Gerard O'Grady

A Grammar of Spoken English Discourse

The Intonation of Increments

continuum

A Grammar of Spoken English Discourse

Continuum Studies in Theoretical Linguistics

Continuum Studies in Theoretical Linguistics publishes work at the forefront of present-day developments in the field. The series is open to studies from all branches of theoretical linguistics and to the full range of theoretical frameworks. Titles in the series present original research that makes a new and significant contribution and are aimed primarily at scholars in the field, but are clear and accessible, making them useful also to students, to new researchers and to scholars in related disciplines.

Series Editor: Siobhan Chapman, Reader in English, University of Liverpool, UK.

Other titles in the series:

Agreement Relations Unified, Hamid Ouali
Deviational Syntactic Structures, Hans Götzsche
First Language Acquisition in Spanish, Gilda Socarras
A Neural Network Model of Lexical Organisation, Michael Fortescue
The Syntax and Semantics of Discourse Markers, Miriam Urgelles-Coll

A Grammar of Spoken English Discourse

The Intonation of Increments

Gerard O'Grady

Continuum Studies in Theoretical Linguistics

continuum

Contents

List of Figures

List of Tables

Acknowledgements

Transcription Symbols

Intonation

/	Rising tone
\	Falling tone
∨	Falling-Rising tone
∧	Rising-Falling tone
–	Level tone
↑WORD	High-Key
↓WORD	Low-Key
↑<u>WORD</u>	High-Termination
↓<u>WORD</u>	Low-Termination
<u>WORD</u>	Tonic word: word containing major tone movement in tone unit
//	Tone unit boundary
. . .	Incomplete Tone Unit

When discussing Brazil's work the following alternate intonation conventions are used:

p	proclaiming/falling tone
p+	proclaiming/falling-rising tone dominant
r	referring/falling-rising tone
r+	referring/rising tone dominant
o	o/level tone

Grammar

N	Nominal element
V	Verbal element
V'	Non-finite verbal element
A	Adverbial element
E	Adjectival element

W	Open selector
CON	Convention
P	Preposition
PHR	Phrase: series of elements treated as a single lexical selection
NUM	Numeral
VOC	Vocative
d	Determiner
d°	Determiner with zero realisation
c	Conjunction
Ø	Element or elements which are unrealized
ex	Exclamation
n	Suspensive nominal element
v	Suspensive verbal element
v'	Suspensive non-finite verbal element
a	Suspensive adverbial element
e	Suspensive adjectival element
w	Suspensive open selector
con	Suspensive convention
p	Suspensive preposition
phr	Suspensive phrase
num	Suspensive numeral
voc	Suspensive vocative
+	Reduplication
#	End of increment
(N)	Bracketed element(s): element(s) did not lead to the realization of a new intermediate state
…	Abandoned increment

This page intentionally left blank

Part I

Setting the Scene

This page intentionally left blank

Chapter 1

Introduction: The Organization
of Spoken Discourse

In 1995, David Brazil published *A Grammar of Speech* which he described as an exploratory grammar and claimed that:

> An exploratory grammar is useful if one is seeking possible explanations of some of the many still unaccounted for observations one may make about the way the language works. It accepts uncertainty as a fact of the linguist's life. Its starting-point can be captured in the phrase 'Let's assume that . . .' and it proceeds in the awareness that any assumptions it makes are based on nothing more than assumptions; the aim is to test these assumptions against observable facts. (1995: 1)

Due to Brazil's untimely death, he was unable to continue his exploration past the point reached in Brazil (1995) namely the testing of his grammar against a small monologic corpus: a retelling of a short urban myth to a listener who had not previously heard the story by a speaker who had him/herself only heard the story shortly before it was retold.[1] This book sets out to update the exploration in two ways. The first, an 'inward' exploration, critically examines the premises on which Brazil's grammar rests and attempts to link these assumptions to the wider literature. The second, an 'outward' exploration, tests the grammar against different data, and seeks possible explanations for a range of attested linguistic behaviour not accounted for by Brazil. Unlike Brazil (1995) this book explicitly considers the role of intonation in helping to segment a stretch of speech into meaningful utterances and in projecting the unity of the segmented unit of speech.

Conversation Analysts e.g. Sacks (1995) and Schegloff (2007), like Brazil recognize that there is a structure and design in spoken discourse. Their famous 'no gap no overlap' model of conversation, centred on the smooth transition of turn-taking, is premised upon the belief that cooperative

interlocutors are so tuned into the discourse that they can effortlessly produce a seamless flow of smooth, pause-free conversation. The studies presented in Couper-Kuhlen and Selting (1996) illustrate clearly how interlocutors utilize intonation and rhythm to manage their conversational contributions by signalling their intention to either maintain or relinquish the floor resulting in a smooth flow of conversational discourse. Yet, by focusing exclusively on turns and potential turns much of the structure and design of spoken discourse is overlooked. This book building on Brazil (1995) aims to describe how speakers design and structure their discourse to suit their own individual conversational needs and not just how they manage the conversational floor.

Since the publication of Brazil (1995) two very influential phonological theories have emerged: Optimality Theory (Prince and Smolensky 2004), and the Tone and Break Index (ToBI) description of intonation based on the autosegmental-metrical model of intonation developed by Pierrehumbert (1980). Much work in Optimality Theory (OT) has focused on tonality and OT theorists have shown how language specific morpho-syntactic structure and information focus interact with universal constraints to create language specific tonality divisions (Gussenhoven 2004: chapter 8). Yet, OT as a theory with generative underpinnings has not involved itself with real language data and is therefore incapable of describing the structure and design of an utterance produced to satisfy a specific communicative need.

Beckman, Hirschberg and Shattuck-Hufnagel (2005) is a revealing account of the motivations which lead to the development of the ToBI transcription system. They remind us that ToBI emerged from a series of interdisciplinary workshops which aimed to create a standard set of conventions for annotating spoken corpora. The standardization of conventions was required for a broad set of uses in the speech sciences such as the development of better automatic speech recognition systems and the creation of speech generation systems (ibid. 10–12). While ToBI is a phonological theory and notates meaningful intonational differences it does not annotate any unit of speech larger than the Intonational Phrase or tone unit. This is undoubtedly because the tone unit is the largest stretch of speech which can be unambiguously defined by phonology alone.[2] Scholars working within the ToBI framework have not concerned themselves with the self-evident fact that humans produce speech in order to achieve a purpose and as a result have not attempted to find regularity in the interaction between the phonology, the grammar and the semantics. Consequently ToBI, like OT descriptions of speech, focuses on the form of utterances rather than on their function and ignores many of the means

speakers employ to structure their utterances in the pursuit of their individual communicative purposes. Brazil's grammar is capable of describing the organization of discourse precisely because it looks for regularity in how the lexicogrammar, the phonology and the context combine to create and structure meaning.

Brazil's grammar rests on four premises, which will be examined and situated within the literature. The four premises are (1) speech is purposeful, (2) speech is interactive, (3) speech is cooperative, and (4) the communicative value of a lexical item is negotiated as the discourse unfolds. For the moment, I will presume that Brazil's premises are well-founded and will instead turn my attention to describing his claim that what he dubs *used language* can be described as a sequence of word-like elements which move from an *initial state* to a *target state*. Brazil (ibid. 48) defines initial state as speakers' perceptions, prior to performing the utterance, of what needs to be told either by themselves to their hearers or by their hearers to themselves, while target state is defined as the modified set of circumstances which have arisen after the telling. The stretch of speech which completes the telling, by moving from initial to target state, is the *increment*. Chapter 1 details the two criteria – one grammatical, the other intonational – which Brazil employed to identify increments. Without, at this point, getting bogged down in the details of how to identify an increment, it is sufficient to propose that an increment is a unit which tells something relevant to the speakers' or the hearers' present informational needs.

The following paragraphs continue the inward exploration of the grammar by sketching a possible model of language processing and arguing that if the model and the assumptions upon which it rests are correct, increments are vital intermediate processing units which bridge the tone/information unit and the achievement of a speaker's ultimate communicative intention. Without speaker/hearer recognition of the achievement of a target state, speakers would be less able to achieve their ultimate communicative intentions.

Increments which consist of a chain of word-like elements simultaneously consist of a chain of tone units. The data studied here consists of eleven readers reproducing two short political monologues unimaginatively labelled as Text 1 and Text 2 – see Chapter 5 for a full description of the corpus. In Text 1, the smallest number of complete tone units found in an increment was 1, the largest 14, and the mean 3.96. The smallest number of complete tone units found in an increment in Text 2 was 1, the largest 10 with a mean of 2.76.[3] Thus, in the corpus studied here an increment was a unit of speech which completed a telling and was on average between 3 and 4

tone units long. Before proceeding with the outward exploration of the grammar it is first necessary to demonstrate that a grammar grounded in increments and not in clauses[4] is a useful way of segmenting and describing the speech signal. The decision to segment the continuous speech signal into discrete units reflects an ideological stance and necessarily imposes a non-neutral perspective on how an act of communication is viewed. To illustrate, adoption of the clause as the unit which primarily generates meaning in a hierarchical grammar such as that proposed by Halliday and Matthiessen (2004) results in a view of language as a series of Matryoshka dolls with smaller units nesting inside larger ones. The usefulness and power of such an approach has been repeatedly demonstrated and this raises the question of why anyone would wish to look at language from a different perspective. This book attempts to demonstrate that looking at language as a process or discourse, and not as a product or text aids the overall explication of the meaning potential of the language.

If speech is viewed as a series of increments it must also be seen as a concatenation of tone units. Halliday and Matthiessen (2004: 88) argue that every tone unit[5] realizes a quantum or unit of information in the discourse and that 'spoken English unfolds as a sequence of information units, typically one following after another in unbroken succession'. Chafe (1994: 66) similarly argues that every intonation unit realizes a single new idea and that speakers build up their discourse idea by idea or, in other words, intonation unit by intonation unit. As a preliminary statement it can be postulated that speakers move from initial to target state by producing a sequence of tone units.

Such a preliminary statement raises two questions: is there evidence in the literature for the unitary nature of the tone unit as a unit of language processing, and even if tone units are units of language processing, is it feasible that an act of telling could be produced tone unit by tone unit? The next paragraph evaluates evidence which supports the view that the tone unit represents a pre-assembled information unit[6] which is inserted into the discourse as a single unit.

As seen above, linguists such as Halliday and Chafe argue that tone units realize a single quantum of information. Laver (1970: 68) offers psycholinguistic support by arguing that the tone unit is a pre-assembled stretch of speech, while Boomer and Laver (1968: 8) claim that evidence from speech errors provides good evidence in support of the view that tone units are handled as a unitary behavioural act by the central nervous system. If this view is correct,[7] then the increment can usefully be described as

a string of information units which move the discourse from an initial to a target state.

The second question is whether it is psychologically realistic to describe an act of telling as a concatenation of tone units which form increments. The work of Levelt (1989) suggests a possible mechanism which may allow us to realistically describe the satisfaction of a communicative intention as a concatenation of one or more tone units which achieve target state. He argues (ibid. 109) that, in order to satisfy their communicative needs, speakers 'microplan' and 'macroplan' the content of their utterances. He defines microplanning as the assigning of information structure within the discourse,[8] and macroplanning as the sum total of all the activities which speakers use to satisfy their individual communicative intentions; speakers macroplan in order to achieve target state and realize their communicative intentions. Thus, it seems feasible to argue that, prior to speaking, speakers set a target which they realize by producing a chain of tone units which form an increment. Calvin (1998: 120) reminds us that working memory is rather limited and that the average person can only hold onto a maximum of nine separate chunks of information at any one time. Thus, if increments are formed out of preassembled chunks we would not expect to find increments of larger than 9 tone units. In the data studied, the mean size of an increment was 3.96 and 2.76 tone units in texts 1 and 2 respectively, well within the capacity of working memory.

Levelt's definition of macroplanning is wider than the planning of an increment. It is easy to imagine communicative intentions, such as the desire of a politician to convince an audience to vote them into power, which could hardly be satisfied by the production of a single increment. Speakers who need to produce more than one increment[9] to satisfy their communicative intentions, are clearly able to do so without any apparent difficulties caused by the attested limitation in the storage capacity of working memory. Levelt (ibid. 109) recognizes that the 'journey from message to intention' often requires more than one step or, in the terminology used here, increment. Accordingly, he argues that speakers realize their goals by producing a series of sub-goals. At the same time, he acknowledges that a major task of a speaker, while constructing a message, is to keep track of what is happening in the discourse. It is proposed here that the increment, by realizing a target state, enables the speaker to successfully achieve a sub-goal and move a step closer to the achievement of the overall communicative goal. Increments produce a target state which is simultaneously the initial state of the immediately following increment

and this concurrent target/initial state allows the speaker to dump the previous increment from working memory in order to make space for the following one without losing track of what has gone before. Thus, it seems that increments may function to: (1) satisfy the speaker's communicative intention; or (2) produce a target/initial state which allows speakers to progress towards the satisfaction of their communicative intentions while keeping track of what is happening in the discourse.

To summarize the preceding paragraphs, an information unit realized phonologically as a tone unit is a preassembled chunk which joins with other tone units to form an increment. A telling increment may satisfy the speaker's communicative intention but if it does not, it results in the creation of a new initial state which speakers use as a springboard to realize their ultimate telling, i.e. the modification in the existing state of speaker/hearer understanding required to achieve their purpose and generate – if appropriate – the desired perlocutionary response.

Much recent linguistic theory, e.g. Sinclair (1991: 110), Wray (2002: 18), persuasively argues that language is, at least partly, formed out of chunks larger than orthographic words and so the outward exploration of the grammar must attempt to encode increments, where possible, as chains comprised not only of orthographic words but also of what we informally label here as chunks. Brazil coded his chains as strings of verbal, nominal, adverbial and adjectival orthographic words but did so with the express proviso that such labelling is no more than 'a temporary expedient' (1995: 43). Similarly, we code the lexical elements which occur in increments in traditional terms but keep an open mind as to whether it may become necessary to abandon traditional classification in order to provide a psychologically more realistic coding of how humans assemble speech. It is clearly true that the categorization of language into nouns and verbs is descriptively useful. Even a scholar such as Elman (1990), who argues against the existence of mental concepts such as nouns and verbs, found it necessary to describe his findings in terms of nouns and verbs. For the moment, there appears to be no other way to describe accurately a concatenation of lexical elements other than by using the traditional codings.[10] Yet it also appears sensible not to attempt to decompose each and every functional lexical element, e.g. idioms, into strings of orthographic words (Thibault 1996: 257–8).

The remainder of the book comprises seven further chapters: the following three are theoretical and represent the inward exploration of the grammar. Chapter 2 describes the formal mechanism of Brazil's grammar of speech and suggests ways in which the grammar can be expanded.

In Chapter 3 we examine the theoretical underpinnings on which Brazil's grammar rests. Some difficulties, chiefly with Brazil's view of *shared knowledge* and how this is projected by tone selections, are highlighted and revisions are offered. Chapter 4 explores the feasibility of encoding speech in a linear grammar and critically examines how to notate lexical elements in the grammar. Chapters 5 to 7 represent the outward exploration of the grammar. Chapter 5 describes the corpus used to test the grammar and details the notation system employed. Chapters 6 and 7 test the grammar against the corpus. The arguments presented in the book are concluded in Chapter 8 which also sets out further areas where the grammar needs to be developed.

This page intentionally left blank

Part II

The Outward Exploration of the Grammar

This page intentionally left blank

Chapter 2

A Review of *A Grammar of Speech*

This chapter, drawing from Brazil's exploratory article *Intonation and the grammar of speech* (1987) and his book *A Grammar of Speech* (1995), summarizes his theory of a linear grammar of spoken English. It will be seen that Brazil's grammar rests upon four premises. In this chapter, only Brazil's first premise is described in detail because the remaining three premises are best described and evaluated after a review of the wider literature which is presented in Chapter 3. Once the theory has been described omissions which are explicitly mentioned by Brazil as worthy of future exploration but not yet incorporated in the grammar, are considered in order to generate proposals suggesting how the grammatical description of speech might be expanded. It is hoped that the incorporation of these omissions will allow the grammar to further describe how speakers employ their grammatical resources to satisfy their communicative needs.

2.1 Starting Premises[1]

The grammar proposed by Brazil aims to describe the observable fact that, in real time communication, speech unfolds word by word. He does not attempt to describe how language is generated or processed in the mind. Brazil (1987: 146–8) postulates five premises on which he bases his theory. However, in line with Brazil (ibid. 26–36) I have combined premises 4 – *talk takes place in real time* – and 5 – *speakers exploit the here and now values of the linguistic choices they make* – into one premise – *existential values*.

  The first premise is that *speakers speak in pursuit of a purpose*; they are not concerned with whether or not their utterances obey de-contextualized abstract syntactic rules but rather with whether or not their speech is able to contribute to the successful management of their affairs. Linguistic competence consists of the ability to engage in the communicative events

with which speakers are faced from time to time (p. 9).[2] Brazil labels such communicatively engaged language as *used language* and defines it as:

> language which has occurred under circumstances in which the speaker was known to be doing something more than demonstrate the way the system works. (p. 24)

Used language, according to Brazil, can be analysed in terms of abstract syntactic constraints, but he claims that such an analysis is an additional fact which arises from the post-hoc examination of an utterance no longer serving any communicative purpose. Such an analysis, he argues, is an acquired skill not required by speakers engaged in successful communication. A grammar which aims to describe the observed workings of speech need not, he claims, concern itself with explicating the inherent possibilities of the language system (p. 16). Traditional approaches to grammar have focused on the workings of formal decontextualized abstract sentences and have assigned the study of how speakers employ sentences to satisfy their communicative needs to the discipline of pragmatics. Competence, according to these traditional views, is independent of and prior to use. Brazil's grammar, unlike traditional grammars, does not draw a distinction between form and use. An utterance, according to Brazil, is ill-formed if it is incapable of satisfying the speaker's communicative needs, regardless of whether or not it breaches formal rules.

A grammar which does not distinguish between form and function is uninterested in any formal classification of sentences into formal categories, i.e. imperative, interrogative, and declarative. Instead it classifies language functionally. Brazil proposed that while there are numerous ways of describing the purpose of any particular utterance, speakers realize their individual communicative purposes either by telling or asking (pp. 27–8). For example, a speaker can warn a hearer planning to go hiking by producing an indicative clause: *Bears have been seen at the bottom of the mountains* or *Watch out for the bears* or an interrogative clause *Have you heard the reports of the bears at the bottom of the mountains?* Brazil's claim is that the mechanisms employed by speakers can be divided into *telling* and *asking exchanges* which speakers employ to fulfil their communicative purposes. Such exchanges are defined as follows:

Telling exchanges: Tellers simultaneously initiate and achieve their purpose; the hearer may (or may not) then acknowledge the achievement.