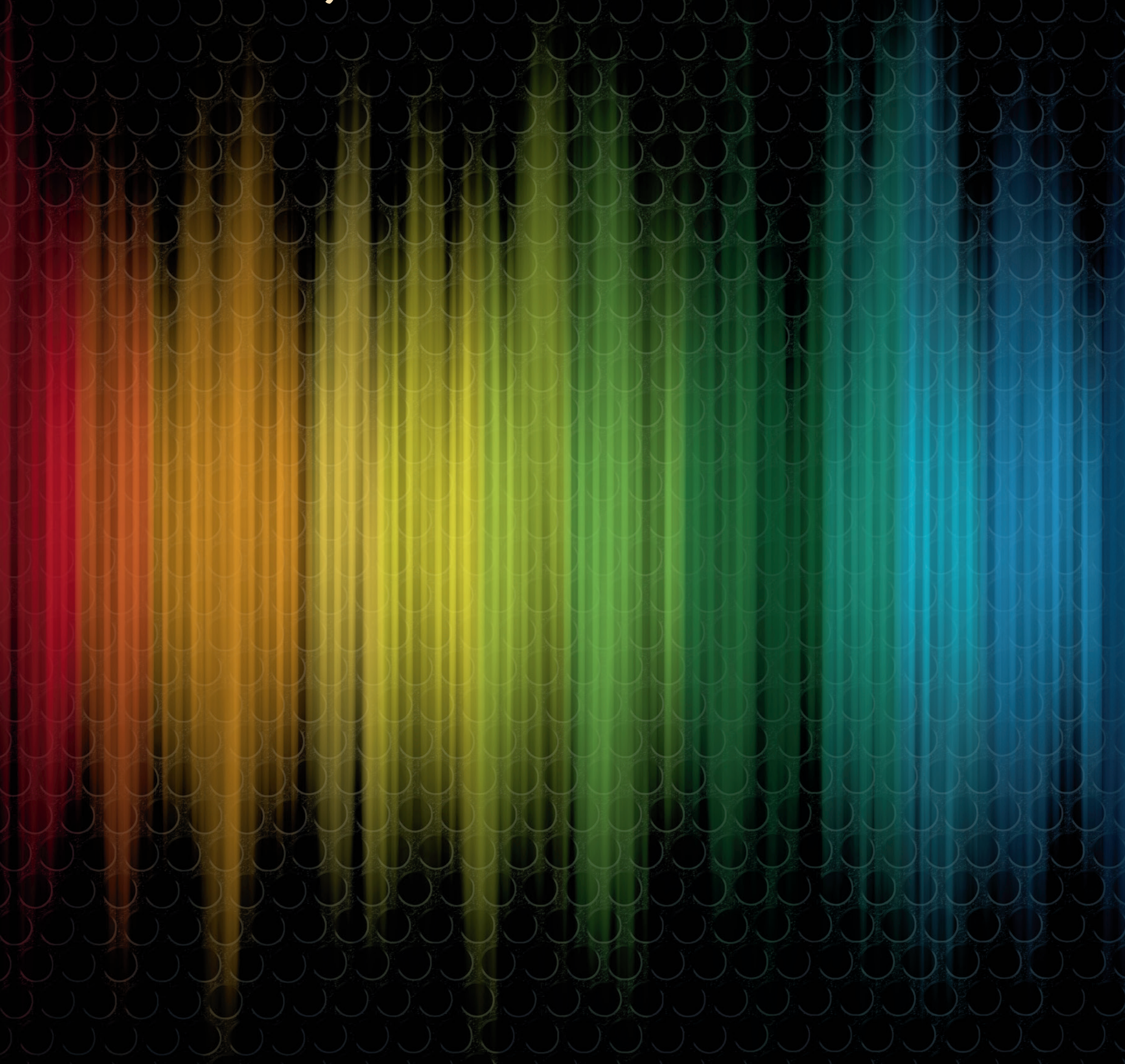
# The Study of Word Stress and Accent

## Theories, Methods and Data

Edited by Rob Goedemans, Jeffrey Heinz
and Harry van der Hulst



# The Study of Word Stress and Accent

Stress and accent are central, organizing features of grammar, but their precise nature continues to be a source of mystery and wonder. These issues come to the forefront in the phonetic manifestation of stress and accent, their cross-linguistic variation and the subtle and intricate laws they obey in individual languages. Understanding the nature of stress and accent systems informs all aspects of linguistic theory, methods, typology and especially the grammatical analysis of language data. These themes form the organizational backbone of this book. Bringing together a team of world-renowned phonologists, the volume covers a range of typological and theoretical issues in the study of stress and accent. It will appeal to researchers who value synergistic approaches to the study of stress and accent, careful attention to cross-linguistic variation, and detailed analyses of both well-studied and under-studied languages. The book is a lively testimony of a field of inquiry that shows progress, while also identifying questions for ongoing research.

ROB GOEDEMANS is Information Manager in the Humanities Faculty of Leiden University. His research focuses on phonetics, phonology and typology of stress in the languages of the world in general, and the languages of Aboriginal Australia and Indonesia in particular. Together with Harry van der Hulst, he has worked on several publications based on the StressTyp database, with which he has been involved since its inception.

JEFFREY HEINZ is Professor in the Department of Linguistics and the Institute of Advanced Computational Science at Stony Brook University. He conducts research at the intersection of theoretical linguistics, theoretical computer science and computational learning theory. With Rob Goedemans and Harry van der Hulst, he helped develop the StressTyp2 database, which organizes and presents information on the stress and accent patterns in hundreds of languages around the world.

HARRY VAN DER HULST is Professor of Linguistics at the University of Connecticut. He specializes in phonology, which is the study of the sound systems of languages, as well as the visual aspects of sign languages. He has published 25 books and over 130 articles and has been Editor-in-Chief of the international linguistic journal *The Linguistic Review* since 1990. With Rob Goedemans and Jeff Heinz he helped develop the StressTyp2 database, which organizes and presents information on the stress and accent patterns in hundreds of languages around the world.

# The Study of Word Stress and Accent

## *Theories, Methods and Data*

---

*Edited by*

Rob Goedemans

*Leiden University*

Jeffrey Heinz

*Stony Brook University*

Harry van der Hulst

*University of Connecticut*

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre,
New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107164031
DOI: 10.1017/9781316683101

First published 2019

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

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

*Library of Congress Cataloging-in-Publication Data*
Names: Goedemans, Rob, editor. | Heinz, Jeffrey, 1974– editor. | Hulst, Harry van der, editor.
Title: The study of word stress and accent : theories, methods and data / edited by Rob Goedemans, Jeffrey Heinz, Harry van der Hulst.
Description: Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2018. | Includes bibliographical references.
Identifiers: LCCN 2018011858 | ISBN 9781107164031 (alk. paper)
Subjects: LCSH: Accents and accentuation. | Grammar, Comparative and general – Phonology.
Classification: LCC P231 .S78 2018 | DDC 414/.6–dc23
LC record available at https://lccn.loc.gov/2018011858

ISBN 978-1-107-16403-1 Hardback

# Contents

## Part III Case Studies

# Figures

# Tables

# Contributors

ROB GOEDEMANS, Leiden University

JEFFREY HEINZ, Stony Brook University

LARRY M. HYMAN, University of California, Berkeley

RENÉ KAGER, Utrecht University

BJÖRN KÖHNLEIN, Ohio State University

HARUO KUBOZONO, National Institute for Japanese Languages and Linguistics

NATALIA KUZNETSOVA, Institute for Linguistic Studies, Russian Academy of Sciences

ANYA LUNDEN, College of William and Mary, Williamsburg, VA

VIOLETA MARTÍNEZ-PARICIO, University of Valencia

KEREN RICE, University of Toronto

IGGY ROCA, University of Essex

NICHOLAS ROLLE, University of California, Berkeley

HISAO TOKIZAKI, Sapporo University

VINCENT J. VAN HEUVEN, University of Pannonia, Veszprém

MARC VAN OOSTENDORP, Radboud University, Nijmegen

HARRY VAN DER HULST, University of Connecticut

ALEXANDRE VAXMAN, University of Connecticut

MARINE VUILLERMET, Laboratoire Dynamique Du Langage, CNRS & Université de Lyon

# Introduction

*Rob Goedemans, Jeffrey Heinz and Harry van der Hulst*

The present volume contains 13 original chapters, each of which is based on a presentation at a three-day conference organized by Rob Goedemans, Jeff Heinz and Harry van der Hulst and held on 15–17 August 2014 at Leiden University. Some chapters are by invited speakers and others by speakers whose abstracts were selected for presentation. All final versions of the chapters were submitted towards the end of 2016, after an anonymous double peer-review procedure. The conference was the fourth in a series of conferences that are part of a large project on establishing a database for word stress systems (StressTyp2).[1] StressTyp2 is a typological database that supplies information about the stress/accent systems of the world's languages. This database is based on two previously designed databases, StressTyp (ST1) and the Stress Pattern Database, while also incorporating information in van der Hulst (1999b) and van der Hulst, Goedemans and van Zanten (2010).[2] The goals of StressTyp2 (ST2) were to improve, verify and enrich the combined datasets in a variety of ways and to develop a web-based interface that makes the information in ST2 easily available to researchers and citizens around the world, and which meets or exceeds professional and scientific standards. The third goal of this project was to adopt (and, where necessary, establish) best practices for the collection, organization, dissemination and presentation of typological data pertaining to sound patterns in natural language. Currently StressTyp2 is available on the Web (http://st2.ullet.net/).

As with the previous volumes, chapters in the present volume are not concerned with the technical details of the StressTyp2 project, but are based on some of the public talks in which more general issues were addressed that

---

[1] Two previous conferences were held at the University of Connecticut, on 30 April 2010 and 3 December 2011, respectively. A volume based on these conferences appeared in 2014 (van der Hulst 2014c). The third conference was held at the University of Delaware on 29 November to 1 December 2012 (Heinz, Goedemans and van der Hulst 2016).

[2] The project is a broad collaboration between Harry van der Hulst (University of Connecticut), Rob Goedemans (Leiden University), developers of StressTyp1 and Jeffrey Heinz (University of Delaware), developer of the Stress Pattern Database. This project is financed by the National Science Foundation, NSF grants NSF#1123661 (PI H. van der Hulst), NSF# 1123692 (PI J. Heinz).

relate to typologically based theoretical work. In this introductory chapter for the present volume, our goal is to provide summaries of the 13 contributions and to briefly discuss the common threads in these chapters.

The first chapter in the first volume in this series (van der Hulst 2014a) provides a broad introduction to the study of word stress/accent, as well a detailed description of the StressTyp2 project. This chapter discusses the terminological issues that arise in the study of word prominence, in particular regarding usage of the terms ‘stress’ and ‘accent’. While many writers (also in the present volume) use the term ‘stress’ as a cover term for all word prominence effects that do not involve tonal or, more generally, predominantly pitch exponents (often regarding the term ‘accent’ as interchangeable), other writers specifically take accent to be an abstract (i.e. phonetics-free) property of one or more syllables (being lexically specified or being the head of a predictable metrical constituent structure), while seeing stress as an exponent of accent (with typical cues such as extra duration, increased intensity and elevated fundamental frequency). This latter view would speak of a *stress-accent* language as being typologically different from a *pitch-accent* language in which the primary exponent of accent is fundamental frequency. Often stress-accent systems differentiate between a primary stress location and secondary or rhythmic stresses which may differ in their precise phonetic exponents, which are usually more clearly detectable for the primary stress.[3]

In the following discussion of the chapters contained in this volume, we will make an effort to clarify how different authors use their terminology. When referring to systems that fall within the stress category, we will sometimes use the expression ‘stress/accent’ instead of stress-accent because not all authors differentiate these two terms. An even ‘vaguer’ term such as *word prominence* will also in some cases come in handy.

The chapters in this volume have been grouped under three main themes:

- Phonetic correlates and prominence distinctions
- Typology
- Case studies.

Of course, some chapters, as we will indicate below, address issues that cross this thematic division.

## Part I Phonetic Correlates and Prominence Distinctions

The first four chapters focus on two general areas: the phonetics of stress/accent and the question of which kinds of prominence effects count as stress/accent. Van Heuven offers a thorough overview of the correlates of stress/accent and the way in which listeners do, or do not, use them. In a sense, this is the base to

---

[3] For a general discussion of ‘word stress’ see also Gordon and van der Hulst (forthcoming).

which Hyman wants to return. His contribution argues against the significance of using (or quarrelling about) labels such as ‘stress’, ‘stress-accent’ or tone- or pitch-accent in favour of analysing the prominence properties that really matter to the language user. Lunden shows that such correlates can play a crucial role in the phonology of stress/accent, even if the presence of these correlates is not actually reflected in metrical representations. What matters, again, is the perception of the listener. Finally, Kuznetsova argues that the distribution of the phonetic correlates forms the basis for the typological classification of the language’s word prosody.

### *1 Vincent van Heuven: Acoustic Correlates and Perceptual Cues of Word and Sentence Stress: Towards a Cross-linguistic Perspective*

In this chapter the focus of interest is the phonetic realisation of stress at the word and sentence level. While not dealing with the physiological basis of stress (but see van Heuven and Sluijter 1996 for more discussion), it concentrates on the acoustic consequences of increased versus decreased effort and asks (i) what acoustic correlates can be found for the difference between a stressed syllable and its unstressed counterpart, and (ii) what the relative importance is of each acoustic correlate in the marking of stress. At the same time, the author considers the question of which acoustic properties are used by human listeners and to what extent these are used to decide whether or not a syllable is stressed. Van Heuven makes a strict terminological distinction between acoustic correlates of stress (which can be used, for instance, to identify a stressed syllable by some computer algorithm) and the perceptual cues used by the human listener, showing that some acoustic correlates, notably the (peak) intensity of a syllable, allow excellent separation of stressed from unstressed syllables but are hardly used by the human listener.

### *2 Larry Hyman: Positional Prominence versus Word Accent: Is There a Difference?*

Hyman addresses one of the major unresolved issues in the study of word-accentual systems, which is determining what exactly counts as accent, a problem which is further complicated in languages with tone or so-called pitch-accent (see the above remarks and the introductory chapter in van der Hulst 2014c for a general discussion of such matters). In this chapter, Hyman analyses three African cases, each of which display diverse positional prominence effects that are clearly word level, reasonably subject to a metrical (accentual) interpretation, but which do not consistently coincide. In Ibibio, a Cross-River language spoken in Nigeria, greater consonant and vowel contrasts suggest that the initial stem syllable is the accented head of a trochaic foot, whose required shape varies, however, by construction. In Punu, a Bantu language spoken in Gabon, tone suggests that the word-penultimate syllable is accented, while vowel length suggests that both stem-initial and word-penultimate syllables are accented. In Lulamogi, a small, understudied Bantu language spoken in Uganda, vowel length suggests that all stem (versus prefix) and word-penultimate syllables are accented, while tone suggests it is the penult. While some, or all, of these instances of positional prominence resemble what is found in stress-accent systems, Hyman concludes that we should focus more on the specific properties of prominent positions and less on what we call them.

## *3 Anya Lunden: Explaining Word-Final Stress Lapse*

Focusing on rhythmic stress, this chapter proposes and examines evidence for a motivation behind the well-known asymmetrical tolerance for a stress lapse word-initially versus word-finally across languages. While many binary-stress languages tolerate a stress lapse at the right edge of the word, very few tolerate a stress lapse at the left edge (see Gordon 2002; van der Hulst 2014b). Lunden proposes that, in the languages that tolerate a final stress lapse, there is nevertheless a rhythmic alternation present at the right edge of the word, due to the phonetic effect of word-level final lengthening. However, if final lengthening were able to perceptually contribute to a word's rhythm in cases of final lapse, we would expect this to only be possible when stress in the language has duration as a stress correlate. Evidence from two different sources is shown to support the connection between languages' tolerance of final stress lapse and their use of duration in stress. Drawing on a database of stress correlates that the author made for this research project, it is shown that languages which tolerate a final stress lapse are indeed extremely likely to have duration as a stress correlate, whereas no such correlation exists for final lapse and the stress correlates of pitch or intensity. Several perception experiments also support this connection: final lengthening was found to be confusable with stress only if stressed syllables included increased duration. Finally, an account is sketched of how this could be captured in an Optimality Theoretic (OT) analysis.

## *4 Natalia Kuznetsova: What Danish and Estonian Can Show to a Modern Word-Prosodic Typology*

The typology of word prosody is still a subject of hot debate (see van der Hulst 2014a). Kuznetsova asserts that tone and stress-accent remain the central units of classification, but shows that there is no established consensus about a definition of these notions. In this chapter, she focuses on two specific word-prosodic units with a non-pitch-based primary phonetic exponent: prosodic quantity in Standard Estonian and prosodic laryngealization in Copenhagen Danish. Kuznetsova summarizes their main phonetic and functional features. She also compares these prosodic units with functionally similar cases of pitch-based word prosody in other languages within what she calls the framework of mainstream word-prosodic typology. Both cases are challenging for the typology, as they do not qualify either as tone or as stress. In the end, she proposes a view on the word-prosodic typology which incorporates a clear separation between the variable of location and ways in which the location is realized or cued (see also van der Hulst 1999a, 2010, 2014). At the word level, we find three logically possible LOCATION values: (i) 'no prosody at the word level', (ii) 'prosodic marks on some syllables of a word', (iii) 'prosodic marks on every syllable of a word'. Case (i) would imply that no phonetic cues have phonological relevance at the word level, the tentative term for this could be 'non-accentual prosody'. This is the case, for example, in French, where all prosody can be defined at the post-lexical levels. Case (iii) is what is prototypically called 'tone'. Case (ii) is what Hulst (2011) proposes to call 'accent'. She concludes that an extensive typology of all the accent varieties should be based on accurate descriptions of the word prosody of particular languages.

## Part II Typology

Clearly, both preceding chapters (3 and 4) deal not only with both phonetic exponents but also typological issues. The following two chapters also address typological issues, both in different ways. Chapter 5 addresses the long-standing debate about foot typology (see van der Hulst 1999a, 2010), adding a new type of foot to the inventory. Issues of foot typology are also taken up in Chapters 7, 8 and 9. Chapter 6 is typological in a different way in that it focuses on correlations between phonological structure and syntactic structure; this line of work falls within the class of so-called holistic typological studies (see van der Hulst 2017).

### *5 René Kager and Violeta Martínez-Paricio: Mora and Syllable Accentuation: Typology and Representation*

This chapter has two goals. First, the authors argue that metrical feet can immediately dominate morae. This situation occurs only under duress of metrical foot form constraints, which impose strict requirements on the number of morae in the head and dependent positions in metrical feet. The authors propose to encode this situation in terms of the internally layered (IL) foot, a minimally recursive metrical foot (Martínez-Paricio 2012, 2013, Martínez-Paricio and Kager 2015). They support this claim with data from Gilbertese, which exhibits a metrical distribution of high pitch and stress that disrespects syllable integrity. They show that this pattern can be analysed straightforwardly using IL feet. A second goal of this chapter is to show that IL feet offer an insightful account of 'mora-counting' metrical patterns in which prominence is located on the syllable that contains the antepenultimate mora, in particular in Tokyo Japanese loanword accentuation and Dihovo Macedonian stress. These patterns pose serious problems for a standard moraic trochee analysis, due to the fact that the weight sequences . . . LH# and . . . HL# behave similarly as trimoraic units in locating prominence on the penultimate syllable. Using IL feet, Kager and Martínez-Paricio propose analysing these systems as 'mixed binary/ternary': IL feet occur in sequences ending in light syllables, non-IL feet elsewhere. Additional evidence for this analysis comes from an innovative pattern of loanword accentuation of Tokyo Japanese, which differs from the conservative pattern in shifting the accent to the antepenultimate syllable precisely in . . . LH# sequences. On this analysis, this pattern has IL feet in all forms. The fact that Tokyo loanword accentuation shows signs of accentual instability and is moving in the direction of the Latin stress pattern can be interpreted as movement towards consistency in foot parsing.

### 6 *Hisao Tokizaki: Word Stress, Pitch Accent, and Word Order Typology with Special Reference to Altaic*

It has been claimed that in some languages the location of word stress correlates with the word order of a syntactic head and its complement – for example, a verb and its object (Donegan and Stamp 1983, Plank 1988) – that is, languages with left-hand stress have head-final order while languages with right-hand stress have head-initial order. Based on an analysis of the data in Haspelmath and Dryer (2005), Tokizaki shows that this correlation generally holds in the world's languages. However, potential counterexamples to this generalization are Altaic languages, a large number of which have been reported to have right-hand stress and head-final order (see Goedemans, Heinz and van der Hulst 2014 and data in Haspelmath et al. 2005). It is argued in this chapter that Altaic languages in fact have word-initial stress as well as right-hand pitch accent. Thus, the general correlation between stress and word order also holds for Altaic languages.

## Part III Case Studies

In this part, Chapter 10 is unique in discussing focus prosody, going beyond the word unit and including phrasal phonology. While focusing on the analysis of specific word-prosodic systems, all other chapters in this part raise specific issues with respect to the role of lexical marking of stress/accent, as well as of morphological structure.

### 7 *Keren Rice: Persistence and Change in Stem Prominence in Dene (Athabaskan) Languages*

The stem has long been identified as a domain of prominence in Dene (Athabaskan) languages. Given this, one might ask if, in addition to the importance of morphology in the placement of prominence, phonological factors also play a role. In this chapter, Rice examines the role of morphology and of phonological constraints in a number of Dene languages, addressing the pathways of change that can be identified in this family. Overall, she notes that there is conservatism in the prominence systems reported for Athabaskan languages, with the root attracting prominence, all other things being equal. The phonological factor that plays a role is the placement of a trochaic foot. While the core of the reconstructed system is maintained, the trochee may shift from syllabic to moraic, with different factors involved in determining weight. Word-level prominence may be either left or right oriented. Prefixes may involve trochees as well. What is overall resistant to change is the domain of root.

### 8 *Iggy Roca: Spanish Word Stress: An Updated Multidimensional Account*

Generative Phonology research output on Spanish word stress spans just over half a century at the time of writing (1965–2016) and is substantial. However, no unanimity of analysis has as yet been achieved. This chapter provides both a précis of the historical peaks of the still ongoing endeavour and a proposal further elaborating the lines in Roca’s (2006, 2014, 2016) recent contributions. For reasons of space, Roca’s analysis is restricted to non-verbs. The proposed grammar uses OT as an analytical tool. It hinges on interactions between foot shape and size, on the one hand, and alignment relations between the metrical and morphological structures, on the other. The account is multidimensional in as much as it provides (i) a full formalisation of Spanish (non-verb) stress, (ii) a typology of its various materialisations, (iii) a historical justification for their emergence, (iv) an evaluation of some previous alternatives and (v) a conclusion bringing together the various strands.

### 9 *Björn Köhnlein: Metrically Conditioned Pitch Accent in Uspanteko*

Uspanteko, a Mayan language spoken in Guatemala, shows a remarkably rich interaction between the location of stress, vowel quality, syllable weight and pitch accent. Commonly, it is assumed that the language has privative lexical tone. Counter to previous analyses of the facts, this chapter proposes that both tonal contrasts and other relevant interactions can be derived from an opposition between trochaic and iambic feet. No tonal information is stored in the lexicon. While improving the empirical coverage of previous analyses with lexical tone, the current analysis adds little additional machinery, since the general distinction between trochees and iambs in Uspanteko has already been motivated on independent grounds. From a broader theoretical perspective, the chapter contributes to ongoing discussions on the phonological nature of tone-accent systems, one of the key issues in debates on prosodic typology.

### *10 Haruo Kubozono: Focus Prosody in Kagoshima Japanese*

This chapter discusses focus prosody in Kagoshima Japanese, a dialect spoken in the south of Japan with a lexical prosodic system remarkably distinct from that of standard Tokyo Japanese. Starting with the phenomenon of question particle incorporation in direct (matrix) Wh-questions, Kubozono considers several apparently different phenomena involving the incorporation of sentence-final particles into the sentence-final prosodic phrase. He proposes that all these phenomena can be generalized as manifestations of focus prosody whereby sentential particles are incorporated into the domain of prosodic phrase in post-focal positions. This process can be attributed to the Obligatory Contour Principle (OCP) whereby sequences of High tones are avoided by a dissimilatory process of H tone deletion and a subsequent process of prosodic rephrasing. This analysis nicely explains why particle incorporation fails to occur in a particular accent class of words as well as in yes/no questions and other syntactic constructions. The chapter also argues that post-focal prosodic incorporation is not an isolated phenomenon in Kagoshima Japanese, but rather, that similar prosodic (re)phrasing phenomena are found in other dialects and languages such as standard Tokyo Japanese, Fukuoka Japanese and the south Kyungsang dialect of Korean.

### *11 Björn Köhnlein and Marc van Oostendorp : Where is the Dutch Stress System? Some New Data*

There is an extensive body of theoretical work on the Dutch stress system, which is, however, mostly built on inspection of existing, sometimes rather exotic, words as found, for example, in dictionaries. In this chapter, Köhnlein and van Oostendorp confront these theories with new data from two online experiments in which participants had to indicate the most likely location for stress in biblical names or nonsense words respectively. The results partially confirm the claims already gathered in earlier work: there is a strong preference for stress on penultimate syllables and quantity plays a role in establishing the preferred location of stress. On the other hand, the way in which quantity works out is slightly different from what previous literature suggests. Furthermore, the authors report that they have not found strong evidence for the so-called three-syllable window.

### *12 Nicholas Rolle and Marine Vuillermet: Morphologically Assigned Accent and an Initial Three-Syllable Window in Ese'eja*

In this chapter, Rolle and Vuillermet argue that Ese'eja demonstrates an unusual initial three-syllable window within which primary prominence must fall, a typologically rare type. Using a corpus of 2,000 elicited verb forms (Vuillermet 2012), the authors show that the position of prominence depends on syllable count and the type of morphologically assigned accent. They posit four types of this morphological accent: inherent transitive accent, dominant indexical accent, recessive accent with one set of tense/mood suffixes and rightmost-preserving accent with another. Further, tense/mood suffixes trigger the creation of iterative trochaic or iambic feet, which the authors capture using cophonology theory employing common OT constraints (Inkelas and Zoll 2007). The authors posit that iterative footing occurs with a leftmost constraint, resulting in primary accent falling on the first, second or third syllable, which is realized with primary prominence. Additionally, because iterative footing occurs prior to primary accent delegation, Ese'eja constitutes a true 'count system' challenging the Primary Accent First model (van der Hulst 1996, 1997, 2012). Finally, they argue that when morphological accent in Ese'eja is assigned outside the metrical window, the position of primary prominence falls on a rhythmically dependent position, termed 'rhythmic repair'. The authors contrast this to Kager's (2012) typology, showing that under these circumstances primary prominence surfaces on a default position within the metrical window, termed 'default repair'.

### *13 Alexandre Vaxman: A Scales-and-Parameters Approach to Morpheme-Specific Exceptions in Accent Assignment*

This chapter addresses the long-standing problem of morphologically conditioned exceptions in accent assignment. Vaxman introduces a new approach, called the Scales-and-Parameters (S&P) theory, a new parametric, non-metrical theory of word accent, which takes as a point of departure the PAF theory of van der Hulst (1996, 1997, 2012, inter alia). The S&P theory is shown to uniformly capture regular and exceptional accent locations both within a given system and across different types of systems in terms of a single accentual grammar. As the author claims, the proposed grammar accurately derives accent location in lexical accent systems with dominant suffixes and in phonological weight-sensitive systems in which certain morphemes violate the accent rule. A core proposal of the theory is to extend the notion of 'weight' to morphemes by treating their accent-attracting ability as 'diacritic weight' (rather than lexical accent). Vaxman shows that since weight is a scalar variable, it allows for novel types of weight scales, that is, those containing diacritic and/or phonological weight. Reference to such scales allows the S&P parameter system to correctly assign word accent and to account for morpheme-specific exceptions, as illustrated here with detailed case studies of Central Selkup and Eastern Literary Mari.

We hope to have shown that the chapters in this volume cover a range of typological and theoretical issues in the study of stress/accent. On the typological side, various authors are concerned with the array of word-prosodic types and their distinguishing phonetic properties. Of specific theoretical concern is the question of foot structure and how it is manifested in different ways in different systems. In addition, several chapters explicitly discuss the role of morphological structure, as well as of lexical marking of accent. The 13 chapters collected here present a lively testimony of a field of inquiry that shows progress, while also identifying questions of ongoing concern.

## The ST2 Book Series

Volume 1: Hulst, H. G. van der. (ed.) 2014. *Word Stress: Typological and Theoretical Issues*. Cambridge University Press.

Volume 2: Heinz, Jeffrey, Rob Goedemans and Harry van der Hulst. 2017. *Dimensions of Phonological Stress*. Cambridge University Press.

Volume 3: Goedemans, Rob, Jeffrey Heinz and Harry van der Hulst. (eds.) 2018. *The Study of Word Stress and Accent: Theories, Methods and Data*. Cambridge University Press.

## References

Donegan, P. J. and D. Stampe. (1983). Rhythm and the holistic organization of language structure. *Papers from the Parasession on the Interplay of Phonology, Morphology and Syntax*. Chicago Linguistic Society, 337–353.

Goedemans, R., J. Heinz and H. van der Hulst. (2014). StressTyp2. http://st2.ullet.net.

Gordon, M. (2002). A factorial typology of quantity insensitive stress. *Natural Language and Linguistic Theory* 20: 491–552.

Gordon, M. and H. van der Hulst. Forthcoming.Word stress. In C. Gussenhoven and C. Aoju (eds.) *The Handbook of Prosody*. Oxford University Press.

Haspelmath, M., M. S. Dryer, D. Gil and B. Comrie. (eds.) (2005). *The World Atlas of Language Structures*. Oxford University Press.

Heinz, J., R. W. N. Goedemans and H. G. van der Hulst. (eds.) (2016). *Dimensions of Phonological Stress*. Cambridge University Press.

Heuven, V. J. van and A. M. C. Sluijter. (1996). Notes on the phonetics of word prosody. In R. Goedemans, H. van der Hulst and E. Visch (eds.) *Stress Patterns of the World, Part 1: Background*. The Hague: Holland Academic Graphics, 233–269.

Hulst, H. van der. (1996). Separating primary accent and secondary accent. In R. Goedemans, H. van der Hulst and E. Visch (eds.) *Stress Patterns of the World, Part 1: Background*. The Hague: Holland Academic Graphics, 1–26.

(1997). Primary accent is non-metrical. *Rivista di Linguistica* 9: 99–127.

(1999a). Word accent. In H. van der Hulst (ed.) *Word Prosodic Systems in the Languages of Europe*. Berlin and New York: Mouton de Gruyter, 3–116.

(ed.) (1999b). *Word Prosodic Systems in the Languages of Europe*. Berlin and New York:Mouton de Gruyter.

(2010). Word accent: Terms, typologies and theories. In H. G. van der Hulst, R. Goedemans and E. van Zanten (eds.) *Stress Patterns of the World, Part 2: The Data*. Berlin: Mouton de Gruyter, 3–54.

(2011). Pitch accent systems. In M. van Oostendorp, C. Even, E. Hume and K. Rice (eds.) *The Blackwell Companion to Phonology: Vol. 2*. Chichester: Wiley-Blackwell, 1003–26.

(2012). Deconstructing stress. *Lingua* 122: 1494–1521.

(2014a). The study of word accent and stress: Past, present and future. In H. van der Hulst (ed.) *Word Stress: Theoretical and Typological Issues*. Cambridge University Press, 3–55.

(2014b). Representing rhythm. In H.G. van der Hulst (ed.) *Word Stress: Theoretical and Typological Issues*. Cambridge: Cambridge University Press, 325–65.

(ed.) (2014c). *Word Stress: Typological and Theoretical Issues*. Cambridge University Press.

(2017). Phonological typology. In A. Y. Aikhenvald and R. Dixon (eds.) *The Cambridge Handbook of Typological Linguistics*. Cambridge University Press, 39–77.

Hulst, H. G. van der, R. Goedemans and E. van Zanten. (eds.) (2010). *A Survey of Word Accentual Patterns in the Languages of the World*. Berlin: Mouton de Gruyter.

Inkelas, S. and C. Zoll. (2007). Is grammar dependence real? A comparison between cophonological and indexed constraint approaches to morphologically conditioned phonology. *Linguistics* 45: 133–71.

Kager, R. (2012). Stress in windows: Language typology and factorial typology. *Lingua* 122: 1454–93.

Martínez-Paricio, V. (2012). Superfeet as recursion. In N. Arnett and R. Bennett (eds.) *Proceedings of the 30th West Coast Conference on Formal Linguistics* (WCCFL 30). Somerville, MA: Cascadilla, 259–69.

(2013). An exploration of minimal and maximal metrical feet. PhD dissertation, CASTL, University of Tromsø.

Martínez-Paricio, V. and R. Kager (2015). The binary-to-ternary rhythmic continuum in stress typology: Layered feet and non-intervention constraints. *Phonology* 32(3): 1–46.

Plank, F. (1998). The co-variation of phonology with morphology and syntax: A hopeful history. *Linguistic Typology* 2: 195–230.

Roca, I. (2006). The Spanish stress window. In F. Martínez-Gil and S. Colina (eds.) *Optimality Theoretic Studies in Spanish Phonology*. Amsterdam: John Benjamins, 239–77.

(2014). Dogma and fact: Spanish stress 30 years on (a personal history). Paper read at the Workshop on Word Stress and Accent, Leiden, 15–17 August.

(2016). Gliding ghosts or ghostly glides, and does it matter which? In R. A. Nuñez Cedeño (ed.) *The Syllable and Stress: Studies in Honor of James W. Harris*. Berlin and New York: Mouton de Gruyter, 51–104.

Vuillermet, M. (2012). A grammar of Ese Ejja, a Takanan language of the Bolivian Amazon. PhD dissertation, Université Lumière Lyon 2. www.marinevuillermet.com/these/complete-thesis/

# *Part I*

# Phonetic Correlates and Prominence Distinctions

# 1 Acoustic Correlates and Perceptual Cues of Word and Sentence Stress

## Towards a Cross-Linguistic Perspective

*Vincent J. van Heuven*

## 1 Introduction

### *1.1 Stress at the Word Level*

The languages in the world can be divided roughly into two types of word-prosodic systems. One type, probably a minority, has tone.[1] A tone language uses different pitches or melodies to differentiate between words in the lexicon, just as the vowels and the consonants do. The second type, which is the type that we address in the present chapter, has stress. When a language has stress, every word has one syllable which in some sense is more important, or more prominent, than any other syllable in the same word. This is also the crucial difference between tone and stress. In a tone language there is no difference in prominence attached to the syllables that make up the word, whereas stress is a culminative property: only one syllable can be the strongest (the prosodic head) within a constituent – such as a word.[2]

Which syllable is the prosodic head of the word is often predictable. For languages with fixed stress there is just one single rule that determines the position of the word stress for the entire lexicon. Hungarian words, for instance,

---

[1] The *World Atlas of Linguistic Structures* (WALS; Comrie et al. 2005) lists 220 tone languages versus 307 no-tone languages (Chapter 13); at the same time it lists 502 stress languages, divided in chapter 14 between 282 with fixed stress (281 in chapter 15) versus 220 with non-fixed stress (219 in chapter 15). Van Zanten and Goedemans (2007: 64) estimate that languages with stress-based word prosody, tone-based systems and languages without word prosody occur in 80, 16 and 4 per cent of the world's languages, respectively.

[2] It appears that the two types of word prosody are mutually exclusive. A language has either tone or stress but not both. To be true, there are so-called restricted tone languages in which only one syllable (i.e. the stressed syllable) in a word may carry different tones (e.g. Swedish, Norwegian) but languages that freely combine stress and tone are highly exceptional and seem to arise only as the result of accidental contact between a stress language and a tone language such as happened in the case of Samate Ma'ya (Remijsen 2002) and Papiamentu (Remijsen and van Heuven 2005).

always have stress on the initial syllable; in Weri (a Papuan language; Boxwell and Boxwell 1966) the stress is always on the last syllable of the word. Other languages may have more complex rule systems for assigning stress to words. In weight-sensitive languages such as English, German and Dutch, the complexity of (the rhyme portion of) the syllables determines where the stress goes, at least in monomorphemic words. For instance, stress in Dutch simplex words goes to the final syllable if it is superheavy (i.e. contains more than two morae in its rhyme); if not, stress goes to the pre-final syllable if this syllable is at least heavy (has two morae in the rhyme portion). It has been estimated that a relatively small portion of the monomorphemic lexicon is stressed by exception, that is, deviates from the weight-sensitive stress assignment (e.g. 15 per cent exceptions in Dutch; Langeweg 1988). The exceptions would be cases of unpredictable (or 'lexical') stress. In some languages, there are so many exceptions to any regularity one might want to formulate that stress rules do not make sense. Russian and Greek are often cited as examples of such lexical-stress languages.

Linguistically speaking, the inventory of stressed syllables in a language is richer (i.e. with a greater diversity of segmental structures) than that of unstressed syllables (see, for instance, the counts for Swedish (and four other languages) by Carlson et al. 1985 and for Dutch by van Heuven and Hagman 1988). Moreover, stressed syllables typically resist deleting or assimilating segments to neighbouring unstressed syllables, and whereas unstressed syllables tend to assimilate to adjacent stressed syllables, are susceptible to weakening processes and deletions. In this chapter we will not, however, be concerned with the linguistic properties of stressed syllables. The focus of interest will be on the phonetic realization of stress at the word and sentence level.

## *1.2 Stress at the Sentence Level*

Prosody is hierarchically structured. Where one syllable is the prosodic head of the word domain, one word will be the prosodic head of the phrase or utterance it occurs in. Typically, when a word receives sentence stress, the marking of this stress will fall on the syllable within the word that carries the word stress. A syllable in a word with sentence stress has all the phonetic markers of a word stress plus some characteristics that mark it as a sentence stress. Which words in an utterance receive sentence stress and which ones do not depends on the syntax-prosody interface of the language. In Romance languages, for instance, the location of the sentence stresses is largely, if not fully, determined by the syntactic structure of the utterance. In Spanish, the sentence stress (indicated by capitals) will invariably be on the nouns in (1) even though the pragmatic contrast (indicated by square brackets) is in the prepositions (Ladd 1996):

(1) ¿Quiere café [con] LEche o café [sin] LEche?
‘Require-you coffee [with] MILK or coffee [without] MILK?’

In other languages, such as those in the Germanic family, sentence stresses are assigned by default to specific words on the basis of the syntactic/prosodic structure of the utterance, but the default rules may be overridden by pragmatic considerations that delete or move sentence stresses so as to express the focus structure of the utterance. Typically only the prosodic head of a prosodic constituent that is in focus, that is, contributes new and contextually unpredictable information to the discourse, receives sentence stress, whereas sentence stresses are deleted (or moved away) from words and phrases that are out of focus, that is, contain relatively unimportant and contextually given information. Thus, in (2a) there is a contrast between two phrases: *the girl* and *the old man*. By default, sentence stress in the latter phrase goes to the noun, which is the prosodic head of the NP. In (2b), however, the pragmatic contrast is between the adjectives *young* and *old*. In this situation pragmatic rules delete the default sentence stress from the noun and reassign it to the adjective.

(2a) Is Lesley [the GIRL] or [the old MAN]?

(2b) Is Lesley the [YOUNG] man or the [OLD] man?

## *1.3 Acoustic Correlates and Perceptual Cues*

The purpose of the present chapter is to present and discuss the way word and sentence stress are phonetically marked. It has been known since the 1950s that stress (whether at the word or sentence level) is never marked by a single acoustical property (for a survey see Lehiste 1970). To make the stressed syllable stand out from its neighbours, it is produced with greater physiological effort on the part of the speaker than its unstressed counterpart (e.g. Ladefoged 1967). The greater effort will be exerted at any stage in the speech production process, that is, by the subglottal mechanism (more air is pushed out of the lungs), by the glottal (laryngeal) system (contraction of laryngeal muscles, generating a change in pitch) and by the supraglottal organs (e.g. larger and faster displacement of lips, tongue and jaw, yielding more clearly articulated vowels and consonants). The greater effort is seen, first of all, in closer approximation of articulatory target configurations for segments in stressed syllables. More extreme articulatory movements require more time than small displacements of the vocal organs. The result of this is that segments in stressed syllables have longer durations – all else being equal – than unstressed segments.[3] Moreover, in terms of the theory of articulatory phonology (e.g. Browman and Goldstein 1992), there is

---

[3] This view on the relationship between expansion of the articulatory space and duration goes back to Lindblom (1963). Given that longer duration is not associated with clear articulation in

relatively little overlap between adjacent segments in a stressed syllable. In contradistinction to this, unstressed segments greatly overlap, which leads to considerable reduction of segmental contrast. This also accounts elegantly for the observation that segments at the edges of stressed syllables tend to maintain their identity (resist coarticulation with an adjacent segment in an unstressed syllable) whilst unstressed segments across the syllable boundary are disproportionally affected by coarticulation (e.g. Dogil and Williams 1999).[4]

Effort expended at the laryngeal level of speech production takes the form of contracting selected muscles that influence the speed with which the vocal folds vibrate during phonation. The result may be a rapid increase (through activation of cricothyroid and vocalis muscles) or decrease (through activation of the sternohyoid muscle) of the repetition rate of the glottal cycle, causing, respectively, a rise and fall of vocal pitch. A secondary effect of laryngeal effort may be a tightening of the vocal folds (musculi vocales), which will then snap together more forcefully than when in a less tightened state. Finally, increased effort at the subglottal level will push more air per unit of time through the glottis, causing, first of all, an increase in intensity of the sound produced by the glottal siren. Secondarily, the greater volume-velocity of the airstream through the glottis boosts the Bernoulli suction effect. The increased suction and the tightening of the vocalis muscles conspire to shorten the closing phase of the glottal cycle, which causes the spectrum to become flatter (boosting the intensity of higher harmonics, thereby generating a louder sound – I will come back to this later).

In this chapter we will not deal any further with the physiological basis of stress (but see Erickson and Kawahara 2016 for a well-documented survey of current issues). We will concentrate on the acoustic consequences of increased versus decreased effort (as foreshadowed in the above) and ask (i) what acoustic correlates can be found for the difference between a stressed syllable and its unstressed counterpart, and (ii) what the relative importance is of each acoustic correlate in the marking of stress. At the same time we will consider the question of what acoustic properties are used by human listeners and to what extent these are used to decide whether or not a syllable is stressed. We will make a strict terminological distinction here between acoustic correlates of stress (which can be used, for instance, to identify a stressed syllable by some computer algorithm) and the perceptual cues used by the human listener. We will see that some acoustic correlates, notably the (peak) intensity of

---

phrase-final syllables (domain-final lengthening), it seems reasonable to assume that clarity of articulation is the primary goal which is subserved by lengthening in stressed syllables.

[4] The coarticulation window of a stressed vowel may extend to an unstressed neutral vowel schwa in the preceding syllable, across an intervening consonant, and yield both acoustic and perceptual effects (e.g. Van Heuven and Dupuis 1991).

a syllable, allow excellent separation of stressed from unstressed tokens but are hardly used by the human listener.

There is no need, a priori, for the three subsystems of speech production to expend extra effort on the production of a stressed syllable in equal proportion. We may speculate, in fact, that languages differ in the way they exploit effort in each subsystem. For instance, Germanic languages seem to exploit the gradation of supralaryngeal effort more than Romance languages do. More generally, we will ask to what extent the acoustic correlates and perceptual cues of stress have the same ranking order across languages or are differently ordered from one language to the next. If the latter should be the case, then we may ask the supplementary question if the order of importance of correlates and cues can be predicted from the phonological structure of the language at issue.

## 2 Acoustic Correlates

### *2.1 Some Methodological Considerations*

When trying to find acoustic correlates of stress, it is generally not a good idea to just compare acoustic properties of successive syllables in a word. If the segmental make-up of the syllables is different, the correlates of stress are obscured by the intrinsic and co-intrinsic properties of the segments. For instance, open vowels have inherently greater intensity (Lehiste and Peterson 1959) and longer duration than close vowels (Peterson and Lehiste 1960), so that an unstressed open vowel may, in fact, seem more stressed than a closed stressed vowel, as may happen in the English noun ***im**pact*. Several tricks have been suggested to eliminate, or correct for, such inherent segmental properties. One is to run some extrinsic normalization procedure by which the intensity or duration of a segment is expressed in standard deviations away from the mean value of that segment (i.e. z-normalization) as produced by the individual speaker in a larger corpus of materials (e.g. Potisuk, Gandour and Harper 1996).[5] Another way out would be to use so-called reiterant speech (Larkey 1982, Liberman and Streeter 1978, Nakatani and Shaffer 1978). In this speech mode the speaker replaces the syllables in a target word by repetitions of the same segmental structure; for example, repetitions of /ma/ or /lɪs/. For instance, the target utterance *please say **im**port again* would be produced as *please say*

---

[5] Extrinsic normalization expresses the location of an object relative to all other objects in a dataset. Z-normalization is a typical example of extrinsic normalization. This is in contrast to intrinsic normalization, which does not compare values across tokens in the dataset but computes relationships (ratios, differences) between variables obtained for a single token. The $V_1/V_2$ ratio computed by Fry (1955) would be an example of intrinsic normalization. For a discussion of normalization procedures see Nearey (1978).

***ma**ma again*, or *please say **lis**lis again*. The claim is that the speaker dubs all (and only) the prosodically relevant variations onto the reiterant version of the original utterance so that no normalization for intrinsic segmental differences is needed. A potential problem with these techniques is that stressed and unstressed syllables are compared syntagmatically, that is, in different linear positions in a larger structure, such as an initial stressed and a final unstressed syllable – so that, strictly speaking, the researcher does not know whether he measures correlates of stress or of sequential position. The safest precaution, therefore, would be to compare stressed and unstressed versions of the same syllables in a paradigmatic way; for instance, by comparing the stressed and unstressed realizations of the first and second syllables in a minimal stress pair such as *the import* versus *to import*. This solution, of course, can only be used if the language has at least one minimal stress pair – which means that it cannot be used in languages with fixed stress.[6]

It has also been found expedient to measure the correlates of stress separately for stress at the word level and at the sentence level. This is generally achieved by (paradigmatically) comparing tokens of stressed and unstressed syllables in a minimal stress pair which was produced in the same position in a surface-syntactically identical sentence with and without focus on the target. Focus on the target word, indicated in (3a–c) in square brackets, is often manipulated by having the speaker answer different questions that highlight one constituent or the other as in (3a–c):

(3a) Q: who borrowed a chainsaw?
A: [OScar] **bor**rowed a **chain**saw
(3b) Q: what did oscar borrow?
A: **os**car **bor**rowed [a CHAINsaw]
(3c) Q: did oscar buy a chainsaw?
A: (no,) **os**car [BORrowed] a **chain**saw

The recordings now contain tokens of the words *Oscar, borrow* and *chainsaw* produced with and without sentence stress, which can be directly compared: any difference between the readings must be the consequence of presence versus absence of sentence stress. The difference between stressed and unstressed syllables in the tokens that are produced without sentence stress (out of focus) will then be a matter of word stress only (indicated by bolded small capitals). Examining the effects of word and sentence stress in a single experimental setup using minimal stress pairs can only be achieved by using

---

[6] Strictly speaking, correlates of stress can be investigated only in a language with non-fixed stress. In languages with fixed stress, such as Hungarian, where every word has stress on the first syllable, it is impossible to separate correlates of stress from word boundary effects.

highly contrived contexts, for instance, with target words used metalinguistically (as citation forms), as in (4a–d):[7]

(4a) Q: did you read ‘the import’ or ‘the sale’ again?
A: i read [‘the IMport’] again
(4b) Q: did you read ‘to import’ or ‘to sell’ again?
A: i read [‘to imPORT’] again
(4c) Q: did you read ‘the import’ again or write it down?
A: i [READ] ‘the **IM**port’ again
(4d) Q: did you read ‘to import’ again or write it down?
A: i [READ] ‘to im**PORT**’ again

We will now briefly review what has been reported in the literature on the acoustical marking of word and sentence stress. I will draw on publications on Dutch and English but occasionally digress to other languages. We will begin by discussing properties that are found equally in word and sentence stress and finish by zooming in on those properties that differentiate word from sentence stress (and are found, therefore, only when a syllable occurs in a word with sentence stress).

### *2.2 Acoustic Properties of Word Stress*

*2.2.1 Temporal Organization* Since the work by Fry (1955) it has been clear that stressed syllables – all else being equal – are longer than their unstressed counterparts. Fry measured the duration of the first and second vowels ($V_1$ and $V_2$) in five English minimal stress pairs (noun-verb pairs *contract, digest, object, permit* and *subject*) spoken once by 12 American speakers in sentence-final position in a fixed carrier *Where is the accent in . . .*, which elicits sentence stress on the target words.[8] With the duration of $V_1$ and $V_2$ as predictors, a Linear Discriminant Analysis (Klecka 1980), a classification algorithm often used for this purpose, yields correct classification of stress pattern in 83 per cent of the cases.[9, 10] After z-normalizing $V_1$ and $V_2$ duration within minimal stress pairs, the percentage of correct classification

---

[7] A very clever but elaborate way of obtaining minimal stress pairs in English with and without sentence stress was used by Huss (1978).

[8] The information on the context sentence can be found in Fry (1958: 135).

[9] This information is not found in the original paper but computed by me (VH). Fry (1955) provides the raw measurements of vowel durations and peak intensities in an appendix. This appendix contains one obvious error in that the duration values for the noun and verb reading of the target word *contract* have been switched (this error becomes apparent when the data are checked against the plot of $V_1$ against $V_2$ for *contract* in Fry’s Figure 2 (left panel), which shows the correct durations).

[10] In meta-analyses of the type performed here, it is more customary to quantify the importance of a parameter in terms of effect size. Effect size can be interpreted as the degree of overlap between two samples, for example for words with initial stress versus final stress. The smaller

of stress pattern increases to 93. Using Fry's data, we may apply intrinsic normalization by computing the relative duration of the first vowel ($V_1$%) as a percentage of the summed durations of $V_1$ and $V_2$. Comparing the $V_1$ percent values for each of Fry's 60 minimal stress pairs, we find just one single case in which $V_1$% was the same for the noun and the verb reading of the pair; in all other 59 cases $V_1$% was larger for the noun (initial stress) than for the verb (final stress) reading (98% correct classification). The conclusion was that vowel duration (especially when expressed relatively within a token) is a very good correlate of stress. Fry (1955: 765), however, remarks that consonant duration ratios were 'not materially affected by the shift of stress'. Since word stress is generally believed to be a property of a syllable, this conclusion deserves further scrutiny. I turn to data on Dutch to examine effects of stress on subsyllabic units, that is, vowels, onset and coda consonants, separately.

An early study that examined the effect of stress on the durations of subsyllabic units in Dutch can be found in Nooteboom (1972: appendices 11–12). Target items were non-words /pɑpɑpɑp/ and /papapap/, with short/lax /ɑ/ and long/tense /a/, respectively. These items were spoken with stress on the first, second and third syllable in turn in carrier sentences such that they were either 'accented' (with sentence stress) or 'unaccented' (word stress only). A large number of tokens were produced by each of two male Dutch speakers for each of the 3 (stress position) × 2 (accentuation) × 2 (vowel length) = 12 non-word types (between 17 and 26 tokens per type by speaker SG; between 12 and 24 by speaker IS). Duration of all plosives /p/ in positions $C_1$ to $C_4$ were measured physiologically (rather than acoustically) using electronic switches that were activated by lip contacts, as were the durations of the vowels in $V_1$, $V_2$ and $V_3$. A summary of the results is seen in Figure 1.1. This figure plots the segment durations, in milliseconds (ms), of $C_1$, $V_1$, $C_2$, $V_2$, $C_3$, $V_3$ and $C_4$, in this order, along the X-axis, with separate lines for items with initial, medial and final stress. The four panels are arranged by vowel length (rows) and by accentuation (columns).

The relative effects of stress on the temporal make-up of the non-words are very similar for accented and unaccented items – although durations are consistently longer overall under sentence stress. Hardly any effects of stress

---

the overlap, the fewer the number of classification errors. When the distribution of values for two (equally large) samples overlap completely, the percentage of classification errors will be 50; when there is zero overlap, there will be no classification errors. Cohen's *d* is commonly used as a measure of effect size. It is basically a z-score: if $d = 1$, then the mean of sample A is one time the common standard deviation away from the mean of sample B; there is considerable overlap between the distributions yielding classification errors (assuming equal variance in the samples) in 27.7% of the cases. For $d = 2$, the overlap is much smaller, yielding less than 5% classification errors (Cohen 1988). When the variance in the two samples is not uniform, more sophisticated procedures have to be applied, such as an LDA.

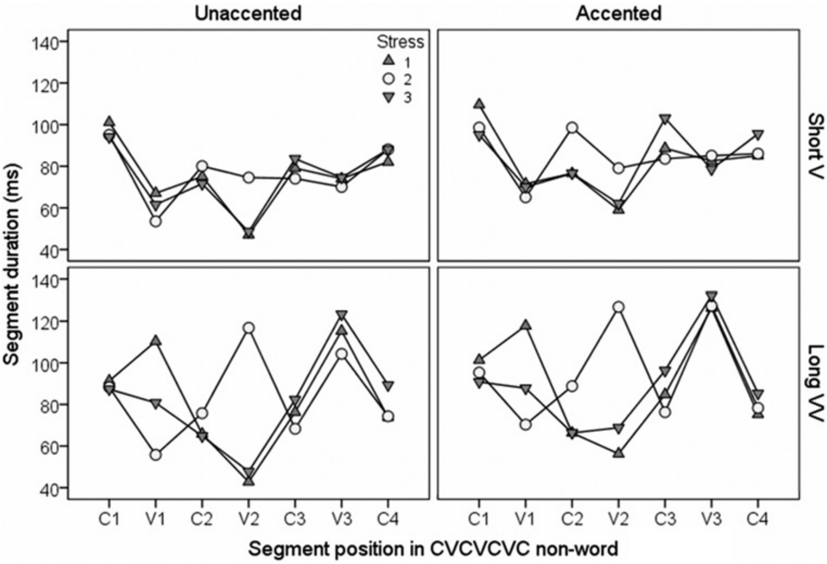


Figure 1.1 Duration (ms) of seven segments in the sequence /pVpVpVp/ as a function of stress position (initial, medial, final) in accented versus unaccented non-words with short (lax) and long (tense) vowels (data from Nooteboom 1972, appendices 11–12).

can be seen in the final syllable.[11] There are very large differences in the durations of $V_1$ and $V_2$ depending on the stress position. When the item is spoken with initial stress, $V_1$ is very long and $V_2$ short (ratio $V_1/V_2 > 1$). With medial stress, this pattern reverses completely, with a very short $V_1$ and a very long $V_2$ (ratio $< 1$), while items with final stress have intermediate vowel durations for $V_1$ and $V_2$ (ratio $\approx 1$).[12] The crucial observation, however, is that the effect of stress position on the duration of the consonant segments, though small in absolute terms, appears to be quite consistent as well: it is nearly always the case that a C, whether onset or coda, is somewhat longer on

---

[11] Segments in a word-final syllable in Dutch are affected by domain-final lengthening but will not be lengthened any further when stressed. Unlike what happens in English, the effects of stress and final lengthening in Dutch are therefore not additive (Cambier-Langeveld and Turk 1999).

[12] The difference between initial and medial stress is very clearly marked by the $V_1/V_2$-ratio. The difference between initial and final stress is less clearly marked, especially when the vowels are lax/short. The appendices in Nooteboom (1972) do not contain data on individual tokens (only means and number of tokens). Therefore no meaningful effect sizes can be computed.

average in the stressed version of the syllable than in the unstressed version (i.e. in a paradigmatic comparison).[13]

An experiment on a smaller scale involving both words and reiterant nonwords in Dutch shows that the lengthening effect of stress is most clearly and consistently seen in the rhyme portions of the syllables (Sluijter and van Heuven 1995). The effect of stress on onset consonants is less systematic or absent.

*2.2.2 Intensity* The intensity of the sound pressure wave has long been considered as an acoustical correlate of stress. Intensity (or sound pressure) is proportional to the square of the amplitude of the speech waveform averaged over a moving time-window that is long enough to include two glottal pulses (typically with an integration time of 20 ms for the male voice range and 10 ms for a female voice). Absolute intensity is expressed in Watts per square inch (or dynes per cm$^2$). However, since in speech we are not so much interested in absolute sound pressures as in relative differences between sound pressures, intensities are usually expressed in decibels (dB). When two intensities differ in terms of Watts by a 1:10 ratio, the stronger of the two has a 20 dB greater relative intensity; when the power ratio is 1:100, the relative intensity difference is 40 dB; and when the ratio is 1:1000, the difference is 60 dB. So each time the absolute intensity difference is multiplied by 10, there is a 20 dB increase in intensity. The perceptual span between the weakest sound pressure that can be detected in silence (the threshold of hearing, axiomatically set at 0 dB) and the strongest sound pressure that can be tolerated without crossing the pain threshold is 120 dB. Generally, the dynamic range of a spoken utterance is rather restricted, somewhere in between 55 and 75 dB above the threshold of hearing. When screaming, intensity levels rise to some 85 dB, and by whispering low intensities in the 40 to 55 dB range are afforded.

Intensities of speech sounds are unstable as they vary considerably (intensity drops in the order of 5 dB) when the speaker inadvertently turns his head or when some object momentarily intervenes between the speaker's mouth and the listener's ears. Intensity differences of similar magnitude have commonly been reported as correlates of stress. These differences are small but prove reliable correlates (i.e. with little variability) of sentence stress but are even smaller and less reliable when word stress is signalled

---

[13] For each of the four consonant positions $C_1$, $C_2$, $C_3$ and $C_4$, 16 paradigmatic comparisons between stressed and unstressed conditions can be made in Nooteboom's appendices. In each of these four positions the stressed version of C is longer than its unstressed counterpart in 15 conditions. A similar count for the vowels in positions $V_1$, $V_2$ and $V_3$ yields longer stressed than unstressed values in 13, 16 and 10 out of 16 comparisons. In this sense, consonant durations are at least as accurate as correlates of stress as are vowels.

(cf. Lea 1977, Beckman 1986 for English; van Katwijk 1974, Rietveld 1984, Sluijter 1995, Sluijter and van Heuven 1996a for Dutch). In all these (and other) studies, peak intensity was measured, which is usually reached shortly after the vowel onset. Lea (1977) and Beckman (1986) suggested alternative correlates of accent, viz. the *intensity integral* (the summation of intensities throughout the stressed vowel) or *average intensity* (as the preceding but normalized for vowel duration). The intensity integral proved a very stable correlate of stress, but it should be pointed out that the intensity and duration correlates are conflated here into one complex cue. Obviously, the combined correlate will be more successful than either of its components. As a general rule, we advocate the use of multiple simplex correlates rather than singular complex indexes as the latter obscure whatever systematic interactions exist among the component correlates.

Since open vowels have more intrinsic intensity than close vowels (see Section 2.1), using raw peak intensity as a direct correlate of stress is rather pointless. In a paradigmatic comparison, that is, comparing the stressed and unstressed reading of the same vowel in the same position in minimal stress pairs (as in Fry 1955), the stressed version had more decibels than the unstressed counterpart in 52 out of 60 $V_1$ pairs and in 55 $V_2$ pairs. Note that the decibel is a logarithmic measure, so that the difference (obtained by subtraction) rather than a ratio (obtained by division) between the (peak) intensities of two vowels (e.g. in a stressed syllable and in an unstressed counterpart) is used here as the correlate of stress. Moreover, it is nearly always the case that the intensity difference between $V_1$ and $V_2$ was more positive in the noun reading (with stress on $V_1$) than in the corresponding verb reading (with stress on $V_2$). Out of 60 comparisons, 58 behaved as predicted, in one case the relationship was reversed and in one more the noun and the verb reading had the same intensity difference between $V_1$ and $V_2$. This makes (peak) intensity, and especially the intensity difference between stressed and unstressed syllables, a very reliable acoustic correlate of stress in English. It should be pointed out in this context that Fry (1955) is often misquoted. It is not the case that his data show that intensity is a poor *acoustic* correlate of stress or that it is a poorer correlate than duration.

*2.2.3 Spectral Balance* Accent in Western Germanic languages has often been equated with the expenditure of vocal effort, which is correlated with perceived loudness. The most obvious acoustic correlate of physiological effort and perceived loudness, it was held, is vocal intensity. As was explained in Section 1.3, increased pulmonary effort causes a larger volume-velocity of airflow through the glottis. The result is not just the generation of larger glottal pulses but also, and more importantly, of a more strongly asymmetrical glottal pulse (Figure 1.2). Typically, the closing phase of the glottal period is

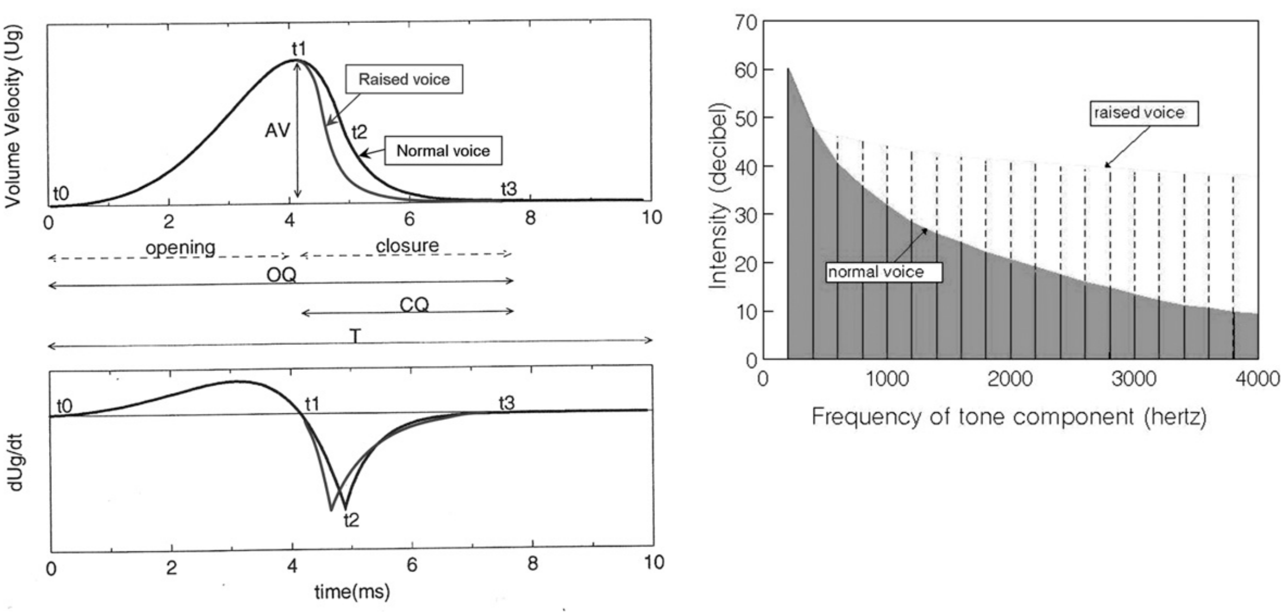


Figure 1.2 Effect of normal versus raised voice on volume-velocity of airflow through glottis (top left) and its first derivative (bottom left). The right-hand panel shows the effect of decreased Open Quotient (OQ) and Closure Quotient (CQ) due to raised voice on the spectral envelop (difference is exaggerated). t1: maximum flow during glottal cycle, t2 fastest decrease of glottal flow, t3 complete glottal closure (no flow). Graphs are based on Sluijter (1995) and van Heuven (2001).

shortened, yielding a smaller opening quotient (the duty cycle of the glottal pulse, that is, the proportion of the time the glottis is open relative to the period duration), and the trailing edge of the glottal period is steeper. The greater steepness of the glottal closure, as well as its more abrupt ending, cause the generation of relatively strong higher harmonics in the glottal pulse. As a result, the spectral tilt of vocalic sounds produced with greater vocal effort emphasizes the higher frequencies. The spectral tilt of the glottal period produced with average effort has a –12 dB/octave roll-off.[14] When speakers (or rather, singers) were asked to produce sustained vowel sounds with great vocal effort, the spectral tilt proved less steep, due to the fact that there was a relative boost of frequencies between 500 and 2000 Hz (Gauffin and Sundberg 1989). It has been shown that a similar phenomenon can be observed during the production of local vocal effort, that is, during the production of a stressed syllable (Sluijter and van Heuven 1996a for Dutch; Sluijter et al. 1995 for American English; Fant and Kruckenberg 1995, Heldner 2003 for Swedish; Campbell 1995 for Japanese; see also Campbell and Beckman 1995, Sluijter 1995).

Measuring the spectral balance (or ‘tilt’) is not without problems. Ideally, one needs to strip away the influence of resonances brought about by cavities in the supraglottal tract from the vocal output radiated from the mouth, so that the spectrum of the unfiltered glottal waveform is recovered. Once a clean glottal spectrum is available, the spectral tilt is a matter of fitting a simple linear regression function through the harmonics (plotted along a logarithmic frequency axis), and measuring its slope coefficient in dB/octave. Undoing the resonance effects of the vocal tract is done by inverse filtering. Inverse filtering software is now readily available (e.g. Airas et al. 2005) but the routines are not included in more comprehensive speech-processing packages. In lieu of full-fledged inverse filtering, some fast-and-dirty approximations have been suggested by Stevens (1998) and were applied in earlier research (Sluijter 1995, Sluijter et al. 1995, Sluijter and van Heuven 1996b). When it is not necessary to know the absolute values of spectral tilt (e.g. when no comparison across different vowels is being made), a simpler approximation of spectral tilt is afforded by measuring intensity in four contiguous filter bands (one base filter 0–0.5 KHz, and three contiguous octave filters: 0.5–1 KHz, 1–2 KHz, 2–4 KHz, cf. Gauffin and Sundberg 1989, Sluijter 1995). A linear regression line fitted through the four intensity levels at the filter bands’ centre frequencies (plotted along a log frequency axis) yields the spectral tilt measure. In fact, we found that the intensity levels in the base and highest octave filter did not vary much as a function of accent level, so that a good substitute of spectral balance

---

[14] When vowel sounds are radiated from the mouth, some +6 dB/octave is added to the spectral slope, so that the spectral tilt of an average vowel equals –12 + 6 = –6 dB.

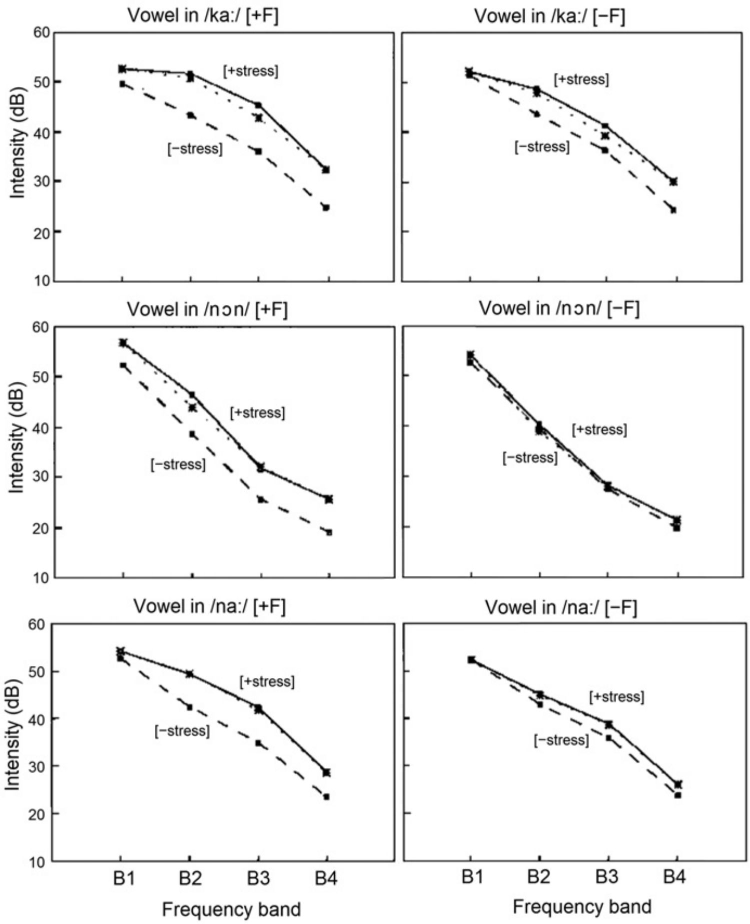


Figure 1.3 Effects of sentence (left-hand column) and word (right-hand column) stress on spectral tilt. Intensity (in dB) is plotted for four frequency bands (B1: <.5 KHz, B2: .5–1 KHz, B3: 1–2 KHz, B4: 2–4 KHz).

was obtained by just measuring mean vowel intensity (at the overall intensity peak) in the 0.5–2 KHz band (Sluijter 1995, Sluijter and van Heuven 1996a).

The effects of stress on spectral tilt at the sentence (left-hand column) and word level (right-hand column) can be seen in Figure 1.3 for a paradigmatic comparison of selected syllables in the Dutch minimal stress pair *canon* ~ *kanon* /ˈkanɔn ~ kaˈnɔn/ ‘round song ~ cannon’ and reiterant mimicry by five male and five female speakers.

Figure 1.3 shows that generally no effects of stress can be observed in the base band (< .5 KHz). Effects are strong in the higher frequency bands, causing flatter spectral tilt, especially under sentence stress, and more clearly so in the initial syllable than in the final syllable.

*2.2.4 Spectral Expansion* Stressed vowels have often been described as 'clear' (or, spectrally expanded), reflecting greater articulatory effort and precision. These vowels lack the spectral reduction that is characteristic of unstressed vowels. The acoustic consequences of vowel expansion and reduction can be examined by measuring the centre frequencies of the lowest two resonances of the vocal tract, the first and second formants, where $F_1$ (the lowest resonance) reflects degree of openness of the vowel and $F_2$ (the second-lowest resonance) reflects vowel backness and lip protrusion (i.e. the length of the oral cavity). Degree of vowel expansion is best expressed in terms of the Euclidean distance of a vowel away from the centre of the (acoustical) vowel space, which is defined by the mean value of $F_1$ and $F_2$ found for the individual speaker, when the speaker has produced an equal number of all the vowels in his language (under identical circumstances). For an average male speaker this will be an $F_1$ at 500 Hz and an $F_2$ at 1500 Hz.[15] Spectrally reduced vowel tokens will then be closer to the centre of the vowel space than their full or expanded counterparts.

An exemplary study of the effects of stress on vowel quality in Dutch was done by van Bergem (1993). In Dutch the acoustical effects of stress on vowel quality are particularly noticeable − maybe more so than in any other language. Figure 1.4 illustrates the effects of word and sentence stress on the expansion/reduction of the long (tense) Dutch vowels /eː, oː, aː/ read by 15 male speakers. The position of the schwa (averaged over 300 tokens across consonant environments and speakers) may serve as the centre of gravity of the vowel space. Spectral expansion is largest for vowels pronounced in isolation ('isol'). Some reduction is visible when these vowels occur in the stressed syllable of focally accented words ('+S+A'). Considerable reduction is observed for stressed vowels in unaccented words ('+S−A') or for unstressed vowels in accented words ('−S+A'). Severe spectral reduction is applied to the unstressed vowels

---

[15] Since the $F_1$ and $F_2$ values differ considerably from one speaker to the next, especially when the speakers have different gender, normalization is called for when individuals are compared. Z-normalization (Lobanov 1971) is generally seen as the most adequate option. The centre of the vowel space is then by definition at $F_1 = F_2 = 0$. Comparisons across larger numbers of speakers can be safely done without normalization. Often, formant frequencies are psychophysically scaled (through Bark or Mel conversion) so as to reflect properties of the human auditory mechanism, which is more sensitive to differences between low frequencies than to (physically equal) differences between high frequencies (for details, see introductory textbooks such as Hayward 2000 or Johnson 2003).

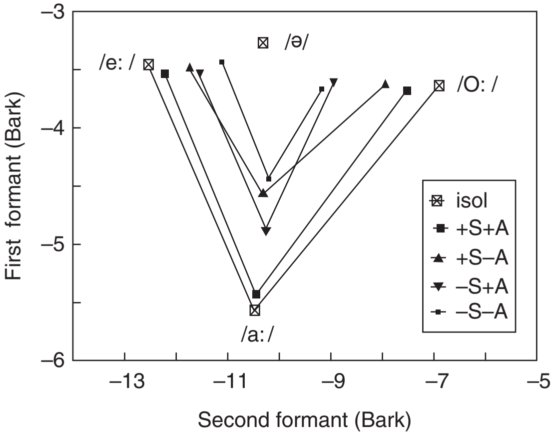


Figure 1.4 $F_1$ and $F_2$ (Bark) of three Dutch tense peripheral vowels produced by 15 male speakers in five stress conditions (after van Bergem 1993).

of unaccented words ('−S−A'): here the spectral distance to the centre of gravity /ə/ is minimal. Similar results were obtained for reiterant American-English non-words by Sluijter et al. 1995 (for details, see Sluijter 1995: 116–17, see also Section 2.3).

Automatic classification of stress by spectral expansion of Dutch vowels was done by Sluijter and van Heuven (1996a) in the minimal stress pair /ˈkanɔn ~ kaˈnɔn/ (see Section 2.2.3) and their reiterant versions (/nana/) produced in a short carrier with and without word and sentence stress (four combinations). Predictors in the LDA were the $F_1$ and $F_2$ of $V_1$ and $V_2$. Percentages of correct stress identification were 84 and 77 for words with and without sentence stress, respectively, and 68 and 71 for the reiterant non-words. These identification scores are better than chance (= 50%) but are poorer than what was observed for most other stress correlates (see following section).

*2.2.5 Resistance to Coarticulation* One characteristic of a spectrally expanded stressed syllable is that it shows minimal influence of coarticulation with abutting syllables, which in turn are strongly influenced by the adjacent stressed syllable. So properties of the stressed syllable are anticipated in the preceding syllable, and perseverate into the following syllable, but the stressed syllable itself is hardly influenced by the abutting unstressed syllables. Resistance to coarticulation was claimed to be the most important correlate of stress in Lithuanian by Dogil and Williams (1999; see also Pakerys 1982, 1987).

One way in which the mutual coarticulatory influence of abutting syllables can be quantified would be to locate the beginning and end of vowel-onto-vowel formant transitions (if the formants do not move in synchrony, study the behaviour of $F_2$ only) from the preceding syllable into the stressed syllable, and from the stressed into the following syllable (cf. Öhman 1967). Then determine the point along the time axis where half of the formant trajectory (i.e. half of the $F_2$ frequency difference between the consecutive vowels) from the stressed to the unstressed vowel (and vice versa) has been covered. The coarticulatory window of the stressed syllable is then expressed as the time interval between the preceding and following 50 per cent points divided by the duration of the stressed syllable. The larger the relative window size, the more resistant the syllable is to coarticulation. I am not familiar with published data on measurements of resistance to coarticulation.

## *2.3 Acoustic Correlates of Sentence Stress*[16]

Theories have been proposed in which there is no principled difference between word and sentence stress. In such views, for example, in American structuralism (Bloch and Trager 1942) and early Generative Phonology (Chomsky and Halle 1968, Halle and Keyser 1971), sentence stresses were seen as merely stronger degrees of stress along a continuum, where degrees of stress differ along all stress-related acoustic parameters in proportion. More recently, phonetic research has brought to light, however, that sentence stresses – used to place constituents in focus – are marked in a principally differently way from mere word stresses. Typically, as long as there is no sentence stress on a word, the speaker makes no effort to change the vocal pitch. To be true, there may well be a small rise–fall contour on any vowel (with or without word stress) but this is due to an involuntary response of the glottal mechanism to the greater transglottal pressure that comes about when the oral tract opens during the articulation of the vowel sound; during the articulation of consonants the oral tract is fully or partially closed so that intraoral impedance yields a transglottal pressure drop causing the vocal folds to vibrate more slowly. It has been estimated that the involuntary effect of mouth opening on the rate of vocal fold vibration does not normally exceed a threshold of four semitones (a frequency rise and subsequent fall of less than 25%). Only when a word is produced with sentence stress does the speaker issue a voluntary

---

[16] This section summarizes work done mainly on English and Dutch, with occasional excursions to other languages, concentrating on research methods and basic findings. In the last 25 years many more languages across the world have been studied using these or similar research methods. It is beyond the scope of this chapter to present a comprehensive overview of findings. The interested reader is referred to work by, for example, Hargus and Beavert (2005), Remijsen and Van Heuven (2006) and Gordon (2011).