VOLUME II

Suprasegmental and Prosodic Phonology

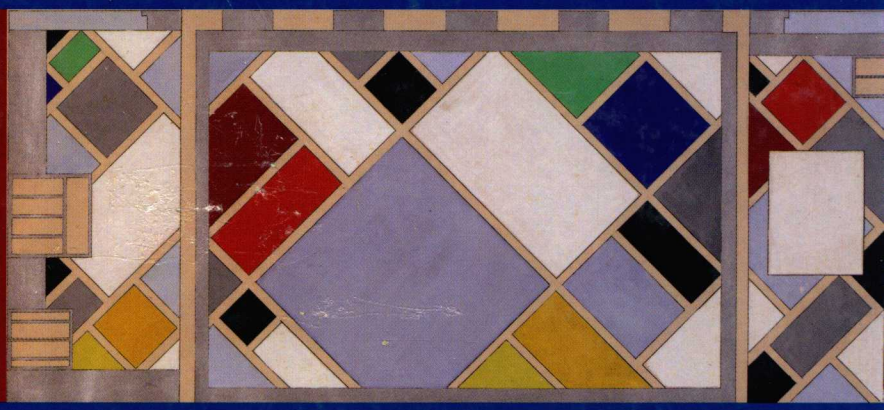
Edited by

Marc van Oostendorp

Colin J. Ewen

Elizabeth Hume

Keren Rice



# THE BLACKWELL COMPANION TO Phonology

WILEY-BLACKWELL

# The Blackwell Companion to Phonology

Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice

Volume II
Suprasegmental and Prosodic Phonology



WILEY-BLACKWELL
A John Wiley & Sons, Ltd., Publication

This edition first published 2011


Blackwell Publishing was acquired by John Wiley & Sons in February 2007. Blackwell's publishing program has been merged with Wiley's global Scientific, Technical, and Medical business to form Wiley-Blackwell.

*Registered Office*
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

*Editorial Offices*
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford, OX4 2DQ, UK
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, for customer services, and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com/wiley-blackwell.



Library of Congress Cataloging-in-Publication Data

The Blackwell companion to phonology / edited by Marc van Oostendorp . . . [et al.].
p. cm. — (Blackwell companions to linguistics series)
Includes bibliographical references and index.
ISBN 978-1-4051-8423-6 (hardcover : alk. paper)
1. Phonetics. 2. Grammar, Comparative and general—Phonology.
I. Oostendorp, Marc van, 1967–
P217.B53 2011
414—dc22 2010042206

A catalogue record for this book is available from the British Library.

Set in 10/12pt Palatino by Graphicraft Limited, Hong Kong
Printed and bound in Singapore by Fabulous Printers Pte Ltd

1 2011

MARC VAN OOSTENDORP is Senior Research Fellow and Head of the Department of Variationist Linguistics at the Meertens Institute of the Royal Netherlands Academy of Arts and Sciences, and Professor of Phonological Microvariation at the University of Leiden. He holds an M.A. in Computational Linguistics and a Ph.D. from Tilburg University.

COLIN J. EWEN is Professor and Chair of English Language and Linguistics at the University of Leiden. He holds an M.A. in English Language and Linguistics and a Ph.D. from the University of Edinburgh. He is editor (with Ellen Kaisse) of the journal *Phonology*.

ELIZABETH HUME is Professor and Chair of the Department of Linguistics at The Ohio State University. She holds a Ph.D. and M.A. in Linguistics from Cornell University, an M.A. in French and Social Psychology of Language from McMaster University, and a B.A. in French from Université Laval.

KEREN RICE is University Professor and Canada Research Chair in Linguistics and Aboriginal Studies at the University of Toronto. She holds an M.A. and Ph.D. in Linguistics from the University of Toronto and a B.A. from Cornell University.

# Blackwell Companions to Linguistics series

Series Editors:
Martin Everaert and Henk van Riemsdijk

**The Blackwell Companion to Syntax**
Edited by Martin Everaert and Henk van Riemsdijk

**The Blackwell Companion to Phonology**
Edited by Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume, and Keren Rice

# Brief Contents

# Full Table of Contents

## Volume I



### *General Issues and Segmental Phonology*

## Volume II

### *Suprasegmental and Prosodic Phonology*



## Volume III

### *Phonological Processes*

## Volume IV

## *Phonological Interfaces*

## Volume V

## *Phonology across Languages*

# 32 The Representation of Intonation

Amalia Arvaniti

## 1 Introduction

It is a well-known truism that no utterance is ever produced in a strict monotone; all utterances, in all languages, show some pitch modulation. Such changes in pitch – impressionistically described as rises and falls – are due to changes in fundamental frequency or F0, the physical property of the speech signal that is determined by the basic rate of vibration of the vocal folds and gives rise to the percept of pitch.

Although pitch modulations exist in all languages, their origin and function differ, in that pitch patterns may be specified either at both the lexical and phrasal levels or only at the phrasal level, resulting in more or less dense tonal specifications, respectively (Gooden *et al.* 2009). The term *intonation* is used to refer to phrasal tonal patterns, while the terms *pitch accent* and *tone* are traditionally used to refer to lexical tonal specifications (CHAPTER 45: THE REPRESENTATION OF TONE). Simplifying somewhat, in languages like English, Italian, Greek, and many other European languages the entire F0 contour is specified at the phrasal level by means of a complex interplay between metrical structure, prosodic phrasing, syntax, and pragmatics; these factors determine where pitch movements will occur and of what type they will be. In languages referred to as *tone languages* – such as Mandarin, Thai, and Igbo – most syllables are lexically specified for tone and tonal changes affect lexical meaning; in languages often referred to as *pitch accent languages* – such as Japanese, Swedish, and Serbian – tone operates in a similar fashion, except that at most one syllable in each word is lexically specified for tone. In both tone and pitch accent languages additional tonal patterns are specified at the phrasal level. Here the focus is on languages without lexical tonal specifications, since it is the intonation of these languages that has been mostly examined.

Determining the structure of pitch modulation and the primitives that make up pitch contours in languages without lexical tone is challenging, since F0 changes are not as discrete and easily identifiable as in "tonal" languages, their connections to segmental material are less easy to determine, and associated meanings are harder to pinpoint since they deal with information structure and pragmatic interpretation rather than lexical semantics. The following examples illustrate these points. In Figure 32.1, two utterances are shown, *Me?!* and *A*

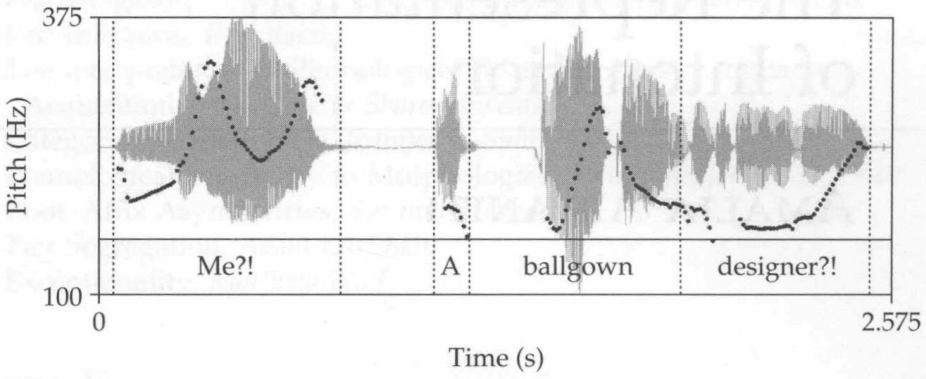



**Figure 32.1** Waveforms and F0 contours of two English utterances illustrating the incredulity contour (Hirschberg and Ward 1992); on the left, *Me?!*; on the right, *A ballgown designer?!*. Vertical lines indicate word boundaries

*ballgown designer?!*, both using the rise-fall-rise melody that implies incredulity (Ward and Hirschberg 1985; Hirschberg and Ward 1992). They are plausible responses to a career advisor's pronouncement that, according to test results, designing ballgowns is the recommended career choice for the speaker, who has all along dreamed of becoming an aerospace engineer (for similar examples, see Ladd 2008: 45–46). Although the short contour can be informally described as rise–fall–rise, the longer contour cannot be described in a similar fashion, as it shows a long low-level stretch between a rise–fall and a final rise. In Figure 32.2, Greek contours very similar to the English ones in Figure 32.1 are shown, though in the case of Greek these contours are used for wh-questions (Arvaniti and Baltazani 2005; Arvaniti and Ladd 2009). As can be seen, the same issue with overall shape arises here as well. Further, as Arvaniti and Baltazani (2005) note, the Greek melody in Figure 32.2 can also be used for polite requests employing an imperative; e.g. ['ðose sti ma'ria 'liɣo ne'raci] 'give Maria some water' (*lit.* give to Maria a-little water-DIM). Finally, Figure 32.3 illustrates two instances of another English melody: unlike the contours in Figures 32.1 and 32.2, which look different from each other but convey the same meaning in each case, the contours of Figure 32.3 are realizations of the same melody but convey different meaning, depending on the utterance: the melody used is

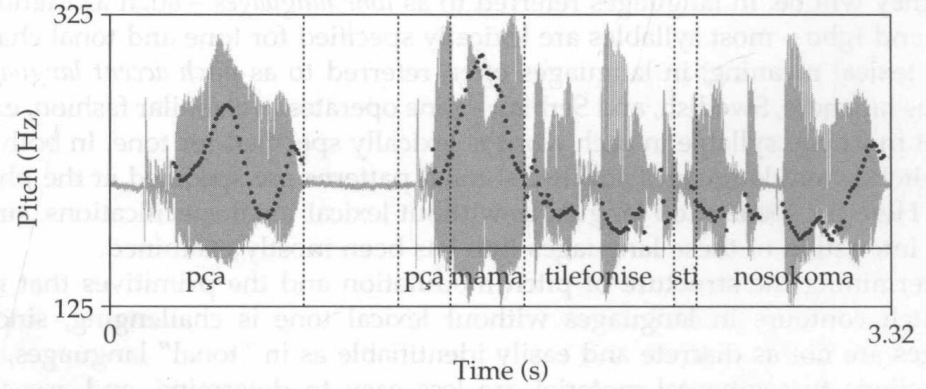



**Figure 32.2** Waveforms, transcriptions, and F0 contours of two Greek wh-questions, on the left, ['pça] 'which (FEM)', on the right, ['pça ma'ma tile'fonise sti noso'koma] 'which mom called the nurse?' Vertical lines indicate word boundaries

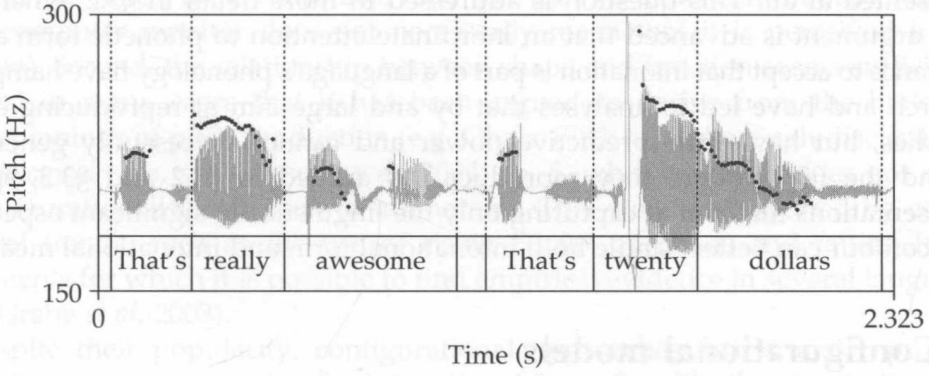



**Figure 32.3** Waveforms and F0 contours of two English utterances; on the left, *That's really awesome*; on the right, *That's twenty dollars*. Vertical lines indicate word boundaries

the default for *That's twenty dollars*, but sounds blasé or sarcastic when used with *That's really awesome*. Note that this is not because the melody is wrong for *That's really awesome*: the use is legitimate and meaningful (if the speaker wishes to be sarcastic or convey her indifference).

These examples illustrate three main points about intonation. First, they show that the shape of intonational contours with a given pragmatic interpretation can vary substantially, depending on the segmental material with which they are uttered. Such differences are not random, but related to the overall prosodic structure of the utterance with which the contours are associated, including the number of syllables and the position of stresses (where applicable). Second, the examples show that contours do not have a constant meaning, either within or across languages; within a language, their interpretation may well depend on lexical and other choices that accompany the use of the melody; across languages differences can be arbitrary. A successful theory of intonation should be able to capture these properties: it should be able to explain the connection between intonation and meaning and make generalizations from surface F0 data with sufficient predictive power to generate new contours of the same basic melody to "fit" new utterances of varying lengths and structures.

Although the above observations are by and large shared by most intonational models, the ways in which they treat these basic properties show substantial differences. As discussed in more detail in §2.1, many researchers have treated F0 contours as gestalts or *configurations*, that is, as holistic pitch movements that encompass entire utterances and have a uniform meaning. In other models, melodies are seen as being composed of primitives of some sort. These primitives are considered to be either local configurations (or *dynamic* tones), such as local rises and falls, or level tones, such as high, mid, and low. Here I review both the controversy between advocates of gestalt approaches to intonation and those who proposed analyses based on the decomposition of melodies into smaller elements, and the disagreement between researchers who use dynamic tones (that is, local configurations) as the primitives of intonational structure and those who advocate the use of level tones instead.

As I show, however, focusing only on the form of intonational primitives avoids an even more fundamental question, namely which parts of a contour should be

represented at all. This question is addressed in more detail in §3.2, where the main argument is advanced that an inordinate attention to phonetic form and a reluctance to accept that intonation is part of a language's phonology have hampered research and have led to analyses that by and large aim at reproducing entire melodies, but have little predictive power and cannot successfully generalize beyond the F0 contours they reproduce. As argued in §3.2 and §3.3, sparse representations that aim at capturing only the linguistically significant aspects of each contour can better handle both intonational form and intonational meaning.

## 2 Configurational models

### *2.1 Melodies as gestalts*

As mentioned above, many researchers have treated intonation contours as gestalts, such as Bolinger (1951), Jones (1972), Cooper and Sorensen (1981), Hirst and Di Cristo (1998), Grabe *et al.* (2003), and Xu (2005). In these works, it is most often the case that intonational contours are seen as holistically and directly reflecting certain functional or structural aspects of speech, such as the distinction between questions and statements or that between levels of phrasing.

According to Jones (1972: 279) – who in the last edition of *An outline of English phonetics* followed several earlier intonational analyses, notably those of Armstrong and Ward (1926) and Kingdon (1958) – English has two basic tunes, Tune 1 and Tune 2. These are a fall and rise respectively, which "may be spread over a large number of syllables, or [. . .] be compressed into smaller spaces." Bolinger (1951: 208) also concludes his critique of level-based analyses (see §3.1) by saying that "intonation could not be a more appropriate illustration of the Gestalt." More recently, Cooper and Sorensen (1981) presented a series of experiments in which contours are treated as undivided wholes and peak height is taken to directly reflect levels of phrasing.

Modern versions of the gestalt approach include INTSINT (International Transcription System for Intonation; e.g. Hirst and Di Cristo 1998; Hirst *et al.* 2000), OXIGEN (Oxford Intonation Generator; Grabe *et al.* 2003), and PENTA (Parallel Encoding and Pitch Target Approximation; e.g. Xu 2005). In INTSINT, entire intonation contours are transcribed, using symbols that represent pitch movements. The movements, however, are not seen as primitives but rather as a means to transcribing the course of F0 over an entire utterance (hence their descriptive labels Higher, Lower, Upstepping, Downstepping, Same, Top, and Bottom, which express the course of F0 relative to preceding points and the overall range of the speaker). In OXIGEN, polynomials are used to model overall contour shape differences between statements and questions in British English. Finally, in PENTA, each syllable in a contour has its own pitch specification, while global aspects of the overall melody are directly linked to functional effects (a feature shared with OXIGEN); e.g. the use of an utterance as statement or question is said to lead to changes in overall pitch shape from fall to rise (Xu 2005).

Configurational approaches have been quite popular for several reasons. First, their appeal is intuitive: F0 contours are more or less continuous, so, as Bolinger (1951: 206) put it, intonation can be seen as "a pattern [. . .] in the fundamental, down-to-earth sense of a continuous line that can be traced on a piece of paper"

(though, as noted in Arvaniti 2007, the fact that F0 looks continuous on paper or a computer monitor does not necessarily mean that it is perceived in this fashion). Second, the relationship between shape and function seems sufficiently natural in many cases that it has been argued to derive from the biological underpinnings of pitch production (e.g. Ohala 1983; for a thorough discussion of the *biological code*, see Gussenhoven 2004: ch. 5, who, however, does not adopt a configurationalist approach to intonation). This view is supported by certain general trends, such as the use of rising F0 for questions and falling F0 for statements for which it is possible to find empirical evidence in several languages (e.g. Grabe *et al.* 2003).

Despite their popularity, configurational approaches face a major problem when it comes to accounting for intonational form. Specifically, if melodies were undivided wholes, they should simply shrink and stretch to "fit" the segmental duration of the utterance they co-occur with. There is plenty of evidence, however, that when tunes are matched with utterances of varying lengths and different metrical structures they are not realized in this simple manner. This was observed by 't Hart *et al.* (1990: ch. 4), among the first researchers to use instrumental rather than impressionistic data for intonation research (e.g. Cohen and 't Hart 1968; 't Hart and Cohen 1973; 't Hart and Collier 1975). They noticed that in their Dutch corpora sequences of pitch movements would appear on a single syllable in some instances, but would be separated by several syllables in others (cf. Figures 32.1 and 32.2). Importantly, 't Hart *et al.* found that this *elasticity*, as they called it, did not affect the melodic identity of the contour (determined by means of perceptual experiments; see §2.3), even though it radically altered the overall contour shape (thus, the concept of elasticity can be juxtaposed with the *compression* envisaged by Jones 1972, which implies greater uniformity in the squeezing and stretching of contours).

Results from several later studies support the original observations of 't Hart *et al.* that certain aspects of the contour are important for listeners, while overall contour shape is not. Pierrehumbert and Steele (1989) varied in steps the position of the pitch peak in English melodies that can be holistically described as rise–fall–rise, and found that listeners imitating these stimuli produced not a continuum but two different contours, one with an early and one with a late peak. Similar results have also been presented by Redi (2003), following Pierrehumbert and Steele's imitation protocol (argued by Gussenhoven 1999 to be the best way to examine categoricality in intonation). Similarly categorical responses to intonational continua that would be holistically seen as instances of the same contour have also been obtained by Rietveld and Gussenhoven (1995) for Dutch, D'Imperio and House (1997) for Neapolitan Italian, and Botinis (1989) for Greek, inter alia. The contours in Figures 32.1 and 32.2 also illustrate this general point. In the monosyllabic utterances, the rise–fall–rise stretches over the entire syllable. In the longer utterances, however, the rise co-occurs with the first stressed syllable (with some peak delay) and the final rise is realized on the last syllable, while the fall and subsequent low-level stretch vary depending on the language and length of the utterance. As a result, the contours of the longer utterances are not stretched-out copies of the shorter contours, nor are the shorter contours compressed versions of the longer contours; these differences, extensively discussed in Arvaniti and Ladd (2009), are illustrated in Figure 32.4, using the Greek contours of Figure 32.2.

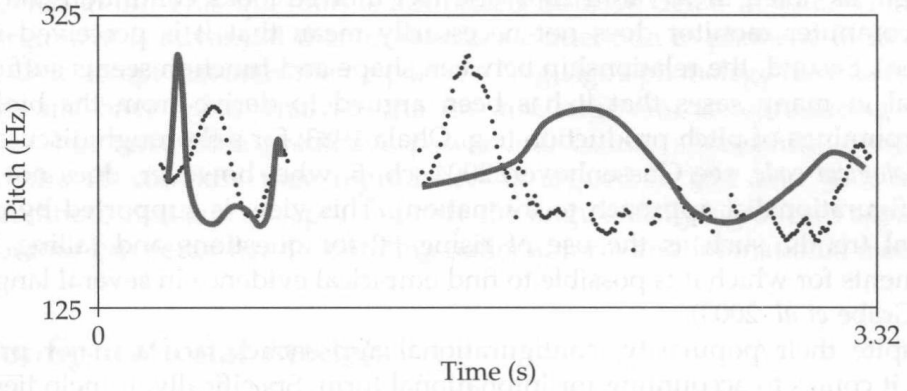



**Figure 32.4** F0 contours of the Greek wh-questions shown in Figure 32.2; the solid gray lines are compressed and stretched-out copies of the long and short contour respectively

Differences in contour shape may also relate to the number and position of stressed syllables in an utterance and the location of the word in focus. This is experimentally shown in Arvaniti *et al.* (2006a), who studied the default melody of Greek polar questions in which the position of focus can vary. They show that the shape of the contour is strongly affected by the position of the stressed syllables and of the word in focus (see also Arvaniti and Baltazani 2005; Arvaniti 2007). The focus effect in particular is illustrated in Figure 32.5, which shows the two contours that can be used with the sentence ['pinun lemo'naða] 'they drink lemonade' when it is uttered as a question with focus on the verb (dotted line) or the noun (solid line). As can be seen, no description in terms of overall shape can possibly cover both contours; at best, the late focus question would be characterized as rise–fall–rise–fall and the early focus question as rise–fall, but this ignores the location of the rise–fall part that the two contours share and the significance of this location both for understanding the pragmatics of the two questions and for their naturalness (for details, see Arvaniti *et al.* 2006a).

Holistic approaches, then, suffer from two problems. First, they cannot represent in the same manner contours that superficially look different, like the contours in Figure 32.5. At the same time, holistic approaches cannot account for systematic differences between realizations: e.g. they cannot account in a principled manner for the fact that the syllable [nun] in Figure 32.5 is low in the dotted, early focus contour, but rising in the solid, late focus contour. In short, configurational approaches cannot represent different instantiations of the same melody in a way that can either capture their similarity or predict their differences, thereby failing one of the main criteria for an adequate theory of intonational phonology mentioned in §1.

## 2.2 *Intonational gestalts and meaning*

In addition to issues with intonational form, gestalt approaches encounter problems with intonational meaning. In gestalt models, overall contour shape is said to associate with differences in meaning. Yet, as illustrated in §1, the relationship between melody and meaning is not one-to-one: the same melody can lend different pragmatic nuances to different utterances, while the same meaning can

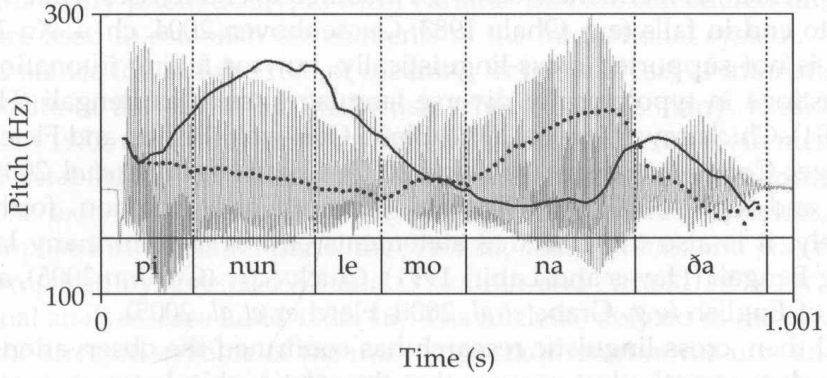



**Figure 32.5** Waveform, transcription, and F0 contours for the phrase ['pinun lemo'naða] 'they drink lemonade', uttered as a question with focus on ['pinun] 'they drink' (dotted contour) or on [lemo'naða] 'lemonade' (solid contour); the former could be glossed as 'is lemonade something they would drink?' and the latter as 'is it lemonade that they are drinking?' Vertical lines indicate syllable boundaries

be expressed by superficially different-looking contours. This lack of one-to-one correspondence has been repeatedly noted over the years (among many, Bolinger 1964; Lehiste 1970: 95ff. and references therein; Ladd 1980; Pierrehumbert 1980; 't Hart *et al.* 1990: ch. 4), and prompted Pike (1945: 23ff.) to strongly caution against the practice of investigating contour meaning on the basis of grammatical structure (such as pitting statements against questions). Overall, then, melodies do not appear to have specific functions, and indeed attempts to describe the melodies of specific pragmatic nuances, such as irony, have proved unsuccessful (e.g. Bryant and Fox Tree 2005).

In addition, cross-linguistic research has shown that functional effects of the sort favored by gestalt approaches are expressed in language-specific ways. Such findings abound and strongly argue against a natural or direct relationship between intonational form and function. For instance, it has been argued that focus is universally expressed as local pitch expansion with a concomitant reduction in pitch range post-focally (e.g. Xu 2005). Yet a review of the literature clearly shows that this is far from a universal mechanism for the prosodic marking of focus. For example, in Greek polar questions the word in focus has the lowest F0 in the entire utterance and pitch expansion is associated with the post-focal region (Arvaniti *et al.* 2006a; see Figure 32.5 for an illustration). Taiwanese relies on duration to mark focus rather than changes in pitch (Pan 2008), while in many other languages pitch expansion is just one, optional, manifestation of an overall prosodic reorganization under focus (e.g. Chen 2006, 2010 for Mandarin; de Jong 1995 and Harrington *et al.* 2000 for English; Baltazani 2006 and Arvaniti *et al.* 2006b for Greek; Jun 2005a for Korean; Venditti *et al.* 2008 for Japanese). Perhaps the strongest counter-argument against the view that the relationship between focus and pitch range is natural and direct is the fact that not all languages can use intonation to mark focus (e.g. Swerts *et al.* 2002 on Italian; see Ladd 2008: ch. 6 for a discussion). If the relationship between focus and intonation is natural and direct, there is no good explanation of why some languages do not avail themselves of this option.

The presence of an arbitrary relationship between intonation and meaning (as in all aspects of linguistic structure) is also evident in cross-linguistic data from