English Speech Rhythm

Elizabeth Couper-Kuhlen

ENGLISH SPEECH RHYTHM

FORM AND FUNCTION IN EVERYDAY VERBAL INTERACTION

ELIZABETH COUPER-KUHLEN

JOHN BENJAMINS PUBLISHING COMPANY
AMSTERDAM/PHILADELPHIA

1993

Library of Congress Cataloging-in-Publication Data

Couper-Kuhlen, Elizabeth.
English speech rhythm : form and function in everyday verbal interaction / Elizabeth Couper-Kuhlen.
p. cm. -- (Pragmatics & beyond, ISSN 0922-842X ; new ser. 25)
Includes bibliographical references and index.
1. English language--Rhythm. 2. English language--spoken English. I. Title. II. Series.
PE1591.C68 1993
421'.6--dc20 93-6530
ISBN 90 272 5037 5 (Eur.)/1-55619-293-2 (US) (alk. paper) CIP

John Benjamins Publishing Co. · P.O. Box 75577 · 1070 AN Amsterdam · The Netherlands
John Benjamins North America · 821 Bethlehem Pike · Philadelphia, PA 19118 · USA

TABLE OF CONTENTS

TABLE OF FIGURES

INTRODUCTION

This work begins with a caveat for the musically minded reader. Any experienced musician will know that there is a difference between *meter* and *rhythm*. In the words of one recent musical scholar:

> The meter is a fixed and steady abstract ([..] "conceptual") norm against which the rhythm, the constantly changing acceleration and slowing, syncopation, anticipation, shift of stress, and so on, is counterpointed. (Childs 1981:36)

According to this distinction, the present work should be entitled *English Speech Meter*, since it is concerned with isochrony or the presence of a regular 'beat' in spoken language.

Yet there is a long tradition in the language and speech sciences of so-called *rhythm* studies, in which the patterns of prominence in *prose* are compared and contrasted with those created by meter in *verse*. That is, in linguistic usage *meter* has until recently been reserved for the temporal or structural organization of prominences in verse, *rhythm* for the same organization in non-verse or prose. The present work is aligned in the tradition of earlier prose rhythm studies, although the prose it is concerned with is that of spontaneous spoken, not pre-planned written English.

Recent prosodic theory in generative grammar has adopted terms such as *meter, metrical grid, metrical tree,* etc. to deal with the structural and/or temporal organization of prominence in any kind of (idealized) language use, whether verse or non-verse. But, as so often happens with the terminology of highly elaborate, explicitly articulated schools of thought, this use does not transfer well. To employ the term *meter* for speech rhythm here would be perceived as a pledge of allegiance to the generative enterprise. Yet the present study, although it is not incompatible with metrical phonology (see Chapter III), takes on the whole a different approach. 'Speech meter' would risk even greater confusion among potential readers than 'speech rhythm'.

As its subtitle suggests, the book falls roughly into two parts, one dealing with phonetic and phonological aspects of English speech rhythm, the other with functional questions related to its occurrence in everyday verbal

interaction. Chapter I outlines briefly what has become known as the isochrony debate and discusses the recent discovery of P-centers, which promises to settle much of the controversy. Chapter II examines speech rhythm in a fragment of English conversation and develops an auditory-acoustic method for discovering and verifying the presence of perceptual isochrony in speech. Chapter III relates the present approach to current models of metrical and prosodic phonology and situates speech rhythm within a larger cognitive framework by comparing and contrasting it with music and verse. Chapter IV exposes three weaknesses in the treatment of time and timing in Sacks *et al*'s model of turn-taking in conversation and advocates, instead of a metric based on absolute duration, a rhythmic view of timing. The options which a rhythm-based metric would provide are spelled out and empirical evidence is adduced on a small scale to support such a view.

With Chapter V the focus shifts to function. After a brief review of socio-psychological and ethnographic accounts of behavioural synchrony and interactional rhythm in speech encounters, three questions raised by the most promising of these accounts, interactional rhythm as contextualization cue, are discussed. The remainder of the chapter follows up the question of distribution of isochronous rhythm in interaction empirically by attempting to validate a hypothesis based on transition type in a large corpus of spoken English. Chapters VI and VII are devoted to situated interpretations of speech rhythm configurations at sequence-external and sequence-internal junctures respectively in the corpus conversations. Chapter VIII looks at speech rhythm contextualization as a function of activity type, specifically of questioning vs. repairing in everyday talk.

The contribution of this work is intended to be threefold: (i) to propose a perceptually realistic method of identifying and verifying isochrony in connected speech; (ii) to advocate and assemble proof for a rhythm-based metric of turn-taking in everyday conversation; (iii) to further an understanding of prosodic contextualization and in particular of rhythm as a context-independent and, at the same time, context-dependent cue to the design and interpretation of speakers' meanings.

This book has grown out of a research project entitled *Prosodic contextualization*, financed over the past four years by the Deutsche Forschungsgemeinschaft. The manuscript was accepted by the Philosophische Fakultät of the University of Zürich as a post-doctoral thesis (*Habilitationsschift*) in Spring 1991. Large portions of the manuscript have been read and criticized by my Konstanz colleagues Peter Auer, Aldo di Luzio and Frank Müller. All remaining errors and inconsistencies are my own. Peter Gebert and especially Allison Wetterlin have helped with the rhythmic analyses

and measurements. Without this support - and the constant encouragement of my Zürich friends and colleagues Silvia Dingwall and Silvia Kübler - the result would have been much the poorer. I am happy to acknowledge my indebtedness to all of the above.

Part of Chapter VIII *Interpreting speech rhythm in specific activity sequences* has appeared as "Contextualizing discourse: The prosody of interactive repair" in P. Auer/A. di Luzio (eds.), *The Contextualization of Language*, Benjamins, 1992. With the permission of the editors and the publisher, it has been included here in the interest of presenting as complete a picture as possible of English speech rhythm function.

I. IS THERE RHYTHM IN SPEECH?

The title of the present work would appear to pre-empt the issue of whether rhythm in speech does in fact exist. But the question is not as foregone as appearances suggest. Much of the relevant literature over the past several decades has been devoted to defending one side or the other in an ongoing debate about isochrony in spoken language. For this reason it is not unfitting to begin by addressing existential questions.

1. The isochrony debate

In retrospect the debate over speech rhythm in English began with the publication of Joshua Steele's now famous *An Essay towards Establishing the Melody and Measure of Speech to be Expressed and Perpetuated by Peculiar Symbols* (1775). Steele's essay was intended as a refutation of claims made by the Scottish Lord James Burnet in a treatise entitled *Of the origin and progress of language* (1774). Among other things Burnet claimed that the English language has no melody:

> We have accents in English, and syllabic accents too; but there is *no change of the tone in them*; the voice is only raised more, so as to be *louder* upon one syllable than another. [...]...the *music of our language* [is], in this respect, *nothing better* than the *music of a drum*, in which we perceive no difference except that of *louder* or *softer*. (cited in Steele 1775:3)

Steele, with musical propensities of his own, set out to disprove this claim by showing that the English language employs more than what Burnet called 'accent' (loud vs. soft) and 'quantity' (long vs. short). He claimed that it has five independent prosodic dimensions:

> *accent* (=pitch: rising, falling or combined)
>
> *quantity* (=duration: longest, long, short, shortest)
>
> *pause* (=silence: semibrief rest, minim rest, crotchet rest, quaver rest)
>
> *emphasis* (=weight: heavy, light, lightest)

> *force* (=loudness: loud, louder, soft, softer; increasing, decreasing)
>
> (1775:24)

By devising a notation system which enabled him to represent these five dimensions and using it to transcribe a then popular stage rendition of Hamlet's soliloquy as well as a (contrived) prosaic dialogue, Steele became the first English prosodist of modern times.[1] Not only is his system astonishingly accurate in many of its details; it also shows great perception in assuming that the prosody of the common man and that of the poet obey fundamentally the same rules.

Judging from Burnet's reply, Steele was instantly successful in convincing the lord that English does have changes in pitch. Lord Burnet, however, was not and never could be fully convinced of the *rhythm* which Steele purported to find in English speech:

> Now as I am no musician, I am not able to make the distinction betwixt *light* and *heavy*, and *loud* and *soft*; and though I have consulted more than one of the greatest musicians here, I cannot discover the difference; nor do they seem to me to understand it any more than I do, even in music. And as to words, I cannot conceive how the *heavy*, or accented syllable, as it is commonly called, should be sounded *soft*, or the *light* syllable *loud*. (cited in Steele 1775:60)

To this Steele replied:

> The variety of *loud* and *soft* should never be considered as (necessarily) a governing principle of *rhythmus;* because though it may, sometimes, be accidentally coincident with rhythmical pulsation, yet it would be offensive if it continued so for any considerable length of time: for the application of *the loud* and *the soft*, both in music and language, either for use or ornament, must not be indiscriminate or periodically alternate, but as occasion calls for it; whereas the rhythmical pulsation is regularly periodical and constant as the swings of a pendulum, but of itself implies no noise or sound at all. (1775:68)

Steele claimed that the dimension heavy/light must be understood in conjunction with a rhythmical pulsation, whose natural origin is the alternation of posing *(thesis)* and lifting *(arsis)* movements of the human foot in walking.

---

[1] The actor David Garrick presumably owes much of his posthumous fame to Steele's transcription of his rendering of 'To be or not to be...'. Steele's prosaic dialogue runs as follows:

- As Peter was going to the hall, he met John.
- Sure, you mistake; you must mean, Peter coming from the hall.
- Coming fróm! No, no. I say going tó. (1775:134)

> Now I say, that the affections of *heavy and light* are the most essential governing powers of *rhythmus*; for, since the accents, *acute, grave,* and *circumflex*, are common both to the *heavy* and to the *light*;
>
> And since quantity, or the *long* and the *short*, are likewise common to each;
>
> And since the accidents of *loud* and *soft* are also common to each;
>
> And lastly, since the accidents of *accent, loudness* and *quantity*, occur not periodically, but occasionally, whilst *cadence* is strictly *periodical*, and divided into *heavy* and *light* alternately; which affections are to be accounted for in the mind, whether *sounding or pausing*, continued or articulated,
>
> It follows, that *heavy* and *light* (as the certain alternate division of cadence) are the most essential governing powers of rhythmus both in poetry and prose. (1775:87f)

Judging from the correspondence which is available to us, Steele was never able to convince Burnet that there is rhythm in speech, although Burnet does admit that, much as the *bourgeois gentilhomme* with prose, he may have been using rhythm all his life without knowing it:

> That language may be divided into bars as well as music, you have shewn very evidently; and it is likely, that a well-taught ear, such as yours, will perceive that division, and will measure speech by it as well as it does a tune. It may also perceive, that those bars proceed either by common or triple time. But I much doubt, whether any man, that is not a musician, can be made to perceive it; the consequence of which is, that it will be of no use. It may, however, be true, that though the division itself may not be perceived by any but those of learned ears, yet the effects of it may be felt by all. For this is generally the case of the popular arts, of which every body feels the effects, but only the learned know the causes which produce them. (cited in Steele 1775:176f)

Burnet's reaction is typical of those who Steele calls the *amusoi* (the unskilled in music) and can be encountered to this day in the reception of speech rhythm studies.

Steele's first ally in the twentieth century was Daniel Jones, who wrote in 1918 that for English "[t]here is a strong tendency in connected speech to make stressed syllables follow each other as nearly as possible at equal distances" (^{9}1960:237). Jones points out that it is this tendency which is to a considerable extent responsible for the variations in duration encountered in English stressed syllables. A 'long' vowel or diphthong in a stressed syllable is shorter if that syllable is followed by an unstressed syllable than it would be if it were final or followed by another stressed syllable. And the more unstressed syllables follow, the shorter the stressed vowel becomes (^{9}1960:237). To demonstrate this, Jones adopts the eighth note from musical notation as a

representation of the time between the 'peaks of prominence' in a series of stressed syllables only. Thus in counting, we have

(a) 

Assuming the intervals of time between these syllables remain constant, any unstressed syllables occurring between the stresses must be fitted in accordingly. Thus:

(b) 

Jones claims, as this notation suggests, that the diphthongs /eɪ/ and /aɪ/ are approximately twice as long in (a) as in (b) ([9]1960:238).

Similar observations concerning isochrony in English speech are found again, several decades later, in Kenneth Pike's *Intonation of American English* (1945). For Pike "[a] sentence or part of a sentence spoken with a single rush of syllables uninterrupted by a pause" is a *rhythm unit* (1945:34); a *simple rhythm unit* contains only one primary contour (strong stress with significant pitch levels at the beginning and end):

> The timing of rhythm units produces a rhythmic succession which is an extremely important characteristic of English phonological structure. The units tend to follow one another in such a way that the lapse of time between the beginning of their prominent syllables is somewhat uniform. (1945:34)

Like Jones, Pike sees one of the consequences of this uniform spacing of stresses as being the necessity to make syllables in rhythm groups with more syllables shorter than those in rhythm groups with fewer syllables:

> Since the rhythm units have different numbers of syllables, but a similar time value, the syllables of the longer ones are crushed together, and pronounced very rapidly, in order to get them pronounced at all, within that time limitation. This rhythmic crushing of syllables into short time limits is partly responsible for many abbreviations -- in which syllables may be omitted entirely -- and the obscuring of vowels; it implies, also, that English syllables are of different lengths, with their length of utterance controlled not only by the lexical phonetic characteristics of their sounds but also by the accident of the number of syllables in the particular rhythmic unit to which they happen to belong at that moment. (1945:34)

Pike goes one step further than Jones, however, in proposing that this kind of rhythm unit be termed *stress-timed* and contrasted with a type in which syllable length is dependent upon the number of syllables rather than the presence of strong stress. The latter he calls *syllable-timed* and describes as follows:

> In these particular rhythm units each unstressed syllable is likely to be sharp cut, with a measured beat on each one; this recurrent syllable prominence, even though the stressed syllables may be extra strong and extra long, gives a "pattering" effect. (1945:35)

According to Pike, both rhythmic types exist in English, although the syllable-timed sort is used only rarely, for instance in spoken chants. Many non-English languages - Spanish, for one, according to Pike - make predominant use of syllable-timed rhythm, a fact which is responsible for considerable interference when e.g. Latin Americans speak English as a foreign language.

Today the two most influential advocates of rhythm in English speech are Abercrombie and Halliday. Abercrombie's approach resembles that of his teacher and mentor Daniel Jones, but also incorporates Pike's typological categories, which are viewed physiologically as different ways of using the breathing muscles to create periodicity:

> Although hesitations and other pauses tend at times to disguise the fact, all human speech possesses *rhythm*. This emerges clearly during those moments when speech is fluent and uninterrupted. Rhythm, in speech as in other human activities, arises out of the periodic recurrence of some sort of movement, producing an expectation that the regularity of succession will continue. The movements concerned in the rhythm of speech are those of the syllable- and stress-producing processes, which together make up the pulmonic air-stream mechanism. [...] Speech rhythm is essentially a muscular rhythm, and the muscles concerned are the breathing muscles. (Abercrombie 1967:96)

According to Abercrombie, it is the way the syllable-producing mechanism, the so-called 'chest-pulse', combines and coordinates with the stress-producing mechanism, the so-called 'stress-pulse', which determines the kind of rhythm a language has.

Abercrombie goes one step further than Pike - who originally formulated the distinction between stress-timing and syllable-timing with respect to rhythmic units only - by extending it to languages in general. The implication is that every language can be clearly assigned to one or the other rhythm type:

> As far as is known, every language in the world is spoken with one kind of rhythm or with the other. In the one kind, known as a *syllable-timed* rhythm,