COMBINED CONSTRAINTS IN SPEECH PRODUCTION: EVIDENCE FROM LINGUISTIC DATA, ORAL POETRY, AND CULTURAL DYNAMICS

by

NICOLE MIRANTE Dottoressa in Lingue Orientali, Università di Venezia, 1994

A THESIS SUBMITTMENT IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR IN PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Comparative Literature)

THE UNIVERSITY OF BRITISH COLUMBIA January 2006

© Nicole Mirante, 2006

Abstract

This work describes a model of speech production based on the central role exercised by a speaker's working memory. It is proposed that speakers make intensive use of their working memory when planning, composing and uttering speech, and that a speaker's working memory is guided in its composition processes by an array of cooccurring cues, or constraints, which determine the selection of chunks of utterances in memory. The constraints are: semantic activation, imagery (i.e. the activation of detailed semantic, visual and spatial information), syntax, speech rhythm, prosody and sound repetitions. Speakers are exposed to the perception of environmental information and to others' speech, and these inputs determine the co-occurring activation and the selection of mnemonic data according to the constraints outlined. Evidence for the model is drawn from linguistic material, research on the cognitive psychology of oral literatures, and studies in social psychology and cultural information transmission.

The model stems from criticism that I direct to the concept of language as it is understood in modern linguistics. It will be shown that the assumptions on which current theories of language rest are at odds with recent developments in philosophy and communication studies. It will be argued that the proposed model is not only more theoretically sound, but also more adequate to describe speech as it is produced by real speakers.

İİ

ABSTRA	АСТ	••••	
TABLE	of co	NTEN	ITSIII
LIST O	F TABL	.ES	V
LIST O	F FIGL	IRES .	VI
ACKNO	WLED	GEME	NTSVII
INTRO	DUCTI	ON	1
CHAPTE	ER ON	E ♦ S	PEECH CONSTRAINTS AND THE COMPOSITION OF UTTERANCES
	1.	Тне	Social Coordinates of Speech
	2.	THE	MATIC CONSTRAINTS
		a.	Semantics25
		b	Imagery Constraints: Details in Conversation
	3.	Sou	ND CONSTRAINTS
		a.	Rhythmic Constraints: Speech Rhythm
		b.	Melodic Constraints: Prosody40
	•	c.	Sound Repetitions: Formulaicity, Repetitions and Echoes47
	4.	Syn	TAX AND A FEW WORDS ON HIERARCHY59
	5.	Sami	PLE ANALYSIS62
	6.	Sum	MARY: SPEECH CONSTRAINTS IN A NUTSHELL79
CHAPTE	ER TW	0 • 9	SPEECH CONSTRAINTS AND ORAL POETRY83
	1.	Defi	INING ORAL TRADITIONS
	2.	The	CONSTRAINTS OF AN ORAL PERFORMANCE
		a.	Meaning Constraints90
	•	b.	Imagery Constraints93
		c.	Sound Constraints95
		d.	Constraint Integration in Oral Performance

Table of Contents

iii

3.	А LOOK AT SOME DATA					
	a.	Counting-out Rhymes				
	b.	Traditional European Ballads114				
4.	SUN	IMARY: FROM ORAL PERFORMANCE BACK TO SPEECH				
CHAPTER THREE						
1.	THE DYNAMICS OF CULTURAL INFORMATION TRANSMISSION					
	a.	Charting the Patterns				
	b.	Explaining the Mechanisms148				
2.	Spe	ECH AND COMMUNICABILITY				
3.	SUN	MARY: CULTURAL TRANSMISSION AND THE PRODUCTION OF SPEECH				
CONCLUSION	۰۰۰۰۰ ۱					
BIBLIOGRAPHY						
APPENDIX O	NE 🔶	THE FIRST 250 MOST COMMON ENGLISH WORDS				
APPENDIX TWO A LOOK AT THE FUTURE						
1.	NEXT STEPS					
	a.	The Treadmill				
	b.	The Metronome				
	c.	The Echo and Syntax Lab216				
	d.	The Melody and Syntax Lab				

iv

List of Tables

Table	1.1	Nota	ati	on marks fo	r prosodic	info	rmation emp	loye	d by	Wennerstrom	(2001	l) in	
	her	r trar	ารด	cription		•••••				••••••		66	
Table	2.1	А	. (comparison	between	the	constraints	of	oral	performance	and	the	
	cor	constraints of speech									137		

List of Figures

.

,

Figure 1.1	Amplitude graph from Wennerstrom (2001, p. 51)	44
Figure 1.2	Pitch graph from Wennerstrom (2001, p. 51)	45
Figure 1.3	Pitch graph for "So, what other"	72
Figure 1.4	Pitch graph for "OK, so you're thinking"	73

Acknowledgements

I would like to express my most sincere gratitude to Dr. Eric Vatikiotis-Bateson (Linguistics, UBC), Dr. Geoffrey Winthrop-Young (CENES/Comparative Literature, UBC) and Dr. Mark Schaller (Psychology, UBC) for their patience and constant support in wielding this many-headed dragon.

Introduction

The purpose of this dissertation is to present a new approach to the study of speech production¹. This approach will emphasise the social dimension of speaking and the role that a speaker's memory plays in composing utterances. As I will illustrate below, society and memory are intimately connected to each other and to the act of producing language. However, their inclusion in a theoretical framework of speech production is a challenge to some of the most fundamental concepts in modern language scholarship. Below, I will introduce some of the Linguistic concepts I wish to examine by providing an abbreviated history of their development.

In 1911, an important shift took place in the study of language. Ferdinand de Saussure gave the third and most important of his lectures at the University of Geneva, which was published in 1915 as the *Course in General Linguistics*; Franz Boas published his *Handbook of American Indian Languages*; and Vilém Mathesius published the first call to action for what would shortly become the Prague School of Functional Linguistics (Sampson, 1980, p. 103). These three otherwise unrelated events mark a shift of focus from historical and comparative linguistics to synchronic linguistics. While nineteenthcentury scholars had been interested primarily in philological work, the twentieth

¹ In current linguistic literature, 'speech production' usually indicates the articulatory and/or acoustic processes involved in producing speech sounds. My use of the words 'speech production' is perhaps more akin to the current meaning of 'language production'. However, the reasons for my apparently inaccurate choice of terminology are deliberate; they will be clarified further in this Introduction, and especially at the beginning of Chapter One.

century saw a considerable development in research that pertained exclusively to the present state of a language. Historical linguists studied, and still study, the development of particular forms, sounds and structures through time. Synchronic linguists investigate the usage of a language as it exists at a specific moment in time. The shift of focus from historical to synchronic interests was so pervasive that modern linguistics studies are generally understood to be synchronic by default, unless otherwise specified.

In order to study the present state of languages, synchronic linguists have to resort to broad generalisations. Studying a language as spoken by a specific individual alone would be of limited interest: the goal, rather, is to study a language as spoken by an entire population of speakers. The three schools of thought founded in 1911, namely Structural Linguistics (Saussure), Descriptivism (Boas) and Functional Linguistics (Mathesius) produced different types of generalisations. As the name suggests, Descriptivists were mostly concerned with producing detailed descriptions of the ways in which a certain language was used; in most cases, their generalisations were limited to assuming that the utterances recorded from a small number of speakers were typical of an entire speaking population. The Descriptivists' theoretical generalisations about language use *per se* were cautious and few (Sampson, 1980, ch. 3). Functionalists produced generalisations about the types of sentence structures that can be produced in a given language and about their conceptual implications (Sampson, 1980, ch. 5).

In contrast, Saussure's structuralist concepts were highly theoretical, envisioning an exquisitely human "meaning faculty" by which people have a natural tendency to engage in interpersonal communication through the systematic production of signs. Saussure proposed the institution of a new science that would study all types of human communicative signs produced by this "meaning faculty"; he proposed to call this

science semiology. Linguistics, as a sub-field of semiology, studies only those signs that are part of a natural language. Saussure considered this particular type of signs the most important of all human signifying practices.

For Saussure, language as a system of signs is distinct from the language that each individual uses on a daily basis. Saussure's systemic view posits language as an independent entity, shared by all speakers but not fully owned by any one of them:

> Language exists in the form of a sum of impressions deposited in the brain of each member of a community, almost like a dictionary of which identical copies have been distributed to each individual [...]. Language exists in each individual, yet is common to all. Nor is it affected by the will of the depositaries. Its mode of existence is expressed by the formula:

> > 1+1+1+1...= I (collective pattern)

(Saussure, 1959, p. 19)

Moreover, his systemic view placed far more stock in the collective aspects of language use than in the practice of individual speech production:

What part does speaking play in the same community? It is the sum of what people say and includes: (a) individual combinations that depend on the will of speakers, and (b) equally wilful phonational acts that are necessary for the execution of these combinations.

Speaking is thus not a collective instrument; its manifestations are individual and momentary. In speaking there is only the sum of particular acts, as in the formula:

For all the foregoing reasons, to consider language and speaking from the same viewpoint would be fanciful. Taken as a whole, speech cannot be studied, for it is not homogeneous [...]. (Saussure, 1959, p. 19)

In other words, Saussure established a fundamental distinction between language as a system and language as individual production. The system enables individual productions. It is not important that these productions be oral or otherwise, since "what is natural to mankind is not oral speech but the faculty of constructing a language, i.e. a system of distinct signs corresponding to distinct ideas" (p. 10).²

Saussure's ideas on human systems of signification proved so fascinating that his theories spread like wildfire among scholars of linguistics, literary criticism and philosophy. Today we no longer think of language as "impressions deposited in the brain", but the fundamental differentiation between language as a general system and language as individual production has persisted. Chomsky has retained this distinction in defining "competence" vs. "performance": the first is a speaker's intrinsic "knowledge" of his/her entire language, the second is the way in which a speaker actually makes use of this knowledge in order to produce utterances. Notably, Chomsky makes the

² Saussure maintained that speech was an unimportant aspect of the systemic nature of language. However, at the same time, he also privileged speech as the most pure of all semiologic signs. I discuss this contradiction in more detail in Chapter 3.

competence/performance distinction in a discussion in which he equates Saussure's ideas with his own (1964, p. 10).³ In the same passage, he adds:

The actual use of language obviously involves a complex interplay of many factors of the most disparate sort, of which the grammatical processes constitute only one. It seems natural to suppose that the study of actual linguistic performance can be seriously pursued only to the extent that we have a good understanding of the generative grammars that are acquired by the learner and put to use by the speaker or hearer. The classical Saussurian assumption of the logical priority of the study of *langue* (and, we may add, the generative grammars that describe it) seems quite inescapable. (Chomsky, 1964, p. 10-1)

Thus, the two scholars agree that language as a system (*langue*) and language as individual utterances (*parole*) are distinct, that a pursuit of the first is more important, and that the second is too "heterogeneous" (in Saussure's terms) to be studied with much success. According to Saussure's philosophy, the nature of speech has no bearing on the nature of language: the only relevant feature of speech is the fact that it is a communal practice. This assumption has been carried over in Chomsky's view of language. Generative (Chomskyan) linguistics, the most prevalent stream of thought in modern language studies, pays little attention to the ways in which language is expressed. It makes no theoretical distinction between spoken and written language, since grammar can be found in both. Indeed, as I will argue later in this work, its focus

³ Sampson (1980, p. 49-50) points out some problems in Chomsky's comparison, but his argument is not important for the present discussion.

on grammar can be more easily traced to a study of written texts rather than oral production.

It is this supposed distinction between medium and message that I find untenable. Postulating a discrepancy between competence and performance, or between *langue* and *parole*, leads to imply that when an utterance is produced, two discrete entities come together: a message (i.e., language) and a medium (i.e., its spoken or written expression). The message is the "content" of the utterance; this content is language, structured by grammar. A speaking voice or a collection of written symbols is simply an expressive medium, detached from and subordinated to language. Voice and written symbols are only a secondary property, which bears no effect on the nature or formal organisation of language.

In effect, this model posits grammar as part of the content inside speech (or writing); grammar can be differentiated and extracted from speech (or writing). Defining grammar as a message seems rather strange: it is more intuitive to equate messages with semantics, with meaning. However, a grammar that uses different media (speech, writing) to get itself expressed while remaining unaffected by such media constitutes a kind of message. In this view, grammar is part of what gets communicated, while being quite detached from any particular form of communication. Conversely, a form of communication (speech, writing) that has little to do with the way in which the linguistic signal is organised is akin to a very transparent, impalpable medium.

By now, a few decades of cultural and media studies have shown that there is no such thing as a transparent medium; absolute form/content divides are nonsensical (McLuhan, 2001; Ong, 1987). The way in which something is said also determines what

is said: there is no possibility to escape the medium in order to obtain a "pure message". Pure messages cannot exist even in the most remote corner of our imagination – imagination itself depends on cognition, and cognition has very clear limits. Sampson (1980) agrees with this view:

> the aprioristic decision to consider language as pure form, divorced from the substance that realizes it, is mistaken; linguistic substance largely determines linguistic form. Our languages are the way they are in large part because they are spoken; any attempt to ignore the medium of speech and to analyse the nature of language in the light of pure logic alone is doomed to sterility. (p. 186)

However, Sampson's position is not widely shared by his colleagues – in fact, it is rare. Modern linguists, and especially generativist linguists, have supported the idea that there can be such a thing as a "pure message". In the generative case, it is "pure grammar". The generative brand of "pure grammar" sits somewhat above all other aspects of speech and has a causal and a shaping effect on linguistic production. This conception of grammar is Cartesian, because as in the Cartesian idea of the *homunculus*, a higher controller is envisioned to sit in the middle of other human intellectual capacities and direct them – although the *homunculus* himself remains mysterious and elusive, like Chomsky's Universal Grammar. Chomsky himself has embraced this Cartesian aspect of his theory (see, for example, his volume *Cartesian Linguistics* of 1966). The success of his idea is baffling for the simple fact that, in any other research discipline, a Cartesian view would be deemed most unscientific. Positing a *homunculus* means positing a higher faculty that will never be fully researchable or

explainable. This faculty is a sort of personal god – individual and discrete, yet intangible, unattainable and superior (Dennett, 1991).

In my view, the concept of "linguistic competence" is both unscientific and unnecessary. Of course, if we wish to investigate a certain phenomenon, we must make some generalisations in order to build our understanding of the phenomenon. It is also obvious that people acquire syntactic rules and use them regularly. Human beings are capable of reasoning, and this faculty is demonstrated on a daily basis by many academic disciplines as well as by common activities – game theory, mathematics, chess. There is no reason to refrain from making generalisations about the fact that people can infer syntactic rules by listening to others' speech, and that they can therefore apply those rules when producing their own speech. However, positing Universal Grammar as the absolute prime mover of our speaking and thinking, as Chomsky does, seems altogether far-fetched. The causative role of Universal Grammar as an underlying agent of language production is unproven. Most importantly, this concept is insufficient in explaining the daily production of speech. At best, the generalisations that pertain to syntactic rule behaviour describe part (and not even a large part) of the nature of speech.

If we re-examine Saussure's (and Chomsky's) assumptions, it seems obvious that their lack of interest in speech as a medium has brought them to make fleeting and superficial considerations. They both view speech as something garbled and haphazard, too unruly to allow for any generalisations. In Saussure's words above, the act of speaking depends on the will of the speaker, and every speaker is bound to have his/her

own individual will; hence, the speaking behaviour of an entire population cannot be summarised or systematised.

Again, much of modern cultural studies, media studies, psychology and discourse analysis provide ample evidence that this view of speech is debatable. To begin with, the purported "will" of individual speakers has been questioned on many fronts. From the transmission of concepts and practices (Dawkins, 1989; Blackmore, 1999) to the production of language (Dennett, 1991), from the proliferation of beliefs (Sperber, 1990) to the symptomatic acquisition of illness (Showalter, 1997), much research has shown that it is perhaps more accurate to view language, concepts and practices as forcing themselves into a person's psychology and behaviour, rather than to assume a wilful agency on the part of the doer/speaker. Many scholars (Boyd & Richerson, 1985; Cavalli-Sforza & Feldman, 1981; Rogers, 1962; Gladwell, 2000; Norenzayan & Atran, 2004; Rosnow & Fine, 1976; Allport & Postman, 1947) have demonstrated that the diffusion of a concept, an item of discourse or a cultural practice follows regular and predictable patterns. On the one hand, the regularity of these patterns again demonstrates that the agents responsible for the diffusion do not operate in a completely independent and wilful manner. On the other hand, this regularity points to communicative practices that must themselves be regular and at least partially predictable.

In fact, many studies in discourse analysis and in linguistics have illustrated that much of the behaviour of the participants in a conversation can be predicted. Speech is also routinely predicted in everyday circumstances, when interlocutors prompt a certain response or finish each other's sentences. Based on these observations and on a variety of studies and experiments, I wish to offer some new considerations on the nature of

speech as a medium. I believe that the production of speech is organised in a systematic manner, to the point that all utterances may be almost entirely predictable.

In order to examine speech, we need to determine what kind of resources it possesses as a medium. Clearly, these resources are extensive. In terms of overt behaviour, speakers produce aural signals. They also produce visual signals (including facial movements, facial expressions, gestures and body movements) in the case of face-to-face or video-mediated exchanges. Hence, a speaker must possess the physiological and psychological skills to process and understand these signals. Covertly, a speaker must also make extensive use of his/her memory and reasoning in order to understand another speaker's meanings and respond appropriately. The process of speaking is quite complex, in terms of both speech perception and speech production; the resources that are employed at each stage of the speaking process are numerous and intertwined.

At present, my interest lies specifically in investigating the covert activity that goes on when a speaker composes his/her speech production spontaneously. In other words, I will examine only a small fraction of the entire speaking process, the portion that pertains to those precise moments in which a speaker starts to produce an utterance. Admittedly, this is a significant limitation, since I will disregard all the complexities of visual and aural perception that an interlocutor must process in the course of a conversation. Despite this limitation, I believe that it makes sense to consider in relative isolation the way in which speakers organise and compose utterances, because this act of production invariably occurs – in fact, it can take place even when a person is alone and has no immediate perception of others' speech.

Since I wish to examine specifically the organisation and composition of utterances as they are produced, it is important to consider the "workspace" in which the process of composition takes place. A likely hypothesis is that this space is a speaker's working memory, which is, in effect, the primary resource that the medium of speech accesses when organising production. Hence, appraising the contents of working memory should be a fundamental step when attempting to predict production. In this appraisal, the analyst must consider the social situation in which a speaker is situated, the utterances that other speakers have produced, as well as the production history of the speaker in question.

Keeping these considerations in mind, I propose to examine the mechanisms that lead to the selection of individual chunks of data in working memory and therefore to the composition of utterances. Several parameters of utterance production must be taken into consideration when making any predictions on the outcome of the composition process. Some of these parameters are of a social nature because a community of speakers follows them almost as if they were strict social norms. I am referring specifically to intonation patterns, conversational rhythm and syntactic structures, which are employed regularly and predictably by all speakers of a language. Other parameters, like semantic activation, are determined by human psychology and the ways in which the brain clusters information into groups and hierarchies.

My aim is to attain a list as comprehensive as possible of probable candidates for those parameters that guide the production of utterances. I call the individual parameters "constraints" and their sum "combined constraints". The term constraint points specifically to a limitation placed on the contents of working memory in order to

allow the selection of a particular word or phrase over others. In this sense, the limiting action of speech constraints is a productive action, because it enables a process of selection that would otherwise be impossible. Hence, speech constraints are akin to helpful mnemonic cues; they should be understood as facilitating agents rather than mere restrictions.

The data on which I base my argument are quite diverse and will have varying relevance for different audiences. Readers should not be discouraged if they find that some of the content is rather alien to them, as these materials are rarely examined together. I will delve into the details of linguistic analyses, literary analyses, psychological research and cultural theory; although readers from both the social sciences and the humanities will be exposed to unfamiliar material, the thread of the argument will remain consistent and unbiased towards any one discipline.

Specifically, in Chapter One the language scholar will find a detailed description of the constraints at work in speech production, as I envision them. The theoretical considerations are my own, and I base my research on data collected through a variety of studies in linguistics and discourse analysis. The constraints I describe are six: semantic activation, imagery, speech rhythm, prosody, repetition (of phrases, words and sounds) and syntax. I close the chapter with my analysis of a segment of conversation. In the analysis, I highlight all the constraints at work and demonstrate the predictability of certain utterances within the conversation.

In Chapter Two, the literary scholar will see that the linguistic considerations of Chapter One are intimately connected to one of the most successful theories in the study of oral traditions: the theory of oral-formulaic composition proposed by Parry and

Lord. I will provide a brief account of the theory and of the explanation it offers regarding certain recurring features in the compositions of oral poets. Most importantly, I will consider the research conducted by Rubin (1995), a psychologist who has explained the theory in terms of cognitive constraints upon the oral poet's working memory. Since the features of oral compositions that Rubin examines are extremely similar to the features of everyday speech, I will argue that the constraints of oral poetry and those of common speech are equivalent. Rubin's work is further evidence that a process of constrained production is both possible and naturally occurring in singers and speakers. The chapter ends with the analysis of several oral texts from two distinct traditions: children's rhymes and European ballads. As in the previous chapter, these analyses aim at demonstrating the ubiquitous nature of cognitive constraints in the act of oral composition.

Finally, Chapter Three is intended for scholars interested in media theory, social psychology and cultural studies. The chapter supports the constraints argument with "circumstantial evidence" that comes primarily from studies in social and cultural information transmission. The studies illustrate universal patterns of social communication that are largely recursive. I call this evidence "circumstantial" because the research I examine here targets not the production of utterances, but rather the transmission of cultural information in general. As my sources will illustrate, such recursive patterns of communication are produced, on the one hand, by recursive psychology, and on the other hand, by recursive communication dynamics. I will close this part of the chapter by arguing that the recursiveness of communication is thoroughly consistent with the constraints argument of Chapter One.

I end the dissertation with a Conclusion, in which I offer a short recap of the combined constraints argument. The Conclusion will also include further discussion on the nature of speech and writing, and on the perils of underestimating their differences.

This work is a first theoretical step towards the study of cognitive speech constraints. The subject is far too intricate to be exhausted here, and its experimental nature ensures that much about it is yet to be discovered. The constraints I will articulate are disparate and the theoretical framework surrounding each of them should be supplemented with more research in linguistics, discourse analysis, psychology, cognitive science, cultural studies, media studies and many other fields. Since the phenomenon of speech production is complex, it seems natural that the theory that accounts for it will also be complex. However, I believe that the difficulties inherent in the subject should not deter any earnest attempts to understand speech production mechanisms, and that such an understanding can lead to more accurate predictions than those based on syntax and semantics alone.

Chapter One

Speech Constraints and the Composition of Utterances

The study of speech and language production includes a variety of related areas of specialization. Two of the main areas of inquiry are psycholinguistics and articulation. Studies on articulation centre on the mechanics surrounding the movement of speech organs and on the neurological mechanisms that control such articulation (see for example Fadiga et al., 2002; C.A. Fowler, 1986; Guenther et al., 1999; Houde & Jordan, 1998; Jones & Munhall, 2000; Kuhl & Miller, 1978; Liberman & Mattingly, 1985; McGurk & MacDonald, 1976; Meltzoff & Moore, 1977; Rizzolatti & Arbib, 1998; Sams et al., 1991; Stevens & Blumstein, 1978). This particular area of inquiry is usually referred to as speech production. On the other hand, the area investigated by psycholinguistics, which includes the storage, retrieval and organisation of linguistic structures in the brain, is usually called language production. This terminology is revealing; it demonstrates once again an *a priori* duality, by which language is understood as an abstract construct that precedes material expression, and speech is understood as expression that has hardly any recourse on the organisation of language per se.

This terminology is not appropriate for the present study. While my area of research is closest to psycholinguistics, the object of my research is the composition and production of utterances. Hence, this is a study in speech production, although its limitations will force me to ignore all the intricacies of articulation. More specifically, the present study aims to address some of the shortcomings that are inherent in the current foundations of psycholinguistics. For this purpose, I base my survey of psycholinguistic research on two anthologies: the Butterworth anthology published in 1980 and the Wheeldon anthology of 2000. The articles included in the two anthologies pursue a great variety of research questions. However, two primary approaches to research are identified. The principal methods involve the production of single words in experimental settings, and the study of speech errors, especially "slips of the tongue" (see also Cutler, 1982). These two methods have been employed primarily to investigate the mental lexicon (i.e., the representation/storage of words in the brain) and its morphological, semantic and phonological attributes.

A number of theoretical views on the production of speech have emerged from psycholinguistic research. In the Butterworth anthology, the editor concludes the volume by proposing a theoretical umbrella under which the articles can be summarised. In contrast, most authors in the Wheeldon anthology, including the editor, adhere to the production model hypothesised by Levelt in his *Speaking: From intention to articulation* (1989). Since the Butterworth model is comparable to Levelt's, I will provide a brief explanation of the latter only.

Levelt's model posits a series of mental modules. The first, called Conceptualizer, produces a "communicative intention" that is not yet expressed in natural language; the intention is manipulated by the Formulator module, in which it undergoes first grammatical and then phonological encoding; finally, a phonetic and articulatory plan is devised in the Articulator module, and the utterance is produced. The "communicative intention" initiated by the Conceptualizer can be described as a combination of both pragmatic intention and semantic (but pre-lexical) content. The Formulator works in parallel with the lexical storage and retrieval system, so that words are picked from the

mental lexicon and immediately manipulated into a grammatical and phonological structure (see also Nickels & Howard, Roelofs in Wheeldon, 2000).

The theoretical assumptions of most of the Butterworth and Wheeldon articles adhere to the same general processing order found in Levelt's model: (a) a semantic/pragmatic "intention" is formulated, which then undergoes (b) lexical selection, (c) grammatical structuring and (d) phonological structuring, before being voiced in (e) articulation. Butterworth briefly considers the possibility of a different succession of processing phases (p. 452-3), but he quickly discards the thought after stating that the evidence in favour of other models is "no more than hints" (p. 453), despite the fact that some of the research in his own anthology (e.g. Schenkein's article) provides a number of such hints.

This model of speech production is problematic. The theoretical, almost philosophical presuppositions hidden within its chain of production processes deserve to be questioned. In particular, the originator of the "communicative intention" is suspect. The nature of this intention and the entity that produces it (i.e., the subject that intends) are left implicit and unexplained. This model of behaviour exalts individual action and relegates social influence to the background. The Conceptualizer is, in effect, a "pure agent" that is free to create (to have intentions) in complete autonomy. However, much of modern theory and philosophy – from Durkheim and Halbwachs to Dennett, Maturana & Varela, Derrida, Dawkins and many more – have converged in establishing the fundamental unfeasibility of concepts such as that of an autonomous subject. They have instead stressed the temporal and causal primacy of social structures over individual identity and behaviour. While the speech production of a subject caught in an experimental situation may well resemble the uncomplicated behaviour of a free

agent, this perspective does not reflect the processing of speech in the real world. In everyday circumstances, a speaker's memory is awash in environmental data and his/her production depends on that data. Social input precedes individual output; this fact is plainly observable in language acquisition, but it is also observable in everyday conversation. For example, people constantly repeat each other's utterances. People also strictly observe social speech norms every time they employ their language's grammar, typical intonation contours, acceptable conversation topics, etc. However, the social dimension of speech is hardly ever observed in production studies:

> The vast majority of the speech we produce is conversational; however, this form of speech receives scant attention in the production literature. One reason for this is that spontaneous speech is messy – filled with paralinguistic phenomena such as stops and starts, ums, and long pauses. (Wheeldon, 2000, p. 5-6)

We are therefore brought back to Chomsky's observations, and Saussure's, that speech is too heterogeneous to be investigated. Hence, although the ostensible focus of many psycholinguistic studies is the production of speech, we should question whether this is indeed the case, or whether these studies are yet another attempt at shunning "performance" for the sake of investigating "competence".

The purpose of this study, then, is to look at the social dimension of speech production in order to provide an alternative in which "pure agents" are not contemplated. Specifically, I will consider a speaker's memory as the repository of various features (lexicon, grammar, prosody, etc.) that are expressed in an utterance. I will posit environmental stimuli, including others' speech, as the origin of mnemonic

cues that trigger the composition of utterances in a speaker's memory. My aim is, ultimately, to determine some of the parameters (constraints) that prompt the mnemonic recall of the elements of speech.

This view of speech production eliminates "pure agents" because recall, and therefore speech, is mostly prompted by perception – including the perception of others' speech. The subject matter of an utterance is generally suggested by environmental cues and by socially acquired habits. However, unlike in a behaviourist model,⁴ recall is not a direct point-to-point equivalence, where, if a person hears "A," s/he will always respond with "B". The process of recall that I will illustrate is much more complex. The recall of speech features will be described as guided by several simultaneous constraints, so that very recent data (i.e., the perception of a content word) is blended with long-standing speech habits or patterns (i.e., habitual preferences in lexical combinations) in ways that are often novel. This type of recall is virtually indistinguishable from a more spontaneous and creative view of speech composition, although it does allow for some measure of prediction. Six simultaneous constraints will be described, and this initial list may well be expanded in the future.

Three primary advantages arise from considering speech composition as the simultaneous recall of distinct speech features. First of all, this view takes into great consideration the social dimension of speaking, to the point that this dimension is invested with a sort of distributed agency. Second, this perspective accounts for the everyday normative dynamics of speech, while other views that posit autonomous "pure agents" would have a hard time explaining the regularities observable from speaker to

⁴ See Sampson 64-9 for a critique of behaviourist enquiry.

speaker. Third, this view values every aspect of speech as invested with potentially creative power: the semantic dimension, the grammatical structure and the sound qualities of speech are all equally important in shaping the composition of utterances.

Below, I start with some general considerations on the placement of speakers in social communities and the placement of utterances in social conversation. I continue by describing those aspects of speech that should be considered as shaping constraints in the cuing of mnemonic material and in the composition of utterances. I conclude the chapter with the analysis of a conversation in order to demonstrate the predictability of some utterances within the conversational context.

1. The Social Coordinates of Speech

When a person opens his or her mouth to speak, a variety of factors determine what utterances he or she will produce. The most general set of factors derives from the current socio-geographical placement of the speaker, who will use the language presently spoken by his or her group. This group is determined by geographical location, age, sex, employment, social class and other characteristics. The speaker will also possess a speaking history. Hence, the speaker may have picked up expressions, prosodic patterns or other characteristics from his/her childhood, or may have acquired a habit of using certain expressions or patterns over time.

A more localised set of factors is determined by the placement of each utterance. The speech may occur in a face-to-face conversation, a monologue, a mediated conversation like a videoconference or a phone call, and so forth. It may be directed at one or more interlocutors, or at silent listeners, or at nobody at all. It may have several pragmatic values, such as conveying a disposition and/or a piece of information. Finally, it may be completely spontaneous, even involuntary, or it may have been rehearsed at length.

These two orders of factors – speaker placement and utterance placement – influence the behaviour of all speakers and the nature of all utterances. It is highly unlikely that any competent speaker will produce an utterance inconsistent with his or her speaking history or incompatible with the speaking situation at hand. Since it is possible to acquire a good knowledge of a speaker and of his/her speaking situation, predicting a speaker's output should prove a relatively simple task. This is the case, for instance, when good friends have a conversation; often, their interactions are so intimately understood that it is normal for one to prompt a certain response on purpose, or to finish the other's sentence.

Nevertheless, scholars have yet to develop a model for predicting the production of specific utterances. The issue has been attacked from many investigative angles, many of which have produced excellent observations and results. However, given the complexities that determine the placement of a speaker and of his or her utterances, it is highly unlikely that one order, or type, of predictor alone will be sufficient to specify a speaker's output in any given situation. Moreover, the predictive models thus far proposed by Linguists, from Generative Syntax to Systemic Functional Grammar, are rule-based systems appropriate for speakers with massive computational resources and no restraints on their utterance output; this, however, is clearly not the case with real speakers.

While speakers do have the capacity to apply syntactic rules, the fast pace of an everyday conversation denies a speaker the luxury of computing thousands of utterances to come up with the one that best expresses the thought at hand. More

importantly, in order to achieve communication within a conversation, speakers must value consistency and continuity over ingenuity or originality. Spoken communication is achieved much more easily when well-known words are used – often repeatedly – to express an idea similar to that expressed in previous utterances; conversely, it is extremely difficult to understand somebody who uses unfamiliar words to express a variety of unfamiliar and unexpected concepts. Communication, especially spoken communication, is redundant. It entails expecting, and perhaps predicting, at least part of the utterances that will be produced by other speakers. This continually elastic bond between the productions of speakers and the expectations of listeners ensures the comprehension of utterances as well as the affective (emotional) participation of the interlocutors. Many of the examples I provide below illustrate these principles in further detail.

When considering the behaviour of speakers, it is important to keep in mind that, aside from a limited set of cognitive mechanisms (e.g. rhythm-keeping) and rule-abiding computational skills (e.g. syntactic skills), all spoken communication must make extensive use of memory. Speakers use a pre-existing vocabulary that they have memorised, and pre-existing speaking patterns that they share to varying degrees with the other speakers of the same language. Speakers also share somewhat similar mnemonic abilities and limitations. For these reasons, speakers ensure the success of their communicative efforts by resorting to commonly used patterns and vocabulary, and by making good use of the constraints of human memory. Therefore, in envisioning a group of factors that determine the predictability of speech, I believe it is paramount to include those mechanisms that make use of memory and mnemonic limitations. Any

factor that constrains the use of memory in speech must be considered as a predictive factor in speech output.

In my view, there are several constraints in the serial composition of utterances that together can predict the output of a speaker in a certain situation. The first type of constraint is *thematic*. By this I mean not only that the interlocutors in a conversation need to keep on track in terms of subject matter, but also that the specific terms and expressions they will select for production will be determined in a semantically continuous fashion. Under the thematic umbrella I include both thematic constraints proper and imagery constraints.

Sound constraints are the second type I will illustrate. My first argument here is that the perception of certain words or sounds acts as a mnemonic prompt, and therefore increases the probability that similar words or sounds will be produced. A second type of sound constraint can determine the selection of certain utterances based on their length, stress pattern, and prosodic contour; this is the case when an utterance must obey a prosodic pattern or maintain a rhythm. Both of these sound constraints can be prompted by other speakers, or by an individual speaker's own utterances.

The third type of constraint is *syntactic*. I argue that syntax is important in determining the order of words in an utterance. However, I also challenge the supposed primacy of syntax among the predictive constraints of speech, and this observation leads me to discuss matters of hierarchy in a system that includes so many heterogeneous constraints.

Each of the three constraint types contains one or more predictive mechanisms. The individual predictors I envision are: semantic theme, imagery, rhythm, prosody, repetition and syntax. In the sections that follow, I will describe each of these

constraints. Some of them have been investigated at length by entire schools of researchers, at times to the point of being fractioned into sub-fields, while others have been merely identified by a few scholars. For example, thousands of studies exist on the many facets of semantics, while the research on speech rhythm has been less prolific, and that on imagery is even less documented or detailed. Some discourse analysts have investigated several of these constraints, and therefore I will refer to their studies more than once; other data will come from completely unrelated areas of linguistics and psycholinguistics. Thus, while I strive to limit my research to linguistic studies, my material remains extremely heterogeneous. My sources are disparate and have evolved differently according to the interests, expertise and politics of each research area. To my knowledge, the bulk of this material has not previously been examined together, and it is challenging to bring these studies in conversation with each other, especially as no single piece of data can truly be regarded as definitive. Progress in all areas of research has been and remains constant, and the studies I examine here are best viewed as examples.

Despite the difficulties, this puzzle is worth piecing together; the bigger picture can yield results and predictions that the more detailed studies are unable to encompass. Below, I investigate each of the individual constraints in detail. Towards the end of this chapter, I return to the more general picture of constraint integration and I examine the overlapping constraints at work in a sample of transcribed conversation.

2. Thematic Constraints

a. SEMANTICS

Semantics as the study of meaning has developed in many different directions. Within the many intricacies that the study of meaning entails, I am primarily concerned with the development of thematic associations and sequences. Themes are studied in psychology under the name of *schemata*, and many models such as scripts, story grammars and associative networks have been proposed to instantiate schemata in memory. But there are also many studies in linguistics that investigate the way in which language progresses along a certain semantic path. In fact, the study of linguistic semantics and that of schemata frequently overlap, because schemata are commonly probed through associations of words.

In Linguistics, a web of lexical associations is called a semantic network (Jackendoff, 2002). These networks are represented as graphs in which each word is connected by a line to other words in the graph. One of the largest semantic networks in existence is WordNet, maintained by the Cognitive Science Laboratory at Princeton University. Another interactive example is the Visual Thesaurus, a commercial application that takes advantage of the ease of visual organisation to facilitate navigation among related words.⁵ Navigating these networks, either online or on a printed page, is simple and fun – and tremendously misleading. As representations of knowledge in general, these networks are very effective. However, when it comes to

⁵ The URLs for both WordNet and the Visual Thesaurus are provided in the Bibliography. The reader is encouraged to explore them.

representing lexical knowledge as instantiated in the brain, the depictions are much too simple. Semantic networks are often used as straightforward illustrations of the way in which lexical access is achieved: the way, for instance, in which the concept of "book" may facilitate access to the concept of "page" (as opposed to the concept of, say, "cactus"). However, matters are not quite as easy as tracing a line, where brain mechanisms are concerned:

> One often sees lexical access characterized simply as something like 'activating a word node' in memory. This is far too crude, and calls for several refinements.

> First of all, [...] there is no single 'word node'. A lexical item is a complex association of phonological, syntactic, and semantic structures. I stress here *structures*, not just *features*. One cannot just think of a word node as activating a number of other nodes that together constitute a collection of independent features. Rather, any adequate account has to include provision for hierarchical phonological structure in complex morphological items, for hierarchical syntactic structure in idioms, and for hierarchical conceptual structure in just about any word [...]. (Jackendoff, 2002, p. 205)

Put in simpler terms, each lexical entry in long-term memory includes phonological (sound), syntactic and semantic information; each of these types of information has its own set of associations. Moreover, the associations are hierarchical, so that a certain type of information and a certain subset of associations are more likely to be accessed than others.

Despite these complications, lexical access is routinely achieved in both the perception and the production of speech. In experimental settings, lexical access has been widely demonstrated by a wealth of experiments on priming (Rubin, 1995; Jackendoff, 2002). Priming refers to the preferential access that a word extends to other words that are semantically related to it. It translates into the fact that, if you hear or read the word *bug*, for example, your response time in deciding that *insect* is a word will be shorter than if you had to decide about *insect* without reading or hearing *bug* first.⁶ Similarly, if you have been mis-primed, i.e., if you read or heard *table* and then the semantically unrelated word *insect*, your reaction time will not be affected by the priming of *table*. Moreover, frequently used words seem to be in a constant state of being primed, because they will be recognised more quickly than infrequent words.⁷

Many experiments have documented the effects of priming in lexical decision tasks. A lexical decision task involves the following procedure. Imagine that you are sitting in a chair in front of either a computer monitor or a loudspeaker. You may read a sequence of letters displayed on the monitor, or hear a sequence of sounds coming from the loudspeaker. You will be required to signal in various ways (e.g. with a button press, or by pulling a small lever) whether you recognise the sequence of letters or sounds as an actual word. You have therefore accomplished the task of deciding whether the

⁶ This example is taken from Jackendoff 209-10.

⁷ However, note the findings by Balota & Chumbley. Although they do not deny the commonness of word frequency effects, the authors also suggest that the nature of lexical decision tasks may skew results to the point of reporting a word frequency effect that is disproportionately large when compared to results from other types of experiments.

sequence you read/heard matches an existing entry in your lexicon (hence, lexical decision task).

This experimental paradigm has been used in countless different scenarios, where variations of the task are often correlated to response times in order to probe the semantic, phonologic or syntactic organisation of the subject's lexicon. Within this and other paradigms, the mechanism of semantic priming has been vastly documented. What is important for the present argument is that semantic priming in lexical access takes place in production as well as in perception. When we produce an utterance, many different concepts (and words) become activated, so that even if only one word is uttered, many more words are on the threshold of consciousness. Ray Jackendoff (2002) describes this mechanism:

Suppose the conceptual department of working memory contains some thought that the speaker wishes to express. [...] The initial event has to be a call to the lexicon: what words potentially express parts of this thought? [...]

[...] it appears that the call to the lexicon results in the activation of a variety of lexical items of varying degrees of appropriateness for the part of the thought in question. In present terms, this can be thought of as binding the candidate items to working memory, where their conceptual structures compete for integration into the thought in conceptual working memory. At some point the best candidate wins the competition (the competition is resolved), based among other things on its match to the intended thought. This last step is standardly called 'lexical selection' in the production literature. (p. 212)

Therefore, lexical selection is the production of lexical activation, but while lexical selection is serial (one word at a time is selected and uttered), activation entails "a variety of lexical items". This means that lexical selection is likely to progress according to previously activated semantic associations: we speak along the lines of what is presently in working memory, and semi-activated lexemes are more likely to make their way into working memory than inactive lexemes.

Wallace Chafe (1994), in particular, proposes the hypothesis that the contiguous production of utterances provides a roadmap for the travels of our conscious focus, which moves between activated and semi-activated information. Every intonation unit⁸ we utter is a step in the roadmap of a semantic network. He posits that the intonation unit "verbalizes the speaker's focus of consciousness at that moment" (p. 63). The information contained in the intonation unit is active at the time of the utterance, but "the active focus is surrounded by a periphery of *semiactive* information that provides a

⁸ Chafe (1994) defines the intonation unit as the amount of speech that a speaker produces between breaths. This is not a new concept: Wennerstrom (2001) reports a list of other scholars that have given different names to the same phenomenon (p. 28). Wennerstrom herself uses the term 'intonational phrase'. But note that Chafe is quite specific in quantifying the intonation unit by the number of words and concepts it contains. Apparently, the typical English intonation unit contains four words, but "it is important to realize that this figure is valid for English only; languages that pack more information into a word show fewer words per intonation unit [...]" (Chafe, 1994, p. 65). I disagree with this quantification. A quick scan of the examples reported by Chafe will reveal that each of them contains two accents (two rhythmic beats); perhaps rhythmic structure may provide a more reliable measure than the debatable quantification of meaning.
context for it" (p. 29). When a sequence of utterances is produced, we can follow the focus of consciousness as it moves restlessly from active to semiactive – but newly activated – information (p. 29).

Put simply, lexical activation is a constraint on the production of speech because it skews the chances of certain lexemes being selected over others when a sentence is composed in working memory. Therefore, the serial selection of lexemes is somewhat constrained by previous activations. This mechanism ensures the coherence of utterances, so that (1) a sentence is made up of words that can be used together, and (2) several sentences are strung together in a semantically meaningful manner.

This, of course, does not mean that we can never change the subject. It also does not mean that if we "wish to express" something (per Jackendoff's terminology above), the words will automatically pour out of our mouths in faithful succession. The question of meaning is much more complicated. As many linguists – especially those working in the Functionalist tradition – have repeatedly pointed out, meaning is a matter of context. So, let us suppose that you owe me a lot of money. I may show up at your house, unexpectedly and uncharacteristically, and engage you in a conversation about the weather. The fact that I mean to tell you that I want my money back is never put into words, but it is clearly conveyed by my decision to visit you. However, as we discuss the weather, my sentences and yours must still be tied together by some measure of coherence. As we talk about our winter, all sorts of words regarding rain and low temperatures will sit in the periphery of our conscious focus, pushing to be uttered. This is where the constraints that I am discussing will come into play: the larger frame is determined by a number of factors, but the step-by-step, sequential frame of sentence

production is determined heavily by the semantics of my words and yours, among other factors.

b. IMAGERY CONSTRAINTS: DETAILS IN CONVERSATION

Imagery is commonly observed in everyday speech. People normally tend to "paint a picture" of what they are discussing, enriching their descriptions with concrete and meticulous details as these are usually preferred to abstract terms. This is especially true when narrative is involved: no matter how short the episode or how long the tale, details are key elements in storytelling.⁹ Often, dialogues are part of such stories, and when interpersonal interaction is described, dialogues are routinely reported in direct speech. This is the most detailed account that can be offered of a dialogue, because it specifies each word that was employed. And yet it is unlikely that the storyteller remembers the dialogue word by word; in fact, direct speech dialogues are usually fabricated (Tannen, 1989, ch. 4). Nevertheless, dialogues are narrated as though they were taking place at that very moment in order to convey immediacy: the rate of gesticulation increases, the tone and the quality of voice changes, and the narrator is momentarily transported to another place and time.

In her book *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse* (1989), Deborah Tannen stresses that one of the primary goals of conversation is to create emotional involvement between the participants. She discusses

⁹ It should be noted that in everyday conversation, storytelling does not need to take the shape of a fairy tale recital; relating the small events of one's day to a friend is also an instance of storytelling.

several strategies that aim at creating involvement: some of these strategies are based on sound, some others on meaning. I will return to the sound strategies later in this chapter. With regard to the strategies based on meaning, Tannen (1989) writes: "The strategies that work primarily (but never exclusively) on meaning include (1) indirectness, (2) ellipsis, (3) tropes, (4) dialogue, (5) imagery and detail, and (6) narrative" (p. 17). Out of this list, the last three items are emphasised as the most important. In particular, chapter five of Tannen's book is entirely devoted to discussing the use of imagery in conversation. The author posits that descriptions or references that use details, rather than abstract notions, create more involvement because they are conveyed as being affectively significant. Details communicate a concept while at the same time offering a sense of intimacy: "[...] specific details trigger memories that trigger emotions" (p. 150). For this reason, speakers often ground their narrations, descriptions or other types of talk by offering details about their subject matter. Often, speakers telling a story appear to scavenge their memory: they insist on determining the name of an old acquaintance, the location of a certain place, or other unimportant information. This searching of one's memory lends to the verisimilitude of the speaker's story; it also displays the common process of struggling to remember a small fact -aprocess with which the listener can relate (p. 140-1). And finally, details add to the memorability of tales, and therefore are likely to be repeated (p. 165).

In other words, the details of imagery are precious to the art of conversation. This fact leads to the conclusion that concrete and specific lexemes are routinely preferred to abstract and generic lexemes. In terms of the semantic structures discussed in the previous section, this preference for detail means that the selection of a lexeme will fall on the lower end of a hierarchical semantic network more often than on the upper end.

A hierarchical network, or tree, is a particular type of semantic network. It illustrates the concept of semantic inheritance,

whereby a more highly specified or highly structured item 'inherits' structure from less specified or less structured items. [...] Inheritance is actually a special case of taxonomic categorization, often expressed in terms of a semantic network along familiar lines like (36)

(36)



[...] Semantic networks in the literature usually express a strict taxonomy like (36). But it is also possible to set up networks with multiple inheritance [...]. (Jackendoff, 2002, p. 184)

In terms of the constraints I am discussing, the imagery constraint simply dictates that as the composition of utterances progresses along a semantic path, lexemes lower (more detailed) in the semantic hierarchy will be preferred to those that are higher (more abstract). So, as I am sitting in your living-room talking about the weather instead of the money you owe me, I am likely to string together coherent utterances that contain (a) weather-related lexemes, (b) detailed rather than abstract lexemes, and (c) common rather than uncommon lexemes (due to the word frequency effect). Therefore, I am likely to talk about rain instead of precipitation, and about clouds instead of condensed atmospheric vapour; I am also likely to offer an illustration of the cold weather by discussing items of clothing, or other observations that are a result of the circumstances. You are likely to respond by using words that are closely related to mine, since "perceived words can prime production" (Jackendoff, 2002, p. 213, footnote), and by offering more details of your own. Therefore, if semantics has already circumscribed the selection of words that you and I are likely to use, imagery will restrict that choice even further, and will also skew the probabilities in favour of using one word rather than another in the same semantic domain.

3. Sound Constraints

a. RHYTHMIC CONSTRAINTS: SPEECH RHYTHM

The rhythmic nature of speech has been observed in a variety of different studies. In discourse analysis, for example, scholars have documented the occurrence of rhythmic phenomena in many languages and many speech settings (e.g., monologues, public readings, lectures, face-to-face conversations, telephone conversations). Speech rhythm provides a temporal organisation to the production of utterances by a single speaker. It also helps in coordinating dialogue by timing turn-taking among two or more speakers, so that speakers synchronise the rhythm of their utterances with one another. In this way, pauses between turns are timed according to the cultural conventions shared by the group. For example, in English, "appropriate" pauses are generally short and longer pauses are immediately recognised as awkward.¹⁰

¹⁰ Note that in-group variations regarding the "appropriate" length for pauses are also common. An example from Tannen on p. 42 below illustrates this phenomenon quite clearly.

Utterances in all languages are perceived as rhythmically aligned. In English, this alignment is perceived as a regular placement of stresses, or pitch accents.¹¹ The rhythm of an English utterance looks something like this:

We're / all in/tuitively fa/miliar with the i/dea of a de/rivative.

This utterance, recorded and transcribed by Wennerstrom (2001, p. 51), comes from a lecture on mathematics. The accents show the places of stress, where the speaker's pitch rises noticeably and quickly, before dropping back towards the baseline of the

¹¹ Not all languages can be thus described. Languages are generally distinguished into three rhythmic classes: stress-, syllable- and mora-timed languages (Abercrombie, 1967; Bloch, 1950). Stress-timed languages such as English and German are thought to employ a regular pattern of stressed syllables; syllable-timed languages, i.e., Romance languages such as Spanish and Italian, are thought to employ a regular pattern of syllables with an equal duration; finally, Asian mora-timed languages such as Japanese and Tamil are thought to employ a regular pattern of morae. (The actual definition of "mora" is somewhat debated. However, typically morae can be described as short, sub-syllabic sounds.)

This categorisation has been repeatedly challenged: while on the one hand experiments like those of Nazzi et al. (1998) have proven that the perception of speech rhythm is a reality, careful measurements on the acoustic signals have failed to show any consistent isochrony between stresses, syllables or morae. Somewhat recently, Ramus et al. (1999) and Shukla et al. (n. d.) have offered an interesting hypothesis by positing that the perception of rhythm may be a product of the placement of vowels. However, consensus has yet to be reached on this or other theories.

For an interesting aside, see also the Patel and Daniele study (2003), which used the Ramus et al. methodology to demonstrate some cultural correlations between speech and musical rhythms.

speaker's range. The slashes identify the metric feet, or the groups of syllables that cluster around a stress and make up a beat. As shown in this example, English metric feet are usually structured in a stressed-unstressed manner, so that the accent falls on the beginning portion of the foot.

Now, let us examine this example a little more closely. It is easy to observe that the beats of the utterance fall exactly on the correct place of stress for each word: "intuitively" is normally pronounced with a stress on the u, "familiar" with a stress on the first i, and so forth. The speaker did not string some words together and then superimpose a rhythmic pattern, nor did the speaker delay or accelerate the pronunciation of the words in order to observe a beat. The speaker unconsciously selected these words so that their stresses would fall on a regular beat. This is the result of speech constraints.

But there is more to be said about speech rhythm. Not only does each speaker produce a rhythm, but also, speakers engaged in a conversation must negotiate for a common beat. When a conversation takes place, speakers engage in a cooperative act of communication. As Paul Fraisse (1974) has pointed out, cooperation in interaction, especially when involving physical effort, usually leads to synchrony in a rather automatic manner:

> Le rythme devient indispensable lorsque le travail est fait en commun, et surtout lorsque les *divers* mouvements sont solidaires les uns les autres. Rameur à plusieurs n'est possible que sur une cadence commune. Haler, comme faisaient des centaines d'Egyptiens déplaçant les énormes blocs de pierre dont il ons fait les temples et les

Pyramides, impliquait sûrement que les efforts fussent synchronises. (p. 171)¹²

A conversation is a physical interaction in which efforts – however minimal, in comparison to hauling blocks of stone – converge on the common goal of achieving some measure of successful interaction. It is in terms of cooperative effort that Tannen (1989) describes a conversation between herself and a new acquaintance at a Thanksgiving dinner:

[...] we achieved a high degree of cooperation. For example, we exhibited a pattern of cooperative sentence-building in which the listener picks up the thread of the speaker and supplies the end of the speaker's sentence, which the speaker then accepts and incorporates into the original sentence without a hitch in rhythm and almost without a hitch in timing. (p. 56)

Speakers tend to align the rhythm and the timing of their utterances in order to achieve maximum group concord. Ron Scollon (1981) has observed and analysed synchrony in a vast array of conversational settings, and has developed the concept of "ensemble". His work shows that rhythmic synchrony in conversation is not limited to speech: "conversationalists also cough, sneeze, clear their throats, blow their noses, and laugh in rhythmic ensemble" (p. 340). Before Scollon, Erickson and Shultz (1982) had

¹² My translation: "Rhythm becomes indispensable when work is performed jointly, and especially when *different* movements are interdependent. It is impossible for a group of oarsmen to row together unless a common cadence is observed. Hauling enormous stone blocks, as hundreds of Egyptians did when they built their temples and pyramids, undoubtedly required a synchronisation of efforts."

also documented the synchrony of speech as well as body movements in the context of face-to-face interviews:

Erickson and Shultz (1982) demonstrate that successful conversation can be set to a metronome; movements and utterances are synchronized and carried out on the beat. This phenomenon is informally observed when, following a pause, two speakers begin speaking at precisely the same moment, or when two people suddenly move – for example, crossing their legs or shifting their weight – at the same moment and often in the same direction. (Tannen, 1984, p. 154)

The Erickson and Shultz study showed that the rhythms of conversation and extralinguistic communication, such as gesturing and body movements, are synchronised down to 1/24th of a second (this is the frame interval in the film strips of the recorded interviews).

Sometimes, however, this cooperation breaks down. The problem may be due to a difference in the perception of topics that are appropriate or inappropriate for conversation. For instance, someone may think that asking personal questions is a sign of interest, while others may find it meddlesome. Other times, the problem may be due to an irreconcilable difference in conversational rhythms, so that some speakers may find a quick pace too pushy, while others may find a slow pace too awkward and reserved. Often the two problems overlap; if speakers are uncomfortable with a topic, the overall rhythm of the conversation suffers a misalignment.

In her volume *Conversational Style*, Tannen (1984) analyses two hours and 40 minutes of conversation at a dinner with five of her friends. She illustrates numerous

examples of rhythm getting in the way of harmonious communication. For example, just after the transcription of an exchange with one of the dinner participants, she reports:

The rhythm of this interchange is significant. [...] the rhythm is a pattern of answer-question, pause, answer-question, pause. Normally, a question and an answer are seen as an 'adjacent pair' [...], and in a smooth conversation, they are rhythmically paired as well. The differences in David's style on the one hand, and Peter's and mine on the other, however, create pauses not between an answer and the following question, but between our rapid questions and his delayed answers. Each resultant rhythmic pair, then, is made of David's answer and the next adjacent question. This is typical of the way in which stylistic differences create obstructions in conversational rhythm. The jerky rhythm is created by the difference in expectations about how much time should appropriately lapse between utterances in the conversation. (p. 71)

The differences in speech rate and pause length between Tannen and her friend Peter on the one hand, and David on the other, seem to be geographically – and therefore culturally – determined. David is from California, while Tannen and Peter are from New York. According to Tannen's account, the three New Yorkers who participated in the recorded dinner had similar speech rates; the three other participants, of whom two were from California and one from England, are portrayed as somewhat intimidated by their fast talk and assertive tone, and contributed much less to the overall conversation. Thus, it appears that the three New Yorkers created a successful ensemble, while the other speakers were more or less unable to align to their dominant

rhythm¹³. This difference in speech rate had a large impact on the level of each speaker's involvement in the conversation, and ultimately determined each participant's overall enjoyment of the dinner party. Thus, Tannen concludes that the ability to join others in a rhythmic alignment carries great affective value:

[...] the rhythmic synchrony basic to conversational interaction contributes to participant involvement much as singing along or tapping one's foot in rhythm with music. The opposite experience – lack of involvement resulting from inability to share rhythm – can be envisioned in the experience of trying to clap along but continually missing the beat. (p. 155)

This observation is not surprising, since many researchers (for example, Havelock, 1986; Fraisse, 1974) have discussed the pleasure that rhythm brings to the experiencer, especially when joining a group rhythm. Anyone who has witnessed the repetitive droning of a drum circle, and wondered what the attraction of such an activity may be, would only have to beat on a drum for a few minutes to realise that the communal rhythm has a genuinely narcotic effect on the players.

b. MELODIC CONSTRAINTS: PROSODY

Speech includes melodic elements that differ for each language and together form the prosodic repertoire of a population of speakers. The melodic phrases observed within a language are called intonation contours. Prosody also includes other

¹³ Note that the New Yorkers dominated the conversation even though the dinner took place in California; David is therefore not an 'outsider' in the overall conversational environment.

characteristics such as pauses, pitch variations and variations in voice quality. I have already mentioned how prosodic characteristics can exhibit some in-group variation. However, regardless of the size of the group and aside from some physiologically induced differences, prosodic characteristics are always culturally determined and their use in speech is highly normative. No matter how syntactically correct a sentence may be, the speaker will not be recognised as fluent in the language unless its prosodic characteristics conform to the expectations of the rest of the speaking population. In fact, I would argue that a speaker with a good command of the prosodic elements of a language would still be perceived as a fluent speaker even after making some mistakes; after all, native speakers also make speech mistakes sometimes.

Prosody is a very important aspect of linguistic expression, because it allows for the communication of extra-syntactic and extra-lexical information. In many cases, we can understand the information conveyed by a remark only because we evaluate its intonation. A simple sentence like 'You are so lucky' may indicate its literal meaning or its opposite, depending on the intonation contour. (Irony is a particularly good example because although it can be conveyed with syntactic structures or specific lexical choices, it exists mostly in the realm of intonation.) As I sit on your sofa and avoid bringing up the money you owe me, I have ample resources to indicate what is on my mind both by selecting a suitable topic (e.g., how richly furnished your living room is) and, especially, by using a tone of voice which indicates that what I am saying is not nearly the whole, or even the main point, of my story.

Apart from conveying meaning, prosody is also fundamental in organising speech into manageable chunks. Speakers employ prescriptive variations in pitch, intonation and pauses to convey a clear organisation of concepts to their listeners. For example,

both Wennerstrom (2001) and Chafe (1994) examine the fact that English speakers employ accents (substantial and sudden raises in pitch) to signal "new" information in a conversation. Conversely, when non-new ("given" or "accessible") information is mentioned, the pitch remains low or is lowered. "Given" information has been previously mentioned in the conversation; "accessible" information has not been previously mentioned but can be implied from other conversational items. Therefore, when I discuss the winter weather with you and I say "cloudy" for a second time, this information will be given; once I start discussing rain, you will have no problem following me because rain is accessible information in a conversation about cloudy weather. The first time I say "cloudy", I will probably employ a high pitch because I am introducing a new referent to you. As the conversation proceeds, it is unlikely that I will use anything other than a low pitch for both "cloudy" and "rain", unless there is something quite unexpected about them that I am trying to convey.

Pitch boundaries – the final tones of an utterance – constitute another organisational element in speech. They usually contain a lengthening of the last syllables (Wennerstrom, 2001, p. 20) and they signal either a temporary pause or the end of a conversational turn. For example, in English yes-no questions, the pitch boundary rises steeply and signals the end of a turn, as another participant in the conversation is required to respond. Other pitch boundaries, such as low rising, partially falling or plateau boundaries signal that the speaker is taking a short pause but intends to continue speaking; they also signal that the speaker is putting the uttered information into some relationship with what is going to follow (see Wennerstrom, 2001, p. 21-2 for a full account of possibilities). Finally, a low pitch boundary signals that the speaker has

concluded his/her speech for the moment; a short pause will follow, and other speakers may take the floor.

Wennerstrom also discusses keys, or the pitch used at the beginning of a clause, as an indication of conjunction or disjunction with what has been uttered before (p. 23-4). A high key indicates a contrast with the previous utterance, while a flat key indicates consistency with it, and a low key indicates a foregone conclusion. Paratones are a different type of onset, employed at the beginning of a whole section of speech after a somewhat long pause. They are effective in marking the organisation of large chunks of speech with regard to topic: high paratones signal a new start and a new topic, while low paratones introduce an aside, such as a series of parenthetical remarks that expand upon the current topic.¹⁴

The observations reported above from Chafe (1994) and Wennerstrom (2001) are valid only for English, of course, although Wennerstrom also illustrates some prosodic analyses in other languages, especially German. Other languages may employ different symbolic meaning for the organisational functions of pitch, clause endings and onsets. However, they will do so in a systematic manner. Across all languages, prosody is employed as an organising device for the harmonisation of speakers and the exchange of information between them.

Finally, in this very brief review on the properties of prosody it is important to mention that prosody is tightly bound to the expression of affect. It both signals and causes involvement in speakers and listeners. The affective mechanisms of prosody in speech have been studied systematically (see, for example, Scherer et al., 2001;

¹⁴ Most of the features from Wennerstrom discussed here can be observed in the transcription of a conversational passage at the end of this chapter.

Wennerstrom, 2001) and have been compared successfully to the affective properties of music (see Fonagy & Magdics, 1963, and, more recently, the excellent study by Cook, 2002). Just like rhythm, prosody involves a speaker in a conversation, and therefore encourages and perpetuates linguistic interaction.

In conclusion, prosody is extremely important in conveying meaning, in organising spoken discourse, and in creating an affective bond between speakers. Most of the characteristics of prosody are culturally determined (but see also Scherer et al., 2001) and are consistent among the speakers of a population as well as of localised groups. Therefore, prosodic behaviour is obviously constrained, because it must conform to a group code in order to convey meaning. However, its constraints on speech production also operate in other, more subtle ways. Let us return to the example I employed above when discussing speech rhythm:

We're all intuitively famíliar with the idéa of a derivative.

To illustrate the nature of rhythmic stresses, Wennerstrom includes pictures of the amplitude and pitch graphs for this utterance. Corresponding to every stress in the sentence, the amplitude graph reports a substantial increase, showing a burst of energy (and volume) with each beat:



Figure 1.1 Amplitude graph from Wennerstrom (2001, p. 51).



The pitch graph, however, is what interests me here:

Figure 1.2 Pitch graph from Wennerstrom (2001, p. 51).

The speaker starts the sentence around 175 Hz. With every stress, the pitch graph forms a chain of descending peaks, so that the first stress reaches 250 Hz, the second stress gets up to about 230 Hz, while the third stress is already much lower (about 170 Hz) and the fourth is about the same height as the third (170 to 165 Hz). The graph does not show the offset of the sentence, but we can be certain that the pitch peak on the fifth and last stress is even lower than on the fourth. The pitch for the overall sentence, in both peaks and valleys, descends along an irregular slope; the offset is lower in pitch than the onset, and the duration of the last syllables in the offset is most likely lengthened. The final pitch of the utterance is probably very close to the baseline of the speaker's range. This is a straightforward intonation contour for declarative clauses in English¹⁵. But once again, what is interesting is that this is not a case of superimposing an acoustic contour to a ready-made clause. The speaker selected this succession of words so that, from the initial tone, it would be timed precisely to describe a regular trajectory that ends at the speaker baseline. The initial and final tones of this contour, as well as its overall trajectory, are fixed by both physiology and cultural convention. The duration of the contour is determined by lung capacity. Yet the timing is impeccable: the contour is completed without effort, and the clause ends at the baseline. Therefore, (1) the

¹⁵ Note that there is some controversy on this specific type of contour between the 'declination' vs. 'breath-group' camps (see Umeda, 1980; Liberman & Pierrehumbert, 1984; Lieberman, Katz et al., 1985a and 1985b; Repp, 1985; and Hart, 1986). However, as Hird & Kirsner (2002, p. 536-7) point out, the controversy is engendered by the view that prosody is subordinated to syntax, which Hird & Kirsner oppose. The main point of contention between declination vs. breath-group regards the supposed correspondence between prosodic structure and syntactic structure.

The controversy, and its views on prosody/syntax subordination, has little import for the present speech constraint argument. Both declination and breath-group supporters agree that there are predictable and relatively consistent intonation contours for declarative utterances, although the nature of these contours varies with each theory. This predictability and regularity therefore remains one of the main points in my discussion on utterance planning.

Moreover, note that I have chosen the example above ("We're all intuitively...") for the sake of simplicity and ease of understanding. Although this example is a complete and grammatical sentence, I do not mean to imply that the organisation of intonation is dependent of, or even constrained by syntactic structures. As I repeatedly explain at many points in this work, I view all speech constraints as co-occurring, yet independent from each other.

meaning of the sentence was determined by the overall environment in which the sentence was uttered; (2) the words in the sentence were probably selected because of semantic activation; (3) the words were selected so that their stresses would align rhythmically; and (4) the words were selected so that they would compose a complete utterance in exactly the time that it took for the speaker to transition from the initial to the final tone, following a normal declarative intonation contour.

c. Sound Repetitions: Formulaicity, Repetitions and Echoes

Many investigations conducted by both linguists and other scholars have targeted the frequent instances of repetition in speech. The two volumes of *Repetition in Discourse: Interdisciplinary Perspectives* edited by Barbara Johnstone (1994) are a good example of the vastness and assortment of these studies; each chapter investigates a different aspect of repetition, from linguistics to literary theory to anthropology. In the very first chapter, we learn from Marilyn Merritt that repetition is both a universal phenomenon and a major resource in communication.

Deborah Tannen has also written a substantial article entirely devoted to repetition, in which she offers a comprehensive classification of the types of repetition commonly employed in speech. Tannen (1987) discusses instances of self and allo-repetition (repetition of others' words):

instances of repetition may be placed along a scale of fixity in form, ranging from exact (the same words uttered in a similar rhythmic pattern) to paraphrase (similar ideas in different words). Midway on the scale, and most common, is repetition with variation – including questions transformed into statements, or vice versa, and repetitions

with change of person or tense or other changes in wording. As repetition I also include patterned rhythm, in which wholly different words are uttered in the same syntactic and rhythmic paradigm as a preceding utterance. There is also a temporal scale ranging from immediate to delayed repetition. (p. 586)

While I am not here concerned with rhythm, all other instances of repetition Tannen mentions are important to the present discussion. Tannen documents her observations extensively, providing several examples from her transcriptions in which speakers in a conversation keep repeating the same sentence back and forth with very small changes; she also quotes instances when one speaker uses the same phrase many times over. She declares that all of her transcripts, which include conversations in American English and in Greek, show instances of repetition. In one of her examples, she discusses the same type of given/new intonation that I described in the previous section, in order to show that an expression can be repeated in an automatic manner. The speaker is a translator of American Sign Language:

(18) a. N: When you speak,

b. you use words to... to recreate that image

c. in the other person's mind.

d. C: Right.

e. N: And in sign language,

f. you use SIGNS to recreate the image.

In 18b, the intonation on *recreate that image* rises and falls. In the repetition 18f, N's pitch rises on *signs* but remains monotonically low and constant throughout *to recreate the image*. This intonation signals

given information, in part by the automaticity of the phrase in its second occurrence. Its meaning does not have to be worked out anew on subsequent reference, but is carried over ready-made. (p. 595)

In other words, the first occurrence of the phrase is stressed because it carries new information; the second occurrence is given, and the phrase is used in a formulaic manner. Through this and other examples of the same nature,¹⁶ Tannen demonstrates that, by analysing intonation, it is possible to determine that a phrase was repeated as a prefabricated chunk. The reiteration happens in an automatic manner, in the sense that the speaker is intent on achieving a certain conversational goal and simply employs words that are already present in memory. There seems to be no awareness of the fact that words are being repeated, and indeed this fact is not important in the context of the utterance. Repetitions are produced constantly in conversation, and although they are chastised in writing, they are hardly detected in speech.

Two more examples, from Tannen's *Conversational Style* (1984), effectively portray the automaticity of repetitions. In the first of these examples, Tannen and her friend Steve are discussing the former location of a New York radio station. Steve is trying to describe the location of the offices to Tannen, who does not know the building. Tannen is struggling to get her bearings and she makes a statement regarding a landmark in the area. Steve, who is very intent on cooperating with her to reach an understanding, starts repeating her statement in agreement. However, he soon has to stop and reverse, since the landmark she identified is incorrect:

¹⁶ Another excellent example is reported on p. 77 of Tannen's *Talking Voices*.

[...] Steve begins automatically to repeat my phrase 'now it's a round building with a movie theater', to ratify my offer of understanding. But in fact he cannot do so, because I have been wrong again [...], so he must stop himself from agreeing, to correct me again. The false start is a testament to the strength of his impulse to repeat an interlocutor's phrase that has been offered as a show of rapport, that is, to incorporate the other's offer into his own statement. It is interesting to note, too, that Steve's correction, (29) 'this is a huge skyscraper', is a repetition of his earlier statement, (17) 'Then they built a big huge skyscraper there?' 'Huge skyscraper' seems to be operating as a formulaic phrase [...]. (p. 76)

In another example, Sally, who is British, is describing the meal she was offered on an airplane. Tannen is confused about the point of Sally's story, and repeats Sally's utterance both as a sign of cooperation and as a sign of poor understanding. However, in her repetition, Tannen makes a small change in order to use an expression – a formula – that is very common in North America ("bagel and cream cheese"). Sally inadvertently follows suit and repeats the phrase, which is not a formula in her own culture:

> It is interesting to note that Sally says 'a bagel WITH cream cheese', her first two utterances, but I say 'a bagel AND cream cheese'. For me this is formulaic. When Sally ratifies my echo of her utterance [...], she switches to AND. This is apparently the effect of the echo. There are numerous examples in the conversation of people repeating things that

they would not ordinarily say, because the person they are echoing said them that way. (p. 121, footnote)

Automatic repetition is not limited to words and phrases; Tannen reports several studies which included the repetition of syntactic constructions (1984, p. 155). A further study is mentioned by Jackendoff (2002), who states, "Bock and Loebell (1990) give experimental evidence that not only do words prime other words, but syntactic structures prime other syntactic structures" (p. 217) .¹⁷ Syntactic repetition is likely to pass even more unnoticed than the repetition of words or phrases, and therefore is even more likely to occur in an automatic, unconscious manner.

All researchers agree in stating that the principal function of repetition is the production of speech with the least amount of effort. Because of its automaticity, repetition is a good stalling mechanism: the speaker can participate in the conversation with minimal effort, while at the same time preparing or deciding on the next utterance (Tannen, 1987, p. 582). Like Tannen (1987, p. 596-7), Merritt also notes that this expedient allows speakers to keep the beat of the conversational rhythm without adding anything to the conversation:

A salient property of linguistic repetition is that it occupies verbal space and therefore can be used as a kind of 'filler.' This is particularly useful when there is an ongoing rhythm in the discourse that is an integral part of sustained group participation and involvement. (in Johnstone, 1994, p. 28)

¹⁷ The reference is to Bock, K. and Loebell, H. "Framing Sentences." *Cognition* 35: 1- 39.

Since it takes up verbal space with information that is already given, repetition dilutes the contents of speech and therefore also facilitates comprehension. The hearer is able to absorb information more slowly and with more ease (Tannen, 1987, p. 582). Repetition also aids coherence, because it allows for an easier connection between utterances produced by different people and at different times.

Finally, repetition serves several interactional functions. The cooperative function of agreeing with somebody by repeating his or her words has already been noted above, as has querying someone's point or showing poor comprehension by repeating another's utterance with a questioning tone. But repetition is also employed for other purposes:

> Some functions observed in transcripts [...] include the following: getting or keeping the floor, showing listenership, providing backchannel responses, stalling, gearing up to answer or speak, humor and play, savoring and showing appreciation for a good line or joke, persuasive effect [...], linking one speaker's ideas to another's, ratifying another contribution [...], and including in an interaction a person who did not hear a previous utterance. In other words, repetition not only ties parts of discourse to other parts, but ties participants to the discourse and to each other, linking individual speakers in a conversation. (Tannen, 1987, p. 583-4)

So far, I have discussed the repetition of words, phrases and syntactic structures within a monologue or a conversation. But there are also a large number of repetitions common to an entire population of speakers. For a long time, linguists have observed the recurrence of lexical patterns in languages; many of them have brushed aside these

observations by cataloguing them under the peripheral category of "idioms". But some linguists have paid more attention than others to these patterns of recurrence. Dwight Bolinger (1976), for example, has striven to emphasise their importance. Alison Wray has chosen to specialise in formulaicity and has been very active in developing this area of research. As she expresses in *Formulaic Language and the Lexicon* (2002), the most inclusive definition of formula is:

> a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar. (p. 9)

With Michael Perkins (2002), she specifies:

This includes, at the one extreme, tightly idiomatic and immutable strings, such as *by and large*, which are both semantically opaque and syntactically irregular, and, at the other, transparent and flexible ones containing slots for open class items, like *NP be TENSE sorry to keep TENSE you waiting* [...]. [...] If we take formulaicity to encompass, as some do, also the enormous set of 'simple' lexical collocations, whose patterns are both remarkable and puzzling from a formal grammatical point of view [...], then possibly as much as 70% of our adult native language may be formulaic [...]. A range of corpus studies [...] have shown that the patterning of words and phrases in ordinary language manifests far less variability than could be predicted on the basis of grammar and lexicon alone, and in fact most natural language, written

or spoken, appears to consist largely of collocational 'sets' or 'frameworks' [...]." (p. 1-2¹⁸)

The lexical collocations discussed here are preferential combinations of words. They occur in all languages and can be more or less exclusive. For example, in English, "foregone" is invariably used before "conclusion" (maximally exclusive collocation), while each of the words in expressions like "to take care of" or "to take advantage of" can also be used in many other situations (minimally exclusive collocation).

Both Wray (2002) and Wray & Perkins (2002) posit that formulas have two principal functions: (1) they are tools for social interaction (they help achieve the manipulation of others, the assertion of individual identity, and the assertion of group identity); and (2) they are a shortcut in processing (they increase production output and fluency, they are time-buyers for planning, and through rehearsal and mnemonics they also allow the manipulation of information for easier retention).¹⁹ With regard to the second point, the authors suggest that

the benefits of prefabricated language in reducing processing effort can account for why an individual or indeed a whole speech community comes to prefer certain collocations and expressions of an

¹⁸ The expression 'NP' used in this passage is common linguistic notation for noun phrase.

¹⁹ Wennerstrom (2001) argues that lexical discourse markers that are common in speech (expressions such as 'like', 'you know', 'I mean' etc.) function as emphasis markers. I would argue that they also function as fillers, in the same way that other formulas do, since they obviously allow the speaker to maintain their rhythm and their turn in the conversation while adding no meaning. I think they are a perfect example of function #2 above.

idea over other equally permissible ones [...]. (Wray & Perkins, 2002, p. 16)

I think that there is a very clear parallel here with Tannen's observations on group cooperation and involvement, and on economy of effort.

With regard to the automatic chunking that Tannen examined, Wray (2002) writes:

There are words and phrases that we are likely to say when we see a particular friend, or find ourselves in a certain situation [...]. If we tell the same story, or deliver the same lecture, more than once, we will soon find that whole ideas are expressed in the same chunks of language each time [...]. We may re-echo a form of words that we used earlier, or which someone else has just used [...]. In the context of 'collocation' we find that some words seem to belong together in a phrase, while others, that should be equally good, sound odd. [...] Whether these preferred strings are actually stored and retrieved as a unit or simply constructed preferentially, it has been widely proposed that they are handled, effectively, like single 'big words' [...]. (p. 5-7)

The fundamental aim of Wray's volume is, ultimately, the development of an updated vision of a speaker's lexicon, where the "big words" mentioned above are not subject to the grammatical analysis imposed by the generative tradition, or if they are, the extent of this analysis is minimal (i.e., it may set the correct tense for a verb within a formula, but it does not break apart the structure of the formula).

Finally, Wray (2002) observes that some idiomatic expressions actually add to the processing efforts of the speaker in terms of sheer number of words uttered (e.g., "at

the end of the day" instead of "really", p. 74). Wray states that these cases may signal that cultural identity and belonging, which entail repetition of current phrases, are as important to a speaker as saving processing time. However, she also adds:

One possible explanation of the preference for a longer over a shorter way of expressing an idea is that it buys time for planning and/or that it contributes to an even rhythm [...]. If so, then the saving of processing effort is not simply about taking short cuts. It is about regulating production so that it is manageable, and this may mean, at times, taking the long way round, adding padding and/or establishing and maintaining a particular preferred rhythm and flow. (p. 75)

In closing this section, I will briefly discuss the repetitions of sounds that are below the level of the word, which I call *echoes*. In my own repertoire, I have observed some constructions with an interesting echoic structure. For example, I seem to be fond of using my own formula "I was wondering whether..." when I wish to make a polite request. This string obviously contains a heavy alliteration in "w". Other scholars have noticed a tendency to produce sound repetitions such as alliteration, assonance, consonance and rhyme. For example, Tannen (1984) reports a passage in which she and Peter are discussing Thanksgiving. Tannen declares that Thanksgiving was a foreign holiday for her and Peter's families, because they both immigrated to the United States; Peter, on the other hand, is concerned with dinner arrangements (he is the cook), but he also manages to express his fondness for the holiday. Below, I copy the beginning of the conversational passage, and then some of Tannen's observations on this passage:

(1) Deborah: Í wonder how óur ... grandparents and parents felt

(2) Peter: /?/ cranberry sauce.

about Thanksgiving.

(3) Peter: Cránberry sauce.

(4) Deborah: It wasn't their holiday.

(5) Peter: It's a wonderful holiday.

[...]

Peter's remark (5) 'It's a wonderful holiday' appears to be a response to my second try (4) 'It wasn't their holiday'. Paralinguistically, it echoes the sound and rhythm of my comment in an almost poetic way. His choice of the word 'wonderful' echoes my verb 'wonder', and the sound of 'wonderful' echoes the initial consonant and the rhythm of my 'wasn't their'. (Sacks [1971] noted the tendency of speakers to choose words just used by interlocutors or that use sounds that appeared immediately prior. He called that process 'sound selection' or 'sound touchoffs'.) (p. 92-3)

As Tannen (1984) observes, the exchange includes alliteration in "w"; this corresponds also to the repetition of a lexical item (wonder/wonderful) and to a tight rhythmic parallelism. Unfortunately, the reference to Harvey Sacks' material is to unpublished lecture notes, so it was not possible for me to examine the material. But on page 155, Tannen discusses his work again, and this time she offers a concrete example:

In his example, a speaker says, 'cause it comes from cold water', and a few lines later says, 'You better eat something because you're gonna be hungry before we get there'. In suggesting why 'cause' occurs in

the first instance and 'because' in the second, Sacks notes that 'cause' appears in the environment of repeated /k/ sounds, whereas 'because' is coordinated with *be* in 'be hungry' and 'before'.

In other words, this is another case of alliteration²⁰. Other instances of echoes appear in Tannen's own transcriptions. Tannen (1989) discusses the repetition of single sounds in another passage of the Thanksgiving conversation (p. 77). Two examples from that conversation are the sentences "it's not a good idea in terms of time" (alliteration in "t"), and "I eat a lot because it's not satisfying" (rhyme in "ot").

Regrettably, overall, the work on echoes is much scarcer than that on repetition and formulaicity. While I have no doubt that echoes are extremely common, and that they carry a considerable impact on lexical selection, more work must be done to determine their actual significance.

To conclude, I have reviewed many studies on repetition – of words, phrases, syntactic structures and individual sounds that occur in every conversation, and of formulas and collocations as part of the repertoire of a language. All of these repetitions occur mostly in an automatic fashion, both because repeated information is readily accessible in working memory, and because repetitions convey the speaker's affective disposition and participation in a group dynamic.

Repetitions interest me because, once again, they help in predicting the output of a speaker. In general terms, the use of formulas and collocations can and should be expected of a fluent speaker. Mastering a language means mastering its syntax as well

²⁰ Of course, since 'because' is longer than 'cause' by one syllable, other constraints such as prosody and/or rhythm may also converge in determining these word choices.

as its formulas and collocations; hence, we can see two approaches to learning a language, the grammar textbook and the phrasebook, neither of which is sufficient without the other. However, the most intriguing type of repetition in terms of predicting output in an everyday conversation is perhaps the automatic repetition of non-formulaic words, phrases and sounds. These repetitions indicate that perceived speech guides output not only in terms of semantic priming, but also in terms of offering ready-made building blocks and strong mnemonic prompts of an acoustic nature. The fact that a speaker's output can easily accommodate chunks of pre-processed speech without a breakdown in other respects, such as its prosodic patterns, rhythm or syntactic structure, shows that the effort in producing an utterance must involve a lot of mnemonic resources that are primarily guided by, and not simply adhere to, several constraints at the same time. These constraints guide the composition of an utterance by selecting chunks of speech that are accessible in working memory and that satisfy several requirements at the same time. The requirements are those of theme and imagery, rhythm, prosody, sound or word activation, and syntax.

4. Syntax and a Few Words on Hierarchy

Naturally, in the production of utterances syntax is also a constraining factor. The predictive value of syntax is important in the case of repeated syntactic structures, which are a type of repetition and therefore clearly constitute a type of mnemonic prompt. If we consider that a speaker is likely to use a structure that others or that the speaker him/herself has recently employed, we could be one step closer to guessing the exact form of the utterance s/he will produce.

The rule-based application of syntactic structures is also very important, independently of whether any repetitions are involved in the construction of an utterance. The capacity to infer syntactic rules from perceived speech is common to all speaker populations. Once they have learned them, speakers must keep the rules of their language in memory, just as they keep their lexicon. Fluent speakers of a language are well versed in following shared syntactic rules, and these rules determine the order of the words in an utterance as well as their conjugation, inflection, combination and so on. Syntactic rules are a very important factor in producing utterances, and in being able to understand perceived speech. Therefore, syntactic rules are also an important factor in predicting some aspects of the shape that utterances will take when produced.

However, it should now be clear that syntactic constraints are only one of several constraints that must be observed when producing an utterance. Speakers usually produce utterances that are syntactically correct, rhythmically aligned, semantically coherent, prosodically predictable, and replete with formulas, repetitions and echoes. All of these constraints are satisfied at the same time, and they are all important in establishing the meaning of an utterance and its pragmatic values. As I have already discussed, fluent pronunciation, use of slang and adherence to prosodic norms are just as important as a correct use of syntax in identifying a fluent speaker. In fact, they may be even more important because they allow for less flexibility; in the casual observation of everyday occurrences, it is rather normal to detect syntactic errors in speech, while in other areas, such as in the use of formulas or prosody, errors are less common and far more noticeable.

The constraints of speech all work towards cuing items in working memory so that an utterance will adhere to them. Some utterances fulfil all constraints very well, while

others fulfil some constraints well and others only partially or not at all. In such a complex system of concurrent and perhaps competing mnemonic cues, it is logical to wonder whether there might be a hierarchy or a processing priority among the constraints. In current linguistic circles of both Generative and Functionalist traditions, the most predictable answer to this question is that semantics and syntax are the twin engines of production, and that other characteristics such as intonation and rhythm develop as secondary effects. I wish to stress that this cannot be the case. The patterns of intonation, rhythm and repetition are ubiquitous occurrences: they must coexist in the same utterances with syntax and semantics, and therefore the utterances must suit these patterns just as much as they must suit syntax and semantics. Clearly, there is an imitative drive in humans that leads to the recurrence of elements such as rhythm, intonation and formulas between speakers. This drive is inherent in communication. Speech is not understandable unless it is organised by intonation; therefore, intonation contours do at the very least constrain the minimum and maximum length for any utterance. If a rhythm is to be respected, then the syntactic structure of an utterance and the selection of its lexemes must allow for semantic cohesion and syntactic agreement, but also for the consistent placement of stresses. If previously heard speech is semi-active in memory and is likely to be repeated, then it must influence semantics as well as syntax in order to be produced again.

I have not established whether there is a hierarchy among constraints, nor did any of the studies I have examined yield any evidence for the primacy of specific constraints – in fact, some of the studies also take issue with the supposed primacy of syntax,

although they do so for their own reasons.²¹ I hope to explore the issue of constraints hierarchy in future studies. For now, the question remains open.

5. Sample Analysis

Thus far I have identified and provided examples for a range of mnemonic constraints: semantic activation, normative intonation contours, syntax, speech rhythm, and repetitions of words, formulas, syntactic structures and echoes. I have stressed the fact that all of these constraints together determine the utterances that a speaker will select when producing speech, and that this process of selection is automatic, involuntary and largely unconscious. While we are obviously somewhat aware of the topics we are discussing in a conversation or monologue, our conscious reasoning is

²¹ Wray and Perkins (2002) offer an interesting point of debate with the generative tradition. I earlier quoted them as noting that everyday speech shows much less variability than would be afforded by the syntax and lexicon of a language. This is a valid observation; if syntax were not kept in check by other factors, there would be no explanation for the repetitiveness of language. Wray (2002) correctly points out that the canonical position of the generative tradition, with its emphasis on analysis and on a minimal lexicon composed almost entirely of primitives, simply does not fit ordinary linguistic behaviour. She advocates a dual processing model in which analysis and formulaicity both have a place, with analysis "filling in" the places where formulas are ineffective.

Other researchers have emphasised repetition and formulaicity at the expense of analysis, both in the domain of adult speech processing and in that of language acquisition. See, for example, Tannen (all works in bibliography) and the chapters by Bennett-Kastor and Shepherd in Johnstone (1994).

usually preoccupied with the ideas that are to be expressed (and, as I have shown, even this serial progression of ideas is quite restrained by semantic processes). The actual selection of each word, or even of entire sentences and sentence structures, does not involve voluntary planning; the strings of words surface to consciousness ready-made. What I am suggesting is that our brain puts together these strings of words largely by isolating lexical and syntactic elements in memory because it is cued by speech constraints. These constraints are active in memory either because of speech that has been recently heard, or because of behaviours that have been learned by mimicking fellow speakers from a very young age. Most likely, both these types of speech patterns are active at the same time, so that recently heard speech blends with self-reinforced speech behaviours.

In the previous sections, I have offered many examples of each constraint in action by employing transcriptions of actual utterances. Now I will concretize my argument by showing the aggregate activity of all constraints at once.

Unfortunately, Tannen does not notate accents or intonation contours in most of her transcriptions; although some of the passages reported in her work would make for extremely interesting analyses, I cannot take advantage of them because this would automatically exclude rhythm and melody from the discussion. For this reason I have chosen to analyse the longest conversational passage in Wennerstrom (2001, p. 68), which consists of 29 lines of dialogue between an instructor of economics and two students. Fortunately the whole passage has not only been reported and annotated with prosodic markers by Wennerstrom, but was also video-recorded for an instructional videotape. I obtained a copy of the videotape and reviewed the lecture in which the passage occurred, as well as the prosody of the passage itself. For my analysis, I will

first provide a plain transcription of the passage. Then, I will give a brief overview of Wennerstrom's notation system for prosody. Finally, I will report the passage with Wennerstrom's prosodic notation, for the reader to deduce the prosodic contours of the passage.

Below is the plain transcription, preceded by the initial of the speaker's name:

- 1 **D** So what other things can you think of... other...
- 2 perhaps other cultural factors that
- 3 influence taking your product abroad.
- 4 **B** C... hum... competition with another
- 5 product. An existing product.
- 6 **D** OK....
- 7 **B** Like you said: a Cuisinart...
- 8 **D** Sure!
- 9 **B** is competing with a small space and
- 10 chopping knife... or... uh... some other...
- 11 hum....
- 12 D OK... so you're thinking of competition...
- 13 you're also thinking of substitute products.
- 15 **B** Substitute products.
- 16 **D** Uh huh.
- 17 K Also, the disposable income of the
- 18 market you're selling to.
- 19 **D** OK. Can you think of any ways that you
- 20 can... hum... if you're- if you're trying to

- 21 market something... you... as a... somebody in
- 22 the United States trying to market something in... uh...
- 23 China. How would you find out about
- 24 disposable income? Where would you go to find
- 25 that information?
- 26 K Uh... in that case I'd assume there'd be
- 27 some numbers from the government
- somewhere.

I will now list the notation marks that Wennerstrom (2001) employs in her transcription.
Notation Mark	Description						
Λ	High paratone: very high pitch employed at the beginning of a						
	passage of speech; usually indicates a change in topic.						
	Low pitch boundary: falling tone at the end of an utterance; usually						
¥	indicates that the utterance is complete.						
	Plateau pitch boundary: flat tone at the end of an utterance;						
	usually indicates that the speaker intends to continue speaking.						
7	Low-rising pitch boundary: tone at the end of an utterance that						
	rises from low to mid-range; usually anticipates further speech.						
	Partially falling pitch boundary: tone at the end of an utterance that						
-	falls from high to mid-range; usually anticipates further speech.						
=	Continuation from one speaker to another, with no pause.						
	High key: high tone at the beginning of an utterance; usually						
┍→	indicates a contrast between what has been said and what is going						
	to be said.						
à	Primary word stress.						
CAPITALS	High pitch accent; it is usually employed with new information.						
SUBSCRIPTED CAPITALS	Low pitch accent; it is usually employed with given information.						
UNDERLINED	Steeply rising high pitch accent; it is usually employed to call						
<u>CAPITALS</u>	attention to an item, i.e., in a contrastive utterance.						
(.1)	Pause, measured in seconds $(.1 = 0.1 \text{ second})$.						

Table 1.1 Notation marks for prosodic information employed by Wennerstrom (2001) in her transcription.

FING	any, i	There is wernerscrom's transcription, with prosourc annotation						
1	D	Î So what <u>Ó(hh)THER</u> THINGS CAN YOU THINK OF↓ other-						
4		(.2) perhaps _{ÓTHER} <u>CÚLTURAL</u> (.3) _{FÁCTORS THAT} (.4)						
5		ÍNFLUENCE TÁKING YOUR PRÓDUCT ABRÓAD \downarrow (1.2)						
4	В	C- um \rightarrow (.7) COMPETÍTION with ANÓTHER (.4)						
5		PRÓDUCT↓ (.5) An EXÍSTING product↓ (.4)						
6	D	OK <i>⊅</i> =						
7	В	= _{LIKE YOU SAID} (.5) a CÚISI NART- (1.3)						
8	D	SURE↓						
9	В	is competing with $a \rightarrow (1.8)$ SMALL SPACE and- (.3)						
12		CHÓPPING knife∍ or- (.5) uh some ÓTHER- (.2)						
13		Um→ (.6)						
12	D	r→ _{OK} ↗ só you're THÍNKING of COMPETÍTION↗ (.2)						
13		you′re <u>ÁLSO</u> _{THINKING OF} SÚBSTITUTE PRÓDUCTS↓						
14		(.6)						
15	В	SÚBSTITUTE PRÓDUCTS↓ (.2)						
16	D	Uh huh⊅ (.7)						
17	К	$ ightarrow \underline{ALSO}$ the DISPÓSABLE ÍNCOME of the $ ightarrow$ (.1)						
19		MÁRKET- (.4) you're SÉLLING to↓ (.3)						
19	D	OK𝔥 (1.5) → Can you THINK of any WAYS that you						
20		c'n- (.1) um- (.2) if you're- if you're TRÝING to						
26		MÁRKET sómething, y- YOU as a- (.5) somebody in						
27		the UNÍTED States $_{TRÝING TO MÁRKET SÓMETHING IN}$ (.3) uh \rightarrow						

Finally, here is Wennerstrom's transcription, with prosodic annotation:

28		(.1) CHÍNA \downarrow (.2) HOW would you find OUT about
29		DISPÓSABLE ÍNCOMEJ (.1) WHERE would you go to FIND
30		THAT INFORMÁTION↓ (1.25)
26	к	Uhh \rightarrow in <u>THAT</u> _{CASE} I'd- (.9) ASSÚME there'd be \rightarrow (.2)
27		some NUMBERS from the GOVERNMENT
28		SÓMEWHERE↓

The constraints at work in this passage are the following:

1. **Semantic Coherence**: The entire passage clearly revolves around the topic of international commerce. The "theme" of the passage is set in lines 2-3 and explored in the following lines. In a teacher/student situation such as this one, a theme exercises a lot of weight in a conversation; it is somewhat unlikely (although certainly not impossible) that the participants will drift off theme and start discussing something else. In a less rigid situation, when themes are not quite so imposed, participants may still discuss a topic for a long time, but may also be more likely to evolve from one topic to other, related ones. Regardless of the situation, it is common for a theme to be stated explicitly by one of the participants before or while it is discussed by the group.

In this passage, the topic evolves from a broad question on culture and commerce to the narrower issue of substitute products, and from there to disposable income. The vocabulary of the passage is composed of related lexemes, such as "product", "competition", "market", and so forth. When B develops the topic of competing products and briefly explores the domain of kitchen tools, the items mentioned (the Cuisinart, small kitchens and chopping knives) are all part of this domain. B is readily coming up with these items because they have been mentioned recently, in the same lecture.

2. **Details/Imagery**: The theme stated in lines 2-3 is rather broad and abstract: there may be cultural factors to be considered when marketing a product in another country. B's first reply to the theme is also broad and abstract (competition with another product), but soon afterward, B becomes more involved with the answer by giving a concrete example (the Cuisinart versus the chopping knife). After the first example B seems to falter, and D cuts in to offer another abstraction ("substitute products"). If D had not cut B short, B would likely have pursued more details in his discussion. Similarly, when K introduces the topic of disposable income, D develops it by offering another example on ways to find information about disposable income in China. I think that this passage is a clear illustration of the fact that, even when exploring rather abstract and academic subjects, it is quite impossible to develop, or even understand, a theme without grounding it in the details of real-life or imaginable situations.

3. **Rhythm**: The rhythm in this exchange is not particularly cohesive, yet I think that the level of cooperation is relatively high. To begin with, the pauses are numerous but somewhat short: in the 63.35 seconds of conversation we find 33 pauses, with an average length of 0.52 seconds (variance = 0.18 sec). The placement of few longer pauses is quite revealing of the situational dynamics. To begin with, both of D's questions are followed by longer pauses; it seems natural to assume that the students are concerned with finding the correct answer, and therefore take a little time to

respond. Another longer pause occurs on line 7, when D cuts B off. B takes a little while to recover and continue the sentence. I think that this occurrence, again, clearly shows that the views expressed by the two speakers do not carry the same weight. Finally, another longer pause happens on line 19, right before D asks a second question introduced by the high key. D is obviously placing some emphasis on this question; not having obtained the desired answer the first time, D produces a contrastive tone, indicating that the required information has not yet been achieved and that another route is being explored in order to get to it.

In terms of the actual rhythm of the utterances, Wennerstrom has not been overly diligent in notating all stresses in the passage. For example, all stressed words (in capitals) are pronounced with a stress, but these stresses are not reported visually in the transcription. For a few periods in the passage, no stress is notated, yet the speakers did employ a stress scheme when producing their sentences. Nevertheless, certain utterances report a continuous stress pattern: lines 2-5, 12-13, 15, 17-18, 22, and 27-28. The patterns of these utterances are quite regular, with most stresses being divided by intervals of two or three unstressed syllables. Occasionally, the intervals are longer (four syllables) or shorter (one syllable), but in these cases the rhythm is still predictable. For example, both lines 12 and 13 include a four-syllable interval in adjoining utterances: "OK, só you're thínking of competition, you're álso thinking of súbstitute próducts." The stress scheme of these two utterances reflects the singsong prosody of the whole sentence. I will show this below by displaying short and unstressed syllables (\bigcirc) and long and stressed ones (-):

\cup		U		U	U	U	U	—	U	
ΟК,	só	you're	ţhín	king	of	com	ре	tí	tion	
U	—	U	U	C	U		U	U	_	U
you're	ál	SO	thin	king	of	súb	sti	tute	pró	ducts

4. **Intonation**: The intonation patterns in the passage are easily identified as the standard intonation contours of the English language. The paratone at the very beginning of the passage indicates that D is now switching to a new topic in the discussion. The prosodic features described in the notation table above the passage, such as keys, intonation boundaries etc., are all used in a manner consistent with the standard prosodic features of English. By far the most common intonation contour, as represented by Wennerstrom's notation, is that of a key (beginning of an utterance) located midway in the speaker's range, followed by a high or very high pitch accent, and either descending directly towards a low pitch boundary, or hitting a few lower pitch accents before ending at the low boundary. This contour is well exemplified by line 1 ("So what other things can you think of"). Let us look at its pitch graph:²²

²² All graphs and pitch measurements were produced with the application Praat. I have imported the sound stream from Wennerstrom's videocassette using Praat's digital recorder. The stream was imported as a 16-bit-per-sample mono track, with a rate of 32,000 samples per second.



Figure 1.3 Pitch graph for "So, what other..."

The beginning corresponds to the high paratone, the peak corresponds to the accent on "other", which is relatively high, the words "things can you think" are low, and the last syllable, "of", is the low pitch boundary.

The graph of a longer utterance with multiple accent peaks, such as lines 12 and 13 ('OK, so you're thinking of competition, you're also thinking of substitute products"), looks like the one below. Notice that the speaker is talking very fast, but she slows down considerably at the end of the sentence (starting at "substitute") because she has started writing on the blackboard at the same time. The text is rather long, so I have reported in the graph only those words that correspond to a heightened pitch.



Figure 1.4 Pitch graph for "OK, so you're thinking..."

The first peak, "OK", goes from mid-range upward and is sustained by the second peak on "so". The two peaks on "thinking" and on the second syllable of "competition" are the highest in the utterance (264.63 Hz and 259.43 Hz respectively), and they establish the main idea that is being pursued. A pause of just 0.2 seconds follows, and then "also" and "thinking" again reach a relatively high peak (246.43 Hz and 250.27 Hz), effectively conveying the idea of adding one more item to the "thinking" list. The last two peaks, on "substitute" (234.81 Hz and 234.58 Hz) and "product" (248.69 Hz) are still quite high, indicating that new and important information is being discussed. "Products" ends low, at 187.96 Hz. The lower offset signals that the speaker has

concluded her utterance and that another speaker may take the floor. Accordingly, B makes another contribution to the conversation at this time, and he repeats the words "substitute products".

5. **Repetitions and Formulas**: The passage is obviously replete with repeated speech. I will group the repetitions by type in order to list them below. In my list, "formulas" refer to expressions that recur in the speech of all fluent English speakers; "groups of words" refers to repetitions of groups of words that are limited to the present situation – in a way, they are "temporary formulas". But also, note that the two phrases "substitute products" and "disposable income" are likely to be formulas in the realm of economics, although they are not necessarily formulas for the population of English speakers at large. In any case, they are certainly used in a formulaic fashion in this particular conversation.

The numbers in parentheses refer to the line numbers in which expressions are to be found. The notation "NP" means noun phrase, "VP" means verb phrase, and "ADV" means adverb.

<u>Formulas</u>: "OK" (6, 12, 19); "Uh huh" (16); "like you said" (7); "sure" (8); "some x... somewhere" (27, 28) – this formula may also be used in its variation "something... somewhere"; "what x... can you think of..." (1); "in that case" (26).

<u>Groups of words</u>: "substitute products" (13, 15) – this formula is particularly interesting because even the intonation is repeated exactly by the student, who is obviously mimicking the teacher; "disposable income" (17, 24) – again, the intonation here is interesting because on first appearance the formula has the typical high accent of new information, while in the repetition it has the low accent of given information

(Wennerstrom herself comments upon this point); "trying to market something" (20, 22); "can you think of" (1, 19).

<u>Syntactic repetitions</u>: "you're (ADV) thinking of + NP" (12, 13); "how/where would you + VP" (23, 24).

<u>Words</u>: "other" (1, 2, 10, and also "another" in 4); "that" (25, 26); "product" (3, 5 twice); "find" (23, 24); "you're" (12, 13, 18, 20, 21) and "you/your" (1, 2, 7, 19 twice, 21, 23, 24); "also" (13, 17) – note that the word is uttered with the same very high accent on both lines; "market" (18, 21, 22) – note that in line 18 the word is used in its noun form by K, while D uses it as a verb in the other two instances; likewise, in "competition – competing – competition" (4, 9, 12) the speakers use the same root in its noun and verb derivations.

<u>Echoes</u>: "<u>things</u> can you <u>think</u> of" (1); "<u>factors</u> tha<u>t</u> influence <u>taking</u> your produc<u>t</u> ..." (2, 3); "<u>c</u>ultural... <u>c</u>ompetition... <u>C</u>uisinart is <u>c</u>ompeting" (2, 4, 7, 9); "<u>s</u>mall <u>s</u>pace... <u>some...</u> <u>so...</u> <u>s</u>ubstitute... <u>s</u>elling" (9, 10, 12, 13, 18); "<u>some</u>thing... <u>some</u>body... <u>some</u>thing... <u>some</u>... <u>some</u>where" (21, 22, 27, 28); "how <u>w</u>ould... <u>w</u>here <u>w</u>ould..." (23, 24); "a<u>ss</u>ume... <u>some</u>" (27, 28).

6. **Syntax**: There are two sections of this passage in which the speakers' utterances are perfectly correct and complete in terms of traditional grammatical analysis: on lines 12-13 ("OK, so you're thinking of competition, you're also thinking of substitute products") and 26-28 ("In that case, I'd assume there'd be some numbers from the government somewhere") speakers express themselves with complete and accurate sentences. Many other utterances in the dialogue are fragments of sentences. These fragments, however partial, are internally accurate and grammatically viable: for

example line 1 ("So what other things can you think of..."), lines 4-5 ("Competition with another product. An existing product"), line 15 ("Substitute products") and so on.

The passage also includes some grammatically incomplete utterances. In particular, the two longer turns from single speakers (B's turn in 7-11 and D's turn in 19-25) are decidedly choppy and ungrammatical.

My analysis shows that in every line of this passage of conversation, multiple constraints are at work. The constraints overlap at every point of the passage and guide the production of utterances by determining the selection of syntactic constructions and lexemes.

Since the reader may still be wondering what I mean by "determining the selection", let us look at an utterance from the passage above: "OK, so you're thinking of competition, you're also thinking of substitute products." The intonation at the end of "competition" rises, and the speaker repeats the wording "you're thinking of (x)"; these two features used together give the impression of a list, to which other things could be added. To obtain this effect, the parallelism between the two sentences is crucial (cf. Tannen "Repetition" 582). But the speaker also wishes to draw attention to the second item in the list, and that is achieved through the high accent on "also", "thinking" and "substitute products". Therefore, the intonation pattern of the second half of the utterance, after the pause, is very tightly linked to its syntax. If we want to examine alternative ways in which the speaker could have expressed the very same concept in the second half, we cannot simply remove "also", especially because the accent would be lost. If the high accent on "also" were eliminated and the intonation on that word were to remain flat, then the highest accent in this clause would have to fall exclusively

on "products". There is no other way in which this utterance would otherwise make sense to the listener. But the rising pitch of "competition" and the pause after this word create an expectation that a new, important element will be added; the early accent in the second utterance is there to fulfil that expectation. If the entire sentence was "so you're thinking of competition [– pause –] and you're thinking of substitute products as well", the end of "competition" would not likely be pronounced with a rising tone. Therefore, if an utterance parallel to the first is to be produced and the accent is to fall early in the utterance, then "also" must be there. A longer synonym, such as "likewise", would take up too much space.

Consider another example, the utterance at lines 17-18: "Also the disposable income of the market you're selling to." I have already pointed out that "also" is a repetition of the other "also" on line 13, and that it is even uttered with the same very high pitch. Again, there is somewhat of a list effect, with participants adding their contribution to the discussion almost in an itemised manner. Since "also" starts on a contrastive high key, the utterance cannot be too short, because it must reach a low boundary in order to signal that the speaker has finished his/her turn. So, the utterance must include at least three or four syllables after the very high accent. But if the speaker is going to use the semi-formulaic expression "disposable income", then the earliest location for a low boundary would be on the last syllable of "income". However, the utterance "also the disposable income" would not be very clear in the context of the conversation; the speaker must qualify the new information "disposable income". That is why "of the market" gets added on, with a plateau boundary at the end of "market". Once again, "market" is relatively new information and must be qualified. To do so, the speaker chooses to use the same syntactic structure as other speakers, who have

offered examples in the "you + VP" form. (This seems a much more common manner of exemplifying actions in English than other alternatives, such as "I + VP" or "one + VP".) Therefore, the qualifying utterance "you're selling to" is also predictable, given the context.

Indeed, not many variations of this utterance would be realistically predictable. Although an utterance such as "in addition, we must take into consideration the disposable income of the buyer that is being targeted" would be perfectly appropriate and would convey the same meaning – possibly even more precisely than the actual utterance does – I think it would be quite unlikely for a speaker in this situation to employ it; the alternate utterance would clash with the overall register of the context, and the overall register is heavily determined by the syntactic complexity, the choice of vocabulary and, most of all, the repetitions that occur in the context. The only feasible and predictable variation I would envision for this utterance pertains to the lexeme "market", which a more naïve speaker could substitute with the generic "people" ("Also the disposable income of the people you're selling to"). In this case, the more frequent "people" would most likely win over other synonyms such as "population"; it could be substituted by "those" ("Also the disposable income of those you're selling to"), but in this case the distribution of pauses would have to change.

In closing, I emphasize that this passage does not seem to indicate any unquestionable primacy among the constraints at work. Syntax certainly does not seem to be a very reliable guiding factor; some sentences are quite correct, while others are fragmentary, and yet others are ungrammatical. Interestingly, the ungrammatical sentences are also the longest ones in the passage. Perhaps, over time, syntax may in

fact be the first constraint to break down. In contrast, there are no inconsistencies or mispronunciations in any of the speakers' intonation patterns. Thus, despite the syntactic errors, the speakers clearly have no problem understanding each other. The repetitions employed in the exchange are also crucial for the coherence and understandability of the dialogue. Furthermore, the speakers are engaged in the conversation to at least some degree; this is a sign that rhythm and intonation are doing their job at keeping participants involved. Therefore, even at those times in the conversation when syntax does break down, communication among the speakers can still continue because of the effect of other cohesive constraints.

6. Summary: Speech Constraints in a Nutshell

In this chapter, I have argued that a speaker's memory is flooded with cues when the speaker is immersed in a conversation. These cues are both generated and kept in check by a variety of independent mechanisms: semantic activation, imagery, rhythm, prosody, repetition of phrases, words or sounds, and syntax. Exposure to environmental cues may activate a network of lexemes in memory. It may also supply the need to use a certain tone of voice, or a certain rhythm. As the conversation continues, a very large number of lexemes will become semiactive in working memory; this peripheral activation is caused by semantic associations, by repetitions and by sound redundancies triggered by the speaker's previous utterances or by others' speech. Other lexemes may already be semiactive simply because of their frequent use in common speech. The rhythmic and intonation patterns in the conversation will be determined much more by automatic processes than by the speaker's own intent.

Before the speaker starts his/her first utterance, the number of candidate utterances (including their lexemes, syntactic structures, prosodic characteristics and so on) that can potentially be selected for production is limited by the available cues, but is still high. However, as the first sound of an utterance is selected, the sound itself becomes both a constraint and a cue for further production, and the number of feasible utterances that can be produced drops sharply. For example, as soon as the speaker selects a key (initial tone) for the utterance, the tonal distance between the key and the speaker's baseline will determine the minimum length of the utterance, provided that the utterance is a declarative sentence. (In the case of questions, the pitch contour will differ, but observations about minimum length determined by the initial key are still valid.) The rhythm of the utterance will also determine which lexemes can or cannot be selected to carry the rhythm to the end of the utterance. The syntactic structure of the utterance will start from the beginning and proceed with each lexeme selection, according to the speaker's acquired rules. Moreover, each produced sound will recall other similar sounds, and each selected lexeme will recall similar lexemes.

In other words, the composition of the utterance is a complex process of selection based on several simultaneous cues. This selection process may happen partially or entirely in working memory, just before the utterance is actually produced. In this process, the mnemonic constraints at work in speech will help isolate certain lexemes and groups of lexemes in working memory, so that the utterance is composed one chunk at a time. Each sound, word, formula, prosodic element or syntactic structure that is selected contributes to the constraints at work in the utterance. This means that each element adds a cue, a call that may help isolate another set of elements in working memory.

The overall process of selection is additive. At the beginning of the selection process, the candidates for selection are many because only a few cues are at work. However, as several cues accumulate, the candidates that can fulfill all cues at once become much more limited. Therefore, it will be easier to predict the offset of an utterance rather than its onset.²³ In the last section of this chapter I have illustrated the predictability of utterances by examining two examples. Note that in the first case, I chose to analyse the second half of an utterance, and in the second case I analysed a full utterance spoken towards the end of the passage. In both cases, my choices have been determined by the availability of previous data: the more cues that can be computed, the more accurately we can predict their cumulative results.

It is important to point out that, although the overall selection process is additive, each constraint can work forward as well as backward, since it can constrain the selection of chunks that come both before and after it in the utterance. This means that the selection of chunks happens both serially and in a holistic fashion. The overall process of composition proceeds serially, going forward from the beginning, but later chunks may also be pre-selected in order to tie in with earlier chunks. For example, the chunks of a hypothetical utterance like "I am extremely sorry to have kept you waiting for so long" may be selected overall in a serial fashion, proceeding from the sound of "I" and filling in the rest of the sentence according to required length, rhythm, semantic activation and syntactic structure. However, those chunks that make up the idiomatic "sorry to keep [tense] you waiting" are the core of the sentence, and may be selected

²³ However, note that onsets are also constrained by previous speech and consequently, they too can be predicted to a certain degree.

earlier than other "filler" chunks such as the adverb or the end of the utterance. The selection of chunks of utterances in the pre-production phase has been amply demonstrated by research in speech errors (Cutler, 1982), whereby a speaker who says "it's all gun and fames" instead of "it's all fun and games" has obviously pre-selected the entire "fun and games" expression before uttering the first mistaken "g" in the sequence.

In Chapter Two, I will provide more evidence for the cognitive and mnemonic constraints of speech production introduced in this chapter. This evidence comes from a study on the production of utterances in a specific context – that of oral traditions – in which poets and performers must spontaneously compose original yet predictable verses. We will see that there is a fitting comparison to be made between the poets who operate within a tradition and the speakers who operate within a community.

Chapter Two

Speech Constraints and Oral Poetry

In Chapter One, I stressed the role of mnemonic cues in producing speech. I am not alone in proposing that verbal production may be generated by an array of mnemonic constraints. David Rubin has documented the production of utterances within constraints very similar to the ones I have described. His area of research is not the production of everyday speech, but the production of songs, rhymes and other performances within an oral tradition. While the two phenomena may at first seem quite unrelated, the analogies between them are remarkable. As I will argue in more detail below, both types of production involve the use of mnemonic constraints to cue material from working memory. The chief difference is that the performer of an oral piece must adhere to much stricter constraints than those observed by the speaker of a language. However, the nature of the constraints and the ways in which they prompt the utterer's memory are the same.

1. Defining Oral Traditions

Martina runs into the kitchen with small bouncy steps, eyes alert and a big smile on her face. She is wearing her precious pink-and-purple Barbie[™] shoes, and her teddy bear trails from her left hand, in defiance of her mother's continuous attempts at keeping him clean. She is happy to see me; she runs to my knees and climbs up onto my lap. The TV is exuding voices from the living room, and the sounds quickly coalesce into a tune. Martina knows the song quite well and confidently joins in the chorus, fully expecting me to do the same. We get to the 3rd verse: "...up above the sky so high, like a diamond in the sky..." Her glee doubles when I start singing with her, and triples – if at all possible – when I demonstrate that I am willing to sing the song not just once, but however many times she will start it again.

Martina knows several Italian songs like "Stella, stellina" and "La tartaruga", as well as a few Italian rhymes, but her repertoire of English songs is more extensive. Interestingly, she knows only one English rhyme so far. She has learned a few Italian rhymes from her mother, and surely the three of us must make for an amusing sight as we recite them at the kitchen table. For my part, I feel mindlessly happy every time I have the opportunity to revive the sounds and memories of a childhood, and with it, of a whole cultural identity, that have become for me inexplicably and irreversibly remote. Martina will not get to learn the more recent material that is currently passed around in Italian preschools and playgrounds, but I am certain that the rhymes and songs she is learning from us still constitute the bulk of the repertoire of all Italian children from the north of Italy.

The English songs she knows have been learned mostly from musical toys, the TV and her father. He also taught her the only English rhyme she knows; the media, even in their programs for children, seem to trade in musical pieces more commonly than rhymes. When Martina starts attending preschool regularly, she will surely learn a great deal more rhymes from her schoolteachers and, especially, from her interactions with her peers. She will also learn a number of traditional North American folk songs, from "Old MacDonald" to "Oh Susanna", that are part of the cultural knowledge of the people

that live on this side of the world. It is virtually assured that she will emerge from her childhood carrying a great familiarity with all these songs and rhymes; it is also virtually assured that she will not be exposed to a written version of any of these pieces.

This is my favourite example of an oral tradition, because it does not involve such far-away notions as ancient peoples from classical Greece or illiterate aboriginal tribes from the depths of the Amazon forest. Songs and rhymes are part of every person's childhood and of every school's playground. They are present and accessible to everyone, and the pleasure they create through their performance is more easily discussed because it can be evoked from personal experience. Pleasure is an important detail, since it is part of all oral performances for both performer and audience. The social dynamic in which the performance takes place is also important, whether it is a matter of providing group bonding through the sharing of an activity, or of asserting group structuring, for example, by highlighting the leader's role in the performance. Again, children's performances are a good example of these dynamics in action: songs are an effective bonding activity, and rituals such as counting-out rhymes provide a code of behaviour to be followed and a leader to help enforce it. However, many other examples of oral traditions exist, and although they may not be as universal as the playful musicality of children, they are nevertheless just as fascinating. These include traditional European ballads, the myths and fables of North American native tribes, the performances of Slavic epics, the poetry of the Homeric classics and of medieval Anglo-Saxon poetry, some of the texts in the classic Chinese canon *Shi Jing*, and many more.

Generally speaking, we can define oral traditions as spoken or sung social performances that remain relatively unchanged in context and form throughout the course of several generations. Oral traditions are found in every culture and, just as

each culture possesses its own distinct characteristics, the oral traditions of the world are also extremely varied. Many features are, however, common among all traditions. In his volume *Memory in Oral Traditions*, the cognitive psychologist David Rubin (1995) has compiled a list of such features. He writes that oral traditions

- Are universal; that is, they appear in all present cultures and past cultures that have been studied.
- Are fixed only within the accuracy of human memory.
- Exist in genres; that is, they appear in restricted, coherent forms.
- Are transmitted in a special social situation, such as a performance or a ritual.
- Are entertaining by modern literate standards, though this is not always their primary traditional function.
- Are considered as special speech, either art or ritual.
- Transmit useful cultural information or increase group cohesion.
- Are poetic, using rhyme, alliteration, assonance, or some repetition of sound pattern.
- Are rhythmic.
- Are sung.
- Are narratives.
- Are high in imagery, both spatial and descriptive. (p. 8)

Some children's counting-out rhymes actually defy two of the points that Rubin proposes: although some rhymes are minimally narrative – to the extent that can be accommodated by four or six verses – many children's rhymes are actually made up of gibberish language and do not carry any meaning. However, gibberish rhymes are

perhaps even richer in sound patterns than are other rhymes, and therefore seem to compensate, in terms of memorability, for their lack of narrative power. Also, rhymes can be sung but they can also be performed as rhythmic speech; in these cases, prosody is often enhanced but no musical melody is actually employed. Despite these observations, I believe that Rubin has succeeded in capturing a general picture of the features common to most oral traditions found worldwide.

2. The Constraints of an Oral Performance

Rubin's definition of oral traditions states that they are "fixed only within the accuracy of human memory". Many oral texts have remained unchanged for a very long time. Surprisingly, children's rhymes are perhaps the most stable oral texts we know. Other types of texts change slowly over time; for example, traditional European ballads seem to vary over the decades according to the geographical regions that they reach during their transmission. Yet other texts are never repeated in exactly the same manner, although their performances are always quite similar. Alfred Lord, and Milman Parry before him, observed this detail in their studies of Slavic epics. This fact may not be surprising if one considers that children's rhymes are much shorter than ballads, and several orders of magnitude shorter than Slavic epics. It is much easier to remember four lines (rhymes) than forty (ballads) or four thousand (epics)! However, it is most interesting to note that Parry and Lord found that epic singers were quite unaware of any differences in their performances: they maintained that the pieces they had recited were exactly the same. Often, ballad singers will also be unaware of any differences in their performances.

The fact is that the concept of "exactness" in the repetition of texts is very different when examined in literate and illiterate societies, because it depends on whether a former and a latter performance can be compared by means of external storage devices such as a written record or an audio recording. If not, then exactness can only be measured according to the normal capabilities of a person's memory. In our literate society, the availability of written records has created an expectation of exact copies, so that a literate listener will perceive as partially incorrect any instance of textual recall that is not verbatim. However, a listener who does not expect verbatim recall will perceive two performances to be identical if they succeed in telling the same story by employing the same references, rhythms, melodies and devices that are common to a genre. For this reason, variation within the transmitted texts of an oral tradition should be considered the norm and viewed as the sign of a healthy, lively oral genre.

In compiling his volume on oral traditions, Rubin set out to investigate the characteristics that allow these practices to harness the strengths of human memory when no written record is available. His findings demonstrate that oral traditions are not the simple repetition of a static text, but rather the product of a much more complex process:

It is not just that verbatim memory is absent. What I claim, and what is more intriguing, is that one specific variant of a song is not being transmitted at all. Rather, what is being transmitted is the theme of the song, its imagery, its poetics, and some specific details. A verbatim text is not being transmitted, but instead *an organized set of rules or constraints that are set up by the piece and its tradition*. In literary

terms, this claim makes the structure of the genre central to the production of the piece. In psychological terms, this claim is an argument for schemas that involve imagery and poetics as well as meaning. (Rubin, 1995, p. 6-7, emphasis added)

I briefly introduced the concept of schema in Chapter One in the context of semantic schemata. The idea was initially advanced in the work of psychologist Frederick Bartlett. Schemata (or schemas, as Rubin calls them) are sets of notions that become associated with each other in memory in the course of an individual's life experience. In advancing his argument, Rubin observes:

> Bartlett's data led many readers to the erroneous conclusion that meaning or gist is the only dominant and possibly the only dimension along which organization that aids memory takes place. It is this claim that schemas for surface structure of stimuli do not aid memory that is contested. (p. 70)

Rubin states that the evidence offered by his and others' experiments supports the idea that the form of an oral piece is an important factor in its recall:

> When the surface form is important, it can be recalled [...]. When surface structure organizes the stimulus or restricts possible responses, it aids memory [...]. [...] Where surface organization is great but meaning is not, recall is organized with respect to surface structure not gist. (p. 72)

In other words, notions can be remembered because they are associated to one another in memory according to the experiences of the individual. In the same way, surface structures also create a web of associations. For surface structures, Rubin

intends that same "organized set of rules or constraints" mentioned in the quote above. Therefore, Rubin's argument is that in recalling a piece in an oral tradition, the memory of the performer will proceed along the organised progression of the story, but also along the organised progression of the performance's formal structure. In this way, the more the performance advances, the more it is easy to remember the next scene but also the next rhyme, the next formula, the next alliteration, and so on. As I will illustrate in more detail below, this type of "remembering" is so centred on adhering to formal constraints, rather than recalling an exact word or verse, that it is essentially indistinguishable from spontaneous composition. The spontaneity of these oral compositions, as well as the ways in which lexemes, formulas, verses and stanzas are cued in memory to advance the composition in a serial manner, seem to describe a performance just as much as an instance of common speech.

Rubin discusses three types of constraints at work in an oral performance: meaning, imagery and sound constraints. The sum of these provides a structural organisation that effectively cues the spontaneous composition/recall of an oral piece. Below I will illustrate first the three types of constraints, and then their integration.

a. MEANING CONSTRAINTS

The meaning constraints of an oral piece are the themes included in its story. Themes are groups of ideas, associations of elements that typically recur together. While the general story of a poem is the overarching theme for a specific performance, its plot is developed with the aid of sub-themes, which help with the development of a specific part or episode of the narration and add detail to the story. At times, subthemes even provide useless digressions and incoherent twists in the plot.

The meaning constraint is the most straightforward of all constraints, because it corresponds to the traditional idea of schema as a set of associated semantic notions. I have already discussed one representation of schemata in Chapter One, in my discussion of semantic networks. Two more examples of theoretical representations of schemata from psychological studies are scripts and story grammars. These representations are logical aids that help illustrate the psychological processes involved in an individual's experience of semantic domains and of cause-effect progressions. They illustrate the fact that an individual's expectations about being in a certain situation (for example, going to a restaurant) include a collection of trivial actions (entering, seating, eating, paying the bill and so on) that take place in a specific order.

Rubin explains the benefits of each type of theoretical representation and discusses some of the experiments that were conducted to prove its applicability. What is important here is that the meaning constraint operates in the recall of a story (and its sub-stories) along the lines of pre-existing schemata. This fact is true both in terms of semantic domains and of cause-effect progressions. This means that a story will be more memorable for the singer, and more understandable for the audience, if its elements are semantically and causally connected in the story as well as in the singer's and audience's everyday experiences.

Elements that are foreign in everyday associations are more likely to be forgotten and misunderstood; moreover, they are more likely to be replaced with elements that belong to the semantic domain and cause-effect progressions presented in the story. If an oral piece follows a pre-existing schema, it will also be easier for the audience to draw inferences about those parts of the narration that were not explicitly mentioned by the singer. Furthermore, if two or more oral pieces follow the same schema or include

the same sub-schema, it is possible and even likely that parts of them will be confused and substituted for each other, even if these parts are not similar in their details. This is because logically, they will serve the same function in the story, which means that they will satisfy the constraints of the same theme and therefore, from the point of view of meaning alone, they are equally likely to be recalled.

The meaning constraint is important for two specific reasons. First, the storyline is an effective cuing mechanism for the singer. This type of cuing is progressive (i.e., from start to finish) and does not allow easy access to parts of the story that are beyond the next step in the sequence. Second, meaning constraints are important because they explain to a certain extent the use of sub-themes. The meaning constraint will allow use of any element that satisfies the particular schema in question; for example, if a message is to be delivered, the messenger can be introduced straight away, or the messenger theme can be explored more extensively. If the singer has listened to another singer's tale and wishes to borrow his messenger theme, he can do so at will as long as the requisite for a messenger is fulfilled in his story. The cause-effect progression will be satisfied whether the messenger is merely mentioned or whether the singer describes him at length. In this sense, the meaning constraint allows for the inclusion of expansions in the performance. At the same time, it keeps each expansion in check by ensuring some level of cohesiveness in the overall performance and within each sub-theme.

The meaning constraint acts very much like the thematic constraint I described in Chapter One. An oral performance, just like any monologue or conversation, has the freedom to expand in a number of different directions. In both cases, the speaker's production will also be prompted by semantic and cause-effect associations. This fact

will ensure that the utterances remain internally coherent, and that the audience will be able to follow their progression as well as, to a certain extent, anticipate their outcome.

b. Imagery Constraints

The second type of constraint that Rubin explores is that of imagery. Rubin observes that accessing a mental picture of certain events provides a quick method for recalling these events in memory. Most oral traditions include graphic descriptions of the surroundings of a story; moreover, as the story develops, the characters move and travel within these surroundings. Rubin suggests that singers routinely create and recall mental images that represent the settings of a story and the travels of characters. These images are powerful memory aids because they allow the use of visual memory in conjunction with verbal memory. Current evidence has shown that visual and verbal data are processed by different areas of our brain; by using a combination of both visual and verbal stimuli, a narration can engage more resources in order to be recalled quickly and effectively.

About mental images, Rubin (1995) states: "the experimental studies provide evidence for viewing imagery as a 'picture or movie created in the head' in which size, distance, color, shape, location, and intermediate steps in the movement of objects all function much as they would in perception [...]" (p. 41). He later continues: "in recall imagery appears to function mainly by providing a form of organization for a set of stimuli that would otherwise be less integrated [...]" (p. 52). In other words, singers create a mental picture or movie of the story that they narrate, and this movie helps to keep the story and its details together in memory. Singers may or may not be consciously aware of accessing an image; in both cases, the image will facilitate recall.

Since mental imagery functions similarly to actual perception, it also produces a sense of immediacy:

In addition, specific details aid in maintaining emotional balance [...] and [...] increase the sense of emotionality, intimacy, and immediacy when compared with abstract statements that remove the events described from particular situations. [...] Concrete details that increase emotionality, intimacy, and immediacy should lead to more frequent tellings. (p. 56)

Imagery and meaning constraints integrate tightly, and the mental movie is obviously a combination of the two. The story progression and the progression of images in the narrator's mind reinforce each other in creating a continuum. Moreover, the use of images brings two advantages. First of all, it is likely that a specific scene in the story will be recalled in more detail if associated with an image. Second, images will allow for a type of recall less serial than that of the story alone; if a scene is forgotten, remembering the setting for a later scene will facilitate the continuation of the story even if the narration is missing a logical and/or temporal passage.

Again, there is a marked similarity between this constraint and the imagery constraint I illustrated in Chapter One. Both of them provide cues for the description and enumeration of details, and both of them create a sense of involvement and intimacy between the speaker and the listener. While I have described the imagery constraint of speech as a purely semantic cue, there is no reason to exclude that visual memory may aid these semantic processes in cuing material for production. With adequate experiments, it may be possible to collect evidence about the use of visual memory in speech.

c. Sound Constraints

The third and undoubtedly most innovative type of constraint investigated by Rubin is that of sound patterns, including rhythm, alliteration, word repetition, rhyme and assonance. Another sound pattern that is important in the recall of a musical piece is the melody that the singer must follow during the performance. These types of patterns function as memory cues in different manners. For example, one of the primary functions of sound repetitions is to limit the choices of the oral performer:

The repetition of a sound is an aid to memory. When a sound repeats, the first occurrence of the sound limits the choices for the second occurrence and provides a strong cue for it. The repetition is emphasized if it occurs in a specific place, such as alliteration [...]. [...] The most common occurrence of repetition of sound in modern English is, of course, rhyme. (Rubin, 1995, p. 75)

According to several experiments conducted by Rubin, "once rhyme organization is apparent at learning, it can have an even longer-term effect than meaning" (p. 83). This means that if several people memorise a rhyming piece, but some of them are aware of the rhyme while some are not, those who are aware of the rhyme will recall the piece for a longer time than those who are only aware of its meaning. Therefore, rhyme is not only effective in cuing memory by restricting the choice for the endings of the next few verses, but is also an important device for recalling a piece over an extended period of time.

Word repetition is also a very important device. The identity of an oral tradition often includes not only melodies and stories, but a specific choice of diction as well. By sharing in a specific vocabulary, each composition is certain to repeat many of the words

that are used in other compositions. Words are also often repeated within an oral piece, and some texts even include the repetition of entire lines; many ballads, for example, contain a refrain that is sung after every stanza.

Moreover, oral traditions have their own particular type of formulas, which are groups of words that are preferentially or exclusively used together. Parry initially hypothesised the use of formulas in the context of his studies on Homer, and subsequently documented it with Lord in the course of their research on Slavic oral epics. In Homer, these groups of words are paired to a specific location within the metre of the verse. In other compositions where the metre is less strict, the location of formulas can vary. An example of formula in the Odyssey is "brilliant Odysseus", which recurs only at the end of a verse (Rubin, 1995, p. 200). In English ballads, the mention of "gold" is often followed by the mention of "fee", where both words are synonyms for "money". Formulas can be common to an entire tradition, but they can also be specific of a certain singer. In their studies on Slavic epics, Parry and Lord found that themes and formulas were shared among singers, but were not ubiquitous in the tradition; every singer employed a combination of formulas and themes according to his taste and habits. Formulas are often associated with a theme and therefore, the number of formulas that a singer employs depends on the number of themes with which he is familiar.

Parry and Lord's work on formulas and oral-formulaic composition is quite extensive, and their findings have been extremely important in propelling the study of oral traditions. Their accounts of a manner of oral composition that is spontaneous and yet highly consistent coincides flawlessly with Rubin's theory of constrained serial recall. In terms of the present discussion, it is important to note that the formulas of oral

traditions are identical to the formulas of common speech: some are highly restricted and can only be uttered in certain contexts, while others are looser and simply indicate the preferential co-occurrence of two or more words. In both cases, they aid memory because a cue to the first word in the formula will automatically trigger the recall of the entire formula. Moreover, formulas can be associated with a specific rhythm or melodic pattern and therefore fill a whole verse, half verse or intonation chunk in an automatic manner.

Rhythm also contributes significantly to the overall organisation of an oral piece:

Rhythm functions like other constraints or forms of organization to limit word choice, in this case to words with the correct number of syllables or stress patterns [...]. In addition, the rhythm provides a global organization, allowing singers to select, substitute, and add or delete whole rhythmic units (e.g., verses) and still continue. Rhythm also emphasizes certain locations within lines, which facilitates other constraints, such as the placing of rhyme and alliteration on stressed syllables. (Rubin, 1995, p. 11-2)

Rhythm brings another benefit to oral performance. Most of the constraints that I have illustrated so far become established as the piece is being sung – for example, the story proceeds serially and cannot be narrated in a random order. In the same way, sound repetitions help in limiting word choice, but only after the first instance of the repetition has been uttered. The more a performer sings, the more constraints s/he will have established with the previous verses. For this reason, a "running start" effect can

often be observed when, in order to recall a verse or a word, it is necessary to begin with performing or reciting the previous passage first. But rhythm works differently:

The meter for every line of an epic, or every stanza of a ballad, is the same. Once the first metrical unit is sung, the meter is set for the whole piece. That is, unlike the other forms of constraint studied, the rhythmic organization of a piece is fixed very early in its singing. For this reason, rhythm is less sensitive to the cues provided by a running start and thus can function in its absence. If the cues needed to recall the next part of a piece are lacking, the meter is still known, often making it possible to start again at another rhythmic unit. (Rubin, 1995, p. 87)

The effects of melody on recall have been documented in a series of experiments centred on the recall of lyrics being sung to a tune.²⁴ Some experiments employed actual lyrics, while others used nonsensical words. All experiments demonstrated that melody is a major contributor in recalling songs and oral pieces, especially when it has a repeating pattern. One experiment showed that repetitive melody is a more effective mnemonic aid than rhythm alone. The evidence also suggests that, conversely, lyrics can be important in the recall of melody. It appears that lyrics and music often form an optimally fitted pair, in which the recall of one element will assist with the recall of the other in a mutually beneficial system of cues and constraints.

²⁴ See Wallace (quoted in Rubin, 1995, p. 289); Crowder, Serafine & Repp, 1990; Serafine, Crowder & Repp, 1984; Serafine, Davidson, Crowder & Repp, 1986.

There is no reason to discount the hypothesis that prosody may work in the same way as melody. Some words and phrases of everyday speech may be more memorable than others because of the prosodic contour with which they have been uttered. Memorability may be enhanced to the point that the phrase is uttered almost exclusively with a specific prosodic contour. This is indeed the case with many instances of slang (e.g., "no kidding?!", "whassup?!", "yo, man") and common formulas (e.g., "what do you know??", "no way!", "are you crazy?!", "as a matter of fact, ...").

d. CONSTRAINT INTEGRATION IN ORAL PERFORMANCE

Meaning, imagery and the sound constraints of rhythm, melody and sound repetition amalgamate in the performance of an oral piece. The sound constraints are usually dictated by the tradition in which the performance takes place, while the meaning constraints are dictated by the specific story to be narrated. Imagery constraints are linked to the story, but are also influenced by the tradition and culture in which the story is told. The combination of the three different types builds a system of constraints that is much more stable than any one constraint taken singularly: "Combined constraints produce effects much larger than those of the individual factors by decreasing the memory load and increasing the number of cues to recall" (Rubin, 1995, p. 90).

In oral traditions, different sets of constraints exist at many levels: the word, the formula, the rhythmic line, the stanza, the theme, the whole piece, the story pattern and the genre. As Rubin points out, "at each of these levels, constraints combine to limit choices for recall and increase stability" (p. 101). The result of these multiple layers of constraints is a built-in system of error correction. In an oral tradition,

the song maintains itself. Oral traditions contain songs that have survived as systems within boundaries for long periods. Although one cannot posit an intentional 'urge for survival,' there is a tendency for songs (as well as genres) to maintain themselves or change slowly, which has been described as *self-correction* or *homeostasis* [...]. By definition, songs and genres that do not have this property could not be part of an oral tradition for any length of time. (p. 97)

The idea of multiple constraints is particularly appropriate within the framework of interference theory, which posits that forgetting is the inability to locate information in memory due to the presence of other similar pieces of information that cause interference. In this scenario, the task of locating a target is made easier by the availability of cues that restrict the choices around the information to be retrieved. The more cues available, the easier it is to discriminate a particular item in memory. However, interference is still present even with a high number of cues:

Such organizational cues do not prevent all failures in recall. In fact, they can produce a change from the original word by cuing a word that was not in the original variant of a song strongly enough. [...] Such 'interference' could result in intrusions of words, formulas, lines, verses, characters, props, and even themes from other songs. Such intrusions do occur when cues facilitate them. (Rubin, 1995, p. 155)

It is also not given that singers will remember the exact cues they have heard in somebody else's song. If they did so consistently, they could also follow the cues consistently, and it would then be possible to repeat an oral piece verbatim. But by chance, or because of personal habit, changes are always introduced and new verses

composed. With their meaning, imagery and sound, these verses prompt the singer's memory for yet more verses that will observe their constraints. Verses are not simply remembered in their entirety from an earlier retelling; instead, each of their parts is prompted from memory in a piecemeal fashion. For this reason, it is impossible to distinguish recall from composition:

It follows from such a process that a pair of lines or words that are linked together by theme, rhyme, or alliteration will not necessarily be recalled more often than other pairs of words. However, if the first member of the pair is recalled, then the second will almost certainly follow. [...] The process of remembering just outlined looks much like the process of composition. In fact, no distinction is made in the basic way cuing and item selection works. (Rubin, 1995, p. 178)

To sum up, Rubin bases himself on interference theory to advance his theory of constrained serial recall in oral traditions:

Recall starts with the first word of the song and proceeds in a linear fashion. Words sung are cues for words yet to be sung. If words are to be recalled, they must be discriminated from other words in memory. The general constraints of the genre and piece, especially rhythm, act as cues from the start, with the singing filling in other cues as it progresses. A piece fitting the constraints of the genre results, not necessarily a verbatim reproduction of a piece produced earlier. Where the constraints are strong, they will limit variation without the help of particular cue-item associations formed when a piece was heard. Where only one variant has been heard, especially when it has been
heard repeatedly using spaced practice, individual cue-item associations will be more important and will further decrease variation. This process, after the initial, often conscious decision to sing a song has been made, can go on without conscious intervention, using what has been called implicit or indirect memory. The serial-recall method, however, means that knowledge in oral traditions is not routinely accessed without the cues provided by a running start and often cannot be accessed without them. Thus questions about the contents of a piece can often be answered only after the piece is sung. (p. 192)

3. A Look at Some Data

To this point, I have reviewed the cognitive constraints at work in an oral performance. The remainder of this chapter is devoted to the analysis of oral texts that have been recorded in written form. These analyses will illustrate the ubiquitous nature of the constraints discussed above. My purpose is twofold: I wish to demonstrate that the singer of each text was aided in his/her performance by the large number of constraints at work in the text; I also wish to show, by the breadth of sources employed, that Rubin's constraints are truly universal features.

Of course, a major drawback in employing written records as opposed to audio or video recordings is that some of the most important features of these texts must be ignored. I will not be able to discuss their melodies, their pitches or tempos, and therefore, I will not be able to illustrate some of their sound constraints. However, this limitation is unavoidable, since to my knowledge no sound or video recording of these texts is available. My first group of examples is from children's counting-out rhymes, which are generally overlooked in academic research, and my second group is from European ballads, which were performed well before audio and video recorders were commonly in use. True oral traditions are produced by illiterate people, whose lives and customs have seldom become objects of extensive inquiry. Hence, my raw data is scarce. However, I believe that my use of these texts is nevertheless justified, given the number of constraints that are clearly observable even in their written forms. Their aural dimension would only reveal more constraints, in further support of my argument.

a. COUNTING-OUT RHYMES

Counting-out rhymes are performed by groups of children at play, at a time when one specific child must be chosen in order to perform a task (usually an unpleasant one). A very comprehensive survey, and one frequently quoted in discussions on counting-out rhymes, is Henry Bolton's 1888 volume *The Counting-out Rhymes of Children*, in which the author advanced the hypothesis that counting-out rhymes are remnants of ancient divination rites. Although the claim is fascinating, the supporting argument proceeds along a line of reasoning that I find questionable at best; the toys of "advanced" cultures are viewed as being derived directly from the implements of "primitive" ones (Bolton's only example is that of toy bows and arrows). The author's mindset is evident in such gems of political incorrectness as the passage in which he states that the counting-out rhyme "is a pastime with the children of civilized and semicivilized races of the most diverse origin" (p. 9). Nevertheless, the book is remarkable for the breadth of the material that it collects, including one rhyme in Penobscot dialect, two in Japanese, one in Hawaiian, five in Marâthi, one in Romany, three in Arabic, ten in Turkish and Armenian, eight in Bulgarian, three in Modern Greek, seven in Swedish, three in Portuguese, three in Spanish, nine in Basque, five in Italian,²⁵ 21 in French, 40 in Dutch, 18 in Platt-Deutsch, 269 in German and 464 in English, for a grand total of 873 rhymes. Clearly, the selection is skewed towards the Anglo-Saxon languages, primarily because of the sources to which the author had access. However, the collection proves effectively that counting-out rhymes are a global phenomenon.

I have no doubt that the following is presently the most popular children's counting-out rhyme in North America. Quite interestingly, as Bolton documented, it already was the most popular rhyme in 1888.

Eenie meenie miney mo Catch a tiger by the toe If he hollers let him go Eenie meenie miney mo

Bolton stated that this rhyme was the most widespread in the United States in its variant that substituted "nigger" for "tiger" (p. 46). In their 1976 volume, Mary and Herbert Knapp report:

No doubt the best-known racist rhyme in the United States, both in the past and today, is "Eeny, Meeny." Versions of this counting-out rhyme go back to revolutionary times, but it was not until the passage of the Fugitive Slave Law in 1850 that the second line was changed to "Catch

²⁵ To my disappointment, the Italian selection comes from a volume on Venetian rhymes, and therefore all rhymes are in Venetian dialect, a selection which is hardly a fair representation of Italian rhymes in general.

a nigger..." [...] As early as the 1880's, genteel mothers were encouraging their children to substitute some other word for "nigger" in "Eeny, Meeny." At first they had very limited success, but slowly replacements like "lion", "tiger", "monkey", "rabbit" and "dummy" began to be accepted. During moments of national crisis, "Hitler", "Tojo", "Castro" and "the Viet Cong" have appeared in the rhyme. Today "nigger" is still used, but less often than numerous other words. (p. 197)

Unfortunately, the Knapps do not disclose the sources of this information. Interestingly they state that, at the time of their writing, the most common rhyme in American playgrounds was "One Potato, Two Potato", and that "Eenie Meenie" (in any choice of its multiple spellings) was second. In 1995, Rubin conducted several experiments with counting-out rhymes involving undergraduate students from Duke University, in North Carolina. Again, he found that "Eenie Meenie" was the most commonly recalled rhyme, followed by "One Potato" (p. 233). Rubin considers the version of "Eenie Meenie" that includes "tiger" as the standard instance of the rhyme, and other versions as derivatives. This fact stands as testimony of the establishment of this specific variant over all others.

Upon examining the text of "Eenie Meenie", it becomes evident that the sound devices used in the rhyme are its preponderant characteristic. The rhyme has little to offer in terms of narration – indeed, the first and last verses are completely nonsensical, and the fate of the tiger is hardly developed into a story. However, the sound design is quite intricate, and rhetorical devices are ubiquitous. These are:

1. Alliteration: meenie-miney-mo, tiger-toe, he-hollers-him.

2. **Rhyme**, both internal and external: eenie-meenie, mo-toe-go-mo.

3. Word repetition: The entire first verse is repeated as last verse.

4. **Assonance**: cAtch-tIger-bY, hollErs-IEt, If-hE-hIm.

5. **Rhythm**: All verses have a 4-beat rhythm and present a coherent syllabic structure of two stressed-unstressed trochaic feet and one stressed-unstressed-stressed cretic foot ($-\cup - \cup - \cup -$).

If my analysis is correct, the only words not part of repeating sound patterns are "a" and "the" in the second verse – although of course, they still contribute to the rhythm of the verse, and therefore constitute part of the sound structure of the piece.

Let us now look at a counting-out rhyme used in a completely different culture. The following rhyme was by far the most common throughout my childhood in the north of Italy, and still remains the most popular to this day:

Ambarabà ciccì coccò Tre civette sul comò Che facevano l'amore Con la figlia del dottore Il dottore s'ammalò Ambarabà ciccì coccò

The rhyme translates to the following: ambarabà ciccì coccò \ three owls on the nightstand \ who were making love \ to the doctor's daughter \ the doctor fell ill \ ambarabà ciccì coccò. The gist of the rhyme is thus quite peculiar: the first verse, which is repeated at the end of the rhyme just as in "Eenie Meenie", is gibberish; the rest of the rhyme narrates a thoroughly inconclusive story. The fact that the meaning of the

rhyme is hardly taken into consideration by children and adults alike can easily be gathered by observing that the story, in its modest act of narration, speaks of sex between three animals and a woman – yet no one has ever found reason to consider the rhyme obscene or to attempt its removal from school grounds. Indeed, parents often teach the rhyme to their children without a second thought. Most often, children are not aware of the meaning of words like *facevano l'amore*, yet they are perfectly happy to recite the rhyme and have no trouble recalling its text. The truth is that this rhyme possesses an extremely ingenious and memorable sound structure, which, in its brilliance, obscures completely its tale of bestiality and illness. The devices in use are:

1. **Alliteration**: *ciccì-civette* ("ch" sound), *coccò-comò-che-con* (k sound), *facevano-figlia*, *del-dottore*, *l'amore-la* (the first article/noun combination is spoken continuously as one sound: l'a = la).

2. **Rhyme**, external: *coccò-comò*, *amore-dottore*, *ammalò-coccò*. Note the pattern aa-bb-aa, further emphasised by the fact that "a" rhymes are made of words with an end-of-word stress.

3. **Word repetition**: The entire first and last verse repeat; also, *dottore* is repeated in contiguous verses. Moreover, the nonsensical words *cicci* and *coccò* are constituted of redoubled sound syllables with an added end-of-word stress in order to respect the rhyme.

4. **Assonance**: AmbArAbÀ, trE-civEttE, trE-chE, fAcevAno-l'Amore-IA-figliA, AmmAlò-AmbArAbÀ.

5. **Rhythm**: This rhyme also has verses of four beats each, but their structure is a little more complicated than that of "Eenie Meenie":

The overall rhythmic construction is shaped like a chiasmus (a-b-c-c-b-a), with verses 1-6, 2-5 and 3-4 sharing the same metre. Verse 1 (and 6) starts with a trochaic foot but then immediately reverses into a series of three iambic feet. Verses 2 and 5 have the same structure of the verses of "Eenie Meenie"; however, in order to allow the rhyme to fall on an accented end of verse, these verses have cut their syllable count short. In fact, verses 2 and 5 contain seven syllables while all other verses contain eight; the shortened count is achieved by employing contractions in the words at end of verse, such as *comò*, which is short for *comodino*, and *s'ammalò*, which is a common contraction of the full form *si ammalò*. Once again, we can see that in this rhyme almost every single word is heavily employed in the production of sound devices. Only the articles *sul* and *il* are not directly involved in any sound framework other than the metre.

In order to provide an example from yet another culture, I asked a friend what was the most common counting-out rhyme in France. He replied without hesitation that the rhyme is called "*Am Stram Gram*". A quick Internet search yielded a huge number of instances of the rhyme, which correspond to one another aurally even though their spelling is quite varied; however, none of the different spellings are truly incorrect, since all the words contained in the rhyme are nonsensical (and thus also untranslatable):

Am stram gram Pic et pic et colégram Bourre et bourre et ratatam Am stram gram (pic dam)

Some of the versions I have found on the Internet do not include the last two words, *pic dam*, while some others do. This fact is due to the common custom of adding a few words at the end of the rhyme to make the choice of the counted-out person more evident. In the same manner, some English rhymes add 'O-U-T spells out', 'out goes he' or 'Y-O-U is you'. Once again, the sound structure is the important feature in this text:

1. **Rhyme**, internal and external: *am-stram-gram*, *gram-colegram-ratatam-gram* (or *dam*, depending on the chosen ending).

2. **Word repetition**: The whole first line is repeated as last line; verses 2 and 3 both contain a double repetition with *pic* and *et* and with *bourre* and *et*; finally, if the longer ending verse is chosen, a further repetition of *pic* occurs.

3. **Parallelism**: Verses 2 and 3 have the same word structure, emphasised by the repetition of *et* in the same locations of the verse.

4. Assonance: Am-strAm-grAm, rAtAtAm, Am-strAm-grAm-dAm.

5. Rhythm: The metre is as follows

 $--- \setminus - \cup - \cup - \cup - \setminus - \cup - \cup - \setminus - - - (\cup -)$

The first and last verses (in the simpler form) have only three beats each, while the centre verses contain four beats. All verses contain an odd number of syllables. The metric structure is again chiasmatic unless the longer ending verse is chosen, in which case the iambic foot that terminates the verse manages to give even more emphasis to the accented syllable at the end of verse.

The most interesting feature of this rhyme is the three-beat rhythm of the first (and sometimes also the last) verse. In all the collections of children's rhymes that I have had an opportunity to examine, four beat verses are by far the most common, while three and five beat ones are rare. Within the collections of French rhymes that I have been able to access through the Internet, I have noticed that four beat verses are still prevalent, but three beat verses seem to be more common than in other languages.

Finally, I offer an example from China. A classmate from Canton (Guangzhou) recited the following rhyme for me, reporting that it is extremely popular with kids all over the country. Despite the fact that her place of residence is home to the Cantonese dialect, the rhyme is in standard Mandarin Chinese, which is spoken throughout China.

一二三四五

上山打老虎

老虎不在家

放屁大王就是他

The Pin Yin transliteration with tones is as follows:

Yī èr sān sì wŭ

Shàng shān dă lăo hŭ

Lăo hŭ bù zài jiā

Fàng pì dà wáng jiù shì tā

The rhyme translates as: one two three four five / let's climb the mountain to hunt the tiger / the tiger is not home / the great king of farts is him (or her – the child who is identified by chanting the rhyme). Once again, the rhyme is minimally narrative. The sound effects are pervasive:

1. **Rhyme**: The scheme is a-a-b-b, or wŭ-hŭ-jiā-tā. Note that in Chinese a rhyme between two words is still defined as the correspondence of all sounds from the last

vowel to the end of the word, as in alphabetic languages. However, in Chinese there is the added difficulty of having also to match the tone of the vowel (for example, ā and à do not correspond to the same sound).

2. Word distribution and repetition: The first verse is a succession of the first five numbers, and is therefore extremely easy to remember even for the youngest child; also, the first and second couplet are nicely tied together by anadiplosis, or the repetition of 老虎 (tiger) at the end of the second verse and at the beginning of the third.

3. **Rhythm**: The rhyme includes three pentasyllabic verses and one heptasyllabic verse in closing. The number of syllables and their metre are both characteristic of Chinese poetry, although in classical poetry penta- and heptasyllabic verses are usually not combined. In both types of verses, the accents fall on the second, third and fifth syllables, and the heptasyllabic verse also includes an accent on the seventh syllable (see Cheng, 1987, p. 213):

∪**-**-∪-/∪--∪-/∪--∪-∪-

Chinese poetry is peculiar for its use of tones. In classical Chinese poetry, the metric structure encompasses not only a framework of accented syllables, such as the metre that I have shown above, but also a precise tone configuration (Cheng, 1987, p. 214). By analysing the counting-out rhyme with the same technique as the one employed for classical poetry, we find that it contains an interesting arrangement of tones. In tonal analysis, the first tone (with a flat inflection: —) is a longer sound and is placed in a category of its own. The other three tones, with shorter inflections (rising ⁷,

dipping , dropping) are grouped together. I will use - and > respectively to visualise

the long-short tonal structure in this rhyme:

->->>> >->>>> >>>>-->>>>

As we can observe, the first three tones of verses 1 and 2 are inverted: verse 1 has a long-short-long structure while verse 2 has a short-long-short one. Both verses, however, end with three short tones. Verses 3 and 4 are more surprising: they are both comprised of short tones up to the very last sound, the rhyme, which is a long tone. This analysis shows that even at the level of tones, the structure of the rhyme is highly symmetrical. Of course, it is illogical to think that someone with formal knowledge of metrics would compose this rhyme with careful attention to the tonal arrangement. However, the length of tones is a feature of the prosody of Chinese verses, and for this reason a highly regular tonal structure shows that the rhyme in question is extremely rhythmic.

4. **Assonance and Alliteration**: In Chinese, the nature of the language and its limited range of syllabic sounds require the scholar of rhetoric to be extremely cautious²⁶ in identifying assonances, alliterations and homoeoteleutons.²⁷ Our rhyme

²⁶ Chen Wangdao (1932), the forefather of Chinese rhetoric, did not mention these figures. Other scholars, such as François Cheng (1987), take the figures into consideration in their own

includes only two alliterations, Sān-Sì and SHàng-SHān, and one rather feeble assonance, fÀng-pì-dÀ.

Counting-out rhymes are extremely memorable. "Eenie Meenie" has remained virtually unchanged for over a century, a remarkable fact considering that the preschoolers who use the rhyme do not yet have the ability to read or write. This is a clear example of an illiterate social environment in which a performance has remained almost completely unaltered from generation to generation.

From the examples above, it appears evident that the facts narrated in countingout rhymes, if indeed the rhyme is made up of intelligible language, are minimal and often nonsensical. Hence, the meaning constraint of a rhyme, if any exists, is hardly a factor in recall. On the other hand, I have illustrated at length the many sound devices that cause the structure of a rhyme to be extremely redundant: its regular rhythm, its end-of-verse and internal rhymes, its alliterations, word repetitions, and assonances, its chiasmatic and parallel configurations. I will now examine another oral tradition in which the same constraints coalesce in different combinations. The resulting texts are quite unrelated to children's rhymes, yet they too remained consistent for centuries.

rhetorical analyses. I have argued elsewhere in favour of their validity, and therefore include them in my present discussion.

²⁷ A homoeoteleuton indicates that two or more words end with the same sound.

b. TRADITIONAL EUROPEAN BALLADS

In the course of my travels, I have had the good fortune to experience instances of the transmission of oral traditions in other cultures. On 30 December 1997, the rain was coming down in buckets over the tiny west-coast community of Liscannor, County Claire, Ireland. Despite the uncooperative weather, my friends and I made our way from the cottage in which we were staying to the centre of the village, some two kilometres away. The houses were scattered within a small and flat area. The school was at the near end of town; the pub, alas, was at the far end. After an interminable walk, we finally reached our destination and slipped inside for a well-deserved pint of beer. As we sat in merriment and banter, we heard a hush zigzagging through the tables, and then a thin male voice made its way to our end of the room. The man was singing in English, and my companions recognised the ballad at once. I could not hear the lyrics very well, but the tune was clear and sad. It lasted for a few stanzas, and as the singer came to the end of the piece, several listeners clapped and the conversation in the room quickly resumed. In the course of that evening, the room fell silent five or six times. Not even the bartender would produce a sound while the male and female voices intoned their melodies; to interrupt a singer was obviously a sign of profound disrespect. The singers would not stand up during their performance, but rather would continue sitting as their song emerged from the midst of a conversation. Later in the evening, a woman from our table also contributed to the tradition. After inhaling deeply she started a song in Celtic. She was the most applauded singer that night, perhaps because of her choice of language. As we made our way home in the crisp and freshly rinsed night, I was acutely aware that what I had experienced was ordinary for many Irish people, but would remain quite unrepeatable for me.

My friends confirmed that the performance of traditional ballads and songs is common in many pubs across Ireland, especially those located in rural communities. As I later learned from research works on this subject, the practice is also common in pubs and households across Scotland and parts of England (see Niles, 1999). While nowadays ballad-singing is mostly a social pastime, consulting Nettel's account of English popular songs (p. 180-6) reveals that as recently as that book's publication in 1956, ballads were still performed for a fee by professional street singers.

The traditional ballads of the British Isles are a kind of popular song with a very distinctive formal structure. The stanzas, which can vary in number from just a few to over 20, are composed most often of either two couplets or a quatrain. A number of ballads, usually identified as the older specimens, include a refrain that is repeated after every stanza. It is unclear whether each ballad was composed to a specific tune, or whether the tune was adapted to the lyrics at a later time. The most popular ballads were often sung to more than one tune. In many cases, we have records of several different versions of the same ballad, and while it is possible at times to identify the relatively older and newer variations, it is incorrect to consider any one version as the original, because ballad-singing was an extremely widespread phenomenon and each ballad singer would compose his or her own ballads by employing well-known popular tales. Various ballads telling the same tale with stylistic differences were often in circulation at the same time; singers would sometimes copy the tales – or just parts of them – from each other, but would adapt the telling to their own personal style. Because of the constant mingling of versions that took place in the social milieu of singers and their audiences, it is best to think of related ballads as a series of

performances centred around the same story, rather than as one primary performance to which secondary pieces, or variations, are subordinated.

The origins of traditional English ballads are a topic of debate. Alan Bold supports the conventional view that ancient religious carols are to be considered the first examples of popular ballads. The first ballad-like carol to be recorded in writing is titled *Judas* and was written in the thirteenth century; Bold affirms that its subject matter and style are close to those of later popular ballads. He differentiates ballads into popular pieces composed by amateurs and professional pieces composed by minstrels, stating that the professional pieces were attempts at copying the popular pieces, and that the results are actually of a lower quality than the amateur compositions. He explicitly rejects Folwer's argument, which I will examine next.

In his extensively researched volume on the literary history of English ballads, David Fowler attributes the birth of the tradition to the court minstrels employed by the barons of the north and west of England. After the War of the Roses, baronial power declined and patrons were no longer able to maintain the minstrels, who were forced to reduce the length of their extensive performances and to adapt them to popular taste in order to sell their services in the public street. Fowler states that the first true instance of popular ballads is to be found in a few pieces that narrate the adventures of Robin Hood; he conjectures that these pieces were recited at first, and only once the tradition became more established did music become part of the performances. Fowler denies that religious carols are to be considered as first instances of popular ballads, attributing this faulty assumption to the fact that Francis Child included some carols in his extensive work *English and Scottish Popular Ballads*, which is now held as the paramount work in ballad scholarship. The Child canori, published in 1882, is an extremely accomplished

attempt at recording all known instances of English and Scottish ballads in all of their variants, and includes both international and insular ballads. Fowler maintains that Child's selection of these pieces stemmed from personal preference, and that the inclusion of carols in his work should not be hailed as evidence for a theory on the origins of ballads. As a result of his argument, Fowler affirms that the first popular ballads only appeared in England in the fifteenth or sixteenth centuries.

However, we should also keep in mind that balladry was primarily a European phenomenon, and only after it had developed on the continent did the tradition cross the Channel. French ballads provided a first model of composition for English poets, but very quickly the newly founded tradition of English balladry adapted to the sounds and prosody of its own language and developed a distinctive style. Some ballads were extremely successful in travelling across Europe in a number of variants, roaming from Scandinavia in the North, to Spain and Italy in the South, to Poland, Hungary and the Slavic populations in the East, to the British Isles in the West. The international ballads, as they are commonly called, were translated into the local language at each step in their travels, adding exponentially to the number of recorded variants. Other ballads were composed and remained popular in the British Isles alone, and have no counterpart in other languages, although they too have often been recorded in more than one variation.

Due to the length of ballad texts, I will examine only two examples in order to expose the principal characteristics of the tradition and their occurrence across languages and styles.²⁸ The two texts that follow belong to a sub-tradition that revolves

²⁸ The reader is invited to consult the extensive bibliography on this subject in order to verify that the features I describe are common to most ballads.

around one of the most famous and well-studied international ballads in the Child canon: the tale of "Lady Isabel and the Elf-Knight", or Child #4. Holger Nygard (1958) tells us that "Lady Isabel" has been in circulation in Europe since 1550. The ballad was extremely popular in all of Europe, and its multilingual sources are disparate; the author mentions a Danish manuscript of 1550, three South German manuscripts of 1555, 1560 and 1570 respectively, a Spanish manuscript of 1550, and yet another manuscript from Iceland written in 1665. The author places the origin of the ballad in Northwest Europe, in its Dutch-Flemish and German versions. A substantial number of variants in virtually all European languages have survived and are available today. The variants are sometimes widely dissimilar from one another, and only a painstaking philological effort could determine that they do indeed belong to the same sub-tradition. In French, the ballad is well known all the way to Québec, under the title "*Renaud le Tueur De Femmes*".

Here, I will quote two versions of the ballad. The first is Child 4E, or the fifth variant reported by Child (out of six versions, 4A through to 4F). This particular variant is titled "The Outlandish Knight" and was originally reported by J.H. Dixon in his *Ancient Poems, Ballads and Songs of the Peasantry of England*, published in 1846. The second version is in Piedmontese, an Italian dialect from the Piedmont region in the northwest of the country. Piedmontese is an ancient dialect that borrows heavily from archaic French. Costantino Nigra originally reported this ballad in his volume *Canti Popolari del Piemonte* (Popular Songs from Piedmont), published in 1888. Alessandra Bonamore-Graves (1986) copied the text in her volume *Italo-Hispanic Ballad Relationships*, and it is from this work that I am quoting it. Bonamore-Graves examines the ballad in relation to

the Spanish side of the tradition, exemplified in her volume by one version of the ballad, "Romance de Rico Franco".

I have chosen these two texts because I believe that they provide two very clear examples of some of the features common throughout the tradition of balladry. As far as the English text is concerned, I could have chosen any of the six variants Child reports, of which the first is in couplets and the other five in quatrains. The quatrain texts are all very similar, and equally apt at exemplifying the art of ballad composition. My choice was dictated partially by personal preference and partially by the fact that the language employed in this version is not so old-fashioned as to require a translation into current English. With respect to a Southern version of the ballad, I have decided not to take advantage of the Spanish "*Rico Franco*" because it presents a variation on the theme that seems quite removed from the plot of the English text. I would rather not puzzle the reader with two ballads that are seemingly (although actually not) unrelated. The story in the Piedmontese version is also quite different from the English one, but the main theme of the tradition is still evident.

The gist of the story is the following: a "false" knight – charming but evil – manages to convince a noble maid to follow him to a faraway land. Upon arriving at the destination, or while still on the road, he makes clear to her that he is about to kill her. The beautiful maid manages to outwit him and kill him instead.²⁹ The sub-tradition is generally cohesive in one stylistic detail: the predicament in which the maid finds herself, and her success at turning around the situation at the expense of the knight, are highlighted by means of symmetrical dialogues between the two protagonists. When

²⁹ In some versions from the Southern countries, the maid does not kill the knight, but saves her honour by killing herself.

the knight informs the maid of her fate his tone is cruel, but when the maid triumphs, it is her turn to use almost exactly the same verses to be cruel and derisive in return.

The Outlandish Knight

- An outlandish knight came from the north lands,
 And he came a-wooing to me;
 He told me he'd take me unto the north lands,
 And there he would marry me.
- 2 'Come, fetch me some of your father's gold,
 And some of your mother's fee,
 And two of the best nags out of the stable,
 Where they stand thirty and three.'
- 3 She fetched him some of her father's gold,And some of her mother's fee,And two of the best nags out of the stable,Where they stood thirty and three.
- 4 She mounted her on her milk-white steed,
 He on the dapple grey;
 They rode till they came unto the sea-side,
 Three hours before it was day.

- 5 'Light off, light off thy milk-white steed,
 And deliver it unto me;
 Six pretty maids have I drowned here,
 And thou the seventh shall be.'
- 6 'Pull off, pull off thy silken gown,
 And deliver it unto me;
 Methinks it looks too rich and too gay
 To rot in the salt sea.'
- 7 'Pull off, pull off thy silken stays,
 And deliver them unto me;
 Methinks they are too fine and gay
 To rot in the salt sea.'
- 8 'Pull off, pull off thy Holland smock,
 And deliver it unto me;
 Methinks it looks too rich and gay
 To rot in the salt sea.'
- 9 'If I must pull off my Holland smock,Pray turn thy back unto me;For it is not fitting that such a ruffian

A naked woman should see.'

- 10 He turned his back towards herAnd viewed the leaves so green;She catched him round the middle so small,And tumbled him into the stream.
- 11 He dropped high and he dropped low,Until he came to the side;'Catch hold of my hand, my pretty maiden,And I will make you my bride.'
- 12 'Lie there, lie there, you false-hearted man,Lie there instead of me;Six pretty maids have you drowned here,And the seventh has drowned thee.'
- 13 She mounted on her milk-white steed,And led the dapple grey;She rode till she came to her own father's hall,Three hours before it was day.
- 14 The parrot being in the window so high,Hearing the lady, did say,

'I'm afraid that some ruffian has led you astray, That you have tarried so long away.'

- 15 'Don't prittle nor prattle, my pretty parrot,Nor tell no tales of me;Thy cage shall be made of the glittering gold,Although it is made of a tree.'
- 16 The king being in the chamber so high,And hearing the parrot, did say,'What ails you, what ails you, my pretty parrot,That you prattle so long before day?'
- 17 'It's no laughing matter,' the parrot did say,'That so loudly I call unto thee,For the cats have gone into the window so high,And I'm afraid they will have me.'
- 18 'Well turned, well turned, my pretty parrot,Well turned, well turned for me;Thy cage shall be made of the glittering gold,And the door of the best ivory.'

The features that we can observe in this ballad are techniques common to a vast majority of the texts in the tradition.

1. **Rhyme and Metre**: The rhymes are simple and predictable, as they use common words and easily repeatable sounds. The rhyme scheme is a-b-c-b. The metre is arranged in verses of four stresses alternating with shorter verses of three stresses. This is a model for the quintessential ballad quatrain. An equally common model employs the same rhyme scheme and the same metre, with the difference that all verses have four stresses.

2. "Leaping and Lingering": The reader might be surprised by the way in which the tale is action-packed at certain points, while it rather dillydallies in others. For example, there is not a lot of action at the beginning of the ballad. In stanza 3 the maid collects some valuables, and in stanza 4 the lovers ride to the water. Then nothing more than dialogue happens until stanza 10, when she throws him into the sea. In stanza 13 she rides back home, and the rest of the ballad up to stanza 18 is taken up by dialogue. For a story of seduction, murder and deception, the actions that are actually narrated are few and distributed in clumps. This is because typically, stanzas are constrained by their individual theme. They contain either an action scene or a piece of dialogue; a mix is rarely found. For this reason, the actions are concentrated in a few stanzas along the composition, while many more stanzas are taken up by dialogue.

3. **Dialogue**: Quoting a dialogue between characters is a common expedient to make the narration more vivid and present for the audience. The performer identifies

with each of the characters in turn and gives the dialogue a more personal quality. Dialogues are also an economical technique because they represent actions or circumstances in an indirect manner, at the same time as they portray a character's opinion or feelings. Dialogues usually unfold in a very precise succession of repetitions as sentences or entire passages are reiterated with minimal variation. This is clearly the case with stanzas 6, 7 and 8 above, and to a lesser extent with the other instances of dialogue in the ballad. As noted above, this ballad is particularly recognisable for the retort that the maid delivers to the knight by employing almost exactly his same words. When the knight reveals his intentions, he imperiously orders the maid to "Light off, light off thy milk-white steed,/ And deliver it unto me;/ Six pretty maids have I drowned here,/ And thou the seventh shall be." After she has managed to turn the tables on him, she mocks him with a reply that is parallel to his initial statement even in its use of repetitions. In particular, the third line is identical to the one he employed. However, the following and closing lines contain a surprising shift from a passive to an active stance, and effectively delivers the realisation that the situation has been reversed: "Lie there, lie there, you false-hearted man,/ Lie there instead of me;/ Six pretty maids have you drowned here,/ And the seventh has drowned thee." A sneer of revenge is clearly discernible in her words.

4. **Repetition**, especially **Incremental**: All stanzas in the ballad above are replete with repetitions. As I have mentioned above, stanzas 6, 7 and 8 are nearly identical recurrences of the knight's order to disrobe, the only difference being that the item of clothing that must be taken off changes from one stanza to the next; presumably, the previous garment has already been removed and it is time to take off

the next. This is a case of incremental repetition, whereby the succession of items in a list creates a sense of drama and an escalation in pathos. In this case, the garments are getting closer and closer to the maid's bare skin; clearly her time is running out and she will soon be killed. This device is extremely common in ballads and it appears in dialogue stanzas as well as in narrative ones. Most often, a ballad will include only one succession of incremental repetitions, usually placed at the most dramatic point in the story.

Many other types of repetitions are also widely used in ballads, from the nonincremental repetition of entire stanzas to the repetition of verses, half-verses, words and single sounds. Of course, single sound repetitions are far less noticeable than repetitions of whole stanzas, but they still contribute to the overall redundancy of the text. Several examples of repetitions are observable in "The Outlandish Knight". For example, stanzas 2 and 3 are almost identical, with the difference that 2 is in dialogic form and 3 is a narration. Stanzas 4 and 13, which describe the travelling to and from the sea, are also nearly identical. Stanzas 15 and 18 contain one matching verse: "Thy cage shall be made of the glittering gold". Shorter repetitions of single words or phrases are found in the verses "He dropped high and he dropped low", "What ails you, what ails you, my pretty parrot" and "Well turned, well turned, my pretty parrot,/ Well turned, well turned for me". Finally, we can find a good example of sound repetitions in the verse "Don't prittle nor prattle, my pretty parrot" where the alliteration in p and the consonance between "prittle" and "prattle" are quite perceptible.

5. **Commonplaces**: The ballad lexicon includes an inventory of formulas that are consistently employed to describe an entity or express an idea. These formulas are

comparable to those described by Parry and Lord in their studies of Homeric and Slavic texts. However, in ballad studies they are called commonplaces. The phrases may be as short as a noun and its qualitative adjective, or they may take up a verse or more. For example, horses are usually "milk-white" (for the first horse, typically ridden by a woman) and "dapple grey" (for the second, complementary horse, typically ridden by a man). "Middles" (waists) are "small", maidens' gowns are "silken" and less-than-noble knights are "false" or "false-hearted". The mention of "gold" begets the mention of "fee" close by. We can find all of these examples in the ballad above, and we could observe many more examples if we were to take into consideration a larger number of texts. Within any anthology devoted to the ballad tradition, it is easy to see that the same commonplaces are used repeatedly across different ballads.

6. **Imitative response**: This expedient denotes a request in dialogic form to do something, followed by an almost identical stanza that describes the actions that are performed as a result. The technique is apparent in stanzas 2 and 3.

7. **Narrative Symmetry**: The tale of most ballads develops according to a narrative symmetry, meaning there is a strong similarity between the first and second halves of the story. Very often, the symmetry is chiasmatic (i.e., of the a-b-c-c-b-a kind), whereby the story is repeated backward in the second half of the ballad. This allows for a great deal of re-use of verses and commonplaces from the first half. Our ballad above does not provide an ideal example of narrative symmetry, yet its tale does follow a chiasmatic structure. The narration proceeds along this organization: dialogue – travel – his speech – trick – her speech – travel – dialogue. In a more accomplished example of

narrative symmetry, we would be able to find more repetitions between the first and second half of the story. In "The Outlandish Knight", the repetitions are few but noticeable, including the two stanzas that describe the travels to and from the sea, and the two stanzas that contain the knight's dooming speech and the maid's revengeful retort.

Let us now turn to the Piedmontese ballad. In Italian, the ballad is known either as "A Heroine", like the version I quote here, or as "The English Maiden". When the first title is employed, the heroine in question is generally identified as a woman of noble local origins. The second title denotes a version in which the maid is identified as a noblewoman from England, usually the daughter of either a knight or the king himself. I reproduce the Piedmontese text below, with an English translation under each verse. The caesura appears as a space between the two halves of the verse.

Un'eroina

A Heroine

El fiöl dij signuri cuntis'a l'è chiel n'in va ciamè,The son of the Countgoes about to ask,Va ciamè d'üna Munfréinala fia d'ün cavajè.Goes to ask a Monferrina³⁰the daughter of a knight.

2

³⁰ "Monferrina" means a woman from Monferrato, an area of Piedmont well known for its wine and its rich countryside.

S'a l'è 'l saba la va 'mpromët-la, On Saturday he gets her promised,

- 4 L'à meinà sinquanta mia
 He took her away for fifty miles
 Prima vota ch'a j'à parlà-je,
 The first time he spoke to her,
- 6 Guardè là bela Munfréina,
 Look over there, pretty Monferrina,
 Mi sinquanta e due Munfréine
 Fifty and two Monferrine
- 8 Le sinquanta e due Munfréine
 To all fifty and two of them,
 N'autertant farai, Munfréina,
 Just as much I will do to you,
- 10 O scutè, lo signur cunte,
 - Oh listen, Count,
 - O dizì, bela Munfréina,
 - Oh tell me, pretty Monferrina,
- 12 Vöi tajè na frascolina
 I want to cut a branch
 Quand la bela l'à 'biü la speja,
 When the maid got the sword,
- 14 O va là, lo signur cunte,Off you go, Count,

dì dumëgna la va spuzè. on Sunday he marries her. sensa mai parlè-je ansem. without talking to her once. s'a j'à ben cozì parlà: he told her right so: cul castel tan ben mürà. that castle so well fortified. mi la drin j'ò già meinà: I already took in there: mi la testa e j'ò cupà. I cut off their heads. quand che vui n'a sari là. when you will arrive there. prëstè-me la vostra spà. lend me your sword. coza mai na völi fà? what will you do with it? për fe umbra al me caval. to shade my horse with it. ant ël cör a i l'à plantà. she plunged it in his heart. o va là 'nt i cui fossà! off you go into that ditch!

	L'à virà al caval la brila,	andarè l'è riturnà.
16	She turned the horse's bridle	and back she went.
	Ël primier ch'a na riscuntra,	so fradel n'à riscuntrà.
	The first one that she met,	she met her brother.
	- O dì 'n po', bela Munfréina,	l'è dassè che 't trove sì!
18	- Oh tell me, pretty Monferrina,	It is strange that you are here.
	- J'ò trovà i sassin di strada,	l'àn massà-me 'I me marì.
	- I came across some bandits,	they killed my husband.
	- O dì 'n po', bela Munfréina,	t' l'avrei nen massà-lo tì?
20	- Oh tell me, pretty Monferrina,	could it be that you killed him yourself?
	- O sì, sì, me fradelino,	la `vrità ch'a fa bel dì;
	- Oh yes, my little brother,	truth is the most honourable account.
	A sun pa i sassin di strada	l'àn massà-me me marì.
22	It was not the bandits	who killed my husband.
	- O dì 'n po', bela Munfréina,	a cà tua venta turnè.
	- Oh tell me, pretty Monferrina,	you've got to go back home.
	- O no, no, me fradelino,	a cà mia vöi pa pi andè.
24	- Oh no, my little brother,	I will not return home.
	Mi na vöi andè a Ruma,	'ndè dal papa a cunfessè.
	I want to go to Rome,	to go to the Pope and confess.

.

Below is a short list of the devices employed in this ballad. A perfect coincidence between this text and the previous one should not be expected; although they belong to the same tradition, they are, after all, records of oral performances that were shaped by

-

two completely different cultures. For one thing, that the Italian version is in dialect while the English version is in standard English suggests that the origins of the Italian ballad might be humbler, from the peasantry, while the English ballad is more likely to have urban origins. However, despite the differences, some important similarities do exist.

1. **Rhyme and Metre**: The rhymes are of minimal length, since all rhyming words end in an accented vowel. Words with an ending accent are very few in standard Italian, but they are extremely common in many dialects. They have proven very convenient in this case, where the author/performer only had to match a single sound, rather than an entire syllable. However, the overall rhyming scheme is quite irregular. The ballad begins with three verses rhyming in "è", followed by an unrhymed one; then we have seven more verses rhyming in "à" and again an unrhymed one; next are four more verses again rhyming in "à", five verses with an "ì" rhyme and three closing verses with the initial "è" rhyme.

The metre, however, is a regular rhythm of four stresses per half-verse. We can actually consider each pair of verses as a quatrain: if the ballad had been written with one half-verse per line, instead of two half-verses separated by a caesura, the structure would have mirrored that of English ballads with quatrains of four-stress verses. Note that if we consider each pair of verses as a quatrain, then the irregularity of the overall rhyme scheme becomes insignificant, because English ballads do not employ a regular rhyme scheme that runs across the full length of one ballad. The overall rhyme scheme is noticeable only if the ballad is written in the way in which I have reported it.

2. "**Leaping and Lingering**": This technique is not nearly as visible here as it was in "The Outlandish Knight", possibly because the ballad is substantially shorter (12 quatrains instead of 18) and contains more dialogue than its English counterpart (8 quatrains of 12, instead of 10 and ½ quatrains of 18).

3. **Dialogue**: As just mentioned, three quarters of the ballad consists of dialogue. The instances of dialogue are shorter and alternate more often than in the English ballad, and here they serve only to tell the disposition of a character, rather than conveying the actions being performed. However, the theme of the retort is still evident. Although the heroine does not employ exactly the same words as the knight, she follows the same theme of his speech. He has shown her his castle "so well fortified", and she sends him into the *fossa'* (short for *fossato*) – this word means both "ditch" and "moat", and is therefore a pun on the theme of fortification. The mocking and revengeful overtones of her reply still come across intact.

4. **Repetition**: There are no instances of incremental repetition in the Italian version, but I have found examples of this device in other Italian ballads (e.g., see Bonamore-Graves, 1986, p. 82). However, this ballad does contain quite a few instances of repetition proper. For example, a few half-verses are repeated: "*Mi sinquanta e due Munfréine*" and "*Le sinquanta e due Munfréine*"; "*O dì 'n po', bela Munfréina*" in stanzas 19 and 22; "*I'àn massà-me 'I me marì*" and "*I'àn massà-me me marì*"; "*O sì, sì, me fradelino*" and "*O no, no, me fradelino*". Also, "*va ciamè*" is repeated and forms an anadiplosis between verses 1 and 2, and "*i sassin di strada*" is repeated in verses 8 and 21.

5. **Commonplaces**: The most evident commonplace in this ballad is of course "*bela Munfréina*" (pretty Monferrina), which is used every time the heroine is mentioned. Another commonplace is that regarding the Count ("*signuri cunti"*, "*signur cunte"*).

6. **Imitative Response**: There are no examples of imitative response in this ballad. However, the expedient is used in other ballads of the same tradition (for example, it appears three times in the ballad "La bella Leandra" in Bonamore-Graves, 1986, p. 79).

7. **Narrative Symmetry**: The extent of the narrative symmetry is less pronounced in the Italian version than in the English, but is still partially present due to the similarity in the story. The structure is: marriage proposal – marriage – travel – his speech – trick – her speech – travel – encounter with brother – his and her speech. The central part of the story is still chiasmatic.³¹

I have attempted to highlight several important features in the analyses of an English and an Italian ballad. I have discussed rhymes and metres, which are present in both ballads and are quite regular. Both ballads are replete with repetitions of various types and with commonplaces, although the Italian version is less redundant than the English. Both ballads make extensive use of dialogue, and we have seen that the main

³¹ Note that this is not always the case: for example, in the Spanish version "Romance de Rico Franco", the twist in the plot that causes the heroine to kill herself does not lead to narrative symmetry.

stylistic feature of the ballad, the heroine's retort, is a distinctive highlight in both texts. Finally, I have shown that the story is build in a chiasmatic manner, which makes it easier to use the same elements of the tale in both halves of the ballad.

4. Summary: From Oral Performance Back to Speech

In this chapter, I have illustrated several findings regarding the cognitive processes that allow an oral singer to perform a piece with spontaneity and originality, and yet remain faithful to his or her tradition. The chapter has provided a very brief overview of David Rubin's findings on the nature of meaning, imagery and sound constraints at play in the working memory of oral performers; it also has highlighted many similarities with the cognitive processes of speech production that I identified in Chapter One. Finally, the chapter has provided an analysis of many oral texts in an effort to illustrate the pervasiveness of the cognitive constraints that Rubin identified.

As we have seen, the thematic constraint of a song is akin to the semantic constraint of speech: they both prompt the production of utterances that are linked semantically and via cause-and-effect relationships. The thematic/semantic constraint ensures that a story, as well as each individual utterance, remains generally coherent, consistent and comprehensible.

The imagery constraint described by Rubin corresponds to Tannen's account of the use of details in storytelling. Both types of imagery constraints are clearly a special sort of semantic cue, although I believe that Rubin's suggestion regarding their connection to visual memory is only partially tested in the realm of speech production.³² In oral

³² See Hampe for an overview of various types of research in this area.

performances and common speech, imagery allows for expansions, progressions or even digressions within the overall theme or semantic domain being explored. It also produces a sense of intimacy and affective involvement among participants.

Formulas are frequent in oral traditions and in common speech. They can be completely fixed, like idioms, or more or less flexible, like collocations and preferential lexical co-occurrences. Formulas are tightly linked to a specific rhetorical domain; in the case of oral performances these domains can be as small as a particular song or as general as an entire tradition, while in the case of common speech, the rhetorical domain can be restricted to a particular social situation or shared by an entire speaking population. In both cases, formulas facilitate the production of utterances because a cue to the first element or sound in the formula will automatically lead to the production of the whole formula. Moreover, formulas can be used as metric and prosodic fillers when there is a need to complete a verse or an intonational phrase without adding much to its content. (Consider, for example, the similar function provided by expressions such as "brilliant Odysseus" and "by and large".)

With regard to sound repetitions, it is evident that most of those found in oral poetry are virtually indistinguishable from the discourse echoes I described in the previous chapter. Rhymes, alliterations and assonances facilitate recall because two or more words contain similar sounds. Accordingly, discourse echoes facilitate the selection of certain words in everyday utterances because these words contain the same sounds (as in the example "cause it comes from cold water" in Chapter One). In fact, terminology such as "rhyme", "alliteration" or "assonance" can be used just as accurately for poetry as for common speech.

Rhythm is a structural element of both oral performance and speech. The rhythm of a song or poem is often regular and predictable so that the audience can follow the performance more easily and feel part of its unfolding. As I have explained above, the rhythm of conversations also displays a regularity that ensures the efforts of all interlocutors are coordinated. Moreover, the capacity to predict and keep with the rhythm also influences the interlocutors' affective involvement, just as it influences the affective involvement of the audience of an oral performance. In production, rhythm facilitates the selection of words or phrases with a specific syllabic length and stress pattern. This mechanism is apparent both in everyday utterances ("We're / áll in/túitively fa/míliar with the i/déa of a de/rívative") and in oral performances ("cátch a tíger bý the tóe / íf he hóllers lét him gó", and "An óutlandish kníght came fróm the north lánds / and hé came a-wóoing to mé").

Finally, the phrasal organisation provided by melodic constraints is extremely similar to the organisation imposed by prosody. The melodic constraints of a verse or a refrain are imposed by the regularity of a specific song and by the expectations of an entire tradition. Similarly, the regularity of prosodic patterns is imposed by the speaking habits of a certain speaker and of an entire population of speakers. The length of a melodic phrase (or prosodic pattern) provides an immediate measure of the time interval that a series of sounds must fill, and therefore of the amount of singing/speaking that must occur contiguously. Therefore, the selection of a melodic phrase is intimately connected to the number of beats/accents that must be produced, and consequently also to the number of words that an utterance can accommodate. Moreover, the development of a typical melodic or prosodic phrase places emphasis on a few higher pitch points, such as the initial tones of a declarative sentence, the end tones

of a question, and any of the higher notes in a chant or song. As we have seen in Chapter One, this point of emphasis must correspond with an already-occurring word stress and therefore limits the words that can be fitted at that particular point in the phrase. The same mechanism of word selection can be observed in songs and chants.

The table below provides a snapshot of the parallels that exist between the constraints system of oral traditions and the constraints system of common speech:

Const	raint	ts in Oral Performance	Speech Constraints		
Theme	9		Semantics		
Image	ry		Imagery in discourse		
Sound					
└ →	Rhythm		Speech rhythm		
	Melody		Prosody		
	Sound Repetitions				
I	L.,	Above the word: Formulas	Repetitions, formulas, collocations		
		Words	Repetitions, lexical activation		
		Below the word: Alliteration,			
		Rhyme, Assonance	Echoes		

Table 2.1 A comparison between the constraints of oral performance and the constraints of speech.

Each of the constraints of common speech has a counterpart in the performance and composition/recall of oral texts. The differences between the constraints of common speech and those of oral performances are solely a matter of degree. Oral traditions
must remain faithful to strict limitations that are apparent to the audience: a singer's verses are constrained by the story, diction, rhythm and melody of a particular song and, more generally, of an entire tradition. The limitations imposed on common speech are less rigid since, for example, a conversation can expand semantically in many directions and it can employ a less prescribed succession of prosodic phrases. However, as discussed in Chapter One, common speech is also constrained. Both the production and the understanding of common speech depend on employing speech constraints in a manner that remains consistent for speakers and listeners. The utterances of a speaker are just as constrained as those produced by a singer, because a speaker is also limited by the necessity to observe a specific vocabulary and established prosodic and rhythmic conventions. These limitations are imposed on the speaker both by the speaking population at large and by the specific situation at hand, in the same way that a certain tradition, song and audience impose their limitations on a singer's performance.

Rubin's work offers a clear explanation of the processes that can be observed when an oral poet performs his or her work. It also explains those recurring features of oral texts that Parry and Lord thoroughly documented in their research, and that others after them have observed in a variety of oral texts from all regions and epochs. However, one question remains in my mind. While Rubin has described the constraints of oral performance as being exclusively serial, I described the constraints of speech production as being additive. These definitions mean that Rubin's constraints summate only in a forward fashion, from beginning to end. On the other hand, the speech constraints I describe can also function backwards, by facilitating the selections of chunks both at the beginning and at the end of an utterance. I question whether this discrepancy is factual, or whether a chunking process may be at work in singing as well

as in speaking. I cannot answer this question at the moment, but I believe it worthy of further exploration.

Above, I have illustrated an analogy between common speech and the performances of an oral tradition. This analogy is meaningful for two reasons. First, it provides further evidence in favour of those processes of constrained selection of mnemonic material that I illustrated in Chapter One. Second, it makes a further case for Rubin's call for enquiry into what he terms surface schemas, or sound associations, and into what I term sound constraints. Rubin documented that the recall of oral information is aided by a similarity of sound or rhythm among utterances. Since these similarities exist not only in oral texts but also in common speech, it must be concluded that the recall and selection of oral data in working memory during speech production is influenced not only by semantic activation and syntactic reasoning, but also by the actual sound quality of the utterance to be produced. To my knowledge, very little research has been or is being conducted upon this point. I intend to continue the investigation of surface schemata and speech constraints in my future research endeavours. For now, I must leave this matter open until further evidence becomes available.

In this chapter, I have discussed Rubin's studies on oral composition as supporting evidence for the speech constraints argument developed in Chapter One. In the next chapter, I will provide a different type of evidence by examining several studies that investigate the dynamics of cultural information transmission. These studies demonstrate that the dynamics of all types of cultural communication – from gossips to

technological innovations – are extremely predictable because they follow a consistent pattern over time, regardless of the regional and cultural differences in the populations under scrutiny. As I will argue, this fact points to a certain recursiveness in people's psychology and communicative practices, and this recursiveness is consistent with the theory of speech production that I have made in Chapter One.

Chapter Three

Speech Constraints and the Transmission of Culture

In the previous chapters, I have explained that the performers of an oral tradition and the speakers of a language are subject to similar cognitive constraints, and that these constraints determine the nature of the utterances that the singers and speakers produce. My considerations have effectively equated the speech of a population with a broadly defined oral tradition, and each speaker with an oral poet. The key to perpetuating the tradition is in the use that each speaker makes of his or her memory in order to produce speech. Just as a tradition of ballad-singers establishes its constraints in the form of tunes, stories, formulas and so on, a tradition of speakers also establishes its constraints in terms of prosody, grammar, idiomatic expressions and so forth. The result is that, on the one hand, the members of each of these traditions have a great deal of freedom in producing utterances. However, on the other hand, some of these utterances will be well constructed while others will be more or less defective. (This value judgement is, of course, relative to the expectations of the other members in the tradition.) Since defective utterances defy commonly employed constraints, they are also

less memorable, less likely to stick in others' (or even their own utterer's) memory. Just like an oral tradition, a population's speech maintains itself relatively unchanged over time because the underlying constraints of speech production are unlikely to accommodate much simultaneous change. In other words, the cognitive mechanisms of speaking are inherently conservative. This conservatism, in turn, is part of broader communication processes that are also conservative and whose outcomes are predictable.

The transmission of cultural information is the overall topic of this last chapter. In the first section, I present ample evidence that cultural communication takes place in regular patterns. I examine many different types of research in cultural transmission; at the end of the section I illustrate the psychological mechanisms underlying social communication, and show that these mechanisms are fully compatible with my speech constraints argument in Chapter One. In the second section, I expand on some theoretical implications regarding the place of speech constraints in cultural communication; in particular, I examine the features that may make utterances more or less communicable in the context of a population of speakers.

1. The Dynamics of Cultural Information Transmission

a. CHARTING THE PATTERNS

The first and perhaps most famous attempt at recording the transmission of cultural information gave rise to a branch of anthropology called Diffusionism. For Diffusionists, the importance of cultural transmission surpassed that of any other type of human interaction. Many studies in Diffusionism were devoted to charting the

distribution of concepts, such as the adoption of a particular tool or practice, across the populations residing in a circumscribed geographical area. For example, Diffusionism found a commercial application in the studies of E.M. Rogers, who investigated the rates and patterns of adoption of new agricultural products within the farming communities of several areas of the United States.

More recently, in a popular volume inspired by the ideas of Diffusionism, Malcolm Gladwell resorted to psychological and sociological literature to explain the mechanisms of a phenomenon that he calls "tipping". Gladwell primarily examines examples from the world of marketing and advertisement. Many commercial successes, he argues, start as trends among a very limited group of people. The news about a certain product, fashion or practice may then spread from the initial group to a larger set of individuals, and may eventually become widespread among the general population. The crucial step that brings a notion from a small, initial group of adopters to a larger audience is the "tipping point": after an idea has "tipped", it is virtually impossible to stop its further spread. The distribution of the idea will continue in an automatic manner.

Gladwell is successful at demonstrating the automatic nature of this occurrence, but he is unable to fully explain its mechanisms. While purpose and personal benefit can be key elements in the conscious decision to adopt a certain notion, it is important to underline that tipping is clearly not a conscious phenomenon, at least not for the majority of adopters. In fact, both Gladwell and Rogers plainly describe the spread of a notion, past its initial or tipping stage, as unavoidable. Inevitability rules out individual choice to a very large degree, because it applies to notions that can be immediately seen as beneficial to the individual as well as to notions that do not appear beneficial, or may even appear or turn out to be detrimental.

If the diffusion of a notion takes place mostly in an unconscious manner, it must follow that mechanisms other than conscious planning or calculation are at work. Clearly, these mechanisms must play a role in the course of the social interaction that takes place between the members of a community. In their research, Rogers and Gladwell found that the reputation and respect a person enjoys within his or her social circle is a determining factor in the success that his/her beliefs will encounter among other people. Another factor that Gladwell describes is emotional contagion, or the instinctual tendency to imitate others' emotions in order to express support and care. Emotional contagion is spread mostly by mimicry of the verbal, facial and bodily expression of emotions. Some people seem to be particularly good at getting others to move their bodies and faces in the same way as they do, and with this expedient, they succeed in transferring part of their emotions directly to their audience. If a person of this type has invested emotionally in his/her adoption of a notion, the person who mimics them will end up feeling at least part of the same attachment and investment in that same notion.

It is tempting to define the overall mechanism of diffusion as a set of complex acts of imitation. We may admire somebody and therefore attempt to be a little more like them by adopting their views or manners; we may want to feel closer to the people around us by mimicking their gestures and sharing their emotions; we may try to gain our peers' esteem by assuming the attitude that we think is ideal in a certain situation. Humans are undoubtedly very gregarious animals, and the tendency to repeat others' words and actions is certainly one of our predispositions. However, this broadly defined type of imitation does not truly explain the complexities of human interaction. For this

reason, many social psychologists have attempted to further clarify the dynamics of communication by viewing the question from an entirely different angle.

In particular, some social psychologists have focussed on the many psychological motivations that fuel the circulation of information in a community. In 1947, Allport and Postman examined the social phenomenon of rumour, a type of transmitted knowledge the dynamics of which are all the more baffling given the uncertain, and often downright incorrect, nature of the information that gets shared. Yet despite the inaccuracies and refutations, rumour is a virtually universal, ever-present phenomenon. Allport and Postman posited that rumour takes place principally in a situation that presents some ambiguity, where the circumstances have created some anxiety for the people involved. In these settings, the information contained in rumours is an attempt to fulfil the need for an explanation of current events. Rumour serves as a rationalising agent: it explains, justifies, and provides meaning for the emotions at work in people's minds. Therefore, the importance of rumours lies not in their informative value, but rather in giving speakers an opportunity to express their emotions, goals or judgements to their listeners.

As for the shape that information assumes when transmitted, Allport and Postman found that, after numerous retellings, the tales transmitted by rumours were subject to the three principles of levelling, sharpening and assimilation. With levelling, the tale loses most of its details and maintains only a very small set of its original features. With sharpening, the few details that are retained become very prominent, often in a manner that is disproportionate to the importance of the detail in the original story. With assimilation, elements that were not part of the initial telling are added to the story in order to build a more memorable and more logical structure. Other expedients often

employed in rumours are: the present tense, to render events in a more immediate manner; the description of movement rather than scenery or inanimate objects, to make the narration more memorable; and the use of verbal labels, to create a quick and accurate referential background. Allport and Postman also found that large objects tend to be remembered more easily than small ones, and that elements that are important to the story often grow in number with successive retellings.

Rosnow and Fine have also investigated the dynamics of rumour. They posit that it is necessary to make a distinction between rumour and gossip, based on the purpose served by the retelling of a story. On the one hand, following Allport and Postman, they argue that rumour fills an unconscious desire for meaning, clarification and closure, but they emphasise to a greater extent the role that collective anxiety plays in generating and facilitating the diffusion of rumours. On the other hand, they instead find that gossip is motivated mostly by the unconscious need to affirm one's ego and status within a confined social group. While gossip always pertain to individuals in the local community and may or may not be unsubstantiated, rumour is always unsubstantiated and often encompasses events and social circles of a larger magnitude. Perhaps the most important distinction that the authors illustrate between rumour and gossip concerns the instruments of their transmission. Since gossip is a local event, it spreads primarily by word of mouth. In contrast, rumour takes place at a larger-than-local scale. For this purpose, it often exploits not just word of mouth but also mass media such as newspapers, films, TV and radio news reports, Internet sites, email and so forth.

For Elaine Showalter, the influence that media-spread rumours can have on people's psychology reaches directly into the depths of our physical and mental health. Showalter studies the contemporary epidemics of phenomena that she identifies, at

bottom, as mass psychoses. She argues that although the use of the word "hysteria" is no longer current, the psychological illness that it identifies is more common than ever. The stress factors that cause hysteria are omnipresent in people's lives, and people find legitimation for the expression of their troublesome feelings by espousing syndromes that they hear about in media reports. "Contemporary hysterical patients blame external sources – a virus, sexual molestation, chemical warfare, satanic conspiracy, alien infiltration – for psychic problems" (Showalter, 1997, p. 4), and, "as the syndromes evolve, they grow from microtales of individual affliction to panics fuelled by rumors about medical, familial, community, or governmental conspiracy" (p. 5), to the point that the panic reaches epidemic proportions. The epidemics "spread by stories circulated through self-help books, articles in newspapers and magazines, TV series and talk shows, films, the internet, and even literary criticism" (Showalter, 1997, p. 5). As a result, "patients learn about diseases from the media, unconsciously develop the symptoms, and then attract media attention in an endless cycle" (p. 6).

The diagnosis of a new syndrome involves advertising the prototypical patient; the prototypes are often vague and broad, and correspond to the description of many different problems. Then, patients are asked to come forward with their symptoms, effectively becoming living proofs of the syndrome. Often the behaviours of clinical centres, professional journals and practitioners substantiate and support a culture's investment in the syndrome as a real disease. Some people go so far as to rewrite their personal narratives in order to fit them to the description of the disease: "Patients are people with a bewildering set of troubling symptoms and a wide range of explanations for them. Once they see their problems reflected in a prototype, [they] come to believe that the laws of a disorder describe their lives" (Showalter, 1997, p. 19).

Corroborating Allport and Postman's original finding, Showalter observes that "syndrome rumours" often develop in communities that are already in a state of tension. Episodes are most severe when thoughts of an obvious enemy loom on the horizon; Showalter takes Gulf War Syndrome as an example. If authorities react to a stressful situation in a worrisome manner, instead of a calming one, fears will automatically escalate and cause the collective anxiety to grow further.

At bottom, Showalter fully agrees with Rosnow and Fine, and Allport and Postman before them, that an underlying feeling of anxiety provides extremely fertile grounds for rumour to spread. However, she also carries the argument one major step further, by showing that the rumour itself produces effects as tangible as physical symptoms, which in turn cause more anxiety, more ostensible substantiation of the rumour, and ultimately more grounds for the same rumour to spread further. In Showalter's account of the development of hysterical syndromes, the successful transmission of the syndrome rumour is a factor every bit as important as the initial anxiety that allowed the rumour to arise in the first place. Consequently, while it is certainly important to study the sociopsychological environment in which rumours (and, in a more general sense, information at large) are passed around, it is equally important to investigate the mechanics of information transmission, because such transmission itself may impact the psychological environment to a significant extent.

A number of scholars writing in the last 25 years have chosen to focus on studying the information that gets transmitted, in an attempt to shed some light on the processes of transmission at large. The advantage to this approach is that transmitted information is far easier to document and observe than the individual psychological processes that produce social communication. Cavalli-Sforza and Feldman developed a mathematical

approach to the study of information transmission. In the course of their study, they point out several important observations. For example, they note that while some information can greatly influence the genetic fitness of a population (e.g., notions of safety, hygiene and so on), other types of information have virtually no influence on fitness (e.g., yoyos, Coca-cola, chewing gum). Yet, whether or not a piece of information is "evolutionarily useful" seems to have little to do with how successfully it will spread. Moreover, they observe that the diffusion of an innovation produces a regular s-shaped curve on a set of two-dimensional axes, the very same s-shaped curve that Rogers produced in his studies on the diffusion of cultural information.

b. Explaining the Mechanisms

So far, we have learned that the process of cultural information transmission includes predictable patterns, that it may look somewhat like a process of imitation, and that a certain type of information – mostly of the "bad news" kind – is likely to spread when the psychological environment is already in a state of anxiety, consequently increasing the level of general anxiety even further. These observations confirm the remarks I made at the beginning of this chapter with regard to the recursiveness of communication; however, they do not explain why communication is recursive. For that explanation, I will delve a little more into social psychology.

Historically, Frederick Bartlett (1967) was the first psychologist to conduct experiments on the mechanisms of social information transmission. Unlike previous experimenters, who had used mostly nonsensical words in order to exclude the social dimension of mnemonic recall, he purposefully employed socially relevant data as transmission tokens. Bartlett developed many experiments, some of which explored the

outcomes of serial storytelling through chains of listeners/re-tellers, while others tested single subjects' ability to recall the details of visual and verbal data. Through his experiments, Bartlett (1967) documented extensively the three principles of levelling, sharpening and assimilation that Allport and Postman had proposed. Bartlett theorised that information that fits into a social framework is recalled more easily than information that does not (such as nonsensical words), because he saw recall as being achieved through a process of reconstruction:

Remembering is not the re-excitation of innumerable fixed, lifeless and fragmentary traces. It is an imaginative reconstruction, or construction, built out of the relation of our attitude towards a whole active mass of organised past reactions or experience, and to a little outstanding detail which commonly appears in image or in language form. It is thus hardly ever really exact, even in the most rudimentary cases of rote recapitulation, and it is not at all important that it should be so. (p. 213)

The reconstruction that happens in recall is based on those same psychological schemata, or sets of associated notions, that I mentioned briefly in Chapters One and Two. The concept of schema is important in understanding Bartlett's theory of remembering; his significant contribution consisted in positing schemata as sets of associations that are constantly revised and re-constituted, and therefore are not static and unchangeable entities.

He proposed that schemata are initially developed within the first few years of an individual's life. Once the individual begins his/her social life, the reactions dictated by his/her schemata are continually checked and updated in relation to those of others. An

input impulse becomes a cue not just to a schema, but also to a specific part of the schema that is relevant to the needs of the moment. In the case of storytelling, people also employ a psychological schema when recalling information about a story. Levelling, sharpening and assimilation modify the story in order to fit the structure of the schema. Social constructs also intervene to change the story in ways that are beneficial to the person or to the group. In fact, since social interaction is an important factor in the constant revision and reappraisal of a schema, it can be argued that the social landscape influences individual schemata to the point of becoming a defining part of their development:

When a number of people are organised into a social group, whether by appetite, instinct, fashion, interest, sentiment, or ideal, this group speedily develops certain characteristics peculiar to itself, which directly constrain the behaviour of its individual members. I have throughout treated these characteristics as the expression of active tendencies, for we have to consider them, not merely descriptively, as they are expressed in institutions, symbols, catch words, codes, and material culture, but also causally, as actual determining conditions of conduct and experience. (Bartlett, 1967, p. 281)

In other words, group communication in the shape of behaviour, artefacts and language informs the schemata of those who are part of or come into contact with the group. Hence, social information transmission cannot be treated as a trade in tokens of information, but must be seen as a strictly social act in which the traded information is shaped by socially determined schemata and in turn contributes to constantly editing and re-shaping such schemata.

With his theory of schemata and the experimental evidence he collected to support it, Bartlett supplied flesh, bones and method to the hypotheses of a scholar who had lived in the first half of the twentieth century, a sociologist who coined the term "collective memory". Maurice Halbwachs had stated that memory is a social act: people create and remember memories within society. Institutionalised social groups are necessary frameworks for human memory; people's memories are always connected to a group, and without belonging to several groups, an individual would be unable to organise his/her memories in a meaningful manner.

According to Halbwachs, the only time in which our memory is not subject to the structure of society is when we dream, and this is why the structure of dreams is most often fragmentary, only partially logical or completely nonsensical. Temporarily disconnected from society, the consciousness of sleepers has no recourse to the framework of social grouping and therefore is unable to organise its images of people, things and events in a coherent manner:

If purely individual psychology looks for an area where consciousness is isolated and turned upon itself, it is in nocturnal life, and only there, that it will most be found. [...] Almost completely detached from the system of social representations, its images are nothing more than raw materials, capable of entering into all sorts of combinations. [...] The dream is based only upon itself, whereas our recollections depend on those of all our fellows, and on the great frameworks of the memory of society. (Halbwachs, 1980, p. 42)

Memories belong to the individual but at the same time they belong to the family, religion, social class, municipality, nationality, football team, Acme Inc. employees and

any other groups of which the individual is a part. The fact of belonging to such groups is what enables the individual to create and possess consistently organised memories.

Bartlett admired Halbwachs' theory, but criticised his lack of an explanation of the mechanisms by which social groups and memory become one and the same. To be sure, Halbwachs' work includes very little discussion on social communication. He dwells only briefly on the use of language and, just as he uses the negative example of dreams to illustrate his notion of organised collective memory, he employs the negative example of aphasics to illustrate the necessity of language in creating the link between individuals and the groups to which they belong. However, he lacks an explanation of the role of communication in the making of collective memories, which is where Bartlett's work comes in to provide the missing link. Bartlett's research centred on remembering, and yet much of his attention focused not just on what was being remembered by his experimental subjects, but also on how these memories were communicated to other subjects or to Bartlett himself. Reading about his experiments, one cannot fail to notice that Bartlett treats recall and communication as two faces of the same coin. This step is what is missing from Halbwachs' theory, the realisation that communication is the organising agent that shapes, and in this sense creates, both social groups and the memories they produce.

In Bartlett's experiments, the primary mode of communication was verbal exchange: the experimental subjects either talked to the experimenter, or to other subjects who were part of the experiment. More recently, Paul Connerton (1989) has drawn attention to other, less obvious modes of communication, and the explicit aim of his book *How Societies Remember* is to investigate the ways in which the collective memory of social groups is transmitted and upheld. Connerton identifies the

performative aspects of social life as the strongest carriers of cultural information. Within these performative aspects he includes both commemorative ceremonies and the less intuitive "bodily practices".

Commemorative ceremonies are rites that happen regularly in time; they are "stylised, stereotyped and repetitive" (Connerton, 1989, p. 44) in order to be as mnemonically effective as possible. Through the use of repetition, which implies continuity in time, and through the explicit claim of direct descent from an original event, the commemorative ceremonies of a culture are quintessential signifiers of permanence, units of memory that link the present society with its own idea of its past. In doing so, they legitimise both the present society and its current notion of past events.

By "bodily practices" Connerton intends both the production of artefacts and the ways in which people learn to habituate their body movements according to the expectation of their social group(s). Examples of this latter notion are table manners, posture and hand gestures. In Connerton's view, the way we change our environment by producing artefacts and the way we learn to use our bodies are, together, the backbone of cultural transmission.

Since both are performative rather than discursive practices, "Both commemorative ceremonies and bodily practices therefore contain a measure of insurance against the process of cumulative questioning entailed in all discursive practices" (Connerton, 1989, p. 102). Connerton argues that particularly in the case of everyday bodily practices, their performance takes place in an almost completely unconscious manner, thus making their questioning virtually impossible.

Ultimately, Connerton too resorts to the notion of schema, although he does not explicitly thus name it. His references are to mental categories ("The body, reduced to the status of a sign, signifies by virtue of being a highly adaptable vehicle for the expression of mental categories" [p. 95]) and to systems of expectation:

Prior to any single experience, our mind is already predisposed with a framework of outlines, of typical shapes of experienced objects. To perceive an object or act upon it is to locate it within this system of expectations. The world of the percipient, defined in terms of temporal experience, is an organised body of expectations based on recollection. (p. 6)

Thus, for Connerton as for Bartlett, communication entails acting according to an organised set of mental categories (a schema) and, in the process, modifying or strengthening the schemas of all communicators involved. However, the difference between the two is that Bartlett implicitly assumes the communication mode to be mostly verbal, while Connerton maintains that the nature of bodily actions provides a more stable communication mode over many generations.

The most important point here is that schemata – intended as networks of organised information in a person's memory – are responsible for the stability and predictability of communication, of both the verbal and the physical, bodily type. This fact supports in a few ways my constraints argument. First of all, it broadly identifies the processes behind social communication as being primarily mnemonic. Second, it coincides with the description that is commonly made of some of the constraints I have

listed, such as semantics, imagery and, to a certain extent, syntax.³³ Third and most important, it makes a further case for those "surface schemata" that I discussed in Chapter Two. Surface schemata facilitate the mnemonic association of speech sounds and prosodic elements of speech, and are instrumental in the recall/composition of utterances.

From the evidence above, it follows that most of the constraints I have identified in Chapter One can be accounted for as schemata that all speakers possess. The only exceptions are (a) rhythm and (b) the reasoning that is involved in applying syntactic rules (i.e., the "transformational" side of grammar, in generative terms).³⁴ The production of an utterance must involve the interaction of many schemata in a synergistic manner, whereby several different types of schemata (sound, semantic, imagery and syntactic) work in parallel to select a chunk of utterance. It seems probable that among different candidates, the chunk that is selected for production is the one that connects simultaneously to the greatest number of schemata.

³³ Semantic and image schemata are by now a commonplace in psychology. In addition, syntax should also be partially reducible to schemata in terms of storage of the syntactic knowledge that is manipulated when using grammar.

³⁴ However, neither of these two exceptions is truly "exceptional". As discussed, reasoning and rule-application are observed in many other human (and animal) activities aside from speech, and rhythm is a natural occurrence in the bodily actions as well as the cognitive experience of both humans and animals of all species (see Fraisse, 1974, ch. 1, and Havelock, 1986, p. 72).

2. Speech and Communicability

It is interesting to speculate on what makes an instance of speech particularly communicable or (with more precision) what makes it connect simultaneously to the greatest number of schemata. As we have seen, several types of constraints must be satisfied, and several types of schemata exist into which the characteristics of an utterance may or may not fit. A word or an expression that is successfully remembered and repeated by a population of speakers must conform to speech constraints produced by semantics, imagery, syntax, rhythm, prosody and repetitions; therefore, a word or phrase will be highly communicable if it fits these constraints particularly well. The word or phrase may be generic enough to be used in many contexts (broad semantic constraint). Possibly, it will cover more than one syntactic function, or it will have derivations that cover different functions, or it will allow for many different syntactic constructions (broad syntactic constraint). It will be well suited to the rhythm of the language; for example, in the case of English this may mean that it will be short and will include only one stress, as opposed to more complicated patterns such as a primary and a secondary stress. It will fit different intonation contours; again, in English this probably means a shorter word with only one stress. The upper limit of word length is obviously given by the intonation unit. Any word that is longer than an intonation unit would be difficult to pronounce, because the speaker would have to resort to breathing in the middle of the word (even the longest German words are kept in check by this limitation).

All the common words of a language fit these characteristics, since these words are all part of the same oral tradition. They are the result of the screening that speech constraints operate on speakers' output. Among all words, some will fit these characteristics better than others, and will therefore be more communicable. Thus, the most frequently used words in English should definitely comply with the characteristics listed above. I have included a list of such words in Appendix One³⁵ so that the reader can compare these words with the criteria I discussed; their correspondence is evident.

The constraints might not be satisfied simultaneously in all cases, and I have no doubt that there must be many words, in the lexica of all languages, that defy one or more of them. However, if I am correct with regard to the role of constraints in lexical selection, there can be no word that will defy all of them.

There might also be words or phrases that are particularly memorable precisely because they defy one of the constraints. I base this hypothesis on a parallel with Norenzayan and Atran's findings on "minimally counterintuitive" beliefs. In their article "Cognitive and emotional processes in the cultural transmission of natural and nonnatural beliefs," Norenzayan and Atran explain that minimally counterintuitive beliefs are mostly consistent with a person's general understanding of natural phenomena, but possess a few features that are inconsistent with this understanding. These features are more memorable because they are imaginable and somewhat easy to believe, yet interesting because unusual. As an example, the authors refer to common religious beliefs, which include many feasible concepts blended together with a few unfeasible ones. In the completely different domain of word use, there could be a comparable

³⁵ An Internet search has provided me with various lists of the most common English words; all lists are somewhat arbitrary, of course, although all of them make reference to sources such as dictionaries and vast corpora of samples, both written and spoken. Nevertheless, all lists are quite similar, and I believe the reader will recognise at once that these words are indeed ubiquitous. The Appendix contains the first 250 words from one of these Internet lists.

phenomenon: catchy expressions may succeed at drawing attention to themselves with the same ratio of feasibility/unfeasibility. These expressions would generally fit the constraints, but would also deviate in some noticeable manner from at least one of them. We could call this the "supercalifragilisticexpialidocious" effect.

Moreover, judging by the examples I have analysed in the domain of oral traditions and by the slang that is currently in circulation, it seems that out of the three general domains of semantic, syntactic and prosodic constraints, prosody is the most successful at facilitating memorability. I would posit that in order to be memorable, an expression will most likely have to satisfy the constraints of prosody particularly well, and will probably be more aurally redundant than other expressions. For example, it will probably include some distinct sound repetitions and a very clear beat. However, this is mere speculation for now, and much more work must be done on speech constraints before something of this nature is ascertained.

As the studies on rumour and gossip have shown, communicable utterances and words must fit into a psychological as well as a speaking environment. The psychological fit of utterances depends in particular on their semantic value, i.e., the idea(s) that they help perpetuate. Utterances may be more communicable if they can fulfil the semantic constraint of speech production particularly well. A few studies have concentrated on this specific dimension of communicability by investigating the semantic content of transmission tokens.

In their article "Selective Pressure on the Once and Future Contents of Ethnic Stereotypes: Effects of the Communicability of Traits", Schaller, Conway and Tanchuk (2002) analyse the relations that exist between the communicability of a stereotypical

trait, its likelihood to occur in common discourse and its persistence over time. After scoring 76 distinct stereotypical traits for communicability, the authors conduct an analysis of stereotypes regarding black Americans that were recorded by five independent research studies and collected over a period of 60 years. The authors' findings show that higher (better) scores consistently predicted the retention of traits over time. Interestingly, persistence over time was more significant when longer time intervals were used for analysis; communicability predicted retention over three or four generations more accurately than over one or two. The authors attribute these results to the fact that "the effects of communication should be compounded [...] as more opportunities for communication-based selection occur" (p. 870). In a separate experiment described in the same article, the authors observe that the effects of communicability on the persistence of traits are accentuated for groups that are more likely to be talked about, while they are almost insignificant for those groups that are seldom the object of conversation.

At a theoretical level, the article discusses the nature of communicability in a general sense, without offering a formal definition. Measures of communicability are exemplified in the ease of pronunciation or in the length of a certain word, in the number of readily available synonyms with which an idea may be expressed, or whether the information that is communicated is perceived as useful or interesting. At the practical level, the communicability of the 76 traits was scored by means of a questionnaire that experiment participants were requested to complete, in which they indicated the potential likelihood of discussing each trait in their daily social interactions. In this way, the authors cleverly sidestepped the thorny issue of defining

communicability and instead armed themselves with experimental data on which to found their observations.

In his chapter on "Unintended Influence" Schaller (2001) further develops the concept of communicability. He stresses that, when dealing with information and the spread of memes,³⁶ a rule must be kept in mind:

Whatever information is more likely to be communicated is more likely to become common within any population of human beings. It is for this reason that attitudes, beliefs and other memes can be thought of as viruses [...]. [...] just as the social influence underlying the transmission of a virus typically operates outside the realm of intention or awareness, so too the social influence underlying the transmission of memes often occurs unintentionally and outside of awareness. It is clear that cultural norms, like viral pandemics, can and do emerge and persist simply as the result of the social influence that accompanies interpersonal communication. [...] Some memes may be more likely to be communicated and thus more likely to become and remain widespread. Anything that is especially 'communicable' is likely to be normative. (p. 80)

³⁶ Memes are units of information transmission. This simple yet broad and hazy concept has been proposed by Richard Dawkins in *The Selfish Gene* and has spurred a fascinating school of studies in culture and communication. I think that many of this area's findings are quite compatible with my views on constrained speech processing. For more details, see the Dawkins, Blackmore, Dennet, Hull, Aunger, Feldman, Heath, Vaneechoutte and Campbell.

As for what constitutes a communicable trait or meme, Schaller explains that both the characteristics of the meme itself and the characteristics of the environment in which the meme exists are important in determining the success of communication. Schaller circumscribes the characteristics of the environment more narrowly to the actions of the person who is performing the communication and to his/her audience. The first characteristic Schaller lists is the perceived popularity of the meme; anyone who adheres to popular notions and beliefs will not only reinforce their allegiance to their social group, but will also convey those pieces of information that the group, in its unanimity, insists on conveying. In many situations, endorsement by a group can be an extremely strong incentive in the adoption and further communication of a meme.

Moreover, it is true that sometimes

when communicating with other people, we strategically choose what to say and what not to say in order to make a good impression. This means that whatever publicly articulated memes best serve individuals' impression management goals will be especially communicable, and most likely to emerge as widespread cultural norms. (Schaller, 2001, p.

85)

Consistency with existing notions, as well as ease in understanding the notion communicated by a meme, are also very important:

> beliefs that are consistent with preexisting cognitive structures are more likely to be successfully communicated than inconsistent ones. Another powerful contributor to epistemic comfort is ease of understanding. Consequently, simpler beliefs are more communicable

than complex ones, and are more likely to become culturally normative. [...]

The communicability of simple ideas is evident not just in casual conversation, but in more formalised forms of communication as well, such as scientific discourse. Scientists typically try to communicate parsimonious explanations for phenomena, and even when more complex explanations are transmitted, they are less readily understood and retransmitted than are simpler but less complete (and often less accurate) explanations. (Schaller, 2001, p. 87)

The studies conducted by Schaller (2001) and by Schaller, Conway and Tanchuk (2002) shed some light on the psychological processes involved in the communication of information, as well as on the relevance of the features of information itself in its quest to be communicated. The results of these studies can be employed to expand my original description of communicability, especially with regard to its semantic features.

At the beginning of this section, I suggested that a communicable utterance may cover more than one syntactic function or have derivations that do, that it will probably be well suited to the rhythm of the language, and that it may fit different intonations or may be associated with a particularly memorable intonation. To these formal features I also added the fact that an utterance should have "broad" semantic adaptability in order to fit in many contexts. I can now refine the definition of this "broad" semantic constraint: an especially communicable utterance conveys a concept that is relatively simple to understand; it can be used in discussing widespread, common topics; it may be endorsed or even modelled by a dominant group; and it conveys a notion that fits well with pre-existing, familiar notions (i.e., with pre-existing semantic schemata).

Again, these characteristics are suited to a type of information transmission that is bound to proceed in recursive, predictable and generally conservative patterns. As discussed in the previous section, the patterns are those of widespread, socially shared schemata.³⁷

None of the conclusions I have drawn above are surprising or unwarranted. These observations provide further insight into a definition of linguistic communicability. They may allow predictions on the communicability of a word or a phrase within a certain cultural domain by analysing its fit with the constraints at work in that domain. On the other hand, this approach certainly has limited use because, realistically, the only speech domains that can be analysed are current ones. Domains not observable in real time may provide insufficient data to determine what speech constraints are at work. It is possible that some intonation contours have changed over time. We certainly know that some of the words and phrases that were regular objects of repetition decades ago in North America are no longer popular today; therefore, the relatively low or high word

³⁷ The communicability features I have discussed in this section are partially confirmed by Metcalf's observations in his volume *Predicting New Words*. I became aware of Metcalf's book (2002) after writing my observations above, and for this reason the correspondence between his remarks and mine seems even more striking. He proposes that five parameters determine the future success of a neologism in a particular "wordscape" (his term): frequency of use, unobtrusiveness, diversity of users and situations, generation of other forms and meanings, endurance of the concept. Some of these parameters (frequency, diversity, endurance) are broad descriptions of my semantic constraint, another (generation) of my syntactic constraint, and yet another (unobtrusiveness) of my sound constraint.

frequency effect of certain expressions may skew the probability for the semantic activation of other words. Similarly, their use may also have other effects on the rest of the utterances in which they were included. I do not think that it is possible to estimate the incidence of these factors without having access to the particularities of the speaking environment itself. Nevertheless, if data is available and observable in real time, some predictions on communicability and on communication patterns may be performed with relative accuracy.

3. Summary: Cultural Transmission and the Production of Speech

In this chapter, I supplemented the constraints argument of Chapter One with additional evidence from studies on the transmission of cultural information. I started by illustrating the fact that cultural information transmission across populations proceeds in a patterned fashion. The patterns of information transmission emerge from research in anthropology (especially Diffusionism), social psychology of rumour, and media-assisted social epidemics, and in mathematical formulations of communication dynamics. All researchers document transmission patterns that are mostly predictable, repetitive and recursive.

I then investigated the underlying mechanisms that generate these predictable transmission patterns. I discussed Bartlett's schemata, intended as networks of associations that organise the storage of mnemonic data in people's brains. With his experiments, Bartlett showed that the organisation of memory is both predictable and dynamic; memory possesses a relatively fixed structure that remains consistent from person to person and over time, but also possesses the capacity to adapt and incorporate new information and experience. Bartlett showed that the social dimension

of memory, i.e. the notions and associations that we acquire from our social environment, have a causal and a creative effect on our recollection, perception and interpretation of reality. Most importantly, Bartlett demonstrated that recall is a process of reconstruction aided jointly by pre-acquired notions and by environmental cues.

This last point parallels the arguments I developed in chapter One, where I argued that speaking occurs via a similar process. Speaking is also a process of composition/reconstruction that is guided by several constraining parameters; it is aided by pre-acquired notions (such as linguistic habits) and by environmental cues, including the speech of others. The resemblance between these two processes is not coincidental; a causal bond exists. Speaking is quite obviously enabled by psychological schemata, since the process of utterance production is a process of creative recall. As I already discussed in the course of this chapter, many speech constraints I identified in Chapter One are determined by mnemonic schemata: semantic information, imagery, syntactic knowledge, and sound repetitions organise themselves into networks of associations, and it is along the lines of these association that the predictable composition of speech proceeds. In fact, I have conjectured that a combined constraints process may be linked to the activation of several schemata at once; specifically, a segment of utterance (e.g. a word, phrase or chunk of speech) may be selected for production because of its simultaneous and convergent association to multiple semiactive schemata.

In this chapter I also discussed the concept of communicability, intended as the frequency with which a population produces a certain utterance. I made some suggestions on the features that highly communicable utterances may possess, such as a broad semantic applicability, a versatile syntactic fitness, a particularly memorable beat or intonation, and so on.

The arguments of this chapter are tightly related not only to the ones explored in Chapter One, but also to the subject matter of Chapter Two. Oral traditions are obvious examples of a particularly successful kind of cultural transmission, in terms of both their longevity and the fidelity with which cultural information is repeated. In fact, the causal connection between the composition/recall of oral poetry and schemata had already been noted in Chapter Two.

Speaking entails a great deal of remembering, although this simple fact is too often overlooked in linguistic literature. The social dimension of memory and its causal and creative effects (i.e., Halbwachs' collective memory) are a crucial component in the production of utterances. Memory and society are indivisible; they create a symbiosis in which each of the two elements has indispensable structural effects on the other. Without the organisation of memory there can be no organised (social) behaviour, and without social behaviour there can be no organised memory. The organisation of memory progresses in a recursive fashion, because new information can only be understood and retained if it maintains some consistency with existing information. Hence, the assimilation of new information in a mnemonic network expands the limits of the network, but also always confirms the viability of the connections that were originally part of the network. The fundamental recursiveness that can be observed in oral traditions and in the transmission of culture in general is an essential feature of human psychology. For this reason, it is also endemic to the act of speaking.

Conclusion

In the chapters above, I have been demonstrating that the process of composing utterances must adhere to several constraints simultaneously. These constraints are determined by the information that is most readily accessible in a speaker's memory and by habits of speaking that are shared by a whole population and its in-groups. An utterance must adhere to syntactic constructions, semantic activation, prosodic contours, rhythm, and acoustic mechanisms of word selection that facilitate the use of repetitions, formulas and echoes.

None of the individual aspects of speech analysis I have illustrated is particularly innovative; many researchers specialise in investigating syntax, semantics, prosody, speech and conversational rhythm, formulas, repetitions and (to a lesser extent) echoes. My contribution has been to point out that (a) each of the mechanisms underlying these characteristics of speech exercises a constraining effect on the selection of chunks in working memory, and therefore on the production of utterances; and that (b) these constraints must work concurrently – and are all the more effective because of their co-occurrence.

I hope that the comparison in Chapter Three between speaking and the spontaneous composition of oral pieces has allowed for a broader understanding of speech constraints at work. This line of reasoning links artistic oral performance with everyday skills, and demonstrates that the difference between the two is a matter of quantity rather than quality. The "genius" of an oral performer is therefore a matter of refining common skills, rather than possessing a wholly uncommon ability. A marked

advantage in considering the similarity between speech and oral traditions is that the latter can offer a conceptual model that is quite foreign to the former. Even after spending much time researching speech constraints, it is still quite difficult even for me to envision speech production as a constrained and internally conservative process. I am so used to believing that I can open my mouth and say what I want, that the whole idea of possessing automatic mechanisms that select my utterances before I am conscious of them is really quite counter-intuitive.

Yet this notion becomes more understandable if I play a trick and imagine the speaking environment as that of a rather strict oral tradition. Visualise, if you can, that everyone around you is speaking in rhymes - not necessarily the nonsensical rhymes of children, but simply the rhymes of limericks. It is not easy to conjure up this thought, and it certainly feels quite silly, yet I have shown above that this scenario is not so far from reality. If we think of speech in this way, then it becomes much easier to understand why repetitions are so often used in production, and why they are so important for the cohesion of discourse. If you had to produce your own limericks on the spot, it would be very handy to be able to borrow pieces of verses and rhyming word pairs from others' compositions or from your past ones. If you were to have a dialogue in limericks, some back and forth borrowing would help the situation guite a bit. You can also imagine how the necessity to respect a certain structure would impact your use of the language; you may find yourself expressing a thought in a roundabout way, in order to employ a formula or a rhyme that is in your repertoire. But most importantly, you can envision more clearly how your memory would play a big part in the assembly of each verse, because you would be searching your memory for words and phrases of the right length, with the right rhythm, and with the right configuration of sounds. Yet this

intense memory search would still feel rather natural because some of its parameters, like verse length or rhyme, would not involve counting syllables or combing the dictionary, but simply keeping a beat and letting the words cluster by themselves.

I think this image facilitates the realisation that speech is not effortless. It also clearly shows that much happens in speech production that depends on mnemonic processes. Finally, it portrays the fact that grammar and semantics are only a few of the mechanisms at work, and that other, "surface" mechanisms are just as central to the assembly of utterances.

The argument I unfolded in this work has centred on the properties of speech as a linguistic medium. I have investigated the oral/aural properties of speech, and the organising mechanisms inherent in the act of speaking. As a result, I have discussed speech primarily as a product of memory, with the additional facilitation of cognitive processes such as rhythm-keeping and rule-based reasoning.

I believe that the considerations that led me to this research are missing from the majority of current linguistic endeavours. As discussed in the Introduction, most formal theoretical approaches to linguistic research posit language as an abstraction entirely detached from its tangible embodiment. While many scholars have attacked this assumption, their criticism has yet to produce much change in the way in which formal theories of language are understood and researched. Nevertheless, it is evident that the organisation of linguistic production depends on its medium. Spoken language is different from written language. This simple fact is disregarded by most current theories of language; moreover, it is possible to demonstrate that these theories are founded on considerations that can be applied more easily to writing than to speaking.

Speaking may seem more "natural" than writing. This notion is probably due to the fact that writing, unlike speaking, necessitates the availability of technology in order to be produced and re-produced (transmitted). Regardless of whether written information is in the form of inscriptions scratched on mud tablets or keystrokes viewed on computer screens, writing has had to rely on both a standardised system of signs and on artefacts to which its signs could be committed. Both types of inventions – writing systems and writing media – are technological innovations. When reviewing the technological progress that has affected the written word in the course of history (from feather and ink to printing presses, to photocopiers, to websites), it is apparent that the primary advantage of each innovation is the new ease with which words can be repeated or reproduced.

By contrast, speech seems quite effortless and unmediated. Indeed, Saussure believed this to be the case when he described language as "the union of meanings and sound-images" (1959, p. 15), and the foundation of language as an ephemeral entity called "thought-sound" (p. 112). Although on the one hand, Saussure stressed that language, as a semiologic system, is thoroughly detached from the practice of speaking, on the other hand he regarded speaking as the direct embodiment of thought, and therefore as the most privileged of all semiologic systems. In his view, writing is derived from speaking and inferior to it:

Language and writing are two distinct systems of signs; the second exists for the sole purpose of representing the first. The linguistic object is not both the written and the spoken forms of words; the spoken forms alone constitute the object. But the spoken word is so

intimately bound to its written image that the latter manages to usurp the main role. People attach even more importance to the written image of a vocal sign than to the sign itself. A similar mistake would be in thinking that more can be learned about someone by looking at his photograph than by viewing him directly. (Saussure, 1959, p. 23-4)

This view is too partial to be correct. As Holdcroft (1991) states in his study of Saussure's theory,

granting that historically speech always antedated writing, and that 'all systems of writing are demonstrably based upon units of spoken language' (Lyons 1968, 39), it does not follow that a written word actually represents its spoken form. And even Saussure would have to concede – particularly as it is part of his complaint about the tyranny of written language – that, once established, the written language would acquire a life of its own. Even if it started out as a parasitic system, it would become increasingly autonomous. (p. 38-9)

Derrida's rebuttal of Saussure's views on writing became one of the ideas at the core of Deconstruction. His argument can be summarised as follows: if systems of signs rest on simply positing signifiers and signifieds (roughly translatable as signs and their corresponding meanings), there is no reason to believe that written signs should be in any way inferior to spoken ones (Holdcroft, 1991, p. 39-41). Derrida challenges the notion that speech should be granted any privilege; instead, he posits writing as an equally valid semiologic system, and perhaps even a superior one (see especially Derrida, 1973, and 1974, p. 6-14, for a rather obscure elaboration on the superiority and antecedence of writing to speech).

· 20 ·

In Chapter 2 of *Of Grammatology*, Derrida (1974) attacks modern linguistics for having inherited Saussure's covert bias towards spoken language.³⁸ He argues that linguistics is founded on phonology, and that this fact is a direct consequence of Saussure's views on the primacy of speech over writing:

Linguistics thus wishes to be the science of language. [...] Let us first simply consider that the scientificity of that science is often acknowledged because of its *phonological* foundations. Phonology, it is often said today, communicates its scientificity to linguistics, which in turn serves as the epistemological model for all the sciences of man. Since the deliberate and systematic phonological orientation of linguistics (Troubetzkoy, Jakobson, Martinet) carries out an intention which was originally Saussure's, I shall, at least provisionally, confine myself to the latter. (p. 29)

Derrida's views on phonology and linguistics in general are only partially correct. He does have good reason to take issue with Saussure's remarks, as I have described above. Derrida also takes issue with Structuralists Troubetzkoy, Jakobson and Martinet. These scholars attempted to describe, in different ways, the use of speech sounds within a language as coherent systems of sound features. Their research centred on speech as a system of oral, arbitrary (semiologic) signs. Therefore, in Derrida's view, their investigations were a direct result of Saussure's preferential treatment of speech as

³⁸ In *Positions* (1981), he seems to develop this argument by stating that writing should be the paramount object of linguistic enquiry: "The gram [written sign], then, is the most general concept of semiology – which thus becomes grammatology – and it covers not only the field of writing in the restricted sense, but also the field of linguistics" (p. 26).

a system of signs, at the expense of writing. However, Derrida's hostility towards phonology is not entirely justified.

In fact, Derrida's argument that linguistics is founded on phonology is at least partially misguided. A large part of modern theoretical linguistics has been devoted to the investigation of syntactic structures without specifying whether these structures are to be found in speech or in writing. By the time Of Grammatology was published in 1967, Derrida's argument against phonology was out-of-date. Chomsky's Syntactic Structures, published in 1957, had launched the so-called "generative revolution"; and his Aspects of the Theory of Syntax (1965) attempted a complete account of generative theory, with syntax at its core. As I briefly discussed in the Introduction, Chomsky's view of language competence as the primary object of scholarly examination automatically confines all instances of performance (i.e., both speech and writing) to the background. Moreover, in generative theory, phonology is definitely secondary. It is intended as a "module" that acts on the speech signal after the primary "syntactic module" has composed an utterance; phonology has no input on the syntactic formation of speech, but it is rather dependent on it (Schlüter, 2005, p. 11). Hence, in much of modern theoretical linguistics, phonology is at worst unimportant and at best of only secondary importance to the study of language.³⁹

The disregard for linguistic media, and for the forms that language as communicative expression can take, is not exclusive to Generative linguistics. In fact,

³⁹ As I will argue, the theoretical separation between "language as a conceptual system" and "language as communicative expression" (or competence/performance) has led to a theory of language that is in fact much more representative of written language than speech. Derrida seems to have missed this point.
the Saussurean claim that writing is simply the representation of speech, and that writing is practically interchangeable with speech, is fairly common. This claim, which infuriated Derrida, actually hides a bias towards written rather than spoken language.⁴⁰ I will demonstrate this bias below.

Let me consider, for example, the ideas of the famous linguist Edward Sapir, a Descriptivist *sui generis* who studied directly in the field many native North American languages. Unlike Saussure, Sapir was aware of the fact that there are substantial formal differences between speech and written language. He welcomed the use of the International Phonetic Alphabet as an attempt to capture those features that are exclusive to speech and that cannot be reproduced by common writing. The crucial point for the present argument, however, is that despite his attention to the spoken signal and its distinctive features, Sapir (1949) still posits an equivalence between speech and its written representation. Sapir proposed the following definition: "Language is a purely human and non-instinctive method of communicating ideas, emotions, and desires by means of a system of voluntarily produced symbols" (p. 8). Thus, language is a system of symbols voluntarily employed by human beings to exchange rationally formulated meanings with one another. Note that the definition does not differentiate between

⁴⁰ I am not entirely sure whether Derrida was oblivious to this hidden bias, or whether he understood it and attempted to expose it in his *Speech and Phenomena*. My sense is that he was oblivious, since he puts much effort into rebutting linguistic claims in general, but little effort into analysing any linguistic study apart from Saussure's. It is therefore somewhat ironic that he attacked linguistics for its speech-centred approach, when it can be demonstrated that much of linguistic theory is more applicable to writing than it is to speech.

spoken and written language. However, after his definition, Sapir (1949) continues by discussing speech in particular:

These symbols are, in the first instance, auditory and they are produced by the so-called "organs-of-speech". There is no discernible instinctive basis in human speech as such, however much instinctive expressions and the natural environment may serve as a stimulus for the development of certain elements of speech, however much instinctive tendencies, motor and other, may give a predetermined range or mould to linguistic expression. Such human or animal communication, if 'communication' it might be called, as is brought about by involuntary, instinctive cries is not, in our sense, language at all. (p. 8)

In Sapir's vision of spoken interactions, each speaker knows when they are speaking, why they are speaking, what meanings they wish to communicate and what expressions they are going to employ in order to communicate them to their listener. Sapir's vision is a tenable description of some types of language, but it is not applicable to speech. Consider, for example:

CREATE OR REPLACE PROCEDURE flattenData IS

Cursor data1 IS

SELECT * FROM sample_sql_table;

old_usr_id number := 0;

BEGIN

FOR attributeRecord IN data1 LOOP

```
IF old_usr_id <> attributeRecord.usr_id THEN
```

COMMIT;

INSERT INTO sample_sql_table2 (usr_id, usr_identifier)

VALUES (attributeRecord.usr_id, attributeRecord.usr_identifier);

old_usr_id := attributeRecord.usr_id;

END IF;

IF attributeRecord.att_name = 'Attribute1' THEN

UPDATE sample_sql_table2 SET Attribute1=trim(attributeRecord.value)

WHERE usr_id = old_usr_id;

ELSIF attributeRecord.att_name = 'Attribute2' THEN

UPDATE sample_sql_table2 SET Attribute2=trim(attributeRecord.value)

WHERE usr_id = old_usr_id;

ELSIF attributeRecord.att_name = 'Attribute3' THEN

UPDATE sample_sql_table2 SET Attribute3=trim(attributeRecord.value)

WHERE usr_id = old_usr_id;

END IF;

END LOOP;

COMMIT;

END flattenData;

/

In the few lines of code above, written in PL\SQL (a database language for Oracle systems), the coder declares that s/he is about to start a procedure called "flattenData". The coder then declares that s/he is going to use the source table "data1" and the variable "old_usr_id", which is a number and is set to zero. The software engine that will

run this code is required to accept this variable at its symbolic face value. The coder continues by instructing the software engine to take attribute values and insert them in a cross-reference table once it encounters the specific circumstances that the coder is describing in the routine (i.e., once the user ID in question changes so that old_usr_id is not equal to attributeRecord.usr_id). Finally, the statement "END flattenData" closes the procedure, effectively separating its instructions and variable declarations from all other procedures that the software engine might encounter.

This example satisfies all the parameters in Sapir's definition of language; the language employed is entirely voluntary, symbolic, non-instinctive and exclusively human. But speech does not adhere to these parameters, and it is surprising that a man who spent the greatest part of his professional life studying the exclusively spoken languages of several American and Canadian native tribes should formulate such a definition of it. Let us examine the fallacies inherent in each of his parameters.

Speaking is not entirely a voluntary action for several reasons. To begin with, a child does not choose to learn his/her mother tongue deliberately; s/he learns it from the surrounding speakers without having the option of not doing so. All children with normal learning abilities will acquire the language(s) they are exposed to in an involuntary and uncalculated manner. Furthermore, speech is the default method of human information transmission, so if we are spoken to in a language that we understand and in an audible and coherent manner, we cannot fail to listen to and understand what is being said to us. No choice is entailed; we simply cannot tune out our brains or ears, at least not without some effort (hence the effectiveness of TV and radio advertising). Most commonly, when we are asked a question we answer without making the conscious choice to do so, concentrating more on the accuracy or gist of our

answer than on the possible option of not answering at all. And when we need to communicate an opinion or a fact to somebody else, we can rarely afford the time to plan every single word that we are going to use. Normally, we learn which words we want to use to express ourselves only after we have uttered them. For all these reasons, there seems to be little volition in the daily act of speaking, or listening to and understanding speech. In fact, in everyday social circumstances, it takes some conscious effort for any able speaker not to speak or not to listen to a language they understand. Speech can be voluntary, but more often it is simply an automatic reaction to other people's speech.

Sapir stated that speech is symbolic. However, it is far more important to point out that speech is combinatorial. By this, I mean that the ability of speech to label objects and concepts, although important, is only secondary to its ability to create new concepts by juxtaposing two existing symbols (words). Even though its interpretations can vary, I feel that Sapir's statement on the symbolic nature of speech is too much of an oversimplification. While symbolism can support a complex information system, Sapir's claim has rather more to do with an iconic idea of language. Upon reading it, one can almost picture two people standing about and coming up with words for each element of their surroundings: there's a tall green thing, let's call it "tree"; here's a short one, let's call it "plant"... Evidently, language did not come about in this encyclopaedic manner, nor is it used nor does it evolve in this way. From an evolutionary point of view, we must consider that the use of very simple sounds, iconic sounds, probably came about first. Many species in the animal kingdom routinely use mating calls, danger calls, grouping calls and so on. It is most likely that human language also evolved from such basic calls, which are undoubtedly iconic in the sense that they entail a one-to-one correspondence

with a certain situation or intention. However, in the course of time, speech evolved in complexity – sentences and entire systems of sentences developed. Due to the combinatorial nature of speech, in which words are put into relation with each other, the immediate iconic link between a word and its counterpart becomes less important. Certainly the link between a word and a physical counterpart is unimportant, otherwise we would not be able to speak of anything that goes beyond our immediate experience of reality. But also the link between a word and a purely conceptual counterpart is less crucial than one might suppose, because the symbolism of the word does not correspond to a static and unequivocally identical notion in each speaker's mind. Each speaker will embody the notion suggested by the word in a personal manner dictated by experience and by the immediate circumstances in which the word is being used. As such, the truly "real" element in a speech act is the utterance itself and its momentary usage; any symbolised counterparts are only fleeting, context-dependent translations effected by each of the speakers and listeners involved.

As for the claim that speech is non-instinctive, this topic has been the centre of a very intense debate in the speech sciences for some time. The Chomskyan School views the ability to acquire and use language as a fundamentally innate skill (Pinker, 1994); others posit a sound processing instinct and a motor processing instinct at the origin of language development (Jusczyk, 1997 and 2002). The arguments of these two schools of thought are too vast to be explained here. The important point is that both opinions assign a large role to instinctual processes in the development of language; therefore, Sapir's view does not hold much credibility in current linguistic studies.

Finally, Sapir posited that language is exclusively human. However, as discussed above, many animals employ acoustic signals to communicate with their conspecifics.

These signals are very basic and stand each as a symbol for one specific meaning. Their use of calls corresponds exactly with Sapir's iconic notion of human language. In his own view, then, the difference between animal and human language should be one of quantity, not of quality; humans merely use more symbols in a more complicated manner.

Sapir had an intimate understanding of linguistics and extensive first-hand experience with the indigenous languages of many native tribes. For these reasons, it is quite unlikely that his theory of language would be applicable to computer code alone. In my efforts to understand Sapir's point of view, I have found the following passage illuminating:

Written language is thus a point-to-point equivalence, to borrow a mathematical phrase, to its spoken counterpart. The written forms are secondary symbols of the spoken ones – symbols of symbols – yet so close is the correspondence that they may, not only in theory but in the actual practice of certain eye-readers and, possibly, in certain types of thinking, be entirely substituted for the spoken ones. Yet the auditory-motor associations are probably always latent at the least, that is, they are unconsciously brought into play. (Sapir, 1949, p. 20)

Here, we are back to Saussure's ideas on speaking and writing: if spoken words are symbols and written words are simply their visible representation – "symbols of symbols" – then the analysis of written words is equivalent to the analysis of spoken words. Although Sapir might be referring here to the phonetic symbols of the International Phonetic Alphabet, instead of common written texts, his view is still at least partially incorrect. The proposed equivalence between speaking and writing is

mistaken, regardless of the type of writing we are to consider. As I have illustrated above, Sapir's qualifications cannot be applied to speech without running into many problems, but they are perfectly accurate if applied to writing – i.e., phonetic as well as common writing. Writing is voluntary; it is quite unlikely that somebody will not be conscious of their act of writing or typing, unless they are drunk, on drugs or in some very peculiar state of mind. Unlike in the case of speech, it is much easier not to write than to write, as any graduate student will be happy to attest.

To continue with Sapir's parameters, writing is symbolic, since it was developed as a stand-in for a series of speech sounds.⁴¹ Writing is certainly non-instinctive, since only a fraction of the world's languages actually encompass a writing system. And finally, it is exclusively human – or at least, we have yet to discover another species that produces marks or engravings in any consistent and combinatorial manner.

Sapir applied his parameters, conceived by a fully literate mind and regarding primarily written records, to all types of language, including the spoken language of the many illiterate individuals he studied. However, although his discussion centres explicitly on speech, his observations are justifiable only if he actually started his analysis by examining a written form of speech and then continued by applying his findings to the oral sources of his written records. The procedural problem in this approach is not immediately evident unless one has had the opportunity to focus on the actual nature of the written sign as opposed to the spoken utterance. Unfortunately, at the time of Sapir's investigations, the only recording medium readily available to the language

⁴¹ However, in the course of time, writing has developed some independence from the speech sounds it is supposed to represent (see Holdcroft's quotation above). Hence, as Derrida argues, the symbolic nature of writing is valid independently of any connection to speech.

scholar was writing. Thus, in spite of Sapir's attention to spoken rather than written forms of language, available technology forced him to examine speech that had been transcribed into a written form.

The nature of written and spoken language is quite different. In the chapters above, I have considered speech as a medium, and I have examined its formative constraints. Writing is also a medium, and it too has its constraints. These two types of linguistic expression are partially different both in terms of their structure and in terms of the cognitive processes we employ to produce and understand them (see Jahandarie, 1999, for a thorough account of current cognitive and linguistic research on speaking and writing). Indeed, throughout my work I have been developing the view that natural language cannot exist as an independent entity outside of these modes of expression.⁴² The assumption that performance does not matter as much as competence has lead to the kind of misrepresentations that I have discussed above. I hope that my analysis of Sapir's parameters has succeeded in highlighting some of the differences between the production of speech and the production of writing. I will now consider briefly some of the most obvious formal differences between these two types of linguistic expression.

Writing must rely on external media in order to be produced. External media allow the offloading of mnemonic data; therefore, the media interfere with the constraints that are at work in the production of speech (Ong, 1987). When memory is no longer a limitation in the production of language, many features of linguistic use become more apparent. For example, I have already discussed in Chapter One the fact that repetitions

⁴² Admittedly, the present discussion ignores sign language, which I have excluded from my current focus.

are extremely common in speech, while they suddenly become obvious and unpleasant in writing. We have also seen in a piece of transcribed conversation that speech is often far more disjointed and syntactically imprecise than written language, although this fact does not seem to harm the goals or the enjoyment of personal communication. When language is static and it is possible to review it and make corrections, grammatical precision becomes more achievable and more important. Finally, since no extra-lexical information can be delivered without tone of voice, rhythm or other acoustic information, the exactness of syntactic structures and lexical choices becomes more important in writing than it is in speech. In fact, it can be argued that grammatical correctness, as it is generally understood, is a construct that originates with writing; speech often does not adhere to such rules of correctness, nor does it need to do so in order to achieve its communicative goals.⁴³

The use of an external storage medium such as paper counteracts some of the constraints of memory and changes some features in the language that is produced. However, even the use of a written medium cannot completely dispose of speech constraints. As Ong (1987) argues, speech remains very important not only because it occurs temporally before writing, but also because "written texts all have to be related somehow, directly or indirectly, to the world of sound, the natural habitat of language, to yield their meanings. 'Reading' a text means converting it to sound, aloud or in the

⁴³ For these reasons, it can be inferred that a view of language that gives preferential treatment to syntax and semantics must originate from written rather than spoken data. Because of this inherent bias, I believe that both the generative school and other syntactic/semantic systems, such as the various flavours of Functional grammar, are severely underequipped to explain the nature of daily spoken interaction.

imagination [...]. Writing can never dispense with orality" (p. 8). However, Ong's views are incomplete. It is true that writing is partially accessed and understood by means of an aural reading, even when this reading may be performed silently. On the other hand, writing is also accessed via orthographic comprehension, which is only partially related to aural (phonological) comprehension. As Jahandarie (1999) explains, the cognitive processes of reading, and their mutual interdependence, are still unclear (see especially ch. 8 and 9).

However, since it has been proven that writing is also accessed through phonological processing, some mnemonic constraints could be active on the aural dimension of reading. For this reason, many speech constraints may still be present in writing, even though they may be less obvious or somewhat distorted. For example, although written sentences are often longer and more complex than spoken ones, they still possess a certain rhythm (see also Schlüter, 2005). In fact, I would argue that one of the most hidden yet most common features of poor writing is a lack of regular rhythm in the word choice and sentence structure. Prosody is also latent in written language, since part of the function of punctuation is to allow the reader to partially recreate an intended prosodic contour in his/her imaginary ear. Other formal features of speech, such as echoes (i.e., alliteration, consonance), are common and can even be desirable in writing.

Speech and writing are interdependent. While the use of a medium may allow for added complexity and precision, the brains that process the linguistic signal still need it to adhere to certain parameters in order to find it formally pleasing and cognitively clear. On the other hand, a person who grows accustomed to the use of writing is bound to

use language in a different way than an illiterate person. Many years of training in reading and writing are likely to influence a person's linguistic habits, in terms of expanding vocabulary and syntactic skills.

Moreover, the understanding of one's own language will be different depending on whether an external medium such as writing is available or not. If writing is indeed available, it seems understandable that its visible representation of language, which becomes detached from its composer and therefore seemingly "more objective", would be a preferred object of inquiry. Yet, as I have explained, a view of language that is based on writing may lead to theoretical constructs that are fundamentally inadequate to explain the nature of speech. In contrast, an investigation into the practices of illiterate and semi-literate societies, such as their oral traditions, yields a much more holistic view of linguistic production, which accounts for both individual language skills and the fundamental role that memory and social interaction play in the production of language. Ultimately, the theoretical model obtained from such an holistic view is better equipped to explain the mechanisms of both spoken and written production, and to provide accurate predictions on the output of real speakers.

Bibliography

Abercrombie, D. (1967). *Elements of General Phonetics*. Edinburgh: Edinburgh University Press.

Allport, G. & Postman, L. (1947). *The Psychology of Rumor*. New York: Holt, Rinehart, & Winston.

Ambarabà ciccì ciccò. Retrieved 10 November 2005 from http://www.filastrocche.it

Am stram gram. Retrieved 13 November 2005 from http://www.filastrocche.it/ leggi.asp?id=3938&posizione=6, from http://www.ecole-plus.com/Chansons-MUSIQUES/chan sons/amstramgram.htm and from http://www.momes.net/ comptines/amstram gram.html

Aunger, R. (Ed.) (2000). *Darwinizing Culture: The Status of Memetics as a Science*. Oxford: Oxford University Press.

Aunger, R. (2002). The Electric Meme. New York: The Free Press.

Bakker, E. J. (1988). *Linguistics and Formulas in Homer*. Amsterdam: John Benjamins Publishing Co.

- Bakker, E. J. (1997). *Poetry in Speech: Orality and Homeric Discourse*. Ithaca: Cornell University Press.
- Balota, D. A., & Chumbley, J. I. (1984). Are Lexical Decisions a Good Measure of Lexical
 Access? The Role of Word Frequency in the Neglected Decision Stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340-57.
- Bartlett, F. (1967). *Remembering: a Study in Experimental and Social Psychology.* London : Cambridge University Press.
- Bateson. M. C. (1970) Structural Continuity in Poetry: A Linguistic Study in Five Preislamic Arabic Odes. Paris: Mouton & Co.
- Beckman, M. (1992). Evidence for Speech Rhythms across Languages. In Tohkura, Y.,
 Vatikiotis-Bateson, & E., Sagisaka, Y. (Eds). Speech Perception, Production and
 Linguistic Structure. Amsterdam: IOS Press.
- Beckman, M. & Edwards, J. (1992). Intonational Categories and the Articulatory Control of Duration. In Tohkura, Y., Vatikiotis-Bateson, E., & Sagisaka, Y. (Eds). Speech Perception, Production and Linguistic Structure. Amsterdam: IOS Press.

Blackmore, S. (1999). The Meme Machine. Oxford: Oxford University Press.

Bloch, B. (1950). Studies in Colloquial Japanese IV: Phonemics. Language, 26, 86-125.

Bold, A. (1979). The Ballad. London: Methuen & Co. Ltd.

Bolinger, D. (1976). Meaning and Memory. Forum Linguisticum, 1, 1-13.

Bolton, H.C. (1888). *The Counting-out Rhymes of Children: their antiquity, origin and wide distribution, a study in folk-lore.* New York: Appleton.

Bonamore Graves, A. (1986). *Italo-Hispanic Ballad Relationships: the Common Poetic Heritage*. London: Tamesis Books Ltd.

- Bourgeois, M. (2002). Heritability of attitudes constrains dynamic social impact. *Personality and Social Psychology Bulletin, 28,* 1063-1072.
- Boyd, R. & Richerson, P. (1985). *Culture and the Evolutionary Process.* Chicago: University of Chicago Press.

Brown, D. (1991). Human Universals. New York: McGraw-Hill.

Butterworth, B. (Ed.). (1980). Language Production. London: Academic Press.

Cavalli-Sforza, L. (2000). Genes, People and Languages. New York: North Point Press.

- Cavalli-Sforza, L., & Feldman, M. W. (1981). *Cultural Transmission and Evolution: a Quantitative Approach*. Princeton: Princeton University Press.
- Campbell, D. T. (1965). Variation and selective retention in socio-cultural evolution. InH. Barringer, G. I. Blankstein, & R. W. Mack (Eds.), *Social change in developing areas.* Cambridge, MA: Schenkman.
- Cerri, G. (1986). *Scrivere e recitare : modelli di trasmissione del testo poetico nell'antichità e nel medioevo.* Roma: Edizioni dell'Ateneo.
- Chafe, W. (1994). *Discourse, Consciousness and Time*. Chicago: University of Chicago Press.
- Chen, W. (1932). *Xiucixue Fafan* 修辭學發及 (Introduction to Rhetoric). Shanghai: Shanghai Jiaoyu Publishing Co.
- Cheng, F. (1987). *La Poesia Tang* (original title: *L'écriture poétique chinoise*). (M.L. Barbella, Trans.). Napoli: Guida Editori.
- Cheng, J. (1985). Shi Jing Yi Zhu 詩經譯註(Shi Jing Translated and Annotated). Shanghai: Shanghai Guji Publishing Co.
- Child, F.J. (1965). *The English and Scottish Popular Ballads*. New York: Dover Publications Inc.

Chomsky, N. (1964). Current Issues in Linguistic Theory. The Hague: Mouton.

- Chomsky, N. (1966). *Cartesian Linguistics: a Chapter in the History of Rationalist Thought*. New York: Harper & Row.
- Cloak, F.T. (1975). Is a Cultural Ethology Possible? Human Ecology, 3, 161-82.
- Connerton, P. (1989). *How Societies Remember*. Cambridge: Cambridge University Press.

Cook, N. D. (2002). Tone of Voice and Mind. Amsterdam: John Benjamins Publishing Co.

- Christophe, A., Nespor, M., Guasti, M. T., & Van Ooyen, B. (2003). Prosodic Structure and Syntactic Acquisition: the Case of the Head-Direction Parameter. *Developmental Science*, *6*(2), 211-20.
- Crowder, R. G., Serafine, M. L., & Repp, B. H. (1990). Physical interaction and association by contiguity in memory for the words and melodies of songs. *Memory & Cognition, 18,* 469-476.
- Cutler, A. (1992). The Production and Perception of Word Boundaries. In Tohkura, Y.,
 Vatikiotis-Bateson, E., & Sagisaka, Y. (Eds). *Speech Perception, Production and Linguistic Structure*. Amsterdam: IOS Press.

Cutler, A. (Ed.). (1982). *Slips of the Tongue and Language Production*. Amsterdam: Mouton.

Dawkins, R. (1989). The Selfish Gene. Oxford: Oxford University Press.

Delamar, G. (1983). *Children's counting-out rhymes, fingerplays, jump-rope, and bounce-ball chants and other rhythms: a comprehensive English-language reference.* North Carolina: McFarland.

Dennett, D. C. (1991). Consciousness Explained. Boston: Little, Brown & Co.

Derrida, J. (1974). Of Grammatology. Baltimore: Johns Hopkins University Press.

Derrida, J. (1981). Positions. Chicago: University of Chicago Press.

Derrida, J. (1973). Speech and Phenomena. Evanston: Northwestern University Press.

- Edwards, V., & Sienkewicz, T. (1991) *Oral cultures past and present: rappin' and Homer.* Oxford: B. Blackwell.
- Erickson, F. (2003). Some Notes on the Musicality of Speech. In Tannen, D. & Alatis, J. (Eds.) *Linguistics, Language and the Real World: Discourse and Beyond:*

Georgetown University Round Table on Language and Linguistics. Washington, DC: Georgetown University Press.

- Erickson, F., & Schultz, J. (1982). *The counselor as gatekeeper: social interaction in interviews.* New York: Academic Press.
- Fadiga, L., Craighero, L., Buccino, G. & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *European Journal of Neuroscience*, 15, 399-402.

Feldman, M. & Laland, K. (1996). Gene-Culture Coevolutionary Theory. Tree, 11, 453-7.

- Foley, J. M. (Ed.). (1987). *Comparative research on oral traditions: a memorial for Milman Parry.* Columbus: Slavica.
- Foley, J. M. (1990). *Traditional Oral Epic: the Odyssey, Beowulf, and the Serbo-Croatian Return Song*. Berkeley: University of California Press.
- Fonagy, I. & Magdics, K. (1963). Emotional Patterns in Intonation and Music. *Zeitschrift für Phonetik Sprachwissenschaft und Kommunikationsforschung*, *16*, 293-326.

Foster, D. W. (1971). The Early Spanish Ballad. New York: Twayne Publishers Inc.

- Fowler, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics, 14,* 3-28.
- Fowler, D.C. (1968). *A Literary History of the Popular Ballad.* Durham, North Carolina: Duke University Press.

Fraisse, P. (1974). Psychologie du Rythme. Presses Universitaires de France.

Gathercole, S.E. & Baddeley, A.D. (1993). *Working memory and language*. Hove: Lawrence Erlbaum Associates.

Gladwell, M. (2000). The tipping point. Boston: Little, Brown & Co.

- Gottsch, J. D. (2001). Mutation, Selection and Vertical Transmission of Theistic Memes in Religious Canons. *Journal of Memetics – Evolutionary Models for Information Transmission, 5*. Retrieved on 10 April 2001 from http://www.cpm.mmu.ac.uk/ jom-emit/2001/vol5/gottsch_jd.html
- Guenther, F. H., C. Y. Espy-Wilson, S. E. Boyce, M. L. Matthies, M. Zandipour, & J. S.
 Perkell. (1999). Articulatory tradeoffs reduce acoustic variability during American
 English /r/ production. *Journal of the Acoustical Society of America, 105*(5), 2854-65.

Halbwachs, M. (1980). The Collective Memory. New York : Harper & Row.

- Hampe, Beate (Ed.). (2005). *From Perception to Meaning: Image Schemas in Cognitive Linguistics.* Berlin: Mouton de Gruyter.
- Hart, J. (1986). Declination has not been defeated—A reply to Lieberman et al. *The Journal of the Acoustical Society of America, 80*(6), 1838-9.
- Havelock, E.A. (1986). *The Muse Learns to Write: Reflections on Orality and Literacy from Antiquity to the Present.* New Haven: Yale University Press.
- Healey, P., Swoboda, N., & Umata, I. *The Role of Interaction in the Evolution of Human Symbol Systems.* (Unpublished manuscript.)
- Heath, Bell, & Sternberg. (2001). Emotional selection in memes: The case of urban legends. *Journal of Personality and Social Psychology*, *81*, 1028-1041.
- Hird, K. & Kirsner, K. (2002). The Relationship between Prosody and Breathing in Spontaneous Discourse. *Brain and Language, 80*, 536-555.
- Holdcroft, D. (1991). *Saussure: Signs, System, and Arbitrariness.* Cambridge: Cambridge University Press.
- Houde, J.F. & Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science*, *279*(5354), 1213-1216.

Hull, D. (1988). Science as a process. Chicago: University of Chicago Press.

Idema, W. (1997). *A Guide to Chinese Literature.* Ann Arbor: University of Michigan Press.

Jackendoff, R. (2002). Foundations of Language. Oxford: Oxford University Press.

- Jahandarie, K. (1999). *Spoken and Written Discourse: A Multi-Disciplinary Perspective*. Stamford, CN: Ablex Publishing Corp.
- Johnstone, B. (Ed). (1994). *Repetition in Discourse: Interdisciplinary Perspectives*. (Vols. 1-2). Norwood, NJ: Ablex Publishing Corp.
- Jones, J. A. & Munhall , K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *Journal of the Acoustical Society of America, 108,* 1246-1251.
- Jones, J. H. (1961). Commonplace and Memorization in the Oral Tradition of the English and Scottish Popular Ballads. *Journal of American Folklore*, *74*, 97-112.
- Jusczyk, P. (1997). *The Discovery of Spoken Language*. Cambridge, MA: Massachusetts Institute of Technology Press.

- Jusczyk, P. (2002). How Infants Adapt Speech-Processing Capacities to Native-Language Structure. *Current Directions in Psychological Science*, *11*(1), 15-18.
- Kiparsky, P. (1976). Oral Poetry: Some Linguistic and Typological Considerations. InStolz, B. A., & Shannon, R. S. (Eds). *Oral Literature and the Formula*. Ann Arbor:University of Michigan Press.
- Knapp, M. & Knapp, H. (1976) *One Potato, Two Potato*. New York: W.W. Norton & Co. Inc.
- Kuhl, P. K. & Miller, J. D. (1978). Speech perception by the chinchilla: Identification
 functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*,
 63, 905-917.
- Latane, B. (1997). Dynamic social impact: The creation of culture by communication. *Journal of Communication*, *46*(4), 13-25.
- Legge, J. (1865). *The Chinese classics with a translation, critical and exegetical notes, prolegomena, and copious indexes.* London: Clarendon Press.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: The MIT Press.

- Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*, 1-36.
- Liberman & Whalen. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4*(5), 187-196.
- Liberman, M. & Pierrehumbert, J. (1984). Intonational invariance under changes of pitch range and length. In Aronoff, M., & Oehrle, R.T. (Eds). *Language Sound Structure*. Cambridge, MA: The MIT Press.
- Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M. (1985a). Measures of the sentence intonation of read and spontaneous speech in American English. *Journal of the Acoustical Society of America, 77,* 649-657.
- Lieberman, P., Katz, W., Jongman, A., Zimmerman, R., & Miller, M. (1985b). Reply to Bruno H. Repp. *The Journal of the Acoustical Society of America, 78,* 1116-1117.

Lord, A. B. (1960). The Singer of Tales. Cambridge, MA: Harvard University Press.

- McLuhan, M. (2001). *Understanding Media: The Extensions of Man.* Cambridge, MA: MIT Press.
- Matthews, P.H. (2003). *Linguistics: A Very Short Introduction*. Oxford: Oxford University Press.

- Maturana, H. & Varela, F. (1987). *The Tree of Knowledge.* (Paolucci, R. Trans.). Boston: Shambhala Publications Inc.
- McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-748.

McNaughton, W. (1971). The Book of Songs. New York: Twayne Publishers.

- Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*, 75-78.
- Merritt, M. (1994). Repetition in Situated Discourse: Exploring its Forms and Functions.In Johnstone, B. (Ed). *Repetition in Discourse: Interdisciplinary Perspectives*.(Vol. 1.). Norwood, NJ: Ablex Publishing Corp.
- Metcalf, A. (2002). *Predicting New Words: The Secrets of Their Success*. Boston: Houghton Mifflin Co.

Mirante, N. (1994). *Le figure del discorso nelle teorie retoriche di Chen Wangdao* (Figures of Speech in Chen Wangdao's Theory of Rhetoric). Unpublished Masters thesis.

Mortara-Garavelli, B. (1988). Manuale di retorica. Milano: Gruppo Editoriale Fabbri.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language Discrimination by Newborns: Toward an Understanding of the Role of Rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766.

Nespor, M. & Vogel, I. (1986). Prosodic Phonology. Dordrecht: Foris.

Nettel, R. (1956). Seven Centuries of Popular Song. London: Phoenix House Ltd.

- Niles, J. (1999). *Homo Narrans : the poetics and anthropology of oral literature.* Philadelphia: University of Pennsylvania Press.
- Norenzayan, A., & Atran, S. (2004). Cognitive and emotional processes in the cultural transmission of natural and nonnatural beliefs. In M. Schaller & C. S. Crandall (Eds.), *The psychological foundations of culture*. Mahwah NJ: Erlbaum.
- Nygard, H.O. (1958). *The Ballad of* Heer Halewijn: *Its Forms and Variations in Western Europe.* Helsinki: Helsingin Liikekirjapaino Oy.

Ong, W. (1987). Orality and Literacy. New York: Methuen & Co.

Opland, J. (1980). Anglo-Saxon Oral Poetry. New Haven: Yale University Press.

- Parry, M. (1930). Studies in the Epic Technique of Oral Verse-Making: 1. Homer and Homeric Style. *Harvard Studies in Classical Philology*, *XLI*, 73-147.
- Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*, B35-B45.
- Payne, J.S. & Whitney, P.J. (2002). Developing L2 oral proficiency through sychronous
 CMC: Output, working memory, and interlanguage development. *CALICO Journal*, *20*(1), 7-32.
- Peradotto, J. Bakhtin, (n.d.) *Milman Parry and the Problem of Homeric Originality*. Retrieved on 1 Oct 2004 from http://www.acsu.buffalo.edu/~peradott/ BakhtinParry.pdf

Pinker, S. (1994). The Language Instinct. New York : W. Morrow and Co.

- Prentice, D. A., & Miller, D. T. (1996). Pluralistic ignorance and the perpetuation of social norms by unwitting actors. *Advances in experimental social psychology*, *28*, 161-209.
- Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of Linguistic Rhythm in the Speech Signal. *Cognition, 73,* 265-92.

- Repp, B. H. (1984). Categorical Perception: Issues, Methods, Findings. In N. J. Lass
 (Ed.). Speech and Language: Advances in Basic Research and Practice, Vol. 10.
 New York: Academic Press.
- Repp, B. H. (1985). Critique of 'Measures of the sentence intonation of read and spontaneous speech in American English' by Lieberman, Katz, Jongman, Zimmerman, and Miller. *The Journal of the Acoustical Society of America, 78*(3), 1114-1116.
- Repp, B. H. (1991). Some cognitive and perceptual aspects of speech and music. In J.Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech, and brain.*Basingstoke, U.K.: Macmillan.
- Repp, B. H. (2000). Introduction: Rhythm and Meter in Music and Speech. In Desain, P.
 & Windsor, L. *Rhythm Perception and Production*. Tokyo: Swets & Zeitlinger
 Publishers.
- Rizzolatti, G. & Arbib, M. A. (1998). Language within our grasp. *Trends In Neurosciences, 21*(5), 188-194.

Rogers, E. M. (1962). *Diffusion of innovations*. London: The Free Press.

Rosnow, R. L. & Fine, G. (1976). *Rumor and gossip: The social psychology of hearsay*. New York: Elsevier. Rubin, D. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting out rhymes.* Oxford: Oxford University Press.

Sampson, G. (1980). Schools of Linguistics. Stanford: Stanford University Press.

- Sams, M., Aulanko, R., Hamalainen, M., Mari, R., & Lounsamaa, O. V. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters, 127*, 141-45.
- Sapir, E. (1949). *Language: an Introduction to the Study of Speech*. New York: Harcourt Brace Jovanovich.

Saussure, F. de. (1959). Course in General Linguistics. New York: McGraw-Hill.

- Schaller, M. (2001). Unintended influence: Social evolutionary processes in the construction and change of culturally-shared beliefs. In J. P. Forgas & K. D.
 Williams (Eds.), *Social influence: Direct and indirect processes*. Philadelphia: Psychology Press.
- Schaller, M., Conway, L., & Tanchuk, T. (2002). Selective pressures on the once and future contents of ethnic stereotypes: Effects of the communicability of traits. *Journal of Personality and Social Psychology*, *82*, 861-877.

- Scherer, K., Banse, R., & Wallbott, H. (2001). Emotion Inferences from Vocal Expression Correlate Across Languages and Cultures. *Journal of Cross-Cultural Psychology*, *32*(1), 76-92.
- Schlüter, J. (2005). *Rhythmic grammar : the influence of rhythm on grammatical variation and change in English.* New York : Mouton de Gruyter.
- Scollon, R. (1981). The Rhythmic Integration of Ordinary Talk. In Tannen, D. (Ed.) Analyzing Discourse: Text and Talk. Georgetown University Round Table on Languages and Linguistics 1981. Washington, DC: Georgetown University Press.
- Serafine, M. L., Crowder, R.G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*, *16*(3), 285-303.
- Serafine, M. L., Davidson, J., Crowder, R. G., & Repp, B. H. (1986). On the nature of melody-text integration in memory for songs. *Journal of Memory and Language*, 25, 123-135.
- Shi, R., Werker, J. F., & Morgan, J. L. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition*, *72*, B11-B21.
- Shukla, M., Nespor, M., Mehler, J. *Grammar on a Language Map.* Unpublished manuscript.

- Showalter, E. (1997). *Hystories: Hysterical epidemics and modern culture.* New York: Columbia University Press.
- Sperber, D. (1990). The epidemiology of beliefs. In C. Fraser & G. Gaskell (Eds.), *The social psychological study of widespread beliefs*. Oxford: Clarendon.
- Stevens, K. N., & Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, *64*, 1358-1368.
- Tannen, D. (1984). *Conversational Style: Analyzing Talk among Friends*. Norwood, NJ: Ablex Publishing Company.
- Tannen, D. (1987). Repetition in Conversation: Towards a Poetics of Talk. *Language*, *63*(3), 574-605.
- Tannen, D. (1989). *Talking Voices: Repetition, Dialogue, and Imagery in Conversational Discourse*. Cambridge: Cambridge University Press.
- Umeda, N. (1980). F0 Declination is Situation Dependent. *Journal of the Acoustical Society of America, 68*(S1), S70.
- Vaneechoutte, M. (1997). Bird song as a possible cultural mechanism for speciation. *Journal of Memetics - Evolutionary Models of Information Transmission*, 1. Retrieved from http://jom-emit.cfpm.org/vol1/vaneechoutte_m.html

Vaneechoutte, M. & Skoyles, J. R. (1997). The memetic origin of language: modern humans as musical primates. *Journal of Memetics - Evolutionary Models of Information Transmission*, 1. Retrieved from http://jom-emit.cfpm.org/1998/ vol2/vaneechoutte_m&skoyles_jr.html

Visual Thesaurus. Accessed on 13 May 2005 at http://www.visualthesaurus.com/

Wang, C.H. (1974). The Bell and the Drum. Berkeley: University of California Press.

- Wang, L. (1980). *Shi Jing Yun Du* 詩經韻讀 (Reading the Rhymes in Shi Jing). Shanghai: Shanghai Guji Chubanshe.
- Waterson , N. (1987). *Prosodic Phonology: The Theory and Its Application to Language Acquisition and Speech Processing.* Newcastle Upon Thyne, UK: Grevatt & Grevatt.
- Wennerstrom, A. (2001). *The Music Of Everyday Speech: Prosody and Discourse Analysis*. Oxford: Oxford University Press.
- Werker, J. F. & Tees, R. C. (1999). Experiential influences on infant speech processing:
 Toward a new synthesis. In Spence, J. T. (Ed.), Darley, J. M., & Foss, D. J.
 (Assoc. Eds.). Annual Review of Psychology, 50. Palo Alto: Annual Reviews.

Wheeldon, L. (Ed.) (2000). Aspects of Language Production. Hove: Psychology Press.

Wheeler, L. (1966). Toward a theory of behavioral contagion. *Psychological Review*, *73*, 179-192.

Wittgenstein, L. (1958). Philosophical investigations. Oxford: B. Blackwell.

WordNet. Accessed on 13 May 2005 at http://wordnet.princeton.edu/

- Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.
- Wray, A. & Perkins, M. (2002). The Functions of Formulaic Language: An Integrated Model. *Language & Communication, 20,* 1-28.

Appendix One

The First 250 Most Common English Words

The list below was copied from the website "1000 Most Common Vocabulary Words in English" (http://esl.about.com/library/vocabulary/bl1000_list1.htm) and reformatted. The full list of one thousand words is available online.

Rank	Word	Rank	Word	Rank	Word
1	the	17	As	33	what
2	of	18	Ι	34	some
3	to	19	His	35	we
4	and	20	they	36	can
5	а	21	Ве	37	out
6	in	22	At	38	other
7	is	23	One	39	were
8	it	24	have	40	all
9	you	25	this	41	there
10	that	26	from	42	when
11	he	27	Or	43	up
12	was	28	Had	44	use
13	for	29	Ву	45	your
14	on	30	Hot	46	how
15	are	31	word	47	said
		1			

Rank	Word	Rank	Word	Rank	Word
16	with	32	But	48	an
49	each	72	Him	95	may
50	she	73	two	96	down
51	which	74	Has	97	side
52	do	75	look	98	been
53	their	76	more	99	now
54	time	77	Day	100	find
55	if	78	could	101	any
56	will	79	Go	102	new
57	way	80	come	103	work
58	about	81	Did	104	part
59	many	82	number	105	take
60	then	83	sound	106	get
61	them	84	No	107	place
62	write	. 85	most	108	made
63	would	86	people	109	live
64	like	87	Му	110	where
65	SO	88	over	111	after
66	these	89	know	112	back
67	her	90	water	113	little
68	long	91	than	114	only
69	make	92	call	115	round
70	thing	93	first	116	man
71	see	94	who	117	year
		-			

Rank	Word	Rank	Word	Rank	Word
118	came	141	much	164	hand
119	show	142	mean	165	port
120	every	143	before	166	large
121	good	144	move	167	spell
122	me	145	right	168	add
123	give	146	Воу	169	even
124	our	147	Old	170	land
125	under	148	Тоо	171	here
126	name	149	same	172	must
127	very	150	Tell	173	big
128	through	151	does	174	high
129	just	152	Set	175	such
130	form	153	three	176	follow
131	sentence	154	want	177	act
132	great	155	Air	178	why
133	think	156	well	179	ask
134	say	157	also	180	men
135	help	158	play	181	change
136	low	159	small	182	went
137	line	160	End	183	light
138	differ	161	Put	184	kind
139	turn	162	home	185	off
140	cause	163	read	186	need
	1	l i		1	
Rank	Word	Rank	Word	Rank	Word
------	---------	------	---------	------	-------
187	house	209	school	231	hard
188	picture	210	grow	232	start
189	try	211	study	233	might
190	us	212	still	234	story
191	again	213	learn	235	saw
192	animal	214	plant	236	far
193	point	215	cover	237	sea
194	mother	216	food	238	draw
195	world	217	Sun	239	left
196	near	218	four	240	late
197	build	219	between	241	run
198	self	220	state	242	don't
199	earth	221	keep	243	while
200	father	222	Eye	244	press
201	head	223	never	245	close
202	stand	224	last	246	night
203	own	225	Let	247	real
204	page	226	thought	248	life
205	should	227	city	249	few
206	country	228	tree	250	north
207	found	229	cross		
208	answer	230	farm		

Appendix Two

A Look at the Future

The constraints argument explored in this work is generally in tune with the findings of many studies in cultural dynamics, anthropology and social psychology. I believe that the small picture of face-to-face dialogue must ultimately coincide with the larger picture that describes communication and information transmission across entire populations. The small picture is included in the larger picture, and the mechanisms of the former must support the trends and patterns that are observable in the latter. For this reason, I believe that the constraints argument has the potential to enrich and expand the ways in which communication studies are currently understood.

The constraints argument also has the potential of bridging the divide between several schools of linguistics, by proposing a framework in which their lines of thought can all have a place. Recognising syntax as an active constraint of speech means that some of the findings of Generative linguists, and the importance they place on reasoning and rule-based behaviour, are not overlooked. However, the fact that syntax is only one of six constraints also places some significant limitations on the formal outcomes of the speaking process. Working in cooperation with the rule-based reasoning of syntactic processing, the remaining five constraints bind the linguistic output of a speaker to the data that is currently present in his/her working memory. In this way, context also becomes an active factor in constructing linguistic output, as Functionalists often argue. Because of these mechanisms, the potentially innovative strength of syntactic reasoning

remains confined. This is why everyday communication adheres to largely recursive patterns despite the highly developed reasoning skills of speakers. The push for innovation on the one hand, and the opposite and generally prevailing push for conservatism on the other, are both necessary in explaining the behaviour of everyday speakers. Achieving their synthesis in a coherent theoretical framework is a priority in understanding the dynamics of language and communication.

Furthermore, the constraints argument can be applied to studies on rhetoric, genre, and culture. It will be very interesting to investigate various types and places of discourse in terms of the mnemonic constraints to which their users are subjected. While I have given a short overview of some of my observations derived from studies in cultural transmission, I am well aware that my efforts in this sense have not been vast and that more comprehensive and more detailed analyses are both possible and desirable. The examination of trends in language use could be enhanced with a deeper understanding of the processes involved in the construction of utterances. A variety of spoken and written domains of language production can supply excellent testing grounds for the dynamics of linguistic composition I have illustrated.

In addition, the constraints argument can be applied to studies on media and mediated linguistic communication. For example, if prosody and rhythm are indeed somewhat important in the production and re-production of written texts, then it might be interesting to research those media that do, or do not, facilitate the use of these constraints. Let us consider, for example, the widespread use of text messaging on cellular phones. Even more so than email, these instances of writing are used as reproductions of speech: grammar and spelling are often mangled, common formulas are abbreviated to minimalist terms and yet instantly recognised, repetitions abound. Is

this written language, or speech? Could it be that today's media are evolving towards facilitating the constraints of spoken language in the production of written language? How do different media fare in dealing with the speech constraints of different languages? These are only a few of the many questions that are applicable to what I envision as a new direction in media studies.

Finally, I believe that the constraints argument could be applied with some success to studies in speech therapy and in machine-human interaction. As far as speech therapy is concerned, it might be productive to investigate whether certain kinds of speech problems are related to a failure in using one or more of the constraints of production. For example, music therapy has already proven helpful in developing the language skills of autistic individuals, as well as of people with Parkinson's and Alzheimer's. It is possible that music therapy accesses resources that are normally used both for music appreciation and for the application of prosodic and rhythmic constraints. As Cook demonstrated, many areas of the brain are indeed used to process the melody and rhythm of both music and prosody, as well as in decoding their affective content. Therefore, it is possible that music therapy provides a jump-start for the processing of rhythmic and prosodic constraints, which in turn facilitates the processing of other constraints and ultimately results in enhanced production. While I presently do not have the knowledge to determine the validity of my hypothesis, I do believe that it is worthwhile to at least pose the question.

In discussing machine-human interaction, I refer to the many attempts that have been carried out in recent times to produce an electronic system capable of assembling natural speech. Looking at these attempts from the point of view of the constraints argument, it is rather obvious that they are doomed to failure. The shortcomings are

evident because only syntax and semantics have been taken into consideration in the design of these systems. First of all, the systems are developed in isolation; they are not taught to converse, but to produce, as if they were "one-way interactors" – a logical impossibility. It seems to me that the first requisite in envisioning a successful system is to make it a two-entity system. Perhaps one of the two entities can be a human, rather than a machine. However, the human counterpart needs to engage in a long history of interactions with the machine, like a trainer. The machine must be taught to borrow words and formulas from its trainer. It must be taught to make lexical association both semantically and acoustically. And, what is most difficult, it must be taught to recognise and reproduce prosodic patterns and rhythms. It must be given a vocal range, a baseline, and a "lung capacity" that will determine the length of its intonation units. Finally, it must be taught to recognise major and minor intonations in prosody, in order to recognise the affective disposition of its interlocutor and match its output in terms of a psychologically adequate response.

I conclude with a long wish-list of possible research directions because I hope to expand my knowledge and my research challenges to these many interrelated domains. In the immediate future, I hope to collect a corpus of data on speech constraints by conducting the experiments I illustrate in the next section. Certainly, my theory of combined constraints will change and become much more refined as a result of additional work and empirical testing. However, I expect and hope that this is simply the beginning of an original and productive line of enquiry.

1. Next Steps

The research efforts that have allowed me to put together my theoretical framework have been substantial, and yet this is only the start. Theories need to be grounded in data; although here I have attempted to use concrete examples as much as possible, I am aware of the fact that my corpus is much too small. To continue down this path, I would like to conduct some additional discourse analyses in order to collect more evidence on speech constraints.

The most common approach in discourse analysis is the non-invasive one, by which conversations are recorded with the list amount of intrusion in order to preserve the naturalness of the speech event. For at least some of my analyses, I would like to try a different route. I wish to set up some experimental situations in which one or more of the constraints will be slightly distorted, because I think this will make it easier to observe the constraints in action. Following are some of the experiments I hope to be able to perform in the future.

a. THE TREADMILL

Speakers will be recorded in a control situation, before the start of the experimental condition. During the experiment, which will last five minutes, speakers will be recorded while walking on a treadmill. The pace of the treadmill will be adjusted so that subjects will be required to walk quite quickly. The recordings obtained in the pre-experimental and experimental conditions will be analysed and compared.

The purpose of the experimental condition is to force the subject to breathe more frequently and therefore to shorten his/her intonation units. My prediction is that when

comparing the experimental condition to the control one, it will be possible to detect acceleration in the rhythm of the utterances, as well as simpler syntactic structures, more fragmentation in syntactic structures and possibly a more semantically disjointed discourse, and an increase in the use of formulas and collocations.

b. THE METRONOME

Speakers will be recorded in a control situation, before the start of the experimental condition. During the experiment, which will last five minutes, speakers will be recorded while a metronome clicks in the background. The pace of the metronome will be set so that it is relatively quick. The recordings obtained in the pre-experimental and experimental conditions will be analysed and compared.

The metronome is expected to influence the rhythm of the speaker's utterances. My prediction is that the rhythm will accelerate, but that other qualities of the speaker's discourse will remain relatively unchanged.

C. THE ECHO AND SYNTAX LAB

Speakers will be recorded in a control situation, before the start of the experimental condition. During the first experimental phase, which will last five minutes, speakers will be required to listen to a recording of a single speaking voice. The passage produced by the voice will contain a high number of words starting with or containing the same consonant sound or syllable (for example, "m", "ike", "k"). During the second phase, speakers will be given a topic completely unrelated to that of the passage they listened to, and will be required to produce five minutes of speech. The recordings

obtained in the pre-experimental and experimental conditions will be analysed and compared.

The prediction is that speakers will produce the target sound or sounds in a proportion significantly higher than that of the control situation. The analysis will also check for repetitions of syntactic structures, formulas and lexemes between the listened and the produced passages.

d. The Melody and Syntax Lab

This experiment will be quite similar to the preceding one on echoes. However, this time, the speakers will listen to a passage that contains a particularly high number of phrases produced with a specific intonation contour. For example, one interesting and not too frequent pattern is that of the English contour for surprise/outrage, which includes a high accent peak and a low offset (e.g., in "you're JOking!", "I would NEVER do that!" or similar). The prediction is that the speaker will produce the intonation contour in the experimental condition in a higher proportion than in the control condition. Moreover, since it is quite possible that the contour could be preferentially coupled with a specific syntactic pattern (as in the examples above), the syntax of the produced speech will also be analysed for repetitions.