

Like as a discourse marker in different varieties of English

A contrastive corpus-based study

Mémoire réalisé par
Marie Gabrys

Promotrice
Prof. Dr. Gaëtanelle Gilquin

Année académique 2016-2017
Master 120 en linguistique, finalité spécialisée LAMUSO

Acknowledgements

My heartfelt thanks go to all the people who helped me with the completion of this dissertation.

First and foremost, I would like to thank my supervisor, Gaëtanelle Gilquin, for her time, guidance and useful feedback from one end of this project to the other.

I also wish to express my gratitude to my parents for always believing in me – sometimes more than I did myself – and for their invaluable support in the course of my studies, each in their own way; and to Mathieu for being endlessly patient, encouraging and reassuring.

Finally, I would like to thank Besa, Catherine, Charlotte, Justine, Justine and Pauline for making those university years so lively, joyful and enriching thanks to their friendship.

Contents

Acknowledgements	i
List of tables	vi
List of figures	vii
List of abbreviations	viii
Introduction	1
1. Discourse markers: theoretical background	3
1.1 Introduction	3
1.2 Terminology and definition	4
1.3 Conceptual frameworks	5
1.3.1 Coherence-based approach	6
1.3.2 Grammatical-pragmatic perspective	6
1.3.3 Relevance theory	7
1.3.4 Grammaticalization theory	7
1.4 Characteristics	8
1.4.1 Multigrammaticality	8
1.4.2 Orality	9
1.4.3 Initiality and prosody	10
1.4.4 Weak clause association and optionality	10
1.4.5 Non-truth-conditionality	11
1.5 Functions in spoken interaction	11
1.5.1 Connectivity	12
1.5.2 Co-construction of meaning	13
1.5.3 Feedback and attitude	14
1.5.4 Social relationships	14
2. <i>Like</i> as a discourse marker	17
2.1 Introduction	17
2.2 Evolution of <i>like</i>	18
2.2.1 History of <i>like</i>	18
2.2.2 Grammaticalization of <i>like</i>	19
2.2.3 Quotative <i>like</i>	20
2.3 Position, scope and prosody	21
2.3.1 Position	21
2.3.2 Scope	22

2.3.3 Prosody	22
2.4 Functions in spoken interaction	22
2.4.1 Marking lexical focus	23
2.4.2 Searching for the appropriate expression	23
2.4.3 Marking an approximate number or quantity	24
2.4.4 Introducing an example	25
2.4.5 Introducing an explanation	25
2.4.6 Discourse link and restart	26
2.5 Sociolinguistic determinants of the use of <i>like</i>	26
2.5.1 Gender	26
2.5.2 Age	27
3. Data and methodology	31
3.1 The corpora	31
3.1.1 International Corpus of English	32
3.1.2 Santa Barbara Corpus of Spoken American English	33
3.2 The software	34
3.3 Disambiguation of the instances of <i>like</i>	34
3.4 Quantitative data analysis	37
3.5 Qualitative data analysis	37
3.5.1 Position	38
3.5.2 Function	39
3.6 Sociolinguistic analysis	41
3.6.1 Gender	42
3.6.2 Age	42
4. Frequency of <i>like</i>	43
4.1 Overall frequency of <i>like</i>	44
4.2 Proportion of discourse marker uses	45
4.3 Discussion	46
5. Position of <i>like</i>	49
5.1 American English	49
5.2 British English	51
5.3 Canadian English	54
5.4 Irish English	55
5.5 New Zealand English	58
5.6 Discussion	59

6. Function of <i>like</i>	61
6.1 American English	61
6.2 British English	63
6.3 Canadian English	66
6.4 Irish English.....	68
6.5 New Zealand English.....	70
6.6 Discussion.....	72
7. Sociolinguistic determinants of <i>like</i>	75
7.1 Gender	75
7.2 Age.....	76
7.2.1 Canadian English	76
7.2.2 American English.....	77
7.2.3 Comparison of the two samples	78
7.3 Individual speakers' use of <i>like</i>	80
7.3.1 Canadian speakers.....	80
7.3.2 American speakers	82
7.4 Discussion.....	84
Conclusion.....	87
References	91
Appendices	97

List of tables

Table 3.1 Structure of the ‘spoken dialogue’ section of ICE.....	32
Table 3.2 Number of conversations and total number of words in the five subcorpora.....	33
Table 3.3 Number of instances of <i>like</i> before and after disambiguation.....	36
Table 3.4 Functions of the discourse marker <i>like</i> (based on Müller 2005).....	40
Table 4.1 Raw and relative frequencies of <i>like</i> as a discourse marker in the five subcorpora.....	44

List of figures

Figure 1.1 Items identified as discourse markers according to various authors.....	5
Figure 4.1 Relative frequency of the discourse marker <i>like</i> in the five subcorpora.....	45
Figure 4.2 Proportion of the discourse marker uses of <i>like</i> in the five subcorpora.....	46
Figure 5.1 Distribution of the different positions of the discourse marker <i>like</i> in the SBC.....	50
Figure 5.2 Distribution of the different positions of the discourse marker <i>like</i> in ICE-GB.....	51
Figure 5.3 Distribution of the different positions of the discourse marker <i>like</i> in ICE-CA.....	54
Figure 5.4 Distribution of the different positions of the discourse marker <i>like</i> in ICE-IR.....	56
Figure 5.5 Distribution of the different positions of the discourse marker <i>like</i> in ICE-NZ.....	58
Figure 6.1 Distribution of the different functions of the discourse marker <i>like</i> in the SBC.....	62
Figure 6.2 Distribution of the different functions of the discourse marker <i>like</i> in ICE-GB.....	64
Figure 6.3 Distribution of the different functions of the discourse marker <i>like</i> in ICE-CA.....	66
Figure 6.4 Distribution of the different functions of the discourse marker <i>like</i> in ICE-IR.....	68
Figure 6.5 Distribution of the different functions of the discourse marker <i>like</i> in ICE-NZ.....	70
Figure 7.1 Distribution of the discourse marker <i>like</i> according to gender in ICE-CA and the SBC.....	76
Figure 7.2 Distribution of the discourse marker <i>like</i> according to age in ICE-CA.....	77
Figure 7.3 Distribution of the discourse marker <i>like</i> according to age in the SBC.....	78
Figure 7.4 Distribution of the discourse marker <i>like</i> according to age in ICE-CA and the SBC.....	79
Figure 7.5 Distribution of the discourse marker <i>like</i> according to gender and age in ICE-CA.....	80
Figure 7.6 Distribution of the discourse marker <i>like</i> according to gender and age in the SBC.....	82

List of abbreviations

Corpora

ICE	International Corpus of English
ICE-CA	Canadian subcorpus of ICE
ICE-GB	British subcorpus of ICE
ICE-IR	Irish subcorpus of ICE
ICE-NZ	New Zealand subcorpus of ICE
SBC	Santa Barbara Corpus of Spoken American English

Other abbreviations

DM	Discourse marker
NP	Noun phrase
VP	Verb phrase

Introduction

Oral communication, more than simply the spoken counterpart of writing, is characterized by a number of distinctive features. Therefore, the spoken medium and its specificities deserve attention as an object of interest in its own right. The immediateness of the spoken exchanges as well as the spatial and/or temporal co-presence of the participants make for a particular type of linguistic interaction, governed by its own rules. In this respect, various linguistic strategies are at play in spoken discourse to make up for the lack of planning inherent in spontaneous speech and to ensure the fluency of the exchange. Among these strategies, discourse markers can be argued to significantly contribute to the smooth progress of the oral interaction thanks to the numerous pragmatic functions they fulfill.

Discourse markers, both in English and in other languages, have received growing academic attention since the late 1980s, and the essential role that they play in spoken communication has been increasingly acknowledged. In her seminal work, Schifffrin (1987) has laid the foundations of a theoretical framework in which English discourse markers can be defined and examined. Other scholars (cf. Fraser 1999; Schourup 1999) have followed in her footsteps and discourse markers have been studied from a number of perspectives over the last few decades. However, this multiplicity of – sometimes diverging – approaches has contributed to the difficulty of providing a single, precise and all-encompassing definition of the linguistic class formed by discourse markers.

The present dissertation sets out to explore one discourse marker in particular, namely *like*. Increasingly prominent in today's spoken practices, *like* is one of the most multifunctional discourse markers in English. In the literature, various aspects of this discourse marker have been investigated. Diachronic studies, such as D'Arcy's (2005), have looked into the grammaticalization process that has led to the current and highly frequent use

of *like* as a discourse marker. Other linguists (cf. Romaine & Lange 1991; Andersen 2001) have dealt with some formal characteristics of the discourse marker such as position and scope. In addition, a number of studies have examined the sociolinguistic component of the discourse marker *like*, trying to identify the factors that condition its use among an English-speaking community (cf. Fleischman 1998; Siegel 2002).

Most studies that have looked into the use of *like* as a discourse marker have done so by focusing on one single variety of English. (e.g. teenage Canadian English (Tagliamonte 2005); American English (Croucher 2004)). Therefore, the primary aim of this dissertation is to address the issue of the use of *like* from a new, contrastive perspective. To do so, five varieties of native English will be compared, namely American, British, Canadian, Irish and New Zealand English. A corpus-based study will be carried out in order to identify the possible differences and similarities in the use of the discourse marker *like* across the five linguistic communities cited above. Given the different stages that these varieties have reached in the development of *like* as a discourse marker, significant differences – at least in terms of frequency of use – can be expected across the five subcorpora.

The present dissertation is structured as follows. The first two chapters lay the theoretical background of this study, with an introduction to discourse markers in general in Chapter 1 and a thorough description of the discourse marker *like* in Chapter 2. Chapter 3 then presents the data used and the methodology adopted in this study. The following four chapters give an account of the empirical part of this dissertation. Chapter 4 looks into the frequency of use of *like* as a discourse marker across the five subcorpora under scrutiny. It is followed by a contrastive analysis of the position of *like* in Chapter 5, and of the functions that the discourse marker fulfills in Chapter 6. Chapter 7 deals with two sociolinguistic factors that are said to influence the use of *like* as a discourse marker, namely gender and age. Finally, a conclusion chapter reflects on the study carried out, its contributions and limitations, and proposes avenues for further research.

Chapter 1

Discourse markers: theoretical background

1.1 Introduction

Before analyzing the discourse marker *like* more specifically, which is the main purpose of this dissertation, specifying what is meant by the term ‘discourse marker’ in the first place seems a necessary step. This task is not an easy one, however, as discourse markers have been defined in a multitude of ways and considered from many different perspectives. At the crossroads of syntax, semantics and pragmatics, discourse markers are difficult to classify into one discrete category and their complex nature contributes to the terminological diversity that can be found in the literature. Indeed, more or less close synonyms of the term ‘discourse markers’ are many and various and the definition of one and the same term often varies from one author to the next. In addition to this conceptual vagueness and ambiguity, discourse markers fulfill a whole range of functions in oral communication. As a result, the classification of discourse markers is open to debate and there is a great deal of variation in the sets of items to be included under the term ‘discourse markers’.

This chapter aims at defining and characterizing discourse markers, as a class of linguistic elements without any real semantic content, nor any true syntactic function, which still “in each individual context [give] it a quite distinct and unmistakable colouring” (Carlson 1984, cited in Aijmer 2002: 19). Section 1.2 deals with the terminological issues related to the label ‘discourse marker’. Section 1.3 then introduces the different conceptual frameworks within which discourse markers have been studied. The subsequent section gives a brief overview of the main characteristics of discourse markers, which serve as criteria to count them as members of the class of discourse markers or not (Section 1.4). Finally, the various functions fulfilled by discourse markers in spoken interaction are summarized in Section 1.5.

1.2 Terminology and definition

There is still strong disagreement on the exact terminology to adopt when dealing with so-called ‘discourse markers’. The lexical subset that they constitute has been given many different names in the literature, some of which are used interchangeably, sometimes even within one and the same article. These include, non-exhaustively, the terms ‘conversational particle’ (Schourup 1999), ‘discourse connective’ (Blakemore 1987), ‘discourse operator’ (Redeker 1991), ‘discourse particle’ (Siegel 2002), ‘pragmatic device’ (Stubbe & Holmes 1995), ‘pragmatic formative’ (Fraser 1987), ‘pragmatic marker’ (Fraser 1996) and ‘semantic conjunct’ (Quirk et al. 1985). Note that ‘discourse marker’ remains by far the most popular label in recent articles. As mentioned by Redeker (1991: 1139), however, most works on the topic tend to be characterized by “a lack of clarity and consistency in the definitions and the use of theoretical terms and analytical categories”. Thus, Blakemore, for instance, shifted from ‘discourse connectives’ in her earlier works (e.g. 1987) to ‘discourse markers’ in her later papers (e.g. 2002).

This terminological vagueness contributes to the conceptual ambiguity that characterizes discourse markers. If some authors use several terms as true synonyms, others make a meaning distinction between them, like Carter & McCarthy (cited in Huang 2013), who consider discourse markers as a (lexical) subclass of pragmatic markers, to which stance markers, hedges and interjections also belong. Aijmer (2002) stresses the importance of the notion of ‘pragmatic’ when defining particles/markers since their meaning mostly derives from the very fact that they are used in interaction. According to Zwicky (1985: 302), discourse markers are rather part of “the great collection of things that have been labeled ‘particles’”, to which interjections belong as well. On the other hand, Jucker & Ziv (1998: 2) opted for the term ‘discourse marker’ as a “single conceptual umbrella” under which a whole range of linguistic elements can be encompassed.

Several authors have attempted to define discourse markers as a distinctive linguistic class – assuming that they do form one at all. In her seminal work *Discourse markers*, Schifffrin (1987: 31) defines them as “sequentially dependent elements which bracket units of talk”. Her definition was later criticized and rejected for being too broad (cf. Redeker 1991) and for leaving out the significant role that markers play in organizing discourse at a more global level (cf. Lenk 1998a). This gap was filled in Fung & Carter’s (2007: 411) definition of

discourse markers as “intra-sentential and supra-sentential linguistic units which fulfill a largely non-propositional and connective function at the level of discourse”. Overall, no single definition so far has turned out to be truly satisfying and accepted unanimously, which leads to a great deal of variation in the items that enter the category of ‘discourse markers’ according to different authors. Figure 1.1 (extracted from Taguchi 2002: 45) below illustrates this variation. The absence of a “generally accepted list of discourse markers in English” (Taguchi 2002: 44) strengthens the need for a set of precise criteria which one could resort to to evaluate whether a particular word should be treated as a discourse marker or not. These are further detailed in Section 1.4.

Study	Items Identified as Discourse Markers
Ostman (1982)	<i>well, like, kind of, sort of, you know, I mean, oh, now, but, and, uh</i>
Schourup (1985)	<i>well, like, kind of, you know, I mean, oh, now, ah, mind you, uh</i>
Schiffrin (1987)	<i>well, you know, I mean, and, but, or, so, because, now then, oh</i>
Frasier (1990)	<i>well, you see, now, but, so, ah, all right, anyway, OK, or, then</i>
Redeker (1990)	<i>well, you know, I mean, oh, but, so, ah, all right, OK, because, mind you</i>
Stenstrom (1994, 1998)	<i>well, you know, you see, I mean, oh, now, all right, anyway, yeah, OK, like, really</i>
Biber, et al. (1999)	<i>well, right, now, I mean, you know, you see, look</i>

Figure 1.1 Items identified as discourse markers according to various authors

1.3 Conceptual frameworks

The field of study composed by discourse markers has gained increasing attention for the last few decades. Through time, discourse markers have been addressed from a number of perspectives and studied within various conceptual frameworks. These multiple approaches – be they conflicting or complementary – have been presented in different manners in the literature, sometimes showing some degree of overlap. The main ones are summarized in the following four sections.

1.3.1 Coherence-based approach

Also called ‘integrational approach’, this theory was developed by Schiffrin (1987). The central idea of her work is that discourse markers add some coherence to discourse by “index[ing] an utterance to the local contexts in which utterances are produced and in which they are to be interpreted” (Schiffrin 1987: 326). Discourse markers thus locate an utterance according to the five planes of talk¹ formulated by Schiffrin (1987). They do so at the local level, since Schiffrin’s (1987: 24) framework is restricted to “relations between adjacent units in discourse”, in line with her definition of discourse markers (see Section 1.2 above). This approach to discourse markers is a broad one, including paralinguistic features and non-verbal gestures as potential discourse markers (Fraser 1999).

Discourse coherence as a whole depends on the coherence relations between and within certain linguistic units, as well as on the relations between these units and (para)linguistic aspects of the situation of communication (Risselada & Spooren 1998). Although discourse markers actively contribute to the global dimension of discourse coherence, their more far-reaching scope was left out of account in the present theory (cf. Lenk 1998a). Schiffrin’s model of discourse coherence was later revised by Redeker (1991), who highlighted the need for clearer boundaries for the linguistic class that discourse markers form.

1.3.2 Grammatical-pragmatic perspective

This approach to discourse markers was mostly developed by Fraser (e.g. 1987, 1996, 1999). His theory is based on the belief that discourse markers – ‘pragmatic markers’ as he calls them – “do not contribute to the propositional content of the sentence but signal different types of messages” (Fraser 1999: 936). He thus distinguishes between content meaning and pragmatic meaning, i.e. the speaker’s evaluation, attitudes and beliefs towards the content of the message, which discourse markers can signal (Taguchi 2002). Discourse markers are then classified into different classes depending on the type of pragmatic relationship they signal (Fraser 1999).

The grammatical-pragmatic perspective restricts the scope of the label ‘discourse markers’ to linguistic expressions only, rejecting non-verbal items which were included in the

¹ Schiffrin (1987: 24-29) argues that the coherence of a given utterance emerges from one’s ability to locate that stretch of discourse within a model that connects and integrates different types of units. She lists five of them, namely the exchange structure, action structure, ideational structure, participation framework and information state.

previous approach. This definition further emphasizes the core, procedural meaning of discourse markers, influenced to some extent by the communicative situation in which they occur (Fung & Carter 2007).

1.3.3 Relevance theory

This theoretical framework is based on Grice's cooperative principle, i.e. the fact that, in conversation, "the hearer's role is to infer what the current speaker means on the basis of what is said, and principles governing the interaction" (Aijmer 2002: 8). In other words, in interaction, the participants make use of and rely on a number of conversational implicatures as a common effort to build the meaning of the exchange (Grice 1975).

Working within this framework, Blakemore (e.g. 1987) argues that 'discourse connectives' impose constraints on the context in which an utterance is to be interpreted. They function as 'signposts' by "indicat[ing] the relevance of one discourse segment to another" (Taguchi 2002: 4). They thus function as instructions about inferential connections, and as such, they help process the propositional content (Fung & Carter 2007). This approach is an explanatory one, trying to provide arguments for the use of discourse markers, such as their pragmatic, procedural meaning, or their ability to connect the hearer's background knowledge and the contextual assumptions (Huang 2013).

Still within the relevance-theoretical framework, Andersen (2001: 39) defines 'pragmatic markers' as "a class of short, recurrent linguistic items that generally have little lexical import but serve significant pragmatic functions in conversation". In other words, their input lies beyond the level of the propositional content as they rather facilitate the exchange by informing the hearer about the speaker's attitudes or intentions, for instance (Andersen 2001).

1.3.4 Grammaticalization theory

The last major and most recent conceptual framework within which discourse markers have been studied is grammaticalization theory, broadly illustrated by Aijmer's (2002) work. This approach focuses on the diachronic development of discourse markers, which may acquire a conventionalized core meaning as a result of grammaticalization. Traugott (1995) defines grammaticalization as the "process whereby lexical items or phrases come through frequent use in certain highly constrained local contexts to be reanalysed as having syntactic and morphological functions" (cited in Aijmer 2002: 16). If discourse markers lack propositional

meaning, as has been stated in the previous approaches, they might originally come from some lexical material whose syntactic-semantic status has evolved (Brinton 1996; Andersen 2001).

In this way, the linguistic expression *like* has undergone successive steps of the grammaticalization process, which have given it its current function as a discourse marker. This specific case of the phenomenon is further detailed in Section 2.2.2.

1.4 Characteristics

This section focuses on the general properties of discourse markers, compiled from the work of a range of scholars. These characteristics show some variation from one author to another, resulting in different lists of items included under the label ‘discourse markers’. In this study, I start from the assumption that discourse markers form a class of linguistic elements which must share a number of common features. This section is based on the list of features set up by Schourup (1999). The following properties will later be used in my analyses as a set of criteria to select the instances of *like* which can be said to function as discourse markers. Note that one criterion alone is not sufficient to classify a particular item as a discourse marker; several features need to be combined, and other sociolinguistic aspects such as context of use need to be taken into account.

1.4.1 Multigrammaticality

A first distinctive feature of discourse markers has to do with their grammatical status: they show a great categorical heterogeneity (also named multicategoriality or multigrammaticality). The functional category that they form is far from being well-defined. They rather constitute “a broad and diverse class of elements with different developmental trajectories” (Koops & Lohmann 2015: 233). This diversity of grammatical and lexical origins includes, non-exhaustively, the following classes: adverbs (e.g. *now*, *anyway*), coordinating and subordinating conjunctions (e.g. *but*, *because*), interjections (e.g. *oh*, *boy*), meta-expressions (e.g. *in other words*), minor clauses (e.g. *I mean*, *you know*), prepositional phrases (e.g. *by the way*), response words (e.g. *yeah*, *no*) and verbs (e.g. *say*, *look*) (Schourup 1999; Fung & Carter 2007). The accepted multigrammaticality of discourse markers contradicts Östman’s (1982: 153) ‘uniqueness criterion’, i.e. the fact that “for an item to be called a pragmatic particle, it should never be able to have any other than a pragmatic-particle

function”. In practice, however, very few cases fulfill this condition, since the majority of the linguistic expressions functioning as discourse markers have – at least – one other, grammatical meaning as well (e.g. adverbs and conjunctions) (Aijmer 2002).

Importantly, a linguistic expression that has acquired a discourse marker function does not lose its initial lexical or grammatical function per se. As explained by Schourup (1999: 234), “an item retains its non-DM [discourse marker] syntactic categorization but does ‘extra duty’” as a connective. In this way, the pragmatic use of an item in spoken discourse often coexists with another more or less closely related meaning in speech or even in writing, as is the case for *I mean*, for instance, which can equally function as a lexical verb phrase in the present tense and as a discourse marker expressing clarification (Huang 2013). Note that the ratio between DM and non-DM use of one and the same expression shows a great deal of variation depending on the item under study (cf. Lenk 1998a). The multigrammaticality of discourse markers – sometimes giving rise to ambiguous cases – strengthens the need for a manual disambiguation based on a careful examination of the context (Fung & Carter 2007).

1.4.2 Orality

Discourse markers form a typical and quite salient feature of oral discourse, in which a single sentence can contain up to several instances. Because of this high frequency of use, discourse markers are often “stylistically stigmatized [...] and deplored as a sign of dysfluency and carelessness” (Brinton 1996: 33), although they are sometimes presented in a more positive light. If they are predominantly associated with the spoken medium, it is because this channel tends to be characterized by a greater informality and familiarity – to which discourse markers happen to contribute – than the written code. Note that the distinction between formal and informal is more a matter of scale than a clear binary opposition.

Besides, Östman (1982: 169) argues that the use of discourse markers is closely related to the lack of planning time inherent to ‘impromptu speech’, and that discourse markers function as a compensating strategy for the grammatical fragmentation which may ensue from limited thinking time. The almost automatic association of discourse markers with oral discourse may be biased, though, given the fact that a vast majority of the studies on discourse markers up to now have been based on speech data (cf. Schiffrin 1987). Some markers are indeed primarily used in spoken discourse, such as *by the way*, which clearly encodes the spontaneity of the oral interaction. On the other hand, other discourse markers

such as *conversely* are predominantly – if not exclusively – used in writing since they “encode a high degree of utterance planning” (Schourup 1999: 234).

1.4.3 Initiality and prosody

Although this feature cannot define discourse markers on its own, it shows their strong tendency to occur in utterance-initial position. Hansen (1997: 156) argues that discourse markers “prototypically introduce the discourse segments they mark”. The position of discourse markers may be flexible: occasionally utterance-medial, as in “Corgis, **now**, are an intelligent breed” (Schourup 1999: 233), less frequently utterance-final, as in “She likes all kinds of music classical er mainly classical **I think**” (Fung & Carter 2007: 413). However, initiality must be at least possible for an item to be qualified as ‘discourse marker’ (Schourup 1999; Fung & Carter 2007). Schourup (1999: 233) hypothesizes that this characteristic of discourse markers is related to “their ‘superordinate’ use to restrict the contextual interpretation of an utterance”.

Due to their typical position outside the syntactic structure of an utterance or at its extremes, discourse markers generally have a particular prosody as well. Indeed, their peripheral location often correlates with phonological independence (Schourup 1999). Discourse markers constitute separate tone groups and must “have a range of prosodic contours” (Schiffrin 1987: 328). These include pauses, phonological reduction and a distinctive intonation, among others (Fung & Carter 2007). Hansen (1997) points out that certain clause-internal discourse markers such as the French *donc* and *puis* show intonational integration. This does not alter their status of ‘discourse markers’ in any way.

1.4.4 Weak clause association and optionality

Although very useful in spoken interaction, discourse markers are an optional device when organizing what one wants to communicate. Their absence does not make the sentence ungrammatical or unintelligible² in any way since discourse markers do not really have a – lexical, at least – meaning of their own. In that, they differ from other commentary pragmatic markers such as *frankly* or *certainly*, which do contribute to propositional content (Fraser 1988). The optionality of discourse markers is related to the weak clause association that they

² Consider, for instance, the following two examples taken from Schourup (1999: 231):

(a) The others are going to Stoke. **However**, I am going to Paris

(b) The others are going to Stoke. I am going to Paris.

The relationship of contrast expressed by *however* in (a) is less explicit in (b), but this second example is still perfectly grammatical and understandable.

display: they are only ‘loosely attached’ to the syntactic structure (Brinton 1996) or even stand outside of it, being “independent of an already well-formed sentence” (Fraser 1988: 22).

Discourse markers rather function as ‘powerful clues’ about the propositional connections that they signal, “guiding the hearer toward a particular interpretation and simultaneously ruling out unintended interpretations” (Schourup 1999: 231-32). Removing them from a stretch of discourse simply makes the relationship between a proposition and the rest of the message less explicit (cf. Schourup 1999; Fung & Carter 2007).

1.4.5 Non-truth-conditionality

This feature is a necessary attribute for an expression to be qualified as ‘discourse marker’ (cf. Schourup 1999). It simultaneously entails and results from another property of discourse markers, namely their grammatical and semantic optionality (cf. Section 1.4.4 above). Non-truth-conditionality refers to the fact that the “existence [of discourse markers] does not affect the truth condition of the propositions” (Fung & Carter 2007: 414). In this way, they do not contribute to the semantic content of the utterance in which they occur.

This feature distinguishes discourse markers from other ‘content’ words such as adverbs, which do add something to propositional meaning (Schourup 1999). In this respect, Gazdar (1979) proposes an approach that opposes *semantics*, i.e. the study of truth-conditional meaning to *pragmatics*, i.e. “meaning minus truth conditions” (cited in Blakemore 2006: 222). Blakemore (2006: 222) further argues that looking into the non-truth-conditionality that characterizes discourse markers has helped impose the idea of the “non-unitary nature of linguistic meaning”.

1.5 Functions in spoken interaction

The functions performed by discourse markers in spoken interaction are many and various. Their multiplicity is partly related to the diversity in their terminology, since linguistic expressions are often labeled and defined on the basis of their function(s) (Brinton 1996). In this respect, there exists a whole range of terms such as ‘connectives’, ‘fillers’ and ‘politeness markers’ which attempt to describe what discourse markers actually *do* in oral communication (Aijmer 2002). However, these expressions are rather imprecise and/or inappropriate, as pointed out by Kroon (1995), who deplores the lack of a “suitable metalanguage” when it comes to classifying the multiple functions of particles (cited in Aijmer 2002: 25). Aijmer

(2002: 24) emphasizes the importance of an accurate classification of what she calls discourse particles, as “[a] discourse function should be defined in such a way that it is clear how a discourse particle is similar to and different from other discourse particles”. Overall, however, establishing a clear taxonomy of the uses of discourse markers turns out to be challenging – if possible at all.³

In addition, a single discourse marker generally fulfills more than one function – be it simultaneously or in distinct occurrences. Interestingly, some of the functions commonly attributed to discourse markers may be performed by other strategies such as intonation and lexical repetition, among others (Brinton 1996). Schiffrin (1987: 57) illustrates the latter case by the example of a speaker who starts making a point with the discourse marker *see*, before further developing her argument through the repeated use of *because*. Finally, the uses that emerge out of the study of an individual marker may show some degree of overlap with those fulfilled by another particle (e.g. the use of *well* to mark hesitation is also one of the functions fulfilled by *like*) (Brinton 1996).

Unsurprisingly, setting up a clear and complete list of discourse markers functions is not an easy task. The study of different types of communicative situation would be necessary to obtain an exhaustive list of the possible uses of a particle and their distribution (Aijmer 2002). However, “the links between a particle and its uses may be conventionalized into immediately recognizable routines because they are the most relevant, salient, frequent or popular” (Aijmer 2002: 26). These common, ‘basic’ functions of discourse markers have been classified in different ways by several linguists. The following sections provide an overview of the variety of possible functions.

1.5.1 Connectivity

Connectivity is undoubtedly the primary function of discourse markers since it also constitutes one of their defining features (Schourup 1999). Schiffrin’s (1987: 31) definition of discourse markers as “elements which bracket units of talk” suggests that connectivity necessarily occurs between two or more textual units. Other linguists do not subscribe to this view, however. Blakemore, for instance, puts forward the following argument:

[...] it is preferable to view certain DMs not as necessarily relating two segments of text, but as relating the propositional content expressed by the current utterance to assumptions that may or may not have been communicated by a prior utterance. (Schourup 1999: 230-231).

³ Andersen (2001), for instance, believes that a particular function is attributed to a given marker through a process of pragmatic inference in the communicative situation.

Discourse markers function as indexical devices bringing coherence between an utterance and its surrounding context (Fung & Carter 2007). They make the relation between old and new information more explicit, thus contributing to the logical development of the communication (Jabeen et al. 2011). Schourup's (1999) 'connective function' corresponds to the first 'textual use' in Andersen et al.'s (1999) taxonomy, i.e. how markers structure discourse by coordinating turns, speech acts, etc.. Similarly, Croucher (2004: 40) classifies these uses of discourse markers under two of his 'formal textual functions', namely "indicate a turn in conversation" and "frame the general conversation". Typically, connective functions are fulfilled by expressions such as *well*, *OK*, *you know*, *now* and *anyway*, among others (cf. Fung & Carter 2007). In the following example, *so* connects the two utterances: "A: You take the first turning on the left. B: **So** we don't go past the university (then)" (Blakemore 1987: 85).

1.5.2 Co-construction of meaning

A second group of textual functions commonly fulfilled by discourse markers is of more cognitive nature. Such discourse markers give clues about the construction of the exchange and help the hearer "construct a mental representation of the discourse" (Fung & Carter 2007: 415). In unplanned speech, digressions and topic shifts are frequent, and discourse markers help the speaker carry across their intended message. *By the way*, for instance, warns the hearer about an unexpected change of topic (Croucher 2004).

Besides, discourse markers may constitute a number of 'traces' signaling the immediate co-construction process taking place and caused by time constraints. These include marks of the thinking process (e.g. *I think*, *I see*), hesitations (e.g. *well*, *sort of*), reformulation (e.g. *I mean*, *in other words*), elaboration (e.g. *like*, *I mean*), etc. (Fung & Carter 2007). The phrase *I mean* turns out to be particularly useful in that it allows the speaker to further modify a statement that they have previously uttered and to clarify some information that might be unclear for the hearer, as in "They all know that MQM can use terrorism. **I mean** ALtaf Hussain is a terrorist in every definition" (Jabeen et al. 2011: 82). Andersen et al. (1999: 1340) consider that discourse markers in such uses "signal production problems on the part of the speaker". A last function attributed to discourse markers in this category is that of gap/pause filler, i.e. the fact that they give the speaker thinking time to formulate their idea (cf. Croucher 2004; Jabeen et al. 2011).

1.5.3 Feedback and attitude

In addition to the textual functions described above, discourse markers also fulfill what Brinton (1996) calls ‘interpersonal functions’, i.e. the uses in which they contribute to indicating the participants’ stance towards the speech event (Alami 2015). On the one hand, discourse markers can “provide feedback from listeners about whether a prior utterance has been understood or not, and whether they agree or disagree” (Andersen et al. 1999: 1340). In this way, discourse markers such as *you know*, *right*, *I see*, *sure* or *great* may be used to indicate shared knowledge, (dis)agreement, acknowledgement, consent, support and confirmation (Fung & Carter 2007; Jabeen et al. 2011). In the following example, for instance, the speaker introduces the reason for his disagreement by the discourse marker *you know*: “I am sorry to take names but **you know** country is in ~ country is probably still developing” (Jabeen et al. 2011: 81). On the other hand, discourse markers (e.g. *like*, *well*, *I think*, *sort/kind of*) can also reflect the speaker’s attitude and stance towards the communicative situation and the propositional content (Croucher 2004; Fung & Carter 2007). This corresponds to the ‘subjective interpersonal use’ of discourse markers in Brinton’s (1996) dichotomy⁴. Hinting at the speaker’s orientation towards the discourse is, according to Alami (2015), an intrinsic characteristic of discourse markers.

1.5.4 Social relationships

A fourth class of functions fulfilled by discourse markers that is less frequently mentioned in the literature is their use as “markers of the social relationships between interlocutors in any speech event” (Andersen et al. 1999: 1340). For example, discourse markers such as *well* and *now*, when used in initial position, may reflect the level of intimacy and the hierarchy between the speakers, as well as the register used in the conversation (Andersen et al. 1999). Not only do such markers reflect some aspects of the communicative situation, they can also affect them. In this way, one can use discourse markers “[i]nteractively to achieve intimacy between speaker and addressee” (Alami 2015: 6) as well as to “mak[e] communication more interactive, involving and informal” (Fung & Carter 2007: 420). In addition, discourse markers can be used for politeness purposes: they allow the speaker to maintain face⁵ to soften their statements and, as a result, to sound more polite (Jabeen et al. 2011).

⁴ Brinton (1996: 38-39) distinguishes between the interpersonal functions of discourse markers and their textual functions (e.g. filler, topic switch, turn-taker or information indicator).

⁵ This notion was formulated by Brown & Levinson (1987) as part of their model of *politeness theory*. ‘Face’ refers to an individual’s self-image, as displayed to others in interaction.

This chapter provided an overview of the main properties of discourse markers in general. After a review of the terminology employed to define discourse markers and the different perspectives from which they have been studied, the main criteria for considering an item a discourse marker were introduced. Then, the various functions associated with discourse markers were described and illustrated. The following chapter looks into the discourse marker *like* specifically, through a detailed account of its evolution, formal features, functions and sociolinguistic determinants.

Chapter 2

Like as a discourse marker

2.1 Introduction

As part of the growing interest for the study of discourse markers, *like* has gained increasing attention over the last few decades. So far, the literature on the discourse marker *like* remains rather sparse, though. This may result from the absence of a unanimously accepted framework within which discourse markers in general can be defined and studied. Overall, the discourse marker *like* bears a rather negative connotation, considered by many as a simple pause filler devoid of any real linguistic interest and contributing to vagueness of expression. In the OED, one finds such terms as ‘colloquial’, ‘dialectal’ and ‘vulgar’ when it comes to defining *like* as a discourse marker. D’Arcy (2005: 230) goes even further by describing it as a “crutch for lexical indecision”. Despite this rather negative picture, *like* has recently gained considerable importance in spoken practices, having become nowadays one of the most frequent discourse particles in spoken communication with up to one occurrence of *like* every twenty words in spontaneous speech (Tagliamonte 2005). The fact that *like* does not disrupt the syntax of the sequence in which it occurs may partly explain its frequency and its convenience of use. In addition, *like* as a discourse marker turns out to be extremely useful in oral discourse thanks to the wide range of pragmatic functions it fulfills.

The objective of this second chapter is to provide an overview of the main characteristics of the discourse marker *like*. Section 2.2 describes the different steps of the linguistic development that has led to the current use of *like* as a discourse marker, until its most recent use as quotative *be + like*. The next section focuses on three of the distinctive features of *like*, namely position, scope and prosody (Section 2.3). In Section 2.4, the various functions that *like* can take on in oral discourse are summarized and illustrated. Finally,

Section 2.5 looks into the sociolinguistic factors that influence the use of *like* as a discourse marker, with a particular focus on cross-generational and gender-related variation.

2.2 Evolution of *like*

2.2.1 History of *like*

Like in its non-verb, pragmatic function has a long history: the OED records instances that date back to more than two centuries. These are based on written examples, however, so it is reasonable to assume that the first discourse marker uses of *like* were to be found much earlier in spoken English (Miller & Weinert 1995; D’Arcy 2005). The use and role of *like* nowadays seem to be the result of successive stages in a long evolution process. Different developmental trajectories have been suggested in the literature, generally opposing the British and the American uses.

Like first used to occur in clause-final position and could be paraphrased as ‘as it were’ or ‘so to speak’, as in the following example which dates back to 1838: “If your honour were more amongst us, there might be more discipline **like**.” (D’Arcy 2005: 4). According to Partridge (1984: 683), such cases belong to the adverbial use of *like* at the end of a phrase or a sentence, used to express “vagueness or after-thoughted modification”. In more modern uses as an adverb, *like* also “tones down an expression or a whole sentence [...] without much affecting the meaning”, as in “you pile it up **like**” (Miller & Weinert 1995: 368).

Like was further used in the sense of ‘similar to’ or ‘approximately’, along with a post-verbal noun phrase. This prepositional use of *like* still exists nowadays, as in examples like: “That sounded **like** a lecture” (Meehan 1991: 40). According to the OED, this comparative sense of *like* derives in turn from a former adjectival use, meaning ‘in the same manner or the same extent as’, which can be traced back to as early as the 14th century.

Subsequently, this application of *like* extended its scope to quantitative noun phrases, which often contain the highest informative value in the sentence (e.g. “I wrote it in **like** ten minutes”) (Meehan 1991: 41). Whether *like* in such cases can be considered as a discourse marker is debated, though. Siegel (2002), for instance, argues that in numerical contexts, *like* affects truth conditions and therefore violates one of the necessary criteria to be called ‘discourse marker’ (see Section 1.4.5 on non-truth-conditionality). Müller (2005), by contrast, acknowledges this use of *like* with numerical expressions as one of its several discourse

marker functions. Later, *like* took on a comparative reading with the meaning of ‘as if’, further expanding its scope over entire clauses, as in “it was **like** I was watching someone else do it” (Meehan 1991: 41). From the 19th century onwards, *like* developed another usage, namely introducing examples (e.g. “Do you have **like** a mint or something?”), in the way it does today to enumerate objects, with a still wider scope and a growing focusing and highlighting function (Meehan 1991: 42-43).

This list of uses shows the multitude of roles that *like* has fulfilled through time, constantly undergoing a slow shift from one usage to the next, or taking on several functions at a time. In this respect, Meehan (1991) observes that change of meaning and change of scope have exerted a mutual influence in the process. In her research on the origins of the pragmatic functions of *like*, D’Arcy (2005: 8) points out that “this type of trajectory, where a lexeme develops a new function while simultaneously undergoing semantic and pragmatic change, may be indicative of grammaticalization”. The following section is devoted to this phenomenon.

2.2.2 Grammaticalization of *like*

Several linguists (cf. Romaine and Lange 1991; Traugott 1995; D’Arcy 2005) have addressed *like* in its discourse marker usage as the result of grammaticalization⁶, defined by Lehmann (1982) as “a process which may not only change a lexical into a grammatical item, but may also shift an item from a less grammatical to a more grammatical status” (cited in Meehan 1991: 37). In other words, the adverb *like* is progressively losing its lexical content in favour of a more functional and interpersonal nature. Interestingly and as pointed out by Hansen (1998: 225), however, “... it may be that, instead of making a binary distinction between lexical and grammatical items, we should make room for a third, possibly intermediate, category of ‘discourse items’”. Aijmer (2002) also emphasizes the graduality and incompleteness of the process when it applies to discourse markers, which – contrary to ‘regular’ grammaticalized elements – do not integrate with the utterance in which they occur.

When it comes to the actual evolution that *like* has been undergoing, D’Arcy (2005) observes a distinction between the British and the American developmental pathways. She argues that, on the one hand, the current, omnipresent discourse marker *like* with forward scope seems to arise from the North American branch of the English language, as the product

⁶ A terminological remark was made by Lenk (1998b), who argues that ‘pragmaticalization’ is a more appropriate term than ‘grammaticalization’ to refer to the functionalization of discourse markers.

of an evolution from adverb to sentential adverb to discourse marker. According to Underhill (1988: 234), this North American innovative use of *like* is ‘non-standard’ and even ‘entirely ungrammatical’. Meehan (1991) rather speaks of a new linguistic trend. On the other hand, the more traditional British use of *like* – in clause-final position with backward scope – appears to be typical of an earlier stage of development, when it functioned as sentential adverb (D’Arcy 2005). As pointed out by Romaine & Lange (1991: 262), however, “[a] simple linear model of grammaticalization is inadequate to account for these developments”; instead, they suggest the idea of a “network of related meanings”. A further development that ensued from the discourse marker use of *like* is the quotative *like*, described in the following section.

2.2.3 Quotative *like*

Another usage of *like* which must be distinguished from the discourse uses described above is the quotative ‘*be like*’. It has become increasingly prominent in speech since the end of the twentieth century, subsequently gaining growing academic attention. This non-traditional quotative is thought to have originated in California, later spreading to other English-speaking regions, and it has typically been associated with the speech of adolescents (Blyth et al. 1990; Macaulay 2001). Besides, research has shown that this phenomenon is mostly prominent among young female speakers (Blyth et al. 1990; Dailey-O’Cain 2000).

The quotative *like* is indicative of a still further stage on the trajectory of grammaticalization. Meehan (1991) hypothesizes that this use derives from the ‘as if’ sense of *like*, already introducing entire clauses. In an example such as the following: “He’s **like**, “You have to calm down”” (D’Arcy 2005: 3), the collocation ‘*be + like*’ functions as a marker of reported speech and thought (Romaine & Lange 1991). In this respect, however, the ‘reported’ clause is not necessarily a verbatim replication of a prior utterance. Indeed, “many quotes in informal spoken discourse are invented and [...] their main function is as dramatizations” (Meehan 1991: 47). Contrary to its discourse marker uses, the quotative *like* cannot be omitted without affecting the grammaticality of the utterance (see Section 1.4.4 on optionality). Therefore, this special use of *like* will be left out of account in the present study.

2.3 Position, scope and prosody

Due to its discourse marker status, *like* can occur in a wide range of contexts and next to various constituents, either syntactically bound (clause internal) or unbound (Tagliamonte 2005). Studies have shown that the placement of *like* is rule-governed and that there is a certain systematicity in the way it behaves as a discourse marker. Its position, scope and prosody are closely interrelated features and both influence and depend on the function that *like* fulfills in a particular utterance (cf. Miller 2009).

2.3.1 Position

When it comes to the position of *like* as a discourse marker, there is general agreement that it can occur almost anywhere in the utterance since it enjoys “syntactic detachability and positional mobility” (Romaine & Lange 1991: 261). However, as stated by Meehan (1991: 40), “the positioning of *like* in the sentences is never random although to some hearers it may seem so”. In line with the general tendency for discourse markers to occur in utterance-initial position (cf. Section 1.4.3), *like* is located most of the time at the beginning of a fragment, introducing constituents⁷ of variable nature and length (Underhill 1988). In her study of discourse markers among Canadian teenagers in Toronto, Tagliamonte (2005: 1902) observes that, in a large majority of cases, *like* occurs before a noun phrase – of any kind: (in)definite or quantified, pronominal or not – in line with the findings of other studies about the syntactic distribution of *like*, viz. “[its] propensity in pre-sentential and pre-noun phrase position” (cf. Tagliamonte & D’Arcy 2009). The second most frequent position for *like* is sentence-initial (Romaine & Lange 1991; Tagliamonte 2005), as is the case in this example: “You seem dissatisfied. **Like** what do you mean?” (Underhill 1988: 244).

Underhill’s (1988) study yields slightly different results with most cases of *like* preceding a VP, as in “There was this guy **like** sleeping in my doorway”, then an NP, as in “Can’t you get **like** a body wave up front?” (Underhill 1988: 237-8). Other – less frequent – constituents that can be introduced by *like* include adjectives, adverbial and prepositional phrases as well as other discourse markers (Underhill 1988; Tagliamonte 2005). *Like* can also occur in sentence-final position, although this represents a small number of cases (Romaine &

⁷ Underhill (1988) uses the term *constituent* in an imprecise manner to refer to any kind of linguistic element – in a broad sense – that can enter the scope of the discourse marker *like* (e.g. NPs, VPs, adjectives, adverbs, subordinates clauses, or even the entire sentence when *like* occurs in initial position).

Lange 1991). Overall, *like* is typically located where the speaker's lexical indecision is highest (Siegel 2002) and rarely occurs within highly fixed phrases (Andersen 2001).

2.3.2 Scope

In the 'American' usage of *like* (cf. Section 2.2.2), the discourse marker has a forward scope, i.e. it qualifies the segment to its right, and its scope can vary from single terms to whole propositions (cf. Andersen 2001). Interestingly, *like* shows an increase of scope typical of the grammaticalization process that it has undergone as a discourse marker (D'Arcy 2005). However, the rather vague notion of 'constituent' introduced by Underhill (1988) indicates "a degree of indeterminacy in both the syntactic status and scope of the elements [*like*] may modify" (Romaine & Lange 1991: 252). If Siegel (2002) argues that sentence-initial *like* generally scopes over entire sentences, Popescu-Belis & Zufferey (2011) claim just the opposite. In this respect, some cases turn out to be ambiguous and difficult to analyze, such as the following: "**Like**, a sort of mini-tornado knocked the tent over" (Siegel 2002: 12). In this sentence, two different interpretations are equally possible when determining the domain of *like*, namely subject scope or sentential scope. This illustrates the complexity of defining the exact scope of an instance of *like*.

2.3.3 Prosody

A last aspect that characterizes *like* as a discourse marker is its distinctive prosody, which – together with its position – turns out to be a helpful feature in identifying discourse markers. In addition, research has shown that duration can help distinguish discourse marker uses of *like* (average duration) from non-discourse marker uses (either short or long occurrences) (Popescu-Belis & Zufferey 2011). The general tendency for discourse markers to be phonologically reduced and independent has already been described in Section 1.4.3 and *like* is no exception to this rule. Here again, the grammaticalization of *like* has been associated with 'phonological attrition', i.e. the gradual loss of phonological substance (D'Arcy 2005). Finally, in spoken discourse, *like* is typically surrounded by pauses, which often translate into commas in the written transcription of an utterance (Siegel 2002).

2.4 Functions in spoken interaction

Like is one of the most versatile discourse markers in English, performing a wide range of pragmatic functions essential to regulate spoken interaction. In this regard, Andersen (1998)

argues that the various – seemingly contrastive – usages of *like* in speech actually represent subcategories of one single, overarching function, namely contributing to relevance in an utterance. Other scholars have attributed different functions to *like*, however, including a discourse organizational function (Dailey-O’Cain 2000) and a focusing/highlighting function (Underhill 1988). In addition to this broad inventory of uses, *like* very often fulfills more than one function at a time, which makes their identification and classification even more challenging. This multifunctionality has been studied by various authors. Among them, Müller (2005) provides a comprehensive account of the functions fulfilled by *like* as a discourse marker. The following six sections draw on her classification, which will further serve as methodological basis for my analyses.

2.4.1 Marking lexical focus

Highlighting or focusing on (new) information is probably the most frequent function of *like* as a discourse marker. It has been studied by many scholars (cf. Meehan 1991; Dailey-O’Cain 2000). The same label ‘focusing function’ is often used to refer to different concepts, however. For instance, Underhill (1988: 238) views focus “[as] the most significant new information in a sentence – often, the point of the sentence”, in which he uses the terms ‘focus’ and ‘new information’ indistinctly. Miller & Weinert (2005), by contrast, argue that *like* can focus on given information as well and they compare *like* to other – more powerful – focusers such as *it*-, *wh*- or reverse clefts. *Like* advantageously fits in a wider variety of contexts, however, thanks to its greater syntactic freedom and the absence of deictic constraint on its use.

As a non-contrastive focuser, *like* can scope over various elements including noun and verb phrases, adverbs and adjectives. It never qualifies full clauses, however (Müller 2005). Typically, *like* occurs at or near the end of the sentence, where the focused information is located (Underhill 1988). A last subtype of the focusing function of *like* is what Andersen (2001) calls its ‘hyperbolic use’, i.e. when *like* highlights a clearly exaggerated expression, in the same way as stress or particular intonation does, as in: “and he still had **like** tons of eyeliner” (Müller 2005: 223).

2.4.2 Searching for the appropriate expression

A second major function fulfilled by *like* is that of ‘signpost’ in spontaneous conversation. In this function, it gives the speaker processing time to make up or refine what they are going to utter next (cf. Miller & Weinert 1995). This function is clearly illustrated in the following

example: “Madonna? Yeah she’s pretty I mean **like**... she looks better with brown hair though” (Andersen 2001: 210). *Like* thus enables the participants to maintain the floor during a turn by filling the pause necessary to find the appropriate expression (D’Arcy 2005). As a result, the use of *like* negatively correlates with thinking time before an utterance: the more time the speaker has or takes, the less frequent the particle is (Siegel 2002). Note that in this usage, *like* is often surrounded by – filled or unfilled – pauses, which clearly indicates that the speaking turn is to be continued (Müller 2005). What Schourup (1983) qualifies as ‘pausal interjection’ thus appears extremely useful to ensure the fluency of the oral exchange.

This function of *like* has sometimes been associated with lexical indecision on the part of the speaker. This corresponds to the ‘hesitational use’ of *like* in Andersen’s (2001) terminology, i.e. the difficulty for the speaker to find the words that accurately reflect their thought. In this respect, Schourup (1983: 42-43) argues that “*like* indicates a possible minor nonequivalence of what is said and what is meant”, and this idea of a ‘slight mismatch’ is shared by many other authors. Müller (2005), for instance, draws on Jucker & Smith’s (1998) idea of a ‘loose fit’ between utterance and thought to explain how *like* “instructs the addressee not to take the following too literally” (Müller 2005: 200). Whether, at the time of the utterance, the speaker’s idea is already well-defined but difficult to word or still fuzzy in their mind remains debated, though. Note that this function of *like* is also interpersonal since it conveys some information about the speaker’s relation to what is being said (Siegel 2002). Besides, it allows the speaker to weaken their commitment towards the propositional content or the manner in which they formulate it.

2.4.3 Marking an approximate number or quantity

A third purpose for which the discourse marker *like* can be used is that of marking an approximate number or quantity, as in: “If there was a lot of traffic, **like** twenty-five minutes maximum” (D’Arcy 2005: 35). This function of *like* appears to be relatively frequent in spoken data (cf. Meehan 1991; Andersen 2001) and D’Arcy (2005: 47) even observes that *like* has become “the preferred adverb for expressing approximation in numerical contexts among speakers under 30”. This function of *like* shares with the preceding one the speaker’s reduced commitment to the exactness of the utterance (Müller 2005). While traditionally, only cases where *like* precedes actual numbers have been taken into account, Meehan (1991) opts for the more inclusive expression “quantitative noun phrases”. *Like* can thus qualify numbers as well

as expressions of quantity, time period and frequency when specifying their exact value is not necessary for the comprehension of the utterance.

2.4.4 Introducing an example

The ‘exemplifier’ function of *like* – equivalent to ‘such as’ or ‘for example’ – has been widely recognized and illustrated in the literature (cf. Schourup 1983; Meehan 1991; Miller & Weinert 1995). Importantly, this function can be fulfilled by the word *like* in two distinct contexts. Therefore, a distinction must be made between, on the one hand, cases where *like* cannot be omitted without making the sentence ungrammatical, i.e. its use as a preposition, and, on the other hand, optional, discourse marker uses of *like*, as in “uh the minor characters would be... **like** the guys on the boat I guess” (Müller 2005: 212). In this latter case, *like* “signals an accurate but selective representation of what the speaker has in mind” (Miller & Weinert 1995: 369). In other words, introducing a particular – well-chosen – example with *like* can exemplify a general concept just as well as a definition or a list does. In this respect, note that the wider category that gets illustrated does not need to be explicitly mentioned before the example; generally, it can easily be inferred through the context. For example, in a sentence such as “you know he started to order **like** coffee and other stuff”, the context of a restaurant makes it clear that ordering food and beverages is the proper thing to do (Müller 2005: 214).

2.4.5 Introducing an explanation

Although relatively frequent, this function of *like* as a discourse marker has only been acknowledged by Müller (2005: 215) in examples such as the following: “...they would show the movie, and then there would actually be someone playing the piano there, **like** along for the accompaniment”. Müller (2005: 215) argues that, here, *like* does not hedge what the speaker says, nor does it introduce new or focused information; rather, it “extends the information given [...] to make it more understandable”. The clarification brought by *like* may take the form of a repetition of the previous extract in slightly different words or of a reformulation of the content with an alternative term expressing the same idea, as is the case in the following example: “I I’d rather ... say it was neutral for me **like** not good not not really bad” (Müller 2005: 218).

2.4.6 Discourse link and restart

In some cases, the discourse marker *like* has no real function other than that of link between two (parts of) sentences. As such, it is found exclusively within the speech of single speakers, since it contributes to the development and the narration of an idea (Müller 2005). In addition, Schourup (1983: 40) observes that *like* can also occur in the context of a restart, i.e. “a point at which the present speaker stops an item under construction and recommences”. *Like* thus indicates the speaker’s self-repair after a false start. This restart can either be expressed by the same words as those used in the part preceding *like* or by a new structure, as in: “it was just funny how he got his ...**like** how he sometimes leaned in there” (Müller 2005: 224).

2.5 Sociolinguistic determinants of the use of *like*

Various studies (cf. Fleischman 1998; D’Arcy 2005; Tagliamonte 2005) have revealed that *like* as a discourse marker is not evenly employed by all the members of a given English-speaking community. The use of *like* appears to be largely influenced by a number of sociolinguistic factors, which results in an unequal distribution of the discourse marker among a heterogeneous group of speakers. This section focuses on two of those factors which appear to be particularly determining in the use of *like* as a discourse marker, namely gender and age.

2.5.1 Gender

Speaker’s gender appears to be a strong sociolinguistic determinant in the use of *like* as a discourse marker. *Like* first originated among the so-called Valley Girls in California in the 1980s (Blyth et al. 1990; Siegel 2002) and seems to have remained a feature essentially characteristic of female speech since then, although not exclusively used by them. Overall, research has shown a higher frequency of use of *like* among females than among males (cf. Andersen 2001; D’Arcy 2005; Tagliamonte 2005), which endows *like* with its frequent status of ‘female marker’ in the literature. In her study, Tagliamonte (2005), for example, found out that, in general, a vast majority of the instances of *like* were produced by girls. However, she also observed that this correlation is only significant among the older teen group, which makes her conclude that “sex differences – at least with respect to these discourse/pragmatic features – are developmental, and [...] are created in the speech community, within the peer group” (Tagliamonte 2005: 1912-13).

One reason for this sex difference in the use of *like* as a discourse marker has to do with the linguistic change hypothesis that will be detailed below (see Section 2.5.2.1). Sociolinguistic studies have shown that linguistic change and gender asymmetry inherently go hand in hand, and that women typically lead linguistic change (cf. Eckert 1988; Labov 1990). Tagliamonte & D’Arcy (2009: 63) further argue that “once a change becomes associated with women, men either retreat from or resist the incoming form”. In the present case, the female tendency to lead innovation may be related to the pragmatic functions fulfilled by *like* as a discourse marker: women tend to use more hedges and intensifiers in their speech (Holmes 1990), which are two core functions of the discourse marker *like*.

2.5.2 Age

From a sociolinguistic perspective, *like* can be addressed as part of the effects of age differences on speakers’ discursive and pragmatic choices. In this respect, note that if lexical, phonological and syntactic phenomena are recurrent objects of study with regard to age, “pragmatics represents an understudied area of cross-generational variation” (Andersen 2001: 2). Still, the effect of age on the use of *like* as a discourse marker has been investigated by a number of linguists (cf. Fleischman 1998; Dailey-O’Cain 2000) and traditionally, *like* has been considered a typical feature of adolescent speech, in line with discourse markers in general. In her study of Torontonians teenagers, Tagliamonte (2005) noticed a strikingly high concentration of *like* among 15 to 16-year olds. This was confirmed by D’Arcy’s (2005) quantitative findings showing that the discourse marker *like* is typically associated with speakers below the age of thirty.

Overall, adolescence correlates with strong linguistic differentiation and considerable influence from the peers (Tagliamonte 2005). These two factors can partly explain why *like* as a discourse marker has been adopted so easily by young communities of speakers. It functions as a linguistic marker that reflects the feeling of belonging to a group, therefore illustrating how language and social relations are closely intertwined, and how linguistic choices and practices contribute to social distinctiveness⁸. As explained by Eckert (1988: 198), “identification within a peer-defined world” lies at the root of the spread of linguistic change. In this respect, two conflicting theories have been put forward to account for the recent rise of

⁸ The notion of social distinctiveness was also studied by Irvine (2001: 21), who argues that ‘style’ – be it in language or in other domains – “crucially concerns distinctiveness; though it may characterize an individual, it does so only within a social framework”.

like as a discourse marker, namely linguistic change and age-grading pattern. These are further detailed in the following two sections.

2.5.2.1 Linguistic change

The sharp increase that *like* as a discourse marker is undergoing nowadays has been considered by some scholars as a manifestation of ongoing linguistic change in the English language. D’Arcy (2005: 204), for instance, argues that “when viewed across the generations, the discourse uses of *like* will be seen to have developed gradually and systematically, arriving at their current state through regular processes of language change”. As far as *like* is concerned, Tagliamonte & D’Arcy (2009: 77) describe this change as “overt and vigorous”.

As mentioned earlier, *like* is typically associated with young speakers, who are, in turn, “unanimously singled out as the innovators” at the origin of linguistic evolution (Tagliamonte 2005: 1897). The age gradation displayed by the discourse marker *like* appears to be an argument in favour of the linguistic change hypothesis. Following this reasoning, one can expect that *like* will further spread to other age classes in the English-speaking population, as “the young speakers mature who are currently its exclusive users” (Blyth et al. 1990: 223), but a diachronic study of the use of *like* would be necessary to back up this hypothesis.

2.5.2.2 Age-grading pattern

An opposite viewpoint thereupon is that of Tagliamonte (2005), who claims that *like* as a discourse marker rather follows a typical age-grading pattern, i.e. “a change correlated with a particular time of life” (Chambers 1995, cited in Tagliamonte 2005: 1910). In other words, a given linguistic practice is steadily associated with a particular age section of the speaking community and does not spread to other generations.

In the present case, the age-grading hypothesis is supported by the fact that the maximal frequency of *like* as a discourse marker appears within a very restricted age cohort, usually concentrated around fifteen years old. *Like* usually gets acquired in primary school, later developing to reach a peak in frequency of use during secondary school. Then, it suddenly drops when adolescents enter university, and as they become more conservative again and adopt a language which is closer to the mainstream, adult norms (Tagliamonte 2005). Besides, D’Arcy (2005: 45) hypothesizes that “higher rates of approximation [might be] characteristic of younger speakers”, therefore explaining the massive use of *like* as a

discourse marker among teenagers. Evidence based on diachronic data would be needed to confirm this hypothesis, however.

This chapter provided an overview of the main formal and functional aspects of *like* as a discourse marker. After describing the long evolution and the grammaticalization process that have led to its current discourse marker use, three features that condition its linguistic behavior were described, namely position, scope and prosody. Then, the various pragmatic functions fulfilled by *like* were summarized and illustrated. Finally, two of its sociolinguistic determinants were expanded upon, namely speaker's gender and age. The following chapter introduces the methodology that has been adopted in the present study in order to investigate empirically how the discourse marker *like* is used in different varieties of native English.

Chapter 3

Data and methodology

This chapter aims at introducing the methodology that has been adopted for the empirical part of this dissertation, as well as the data that will be used for the analyses. Section 3.1 gives an overview of the two corpora used in this study. Then, the software employed to extract the instances of *like* is briefly presented (Section 3.2). Section 3.3 deals with the disambiguation of the automatically retrieved instances of *like*. Section 3.4 and Section 3.5 present the methodology that has been applied to the data for the quantitative and qualitative analyses, respectively. Finally, Section 3.6 deals with the sociolinguistic part of the analyses.

3.1 The corpora

In the framework of this dissertation, resorting to spoken corpora turned out to be the most efficient way to compare the use of the discourse marker *like* in different varieties of English. As stated by Adolphs & Knight (2010: 38), “spoken corpora provide a unique resource for the exploration of naturally occurring discourse, and the growing interest in the development of spoken corpora is testament to the value they provide to a diverse number of research communities”. In this respect, spoken corpora have made it possible to study “less salient phenomena that used to be overlooked or whose importance used to be downplayed” (Gilquin & De Cock 2011: 8), to which discourse markers – and consequently *like* – obviously belong.

This dissertation investigates the use of *like* as a discourse marker across five varieties of English on the basis of two corpora: the International Corpus of English (ICE) and the Santa Barbara Corpus of Spoken American English (SBC). Both corpora contain material

from naturally occurring language use and the data that they contain are comparable. These databases are further described in the following two sections.

3.1.1 International Corpus of English

The ICE project was initiated in 1990 by Sidney Greenbaum. The corpus is a collection of spoken and written data from a number of Inner and Outer Circle varieties of English around the world (cf. Kachru 1991). So far, twenty-three subcorpora⁹ based on national or regional varieties of English have been compiled. To allow comparison between the different varieties, a common corpus-design was used. Furthermore, a majority of the subcorpora are POS-tagged, following a common annotation scheme. The present study focuses on four components of ICE, namely the British (ICE-GB), Canadian (ICE-CA), Irish (ICE-IR) and New Zealand (ICE-NZ) varieties of English.

Each subcorpus contains one million words (i.e. 500 texts of 2,000 words each), out of which 600,000 words are spoken data. The ‘spoken dialogue’ part of ICE includes both private discourse (direct conversations and telephone calls) and public discourse (e.g. class lessons, parliamentary debates and business transactions). Table 3.1 below represents the breakdown of this section into different genres together with their corresponding conversations:

PRIVATE	Direct conversations	S1A-001 to S1A-090
	Telephone calls	S1A-091 to S1A-100
PUBLIC	Class lessons	S1B-001 to S1B-020
	Broadcast discussions	S1B-021 to S1B-040
	Broadcast interviews	S1B-041 to S1B-050
	Parliamentary debates	S1B-051 to S1B-060
	Legal cross-examinations	S1B-061 to S1B-070
	Business transactions	S1B-071 to S1B-080

Table 3.1 Structure of the ‘spoken dialogue’ section of ICE

In this study, to examine the use of *like* in the context in which it is mostly likely to occur, namely face-to-face interaction (cf. Fuller 2003), the analysis of the four subcorpora cited above was restricted to data from the ‘direct conversations’ section of ICE, which consists in orthographic transcripts of 90 spoken conversations produced from the 1990s onwards. The

⁹ These represent English as used in: Australia, Cameroon, Canada, East Africa (Kenya, Malawi, Tanzania), Fiji, Great Britain, Hong Kong, India, Ireland, Jamaica, Kenya, Malta, Malaysia, New Zealand, Nigeria, Pakistan, Philippines, Sierra Leone, Singapore, South Africa, Sri Lanka, Trinidad and Tobago, USA.

speakers recorded in the various components of ICE are adults whose language of instruction has been English until at least the end of secondary school. Metadata are available for the Canadian subcorpus of ICE (ICE-CA).

3.1.2 Santa Barbara Corpus of Spoken American English

The fifth variety that is investigated in this dissertation is American English. To do so, data comparable to those of ICE have been accessed in the Santa Barbara Corpus of Spoken American English. This corpus contains transcripts of spoken conversations recorded all over the United States between 2000 and 2005. The SBC was designed in the same manner as ICE and follows similar annotation guidelines, thus allowing comparison between the two corpora.

This study is based on the first four parts of the SBC, which contain, in their entirety, a total of about 249,000 words. The corpus includes transcripts of various types of oral interaction, such as telephone conversations, card games, food preparation, classroom lectures, story-telling and town hall meetings. However, a large majority of the transcripts correspond to spontaneous face-to-face conversations. The analyses in this study were restricted to this last subtype of interaction, in line with the ‘direct conversations’ section in ICE. As a result, fifteen files out of sixty had to be discarded as they did not fit in this category. The speakers recorded in the SBC represent a wide variety of social and ethnic backgrounds, ages, genders, occupations, regional origins, etc. Metadata¹⁰ can be accessed together with the written transcriptions and their audio files.

Table 3.2 summarizes the number of conversations analyzed in this study and the corresponding number of words contained in each of the five subcorpora.

	SBC	ICE-GB	ICE-CA	ICE-IR	ICE-NZ
Number of conversations	45	90	90	90	90
Total number of words	189,293	163,899	188,051	186,146	210,125

Table 3.2 Number of conversations and total number of words in the five subcorpora

The total number of words per variety was calculated with the ‘Statistics’ function of *WordSmith Tools’ WordList*. *Like* was typed in as search word and the ‘Mark-up to ignore’ tab was set on <*> for the four components of ICE. For the Santa Barbara Corpus, however, the total number of words had to be calculated in several steps. For each speaking turn, the

¹⁰ Metadata in the SBC provide information on the speaker’s: name, gender, age, hometown, homestate, current state, education, years of education, occupation and ethnicity.

files contained the speaker's name and the number of the turn, which were left aside in the word count. In addition, the files also contained a number of annotations that provide comments on the interaction situation and which, therefore, do not belong to the transcription as such. This was the case for angle brackets and parentheses, whose content was excluded from the word count. Subsequently, the number of words corresponding to the following symbols was further subtracted: <@; @>; <%; %>; <BR; BR>; <F; F>; <FOOD; FOOD>; <HI; HI>; <L; L>; <MRC; MRC>; <P; P>; <PAR; PAR>; <Q; Q>; <SING; SING>; <SM; SM>; <VOX; VOX>; <WH; WH>; <WHISTLE; WHISTLE>; <X; X>; <YWN and YWN>. These successive operations resulted in the – somewhat approximate – final word count given for the SBC in Table 3.2 above.

3.2 The software

The data in this study were extracted and analyzed with *WordSmith Tools*, Version 6.0 (Scott 2014). This software is used by many researchers worldwide for the lexical analysis of corpus data. *WordSmith Tools*' *Concord* function turns out to be particularly interesting in that it provides a concordance of all the instances of a given search word, giving valuable information on its actual use. In this study, the *Concord* tool was used to automatically retrieve all the occurrences of *like* together with their context of use¹¹. However, the software displays all the instances of *like*, irrespective of the different functions that they fulfill depending on the context in which they are used. Therefore, a manual disambiguation of the data is a necessary step (cf. Section 3.3). In addition, the *WordList* function is particularly useful to automatically generate a list of the most frequent words or word-clusters in a given text. Finally, the *KeyWords* tool enables the user to find out which words characterize a text or a genre thanks to their unusually high frequency, compared to a reference corpus.

3.3 Disambiguation of the instances of *like*

Concordancing by means of a text retrieval software like *WordSmith Tools* is useful for data analysis in that it makes it possible “to find every occurrence of a particular word or phrase” (O’Keeffe 2009: 8). Here, the *Concord* tool of *WordSmith Tools* retrieved a concordance of all the instances of *like* in the five subcorpora under study. However, depending on the context in which it appears, *like* has, in addition to its discourse marker function, several

¹¹ A 200-character window was selected to provide the context around the targeted form *like* in each concordance line.

other, non-pragmatic functions, including those of quotative, verb, noun, adjective, adverb, preposition, conjunction, and part of a fixed expression (cf. Müller 2005).

In order to select only the cases in which *like* functions as a discourse marker – which constitutes the object of the present dissertation – a manual disambiguation of all the instances of *like* was carried out. Every concordance line was examined to check whether *like* in that context functioned as a discourse marker or not. To do so, I resorted to some of the defining criteria for an item to be considered a discourse marker such as weak clause association and non-truth-conditionality (cf. Section 1.4). In addition, testing the optionality of a given occurrence of *like* turned out to be a handy way to determine its grammatical nature, since, in its discourse marker use, *like* is necessarily non-obligatory. This resulted in a set of instances from which all non-discourse marker uses of *like* were discarded.

The discourse marker use of *like* is illustrated in examples such as (1) and (2):

- (1) they're set up so you can **like** learn the Jewish language <ICE-IR235>
- (2) they have a little board out at the front there it only had 1 **like** two or three sold out stickers on them <ICE-NZ369>

In a large number of cases in these data, however, *like* did not function as a discourse marker, but rather fulfilled one of its several other functions (cf. Müller 2005). As previously mentioned, such cases were discarded from the present study, since they do not meet – among other criteria – the ‘optionality’ requirement characteristic of discourse markers. For instance, the use of *like* as a verb is illustrated in example (3), while example (4) contains an occurrence of *like* as a preposition:

- (3) but uhm I don't know I I think I would **like** to go on as long as I feel that I'm enjoying it <ICE-GB18>
- (4) it's white and it has ah kind of a thing on the shoulders to make it look **like** a bride <ICE-CA1547>

Another usage of *like* worth mentioning because of its relative prominence is that of quotative, when preceded by a form of the verb *be*. This use was excluded from the analysis on the same grounds as the previous two examples. In line with the literature on the topic (cf. Section 2.2.3), this use of *like* turned out to be strikingly frequent in the speech of American speakers. Quotative *like* is illustrated in the following two examples, extracted from the Santa Barbara Corpus:

(5) We're on the way home, I was **like**, let's not take another detour.

She's **like**, Unh-unh. <SBC923-24>

(6) I was **like**, you know I'm **like**, we don't spend enough time together.

..And then I was all- -- I did that thi- -- Yeah I was j- **like**, you know all --

I'm **like** well well fine, all on Saturdays I wanna go out with my friends every Saturday then. With my -- The girls then.

He's **like**, that's fine. <SBC1212-16>

Note that in a number of ambiguous cases, it turned out to be difficult to determine whether a given instance of *like* was used as a discourse marker or not, because of insufficient context and/or fragmentary or overlapping utterances. Besides, slight differences appeared in the way the five subcorpora are annotated. In the ICE components, for instance, unintelligible utterances are recurrently transcribed according to the number of syllables or words which they are supposed to be made up of, without any clue regarding their meaning, making the disambiguation task even more challenging. Unclassifiable cases were left out of account for the analysis. In example (7), for instance, one can consider *like* either as a discourse marker focusing on *green*, or as the first occurrence of the preposition that introduces *a park*:

(7) It's a huge great expanse of green you know

It's very it's very big

I mean it's not like a

I mean is there is there much on it?

Is it just green **like** just like a park?

Well it's very hilly <ICE-GB37>

Table 3.3 below compares the total number of concordance lines initially generated by *Concord* with the number of instances of *like* that were kept for analysis after the stage of disambiguation.

	SBC	ICE-GB	ICE-CA	ICE-IR	ICE-NZ
Before	1888	910	1852	1774	1378
After	830	182	958	1077	535

Table 3.3 Number of instances of *like* before and after disambiguation

3.4 Quantitative data analysis

The first part of this study consists in the analysis of the frequency with which *like* occurs as a discourse marker in the five subcorpora investigated here. The aim is comparative, looking for potential differences in the extent to which the discourse marker is represented across the varieties. Since the subcorpora do not contain the same numbers of words, relative (or normalized) frequencies were used (i.e. expressed per 100,000 words), rather than absolute frequencies.

As a second step, the multigrammaticality of the word *like* was examined, with an analysis of the individual proportions between the discourse marker use and the non-discourse marker uses of *like* in each of the five varieties of English under study. The aim of this part is to gain an insight into the frequency of the discourse marker usage among all the other non-pragmatic functions that *like* can fulfill.

The study of the quantitative variation in the use of *like* as a discourse marker constitutes the object of Chapter 4. All the findings resulting from the quantitative analyses were statistically tested in order to determine whether the potential differences between the five subcorpora were significant or not. To do so, a chi-square test of independence¹² was applied to contiguous datasets, thus comparing the five varieties two by two. The chi-square test is frequently used in corpus linguistics as it makes it possible to “compar[e] the difference between the actual frequencies which have been observed in the corpus (the *observed* frequencies) and those which one would expect if no other factor than chance had been operating to affect the frequencies (the *expected* frequencies)” (McEnery & Wilson 2001: 84). A significance threshold of 0.05 was adopted for all the statistical analyses.

3.5 Qualitative data analysis

The second part of this study consists in a qualitative analysis of the position and the function of *like* as a discourse marker in the corpus data. The encoding of the data for these two parameters was carried out manually. Due to the relatively large number of concordances (cf. Table 3.2) and the extremely time-consuming nature of this task, a random sample¹³ of 100 instances of *like* as a discourse marker was selected for each of the five subcorpora. Note that

¹² The chi-square and the p-value were calculated with an online calculator (www.quantpsy.org/chisq/chisq.htm).

¹³ The =RAND() function in Microsoft Excel 2010 was used to generate these random samples.

in a number of cases, the stretch of discourse surrounding the target *like* did not provide sufficient context and looking back at the corpus files was necessary to be able to encode these two variables. In addition, some particularly unclear, incomplete and/or incomprehensible utterances were put in an ‘unclassified instances’ category when it was not possible – even with wider context – to determine their exact position and/or function.

3.5.1 Position

The position of the discourse marker *like* was coded in three possible ways, namely utterance-initial (8), utterance-medial (9) and utterance-final (10) position.

- (8) You know, does it have to be stuck in stone, that you have to go out with the boys on Friday?

... I said, cause I like to go out with a=ll of our friends you know.

Like a group of us? <SBC557>

- (9) It’s not even that tiny, I just hate the subject so much

I just sit here **like** scribbling the whole thing down you know

And it’s just everything is so repetitive, yet there are very very subtle differences in each sentence <ICE-IR506>

- (10) It means some just just what you said like

Oh it’s sort of a kind of a a sort of a a negative kind of a word in a way **like** <ICE-GB95>

The transcriptions were used as a clue for this distinction. Note, however, that the boundaries of the utterances and speaking turns were not necessarily signaled in the same way in all the five subcorpora (punctuation, new line, participant’s name, capital letters, etc.)¹⁴. As a result, some decisions regarding the position of *like* in an utterance were based on personal interpretation to some extent. The term ‘utterance’ is deliberately used here rather than ‘sentence’ since the scope of *like* can consist of one or several words only and, as a result, it does not necessarily coincide with a complete sentence. Besides, the utterance-initial position should not be taken in a strict sense as instances in which *like* was preceded by a pause or another discourse marker were still included in this category. This was the case in example (11), for instance, in which the discourse marker *well* does not affect the meaning of the utterance:

¹⁴ The original layout of the transcriptions was kept in all the examples included in this dissertation.

(11) Well you can't get batteries in the Soviet Union

No, that's right, but as she said

Well **like** I sold a cheapo ten pound Walkman and I'm ashamed of this <ICE-GB14>

In a number of cases, *like* occurred within an incomplete utterance: when the expected complement followed, even after a pause, the occurrence was coded as 'utterance-medial'. In example (12), for instance, *like* is followed by a pause – signaled by the word-wrap – but still interrupts the combination *copula* + *predicate*, and is therefore classified as 'utterance-medial':

(12) You know,

I mean,

he was **like**,

really off into wonderful abstract notions,

that were all based in a peaceful ... world. <SBC306>

On the other hand, if *like* occurred at the end of an incomplete utterance and the content of its scope was truly missing, the occurrence was classified as 'utterance-final'. Example (13) illustrates such a case:

(13) Yeah,

like if you're approaching them,

and their brights are on,

you just ... flick your brights on,

and then turn them off,

and then they're **like** <SBC451>

The fourth category of 'unclassified instances' cited above was added for ambiguous cases which could not possibly fit into any other category, as was the case for example (14):

(14) But I think I might do it um like for longer

like

cos we didn't get very brown really <ICE-NZ382>

3.5.2 Function

The coding of the function of the discourse marker *like* was based on Müller's (2005) taxonomy (cf. Section 2.4), that distinguishes between six main functions that *like* as a

discourse marker can fulfill in an utterance. The classification adopted here draws on these six functions, to which a seventh category was added for instances whose exact function could not be determined due to lack or ambiguity of context. In a number of cases, the impossibility to classify the function of *like* resulted from the fact that its position could not be accurately determined either. The seven possible options according to which the function of the discourse marker *like* was coded are summarized in Table 3.4 below:

1. Searching for the appropriate expression [Search]
2. Marking an approximate number or quantity [Number/quantity]
3. Introducing an example [Example]
4. Introducing an explanation [Explanation]
5. Marking lexical focus [Focus]
6. Discourse link and restart [Restart]
7. Unclassified instances [Unclassified]

Table 3.4 Functions of the discourse marker *like* (based on Müller 2005)

Note that the multifunctionality of the discourse marker *like* complicated the classification of the variable. In the samples analyzed here, single occurrences of *like* regularly fulfilled more than one function within one and the same utterance. Therefore, only the dominant function of each occurrence was taken into account for the coding. In example (15), for instance, *like* both introduces an example of “those Indians who speak better English than the anglophone Canadians” and expresses hesitation. However, one can assume that the speaker’s initial intention was to exemplify their previous assertion, so the occurrence of *like* was classified as such:

- (15) Like a couple of the Indians spoke better English than any of the anglophone Canadians in the group
Like like hew one fella fellow was the president of his college debating society back home and
 he could see through every nuance that you put to every sentence <ICE-CA795>

Regarding the use of *like* as marking an approximate number or quantity, Meehan's (1991) inclusive approach was adopted here (cf. Section 2.4.3). As a result, expressions denoting time periods (16) and proportions (17), for instance, were included in this category as well:

(16) But I'm I'm finding you know it's, I don't know it's just so rushed

You know you have **like** about a week to prepare something together in a group
<ICE-CA414>

(17) Do you guys each want **like**, ... half of that?

Sure

Or will you eat more <SBC75>

3.6 Sociolinguistic analysis

The third and final part of this study consists in the analysis of two sociolinguistic factors that influence the use of the discourse marker *like*, namely speaker's gender and age. These appear to strongly impact on the frequency with which members of an English-speaking community use *like*, resulting in an unequal distribution of the discourse marker in a given population (cf. Section 2.5). In this respect, note that the study of corpus data is useful in that it gives an insight into some linguistic phenomena and trends shared by the corpus as a whole, but it tends to overlook the fact that "corpora are inherently variable internally" (Gries 2006: 110). To account for this internal variation, investigating individual speakers' use of *like* turns out to be necessary and will constitute a part of this sociolinguistic analysis, in addition to the study of more general frequencies at the corpus level.

Sociolinguistic data could only be accessed for the Santa Barbara Corpus of Spoken American English and the Canadian subcorpus of ICE. Therefore, the comparison made in Chapter 7 will be restricted to those two varieties of English. To do so, a sample¹⁵ of 300 instances of *like* as a discourse marker was randomly selected for each of the two subcorpora under study. The coding process of the variables is detailed in the following two sections.

¹⁵ The =RAND() function in Microsoft Excel 2010 was used to generate these random samples.

3.6.1 Gender

The gender variable is a binary one. It was coded as ‘f’ for female speakers and as ‘m’ for male speakers. To do so, it was necessary to identify the speaker by whom a given instance of *like* was produced by looking back at the transcriptions. The speaker’s gender could then be accessed in the corresponding metadata. The findings resulting from these analyses were subsequently tested with a chi-square test of independence¹⁶ in order to determine whether the potential differences between the American and the Canadian samples were statistically significant or not. Here again, a significance threshold of 0.05 was adopted for the analyses.

3.6.2 Age

To code this variable, age classes were used. Speaker’s age in the metadata from ICE-CA was already classified according to eight classes, namely under 10, 10-18, 19-24, 25-30, 31-40, 41-50, 51-60 and 61+. The SBC, on the other hand, used exact numbers to code the age of its speakers. These were thus recoded for each speaker included in the sample according to the age groups mentioned above. Since neither ICE-CA nor the SBC contained instances of speakers under ten years old, this category was put aside for the analysis.

This chapter first introduced the data that were used for the empirical part of this dissertation through a brief description of the five subcorpora, the software and the disambiguation process. Then, the methodology that was adopted for the quantitative, qualitative and sociolinguistic analyses in this study was presented. The following chapter inaugurates the empirical part of this dissertation by looking into the frequency of use of the discourse marker *like* across the five varieties of English under study.

¹⁶ The chi-square and the p-value were calculated with an online calculator (www.quantpsy.org/chisq/chisq.htm).

Chapter 4

Frequency of *like*

As previously mentioned, research on discourse markers as a linguistic class has become increasingly prominent since the late twentieth century. Interestingly, their use is not equally distributed across speech contexts, and spontaneous conversations appear to strongly favor their presence (Fuller 2003). As a result, *like* displays a remarkably high frequency of use in spoken conversations (cf. Tagliamonte 2005), as can be observed in this particularly striking example provided by Siegel (2002: 3): “She isn’t, **like**, crazy or anything, but her and her, **like**, five buddies did, **like**, paint their hair a really fake-looking, **like**, purple color”. In their study of bilingual Canadians, Sankoff et al. (1997) found out that *like* ranks second among the most frequently used discourse markers in the speech of their participants. Fuller (2003) obtained similar results for American native speakers of English, with *like* displaying by far the highest frequency of use in conversational data.

The frequency with which the discourse marker *like* occurs constitutes the object of this chapter, which is the first of the empirical part in this dissertation. It looks into the frequency of use of *like* in its discourse marker usage in subcorpora representing American, British, Canadian, Irish and New Zealand English. Section 4.1 compares the frequency with which the discourse marker *like* occurs in the five subcorpora. Section 4.2 analyzes the proportion of discourse marker uses out of the total number of instances of *like* in each of the five varieties. The findings resulting from these two sections are further discussed in Section 4.3.

4.1 Overall frequency of *like*

Table 4.1 below presents the number of occurrences of the discourse marker *like* in each of the five subcorpora, together with the normalized frequencies per 100,000 to allow comparison.

	Raw frequency	Relative frequency (per 100,000 words)
SBC	830	438.47
ICE-GB	182	111.04
ICE-CA	958	509.44
ICE-IR	1077	578.58
ICE-NZ	535	254.61

Table 4.1 Raw and relative frequencies of *like* as a discourse marker in the five subcorpora

The variety that scores highest regarding the frequency of the discourse marker *like* is Irish English, with 1077 occurrences in the section of ICE-IR analyzed, which amounts to a relative frequency of 578.58 occurrences per 100,000 words. It is followed in decreasing order by the Canadian, American and New Zealand subcorpora. Finally, British speakers use the discourse marker *like* strikingly less often, with a relative frequency as low as 111.04 occurrences per 100,000 words, i.e. over five times less than Irish speakers. In this respect, it is important to note that the five subcorpora were not compiled at the same time. The earliest data come from ICE-GB, which can partly explain the low frequency of *like* in that subcorpus, since the boom in the use of the discourse marker is a relatively recent phenomenon. The inter-group differences are statistically highly significant ($p < 0.05$)¹⁷. These results are represented visually on Figure 4.1.

¹⁷ ICE-IR vs. ICE-CA: $\chi^2 = 8.179$, $p = 0.0042$; ICE-CA vs. SBC: $\chi^2 = 9.978$, $p = 0.0016$; SBC vs. ICE-NZ: $\chi^2 = 98.163$, $p = 0$; ICE-NZ vs. ICE-GB: $\chi^2 = 98.829$, $p = 0$.

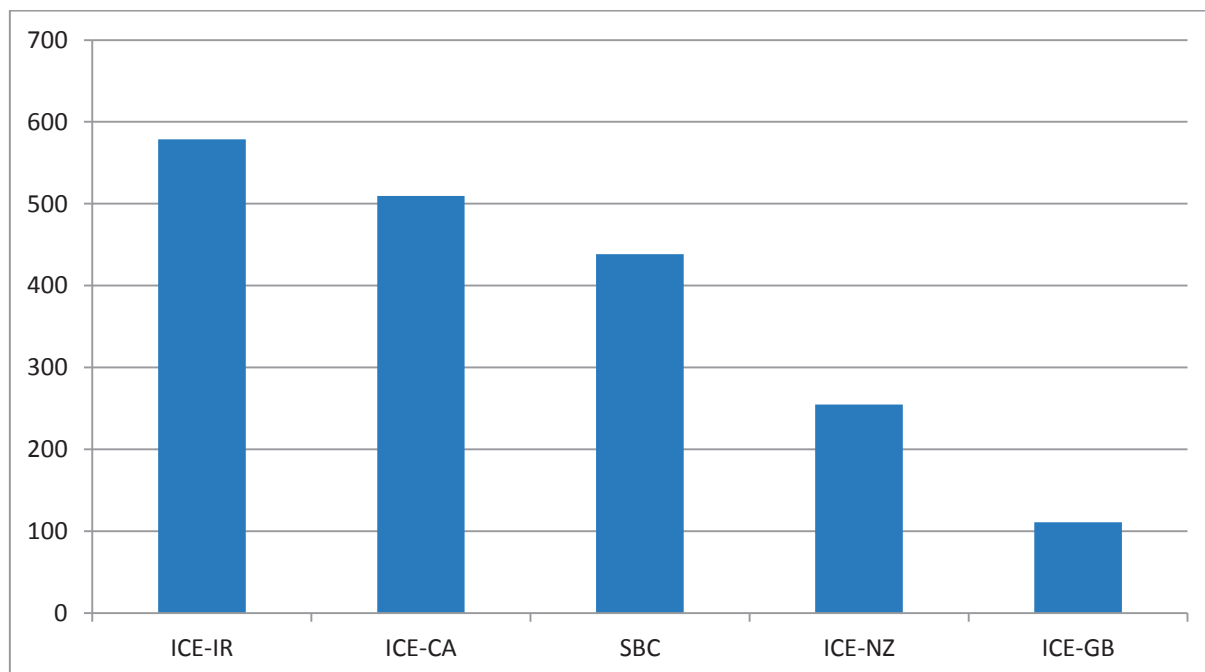


Figure 4.1 Relative frequency of the discourse marker *like* in the five subcorpora

4.2 Proportion of discourse marker uses

The ratio between the number of cases in which *like* functions as a discourse marker and those in which it fulfills a non-pragmatic function differs between the five varieties of English as well. These differences are statistically significant ($p < 0.05$)¹⁸. Figure 4.2 shows this variation of proportion between the five subcorpora.

¹⁸ ICE-IR vs. ICE-CA: $\chi^2 = 29.687$, $p = 5e^{-8}$; ICE-CA vs. SBC: $\chi^2 = 22.598$, $p = 0.000002$; SBC vs. ICE-NZ: $\chi^2 = 8.643$, $p = 0.0033$; ICE-NZ vs. ICE-GB: $\chi^2 = 90.259$, $p = 0$.

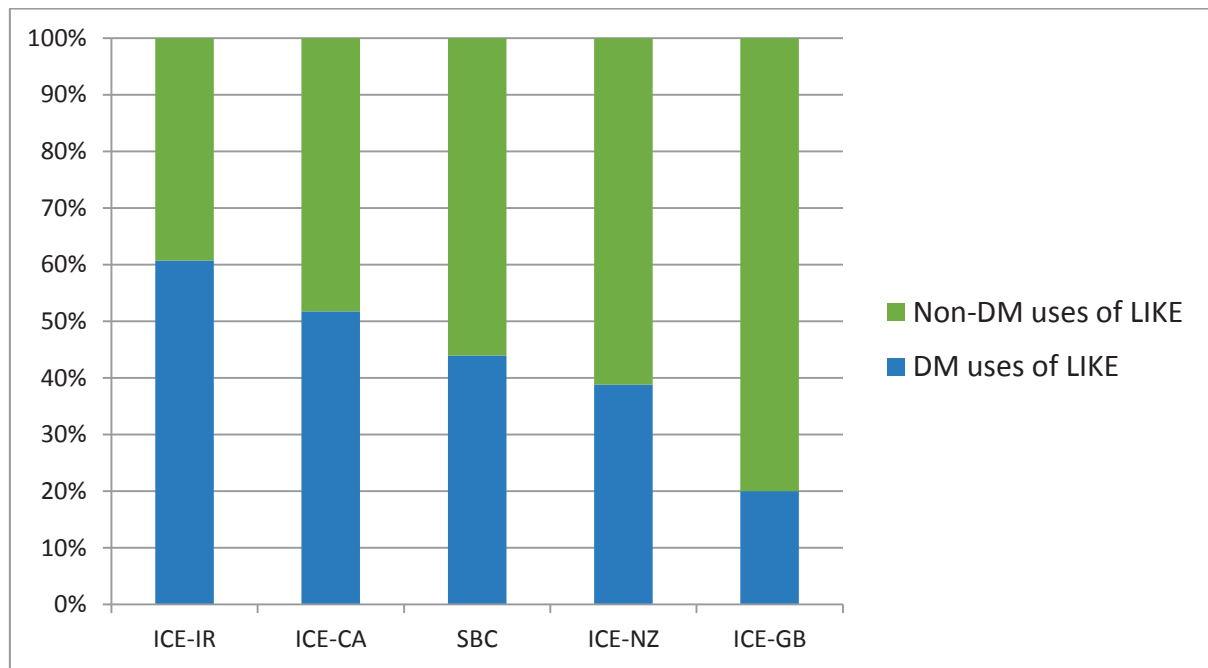


Figure 4.2 Proportion of the discourse marker uses of *like* in the five subcorpora

The proportion of discourse marker uses of *like* among all the different functions that it can fulfill appears to follow a similar pattern to Figure 4.1. Here again, Irish English comes in first position, with more than 60% of its occurrences fulfilling a discourse marker function (60.71%). In ICE-CA, about half of the occurrences of *like* function as a discourse marker (51.73%). The American and the New Zealand varieties have a ratio of 43.96% and 38.82% respectively. Finally, the discourse marker use of *like* by British speakers only represents 20% of the total number of occurrences of *like*. This further emphasizes the multigrammaticality of the word *like*, which, next to its discourse marker usage, also fulfills a whole range of non-pragmatic functions (cf. Section 3.3).

4.3 Discussion

These quantitative analyses have revealed that Irish English stands out as the variety in which *like* as a discourse marker occurs most frequently, both in terms of relative frequency and proportionally to other, non-pragmatic uses of the form. Unsurprisingly, the phenomenon is relatively prominent in the two North American varieties as well (Canadian and American English). This corroborates the tendencies that emerged from previous studies (cf. Fuller 2003; Tagliamonte 2005). By contrast, *like* as a discourse marker is much less frequent in ICE-GB and only accounts for a very small proportion of all the occurrences of the form *like*. All in all, the distribution of the discourse marker *like* among these data illustrates D'Arcy's

(2005) distinction between the North American and the British usage of the discourse marker (cf. Section 2.2.2).

This chapter investigated the frequency of use of *like* as a discourse marker in the five subcorpora, as well as the proportion that this usage of the form represents among all its other possible functions. The following chapter looks into the various positions in which the discourse marker *like* can be located within an utterance, and illustrates these with examples from the data.

Chapter 5

Position of *like*

The literature on the position of *like* has revealed a great deal of variation in the way the discourse marker behaves with respect to this variable. If, overall, the utterance-initial position has turned out to be the most frequent one (cf. Underhill 1988 for American English; Tagliamonte 2005 and Tagliamonte & D’Arcy 2009 for Canadian English), other positions are possible as well (cf. Romaine & Lange 1991 for American English). Besides, the position of *like* as a discourse marker appears to be influenced by the nature and the scope of the constituent it modifies, which, in turn, can vary to a great extent.

This chapter investigates and compares the position of the discourse marker *like* in the American, British, Canadian, Irish and New Zealand varieties of English represented in this dissertation. Each of the first five sections (Section 5.1 to Section 5.5) looks into an individual variety, describing and illustrating the distribution of the position of *like* in one hundred of its occurrences selected from the corresponding subcorpus. The findings resulting from these analyses are further discussed and compared in Section 5.6.

5.1 American English

The position of *like* in American English was examined on the basis of the random sample from the Santa Barbara Corpus of Spoken American English. The distribution of the three possible variants is represented on Figure 5.1 below.

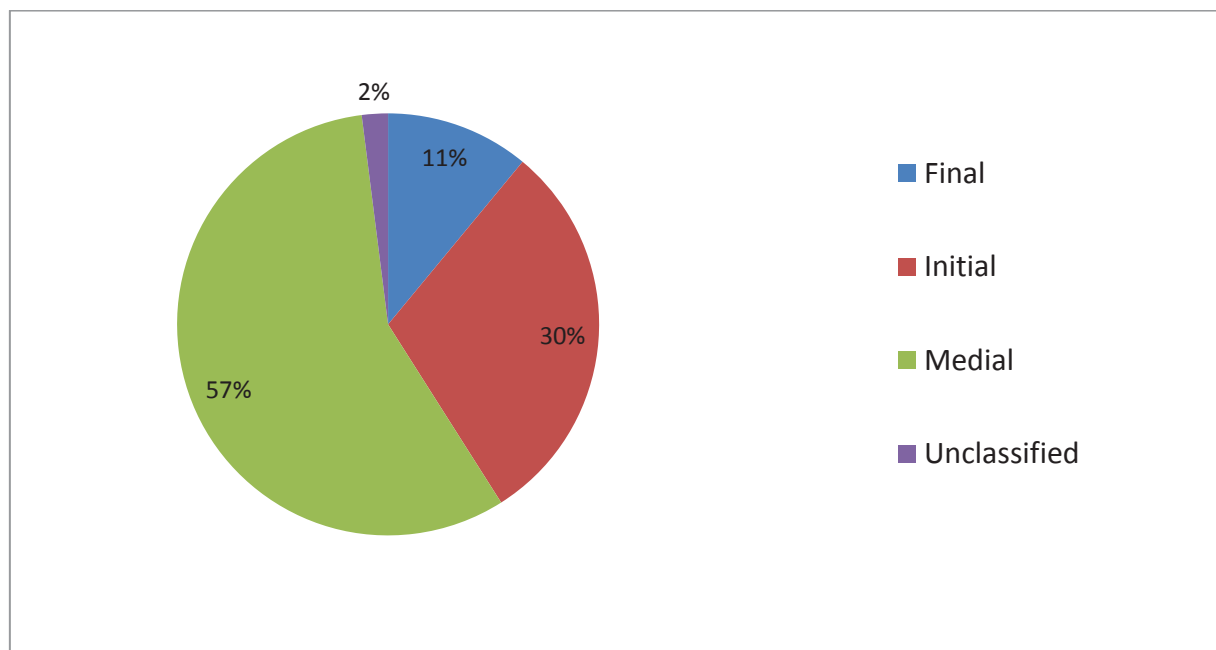


Figure 5.1 Distribution of the different positions of the discourse marker *like* in the SBC

As can be observed on this figure, the most frequent position of *like* as a discourse marker is by far the utterance-medial one with 57% of the instances. Within this category, *like* is often located between a transitive verb and its direct object or between an auxiliary and the following lexical verb. These two cases are illustrated in examples (18) and (19), in which *like* modifies a noun phrase and a verb phrase respectively:

(18) And I think, I think uh, you-, you picked up some virus, like, uh, he named **like** half a dozen viruses, and said, Ah, I could do some more tests, and find out which one, but, by the time we got the test ... results back, you'd probably be over with it anyway, so why bother. <SBC243>

(19) I would take it home, ... and I would **like** put it in, I would leave it in its ... plastic bag, put it in a bunch of water, in the sink. <SBC275>

The second most frequent position in which *like* occurs is utterance-initial, in almost one third of the instances (30%). The following two examples illustrate this position of the discourse marker. Note that in (13), *like* was still considered initial although it is preceded by the discourse marker *so* (cf. Section 3.5.1).

(20) Hard times do train you.

Yep.

They do.

Like I came over here to work with Danae, which is what I'm going to do, I'm going to do some ... translations for her and stuff <SBC159>

(21) And I guess like in Minnesota it's real wet, ... and stuff you know?

So **like**, ... they really have to watch their shoes, for not coming off, because the hoof wall, is so much softer <SBC23>

In both examples, the scope of *like* is the entire proposition to the right of the discourse marker.

In 11% of the instances, *like* occurs in utterance-final position. In most cases, this is due to an incomplete utterance, as in this example:

(22) You don't have to spend money on goodies.

You can have **like**,

You can make these little, um,

Do they have a laminating machine in the school? <SBC121>

Finally, two cases were not clear enough to be coded on the position of *like* and remained unclassified.

5.2 British English

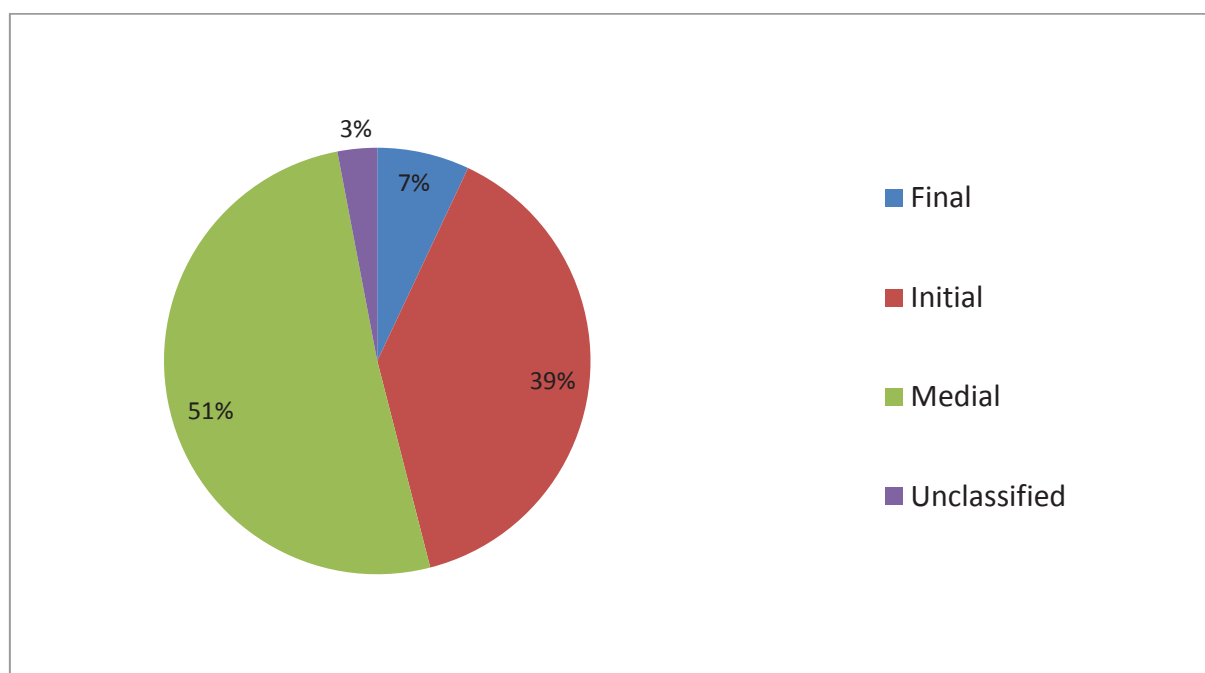


Figure 5.2 Distribution of the different positions of the discourse marker *like* in ICE-GB

As can be observed on Figure 5.2, the three possible functions of *like* as a discourse marker are not equally distributed in British English either. In the sample extracted from ICE-GB, about half of the instances (51%) occur in utterance-medial position. In those cases *like* can modify segments from variable length and nature, including noun, verb, adverbial and prepositional phrases, among others. In example (23), for instance, *like* modifies the predicate *worried* introduced by the copula *were*. In (24), the scope of *like* is the prepositional phrase *up to here*:

(23) And I thought that would be just,
 You were **like** really worried on the phone,
 And I thought oh no <ICE-GB83>

(24) It's very cute,
 It's got a grin that's sort of **like** up to here I mean,
 It says mmm like this <ICE-GB79>

In 39% of the instances, *like* was classified as utterance-initial. This includes cases in which *like* is truly the first word of the utterance, as in (25), as well as cases in which it is preceded by another discourse marker. Most of the time, this is *so* or *well*. In example (26), *like* comes after *I mean*:

(25) But now that we know how we've each come to the conclusion as to what we've done,
Like what's your conversation like
 Yes
 Uh my conversation was going in on Saturday morning into a shop where I was picking up some things that were due to be framed <ICE-GB133>

(26) They're there 're there 're all kinds of bits where somebody's got the words for the others
 I mean **like** you've got them to start Tim
 and we haven't
 So we're like accompanying there
 And then Roger's got a little bit and Gavin's got a bit further on a and so on <ICE-GB41>

Interestingly, *like* appears to occur relatively frequently in combination with the modifier *sort/kind of* which further reinforces the mitigating function of the discourse marker.

This sequence occurs both when *like* is located in utterance-initial position and in utterance-final position. In this data, 11 occurrences of *sort/kind of + like* could be observed, among which the following two examples:

(27) When I was at primary school uhm I really didn't like going to school and I used to pretend that I'd be sick

As I remember it used to be *sort of like* fairly common for a Tuesday, that I'd pretend to be sick, and so I didn't have to go to school <ICE-GB154>

(28) I kind of, I mean, I I started the course, thinking that uhm I'd sort of do the full seven years and stuff

But like I'm just going through the course

I just just realised that it was actually the st study of architecture I really enjoyed

And uh you just *kind of like* get a a few hints at what actually working in the profession's like

And I didn't really like the idea of it <ICE-GB63>

British speakers only used *like* in utterance-final position in 7% of the cases, usually due to incomplete utterances, as is illustrated by example (29) in which the expected final verb is missing:

(29) I'm therefore put into a position to try and bring out some new words and to express myself well and start speaking you know be to the point

Yeah

I always used to **like**

I remember sitting I was in here once and there were some people talking about, what was that film, Dead Poets Society, by Robin Williams <ICE-GB173>

Finally, three instances of *like* were labelled 'unclassified', because unclear segmentation of the utterances and the speaking turns in the transcriptions did not make any other categorization possible.

5.3 Canadian English

In the sample from the Canadian subcorpus of ICE, every instance of the discourse marker *like* could be classified regarding its position within the utterance. The distribution of the three possible positions is very similar to that of ICE-GB. It is represented on Figure 5.3 below.

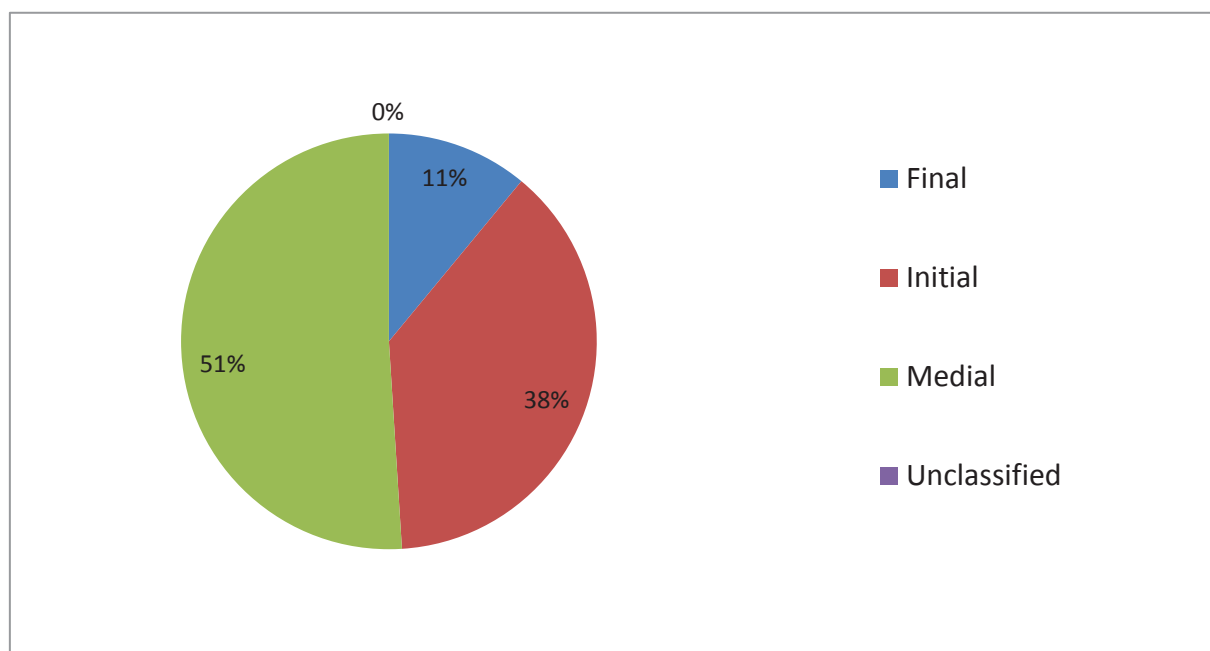


Figure 5.3 Distribution of the different positions of the discourse marker *like* in ICE-CA

Here again, the most frequent position of the discourse marker *like* is the utterance-medial one with 51% of the instances, i.e. exactly the same percentage as in the British subcorpus. Within this category, *like* modifies a whole range of constituents, such as verb phrases as in (30) and adverbs as in (31). Not that in about one fifth of the utterance-medial cases, the scope of *like* is a numerical or quantitative expression, as illustrated by example (32).

(30) And we had planned to go to California the year before and I **like** put the brakes on
it sort of at the last minute
Oh ya
Just decided that either time money everything it was just not the right time for me
<ICE-CA504>

(31) Oh my god

Let's go there **like like like**¹⁹ soon

Like when can we go there <ICE-CA794>

(32) So in other words let's assume your return on investment is gonna going to be twenty percent in the business when all is said and done

So if you were to put it in the bank you might get **like** eight percent <ICE-CA360>

In 38% of the cases, *like* occurs in utterance-initial position. When it follows another discourse marker, it is most of the time *you know*, as in example (33):

(33) Do you remember that from when you were a kid

You know **like** getting into your snowsuit after school

And just so you yank the thing and rip the zippers open and slide into it <ICE-CA740>

Finally, only 11% of the instances occur in utterance-final position. Here again, *like* mostly has a forward scope, but the utterance within which it occurs is incomplete. In such cases, the categorization of an instance as utterance-final was based on the graphic segmentation in the transcriptions, which, we may assume, represents prosodic breaks, but, from a semantic point of view, *like* could also be considered as the beginning of the next utterance. In example (34), for instance, *like* was considered as the last word of the first utterance, whose expected verbal complement is missing. However, it is reasonable to believe that it could have belonged to the next utterance as well.

(34) Cos if you just put sugar we'll just **like**

you'll taste the granules of sugar on it <ICE-CA629>

5.4 Irish English

The coding of the sample for the position of *like* as a discourse marker in Irish English reveals a markedly different distribution. This can be observed on Figure 5.4 below.

¹⁹ Note that this sample contains five cases of repetitions of *like*, mostly as part of the speaker's hesitation and restart.

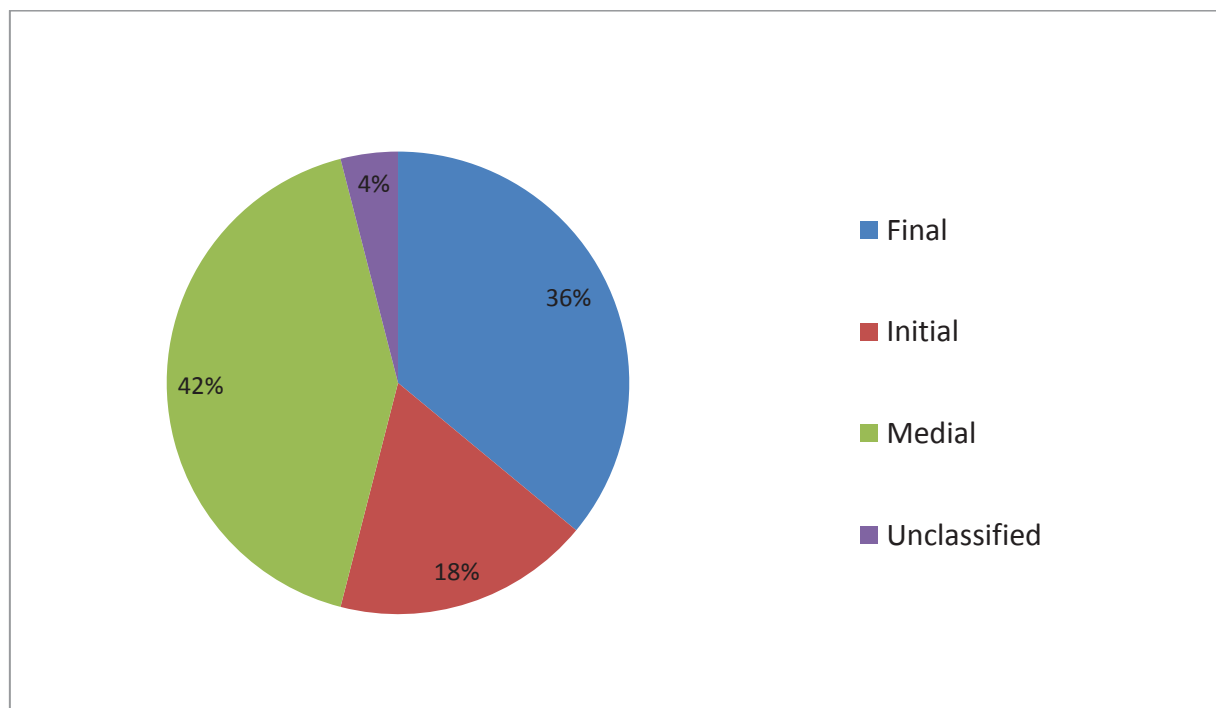


Figure 5.4 Distribution of the different positions of the discourse marker *like* in ICE-IR

While the utterance-medial position is still the most frequent one, its percentage is much lower than in the three subcorpora previously described (42%). In these instances, *like* modifies various constituents, including verb and noun phrases, adjectives, adverbs and prepositional phrases. These last two types are illustrated in examples (35) and (36) respectively:

(35) I take it you didn't go for a drink or anything afterwards

No we, we disappeared **like** directly afterwards cos Tom and Saoirse

Yeah we went to the Errigle <ICE-IR318>

(36) So they got a search warrant to search the house

But what made them go to the house in the first place

Oh they knew something was up

They had reports **like** on beating young boys and things <ICE-IR987>

The second most frequent position for the discourse marker *like* in ICE-IR is utterance-final, with 36% of the instances. Interestingly, *like* appears to behave differently here than in the other varieties of English previously analyzed. In a majority of cases within this position, *like* does not occur at the end of an incomplete utterance. Rather, it has a backward scope, modifying the constituent to its left. This phenomenon is illustrated in the following examples:

(37) Yeah God I remember serving her one day

She's a cow **like**

And I got everything wrong and I forgot to bring stuff down to her <ICE-IR291>

(38) One of those hours I was lifeguarding so I got three pounds an hour

That's grand **like**

But teaching is better you know

Passes the time much more quickly <ICE-IR450>

(39) No no young women'd knit now you know

No no. And it's no often you get stuff knitted for weans when they're born **like**

So, there's there I'll have to show you there's a beautiful young, a a girl Jean Cooper knitted from Ballyhalbert <ICE-IR189>

Eighteen instances of *like* as a discourse marker were classified as utterance-initial, out of which about half occur after another discourse marker. In example (40) below, *like* is the fourth discourse marker in a row. All of them are optional, hence the classification of this instance as utterance-initial:

(40) But he was doing, I think he was lecturing on this kind of thing and he was doing

See how it could be like uh Western conflict resolution theories

How they would integrate into Serbian culture

No not, yes yes, Eastern European sort of yeah

Yeah so I mean **like** I suppose you could get direction in what type, okay you would need a certain, sample of conflict resolution theories <ICE-IR760>

Finally, four instances in the sample remained unclassified because they were unclear or ambiguous.

5.5 New Zealand English

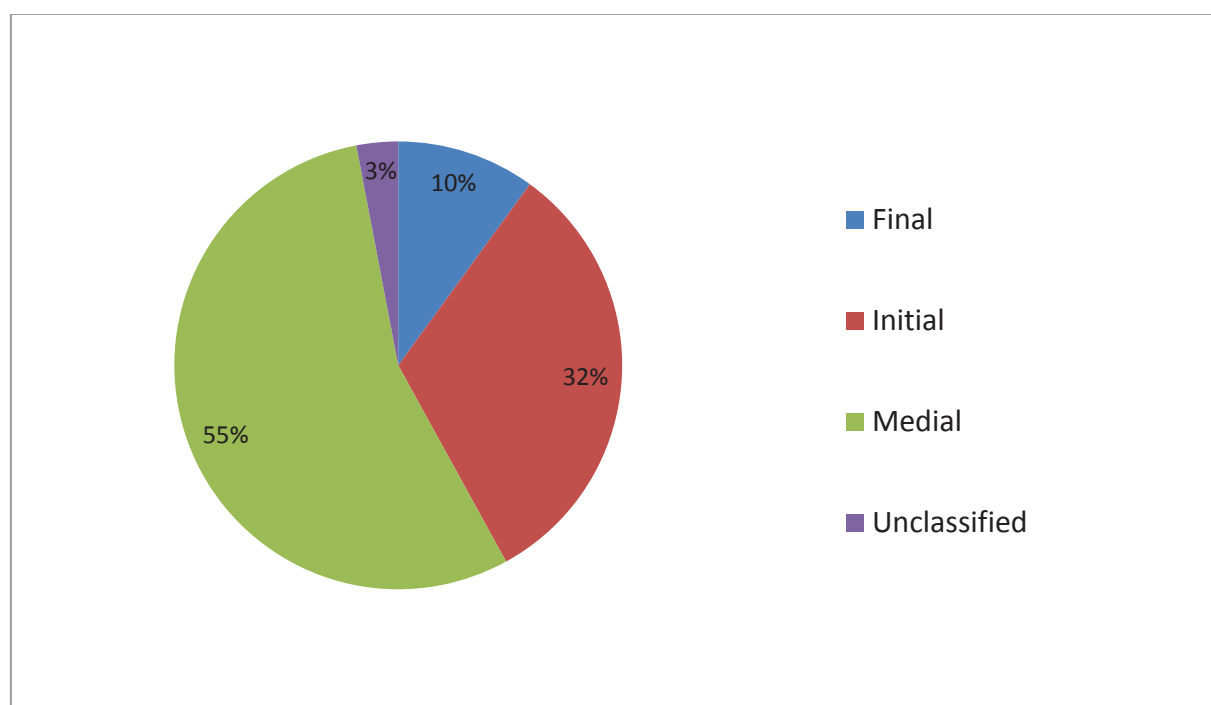


Figure 5.5 Distribution of the different positions of the discourse marker *like* in ICE-NZ

As can be observed on Figure 5.5, the position of *like* as a discourse marker that ranks first in terms of frequency is the utterance-medial one with 55%. In a similar way to ICE-IR, the most frequent constituents modified by the discourse marker *like* are noun phrases as in example (41) and adjectives as in example (42):

(41) They, they have special bins on the street like you know for all your
cos in Vienna and stuff because they're all in apartments rather than houses and
they don't have gardens
they have **like** compost bins on the street where you put all your your food waste
and everything <ICE-NZ203>

(42) Anyway er we walked and there's like the bush is **like** really all clean underneath
and the trees the trees are like scratched with these bloody possum traps
<ICE-NZ249>

When it comes to the utterance-initial position of *like* as a discourse marker, it represents 32% of the instances in the sample from ICE-NZ. Contrary to the other varieties of English this sample barely contains any instance of *like* preceded by another discourse marker. In example (43), *like* is the first word of the utterance. Example (44), on the other

hand, is one of the rare cases in which *like* follows another discourse marker. Note that the use of *I mean* appears to be particularly recurrent in this speaker's discourse:

(43) There's just an awful lot of people skiing there

I can handle people, people

Like the bits where they have the the the rocks,

like on the lefthand side of um, knoll ridge tee bar

Like I'd ski down there on a fine day but I'd never ski down there in white white
outs <ICE-NZ512>

(44) I mean he's always had a job all his life and that's all he's got to show for it

And then when he retires he'll get a gold watch

Gold gold ...

And he's a skilled, skilled tradesman as well

I mean he's a he's a builder,

but I mean he's happy eh

I mean **like** he could go out and get bigger jobs or whatever but

I mean he's happy doing what he wants to do I suppose <ICE-NZ399>

Ten instances of the discourse marker *like* occupy the last slot of an incomplete utterance and were therefore classified as utterance-final, as in the following example:

(45) That, that's the same as um Darren he's probably going to work in Waiuku

But I don't want to live in Waiuku

The nearest town is **like**

Oh God

Even to Auckland it's an hour and a half and that's the closest city <ICE-NZ236>

Finally, 3% of the instances were labelled 'unclassified' because the unclear segmentation in the transcription did not make it possible to determine their exact position within the utterance.

5.6 Discussion

The results emerging from the analysis of the position of the discourse marker *like* across the five subcorpora have revealed a common pattern shared by the American, British, Canadian

and New Zealand varieties of English. By contrast, Irish English displays a quite different distribution of the three possible positions of *like* as a discourse marker.

On the one hand, the first four varieties mentioned above show a clear tendency towards the utterance-medial position with more than half of the samples belonging to this category. Interestingly, when occupying that position, *like* very often interrupts close-knit structures, contrary to what Gilquin & Granger (2015) found out for the use of two-word discourse markers by native speakers of English. The present findings also contradict the strong propensity for *like* to occur in utterance-initial position that was suggested by several studies (cf. Underhill 1988 for American English; Tagliamonte 2005 for Canadian English). When utterance-medial, however, *like* mostly modifies noun and verb phrases, in line with the literature (cf. Underhill 1988; Tagliamonte & D’Arcy 2009). To a lesser extent, its scope can also consist of adjectives, adverbs and prepositional phrases. In these four subcorpora, the utterance-initial position is the second most frequent one, with a proportion ranging between 30% and 39%. In that position and to varying degrees across the four samples, *like* sometimes occurs together with another discourse marker. An interesting observation to mention is the tendency for *like* in the British subcorpus of ICE to occur in combination with *sort/kind of*. Finally, in the four varieties, the utterance-final position of *like* only accounts for about 10% of the cases, mostly as part of an incomplete utterance.

On the other hand, the distribution of the different positions of the discourse marker *like* in Irish English stands apart from the four varieties described above. While the utterance-medial position is also the most frequent one in ICE-IR, its frequency is somewhat lower. In addition, by contrast to the previous findings, the utterance-final position ranks second in the Irish subcorpus, with over one third of the instances. Another difference concerns the scope of *like* in that position, namely backward rather than forward. Finally, the utterance-initial position displays a much lower frequency in ICE-IR than in the four other subcorpora, with less than one fifth of the instances classified as such.

This chapter investigated the position in which the discourse marker *like* occurs across the five varieties of English under study in this dissertation. The following chapter looks into the various functions that *like* can fulfill as a discourse marker, trying to point out the differences and the similarities in its use across the five subcorpora.

Chapter 6

Function of *like*

Different accounts of the pragmatic functions fulfilled by *like* as a discourse marker have been put forward in the literature, every author generally bringing out one of them over the others (cf. Underhill 1988; Andersen 1998; Dailey-O’Cain 2000). Overall, what has commonly been acknowledged is the particularly developed multifunctionality of the discourse marker *like*. In the present study, the six main functions of the discourse marker *like* in Müller’s (2005) classification will be looked into, namely hesitation, quantitative expression, example, explanation, lexical focus and restart.

This second chapter of the qualitative analysis investigates and compares how the functions cited above are represented across the American, British, Canadian, Irish and New Zealand varieties of English studied in this dissertation. Each of the first five sections (Section 6.1 to Section 6.5) focuses on an individual variety, with an analysis of the distribution of the different functions of *like* as a discourse marker within one single subcorpus. Section 6.6 further discusses the findings emerging from these analyses.

6.1 American English

One hundred occurrences of the discourse marker *like* extracted from the Santa Barbara Corpus of Spoken American English were examined in order to gain an insight into the function(s) fulfilled by the discourse marker in American English. The distribution of the different functions can be observed on Figure 6.1 below:

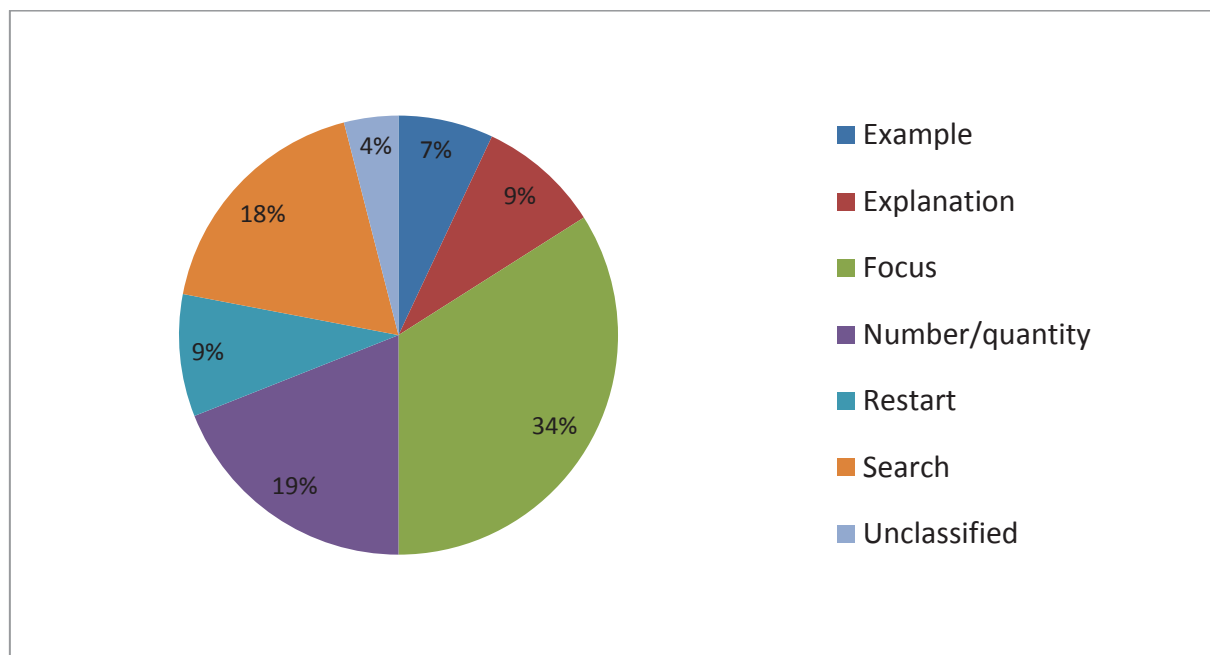


Figure 6.1 Distribution of the different functions of the discourse marker *like* in the SBC

The focusing function of *like* clearly stands out on this graph with about one third of the occurrences (34%). Interestingly, a strong link can be observed between this function and the position of the discourse marker within the utterance: 29 cases out of the 34 with a focusing function occurred in utterance-medial position, mostly before verbs (46), noun phrases (47) and adjectives (48):

(46) And then he goes I'm gonna give you five at a time,
and you'll have to like perform for it?
Nancy and I were like, woah.
And she thought it was so funny.
She was **like** cracking up. <SBC757>

(47) There's **like** a whole strip.
That's like that whole strip right there.
There's all the—
the classic type of prostitutes. <SBC635>

(48) Like he has a lot of money but he's, like, like
I don't know, like Blane, he's **like** kind of rude.
Isn't he? <SBC733>

In 19% of the cases in this sample, *like* occurred in combination with a numerical or quantitative expression, as is illustrated in example (49), in which an approximation of the duration does not hinder the comprehension of the utterance:

- (49) And on the last day, I kept th- four kids after school.
... Yeah.
... You know, but you can only actually keep them over, for **like** five minutes.
Cause they have like buses to catch and stuff <SBC119>

The third most frequent function of *like* is searching for the appropriate expression with 18% of the occurrences. This use can be observed in the following example, in which the speaker's hesitation is also marked by unfilled pauses, false starts and repetitions:

- (50) I know the Caribbean in incredible
beau- -- ... beautiful beautiful blue --, blue water,
and and, warm water, and like ... coral, and tropical fish,
and incredible, r- like resort, ... **like** um, ... hotels, and restaurants <SBC236>

The functions of restart and explanation are both represented by 9% of the instances, followed by that of example in 7% of the cases. Finally, the function of four occurrences of *like* could not be determined because the discourse marker occurred in an incomplete utterance.

6.2 British English

The function of the discourse marker *like* in British English was examined on the basis of 100 occurrences from the British subcorpus of ICE. The various functions that *like* fulfills appear to be represented in varying proportions in the sample. Their distribution can be observed on Figure 6.2 below:

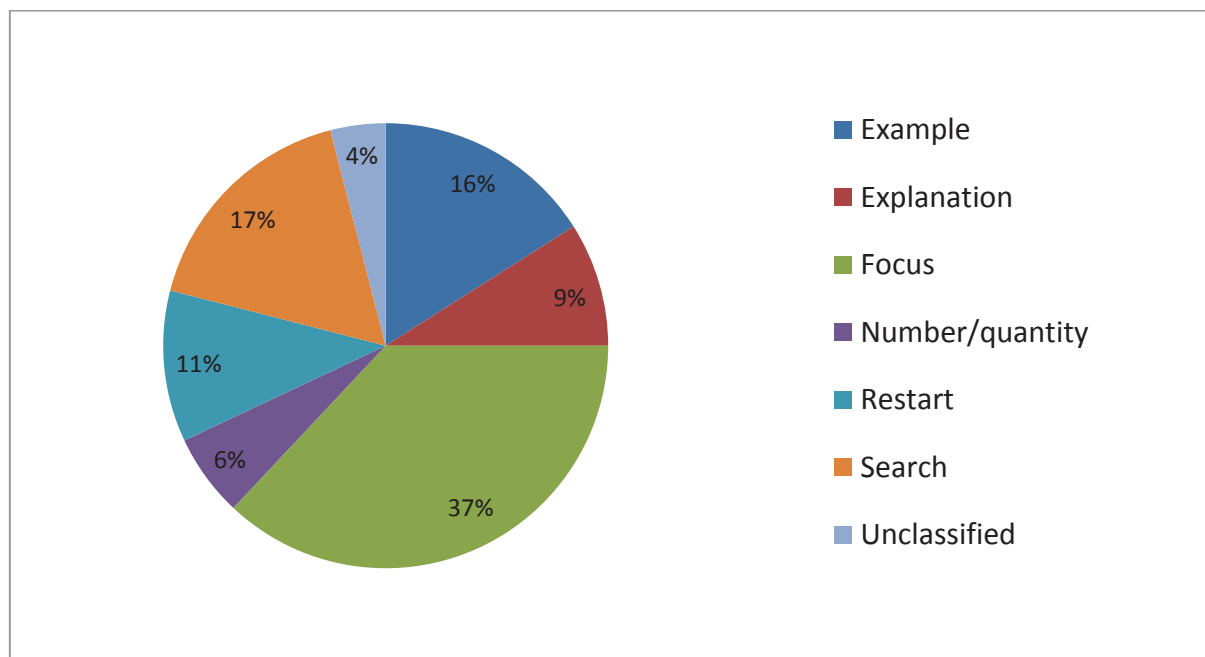


Figure 6.2 Distribution of the different functions of the discourse marker *like* in ICE-GB

Here again, the dominant function of the discourse marker *like* is marking lexical focus with over one third of the instances (37%). Within this category, *like* can both focus on new information as well as on given information (cf. Section 2.4.1). In example (51), ‘Sidney’ has not been introduced yet, while in (52), the referent of ‘the shawl’ on which *like* focuses has been mentioned earlier in the discourse:

(51) Do you know,

I think there’s a deadline approaching so I thought it might be quite useful to talk
cos **like** Sidney mentioned some possibility

Right

Shall we try and get together sometime cos we’ve got one or two possible projects
<ICE-GB44>

(52) There were some headdresses in Debenham’s actually we which had sort of like
half of a shawl on them

I couldn’t really work out <unclear>

I horrid dresses that they had which they had **like** a massive shawl bit in velvet and
then nothing on the other side

and it wa oh it looked like they’d forgotten half of it or something

It was horrid <ICE-GB85>

A special case within the focusing function of the discourse marker *like* is its so-called ‘hyperbolic use’ (cf. Andersen 2001). This subtype occurred once in the data, and is illustrated in (53). In this example, the emphasis is on the large number of books, but the expression ‘tons’ is clearly exaggerated:

- (53) But I mean we we’re in the throes of moving house, and we just emptied the bookshelves ... books that are probably to somebody priceless items
Come on
Can we root amongst them first
You are welcome. I’ve got a bloody box full of them
Uhm we uh Chris has done that and we’ve done that when we’ve moved or he’s moved on. I mean he’s done different things and we’ve done different things so **like** tons of psychology books went because now he’s moving on or teaching s something else <ICE-GB15>

The second most frequent function of *like* is searching for the appropriate expression in 17% of the cases, followed by 16% of instances in which the discourse marker *like* introduces an example. In (54), the example introduced by *like* illustrates the ‘experience’ that is mentioned earlier by the speaker:

- (54) The general impression I got from the talk was that, uh before you sort do the course, you really have to have, quite of bit it’s preferable to have quite a bit of experience don’t you think
Yes yes
I think **like** an MBA uh in industry and commerce it’s as well to have got some feel for what what life out there is like <ICE-GB72>

In 11% of the cases, the discourse marker *like* occurs in the context of a false start and a restart, as is the case in example (55), in which the speaker first produces an assertion before reformulating it into a question:

- (55) I thought you meant language like a language like you know you have different **like** don’t you have different languages that you speak to different people
You know well certain certain parts of it that other people wouldn’t normally understand <ICE-GB21>

Like introduced an explanation in 9% of the instances and a number or a quantity in 6% of the instances. Finally, 4% of the cases remained unclassified regarding the function fulfilled by *like* in the utterance.

6.3 Canadian English

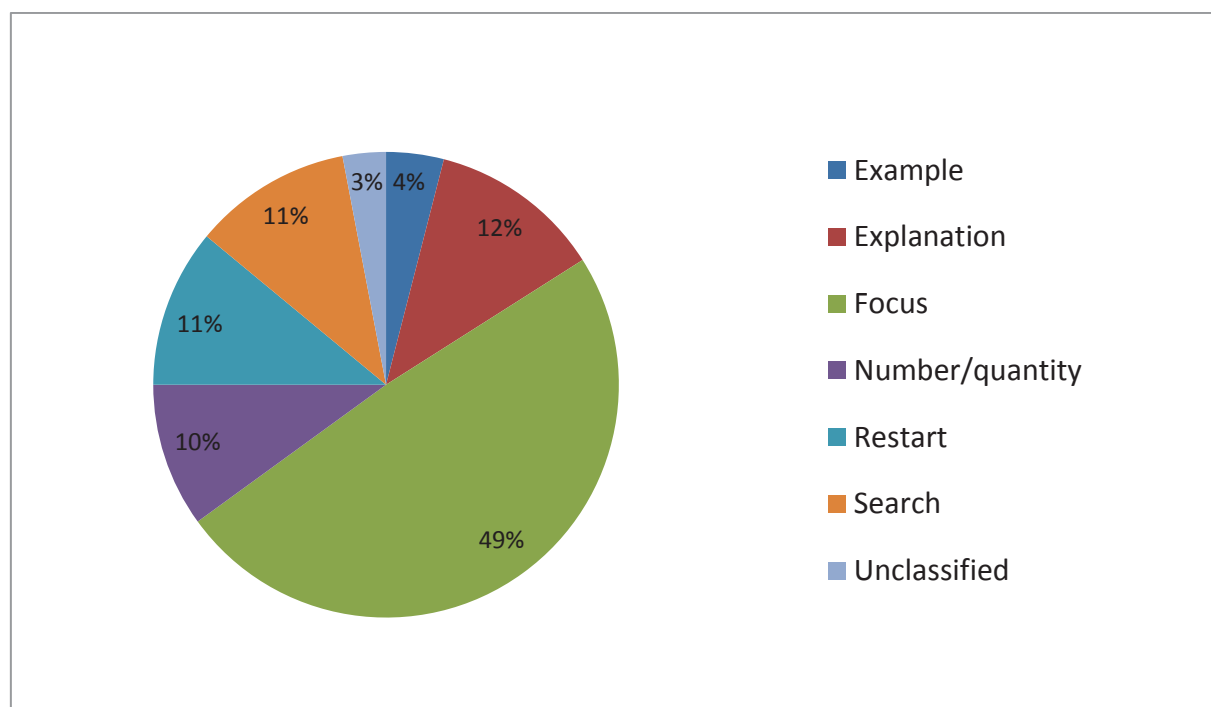


Figure 6.3 Distribution of the different functions of the discourse marker *like* in ICE-CA

As can be observed on Figure 6.3 above, almost half of the occurrences of *like* in the sample from ICE-CA have a focusing function (49%). These mostly occur in utterance-medial and utterance-initial position. The latter is illustrated in example (56):

- (56) And I and so she was telling me this and uh and I said well what did you say
 And she goes, well I told them that you're sort of you know a strange bird and
 And I went, ha excuse me
 <laugh>
Like the tears just came to my eyes, and I was just like, well, what do ya you mean,
 like <ICE-CA773>

In 12% of the instances, *like* was used to introduce an explanation. In 11 cases out of 12 with this function, *like* occurred in utterance-initial position, as a kind of afterthought on the preceding utterance. This is the case in the following two examples:

(57) So she gave she gives me this, set of twelve

Like, I don't have, what is that thing called where you put all, you know it's, where you put all your, forks and spoons and knives

Like the little organizer that you put in your door, drawer <ICE-CA659>

(58) And, I meant to call her and the thing is you know what you know those lucid dreams where you think you've called her but she's like, I'm calling her and all of the sudden you realize

there's no phone in your hand

Mhmm, yes yes, oh no <laugh>

Like when you're waiting for a ring, or like, you you've dreamt that you've called her already, I don't know <ICE-CA764>

The next two functions of *like* in terms of frequency show identical percentages in the data. Searching for the appropriate expression represents 11% of the instances, just as the use of *like* in the context of a restart. Both functions can be considered as part of a more general use of *like* to mark hesitation. These are illustrated in examples (59) and (60) respectively:

(59) But it was a lovely wedding

It was they had roses everywhere

And uh it's a beautiful beautiful huge **like** park-like uh, uh uh yard garden you know in their, uh around their house <ICE-CA497>

(60) Cos if you just put sugar we'll just **like** you'll taste the granules of sugar on it

Ya it'd be kind of, weird <ICE-CA629>

In 10% of the cases, *like* was used in combination with a numerical or quantitative expression, as in example (61). In this use, *like* appears exclusively in utterance-medial position.

(61) Three days on one day off two days on one day off, three days on one day

Right

Hmm

So it's **like** five out of seven

Because you do have to let your body rest because your performance will be that much de uh less you know if you, you know

if you don't allow your body to recuperate <ICE-CA579>

Finally, the least frequent function of *like* in ICE-CA is introducing an example with only 4% of the instances. 3% of the occurrences of *like* remained unclassified.

6.4 Irish English

The six functions that the discourse marker *like* can possibly fulfill are distributed as follows in the sample analyzed from ICE-IR:

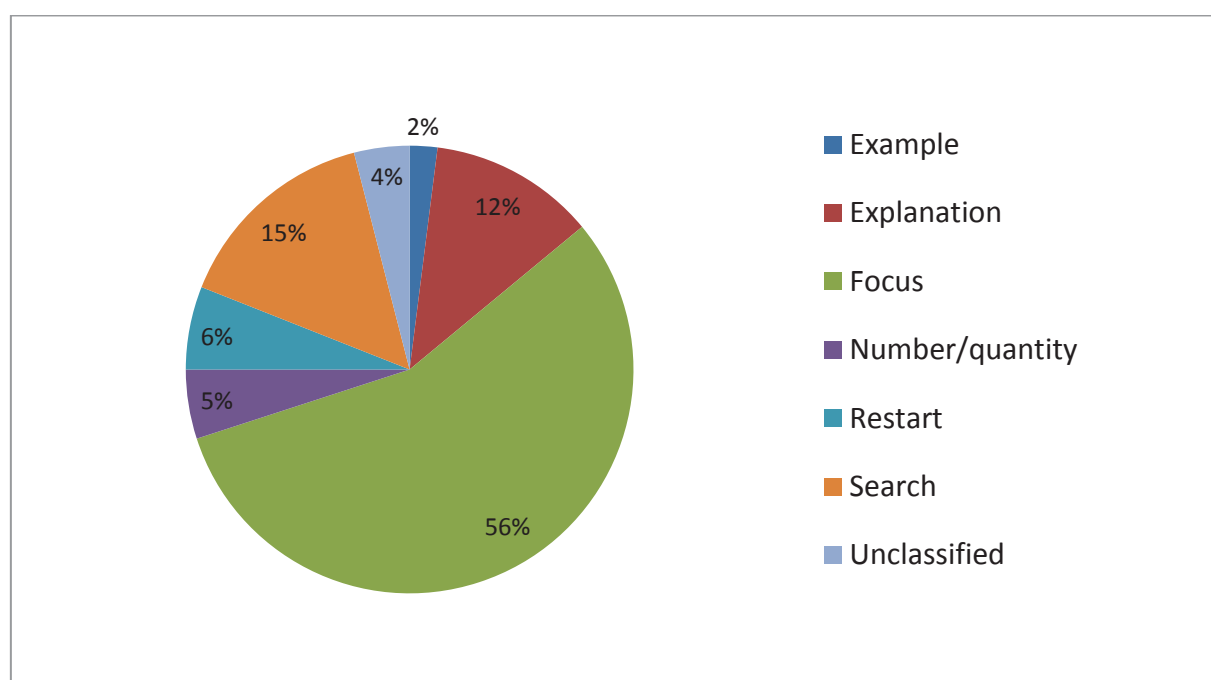


Figure 6.4 Distribution of the different functions of the discourse marker *like* in ICE-IR

As can be concluded by looking at Figure 6.4, *like* functions as a marker of lexical focus in the large majority of the instances (56%). This percentage is much higher than in the other varieties of English investigated so far, and appears to be related to the tendency for the discourse marker *like* to be used with a backward scope in Irish English (cf. Section 5.4 on the position of *like* in ICE-IR). In the present data, about half of the occurrences of *like* within this category were utterance-final, thus modifying the constituent to their left. This can be observed in examples (62) and (63), in which *like* focuses on a noun phrase and an adjective respectively. Note that in (63), the focusing function of *like* is further reinforced by the intensifier ‘really’.

- (62) Well she was home yesterday for she and Rosie to look for bridesmaids dresses for Rosie, uhm but Rosie’s away down to Carlow today, and Peggy’s staying til tomorrow morning

But she was really, well she was sick last night well she just had a bad cold **like**
<ICE-IR260>

- (63) But if things w didn't work out would you be straight back home
No, well if it was, if it wasn't too bad, if it was bad like, I wouldn't get
That you couldn't get work and you were just stone broke
Well like that would be really bad **like**, right if I couldn't if I'd no money, and it
was really bad, I'd have no choice but go home <ICE-IR140>

The second most frequent function of *like* is searching for the appropriate expression with 15% of the instances. In example (64), the discourse marker *like* weakens the speaker's commitment towards the way the utterance is formulated. Note, in addition to the mitigating function of *like*, the presence of other hesitation markers.

- (64) Do you think the constraints are internal, no. You think not
I think uh like, I think he's saying you know that we are we are limited like, in in
this century, but we have multiple choice in like, in those terms you know, **like**,
like we we can kind of marry who we want, things like that you know <ICE-
IR719>

In 12% of the cases, the function of *like* was to introduce an explanation to some information that the speaker thinks might need clarification. Here again, the occurrences of *like* with this function were essentially found in utterance-initial position, as is illustrated by example (65) below:

- (65) But how's the nursing going
Great
She's really bogged down
Like if she comes over Friday night she's going to have to go home and study all
day Saturday <ICE-IR342>

The next functions are much less frequent in the data from ICE-IR. Indicating a restart accounts for 6% of the instances and marking an approximate number or quantity for 5%. The least frequent function of the discourse marker *like* is introducing an example with 2% of the cases. Finally, four cases were too unclear or ambiguous to determine the exact function of *like* in those utterances and were coded as 'unclassified'.

6.5 New Zealand English

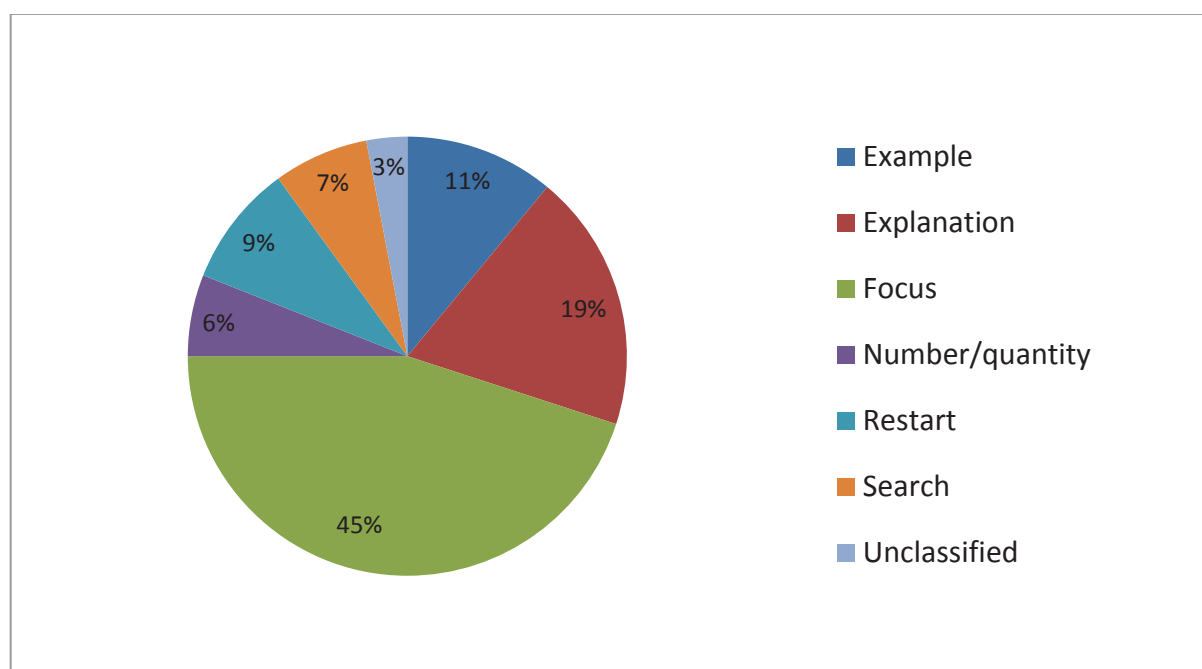


Figure 6.5 Distribution of the different functions of the discourse marker *like* in ICE-NZ

As can be observed on Figure 6.5 above, the most frequent function of *like* in the sample from ICE-NZ is marking lexical focus, just as in the other varieties of English investigated in this study. In the data from New Zealand speakers, the focusing function accounts for 45% of the occurrences of *like*, which mostly occur in utterance-medial position and to a lesser extent in utterance-initial position. These are illustrated in examples (66) and (67) respectively:

(66) It's quite funny really but I had I had a much better time when we went with like
the secondary students choir really just
<laughs> oh right
When we'd just we'd just **like** come in you know turn round and we'd just go out
again and you know and be out all night you know
and just go shopping till we dropped and you know <ICE-NZ318>

(67) I think it's really I ca I can see myself getting really really high you know well
I mean with the Lord eh you know
just really go for God you know his spirit and the witnessing to a lot of people and
preaching a lot
and **like** he really brought me down eh
it brought me down

it was like oh sometimes you know sometimes you know when there are times
when you don't really need to say much
Yeah <ICE-NZ25>

In 19% of the instances, the function of *like* was to introduce an explanation. When functioning as such, *like* mostly occurs in utterance-initial position, adding some information to clarify a previous statement. Note that in several cases, *like* occurs in combination with the discourse marker *you know* or *I mean*, which further reinforces its explanatory function. This is the case in the following two examples:

(68) There are some people who won't know what you're going on by what the rapture
all that stuff look
and the peace treaty peace treaty
are you aware of that
Yeah, yeah, yep
It's really hard
Like *you know* after three and a half years it's gonna be broken and blah blah blah
<ICE-NZ196>

(69) Like when I was really sick last week I rang I you know toll bill was nothing cos I
just needed you needed to talk to people
and you need like you need that connection to to to home and
Mm, mm, yeah
mm to be around
Like *I mean* I just had I had to talk to my twin sister for like I just talked to her for
hours and hours <ICE-NZ60>

The third most frequent function of *like* is introducing an example with 11% of the instances. In (70), for instance, the general category of 'high numbers' to which the examples belong is explicitly mentioned. Besides, the phrase 'or something' that follows indicates that any other member of the general concept would be suitable as well in that context:

(70) The salvation army shop
where is it now
is it way down near the end of the street
no yeah yeah

is it at number twenty six

because it used to be a high number **like** two five six or something and now they've got on their card twenty six <ICE-NZ440>

The next function of the discourse marker *like* in terms of frequency is that of restart with 9% of the instances. Searching for the appropriate expression represents 7% of the instances, and introducing an approximate number or quantity 6%. Finally, three occurrences of *like* remained unclassified because the incomplete utterances in which they appeared did not allow to determine their exact function.

6.6 Discussion

The results emerging from the analysis of the function of *like* as a discourse marker have revealed a number of commonalities and differences between the five varieties of English investigated in this dissertation.

Firstly, the five subcorpora show a common tendency towards the use of *like* to mark lexical focus, with a proportion ranging from around one third to over half of the occurrences of *like*. Data show that *like* can equally focus on new as well as on given information, thus contradicting Underhill's (1988) claim that this function of the discourse marker is restricted to new information only. When used as a focuser, *like* predominantly occurs in utterance-medial or utterance-final position. In this respect, the sample from ICE-IR stands out once again as about half of its instances are utterance-final. This can be explained by the propensity of the discourse marker *like* to be used with a backward scope in Irish English (cf. Section 5.4).

When it comes to the other functions of *like*, the picture is somewhat more contrasted as there is a great deal of variation in their distribution across the five varieties of English under study here. Overall, searching for the appropriate expression is relatively frequent and the function ranks second in the British and the Irish subcorpora. The combination of *like* and a numerical or quantitative expression displays a strikingly high frequency (19%) in the sample from American speakers, while its use is rather limited in the other four varieties. Similarly, introducing an explanation is the second most frequent function in ICE-NZ with 19% of the occurrences of *like*, while its frequency does not exceed 12% in the other subcorpora. Introducing an example has a low frequency in all the subcorpora, except in the

New Zealand sample in which it directly follows the focusing function in terms of frequency. Finally, the use of *like* in the context of a restart is among the least frequent functions of *like* in all varieties of English, with a percentage ranging between 6% and 11%.

This chapter investigated the function fulfilled by the discourse marker *like* in discourse produced by speakers of American, British, Canadian, Irish and New Zealand English. The following – and final – chapter of this dissertation looks into two sociolinguistic variables, namely speaker's gender and age, and the way in which they affect the use of *like* as a discourse marker in these five varieties of English.

Chapter 7

Sociolinguistic determinants of *like*

Various studies that investigate the discourse marker *like* contain a sociolinguistic dimension, looking at the factors that influence its use. As far as the age variable is concerned, research has shown that *like* as a discourse marker tends to be more frequent among females than among males (cf. Andersen 2001; D'Arcy 2005). However, this ratio is also influenced by the speaker's age. Tagliamonte (2005), for instance, observed that this higher frequency of use of *like* by female speakers is only significant among the age cohort of old teenagers. Several other scholars (cf. Fleischman 1998; Dailey-O'Cain 2000) have looked into the influence of the speakers' age on the presence of *like* in their discourse, drawing the conclusion that, overall, the discourse marker is most prominent in adolescents' discourse.

This final chapter of the empirical part of this dissertation investigates how the sociolinguistic variables of gender and age combine to determine one's linguistic behavior regarding the use of the discourse marker *like*. Section 7.1 compares the effect of gender on the use of *like* in data from Canadian and American English. Section 7.2 then looks into the distribution of *like* across different age classes in the two samples. The combination of these two factors is described in Section 7.3, resulting in a specific profile of speakers with the highest propensity to use *like* as a discourse marker. This section is further illustrated by examples from individual speakers. Finally, the results emerging from all these analyses are discussed in Section 7.4.

7.1 Gender

Figure 7.1 below represents the distribution of the discourse marker *like* according to speaker's gender in the samples from ICE-CA and the SBC.

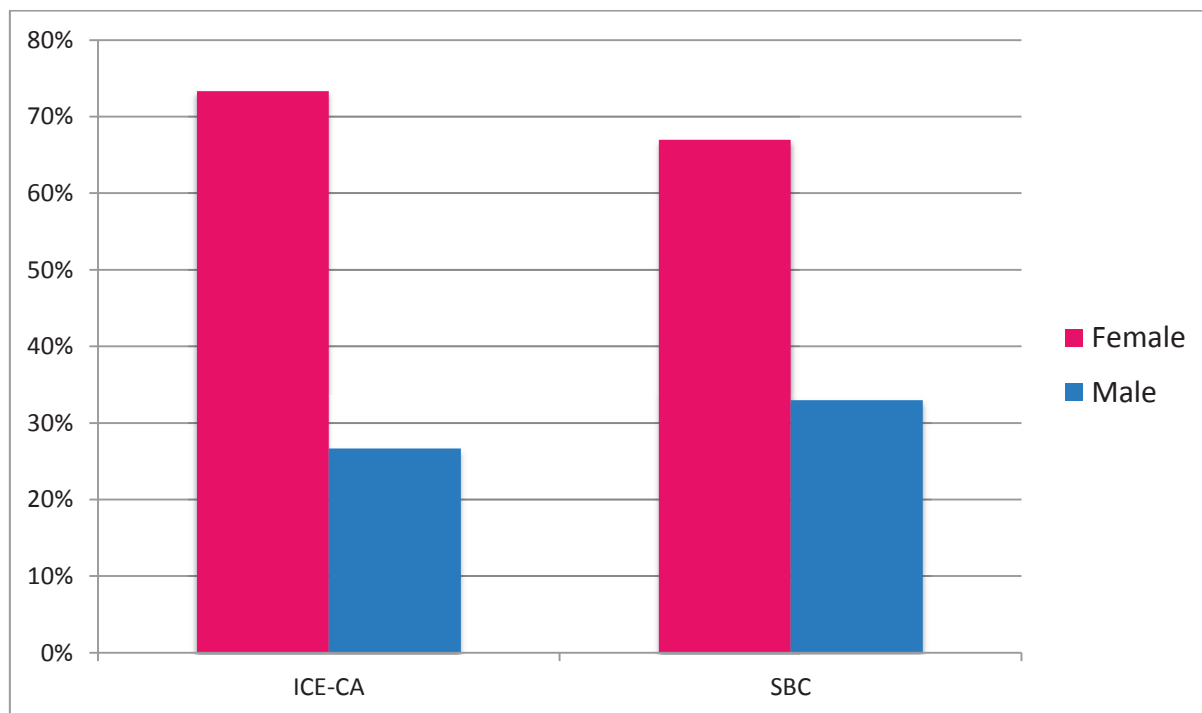


Figure 7.1 Distribution of the discourse marker *like* according to gender in ICE-CA and the SBC

A large difference between the female and the male use of *like* immediately stands out in both subcorpora. In the sample from ICE-CA, 73.33% of the occurrences of *like* were produced by women compared to 26.67% only by men. This gap is slightly weaker in the data from American speakers, with two thirds (67%) of the instances produced by females, and 33% by males. This difference of proportion between the two groups is statistically non-significant ($\chi^2 = 2.874$, $p = 0.09$). Therefore, one can conclude that both varieties investigated here follow a common trend towards a higher frequency of use of the discourse marker *like* among female speakers than among male.

7.2 Age

7.2.1 Canadian English

The 300 occurrences of the discourse marker *like* in the sample from ICE-CA were coded regarding the age of the speaker by whom they were produced. Figure 7.2 below represents the distribution of the seven classes according to which speaker's age was classified.

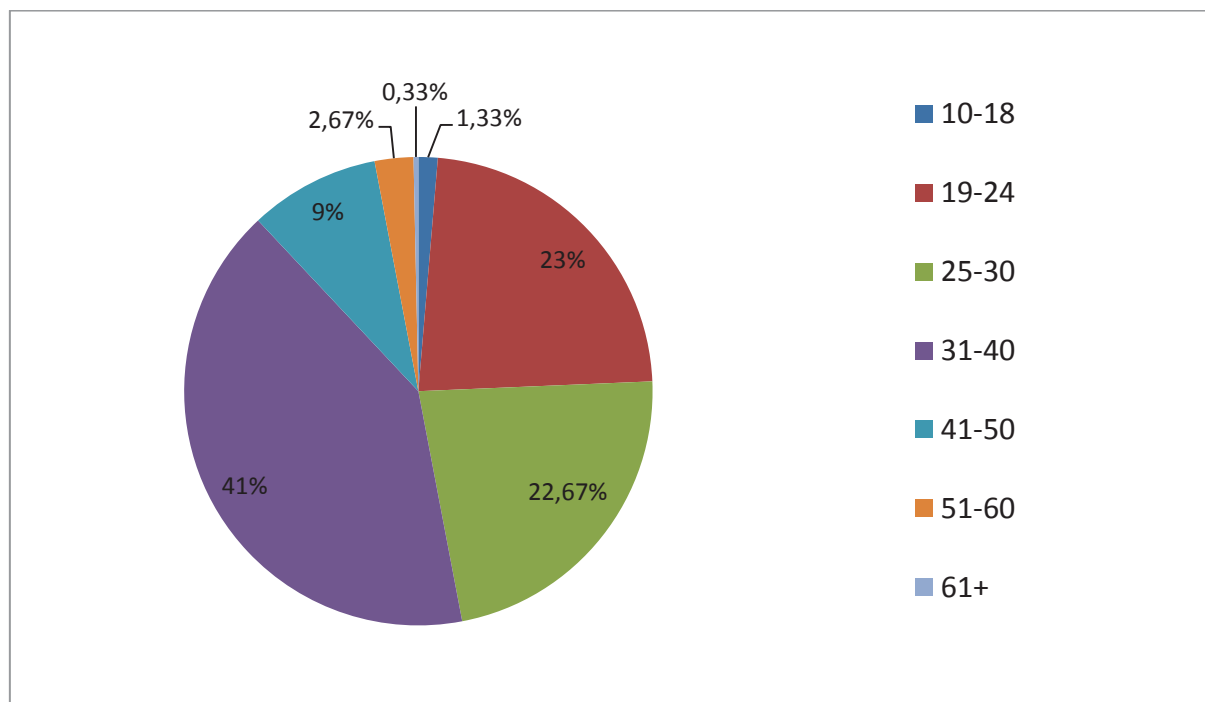


Figure 7.2 Distribution of the discourse marker *like* according to age in ICE-CA

As can be observed on this figure, the 31-40 category is by far the most frequent one with 41% of the occurrences of *like* in the sample. Speakers aged 19-24 and 25-30 come in second and third position with roughly the same proportion (23% and 22.67% respectively). Note, however, that when these two age classes are grouped together, speakers between 19 and 30 years old actually outnumber their older counterparts regarding their frequency of use of *like*. 9% of the instances were attributed to middle-aged speakers, i.e. in the 41-50 category. Older speakers (aged 51 or above) appear to use *like* as a discourse marker strikingly less frequently (2.67% for the 51-60 age group and as little as 0.33% for the 61+ class). Similarly, adolescent speakers (aged 10-18) only account for a very little percentage of all the occurrences of *like* (1.33%). Note, however, that these last three age classes are underrepresented in data from ICE-CA, as can be concluded by looking at the corresponding metadata. This may – at least partly – explain the lower percentages displayed by the speakers of those categories.

7.2.2 American English

Like is not equally distributed across the different age classes in the sample from American English either. The seven age groups display variable frequencies of use of the discourse marker. These are represented on Figure 7.3 below.

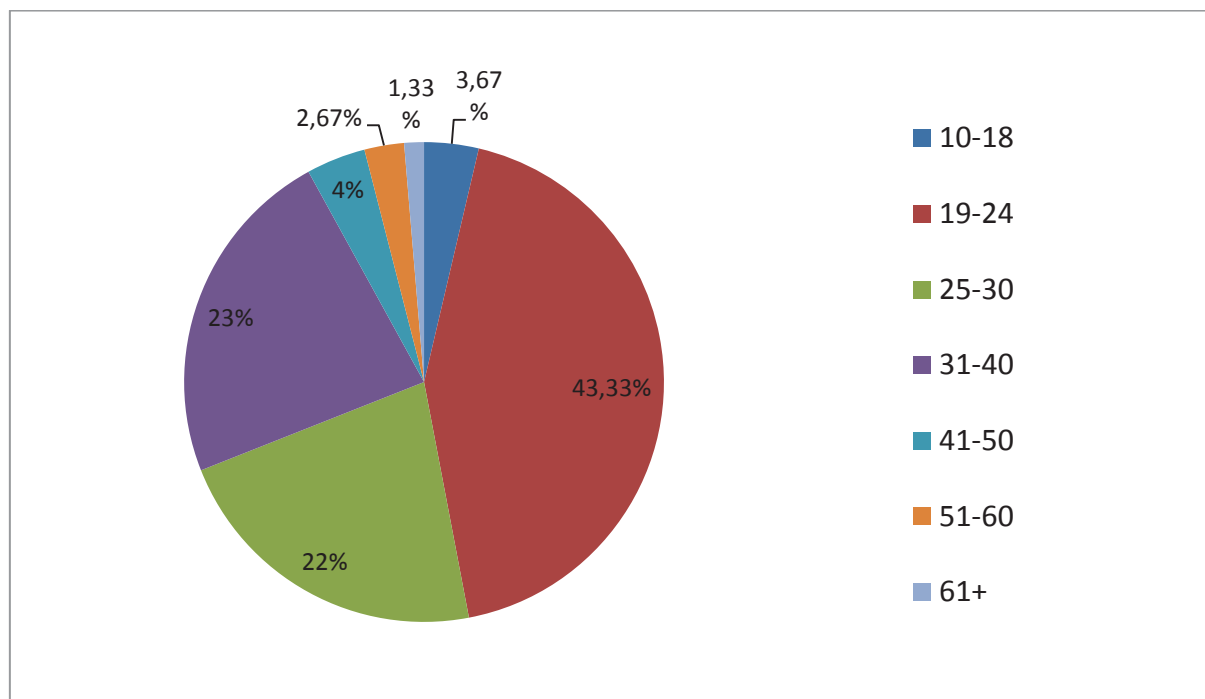


Figure 7.3 Distribution of the discourse marker *like* according to age in the SBC

As can be seen on this graph, almost half of the occurrences of *like* in the sample (43.33%) were produced by speakers between 19 and 24 years old. This class is followed by the 31-40 category with 23% of the instances and the 25-30 category with 22%. The discourse marker *like* appears to be strikingly less frequent among speakers aged 41 and above. The 41-50 age group accounts for 4% of all the instances in the sample, while speakers between 51 and 60 years old only produced 2.67% of all the occurrences of *like*. Similarly, the use of the discourse marker among the 61+ category does not exceed 1.33%. Finally, teenagers (10-18) also display a relatively low frequency of use of *like* as a discourse marker with 3.67% of its total occurrences. Here again, however, both very young and relatively old participants only represent a minor proportion of the transcriptions in the subcorpus, which can contribute to the lower figure that they obtain regarding the frequency of use of *like* as a discourse marker.

7.2.3 Comparison of the two samples

The following figure compares the distribution of the discourse marker *like* across the different age classes in the sample from ICE-CA and that from the Santa Barbara Corpus of Spoken American English.

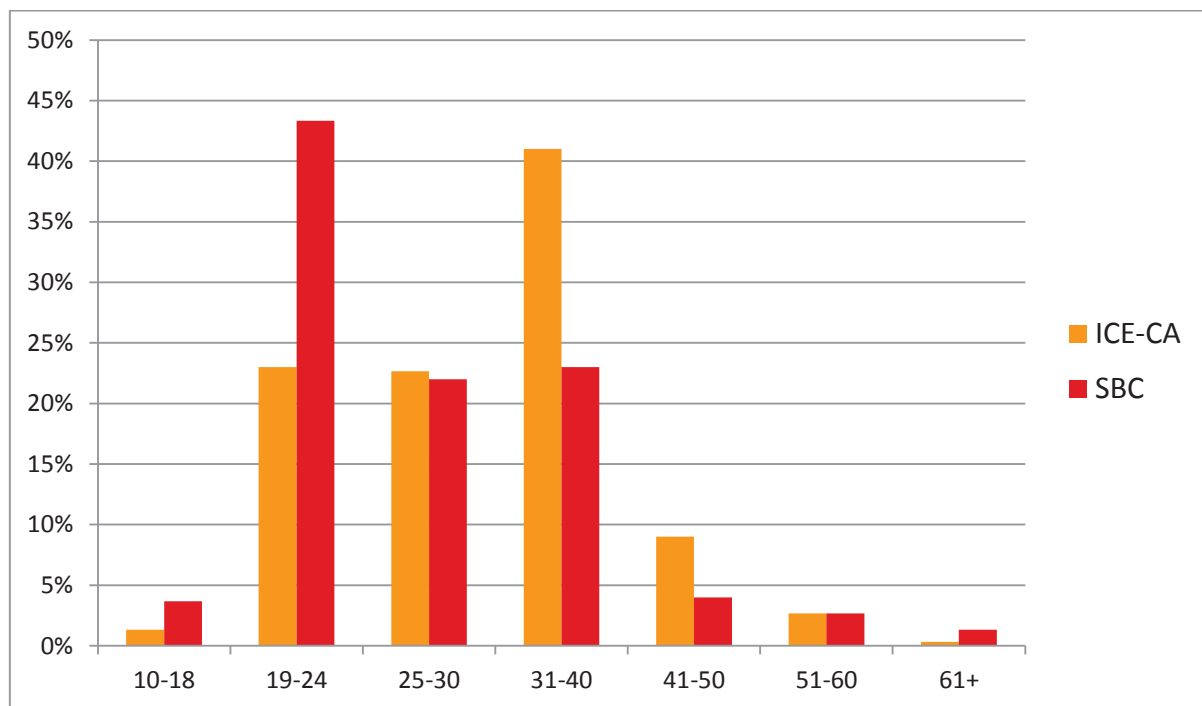


Figure 7.4 Distribution of the discourse marker *like* according to age in ICE-CA and the SBC

Some age classes on this graph display similar frequencies across the two subcorpora. The 51-60 and the 61+ categories, for instance, represent a very little percentage of all the occurrences of *like* in both samples. The age class composed of adolescent speakers (10-18) shows a somewhat higher percentage in the Canadian sample than in the American one, but overall, it represents a minor proportion of all the instances in both varieties. Speakers between 25 and 30 years old produced a bit less than a quarter of the total occurrences of *like* in ICE-CA (22.67%) as well as in the SBC (22%).

On the other hand, three groups show marked differences of proportion between the two samples investigated in this chapter. In this respect, the data from ICE-CA and the SBC appear to follow almost an opposite trend. The Canadian speakers who use the discourse marker *like* most frequently are between 31 and 40 years old (41%), followed by the 19-24 category with 23%. In the American sample, by contrast, speakers aged 19-24 score highest in terms of frequency of use of *like* (43.33%), while the 31-40 age group ranks second with 23% of the instances. Finally, *like* as a discourse marker was rarely used by middle-aged adults (41-50), although somewhat more frequently in the sample from ICE-CA (9%) than in that from the SBC (4%).

7.3 Individual speakers' use of *like*

Combining the distribution of both the gender and the age variables on a single graph brings out certain sections of the population under study for their particular use of *like* as a discourse marker. Examples of individual speakers are included here to illustrate those sections of ICE-CA and the SBC.

7.3.1 Canadian speakers

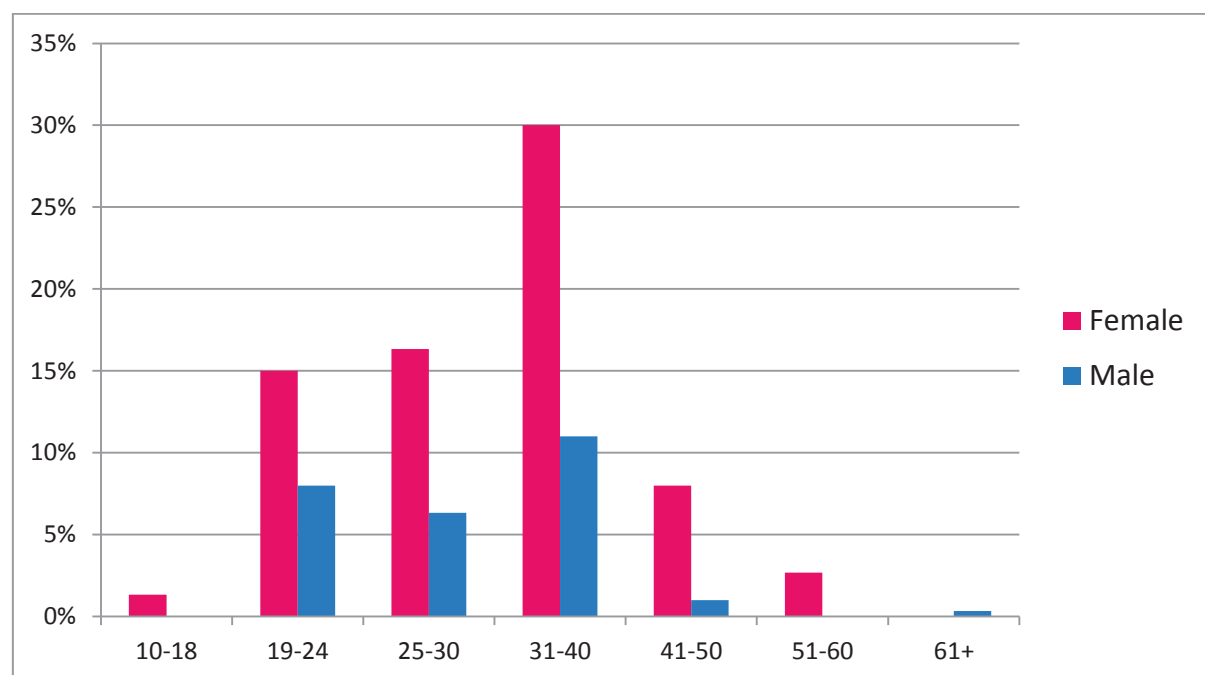


Figure 7.5 Distribution of the discourse marker *like* according to gender and age in ICE-CA

As can be observed on Figure 7.5 above, female speakers outnumber male speakers in all age groups except among speakers above 61 years old. Overall, the category that displays the highest frequency of use of *like* as a discourse marker is that composed of women between 31 and 40 years old, with 30% of the total number of occurrences of *like* in the sample.

When looking more closely at this age group, one individual speaker clearly stands out for her unusually highly frequent use of the discourse marker, as can be seen in example (71) below, full of instances of *like*:

(71) So you've never heard from him again or

Oh I once or twice just **like** made eye contact with him **like** somewhere at a crowded place and **like** just left and

He leaned over from the chairlift and whipped me on the leg with his ski pole and I had this bruise that went from **like** here, around to my buttock, this welt and bruise and

The morning the day I moved out he, we were having this **like** screaming fight right

Like I never cry any more even when I'm arguing with Jeff <speaker 1A: file S1A-078>

This speaker alone accounts for thirteen occurrences of *like* out of the ninety in the section of the sample to which she belongs. She mostly uses *like* in utterance-medial position, to mark lexical focus on verbs, as in examples (72) and (73). To a lesser extent, *like* focuses on other components such as noun phrases and adjectives, or occurs in combination with a numerical expression.

(72) Now where did the Dalhousie apartment get in there

I don't know

I'd have to actually **like** look at a calendar

I'd I'd have, I'd have to look at a date book and sort of figure like, there were like five moves in a year there <ICE-CA20>

(73) So anyways by the end of that, I kept thinking like he's not going he's not going yet

And I was just **like** trying to break up the relationship cos I didn't wanna want to be I knew I had to but I didn't do it <ICE-CA635>

Similarly, over a quarter of the occurrences of *like* among 25 to 30-year-old women were actually produced by a single speaker. Contrary to the pattern of use described above, in the present case, the speaker mostly uses *like* in utterance-initial position to introduce an explanation (74) or in the context of a restart (75):

(74) Tara and I asked him how, Tara and I asked him how he made his squid, because

I'd never had squid that was so tender

It was so beautiful,

Like you your teeth melted in it <ICE-CA145>

(75) So how much

A teaspoon

Just put that in first

Ya like a teaspoon
 And then you just do,
Like you need just a little bit of water <ICE-CA257>

7.3.2 American speakers

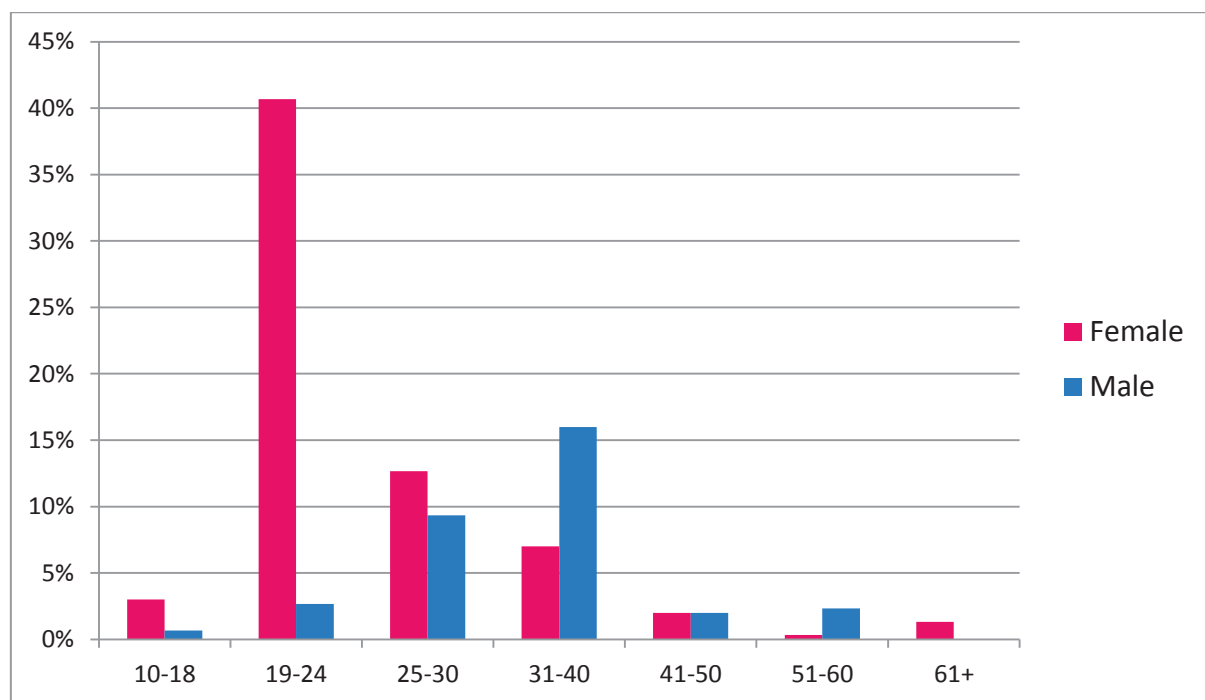


Figure 7.6 Distribution of the discourse marker *like* according to gender and age in the SBC

The picture is more contrasted when combining the gender and the age variables in the sample from the SBC. Female speakers show numerical superiority in four age classes out of seven, both genders display an identical percentage in the 41-50 category, and men outnumber women in the 31-40 and 51-60 age groups.

Overall, the category that shows the highest frequency of use of *like* as a discourse marker is undoubtedly that of females between 19 and 24 years old, with 40.67% of the total number of instances. Three individual speakers appear to be largely contributing to this high proportion due to their strong tendency to use the discourse marker *like*.

Almost one fifth of the occurrences of *like* in this group were produced by Lynne, a 19-year-old student of equine science, describing in detail the degree she is doing, as in:

(76) When I get done with them,

A lot of times, **like** I'll get done, and I'll think I'm done, and I'll look at – look down at the horse's hoof, and it's still, it's too long.

And then **like**, I would never, ever, ever, trust myself to shoe a horse.

See we did that too.

We did – we did a lot of stuff.

We did a lot of stuff with the – **like** we had the, um, the burners, you know

<Lynne: file SBC001>

She mostly uses *like* to search for the appropriate expression (77), often together with an unfilled pause, or in the context of a restart (78). Those two functions of the discourse marker convey the speaker's uncertainty and weak commitment towards both the content and the formulation of her discourse:

(77) Oh, and it's really tiring though.

And it – you know **like**, ... you get so –

I've ... only done like, ... well, ... at the end of the year, now see, I took the second half of the course.

And, ... right now, I've probably, ... only shod about, ... five horses. <SBC16>

(78) What we do, then that's ... that's where the ferrier comes in.

Every shoe, is **like**, ... you get order, you know, I would like a b- case of double-aught shoes.

You'd get -- ... the ferrier gets them. <SBC12>

The other two participants are 20-year-old women, belonging to the same conversation. Their use of *like* represents almost half of the total occurrences of the discourse marker in that section of the sample. *Like* occurs in varying positions within their utterances and fulfills different functions. In example (79), for instance, Lisa uses *like* because she cannot find the appropriate word to accurately express what she means. In (80), Marie's use of *like* marks lexical focus on the following noun phrase:

(79) And he was wearing like a – **like**, you know, like an ... shirt?

You can ... buy em like.

Shirts like, ... you know, for the police, anywhere. <SBC457>

(80) Well, they were like, her paren –

Her mom's like an alcoholic, and stuff like that, and her dad's **like** a biker now, and I'm like, God, you know. <SBC503>

Another group worth mentioning is that made up of 31 to 40-year-old men. Among them, one speaker stands out for his particularly frequent use of *like* (one quarter of all the instances in that category). In his discourse, *like* fulfills various functions, from introducing a numerical expression, as in example (81), to marking hesitation, as in example (82):

(81) They're these big gnarly suckers.

They're like um,

Yeah, I remember what they look like.

They're **like** m – f – I don't know, six or eight inches,

big eyes,

and they've got, ... they only eat like other fish. <SBC282>

(82) I was just imagining what was going through its little mind, as the Oscar was swimming around, frantically trying to dislodge it, so it could swallow it.

You know?

And um, ... finally it d- did manage to, you know, v- **like** loosen it <SBC290>

7.4 Discussion

On the basis of this sociolinguistic analysis carried out on samples from ICE-CA and the SBC, some conclusions can be drawn as to how the use of *like* is distributed in the two English-speaking communities in terms of gender and age.

As far as gender is concerned, both samples show a common trend towards a much higher use of *like* by women than by men, even more markedly among Canadian speakers than among their American counterparts. This result thus confirms what previous studies revealed (cf. Andersen 2001; D'Arcy 2005). When it comes to the reasons for this gap between genders, this analysis did not completely shed light on the issue. Women's propensity to use more hedges and intensifiers (cf. Section 2.5.1) did not clearly appear in the present data since, overall, they did not use *like* to search for the appropriate expression and to mark lexical focus more than men did.

Regarding the age variable, the two varieties of English display a different distribution of *like* as a discourse marker. In the sample from ICE-CA, the age group that uses *like* most frequently is that between 31 and 40 years old, followed by speakers in their thirties. By contrast, speakers aged 19-24 score highest in the sample from the SBC, followed the 25-30

and 31-40 age classes. Note, however, that in both subcorpora, the combination of these three “middle-age” categories accounts for almost 90% of all the occurrences of *like*. Teenagers’ and elderly speakers’ use of *like* remains relatively infrequent, thus departing from the idea commonly found in the literature that *like* is a typical feature of adolescents’ speech.

Finally, looking qualitatively at individual speakers’ discourse reveals that there is a great deal of variation in the use of *like* even among speakers belonging to the same age/gender group. If classifying the occurrences of the discourse marker *like* according to those two sociolinguistic parameters turns out to be useful to bring out some general tendencies, its idiosyncratic character should not be overlooked.

This chapter completed the empirical part of this dissertation with the analysis of the sociolinguistic variables of speaker’s gender and age regarding the use of *like* across the Canadian and the American varieties of English. Next comes the conclusion to this dissertation, reflecting on the study carried out, its contributions and limitations, as well as avenues for further research.

Conclusion

The present dissertation set out to investigate the use of the discourse marker *like* across the five native varieties of American, British, Canadian, Irish and New Zealand English. The method chosen to carry out this comparison was a corpus-based study. Spoken corpora, by providing access to authentic data, made it possible to examine the actual use of the discourse marker *like* by native speakers, a facet of the English language that does not necessarily stand out in other types of data. The innovative character of this study lay in its contrastive dimension. Extending the analysis to several varieties of native English served as a demonstration of the internal diversity that exists within ‘the English language’. After reviewing the literature on the topic in terms of the evolution, the formal characteristics, the functions and two of the sociolinguistic determinants of the discourse marker *like*, the corpus analysis itself was performed on data from the SBC, ICE-GB, ICE-CA, ICE-IR and ICE-NZ.

From a quantitative point of view, the analysis of the discourse marker *like* has revealed differences in frequency across the five varieties. D’Arcy’s (2005) argument on the marked prominence of *like* as a discourse marker in North American English compared to British English was only backed up to a certain extent by our findings. Canadian and American speakers did use *like* as a discourse marker significantly more often than their British counterparts did. However, surprisingly, the relative frequency of *like* as a discourse marker was significantly higher in Irish English than in any other variety under scrutiny. A possible explanation for these quantitative divergences lies in the fact that the subcorpora were compiled at different times, starting with ICE-GB, therefore possibly downplaying the extent of the phenomenon in British English. One can thus wonder whether replicating this study on data collected contemporaneously would yield similar results.

Not only do speakers from the five English-speaking communities use the discourse marker *like* to varying degrees, they also use it in different ways. The analysis of the position and the frequency of *like* as a discourse marker has shed light on these qualitative differences. Firstly, the tendency of *like* – and of discourse markers in general – to occur in utterance-initial position largely depicted in the literature (cf. Hansen 1997; Tagliamonte 2005) did not hold true in the present case as the discourse marker *like* was predominantly used in utterance-medial position across the five subcorpora. However, Irish speakers stood apart from the four other groups in that they tended to use *like* in utterance-final position with a backward scope and a focusing function. In this respect, the dominant function of *like* in all the subcorpora happened to be marking lexical focus, while the two functions related to hesitation, i.e. searching for the appropriate expression and introducing a restart, only accounted for a minor proportion of all the functions. These findings thus strongly contradict the belief that *like* is a marker of indecision contributing to vagueness of expression (cf. D’Arcy 2005).

From a sociolinguistic perspective, the analysis of the gender and age variables in ICE-CA and the SBC has given an insight into the relatively limited role played by sociolinguistic factors compared to idiosyncrasy in the use of *like* as a discourse marker. Overall, women’s propensity to use *like* as a discourse marker more than men (cf. Andersen 2001; Tagliamonte 2005) was confirmed by our results, in both ICE-CA and the SBC. The analysis of the age factor yielded a more contrasted picture. Surprisingly, Canadian speakers who used the discourse marker *like* most frequently were in their thirties. By contrast, *like* was most frequent among 19 to 24-year-old Americans. Whether this observation is symptomatic of ongoing linguistic change or of age-grading (cf. Tagliamonte 2005) could be answered by carrying out a diachronic study on our data. In the present analysis, however, general patterns common to the group as a whole only played a limited role in determining individuals’ use of the discourse marker *like*, especially qualitatively. The position and function fulfilled by the discourse marker *like* in its occurrences largely varied from one individual to the next, irrespective of the specific class to which the speakers belonged, so much so that no straightforward pattern of use could be established on these aspects of the discourse marker.

The findings resulting from the present study can be considered as avenues to be further investigated in the future. This analysis was restricted to five Inner Circle varieties of English. Therefore, it would be interesting to extend the comparison to other – possibly Outer Circle – varieties of English in order to determine to what extent the observations made on the use of the discourse marker *like* can be transposed to different linguistic communities.

Furthermore, one could look into other discourse markers in the same five subcorpora to evaluate the prototypicality of *like* as a member of the linguistic class composed by discourse markers. Finally, comparing the use of *like* in face-to-face conversations with data from other types of spoken discourse, such as telephone conversations, would most probably bring out other aspects of the communication situation that play a role in the use of the discourse marker *like*.

References

- Adolphs, S. & Knight, D. (2010). Building a spoken corpus: What are the basics? In A. O’Keeffe & M. McCarthy (Eds.). *The Routledge Handbook of Corpus Linguistics* (pp. 38-52). London/New York: Routledge.
- Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Alami, M. (2015). Pragmatic functions of discourse markers: A review of related literature. *International Journal on Studies in English Language and Literature*, 3(3), 1-10.
- Andersen, E. S., Brizuela, M., DuPuy, B. & Gonnerman, L. (1999). Cross-linguistic evidence for the early acquisition of discourse markers as register variables. *Journal of Pragmatics*, 31, 1339-1351.
- Andersen, G. (1998). The pragmatic marker *like* from a relevance-theoretic perspective. In A. H. Jucker & Y. Ziv (Eds.). *Discourse markers: Description and theory* (pp. 147-170). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Andersen, G. (2001). *Pragmatic markers and sociolinguistic variation: A relevance-theoretic approach to the language of adolescents*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Blakemore, D. (1987). *Semantic constraints on relevance*. Oxford: Blackwell.
- Blakemore, D. (2002). *Relevance and linguistic meaning: The semantics and pragmatics of discourse markers*. Cambridge/New York: Cambridge University Press.
- Blakemore, D. (2006). Discourse markers. In L. R. Horn & G. Ward (Eds.). *The Handbook of Pragmatics* (pp. 221-240). Malden/Oxford/Victoria: Blackwell Publishing.
- Blyth, C., Recktenwald, S. & Wang, J. (1990). I’m like, “Say what?!”: A new quotative in American oral narrative. *American Speech*, 65(3), 215-227.
- Brinton, L. J. (1996). *Pragmatic markers in English: Grammaticalization and discourse functions*. Berlin/New York: Mouton de Gruyter.
- Brown, P. & Levinson, S. (1987). *Politeness: Some universals in language*. Cambridge: Cambridge University Press.
- Croucher, S. M. (2004). Like, you know, what I’m saying: A study of discourse marker frequency in extemporaneous and impromptu speech. *National Forensic Journal*, 2, 38-47.

- Dailey-O'Cain, J. (2000). The sociolinguistic distribution of and attitudes toward focuser *like* and quotative *like*. *Journal of sociolinguistics*, 4(1), 60-80.
- D'Arcy, A. F. (2005). *Like: Syntax and development*. Unpublished doctoral thesis, University of Toronto.
- Eckert, P. (1988). Adolescent social structure and the spread of linguistic change. *Language in Society*, 17(2), 183-207.
- Fleischman, S. (1998). Des jumeaux du discours. *La Linguistique*, 34(2), 31-47.
- Fraser, B. (1987). Pragmatic formatives. In J. Verschueren & M. Bertuccelli Papi (Eds.). *The Pragmatic Perspective* (pp. 179-194). Amsterdam: Benjamins.
- Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1-4), 19-33.
- Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6(2), 167-190.
- Fraser, B. (1999). What are discourse markers? *Journal of Pragmatics*, 31, 931-952.
- Fuller, J. M. (2003). Discourse marker use across speech contexts: A comparison of native and non-native speaker performance. *Multilingua*, 22, 185-208.
- Fung, L. & Carter, R. (2007). Discourse markers and spoken English: Native and learner use in pedagogic settings. *Applied Linguistics*, 28(3), 410-439.
- Gilquin, G. & De Cock, S. (2011). Errors and disfluencies in spoken corpora: Setting the scene. *International Journal of Corpus Linguistics*, 16(2), 141-172.
- Gilquin G. & Granger, S. (2015). Learner language. In D. Biber & R. Reppen (Eds.). *The Cambridge Handbook of Corpus Linguistics* (pp. 418-435). Cambridge: Cambridge University Press.
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgans (Eds.). *Syntax and semantics 3: Speech acts* (pp. 41-58). New York: Academic Press.
- Gries, S. T. (2006). Exploring variability between and within corpora: Some methodological considerations. *Corpora*, 1(2), 109-151.
- Hansen, M.-B. M. (1997). *Alors* and *donc* in spoken French: A reanalysis. *Journal of Pragmatics*, 28, 153-187.
- Hansen, M.-B. M. (1998). *The function of discourse particles: A study with special reference to spoken standard French*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language and Communication*, 10(3), 185-205.
- Huang, L. F. (2013). The use of Linear Unit Grammar (LUG) in the investigation of discourse markers in spoken English. *International Journal of Language Studies*, 7(3), 119-136.
- ICE. (1990-2017). *International Corpus of English*.

- Irvine, J. T. (2001). "Style" as distinctiveness: The culture and ideology of linguistic differentiation. In P. Eckert & J. R. Rickford (Eds.) *Style and sociolinguistic variation* (pp. 21-43). Cambridge: Cambridge University Press.
- Jabeen, F., Rai, M. A. & Arif, S. (2011). A corpus based study of discourse markers in British and Pakistani speech. *International Journal of Language Studies*, 5(4), 69-86.
- Jucker, A. H. & Smith, S. W. (1998). And people just you know like 'wow': Discourse markers as negotiating strategies. In A. H. Jucker & Y. Ziv (Eds.). *Discourse markers: Description and theory* (pp. 171-202). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Jucker, A. H. & Ziv, Y. (1998). Discourse markers: Introduction. In A. H. Jucker & Y. Ziv (Eds.). *Discourse markers: Description and theory* (pp. 1-12). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Kachru, B. B. (1991). World Englishes and applied linguistics. In M. L. Tickoo (Ed.). *Languages and Standards: Issues, Attitudes, Case Studies* (pp. 178-205). Anthology Series 26. Singapore: SEAMEO Regional Language Centre.
- Koops, C. & Lohmann, A. (2015). A quantitative approach to the grammaticalization of discourse markers: Evidence from their sequencing behavior. *International Journal of Corpus Linguistics*, 20(2), 232-259.
- Labov, W. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change*, 2, 205-254.
- Lenk, U. (1998a). Discourse markers and global coherence in conversation. *Journal of Pragmatics*, 30, 245-257.
- Lenk, U. (1998b). *Marking discourse coherence: Functions of discourse markers in spoken English*. Tübingen: Gunter Narr Verlag.
- Macaulay, R. (2001). You're like 'why not?': The quotative expressions of Glasgow adolescents. *Journal of Sociolinguistics*, 5(1), 3-21.
- McEnery, T. & Wilson, A. (Eds.). (2001). *Corpus Linguistics: An Introduction*. 2nd edition. Edinburgh: Edinburgh University Press.
- Meehan, T. (1991). It's like 'What's happening in the evolution of like?': A theory of grammaticalization. In K. Ichihashi & M. S. Linn (Eds.). *Kansas Working Papers in Linguistics*, 16, 37-51.
- Miller, J. (2009). *Like* and other discourse markers. In P. Peters, P. Collins & A. Smith (Eds.). *Comparative Studies in Australian and New Zealand English: Grammar and beyond* (pp. 317-337). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Miller, J. & Weinert, R. (1995). The function of LIKE in dialogue. *Journal of Pragmatics*, 23, 365-393.
- Müller, S. (2005). *Discourse markers in native and non-native English discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

- OED. Simpson, J. & Weiner, E. (Eds.). *The Oxford English Dictionary*. 2nd edition. Oxford: Clarendon Press.
- O’Keeffe, A., McCarthy, M. & Carter, R. (2009). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Östman, J.-O. (1982). The symbiotic relationship between pragmatic particles and impromptu speech. In N. E. Enkvist (Ed.). *Impromptu speech: A symposium*. (pp. 147-177). Åbo: The Research Institute of the Åbo Akademi Foundation.
- Partridge, E. (1984). *A Dictionary of Slang and Unconventional English*. 8th edition. Edited by P. Baele. London: Routledge.
- Popescu-Belis, A. & Zufferey, S. (2011). Automatic identification of discourse markers in dialogues: An in-depth study of *like* and *well*. *Computer Speech and Language*, 25, 499-518.
- Preacher, K. J. (2001). *Calculation for the chi-square test: An interactive calculation tool for chi-square tests of goodness of fit and independence*. [Computer software]. Available from <http://www.quantpsy.org>.
- Quirk, R., Greenbaum, S., Leech, S. & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London/New York: Longman.
- Redeker, G. (1991). Linguistic markers of discourse structure. *Linguistics*, 29(6), 1139-1172.
- Risselada, R. & Spooren, W. (1998). Introduction: Discourse markers and coherence relations. *Journal of Pragmatics*, 30, 131-133.
- Romaine, S. & Lange, D. (1991). The use of *like* as a marker of reported speech and thought: A case of grammaticalization in progress. *American Speech*, 66(3), 227-229.
- Sankoff, G., Thibault, P., Nagy, N., Blondeau, H., Fonollosa, M.-O. & Gagnon, L. (1997). Variation in the use of discourse markers in a language contact situation. *Language Variation and Change*, 9, 191-217.
- SBC. Bu Bois, J. W., Chafe, W. F., Meyer, C., Thompson, S. A., Englebreton, R. & Martey, N. (2000-2005). *Santa Barbara Corpus of Spoken American English, Parts 1-4*. Philadelphia: Linguistic Data Consortium.
- Scott, M. (2014). *WordSmith Tools*, Version 6.0. Oxford: Oxford University Press.
- Schiffrin, D. (1987). *Discourse markers*. Cambridge/New York/Melbourne: Cambridge University Press.
- Schourup, L. C. (1983). Common discourse particles in English conversation. *Working Papers in Linguistics*, 28, 1-119.
- Schourup, L. (1999). Discourse markers. *Lingua*, 107, 227-265.
- Siegel, M. E. A. (2002). Like: The discourse particle and semantics. *Journal of Semantics*, 19(1), 35-71.

- Stubbe, M. & Holmes, J. (1995). M. & Holmes, J. (1995). *You know, he* and other 'exasperating expressions': An analysis of social and stylistic variation in the use of pragmatic devices in a sample of New Zealand English. *Language & Communication*, 15(1), 63-88.
- Tagliamonte, S. (2005). *So* who? *Like* how? *Just* what? Discourse markers in the conversations of Young Canadians. *Journal of Pragmatics*, 37, 1896-1915.
- Tagliamonte, S. & D'Arcy, A. F. (2009). Peaks beyond phonology: Adolescence, incrementation and language change. *Language*, 85(1), 58-108.
- Taguchi, N. (2002). A comparative analysis of discourse markers in English conversational registers. *Issues in Applied Linguistics*, 13(1), 41-69.
- Traugott, E. C. (1995). The role of the development of discourse markers in a theory of grammaticalization. Paper presented at the Twelfth International Conference on Historical Linguistics, University of Manchester.
- Underhill, R. (1988). Like is, like, focus. *American Speech*, 63(3), 234-246.
- Zwicky, A. M. (1985). Clitics and particles. *Language*, 61(2), 283-305.

Appendices

Appendix 1: SBC

Appendix 2: ICE-CA

Appendix 3: ICE-GB

Appendix 4: ICE-IR

Appendix 5: ICE-NZ

The concordance lines can be found on the enclosed CD-ROM. Each Excel file contains two sheets. The first one is called *WordSmith Concordance list* and contains all the instances of *like* that were extracted with *WordSmith Tools' Concord* and subsequently disambiguated. The second sheet is labelled *Position & Function* and contains the 100 occurrences of *like* that were randomly selected, together with their function and position.

In addition, Appendix 1 (SBC) and Appendix 2 (ICE-CA) contain a third sheet called *Gender & Age* in which the 300 instances randomly selected from those two subcorpora were coded for speaker's gender and age.

