queens (New York) whereas in the Italian version her origins are from Southern Italy. In the US version Fran speaks with a strong accent which immediately makes the audience think of the stereotypical Jewish / New York woman. In the Italian version Fran does not speak with any particular accent but language play and manipulation are achieved through the use of some specific dialectal expressions which the viewers associate with Southern Italy and stereotypes. In this case dubbing contributes to maintaining and, to some extent, reinforcing typically-voiced national stereotypes.

Dubbing is a perfect example of domestication in translation. It has historically been associated with nationalism, and it is not surprising indeed that Germany, Spain and Italy are amongst those countries which have adopted dubbing for socio-political and ideological reasons related to their strong sense of patriotism and their political regime of the time. It would be wrong, though, to think of dubbing only in terms of nationalism and xenophobic purism. On the contrary, dubbing can play a very important role in international communication transfer as a means to dilute the degree of racism which can be noticed in some audiovisual materials. This could be seen, for example, in some of the most famous Walt Disney movies.

6. CULTURAL TRANSFER IN TRANSLATION

It is undoubtedly a very challenging enterprise to make sure that a film released nationally in a specific country can eventually enjoy success abroad. Both linguistic as well as cultural differences are among the most common obstacles to be overcome. This contributes to raising cross-cultural issues which cannot be avoided or ignored if a translation is to be successful. In the past, but still nowadays, cultural issues have been at the core of heated debates which focused their attention on the issue of translatability. This issue was approached from two dif-

ferent perspectives, that is, the universalist theory and the monadist one. According to the first theory, supported by Jakobson, Bausch, Hauge and Nida among others, translatability is possible thanks to the existence of linguistic universals. Whereas, on the other hand, the supporters of the monadist approach, such as Edward Sapir for instance, assert that translatability is jeopardised by the way each linguistic community interprets reality in its own way. In the last few years there has also been another important approach to the issue of translatability coming from the Deconstructionist theory where the notion of translation as transfer of meaning is questioned.

There is an inevitable link between language and culture and this can be seen, for instance, in fairytales and children's literature. Disguised as education tools, fairytales can sometimes twist truth by perpetrating and reinforcing wrong prejudices among children. Some fairytales tend to reflect a society's culture and bias against some other cultures. Translation of fairytales can either leave these stereotypes and prejudices unchanged or eliminate them altogether if they are deemed to be offensive for the target culture. It is therefore the aim of this paper to show how a dubbing translation strategy can indeed soften and, in some cases, completely eliminate racist stereotypes originated in the USA according to Western thinking and presented in Italy which is not traditionally a multilingual community, although things are beginning to change.

7. LANGUAGE, IDEOLOGY AND MANIPULATION

The use of language is inevitably linked to the issues of power, ideology and manipulation. Simpson defines ideology as 'the taken-for-granted assumptions, beliefs and value-systems which are shared collectively by social groups' (1993:3). It is worth noting, in this respect, that Walt Disney animated movies tend to reflect the American attitude towards particular races and ethnicities. Some of these assumptions may also be shared by some other Western countries and the role played by AVT in this respect is fundamental because it can help decrease the sense of 'otherness' through the use of language. Language should not be regarded as being merely a tool for communication, but it carries with it moral, religious and socio-political values which could reflect or manipulate particular ideologies. The term ideology is not easy to define since it is a very controversial issue which is interpreted and consequently defined in different ways. According to Thompson (1984:3) 'to study ideology is, in some part and in some way, to study language in the social world'. A similar position is taken by Gruber (1990: 195) who claims that 'ideology manifests itself linguistically and is made possible and created through language'. Despite numerous definitions of the term 'ideology', there seems to be a common agreement that all ideologies find their clearest expression in language as also acknowledged by Hodge and Kress (1993) who made a distinction between different levels of language, such as the lexical-semantic level and the grammatical-syntactic level. Hatim and Mason also declared that:

Ideologies find their clearest expression in language. It follows, therefore, that the analysis of linguistic forms is enriched by the analysis of those ideological structures

which underpin the use of language. [...] Behind the systematic linguistic choices we make, there is inevitably a prior classification of reality in ideological terms. The content of what we do with language reflects ideology at different levels: at the lexical-semantic level, and at the grammatical-syntactic level. [...] Whatever is said about the degree of freedom the translator has, the fact remains that reflecting the ideological force of the words is an inescapable duty. (1990:161)

It is, indeed, through language that such ideologies are shaped and reality itself is shaped. Language is not neutral and through the use of particular words or constructions it can distort messages and realities. There are obvious differences in the use of language that people select to create ideas, express opinions and interpret the world. Language is linked to power and power relations and it helps control discourse and, thus, reality. The best way to untangle discourse is through a careful analysis of the language used.

For the purpose of this study, the term ideology simply refers to a range of ideas, beliefs, value systems, prejudices, stereotypes, legends and any other notions of language that can be seen through the use of particular accents and/or dialects in the original movies and the use of standard language in the dubbed versions.

7. REDUCING RACIST STEREOTYPES IN DUBBED FILMS

Although dubbing may be accused of domestication and, in some cases, a deep sense of nationalism, it is worth noting that it could prove to be a very useful means of reducing or avoiding racist stereotypes. This is particularly true in the case of Walt Disney feature films whose importance is fundamental in that they directly target children.

The Walt Disney Company is one of the most powerful and best-known media corporations which targets children, parents and teachers from all over the world and presents a worldview based on innocence, magic, and fun. Its main and declared aim is to educate children through entertainment. However, behind these images of innocence and fantasy there is a hidden ideology aimed at reinforcing class, age, gender and racist stereotypes.

There are several examples of racism that can be found in Walt Disney's animated features, such as 1) the native cannibals in *Alice Cans the Cannibals* (1933) 2) the Indians from *Peter Pan*, 3) the merchant and evil Jafar from *Aladdin*, 4) the crows from *Dumbo*, 5) Sebastian from *The Little Mermaid*, 6) Sunflower the centaur from *Fantasia* and 7) King Louie from *The Jungle Book* and 8) the Siamese cats in *Lady and the Tramp*. 'Otherness' seems to be one of the major recurring themes in Disney's production, as also acknowledged by Di Giovanni who claims that 'since the primary aim of Disney films is to appeal to large audiences worldwide, especially children, representations of the Other are necessarily smoothed and simplified by the selection of exotic elements which are well-known to the Western world, being part of a more or less fixed repertoire' (2003:211). Unfortunately, in their attempt to 'simplify' concepts, Walt Disney productions tend to stereotype all those typical features of culturally different 'Others' towards a certain degree of racism, whereas the Americans tend to be portrayed mainly as the 'good people'.

Children's minds are very easy to control and when they watch these cartoons they make connections to their favourite characters and relate to them. Garrett and Bell (1998:3) claim that '[...] media use can tell us a great deal about social meanings and stereotypes projected through language and communication [...]'. People are usually taught from childhood to make social distinctions and this will, in turn, have a significant impact on their own cultural tolerance and social behaviour in the future. Stereotypes are reinforced through actions and language variations as also acknowledged by Lippi-Green (2005). Racist stereotypes can, indeed, be expressed through accents in non-native speakers. Native speakers primarily play the main roles, while non-native speakers act as servants or villains from the lower class. Children eventually learn to associate foreign accents with evil and native standard accents with good. This is mainly due to the fact that children tend to passively absorb this kind of information and they do not actively analyse it. It is not surprising, therefore, that all Disney characters that speak with a foreign accent and are characterized by a foreign look are mainly associated with negative features. Native Americans, for instance, are clearly portrayed as savages in *Peter Pan* and *Pocahontas*; Arabs act just like barbarians in *Aladdin*; African-Americans are portrayed as jive crows in *Dumbo* and as human-wannabe orangutans in *The Jungle Book*; strangely enough, they seem to be totally absent in *Tarzan's Africa*, where people would probably expect them to be; both Latinos and African-Americans become street criminals in *The Lion King* and, finally, the unfaithful Siamese cats in *Lady and the Tramp* resemble Asians. Stereotypes can be stressed by both images and verbal language although the latter seems to have a much stronger impact. Hudson (1980: 193) also claims that 'language is one of the most important factors by which social inequality is perpetuated from generation to generation'.

Walt Disney feature films are famous all over the world and, as such, they are translated and shown either in dubbed or subtitled versions in many different languages. Nevertheless, in most dubbed Italian versions, many of these stereotypes are completely eliminated and a standard language is used for all the characters. Thanks to this strategy racist stereotypes are omitted and children are somehow unaware of them. This could be a positive aspect of dubbing which, in a nationalist effort to ease the comprehension for the target audience, eliminates stereotypes and prejudices. Let us look at some of these examples in the following sections.

7.1 DUMBO (1941)

Dumbo was released in 1941 and it was Walt Disney fifth animated movie. It tells the story of a 'peculiar' and, at the same time, 'special' little elephant that is laughed at by others because of his big 'funny' ears. Thanks to his friend, Timothy Mouse, *Dumbo* is able to transform his weakness into a special gift as he is taught how to fly by a group of crows. Apparently, this movie seems to give a very positive message, that is difference is not bad but a good thing to be appreciated. Diversity seems to be accepted, tolerated and praised in a way. However, if one looks a little deeper into this movie, s/he would realise that the crows are 'different' from all the other characters. They resemble Afro-Americans whose accent is

unquestionably typical of black people and, more precisely, they tend to portray proletarian blacks as also acknowledged by Wilmington (1980). Expressions such as 'I seen a horsefly [...]' or 'Uh, what's all the rookus? C'mon, step aside brothuhs, uh, what's cookin' around heah? What new? What fryin', boys?' are obviously very typical of black colloquial dialects. It is not surprising, then, that when it comes down to singing and dancing they do it in a typical jazz style and, furthermore, one of them is called Jim Crow¹. In the dubbed Italian version all these references are completely lost. The accent is standard or, at least, cannot really be associated with any particular stereotype. The reference to Jim Crow is pointless and smoothed in the dubbed version. There does not seem to be any way in which Italian children can possibly perceive any form of racism in this film based on language differences.

7.2 *THE JUNGLE BOOK* (1967)

In 1967 Walt Disney decided to release its version of Rudyard Kipling's *The Jungle Book*. It tells the story of a young boy called Mowgli who, after losing both parents in an accident, is raised by some wolves in the jungle. When he starts growing up, panther Bagheera decides that it is time for the boy to return to the 'Man Village'. Mowgli, however, seems determined to remain in the jungle and does not want to live in the 'Man Village'.

There is a scene in this movie where Mowgli is abducted by a group of apes that resemble Afro-Americans in both the way they talk and the way they sing and dance. Metcalf (1991) claims that apes in this movie represent the unhappy situation of black people living in the States and their will to be treated just like white people. The song performed in this scene is a clear example of this and the way they sing and dance over this tune resemble the typical Jim Crow dance. Their accent is typically black whereas all the other animals in the jungle speak in a British accent.

In the Italian version the accent chosen is just a standard one where there is no way to associate the apes with any ethnic community. As a result, both the song and the dance performed by the apes acquire no specific meaning and they only serve the purpose of entertainment for children. This means that references to Jim Crow are lost in the dubbed Italian version thanks to the decision to opt for a standard accent.

7.3 *THE LITTLE MERMAID* (1989)

This movie was released in 1989 and was adapted from Hans Christian Andersen's original version. *The Little Mermaid* tells the story of Ariel, King Triton's youngest daughter, who is fascinated with the human world and has fallen in love with the human Prince Eric. Ariel goes to the evil sea witch, Ursula who promises to turn her into a human for three days in exchange for her voice. If Ariel can obtain a kiss of true love from Prince Eric within the three days, she can remain human forever. If not, she will have to become Ursula's slave.

One of the main characters of this movie is Sebastian the crab who is the only one to speak with a typical Jamaican accent. There is a particular scene in this movie where Sebastian tries to persuade Ariel that life is much better 'under the sea' because at least nobody has to do any work or get any job compared to what

happens on the surface. He then sings a song that goes like this:

Up on the shore they work all day, out in the sun they slave away
While we devoting full time to floating under the sea

This could be a typically voiced stereotype about Jamaican people who are regarded as lazy people who do not like to do any work. In the Italian version this stereotype is not shown thanks to the decision to opt for a standard accent. There could have been a way to maintain it even in the Italian version by opting for a Southern Italian accent since it is commonplace in Italy to regard people from the South as lazy people who do not like working. This strategy would have respected the English version but, at the same time, would have reinforced negative stereotypes commonly voiced in Italy. Negative associations are smoothed in the dubbed Italian version and children seem not to be negatively influenced as compared to their English counterparts.

If we compare this movie to *Lady and the Tramp*, it is interesting to note how the Siamese cats that speak with a typical Asian accent are dubbed into Italian with a clear Chinese accent. This decision may be coherent with the original version but it is not definitely coherent with previous choices to opt for a standard accent for all the characters. Moreover, their slanted eyes and buck-teeth contribute to reinforcing existing stereotypes of Asians.

7.4 ALADDIN (1992)

Aladdin was released in 1992 and it tells the story of a poor and homeless boy who lives in a large and busy town with his faithful monkey friend Abu. One day he accidentally meets Princess Jasmine at the marketplace and he saves her from a merchant who wants to cut off her hand for stealing an apple to give to a starving child. Under the orders of the evil Jafar, Aladdin is taken to the castle and thrown in jail. Thanks to the help of the genie of the lamp, Aladdin gets rid of Jafar and marries the princess.

Several examples of stereotypes of Arabs or Moslems can be found in this film and this can be seen in the very first scene where there is a man sitting on a camel who sings the following song:

‘Oh I come from a land/ from a faraway place/ where the caravan camels
roam/ **where they cut off your ear/ if they don’t like your face/** it’s barbaric,
but hey, it’s home’.

This lyric makes a clear reference to a law in the Middle East known as The Sharia Edict of cutting off the hand of a thief. Because of these words, the movie was heavily criticised by the Arab communities and Walt Disney was forced to change the sentences in bold to ‘Where it’s flat and immense/ and the heat is intense’. It is interesting, though, to note that the final verse ‘it’s barbaric, but hey, it’s home’ is left in the revised American version. In the Italian version, this lyric is fully re-

produced, except from the final verse which is changed to *non è facile, ma io ci vivo laggiù* (It is not easy, but I live over there).

It is worth noting how, in the original version, only the villains look different as compared to Aladdin or Jasmine who look very anglicized. Jafar and the other villains speak with an accent, whereas both Aladdin and Jasmine speak very good American English. Their accent, somehow, seems to help them get closer to the viewers who, on the other hand, too often make an association between foreign accent and evil.

In the dubbed Italian version, there is no such a difference between the villains and the main protagonists. All of them speak with a proper standard accent and therefore the sense of 'otherness' is domesticated in a positive light. As a matter of fact, it is worth noting that Aladdin came out during the Gulf War and provided the viewers with an exotic but, at the same time, a barbaric reality of people living in the Middle East. Children cannot obviously link the movie to the Gulf War, but by watching this movie they could learn that strange things happen in those 'places' where there does not seem to be any kind of 'civilization'. In other words, children learn that people with an Arab look and accent should not be trusted.

7.5 THE LION KING (1994)

The Lion King was released in 1994 and it tells the story of a young lion in Africa named Simba, who learns of his place in the great 'circle of life' and overcomes many obstacles to claim his place as the rightful king. Son of the king of all lions, Mufasa, his birth angers Mufasa's brother Scar, who would have been next in line to the throne, and so he plots to kill both of them. Mufasa is eventually killed by a herd of Wildebeest whereas Simba is led to believe by Scar that he is responsible for his father's death. Simba decides to run away and Scar becomes king of the Pridelands. After several years, Simba is persuaded to return to the Pridelands to help his friends and get rid of Scar.

Although this movie was very well acclaimed, it also received several serious criticisms in terms of racist, sexist, homophobic, violent and stereotyping issues contained in it. The racist issues can be clearly seen in the portrayal of both the hyenas as well as Uncle Scar. The Lion king is set in Africa but whereas Simba, the protagonist, speaks with a good and clear English accent, the hyenas or 'bad characters' speak non-standard English typically spoken in the street and it resembles the inner city African American dialect. Two white American actors were chosen for the voice of Simba whereas the hyenas are voiced by Hispanic and Black actors.

In the Italian dubbed version, the choice of opting for a standard accent once again smoothes the sense of otherness and all the xenophobic considerations which can be found in the original version. Italian children only perceive the hyenas as evil only because of their behaviour but there is no connection between 'foreign' accent and evil. There are no racist representations or stereotypes in the Italian version which could exert influence over children.

8. CONCLUDING REMARKS

The media plays a very important role in terms of ideological manipulation of reality. This is particularly important when racist stereotypes or comments may damage linguistic and cultural pluralism. In an increasingly globalised world there is a need to unite people and eliminate all forms of discrimination and AVT can certainly help in this respect.

This paper attempted to show the power of AVT in increasing, decreasing or even maintaining the sense of 'otherness' through carefully selected linguistic choices. Except from *Lady and the Tramp*, all the Disney animated movies analysed in this paper showed how the racist issues related to language in the original version disappeared in the dubbed Italian versions thanks to the dubbing strategies adopted. It is not really possible to assert firmly that such choices were made in order to eliminate racism, but this is the final result in the dubbed versions.

Dubbing, therefore, should no longer be considered as something negative aimed at reinforcing nationalist values against foreign different 'others'. It is nowadays a means of helping the 'others' fit in with the socio-cultural and linguistic community they come in contact with. In the case studies presented in this paper dubbing served the role of introducing diversity in a very mild manner away from stereotypes and racist attitudes. The dubbed Italian versions analysed in this paper offer children as well as adults a new perspective of 'others' which is not based on prejudices. The villains are presented as bad characters only on the basis of their behaviour not because of their 'foreign' accent.

However, as previously mentioned, it is not only the verbal language but also the images which manipulate viewers and unfortunately this element cannot be avoided in the dubbed versions. Racism is, nevertheless, smoothed thanks to the adoption of standard language and Disney animated movies lose most of their racism for the Italian audience.

It was not the aim of this paper to explore the reasons why Walt Disney movies contain all these forms of racist and sexist stereotypes, but they certainly deserve a further investigation.

NOTES

1 Jim Crow arose in the wake of the Civil War and it can be defined as an informal term for the practice of segregation. Jim Crow's name is taken from 1838 minstrel show character, played by a white actor called Thomas Rice who used to sing songs in blackface. He wrote a song called 'jump Jim Crow' where he depicted blacks as lazy, singing and dancing fools. The term 'Jim Crow' became a racial epithet for blacks and by the end of the 19th century, it was used to describe laws and customs that oppressed blacks, particularly in the South of the United States.

REFERENCES

PRIMARY SOURCES

Aladdin (1992) released by Walt Disney and distributed by Buena Vista Home Entertainment.

Dumbo (1941) released by Walt Disney and distributed by Buena Vista Home Entertainment.

Lady and the Tramp (1955) released by Walt Disney and distributed by Buena Vista Home Entertainment.

The Jungle Book (1967) released by Walt Disney and distributed by Buena Vista Home Entertainment.

The Lion King (1994) released by Walt Disney and distributed by Buena Vista Home Entertainment.

The Little Mermaid (1989) released by Walt Disney and distributed by Buena Vista Home Entertainment

SECONDARY SOURCES

Baker, Mona and Braño Hochel (2001) 'Dubbing', in Mona Baker

(ed) *Routledge Encyclopedia of Translation Studies*, London and New York: Routledge.

Barbe, Katharine (1996) 'Dubbing in the Translation Classroom', *Perspectives: Studies in Translatology* 4 (2): 255 - 274.

Danan, M. (1991) 'Dubbing as an Expression of Nationalism', *META*, 36(4): 606-614.

Delabastita, Dirk (1989) 'Translation and Mass-Communication: Film and TV Translation as Evidence of Cultural Dynamics', *Babel* 35(4): 193-218.

Di Giovanni, Elena (2003) 'Cultural Otherness and Global Communication in Walt Disney Films at the Turn of the Century', *The Translator* 9(2): 207 - 223.

Dries, Josephine (1995) *Dubbing and Subtitling: Guidelines for Production and Distribution*, European Institute for the Media.

- Garrett, P. and A. Bell (1998) 'Media and Discourse: A Critical Overview', in A. Bell and P. Garrett (eds) *Approaches to Media Discourse*, Oxford and Malden Massachusetts: Blackwell, pp. 1-20.
- Gottlieb, Henrik (1992) 'Subtitling. A New University Discipline', in Cay Dollerup and Anne Loddegaard (eds) *Teaching Translating and Interpreting. Training, Talent, Experience. Papers from the First Language International Conference*, Amsterdam: John Benjamins, pp. 161 - 170.
- Gruber, H. (1990) 'Ein Gespenst geht um in Österreich', in Ruth Wodak and Florian Menz Editorial 113 (eds) *Sprache in der Politik: Politik in der Sprache. Analysen zum öffentlichen Sprachgebrauch*, Klagenfurt: Drava, pp. 191-207.
- Hatim, Basil and Ian Mason (1990) *Discourse and the Translator*, London and New York: Longman.
- Hodge, R. and Kress, G. (1993) *Language as Ideology*, second edition, London and New York: Routledge.
- Hudson, Richard A. (1980) *Sociolinguistics*, Cambridge: Cambridge University Press.
- Jakobson, Roman (1959) 'On Linguistic Aspects of Translation', in R. A. Brower (ed.) *On Translation*, Cambridge, MA: Harvard University Press, pp. 232-39.
- Kovačič, Irena (1994) 'Relevance as a Factor in Subtitling Reductions', in Cay Dollerup and Annette Lindegaard (eds) *Teaching Translation and Interpreting II: Insights, Aims, Visions. Papers from the Second Language International Conference*, Amsterdam: John Benjamins., pp. 245 - 251.
- Metcalf, Greg (1991) 'It's a Jungle Out There, Kid: The Sixties in Walt Disney's *The Jungle Book*', *Studies in Popular Culture* XIV: 1: 85-97.
- O'Connell, Eithne (2007) 'Screen Translation', in Piotr Kuhiwczak and Karin Littau (eds) *A Companion to Translation Studies*, Clevedon, Buffalo, Toronto: Multilingual Matters Ltd, pp. 120 - 133.
- Perego, Elisa (2005) *La Traduzione Audiovisiva*, Rome: Carocci.
- Shochat, Ella, and Robert Stam (1985) 'The Cinema after Babel: Language, Difference, Power', *Screen* XXVI: 35-58.
- Shuttleworth, Mark and Moira Cowie (1997) *Dictionary of Translation Studies*, Manchester: St Jerome Publishing.
- Simpson, Paul (1993) *Language, Ideology and Point of View*, London and New York: Routledge.
- Lippi-Green, Rosina (2005) 'Teaching Children How to Discriminate: What We Learn from the Big Bad Wolf', in Gail Stygall (ed) *Reading Context*, Boston: Thomson, pp. 405-428.
- Thompson, J.B. (1984) *Studies in the Theory of Ideology*. Oxford: Polity Press.
- Wilmington, Michael (1980) 'Dumbo', in Danny Peary and Gerald Peary (eds) *The American Animated Cartoon: A Critical Anthology*, New York: E.P. Dutton, pp. 80-81.

Mediated Language in Non-native Speaker Texts from the European Commission

AMANDA C. MURPHY

1. INTRODUCTION

The aim of the present article is to examine the English of non-edited texts from the European Commission and compare it with that of the same texts edited by the DGT Editing Service. The interest of this is to investigate editing as a process of mediation, during which language undergoes a process of rewriting and revision wrought by someone who did not write the text. Lefevere (1992: 9) includes editing in his list of rewriting activities, alongside translation, historiography, anthologizing and criticism, and rewriting is interpreted by Ulrych and Anselmi (2008) as a means of mediation, which is extremely important for texts of all types, since mediated texts are actually the form of texts which most readers encounter. In the present paper, mediated texts are investigated in a comparative light, both against the same text previous to their revision, and against the general reference corpus of the BNC.

2. LANGUAGES, MULTILINGUALISM AND TRANSLATION IN THE EUROPEAN UNION

When the six founder Member States signed the treaties of the European Coal and Steel Community (1951) and the European Economic Community (1957), the predecessors of today's European Union, they acknowledged four official lan-

guages: German, French, Italian and Dutch. In the EC Treaty, which founded the EEC, Article 248 declares:

The present Treaty, drawn up in a single original in the German, French, Italian and Dutch languages, all four texts being equally authentic, shall be deposited in the archives of the Government of the Italian Republic which shall transmit a certified copy to each of the Governments of the other signatory States.

The EC Treaty (formally known as the Treaty of Rome) 1957, which established the EEC

This article set the precedent for other Treaties of the Union, and indeed for all legislative documents: it is standard practice in the European Union (henceforth EU) for the different language versions of a legislative document to be considered “originals”: there is no source document and subsequent translations.

The language scenario of the EU in 2008 has changed as much as the geographical reach of the EU: there are now 23 official languages for 27 Member States. The current President of the European Commission, José Manuel Barroso, in office since 2004, has shown particular sensitivity to the issue of interlingual communication, and has instituted two new portfolios in this regard. In his first year of office, he created a portfolio of multilingualism and assimilated it to that of the Commissioner for Education, Training and Culture, Ján Figel'. This led to a Communication by the Commission to the other institutions outlining a new framework strategy on multilingualism. Barroso subsequently appointed a special Commissioner for Multilingualism, a post currently held by Leonard Orban.

Multilingualism is seen as the strong point of the European Union, the emblem of its linguistic and cultural diversity, the “key feature of Europe” according to the report of the High Level Group on Multilingualism of the European Commission. Of course, full multilingualism with translation and interpreting into and out of all languages is impossible, because of a shortage of skilled translators in the languages of the most recently arrived Member States. Given the practical issues involved, its desirability is also questionable.

Following Gazzola (2006), a useful distinction can be made between the practices of translation and interpreting towards citizens and Member States, where fully multilingual communication is in place, and translation and interpreting within the institutions themselves, where communication is partially multilingual. The institutions that are representative of European peoples, governments and regional and local authorities, that is, the European Parliament, the Council and the Economic and Social Committee and the Committee of the Regions, are all fully multilingual, and offer full translation and interpreting services. On the other hand, the Commission, the European Central Bank, the Court of Auditors and the Court of Justice use a limited number of ‘working’ languages. Council Regulation 1/1958, recognised by some working within the EU (eg. Wagner et al. 2002) as the ‘EU Language Charter’, does not stipulate which languages shall be the working languages: English, French and German (in that frequency order) are called the working languages simply because they are the most commonly used for internal activities (Gazzola 2006: 397).

Among the European Institutions, the European Commission has the largest translation service, the Directorate-General for Translation (DGT). DGT employs over 2000 translators, located in Brussels and Luxembourg, and also outsources a considerable amount of work. Under its present structure, DGT is organised into departments, one for each language, and the departments in turn are made up of units which are topic specialised. In this way, translators only work on texts from certain domains, and this has proved to be more effective.

Translators nowadays have sophisticated aids, such as the online interinstitutional terminology database, IATE¹ (interactive terminology for Europe), which gives free access to thousands of terms translated into 22 of the 23 official languages (Irish is at present not yet included). EUR-Lex,² another online database, gives access to all European Union legal documents in all available languages. Besides these resources, which are available for public access, DGT translators also have translation memory tools, such as TRADOS, DGT-VISTA or EURAMIS. The former, a tool which stores translations in real time as they are written, is available commercially, while the latter two are specific to the EU and only draw on a database of EU translated texts. Using these tools, the translator can call up previous translations of phrases from EU texts in specified language pairs via a multilingual concordancer.

DGT does not translate all documents produced within the Commission: the requirement that EU law must exist in all the official languages means that the core of their work deals with proposals for legislation, green and white papers, Commission communications, and scientific and economic reports required prior to legislation drafting. Speeches by Commissioners and personal correspondence are a marginal part of their work.

While in theory all the official languages of the EU can be considered working languages, in practice, the vast majority of texts are drawn up in English. This was estimated at 72% in 2006 by DGT. The prevalent use of English by those working within the EU institutions is a clear sign that practical concerns have the strongest influence on daily lives and working habits, and that English serves as the *lingua franca* of the 21st century, particularly within Europe. It follows that those drafting the documents are, in most cases, not native speakers. The quality of the English in these documents has been cause for concern in recent years, and has led to the institution of an editing service within DGT. The concern about the language of the documents which are translated into other languages was publicly acknowledged by the General Director of DGT, Karl-Johan Lönnroth, at a conference held by the Center for International Cooperation in E-Business in the following terms:

Another challenge which haunts us is quality. This does not only concern the need to develop and update terminology, but also to watch the quality of the originals. In translation we have introduced editing services for French and English to ensure that the source documents (mostly drafted by non-native speakers) are of high linguistic quality and as translatable as we can make them.³

2. The EuroCom corpus and research methodology

EuroCom is a corpus containing documents whose common characteristic is that they originate within the European Commission and have passed through the Editing Unit at DGT⁴. It is a monolingual comparable corpus of specialized texts; the two sub-corpora contain different versions of the same texts, non-edited (EuroCom Non-edited) and edited (EuroCom Edited). It covers a span of 4 years: 2005-2008, and contains 156 documents in their non-edited and edited versions. EuroCom Non-edited corpus totals 1,001,804 tokens, while EuroCom Edited corpus is slightly smaller, standing at 995,451 tokens.

The documents in EuroCom have all been drafted by staff in Directorate-Generals or Services within the Commission. They cover a wide range of domains, coming from DGs as various as Agriculture and Rural Development, Competition, Economic and Financial Affairs, Energy and Transport, Information Society and Media, Justice, Freedom and Security and Taxation and Custom Unit, to name just a few. The text types in the corpus vary, but consist mostly of Communications from the Commission, which aim to open up debate within the institutions prior to proposals for secondary legislation (Regulations, Decisions and Directives), reports for the European Council and the European Parliament on the implementation of Regulations and Directives, documents for internal use only, such as working papers, work programmes, minutes of meetings, staff notices, and brochures, guides, and press releases produced for the general public, as well as texts to be published on the Europa website.

The methodology adopted for the present study fits into the framework of corpus-assisted discourse analysis (Partington 2004). This involves both quantitative and qualitative analyses, using the software *Wordsmith Tools 4* (developed by Scott 2006) for the former, and reading and comparing the texts in the two corpora for the latter. The importance and advantages of using a complementary approach have been emphasized by many linguists (Biber, Conrad, and Reppen 1998; Kennedy 1998; McEnery and Wilson 1996; Sinclair 2006 *inter alia*): a quantitative analysis counts and classifies features, and pays considerable attention to frequency, while a qualitative analysis aims to provide complete and detailed description within a context. The investigation of the features of the texts started from a quantitative analysis, examining wordlists of the two corpora, divided into one, two, three and four-word clusters. Comparisons of frequency of occurrences led to their investigation in the two versions of the texts. The qualitative investigations of the text, on the other hand, involve examining the texts in the form they leave the Editing Unit. This is as a Microsoft Word document, where the revision and tracking functions are activated, allowing the editor's interventions to be displayed visually either within the text, or in 'text bubbles' that appear in the right-hand margin of the document.

3. GUIDES FOR WRITERS OF EC DOCUMENTS AND THE EDITING SERVICE

All writers of EC documents have access to the online Interinstitutional Style Guide, produced by the European Union Publications Office in all the official languages. This guide provides the conventions and rules for documents from all

the EU institutions, and advises that a consistent style in documents will contribute to cohesion between them. The appropriate way to refer to types of legal acts is listed, together with conventions for referring to the institutions, official titles within the institutions, references to countries, languages, currencies, etc. For English, the Guide explains punctuation rules and British spelling conventions which are adopted within the documents, and also warns against interference from French, with a list of the most common false ‘spelling’ friends, such as *correspondance* / *correspondence*, and *defense* / *defence*.

More specifically for English, two documents for writers have been written by translators at the European Commission: *How to write clearly*, which was drafted during the *Fight the Fog* campaign for writing clear English that started from the English section of DGT in Luxembourg in 1998, and *The English Style Guide*. The basic message of *How to write clearly* includes the following points:

- a) put the reader first – the general lay reader is the one who will find documents about the European Institutions hardest to understand; avoid gender-specific terms; use plurals instead;
- b) use verbs, rather than abstract nouns; avoid nominalisation;
- c) use concrete nouns rather than abstract ones;
- d) use active verbs, not passives, unless they are deliberately chosen to avoid mentioning responsibility or agency; don’t be afraid to use first person agents for the verb;
- e) place old information at the beginning of the sentence and new, important or complex information at the end;
- f) KISS – keep it short and simple;
- g) avoid false friends between French and English.

The *English Style Guide*, on the other hand, is a handbook for authors and translators in the European Commission, written by translators working at DGT. It offers the consolidated experience of more than 25 years translation of EU documents – the first edition is dated 1982 – and is updated regularly online (latest version April 2008). In its introduction, it clarifies the sense in which “style” is used: “a set of accepted linguistic conventions; [...] recommended in-house usage”. Its aim is to set a good example by using English that is “as clear, simple, and accessible as possible out of courtesy to our readers and consideration for the image of the Commission”. It clearly reflects the preoccupations of the briefer *Fight the Fog* booklet, and cites the same reference works (Cutts 1996, Williams 1995).

The *English Style Guide* is divided into two parts: one deals with linguistic conventions applicable in all contexts – spelling conventions, information on using upper and lower case, geographical names, hyphens and compound nouns, punctuation, numbers, dates and times, abbreviations, and acronyms. There is a section on the agreement of verbs with singular or collective nouns, verb tenses (past simple versus present perfect), notes on the tenses of verbs to be used in minutes of meetings, and on appropriate verbs in legislation, as well as advice on the split infinitive. It has a section on scientific language, footnotes, translating correspondence, personal names and titles, and gender-neutral language. The second part has a more encyclopaedic function, and deals with the workings of

the European Union, and how they are expressed in English, sometimes accounting for them by referring to the original French.

Despite the guides that are available, a few years after the *Fight the Fog* campaign, there was a general consensus within DGT that translators were spending an excessive amount of time editing texts written in poor English. A pilot service was thus set up to start *editing* documents written in English and French from two DGs, Economics and Finance, and Environment. Staffed by ex-translators with considerable experience of translating for the EU institutions, the Unit met with immediate success, and this led to the establishment of a dedicated Unit – the Editing Unit – within DGT.

4. CLASSIFYING THE REVISIONS

Revisions to translated texts have been studied by Rega (1999), Scarpa (2008), Mossop (2001) and Cosmai (2007). Cosmai, who works as a translator/reviser at the Committee of the Regions at the European Union, classifies revisers' interventions into three categories: subjective, objective and specialized (2007: 102). In the case of editing non-native English, however, the question is perhaps more complex. Some revisions are indeed objective, particularly those referring to the house-style of the European Union institutions (capitalisation, names, dates, numbers, etc.), and grammatical revisions (prepositions, verb tenses, concord of noun and verb). But many revisions can be described simply as ways of improving the text, rendering it more natural, or smooth. Changing a post-modifying phrase into a pre-modifying phrase, or a passive verb into an active one may be ways of making the text more incisive, more concise, but they are not corrections of elements that are incorrect.

Two approaches were taken here to the description of the revisions. Firstly, on a sample of 15 documents, all the revisions were categorized under a series of labels. These labels are:

- a) *Objective revisions*: including layout (capitalisation, acronyms, etc), grammar, punctuation, syntax and spelling
- b) *Overall improvements*: including passages that are rewritten for content, wrong lexical choices, and changes in style (e.g. changing passive verbs to active verbs, post-modifying phrases to pre-modifying phrases).

Secondly, following the guides recommended by the EU Institutions (such as the *English Style Guide*, and *How to write clearly* booklet, and some *Tips for Writing English* drawn up by translators), the whole corpus of edited texts was analysed for certain overall improvements. Apart from the features listed in b), elements such as foggy phrasing, Euro-speak, false friends, and bureaucratic turns of phrase were examined.

On the first sample of 15 documents, 'objective' revisions accounted for 69% of revisions, whereas those which improve the text overall account for 31%. Excerpt 1 shows a text that has been edited in which both types of revisions are visible. Each revision is commented on briefly.

Visits to DGT in September and October 2006

On 18 September 2006, DGT received a group of students from ~~the~~ Drew University, in the USA. It is worth quoting their ~~reasons for wishing motivation~~ to visit our service: “We do understand the importance of teaching our students about learning to communicate in another language and to understand its culture. The European Union is the perfect example and we believe that this visit to your institution could change how American students view the importance of multiculturalism and multilingualism in our world”. ~~These~~ students were particularly curious to learn more about our working methods and computer-assisted translation.

Excerpt 1. Objective Revisions and Overall Improvements

In Excerpt 1, the objective revisions (which refer to layout elements such as capitalisation and acronyms, grammar, punctuation, syntax and spelling) include the deletion of the definite article *the* before the name of Drew University, the added comma and the deletion of *A* from the abbreviation *USA*, the change to single inverted commas from double, and the replacement of the demonstrative adjective *these* by the definite article *the*. The overall improvements are the added prepositional phrase *in the* before *US*, and the lexical change of *motivation* to the noun phrase *reasons for wishing*. While the former type of revisions are objectively explicable (e.g. it is the policy in EU documents to refer to the *US*, not the *USA*), the latter are more complex and subjective interventions carefully weighed. The added prepositional phrase *in the* fleshes out the text, creating better balance, but the original version cannot be said to be wrong. Similarly, the change from the Latinate nominalization *motivation* to the more phrasal *reasons for wishing* marks a change from longer words to shorter words, and from Latinate to German compound phrases, which is intended presumably to make the text more reader-friendly (the text was destined for the DGT website, so for a non-specialized readership). Other types of lexical revisions are classed as objective revisions, if for example they regard typical EU terms. An example is *services* being changed to *departments*: while there are some services which have kept their original French name (such as *Interservice consultation*) or the Legal Service, generally most services are called departments within the EC. Changing the word is thus not a lexical choice, as it were, but a conforming of terms to the objective norms of the Commission.

5. REFLECTIONS ON THE EDITED-MEDIATED CORPUS

In this section, selective observations will be made about the results of the quantitative analyses of the two corpora. Further findings are reported in Murphy (2008).

From a corpus-driven perspective (Tognini-Bonelli 2001), one way of investigating comparable corpora is to compare single-word lists and cluster word lists. Clusters of two and three words were examined across the EuroCom Edited and Non-Edited corpora, and findings from each examination will be reported here.

From the comparison of two-word clusters, it emerged that the two-word cluster *of the* is edited out more than 700 times in the EuroCom edited corpus. Occurrences in EuroCom Non-edited stand at 11,996, and in EuroCom Edited at 11,286. Close examination of some of the occurrences reveal that *of the* is often edited out when it is part of a post-modifying phrase. The editors frequently turn the *of the* phrases into pre-modifying phrases. To find examples of this, the three-word wordlist was useful. In this wordlist, it was seen that phrases beginning with *of the* + *adjective/noun* were fewer than in the edited corpus, examples from the non-edited corpus being *of the European, of the EU, of the Member, of the Council, of the Commission, of the Directive*. Examples 1 and 2 show how one such phrase *of the European citizens* is changed by the editors from a post-modifying position to a pre-modifying one:

1) It will provide a clear signal of support to the **rights** to information and freedom of expression **of the European citizens** and confirm communication as an EU policy at the service of citizens and democracy.

(EuroCom Non-edited Corpus)

>

2) It will provide a clear signal of support for **European citizens' rights** to information and freedom of expression and confirm communication as an EU policy, at the service of citizens and democracy.

(EuroCom Edited Corpus)

Another finding from the comparison of three-word clusters was the pattern *the + noun + of*, with tokens such as *the implementation of, the application of, the protection of, the creation of, the adoption of, the definition of, the admission of*. These patterns involve a nominalised verb. In EuroCom Non-edited, there are 3012 occurrences of these phrases, and 2651 in EuroCom Edited. This shows a tendency to eliminate nominalisations, although they are still present in the Edited corpus. Example 3 shows an example of a nominalisation in the pattern *the + noun + of* (i.e. the provisions of) that has been edited out:

~~Moreover, all sStaff members are in particular reminded in particular of:~~

- ~~the obligations laid downser out in the Staff Regulations of officials and the cConditions of employment of other servants of the European Communities, and, notably, the provisions ofin particular Article 12 of the Staff Regulations, which states that an official "shall refrain from any action or behaviour which might reflect adversely upon his position". Therefore, any abuse linked to personal use of the ICT services at the disposal of themade available to staff member is prohibited (see Report on the work of the Investigation and Disciplinary Office (IDOC)² and relevant case law);~~

Example 3 *the provisions of* (line 4) edited out

The third observation reported here regards the comparison of phraseology in EuroCom Edited corpus and the BNC. By comparing the two versions of EuroCom, it came to light that the three-word clusters *in order to* and *as well as* are among the most frequent in both corpora. Although on some occasions they are added by the editors, the tendency is for editors to edit them out: occurrences in EuroCom Non-edited amount to 183 and 125 respectively, and 139 and 99 in EuroCom Edited. They remain among the 9 most frequent three-word clusters in the edited corpus. A look at the most frequent three-word clusters in the BNC reveals that neither of these phrases occur within the top 1000 3-word clusters in the BNC. This is an interesting finding, in that it would appear to characterize the EuroCom Edited corpus, but not the BNC. It becomes more significant, perhaps, if one reflects that both phrases have *filler* functions: *in order to* (a complex preposition) makes the purpose of something explicit, it extends and clarifies the meaning of the to-form of the verb. Example 4 illustrates one such case from the corpus, where the complex preposition has been added in by the editors, presumably for the sake of clarity:

4) With a view to further development of a coherent immigration policy, **in order to narrow** the rights gap between EU citizens and third-country nationals legally working and to complementing the existing immigration *acquis*, a set of rights should be laid down, in particular in the form of specifying the policy fields where equal treatment with nationals is provided for third-country workers legally admitted into a Member States but not yet long-term residents. (EuroCom Edited corpus)

The interpretation given here is that the editors are making the purpose of *narrow the rates gap* clearer, more explicit, by lengthening *to* to *in order to*. Explicitation is also seen as the process behind the phrase *as well as*. In Excerpt 2 it can be seen that one example of *as well as* has been kept (in line 8), whereas in line 14 it has been deleted.

4. FRANCE

A new law on equal rights and opportunities, participation and citizenship of persons with disabilities came into force on 12 February 2005 (Loi No 2005-102 of 11 February 2005). This law **puts in place-establishes** the principle of non-discrimination on grounds of disability, for both the private and public sectors. It obliges employers to provide reasonable accommodation for disabled workers, as long as the cost is not disproportionate. Special arrangements can be made for disabled candidates to take entrance exams for the public service, and it will be possible to employ disabled civil servants under contract as well as by entrance **competitionsexaminations**. Disability associations which have been legally established for at least five years will be able to give legal help to victims of discrimination. **Measures of p**Positive action in favour of disabled workers **are-is** not considered discriminatory. Employers will have to negotiate annually on **disabled persons' the** conditions of access to employment of **disabled persons, of their** vocational training and working conditions, **as well as and** actions to raise awareness **ofamong** all workers on disability issues. The Government must produce an annual report on the employment of disabled people in the public service. A national Observatory on training, research and innovation in the field of disability will be created.

Excerpt 2 Examples of *as well as*

This complex conjunction is also a filler, a longer way of saying *and*, and it is eliminated to a greater extent than *in order to*, but it still occurs within the 9 most frequent three-word phrases in the Edited corpus, but not among the 1000 most frequent three-word phrases in the BNC.

6. CONCLUSIONS

Research into edited-mediated language in the EuroCom corpus is still in its infancy. However, a few preliminary conclusions may be drawn on the basis of the empirical investigations. Firstly, from the point of view of the editors, by looking at the revisions of texts overall, it appears that approximately two thirds of the alterations made to the text regard objective matters: i.e. matters of house style (capitalisation, layout, abbreviations), grammar or spelling. While grammar remains something of which the individual writer may have a shaky grasp, house style and spelling could be corrected by a computer programme. Intervening in many situations where the accepted version is an objectively established one is perhaps not the best use of editor's time. Secondly, at a higher level of mediation of the text, the overall improvements wrought to the text are rather personal. Changing post-modifying phrases to pre-modifying ones or eliminating nominalisation is certainly a subjective exercise, which undoubtedly varies from one editor to another. Further research might compare the work of individual editors, and gauge the extent to which such changes are uniform across the service.

From the descriptive point of view, it can be said that two tendencies have emerged through comparing the EuroCom Edited Corpus with the EuroCom Non-edited one: firstly, a move towards conciseness. Moving post-modifying phrases to a pre-modifying position makes for tighter syntax and tauter prose. Conciseness may be a characteristic of mediated text. Contemporaneously, an opposing tendency is noted: that of explicitation. The complex phrases *in order to* and *as well as* lengthen and make links that are already in the text more explicit, and characterise the EuroCom Edited corpus significantly in terms of phraseology, whereas they do not characterize the reference BNC. Further research into such issues within these corpora may explain how two opposite tendencies happily co-exist within mediated text.

NOTES

- 1 URL: <http://iate.europa.eu>.
- 2 URL: <http://eur-lex.europa.eu>.
- 3 http://ec.europa.eu/translation/reading/articles/pdf/20060921_ciceb-translation_practices_en.pdf
- 4 I gratefully acknowledge the permission of the Head of the Editing Unit, David Crowther, for the opportunity to study these documents.

REFERENCES

- Biber, Douglas, Conrad, Susan and Reppen, Randi. 1998. *Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Cosmai, Domenico. 2007. *Tradurre per l'Unione europea. Prassi problemi e prospettive del multilinguismo comunitario dopo l'ampliamento ad est*. Milano: Hoepli.
- Cutts, Martin. 1996. *The Plain English Guide*, Oxford: Oxford University Press.
- EUR-LEX. Joint Practical Guide of the European Parliament the Council and the Commission for persons involved in the drafting of legislation within the Community Institutions. Available online at <http://eur-lex.europa.eu/en/techleg/1.htm>. Accessed June 16, 2008.
- European Commission Directorate-General for Translation. 2006. *English Style Guide. A handbook for authors and translators in the European Commission*. 5th edition. Available at ec.europa.eu/translation/writing/style_guides/english/style_guide_en.pdf. Accessed June 16, 2008.
- Fight the Fog Campaign, How to Write Clearly. Available at http://ec.europa.eu/translation/writing/clear_writing/fight_the_fog_en.pdf.
- Gazzola, Michele. 2006. *Managing Multilingualism in the European Union: language policy evaluation for the European Parliament*. *Language Policy* 5/4: 395-419.
- Kennedy, Graeme. 1998. *An introduction to Corpus Linguistics*. London: Longman.
- Lefevre, André. 1992. *Translation, Rewriting and the Manipulation of Literary Fame*, Routledge, London and New York.
- McEnery, Tony and Wilson, Andrew. 1996. *Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Mossop Brian, 2001. *Revising and editing for translators*. Manchester: St Jerome Publishing.
- Murphy, Amanda C. 2008. *Editing Specialised Texts: a corpus-assisted analysis*. Milano: Led.
- Partington, Alan. 2004. *Corpora and discourse, a most incongruous beast*. In Partington, Alan, Morley, John, Haarman, Louann. (eds.)

Corpora and Discourse. Bern: Peter Lang. 11-20.

Rega, Lorenza. 1999. Alcune considerazioni sul problema della revisione nell'ambito della traduzione. *Rivista internazionale di tecnica della Traduzione / International Journal of Translation* 4: 67-78.

Scarpa, Federica. 2008². *La Traduzione Specializzata*. Milano: Hoepli.

Scott, Mike. 2006. *Wordsmith Tools. Version 4.0*. Oxford: Oxford University Press.

Sinclair, John. 2006. The Search for Units of Meaning. *Textus*. 9/1: 75-106.

Tognini-Bonelli, Elena. 2001. *Corpus Linguistics at work*. Amsterdam / Philadelphia. John Benjamins Publishing Company.

Ulrych, Margherita and Anselmi, Simona. 2008. Towards a corpus-based distinction between language-specific and universal features of mediated discourse. In Martelli, Aurelia and Pulcini, Virginia (eds.) *Investigating English with Corpora. Studies in honor of Maria Teresa Prat*. Monza: Polimetrica, 257-273.

Wagner, Emma, Bech, Svend and Martínez, Jesús M. 2002. *Translating for the European Institutions*. Manchester: St Jerome Publishing.

Williams, John. 1995. *Style: Toward Clarity and Grace*. Chicago: The University of Chicago Press.

The get-unit in Corpora of Spontaneous and Non-spontaneous Mediated Language: from Syntactic Versatility to Semantic and Pragmatic Similarity

PIERFRANCA FORCHINI

Università Cattolica del Sacro Cuore (Milano)

1. INTRODUCTION

Several interesting observations have been made about the multiple facets of the verb *get*. However, interest has mainly been focused on the degree of difference between structures such as the *get passive* and the *be passive* (cf. Hatcher 1949, Gee 1974, Standwell 1981, Haegeman 1985, Collins 1996, Rühlemann 2007, *inter alia*). Less frequently, comparative studies between *get* and other verbs like *have* (cf. Kimball 1973, Pizzini 1975, Johansson and Oksefjell 1996, Gilquin 2003), *give* (cf. Cattel 1984), *become* (cf. Quirk *et al.* 1991), *be* and *keep* (cf. Johansson and Oksefjell 1996) have been performed in order to investigate some specific features of *get*. Little attention seems to have been paid to what this study calls the *get-unit*, namely, the framework determined by *get* and the *environment* in which it occurs. The present paper explores the nature of this *get-unit* in American English, with particular regard to the functions it performs in spoken language, both spontaneous and non-spontaneous.

The primary aim is to demonstrate that, despite its syntactic versatility (cf. also Quirk *et al.* 1991 and Biber *et al.* 1999), *get* can be depicted semantically and pragmatically as a general *result marker* and as mainly displaying a *negative semantic prosody* (cf. Louw 1993, Stubbs 2001, Sinclair 2004), on the basis of the resultative quality of the sentences in which it locates, and the negative contexts in which it has a propensity to occur. Secondly, the paper intends to provide an explana-

tion for these semantic and pragmatic similarities. Accordingly, two interrelated and complementary meanings on which the semantics of the *get-sentence* may exclusively depend are here suggested: the *core* and the *peripheral meaning* of the *get-unit*. This underlying theory, which originates from Lindstromberg's (1991) claim that *get* should not be seen as "polysemic in the common sense of the term" but "as having different shades [...] of meaning which stand in a non-complex, semantically motivated relation to each other" (Lindstromberg 1991:285) and Johansson and Oksefjell's (1996:73) intuition that "despite the variety in syntax and semantic content, there appears to be a prototype to which all constructions conform more or less closely", reflects a tradition which does not perceive lexis and grammar separately, but rather interactively and complementarily (cf. Firth 1957a, 1957b; Sinclair 1991, 2003, 2004; Halliday 2003, 2004). *Get* plus the environment in which it occurs (or, in Hunston and Francis' 1999 terminology, plus its *complementation patterns*) are, indeed, perceived here as a *unit of meaning* (cf. Sinclair 1996, 1998, 2004).

Thirdly, the paper investigates the extent to which spontaneous and non-spontaneous conversation differ with particular regard to the features and functions of the *get-unit*. To do so, movie conversation is analyzed as an example of non-spontaneous conversation, both because it is written-to-be-spoken (cf. Gregory 1967, Nencioni 1976, Taylor 1999, Rossi 2003, Pavesi 2005) and because it is mediated, i.e. it derives from a script and it is broadcasted by a televised medium (cf. Ulrych & Anselmi 2008).

The analyses, which are based on authentic data retrieved from the US spoken sub-corpus of the *Bank of English* (henceforth USBoE; i.e. about 30 million words) and on an American movie corpus made up of 6 manually transcribed scripts (henceforth AMC6; i.e. about 60,000 words)¹, are conducted according to both corpus-based and corpus-driven methodologies (cf. Tognini-Bonelli 2001). Due to the large number of occurrences of *get* in the USBoE subcorpus (i.e. 35,860) and to the impossibility of manually checking all of them in context, the corpus-based approach is followed to verify the presence of the patterns of *get* which are usually described in reference grammars (cf. Quirk *et al.* 1991 and Biber *et al.* 1999) in American conversation, whereas the corpus-driven approach is employed to explore the functions that the *get-unit* performs in the data retrieved from the corpora. More specifically, the USBoE corpus is investigated according to sample selection criteria (cf. Sinclair 1999 and Hunston 2002), while the AMC6 corpus is investigated exhaustively. This choice depended on the exceedingly large size of the former compared to the manageable size of the latter; numbers are normalized to 100% so as to allow comparability.

Conceptually, the present account is divided into two parts: Section 2 offers a syntactic description of the types of the *get-sentences* present in spoken American English, whereas Section 3 is a tentative explanation of their semantics and pragmatics. The descriptive section gives an account of the syntax of all the possible clause patterns and classes of verb complementation² in which *get* locates as a full and as an auxiliary-like verb (cf. Quirk *et al.* 1991 and Biber *et al.* 1999); the explanatory section, instead, illustrates the resultative character of what I call the *result marker get* by hypothesizing two basic meanings responsible for the seman-

tics of the *get-sentence* (i.e. the *core* and the *peripheral meaning* of the *get-unit*). Section 3 also highlights the tendency of *get* to occur within negative contexts, and the causative mark it acquires when preceding a noun phrase (henceforth NP).

2. SYNTACTIC VERSATILITY

The spoken data from the USBoE and the AMC6 corpora qualitatively demonstrate that the *get-sentence* is extremely variable in syntax. In particular, syntactic flexibility is manifested in the multiplicity of the uses of *get* both as a full and as an auxiliary-like verb. As examples 1-7 illustrate, indeed, the full verb *get* can locate in all the seven clause types present in the English language posited by Quirk *et al.* (1991) (cf. also Biber *et al.* 1999³): SVC and SVA types (i.e. copular or linking complementation); SVO, SVOO, SVOC, and SVOA types (i.e. transitive complementation); and SV types (i.e. intransitive complementation).

1. (S)VC⁴:

(a) A lot of times in the early morning, like if I'm on a train or something, I **get** nauseous; I **get dizzy**. If I keep my hands up, it seems like I **get** real dizzy. [USBoE]

(b) I may not be tough, but, damn it, I can **get good** and mad! Come on! [AMC6]

2. (S)VA:

(a) There are only a few ways to **get to the bottom of the Los Angeles River**, either through a hole in the fence or by driving or walking down one of the official ramps or tunnels. Just give me all the details. [USBoE]

(b) And I'll **get to the bottom of it**. All right? [AMC6]

3. (S)VO:

(a) Just a few blocks away on West 44th Street, there's a place where the homeless can **get attention** not available anywhere else in this often harsh city. [USBoE]

(b) Well, I kind of fudged my resume a little bit to **get that job**. [AMC6]

4. (S)V_{OO}:

(a) I remember, you know, I helped get him a job. [USBoE]

(b) Listen, you gotta get a message to the colonel for me. [AMC6]

5. (S)V_{OC}:

(a) they're moving the 'Simpsons' up against 'Cosby,' but they're going to be running reruns until mid-October because that's how long it will take to get their shows ready, and there won't actually be any new 'Cosby Show' episodes until the end of September. [USBoE]

(b) Get your foot long and a bag of nuts. [AMC6]

6. (S)V_{OA}:

(a) Do you think it's possible to get him to the table at all? [USBoE]

(b) Yes. And you get us to safe place with them in Atlanta, thank God. [AMC6]

7. (S)V [?'s]:

(a) Why don't you get away. **Get away** from me. [USBoE]

(b) Everybody down! Get down! [AMC6]

Similarly, the auxiliary-like *get* can occur in six out of eight catenative structures (cf. Palmer 1988, Huddleston and Pullum 2002), both simple (as in examples 8, 10, and 12) and complex⁶ (as in examples 9, 11, and 13). The only structures with which it does not occur are the simple and complex bare infinitives. It is worth noting, however, that this lack does not seem to have repercussions on the versatile nature of *get* in that, as Huddleston and Pullum (2002:1244) maintain, "only a relatively small number of catenatives take bare infinitivals". Consequently, catenative *get* can still be considered versatile in its nature by covering the most common categories of catenative verbs.

8. **GET + PAST PARTICIPLE:**

(a) When we got on the Internet, on CNN, we saw the second tower get hit. [USBoE]

(b) You get robbed or something. [AMC6]

9. **GET + O + PAST PARTICIPLE:**

(a) She hopes to contact more authors in the future, and get them involved in the project as well. [USBoE]

(b) I'm gonna get my car washed. [AMC6]

10. **GET + PRESENT PARTICIPLE:**

(a) Well, if the Braves don't get moving, they're going to run out of second chances. [USBoE]

(b) Should we get going? [AMC6]

11. **GET + O + PRESENT PARTICIPLE:**

(a) That's pretty tough when they're bored, they're tired, they don't like the book they read to begin with, and you want to get a discussion going with 20 15-year-olds. [USBoE]

(b) Let's get this going. [AMC6]

12. **GET + INFINITIVE:**

(a) Most Americans never get to rest, and many of us have even lost the ability to relax. [USBoE]

(b) But in return they get to be a lot thinner. [AMC6]

13. **GET + O + INFINITIVE:**

(a) Foreign Minister Levy was unable to **get Secretary Baker** to approve the loan guarantees yesterday. [USBoE]

(b) I don't think I can get her to do it. [AMC6]

Another mark of *get* versatility is encountered in its passive forms: even though, by way of simplification, the *get passive* is included here in the more generic category of *catenatives* (i.e. in the *get + past participle unit*), it can be further categorized into five types of passive constructions. More precisely, as suggested by Collins (1996:45-49), *get passive* can be described as:

1) *Central*, when it may be related to an equivalent active clause⁷, with or without an agent, as in:

14. (a) A problem especially because federal workers often **get paid** less than their counterparts in the private sector. [USBoE]

(b) Why, I **get paid** to count 'em [AMC6]

2) *Psychological*, when it occurs with agent-phrases which refer to various entities or phenomena which initiate psychological processes, as in:

15. (a) This Chicago policeman, who asked not to be named, said that many policemen **get frustrated** because it's so hard to make a legal arrest. [USBoE]

(b) Come on, man, nobody's gonna **get hurt**. [AMC6]

3) *Reciprocal/Reflexive*, when it occurs with reciprocal past participles such as *married* or reflexive past participles such as *dressed*, as in:

16. (a) After a 10-day whirlwind courtship, they decide to **get married**. [USBoE]

(b) Hey mudwhistle, **get dressed** [AMC6]

4) *Adjectival*⁸, when its central members are adjectives (such as *ready*, *angry*, *lost*, *drunk*, *burned*, etc.), as in:

17. (a) They're very very hopeful, but they don't want to **get burned**.
[USBoE]

(b) Good. Let's all finish up and **get ready** to go. [AMC6]

5) *Formulaic*, when it is an idiomatic expression such as *get accustomed to*, *get used to*, etc., as in:

18. (a) Sure do, yeah. The more time--you **get accustomed to** the climate and to the environment around here, the better off you are.
[USBoE]

(b) Well, he wants me to rest while I'm **getting used to** the medication. [AMC6]

3. SEMANTIC AND PRAGMATIC SIMILARITY

Although from a syntactic point of view, *get* shows versatility with respect to the contexts in which it can locate, from a semantic and pragmatic perspective, it seems to express a significant degree of similarity. The data from the two corpora investigated, indeed, strongly suggest that *get-sentences* usually share a constant resultative quality, that *get* tends to occur within negative contexts, and that, when it is followed by an NP, it always acquires a causative meaning keeping, at the same time, the pragmatic functions of the corresponding sentences without NP. This may be due to the fact that the semantics and pragmatics of the *get-sentence* depend on two interrelating and complementary levels of functional meaning, namely, on what I call the *core* and the *peripheral meaning*, which are common to all *get-units*. More precisely, by *get-unit*, I mean the framework determined by *get* and the *environment* in which it appears; a unit whose parts "cannot retain independent meaning" (Tognini-Bonelli, 2001:101), even though each individual part carries a different function: *get* determines the *core meaning*, whereas the *environment* in which *get* occurs determines the *peripheral meaning* of the *get-unit*. The former meaning is constantly characterized by the feature *result*, which is typical of the *result marker get*, while the latter acquires variable semantics according to the specific *environment* in which *get* locates. In the binary notion I am suggesting, the "prototype to which all constructions conform more or less closely" posited by Johansson and Oksefjell (1996:73) may be envisaged.

Indeed, both the USBoE and AMC6 data demonstrate that the *resultative* feature, the tendency to occur in negative contexts, and the causative meaning acquired by the NP-sentences can be applied to any structure related to *get*. Sections 3.1 and 3.2 will qualitatively and quantitatively demonstrate this in detail.

3.1 GET AS A FULL VERB

Copular or *linking verbs* must be followed by a complement for the sentence to be complete. Specifically, there are two cases in which verbs may be said to have *copular* or *linking complementation*: when they are followed by a *subject/object complement* in types which belong to the SVC pattern, and when they are followed by a *predication adjunct* in types which are of the SVA pattern (cf. Quirk *et al.* 1991). In the former context, the *peripheral meaning* of the *get-unit* is claimed here to be determined by the *complement*, whereas in the latter, by the *predication adjunct*. In both cases the *core meaning* of the *get-unit* is contingent on the *result marker get*. When occurring in SVC types, both the USBoE and movie data show that the meaning of the *get-unit* is resultative in that it implies some change, or transformation. As illustrated in examples 19 and 20 from the USBoE and AMC6 respectively, *get* functions as a *result marker*, or a *resulting copula* (cf. Cattell 1984, Quirk *et al.* 1991, Johansson and Oksefjell 1996, Biber *et al.* 1999), which emphasizes a change of state (i.e. *getting nauseous, dizzy, and sloppy*); the *subject/object complement*, instead, specifies the kind of transformation brought about by *get*, which does not need to be an actual change (as in *get nauseous, get dizzy*), but can simply be, for instance, a metaphorical or hypothetical mutation (as in *get sloppy*), etc.

19. A lot of times in the early morning, like if I'm on a train or something, I **get** nauseous; I **get** dizzy. If I keep my hands up, it seems like I **get** real dizzy. [USBoE]

20. Guys like us don't change, Saul. We stay sharp or we **get** sloppy. We don't change. [AMC6]

The SVC *get-unit* not only shares the same resultative connotation in the two corpora under investigation, but also its semantic prosody: as Table 1 illustrates, the SVC *get-units* present in the USBoE tend to display a negative semantic prosody: the transformation/change of state they indicate is mostly negative (736 negative vs. 365 positive out of 1101, i.e. 66.84% vs. 33.15% respectively) in that it is expressed mostly by complements like *worse, hurt, sick, angry, mad*, etc.; similarly, those present in the AMC (cf. Table 2) show a preference of occurrences in negative contexts (13 negative vs. 3 positive out of 16, i.e. 81.25% vs. 18.75% respectively) collocating especially with complements like *hurt, lost, old, sick, tired*, etc..

RIGHT COLLOCATES	#
worse	171
better	152
ready	109
hurt	97
sick	92
tough	79
angry	46
mad	38
rich	31
older	30
nervous	28
serious	27
right	24
tougher	22
old	21
bigger	20
richer	16
tired	13
hot	11
emotional	9
(PRETTY¹)	(8)
wet	7
dirty	6
smaller	6
tense	6
negative	6
greedy	6
violent	5
weaker	5
dizzy	5
desperate	5
TOTAL	1101

Table 1. USBoE right collocates of *get* in the SVC-unit (negative changes of state in bold)

RIGHT COLLOCATES	#
hurt	4
lost	2
ready	2
old	1
sick	1
tired	1
sloppy	1
real	1
squeamish	1
chapped	1
good	1
TOTAL	16

Table 2. AMC6 right collocates of *get* in the SVC-unit (negative changes of state in bold)

When occurring with place adjuncts in SVA types, both the USBoE and movie data show that the meaning of the *get*-unit expresses *movement* and *result*. As for the former type of *peripheral meaning*, i.e. *movement*, this is not surprising in that in such a complementation pattern *get* occurs only with *place adjuncts*, and not with *time*, as *be*, for instance, may do. The latter type, instead, i.e. *result*, raises interest: it seems that *get*, unlike *go* (the prototypical verb of motion), for instance, does not only imply a change of location, but it also adds some information about the movement it implies. This may be ascribed to the fact that the non-neutral *get* acts like a *result marker* by focusing on the difficulty or the unfavourable situation in reaching B and, consequently, on the result of the action. It may be speculated that this focus on the result of the action expressed by *get* may be brought about by at least two main strategies: either by the speaker wanting to create some expectation about the result of the action (i.e. meaning *will the subject be able to reach X?*) or by his/her underlining that despite (or because of) the difficult situation, the subject has (not) been/will (not) be able to reach X; indeed, the most frequent occurrences in both corpora of the SVA *get*-unit in negative contexts (i.e. 73% in the USBoE and 54% in the AMC6) show that *get* is employed, under both the circumstances mentioned, to highlight the doubt, effort, fatigue, or danger of the situation.

This difficulty in reaching the point is illustrated by the few chances (i.e. *a few ways* and *the only way*) to manage to do so (i.e. to reach *the bottom of the Los Angeles River* and *the future*) in examples 21 and 22:

21. There are only *a few ways* to **get** to the bottom of the Los Angeles River, either through a hole in the fence or by driving or walking down one of the official ramps or tunnels. [USBoE]

22. We're all here to do what we're all here to do. I'm interested in one thing, Neo, the future. And believe me, I know - *the only way to get there* is together. [AMC6]

Essentially, there are three kinds of verb complementation that correlate with *get* transitivity: monotransitive, ditransitive, and complex transitive verb complementation. The first kind is encountered when *get* appears in SVO clauses; the second when it locates in SVOO clauses; and the third when it is used either in SVOC or SVOA clauses. It is assumed here that in transitive verb complementation, the *core meaning* of the *get* unit is determined by *get*; whereas the *peripheral meaning* is determined by the object(s), by the object plus the complement, or by the object plus the adjunct respectively.

Interestingly, when *get* occurs in SVO clauses, the *get-unit* can express, in both the USBoE and movie corpus, a binary notion of *possession* for, this time, *get* can express two (and not one like in the SVC and SVA types) *core meanings*, namely, *stative* and *dynamic possession* (cf. LeSourd 1976, Quirk *et al.* 1991, Tobin 1993, Biber *et al.* 1999). In particular, when the *get-unit* expresses *possession* in the most general sense, *get* is *stative* without being marked by the *resultative* feature it would normally display; indeed, in examples 23 and 24, *get/got* means *possess*, as *have* does in examples 25 and 26:

23. They got a chance to use it now [USBoE]
24. I'm sorry you lost him. Hey, listen, I know you got a great heart. You're just fixated on the outside appearances of people so. Hal Larson, I'm gonna do a great favour. [AMC6]
25. Then u have a chance to win everything. [USBoE]
26. Well, you have a better idea? Come on. [AMC6]

On the other hand, when the *get-unit* expresses *dynamic possession*, *get* can be paraphrased by *come to possess/acquire*; indeed, it is marked by the *resultative* feature that already emerged in the SVC and SVA types, as examples 27 and 28 demonstrate:

27. Just a few blocks away on West 44th Street, there's a place where the homeless can get attention not available anywhere else in this often harsh city. [USBoE]

28. You mean you **get** the hyena, and I choose between the hippo and the giraffe? [AMC6]

As Table 3 shows, this binary notion of possession does not emerge by looking at the collocates of dynamic and stative *get*; indeed, in both cases it occurs with similar objects like *job*, *chance*, *place* and *gun*; conversely, it does emerge by investigating the context in which *get* locates. More specifically, when dynamic *get* is employed, there is some acquisition involved: dynamic *get* is often found in sentences containing *if*, *when*, and modal verbs, for instance, which suggest that there is no actual possession, but only some hypothesis or future referent about it. On the contrary, when stative *got* is employed, there is no acquisition involved: *got* does not collocate with the linguistics features just listed, unless it is the past participle of dynamic *get*.

RANK	GET		(HAVE) GOT	
1	150	job	59	job
2	140	chance	41	chance
3	36	place	27	problem
4	23	life	25	people
5	22	score	23	choice
6	22	copy	16	plan
7	21	vote	15	situation
8	20	bill	15	proposal
9	19	break	14	gun
10	18	lawyer	13	group
11	17	picture	12	playground
12	17	peace	12	petition
13	17	gun	12	guy
14	16	letter	11	report
15	16	list	11	place
16	15	loan	10	program
17	14	message	9	mother
18	13	response	9	house
19	13	resolution	9	boat
20	13	license	8	question

Table 3. USBoE right collocates of *get* and *got* in the SVO-unit

Another particularly relevant aspect linked to this binary notion that distinguishes current possession from *succeeding in attaining* or *coming to possess* is also mirrored in the opposition that distinguishes stative *have got* from the perfective/resultative *have gotten* shown in example 29 (cf. also LeSour 1976, Trudgill and Hannah 1985, Leech 1989, Quirk *et al.* 1991, Gramley and Pätzold 1992, Wilson 1993, Biber *et al.* 1999, Tottie 2002). This aspect is rather compelling for it strongly suggests that stative *have got* may not be so closely related to the result marker *get*, as *have gotten* may, instead, be.

29. So well, when they beat the Oilers, he presented a game ball to SMU. So I--I think it's the first time that I can recall where a college team has **gotten** a game ball for a pro team's victory [USBoE]

In terms of comparison between the two conversational domains investigated here, it emerges that the SVO *get-units* mostly occur within neutral contexts in both corpora (i.e. 58% in the USBoE and 83% in the AMC6), although the SVO *get-units* of the USBoE are much more negative than those present in the AMC6 (i.e. 42% and 17% respectively). It is, in fact, the high occurrence with *job* (cf. Table 3 above) which makes the semantic prosody of the SVO *get-units* of the USBoE more negative than those present in the AMC6: *get* mostly collocates with *job* in the USBoE, which is the most frequent collocate which occurs within a negative semantic prosody (70% of the time).

When *get* collocates with two object noun phrases in SVOO types, the *get-unit* still expresses *result* and *acquisition*, as in SVO types. However, unlike SVO types, it is not the subject which acquires something, but it is the subject that makes the indirect object (e.g. *your son*, *him*, and *myself* in examples 30 and 31) acquire a direct object (e.g. *an interview*, *a job*, and *another B plus* in examples 30 and 31). *Get* is thus not only resultative, but also causative by causing somebody to acquire something. More precisely, it is the presence of the object that makes *get* become causative; the *get-unit* semantics and pragmatics can, then, be said to be determined by three items: the *result marker get* which conveys *result*; the *indirect object* of the clause which makes *get* function as a *causative verb*; and the *direct object* of the clause which is the object of *acquisition*. Another constant item (54% in the USBoE and 82% in the AMC6) present in both corpora is the idea of effort, fatigue, and/or danger (cf. *very difficult* in example 30 and *If I don't buckle down* in example 31) of the situation linked to this type of *get-unit*.

30. So, to **get** your son an interview or **get** him a job is very difficult.
[USBoE]
31. If I don't buckle down, I'm gonna **get** myself another B plus.
[AMC6]

In *complex transitive verb complementation*, the *get-unit* is made up of *get* plus an object and a complement clause in SVOC types, or *get* plus an object and an adverbial in SVOA types. As concerns semantics, the *core meaning* of the unit is ascribed here to *get*, which functions as a *result marker* either by causing someone to enter a new state or by causing someone to be in a certain place; the *peripheral meaning*, instead, is linked to the object plus the complement clause or the adverbial and expresses either *causation* and *transformation* (both in SVOC and SVOA types) or *causation* and *movement* (in SVOA types). In other words, the semantics of the SVOC and SVOA *get-unit* is determined, like in SVOO types, by three units: the *result marker* *get* which conveys *result*; the *object* of the clause which makes *get* function as a *causative verb*; and the *complement* or the *adjunct* of the clause which determine *transformation* or *movement*.

As the numbers in Table 4 indicate, there are very few occurrences of the SVOC *get-unit* in the USBoE sample investigated. This, together with their absence in the movie corpus, suggests that the SVOC *get-unit* might be rare; however, further research on other samples is undoubtedly required. As regards semantics, the few occurrences found show that, when this structure occurs, it usually does in a negative situation (69%), like the inconvenient length of time (i.e. *60 to 90 days*) mentioned in example 32:

32. ...it would take us probably 60 to 90 days to **get** them ready. [USBoE]

Prosody	Get R2 Collocates	Occurrences
POSITIVE	ready	5
NEGATIVE	ready	6
POSITIVE	interested	2
NEGATIVE	interested	7
POSITIVE	free	2
NEGATIVE	free	1
POSITIVE	organized	1
POSITIVE	accurate	1
NEGATIVE	drunk	2
NEGATIVE	lost	1
NEGATIVE	angry	1
NEGATIVE	mad	1
NEGATIVE	pregnant	1
NEGATIVE	confused	1
TOT POSITIVE		10
TOT NEGATIVE		22
TOT		32

Table 4. USBoE right 2 collocates⁹ of *get* in the SVOC-unit

As for the SVOA *get-unit*, the 2 occurrences found in the AMC6 demonstrate it is very rare in movies; on the other hand, it is not rare in the spontaneous conversation sample analyzed (i.e. 137 occurrences). In both corpora, it is resultative and causative in that the subject causes somebody/something to change place; and, like the SVOC *get-unit*, it generally occurs in situations which are negative, namely, where the effort, fatigue, and/or danger to make somebody/something reach a certain place is highlighted (59.12% vs. 100%, i.e. 81/137 vs. 2/2 in the USBoE sample and AMC6 respectively). This occurrence within a negative context is illustrated by the speaker's doubt in example 33 and the lack of time in example 34:

33. Do you think it's possible to get him to the table at all? [USBoE]
34. We've got nineteen hours and fifty eight minutes. I'll get Belairiform into your system before them. Just stay alive. I'm not going to lose you. [AMC6]

Finally, with regard to intransitive complementation, i.e. SV types, Quirk *et al.* (1991:722) exclude the use of *get* from this structure, whereas Biber *et al.* (1999:391) attest it in sentences like: *Don't ask, you don't get*. Neither the USBoE nor the movie data seem to offer such examples: the only constructions found in the sample data are SV clause types in which *get* is usually employed as a phrasal or a prepositional verb of movement. Although the present paper is not concerned with *get* either as a phrasal or a prepositional verb, it is worth noting that following the view introduced here, in such constructions the *core meaning* of the *get-unit* would be considered to be determined by the phrasal or prepositional verb *get*, whereas the *peripheral meaning* by its preposition.

3.2 GET AS AN AUXILIARY-LIKE VERB

As the examples in Section 2 have illustrated, *get* shares all the characteristics of catenative verbs except the occurrence with the *bare infinitive*; and this lack does not influence its versatile syntactic nature. Despite this flexibility in syntax and the fact that catenative *get-units* display a variation also in semantics since their *peripheral meaning* depends both on the kind of the verb that follows *get* and on the presence or absence of an intervening NP (i.e. when an NP intervenes, catenative *get* expresses causation similarly to SVOC and SVOA clause types), there is still similarity in semantics and pragmatics. Indeed, the traits that constantly mark its uses as a full verb (i.e. the resultative nature, the negative semantic prosody, the correspondence of meaning between the structures with and without an intervening NP, and the causative mark the former acquire) are also present in almost all its auxiliary-like uses.

More specifically, the *get + past participle* (together with its corresponding *get + o + past participle unit*) and the *get + infinitive* (together with its corresponding

get + o + past infinitive unit) are the only constructions which do imply a change of state and can occur in negative contexts, but preferably occur in neutral ones. As both the USBoE and movie data indicate, the *get + past participle unit* introduces a type of transformation which is generally not negative (66% and 62% of times respectively, cf. also Table 5), as in *get dressed* in example 36; even though it can also be negative (34% and 38% of times respectively), as in *get hit* in example 35:

- 35. When we got on the Internet, on CNN, we saw the second tower get hit. [USBoE]
- 36. Pam: Go take a shower, get dressed and come back down. [AMC6]

USBoE Corpus		
PAST PARTICIPLE	NEGATIVE	NON-NEGATIVE
Ending in -ed	639	1341
Ending in -en	59	30
TOTAL	698	1371
AMC6 Corpus		
PAST PARTICIPLE	NEGATIVE	NON-NEGATIVE
All types	5	8

Table 5. Types of past participles belonging to the *get + past participle unit* and their semantic prosody checked in the USBoE and in the AMC6

In much the same way, its corresponding causative *get + O + past participle unit*, which causes somebody or something to enter a new state, preferably occurs in neutral situations, either in the USBoE (78% of cases) or in the AMC6 (57% of cases), as illustrated in example 37; but it can also occur in negative ones (22% and 43% respectively), as shown in example 38:

- 37. She hopes to contact more authors in the future, and get them involved in the project as well. [USBoE]
- 38. You keep fucking around, you gonna get that scholarship to Yale taken away from you [AMC6]

With respect to the *get + present participle unit*, it is worth underlining that despite its low frequency in movies (only 3 occurrences), it still resembles the structures

found in the USBoE, for two reasons. First, in terms of semantic function, this unit usually implies, in both corpora, entering a new state, especially concerning movement and locating within difficult situations (i.e. 78% in the USBoE and 67% in the AMC6), as *going to run out of second chances* in example 39 and *trying to catch her on the way home* in example 40:

- 39. Well, if the Braves don't **get** moving, they're going to run out of second chances. [USBoE]
- 40. She's hostessing until 11. Actually oh I better **get** going I'm gonna try to catch her on the way home see if she wants to get a drink or something. [AMC6]

Second, the occurrences present in the AMC6 are all *get + going*, which, as Table 6 illustrates, is the most frequent *get + present participle* unit in the USBoE.

GET + PRESENT PARTICIPLE	
going	38
moving	16
cracking	7
training	5
working	3
rolling	3
parking	3
talking	1
spending	1
shaking	1
running	1
overflowing	1
happening	1
flying	1
flooding	1
TOTAL	83

Table 6. Most common present participles belonging to the *get-unit* in the USBoE corpus

A similar case concerns the causative *get + O + present participle* unit: first, it still implies entering a new state, especially concerning movement and locating within difficult situations (i.e. 94% in the USBoE and 100% in the AMC6); example 41, for instance, is explicitly negative, i.e. *it's pretty tough*, whereas example 42 is negative, if perceived in the movie context, i.e. the situation is tough because there is no time.

41. That's pretty tough when they're bored, they're tired, they don't like the book they read to begin with, and you want to get a discussion going with 20 15-year-olds. [USBoE]
42. Let's get this going¹⁰. [AMC6]

Second, there is only one occurrence of the *get + O + present participle* in the movie corpus and the verb which is used is still *going*, like in the *get + present participle* unit. As illustrated in Table 7, this is still the most frequent present participle of *get + O* in the USBoE sample.

GET + O + PRESENT PARTICIPLE	
going	35
moving	22
working	13
thinking	4
talking	5
rolling	3
flying	1
swimming	1
yelling	1
walking	1
TOT	86

Table 7. Most common present participles belonging to the *get + O* unit in the USBoE corpus

The *get + infinitive unit* shows slightly different patterns: in the first place, it usually implies a change, either of state or of place, which tends to occur in neutral situations both in the USBoE (80%) and in the AMC6 (67%), meaning *manage to*; even though it can also occur in negative ones such as those illustrated by examples 43 and 44 (i.e. *It'll take a while* and the *never did* respectively):

43. It'll take a while for people to **get** to know it. [USBoE]
 44. Hey you know, we never did **get** to finish that little "convo" back at the den. [AMC6]

Secondly, although it is rather rare in movies (only 6 occurrences), similarly to the preceding construction, the most frequent verb is *know* (2 occurrences), which is also the most frequent following *get* found in the USBoE sample (cf. Table 8).

GET TO RIGHT COLLOCATES	
know	99
see	79
work	68
play	65
be	49
go	27
do	18
decide	8
TOT	413

Table 8. Get to right collocates (i.e. most common infinitives) belonging to the get-unit in the USBoE corpus

Finally, its corresponding causative *get + O + infinitive unit* still conveys a kind of change of state or of place, but this time it underlines the effort/difficulty of the situation in both corpora (i.e. 68% in the USBoE and 67% in the AMC6), like in:

45. Foreign Minister Levy was unable to **get** Secretary Baker to approve the loan guarantees yesterday. [USBoE]
 46. I don't think I can **get** her to do it. [AMC6]

4. CONCLUSIONS

Both the USBoE and movie data have shown that the *get-sentence* is extremely variable in syntax, but not in semantics and pragmatics. In particular, syntactic flexibility has been demonstrated by the multiplicity of uses of *get* both as a full and as an auxiliary-like verb. Indeed, as a full verb, *get* locates in all the clause types present in the English language; and, as an auxiliary-like verb, it appears in three out of four of both simple and complex catenative constructions, and in five kinds of passive sentences. Semantic and pragmatic similarity, instead, has been demonstrated, first of all, by the resultative feature the *get-sentences* constantly display. Specifically, *get* as a full verb may exhibit a resultative change of state in SVC, SVOC, and SVOA clauses; a resultative change of location in SVA, SVOA, and SV clauses; resultative achievement in SVO, SVOO, and SVOA clauses; and resultative causation in SVOO, SVOC, and SVOA clauses. Whereas *get* as an auxiliary-like verb may be classified into various semantic domains, all resultative in nature, depending on the meaning of the lexical verb which follows *get*. Second, semantic and pragmatic similarity has emerged from the constant negative/difficult situation often highlighted by the effort to reach the result brought about by *get*; in particular 7 *get-units* out of 11 have shown a negative semantic prosody). Finally, such similarity has been demonstrated by both the causative mark *get* acquires every time it is followed by an NP, and by the semantic and pragmatic features it displays under this circumstance which are almost always identical to those expressed by its corresponding sentence without the NP (the only exceptions to this are the SVO and *get + infinitive units* and their corresponding NP-clauses).

The constant resultative quality of the *get-sentence*, together with the variable syntactic context in which *get* can locate, have led me to hypothesize an explanation for this semantic and pragmatic similarity by considering two interrelating and complementary levels of meaning that may play a fundamental role in the semantics and pragmatics of the *get-sentence*. These levels of meaning are the *core* and the *peripheral meaning* of the *get-unit*. In particular, the idea of the *core meaning* has suggested that, due to the presence of the *result marker get*, any potential meaning of the *get-sentence* must necessarily be characterized by the feature *result* it constantly conveys. On the other hand, the idea of the *peripheral meaning* has justified the different shades of meaning determined by the variable environment in which *get* occurs (e.g. movement in SV, SVS, and SVOA types; transformation in SVC and SVOC types; and achievement in SCV and SVOO types).

Regarding the comparison between spontaneous and movie conversation, the present data have empirically demonstrated that the two conversational domains do not differ much (cf. Table 9). Indeed, despite the discrepancy of the uses of *get* in the *SVOC-unit*, despite the difference in semantic prosody percentages, and despite the fact that movie language is a type of conversation which is not spontaneous (cf. Gregory 1967, Nencioni 1976, Taylor 1999, Rossi 2003, Pavese 2005) in that it is both written-to-be-spoken and mediated, as far as the general syntactic, semantic and pragmatic uses of *get* are concerned, movie language has turned out to exhibit similar characteristics to spontaneous conversation.

Table 9. Summing up table of the *get*-units found in the USBoE and AMC6 corpora (*get*-units within a negative context in bold) and the percentage of their negative semantic prosody.

USBoE	AMC6	vs.	USBoE	AMC6
GET as a FUL VERB			Negative Prosody %	
(S)VC	(S)VC	similar	67%	81%
(S)VOC	—	<i>different</i>	69%	—
(S)VA	(S)VA	similar	73%	54%
(S)VOA	(S)VOA	similar	59%	100%
(S)VO	(S)VO	similar	42%	17%
(S)VOO	(S)VOO	similar	54%	82%
GET as an AUXILIARY-LIKE VERB			Negative Prosody %	
GET + V-ED	GET + V-ED	similar	34%	38%
GET + O + V-ED	GET + O + V-ED	similar	22%	43%
GET + V-ING	GET + V-ING	similar	77%	67%
GET + O + V-ING	GET + O + V-ING	similar	94%	100%
GET + TO V	GET + TO V	similar	20%	33%
GET + O + TO V	GET + O + TO V	similar	68%	67%

More specifically, this similarity has emerged, first and foremost, from the fact that in both the domains under investigation, *get* is extremely versatile by occurring in various syntactic structures both as a full and as an auxiliary-like verb; as already mentioned, the only *get*-unit which has not been attested in movie conversation is the SVOC one. Moreover, in both spontaneous and movie language, *get* has been shown to function as a *result marker* by implying some change or transformation. This change or transformation has been found to take place mostly in negative contexts in almost all the *get*-units analyzed; besides, the units which have not turned out to occur in negative contexts (i.e. the SVO-unit, the *get* + *past participle* unit with its corresponding *get* + O + *past participle* unit, and the *get* + *infinitive* unit with its corresponding *get* + O + *infinitive* unit) have appeared to be positive both in spontaneous and movie conversation. Finally, the occurrence of the SVOO and *get* + *infinitive* units within negative contexts (which was not expected since all the other O-structures reflect the prosody of the corresponding structure without an O, or NP) has proved to be the same in both spontaneous and movie conversation.

Since the characteristics concerning the *get*-unit have turned out to be similar in both the conversational domains explored, it can be concluded not only that the two registers are similar with regard to this unit, but also that the features emerged may be typical of the *get*-unit, regardless of the register in which it may appear. In other words, the present research confirms both Lindstromberg's (1991:285) claim that *get* should not be seen as polysemic, but as having different

shades of meaning semantically related to each other, and also lends strength to Johansson and Oksefjell's (1996:73) intuition regarding the existence of "a prototype to which all constructions conform more or less closely". Furthermore, it has been proved that *get* must enjoy a very special status, which derives not only from its syntactic versatility or from the semantic and pragmatic similarity shared by the sentences in which it appears, but also from the fact that such peculiarities of *get* coexist.

1 The AMC6 corpus (where 6 stands for the number of the movies taken into account) is part of the **American Movie Corpus**, namely, a *sample parallel bilingual* database under development for the study of movies (cf. Forchini *forthcoming*) as a form of mediated language both because movie language is constructed to appear spontaneous (Pavesi 2005) and because it is dubbed (i.e. translated) and by extension mediated (Ulrych and Anselmi 2008). For the present research the following 6 movies have been taken into account: *Mission:I-2* (John Woo 2000); *Me, Myself & Irene* (Bobby & Peter Farrelly 2000); *Meet the Parents* (Jay Roach 2000); *Shallow Hal* (Bobby & Peter Farrelly 2001); *Ocean's Eleven* (Steven Soderbergh 2001); and *The Matrix Reloaded* (Andy & Larry Wachowsky 2003).

2 Quirk *et al.* (1991) reserve this term for the function of a part of a phrase or a clause which follows a word and completes the specification of a meaning relationship which that word implies.

3 Quirk *et al.* (1991:722) consider seven major clause types (i.e. SV, SVO, SVC, SVA, SVOO, SVOC, SVOA) according to the “permissible combinations” of clause elements. They exclude the use of *get* from type SV, whereas Biber *et al.* (1999) include it (cf. *Don't ask, you don't get* in Biber *et al.* 1999:391).

4 S stands for *subject*, V for *verb*, O for *object*, C for *complement*, and A for *adverbial*.

5 As explained in note 3, there are doubts about *get* belonging to this structure. Both the USBoE sample and the movie data seem to offer only constructions in which *get* is usually employed as a phrasal or a prepositional verb of movement (cf. Section 3 for details).

6 What is traditionally called a *catenative* verb is classified according to the construction it occurs with, namely, the kind of non-finite

complement the *catenative* is followed by (i.e. the *bare infinitive*, the *to-infinitive*, the *-ing form*, or the *-en form*, cf. Palmer 1988 and Huddleston and Pullum 2002) and the possibility of having an intervening NP which functions as complement in the clause. Specifically, with intervening NP, the *catenative* construction is said to be *complex*; otherwise, it is said to be *simple* (cf. Huddleston and Pullum 2002).

7 Quirk *et al.* (1991) call it *central* or *pure passive*.

8 Palmer (1988) and Quirk *et al.* (1991) call them *pseudo passives*, since it is only their superficial form that recommends them for consideration as passives. Similarly, Huddleston and Pullum (2002: 1443) point out that “it is only the embedded complement that is passive, with the matrix *get* clause being active”.

9 Since *pretty* is a right collocate of *get*, it has been included in the table. However, it has not been counted because it is used as a pre-modifying adverb in the sample corpus.

10 I.e. those which occur in second position, namely, after the object.

11 This example is the same as example 11 (b) for it is the only occurrence of the *get* + O + *present participle* found in the AMC6 corpus.

- Biber, D., et al. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Cattel, R. (ed.). 1984. *Syntax and Semantics*, Vol. 17: *Composite Predicates in English*. London, New York: Academic Press.
- Collins, P. C. 1996. "Get-passives in English". *World Englishes* 15(1): 43-56.
- Firth, J.R. 1957a. "A synopsis of linguistic theory, 1930-1955". In J.R. Firth et al. *Studies in Linguistic Analysis*. Special volume of the Philological Society. Oxford: Blackwell.
- Firth, J.R. 1957b. *Papers in Linguistics 1934-1951*. London: Oxford University Press.
- Forchini, P. (forthcoming 2009), "Well, uh no. I mean, you know". *Discourse Markers in Movie Conversation*. In *Lodz Studies in Language*. Bern: Peter Lang.
- Gilquin, G. 2003, "Causative *get* and *have*: So close, so different", *Journal of English Linguistics*, vol 31, n. 2.
- Gee, J.P. 1974. *Get passive: On some Constructions with Get*. Bloomington: Indiana University Linguistics Club.
- Gramley, S. and Pätzold, K. M. 1992. *A Survey of Modern English*. London, New York: Routledge.
- Gregory, M. 1967. "Aspects of Varieties Differentiation". *Journal of Linguistics*, 3:177-198.
- Haegeman, L. 1985. "The Get-Passive and Burzio's Generalization". *Lingua* 66:53-77.
- Halliday, M.A.K. 2003. *On Language and Linguistics*. London/New York: Continuum.
- Halliday, M.A.K. and Matthiessen C. 2004. *An Introduction to Functional Grammar*. London: Arnold.
- Hatcher, A. 1949. "To be/get invited". *Modern Language Notes* 64 (7), 433-434.
- Huddleston, R. and Pullum, G. 2002. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Hunston, S. 2002. *Corpora in Applied Linguistics*. Cambridge: Cambridge University Press.
- Johansson, S. and Oksefjell, S. 1996. "Towards a Unified Account of the Syntax and Semantics of *Get*". In J. Thomas and M. Short (eds.). 57-75. *Using Corpora for Language Research*. London: Longman.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London, New York: Longman.
- Kimball, J. 1973. "Get". In: Kimball, John. (ed.), *Syntax and Semantics*. Vol. 2, 205-215. London, New York: Seminar Press.
- Leech, G. 1989. *An A-Z of English Grammar & Usage*. London, Melbourne, Auckland: Edward Arnold.
- LeSourd, P. 1976. "Got Insertion". *Linguistic Inquiry* 7(3): 509-516.
- Lindstromberg, S. 1991. "Get: Not Many Meanings". *IRAL* XXIX/4: 285-301.
- Louw, B. 1993. "Irony in the Text or Insincerity in the Writer? The Diagnostic Potential of Semantic Prosodies". In M. Baker, G. Francis and E. Tognini-Borelli (eds.), 157-176.
- Nencioni, G. 1976. *Di scritto e di parlato*. *Discorsi linguistici*. Bologna: Zanichelli.
- Palmer, F. 1988. (2nd ed.). *The English Verb*. London: Longman.
- Pavesi, M. 2005. *La Traduzione Filmica. Aspetti del Parlato Doppiato dall'Inglese all'Italiano*. Roma: Carocci.
- Pizzini, Q. A. 1975. "Have and Get in English Syntax". *Papers in Linguistics* 3/4:233-254.
- Rossi, A. 2003. *La lingua del cinema*. In I. Bonomi, A. Masini, S. Morgana (eds.) *La lingua Italiana e i Mass Media*. Roma: Carocci Editore.
- Quirk, R. et al. 1991. *A Comprehensive Grammar of the English Language*. London: Longman.

- Rühlemann, C. 2007. "Lexical Grammar: The get-passive as a case in point". *ICAME Journal* 31:111-127.
- Shibatani, M. 1976. "The Grammar of Causative Constructions: A *Conspectus*". In M. Shibatani (ed.), 1-39. *Syntax and Semantics*, Vol. 6 (*The Grammar of Causative Constructions*). London: Academic Press.
- Sinclair, J. M. 1991. *Corpus Concordance Collocation*. Oxford: Oxford University Press.
- Sinclair, J. M. 1996. "The search for units of meaning". *TEXTUS*. 9,1. Special Volume L. Merlini and J. M. Sinclair (eds.) *Lessico e Morfologia*, 75-106; reprinted in G. Corpas (ed.) *Las Lenguas de Europa: Estudios de Fraseología, Fraseografía y Traducción*; Granada: Editorial Comares, SL, 7-38.
- Sinclair, J. M. 1998. "The lexical item". In E. Weigand (ed.) *Contrastive Lexical Semantics* Amsterdam/Philadelphia: John Benjamins. 1-24.
- Sinclair, J. M. 1999. "A way with common words". In H. Hasselgård and S. Oksefjell (eds.) *Out of corpora: studies in honour of Stig Johansson*. Amsterdam: Rodopi. 157-179.
- Sinclair, J. M. 2003. "Lexical grammar". In B. Hary (ed.) *Corpus Linguistics and Modern Hebrew*. Israel: Chaim Rosenberg School of Jewish Studies.
- Sinclair, J. M. 2004. *Trust the Text: Language, Corpus and Discourse*. New York: Routledge.
- Standwell, G. J. B. 1981. "The Get-Passive and the Be-Passive". *British Journal of Language Teaching* 3:163-164.
- Stubbs, M. 2001. *Words and Phrases: Corpus Studies in Lexical Semantics*. Oxford/Massachusetts: Blackwell.
- Taylor, C. 1999. "Look who's talking. An analysis of film dialogue as a variety of spoken discourse." In L. Lombardo, L. Haarman, J. Morley, C. Taylor, (eds.). 1999. *Massed Medias. Linguistic Tools for Interpreting Media Discourse*. Milano: LED.
- Tobin, Y. 1993. *Aspect in the English Verb: Process and Result in Language*. London: Longman.
- Tognini-Bonelli, E. 2001. *Corpus Linguistics at Work*. Amsterdam, Philadelphia: John Benjamins.
- Tottie, G. 2002. *An Introduction to American English*. Malden: Blackwell.
- Trudgill, P. and Hannah, J. 1985. *International English*. London: Edward Arnold.
- Ulrych, M. and Anselmi, S. 2008. "Towards a corpus-based distinction between language-specific and universal features of mediated discourse". In A. Martelli, and V. Pulcini (eds.) *Investigating English with Corpora. Studies in honor of Maria Teresa Prat*. Monza: Polimetrica. 257-273.
- Wilson, K. 1993. *The Columbia Guide to Standard American English*. New York: Columbia University Press.

Finito di stampare nel mese di gennaio 2008
presso il Servizio Grafica e stampa
dell'Ufficio di Staff Comunicazione e Organizzazione eventi
dell'Università degli Studi di Trieste