Topicalization and Stress Clash Avoidance in the History of English

Topics in English Linguistics 69

Editors

Elizabeth Closs Traugott
Bernd Kortmann

De Gruyter Mouton

Topicalization and Stress Clash Avoidance in the History of English

by

Augustin Speyer

De Gruyter Mouton

ISBN 978-3-11-022023-0
e-ISBN 978-3-11-022024-7
ISSN 1434-3452

Library of Congress Cataloging-in-Publication Data

Speyer, Augustin.
Topicalization and stress clash avoidance in the history of English / by Augustin Speyer.
p. cm. − (Topics in English linguistics; 69)
Includes bibliographical references and index.
ISBN 978-3-11-022023-0 (alk. paper)
1. English language − Grammar, Historical. 2. English language − Syntax. 3. English language − Word order. 4. English language − History. I. Title.
PE1075.S64 2010
420'.9−dc22
2010002363

Bibliographic information published by the Deutsche Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Cover image: Brian Stablyk/Photographer's Choice RF/Getty Images
Printing: Hubert & Co. GmbH & Co. KG, Göttingen
∞ Printed on acid-free paper

Printed in Germany

www.degruyter.com

Preface

This book is a modified and slightly extended version of my PhD thesis which I submitted at the University of Pennsylvania in 2008. The basic ideas however even go back to my time as Visiting Scholar at the Linguistics Department of the University of Pennsylvania in Fall Term 2002. Anthony Kroch taught a seminar on historical syntax at that time in which I discovered the basic impetus for this work – the decline of topicalization. The other participants were Silvia Cavalcante, Daniel Ezra Johnson, Beatrice Santorini and Laura Whitton. Their comments helped me immensely in the early steps of this work.

The biggest thanks go, of course, to the supervisor of the thesis, Anthony Kroch, for his commitment in supervising this thesis. Many of the ideas expressed in this study go originally back to him. But the contribution of the other committee members, Eugene Buckley, Rolf Noyer, Donald Ringe and Jiahong Yuan cannot be valued too highly either, and I wish to thank them for their commitment. I am also grateful that I could discuss parts of the work at various stages with Werner Abraham, Brian McHughes, Ellen Prince, Marga Reis, Arnim von Stechow, Hubert Truckenbrodt and William Barry. Their comments were extremely helpful. All remaining errors are my own, of course.

As the project went on, it became necessary to conduct experiments in countries with a strong supply of German and English native speakers. The preliminary German experiment was conducted in the Phonetics Lab of the Universität des Saarlandes at Saarbrücken in summer 2004. My thanks go to William Barry who made it possible for me to use the equipment and who also lent a willing ear to discussions, further Uta Panten and Dominik Bauer, who assisted me in doing the recordings. The preliminary English experiment was conducted in the Phonetics Lab of UPenn in spring 2005; many thanks to Maciej Baranowsky for technical support. The main experiments, whose results are recorded in this thesis, have been made throughout the year of 2006 and in the beginning of 2007 at the Phonetics Labs of UPenn, of Tübingen University and ‘on the street’. Many thanks go to Somdev Kar, Marjorie Pak and Jonathan D. Wright for their technical support. In this context I also wish to thank my mother, Dietlinde Speyer, among other things because she supplied me with a large pool of linguistically naïve German native speakers among her colleagues.

Preliminary versions of parts of this study were read at various conferences, PLC 27 (February 2003, Philadelphia), NWAV 32 (October 2003, Philadelphia), International Conference on Linguistic Evidence (January 2004, Tübingen), TaCoS 2005 (June 2005, Stuttgart) and Interspeech 2006 (September 2006, Pittsburgh). I want to tank the audiences of these conferences for innumerable useful hints and comments in the respective discussion periods.

Also I wish to thank my fellow-knights of the 'Dr.-Cardona-Happy-Hour', Jonathan Gress-Wright, Neville Ryant, Joel Wallenberg, and especially Jean-Francois Mondon for their friendship and for proofreading a draft of this opus and correcting my English, if it proved to be too baroque to be of any practical use to the reader.

The final thank goes to the editor of TiEL, Elizabeth Traugott, who made a tremendous impact by her comments throughout the time in which I was preparing the manuscript for publication. It was her suggestions about material that goes beyond the original thesis that allowed the book to be a comprehensive study.

Frankfurt / Main, January 2010 *Augustin Speyer*

Contents

Chapter 1
Introduction

1.1. Overview

The main concern of this study is to demonstrate how a general phonological, or more specifically, a prosodic requirement – the Clash Avoidance Requirement (= CAR) – can influence the syntactic usage of a given language, English. So it is, on a more abstract level, about the interaction of seemingly disparate aspects of the language, namely phonology and syntax. The way they interact is highly dependent on principles of information structuring, the effects of the interaction are observable over a given time span, further insights come from a comparison with German, the close relative. It is consequently fair to say that this study touches on four linguistic disciplines, Syntax, Phonology, Pragmatics and Historical Linguistics.

Topicalization is an exemplary case for demonstrating this interaction and the power of the Clash Avoidance Requirement, and therefore much of this text will be devoted to a discussion of topicalization in the history of English. In the second part of the study we will see that the Clash Avoidance Requirement is responsible for a gradual decrease in the rate of topicalization in Middle and Early Modern English to a stable, yet low, frequency. This decrease in topicalization is observable only in cases in which the loss of the verb second word order option (= V2), which happened in the same time span, leads to potential violations of the Clash Avoidance Requirement. They can occur when two full noun phrases come to stand adjacent to each other, because then both noun phrases have a certain likelihood of bearing focal emphasis. In this case, that is, when there are two phrases with focal emphasis in a sentence, the Clash Avoidance Requirement requires that they must be separated by at least one element of minor prominence. In this study the decline of topicalization will be attributed to the danger of CAR-violations in the wake of the loss of the V2 word order option. Alternative explanations, such as the idea that the decline in topicalization has to do with the growing rigidity of word order in English, or that the decline in topicalization is due to the gradual loss of pragmatic contexts in which topicalization was used, will be argued against.

The study begins with some definitions and an overview over concepts mentioned throughout the study in chapter 1. After having shown in the

second chapter how the Clash Avoidance Requirement influenced syntactic usage in earlier periods of English, the third part of this investigation will be devoted to the Clash Avoidance Requirement in present day English and German and its technical description. I will present experimental data which shows that speakers of English and German prefer to avoid uttering two foci adjacent to each other, but if they are forced to do so, they rescue the Clash Avoidance Requirement by inserting a pause.

In a fourth more theoretically oriented part, I will discuss the reasons why speakers typically choose pause insertion and not other clash resolution mechanisms in situations of focus clash. The properties of rule-governed metrical prominence and semantic focal prominence are so different on a descriptive level that focus cannot simply be reduced to being a continuation of the metrical prominence system. Moreover, different rules are used to generate them which interact in a typical way, but remain quite distinct. A focus indicator is only assigned if there is a narrow focus on a word; otherwise, rule-governed metrical prominence takes care of the assignment of prominence up to the topmost level. The Clash Avoidance Requirement holds on this topmost level, the clause level, both in the presence and absence of focus, and can be easily formalized in the framework of Metrical Stress Theory, following Hayes (1995), as a ban on non-branching feet.

In the fifth part, I will turn to Old English and show that here also the Clash Avoidance Requirement plays a central role in the interaction between syntactic usage and phonology. This is especially obvious in a hallmark problem of English syntax, the alternation of surface V2 and V3 word order. This alternation will be shown to be governed by the CAR: As we can observe, the alternation appears in such a way that the element with the least likelihood of bearing focus always immediately follows the topicalized phrase, either the subject if it is topical (most often realized as pronoun), or the verb if the clause has a full noun phrase, non-topical subject. The former case yields V3-sentences, the latter V2-sentences. This pattern corroborates the view (cf. Haeberli 2002) that Old English syntax was not a strict V2-syntax in the fashion of Modern German, but that Old English had two subject positions for subjects of a different (information-structural) shape, and thus resembled much more Modern English syntax than the classic West-Germanic (= German) type.

1.2. Some background

As it should be useful to give some preliminary definitions of notions and ideas that this study makes use of, let me briefly introduce some relevant concepts. I will devote two sections to this end. The first section (1.2) touches on the theoretical frameworks to be applied. In this section the pragmatic dimensions are introduced that will be discussed, the model of grammar and the metrical theory which I assume, and the German field-model, whose terms we will encounter frequently. The second section (1.3) discusses more specific concepts, viz. what we mean when we say "verb second", and how it is possible at all to determine prosodic properties in written texts, even written texts of a bygone stage of the language.

1.2.1. Pragmatic dimensions

One does not need to be a functionalist to recognize that in a number of languages one of the most important factors determining surface word order is discourse and information structure. Latin is certainly among those languages, but so is German, and, to some extent, even a language like English (Mathesius [1928] 1964).

But information structure is not a unitary notion that always influences word orders in the same way. The term *information structure* is rather a cover term for several ways in which information can be ordered. In the 1960s and 1970s, in the wake of the teachings of the so-called Prague school (e.g. Firbas 1974), it was assumed that there is only one information structural dimension – a 'communicative dynamism', which subsumed theme-rheme, background-focus, given-new, frame-proposition. But at present many researchers assume that there are indeed several information structural ordering principles (cf. to a similar multi-layered conception of information structure Féry and Krifka [2008]). Let us call these principles 'pragmatic dimensions'. It is important to note here that these dimensions are not reducible to one another (as the length of a physical object cannot be traced back to its depth, for instance), but exist independently and try to order the information in their own way, in consequence sometimes coming into conflict with other dimensions, of course.

Four dimensions are relevant here. I do not wish to imply that there are not more dimensions, but these have been selected, partly because they proved to be of importance, partly because they influence the prosody of a clause directly. They are the following:

– newness: old versus new information,
– topicality: topic versus comment (roughly = theme versus rheme),
– focus: focus versus background,
– scene-setting: scene-setting versus proposition-internal.

In the following the definitions are given for each dimension. The definitions depend basically on Féry and Krifka (2008).

Newness is a rather self-explanatory concept, although one has to ask, what the scope of 'new' or 'old' is – new/old for the hearer, new/old for the speaker or new/old in the discourse. In this study I use the old-new-distinction exclusively in the sense related to the discourse: Information that has been previously mentioned in the discourse counts as *old* (or *given*, or, as Prince (1981a) calls it, *evoked*), whereas information that is mentioned for the first time counts as *new*. Examples for discourse-givenness and newness are given under (1). There are practical reasons for that choice, in that in dealing with written texts we may on the one hand assume that the writer only uses entities which are old to him, on the other hand we can trace only newness or evokedness within the discourse – we have no idea what would be old or new for the typical recipient of such literature in the time in which it was composed.

(1) Rudolf Bupfinger, inspector of the state's criminal investigation unit, was sitting in his office. All of a sudden the door was flung open and a young man stumbled into *the room*. *He* held a hatchet in his hand. What *the inspector* found even more remarkable was the knife which was protruding from the back of *his visitor*, *who* fell down, pale-faced.
– Discourse-new information: underlined
– Discourse-old information: in *italics*.

There are several intermediate stages to the old-new-distinction, either to be conceived of as different points on a scale, as in Gundel, Hedberg, and Zacharski (1993),[1] or as different entities altogether, as in Prince (1981a). One of the intermediate stages is the status that Prince (1981a) calls 'inferable', which means that a given entity has not been mentioned in itself before, but other entities which are typically associated with this entity are present in the discourse universe, so that the hearer can infer it via logical or plausible reasoning. An example is given under (2). Here we know from world knowledge that rescue squads typically contain at least one para-

medic, so the mention of a paramedic is in some ways premediated by the mention of *rescue squad*. Inferable information normally patterns with old information.

(2) The first thing the inspector did was calling the rescue squad. On arriving, the *paramedic* felt for the pulse.

Old and new information are often encoded differently; old information tends to be realized by pronouns (if felicituous reference is guaranteed or at least likely), whereas new information is realized by phrases containing 'real' lexical material. Example (1) follows this pattern to some extent; it is obvious that the referent of *he* must be a person that is salient in the discourse. The fact that old information patterns with pronouns in general will prove to be relevant in the further course of this study.

Let us turn to topicality. What counts as a topic has been a matter of debate, partly because there is a great deal of terminological insecurity connected with this concept. Some studies define 'topic' as the element which is at the leftmost position of the sentence (hence the term 'topicalization' for movement of elements to the left periphery).[2] This is not the sense in which the term 'topic' is used here. Other studies (e.g. Chafe 1976) equate topic with old information. As I have introduced old information as an independent notion, I obviously do not follow this usage either. In this study, topic is understood in a non-structural, pragmatic sense as the entity that the sentence is 'about' (following Reinhart's [1981] definition, which is the standard definition of theme in the Prague school tradition and which in the end goes back to Paul [1875: 125]); the rest of the sentence adds information to this particular entity. An example is offered in (3), in which all sentences except the first add information to the ominous young man, who is referred to by a pronoun, as is typical for topics.

(3) Bupfinger looked sadly at *the young man*. Obviously *he* had been in a hurry to come here, but before *he* reached his victim, someone thrusted the knife into *his body*. *He* was clad in a blue jeans and a T-shirt, very unobtrusive.

To determine what the 'topic' of the sentence is, therefore, requires a certain amount of intuition, which most people however possess. An attempt to cast these intuitions into a more formal framework was made by Centering Theory (Grosz, Joshi, and Weinstein 1995; Walker, Joshi, and Prince 1998), which makes crucial use of the fact that topics are usually old in-

formation, and that topics tend to be realized by predictable syntactic means. In English, for instance, topics tend to be realized as pronouns and frequently function as the subject of the sentence. This latter property probably is true for all Indo-European languages (cf. Lehmann 1976).

Focus is strictly speaking not a purely information structural term, but rather a semantic term, because we can identify a semantic operation that is associated with the presence of focus (Rooth 1985). We can distinguish several kinds of focus, e.g. presentational focus (4a), contrastive focus (4b; Rochemont 1986), verum-focus (4c, see e.g. Höhle 1992), and probably more.

(4) a. One thing Bupfinger found strange: *The leather boots* which the young man was wearing.
b. Normally men of his age preferred *sneakers*. *Such leather boots* Bupfinger only knew from Jane-Austen-movies.
c. But this guy *WAS* wearing them, that was the weird thing.

For English, focus is associated with prominence on the focalized element, and this prominence is the highest one in the sentence (see Jackendoff 1972). This means that focus is, in contrast to e.g. old/new information or topics, explicitly marked in the linguistic output. We assume, following Jackendoff (1972) and subsequent literature, that focus is realized by an abstract [+ focus]-feature that is associated at PF with an extra layer of prominence (more detailed see section 4.1.2). Other languages use other strategies to mark focus, e.g. focus particles (e.g. Japanese), pre-specified focus positions (e.g. Hungarian), or a combination of prominence and particle (e.g. German). A presentational focus falls on an element that is new to the discourse and whose newness should be emphasized at the same time. Contrastive focus falls on elements that stand in a partially ordered set (henceforth *poset* for short) relation to each other as members of a set that is either evoked previously in the discourse or is evoked by the first mentioning of one of its members. Verum-focus is a very specialized type of focus; it lies on the verb and emphasizes the claim that the proposition is true. All these different kinds of focus can, in the end, be reduced to contrastive focus, as Rooth (1985) showed: in all cases of focus a set, consisting of salient entities, is evoked of which the focused element is a member. The meaning of focus can be summarized as ‘it is X, and not other members of the salient set containing X, although they would have been equally eligible’. I want to mention here a point that I elaborate on later, viz. that I reserve the notion of focus to cases in which a salient set is clearly identifi-

able, which as a rule coincides with what is known as instances of 'narrow focus'. 'Wide focus', where the set would consist on possible propositions or object-verb pairs does not fall under this strict definition of focus. Therefore not all English sentences do have a focus, under this view; the highest prominence in a sentence is not automatically associated with focus.

Scene-setting, finally, is an information-structural dimension, but with a semantic side to it. As opposed to, say, concepts such as topic-comment or newness, scene-setting elements have direct implications for the truth value of a sentence (whereas, e.g., it is irrelevant for truth conditional purposes whether a given expression is thematic or rhematic, for instance). We can define scene-setting elements as elements that specify the situation under which the truth value of the proposition has to be evaluated (definition following Jacobs [2001]). They do not belong to the core proposition. Examples can be found in (5).

(5) *In the year 2008*, wearing such shoes was most remarkable. All the more since it was a hot summer day. *Only the day before* a heavy thunderstorm struck the town with unwont violence.

Although these four pragmatic dimensions are independent of each other, there are certain typical intersections (see also Speyer 2008a). Topics are, as a rule, also old information.[3] Not all old information functions as a topic, however. New information is often focused, but it need not be. Foci can be new information or old information (this is often the case with contrastive foci). A phrase can be topic and focus at the same time under certain circumstances. We will encounter the intersection between 'topic' and 'focus' in section 2.2 of this study. Scene-setting elements tend to be old information. New scenes can be introduced, though, and in that case these expressions usually receive focus.

The dimensions are often in conflict with each other. This is because each dimension poses certain requirements on the linguistic output and speakers tend to follow these requirements: Old information is likely to be placed before new information, topics are put before their comment, foci are preferably realized at one of the edges of the utterance, and scene-setting elements are usually positioned before the proposition. All of these ordering requirements make sense independently from the point of view of sentence processing: it eases processing if old and new information are not jumbled together but are ordered somehow (Musan 2002). Also, it is more sensible to first evoke the 'filecard' (= topic) and only afterwards the material that has to be added to this filecard (= the comment), if we want to use

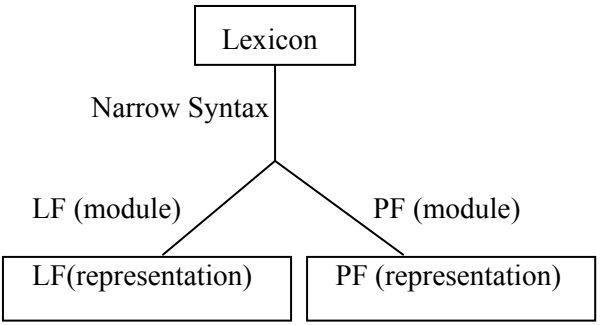
Heim's (1982) famous metaphor.[4] It is better, if one wants to emphasize something, to put it in a position where it coincides with one of the clausal edges and therefore can be treated as separate processing unit. And if a situation's truth value is to be evaluated, it is more practical to know the situation before hearing the material that is to be evaluated. So each dimension has a certain 'claim' on sentence structure and order, so to speak. Which one of these claims determines the shape of the output varies by cases, although languages tend to have a ranking of the dimensions (see Speyer 2008a).

1.2.2. Modularity of Grammar

I assume a modular model of grammar in the tradition of Chomsky (1995, 2001). I assume the modified T-model (or rather: Mercedes-star-model; [6]) in which there are three components: Narrow Syntax, Logical Form (LF) and Phonetic Form (PF). In this study we are mostly interested in PF. In Narrow Syntax we are interested only insofar as it contributes to the PF-representation.

Narrow Syntax is the module in which material from the lexicon – in Minimalism referred to as 'enumeration', at this stage represented as abstract concepts and feature bundles – is assembled and in which the first transformations take place, such as movement of the subject to SpecIP, for instance. At the place at which Surface Structure used to be in the Extended Standard Model (e.g. Chomsky 1981) is now a bifurcation that does not count as an independent level of representation. The output which narrow syntax has produced feeds into two modules, LF and PF. LF is the module where movement operations take place that are not represented in the form of the sentence that is uttered (since the branch leading to the actual utterance is PF, and we have left this track at the bifurcation) and that concern mostly the correct semantic representation of the utterance, e.g. scopal properties. PF, on the other hand, is the module in which the syntactic structure is eventually flattened out, transformed into a linear string. Lexical Insertion takes place (see Halle and Marantz 1993) and purely phonological operations are performed, such as the assignment of prosody and the adjustment of the rhythmic structure. These are in principle not relevant for the semantic interpretation (with the apparent exception of focal emphasis, of course),[5] but they make possible the vocal production and give cues to the syntactic structure which, after reduction of the two-dimensional structure into a one-dimensional string, is no longer directly observable.

(6)



In contrast to Chomsky (1995), but in accordance with many other generative grammarians (see Fanselow 1991; Haider 1997; Rizzi 1997; Haider and Rosengren 2003: 206; Erteschik-Shir 2005), I assume that there are also movement operations that are not governed by strictly syntactic features, but that are discourse-structurally motivated. This implies that there are also functional projections that can host phrases with a certain discourse structural status, such as the ones identified by Rizzi (1997). Movement to these projections is not warranted by Narrow Syntax, if information structure is not considered as part of the semantic representation (but cf. Asher and Lascarides [2003] for a 'semantic' view of information structure). For this reason it should be considered whether the place where such movement operations take place is perhaps PF rather than narrow syntax (even more radically Erteschik-Shir 2005). We could view PF procedurally as consisting of several sub-modules, one in which additional, non-syntactically motivated and non-semantically interpretable movement operations take place, one in which the structure is reduced to a string, one in which Lexical Insertion takes place, one in which the rhythmical structure is assigned and one in which the well-known phonological rules of sandhi, assimilation etc. take place. But this question is beyond the scope of this study.

1.2.3. Prominence

Prominence is used here as a cover term for the property a linguistic entity has (usually a syllable) to be perceived as 'stronger' than other linguistic entities of the same sort. I will make a distinction between the phonological and the acoustic aspects of this concept. Acoustically, a syllable A is more prominent than a syllable B if A has higher values than B on certain meas-

urements – pitch especially, but also volume and duration. In other words: A syllable A is more prominent than a syllable B if it is higher-pitched, louder, and possibly takes more time to articulate, such as *REE* in *refeREE*, or *CAT* in *a tortoise-shell CAT*. One can say that syllable A is also more prominent, that is, higher, louder and longer, than a non-prominent instance A' of the same syllable. *CAT* in *a tortoise-shell CAT* is more prominent than *cat* in *the cat with the HAT*.

Phonologically speaking, prominence can be represented by constructing a metrical tree and/or building a grid in which strong and weak marks are assigned; the more strong marks are assigned to a syllable, the more prominent this syllable is. The grid reflects the grouping of syllables and larger units into feet; the prominence that is assigned is dependent on the headedness of the feet. Further below a distinction will be made between prominence that is assigned by rules and prominence that is the outcome of focus. I will distinguish these types of prominence terminologically in the following way.

On the phonological level, prominence assigned by the metrical calculus (the system that is described by rules of prosody and grid production) will be referred to as *metrical prominence* (or simply *prominence*). The rule-governed construction of metrical prominence can be disturbed by a *focus indicator*, which is prominence (or, as I will often call it in order to distinguish it from metrical prominence, *emphasis*) associated with a focus feature.

The highest prominence assigned by the metrical calculus of a given unit will be called its *prominence peak*. The highest clausal prominence will be called the *clausal prominence peak*.

On the level of phonetic representation, the term *stress* will be used for the acoustic correlate of metrical prominence, and the term *focal emphasis* or simply *focus* for the acoustic correlate of the focus indicator (for the usage of focus in this sense see e.g. Wells [2006]). By use of these terms I do not wish to imply that one of these phonetic entities has fundamentally different properties from the other (e.g. that stress is louder than the rest, and focus is higher pitched than the rest, or the like); 'stress' in my usage can include pitch movement, longer duration etc. The phonetic correlate of the clausal prominence peak is called *sentence stress* or *nucleus*.

In making this distinction I follow Ladd (1996: 160), who seems to be quite close to the consensus of the last few years. Ladd makes a distinction between 'normal stress' and 'focus-to-accent'. 'Normal stress' is rule-governed and thus prominence that can be calculated. Normal stress applies to all domains, including the clause. The highest stress of the clause is re-

ferred to by Ladd as *sentence stress*; Newman (1946: 176) calls it *nucleus*, and this term has often been used to denote this concept (e.g. Chomsky and Halle [1968] in their Nuclear Stress Rule). Ladd (1996: 293 n.2) points out that often the term *default accent* is used. This usage is, however, due to misunderstanding of the term as he himself coined it in Ladd (1980), where it denotes a completely different concept: it is used only in words that are deaccented to refer to the position on which the accent would fall if the word under discussion were not deaccented.

The prominence associated with focus does not have an accepted designation; Ladd (1996: 161) refers to it as *accent*, *focus*, or *emphasis*. This kind of prominence obviously has a semantic side to it, which metrical prominence does not have. Connected with this usage is the idea that every utterance has a focus somewhere, either a 'wide / broad focus', meaning focus on the clause as a whole, the verb phrase or some other relatively large unit, or a 'narrow focus', meaning focus on just a word or an even smaller unit. The unit for semantic focus-assignment is variable; most often it is a whole word, although the focal emphasis is of course realized only on one syllable of this word, usually the syllable which would be the most prominent one anyway. A consequence of this perspective is that sentence stress always coincides with focal emphasis, as this is where the highest prominence of the sentence is, if the sentence or the biggest part of it is in wide focus. Ladd (1996: 161) describes the matter in this way:

> 'Given the idea of broad focus, 'normal stress' rules can be seen as a description of where accent is placed when focus is broad.'

If we have narrow focus, the rules for sentence stress are blocked from applying in a regular fashion, as here the "accent goes on the focused word" (Ladd 1996: 161).

There are other definitions of 'stress' and 'accent'. Ladd's definition depends on Bolinger's (1961, 1972) distinction and is more or less identical to the distinction used by Sluijter (1995). For Bolinger, and the tradition of phonologists before him, *accent* is the term used for the highest prominence in a given unit, whereas *stresses* are the prominences on lower levels (the word, the phrase). He was perhaps the first to draw attention to the fact that it is exactly the highest prominence peak that often is not predictable by rules, but reflects semantic and pragmatic notions such as emphasis, newness, contrast, etc., what was termed *focus* soon thereafter (Jackendoff 1972). This development, of course, caused a certain terminological insecurity, as there were now two competing meanings of the term 'accent':

1. highest prominence in the clause, or
2. prominence associated with focus.

These meanings coincide exactly then when we assume that each sentence has a focus, and this is the line taken by e.g. Schmerling (1976); Ladd (1980); Selkirk (1984). Without the idea of broad focus, these meanings coincide only then when there is a narrow focus on some word. In other words: Only when a word is focused in a clause, this clause will have focal emphasis. Otherwise it may have an accent in the sense of (1.) in the quote above, but if we assume that both definitions must hold for focal emphasis, sentences without narrow focus do not have focal emphasis at all, but simply sentence stress. This is the line I will take in later sections.

One consequence of the terminological complexities sketched here is that there are many special uses of the terms *stress* and *accent*. Schane (1979: 485), for instance, defines *stress* as the phonetic manifestation of prominence and accent as the underlying representation of it. In other studies, *accent* is the term used on the production side. Wells (2006) for instance uses *accent* only as the phonetic realisation of prominence associated with a pitch gesture, whereas the underlying prominence associated with focus is simply called *focus*. Sentence stress is called *nucleus*, which has the advantage that one does not have to commit oneself to the question whether the nucleus is a kind of metrical prominence (rule-generated, no focus) or a kind of focal emphasis (broad focus).

1.2.4. Grid construction

The theory of grid construction used in this study is based on Metrical Stress Theory (Hayes 1995) with elements of Idsardi (1992); cf. also Halle and Idsardi (1995). The grid is constructed in the following way: each relevant element (in this study, the lowest relevant level is the word level, but the theory works the same way below the word level) is assigned a strong grid mark. In this study, asterisks are used for strong grid marks, and dots for weak grid marks. The next higher line adds alternating strong and weak marks following certain rules. This process is equivalent to the bracketing in Idsardi (1992). The lines are not simply a continuum, but (at least) three disctinct levels can be identified which serve as the domains for prominence assignment and for metrical rules. These levels are the word, the phrase, and the clause level (corresponding in conception, but not necessar-

ily in detail, to the levels of word, phonological phrase and intonational phrase of the Prosodic Hierarchy, cf. Truckenbrodt 2007: 436). In this introduction I use a simplified version with a continuous grid, for ease of explanation. The rules for the assignment of strong and weak marks are parametrized, that is, different in a limited way for different languages. In English and German, the two languages that are the focus of this study, the rules are different for the domains below the word and the domains higher than the words. The basic rule for grid construction on levels higher than the word in English is as follows, cast in terms of Metrical Stress Theory:

Iamb Construction Rule:
Assign iambs from right to left.

It is easy to see that this is an iterative version of the Nuclear Stress Rule, as we know it from e.g. Newman (1946: 176) and Chomsky and Halle (1968: 90). It means that the assignment process starts at the rightmost word of the clause, assigning a strong mark to it, assigning a weak mark to the penultimate word, assigning a strong mark to the third-last word and so on, until the clause has been scanned completely. The next higher line uses the same assignment rule, and puts alternating strong and weak marks on the grid. It is not simply a copy of the line below, as the only positions that are available for assignment are the ones with strong marks on the lower line. The assignment process for this level goes on, until the clause has been parsed completely. In this fashion, line after line is added until further assignment would be vacuous, i.e. until a line is reached where only one iamb can be assigned. We will say that the parse is exhausted on this level. The relative prominence of the elements in the clause is a result of the relative number of strong grid marks each element has received. Schematically, the assignment process is shown in (7).

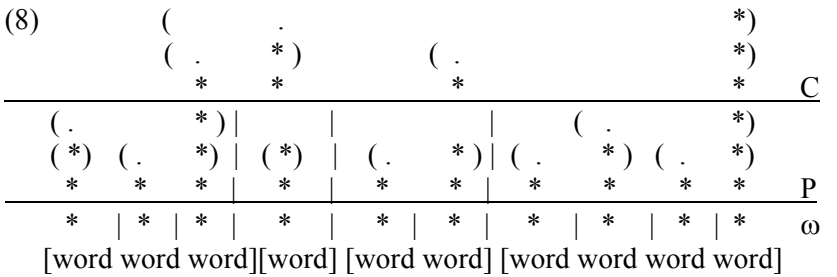
(7)
```
  .           *
  *     .     *
  *  .  *  .  *
  *. *. *. *. *
  * * * * * * * * *
```

There are two factors that can interfere with this strict assignment. One is phrasing, the other eurhythmy. By ‘phrasing’ I mean the fact that not only the word and the clause are relevant domains for prominence assignment, but also the phrase. We thus need an intermediate level of representation.

Each phrase must contain at least one strong mark (Truckenbrodt 2006), with the sole exception of functional elements such as pronouns. Besides the word, the (phonological) phrase and the clause (= intonational phrase), probably no other members of the Prosodic Hierarchy (cf. Nespor and Vogel 1986) are relevant for prominence assignment (cf. also Truckenbrodt 2006). And the 'phrase' I am talking about here is not necessarily the Phonological Phrase of Selkirk (1984) and Nespor & Vogel (1986), but rather a phrase that is roughly identical to a syntactic constituent: either an immediate constituent, that is, a syntactic phrase immediately dominated by VP (in its base-generated position, i.e. before movement of material to functional projections such as IP and CP), or the head of a VP, also in its base-generated position. Precedents for such a 'direct correspondence approach' are e.g. Cinque (1993) and Seidl (2001).

As pointed out above, I assume that there are three relevant levels for asignment of prominence: The word evel (ω), the phrase level (P) and the clause level (C). Each level consists of one or more lines. On each level, a different set of rules for grid construction applies. First, the grids for single words are constructed, by the general rules for grid construction as given in e.g. Hayes (1984: 35), following Liberman and Prince (1977: 315–316, 322), and by the relevant rules for the word level. The peak mark of each word is projected on the next higher level, the starting point for phrase grid production. The relevant rules add lines to the grids of individual phrases, until the level is exhausted, i.e. until a line is reached on which only one foot can be assigned. The strong marks of the phrases are projected to the first line of the next higher level, the clause level, and serve as starting line for the production of the final grid, following the relevant rules on the clause level. Again, lines are added, until the level is exhausted. Every phrase that is dominated by VP and its extended projections IP and CP projects one strong mark onto the bottom line of the clause level (see Truckenbrodt 2006; with the exception of phrases that consist only of intrinsically weak elements, such as pronominal DPs). In this study, no higher unit than the clause is taken into account, although the sentence (= Utterance) constitutes a higher level.

The idea that for each level several lines can be constructed until the level is exhausted goes back to the notion that phrasal (and clausal) metrical prominence assignment happens cyclically (see e.g. Selkirk 1984). So the assignment process would proceed as in (8). Since what the metrical calculus basically does is assign feet, we may as well mark the feet in the grid.



This kind of representation takes account of two requirements on grid production that seemingly are in conflict with each other: On the one hand, there is the additive nature of metrical prominence, in the sense that every level builds on former levels, i.e. that a more prominent metrical prominence is the result of the addition of prominence marks on different levels. This implies that metrical prominence on the word level and on the phrase level must be represented in the same grid (cf. Truckenbrodt 2006), as the final audible gradation in prominence is the addition of metrical prominence marks on the word, phrase and clause level. On the other hand, there is the fact that the assignment rules are potentially different for each of the three levels. Take for instance metrical prominence in German: The rule for the assignment of metrical prominence on the word level is identical to the Latin rule, namely that the first moraic trochee, counting from the right, under extrametricality of the final syllable, receives the main prominence (Speyer 2009b). The rule for metrical prominence on the phrasal level, on the other hand, also counts from the right, but here it is iambs and not moraic trochees that are assigned – see the version of the Iamb Construction Rule above. The metrical prominence assignment rule for the clausal level, in the end, is similar to the rule for the phrasal level, but it treats verbal material at the edge as extrametrical.

We have to bear in mind furthermore that a metrical grid can be subject to another process, namely eurhythmy (cf. Hayes 1984). Eurhythmy is basically a well-formedness condition on grids; the basic rules are, freely after Hayes (1984), that the highest prominence marks should be kept as far apart as possible ('Phrase Rule'), and that in-between a strict alternation of strong and weak marks should be strived for. The grid in (7) would be perfectly eurhythmic. A grid like (8), on the other hand, would not be eurhythmic. The processes trying to obtain eurhythmy would first push the second highest mark to the first constituent (9a), thereby making the grid conforming to the Continuous Column Constraint (Hayes 1995: 34–37),

then repair the equal heights of the intervening material by destressing the column which is closer to the next highest prominence peak – in that case the left of the two constituents (9b). Then the grid will be eurhythmic and an adequate metrical representation of an English sentence with the constituent structure given in (8). Note in this connection that certain function words such as the article or personal pronouns are not counted into the computation normally because they do not have word stress and therefore do not receive a strong mark even on the word level. They are only included into the computation when they happen to bear focal emphasis. In this case they of course receive a grid mark motivated by the focus feature, the 'credit grid mark' which I will elaborate on in section 4.1.

```
(9)  a.        (   .                                          *)
                (  *  ) ( .              .                    *)
                   *      *              *                    *     C
     ____________________________________________________________________
     ( .           * ) |       |           |      (  .        *)
     ( *)  ( .     *)  |  (*)  |  ( .    * ) | ( .     * )  ( .   *)
       *     *     *   |   *   |   *     *   |  *      *      *   *     P
     ____________________________________________________________________
       *   | * | * |   *   |   *  |  *   |   *   |  *  |  * | *    ω
     [word word word][word] [word word] [word word word word]

     b.        (   .                                          *)
                (  *  )             (  .                      *)
                   *     .             *                      *     C
     ____________________________________________________________________
     ( .           * ) |       |           |      (  .        *)
     ( *)  ( .     *)  |  (*)  |  ( .    * ) | ( .     * )  ( .   *)
       *     *     *   |   *   |   *     *   |  *      *      *   *     P
     ____________________________________________________________________
       *   | * | * |   *   |   *  |  *   |   *   |  *  |  * | *    ω
     [word word word][word] [word word] [word word word word]

     b.        (   .                                          *)
                (  *  )             (  .                      *)
                   *     .             *                      *     C
     ____________________________________________________________________
     ( .           * ) |       |           |      (  .        *)
     ( *)  ( .     *)  |  (*)  |  ( .    * ) | ( .     * )  ( .   *)
       *     *     *   |   *   |   *     *   |  *      *      *   *     P
     ____________________________________________________________________
       *   | * | * |   *   |   *  |  *   |   *   |  *  |  * | *    ω
     some dark stranger gave  blue flowers to-the mildly surprised girl
```

Eurhythmy, however, is a special effect of a much more basic requirement of language, the Principle of Rhythmic Alternation (on its importance for grammar see e.g. Schlüter 2005). I will postpone a discussion of this principle to later sections and chapters (especially 2.4, 3, and 4). The theory as it is sketched out here is probably too simple and would need additional features if applied to other problems. But it seems to be accurate for the domain which we are mainly interested in: the domain of phrasal and clausal metrics, which I may call summarizingly *supraverbal metrics*.

It would have been possible to use Optimality Theory with the appropriate metrical formulations (for an overview over such systems in OT see e.g. Truckenbrodt 2007), but I decided to stay within the frameworks of classical autosegmental and suprasegmental theories, as a reformulation in OT terms would have no effect on the results to be described or on the explanations I will be proposing. A short sketch of an OT variant is outlined at the end of section 3.2. The grid serves as input for the assignment of intonation contours, indicating the positions of the different low and high pitch accents and boundary tones (on intonational contours see e.g. Pierrehumbert [1980] for English, Féry [1993] for German). I assume that grid production and the assignment of intonational contours are two distinct processes (cf. also Truckenbrodt 2006). Therefore I will not treat questions of intonation proper (i.e. contour formation, pitch accent realization) here, but confine myself to the construction of the grid, as this is sufficient for the purposes of this study.

1.2.5. The syntactic field model

The *Feldermodell* ('field model') dates from the early years of German linguistics as a mode of representation for the sentence patterns of Modern German. It was introduced in the 1820s by Simon Herling (Herling 1821; see Abraham and Molnarfi 2001), and gained momentum especially under the influence of Drach (1937). According to the most common versions of the field model (cf. e.g. Höhle 1986; Grewendorf, Hamm, and Sternefeld 1987; Reis 1987: 147–148; Abraham & Molnárfi 2001), a sentence can be divided into the following parts which stand in the order given here:

Vorfeld	–	linke Satzklammer	–	Mittelfeld	–	rechte Satzkl.	–	Nachfeld
prefield		*left sentence bracket*		*middle field*		*right sent. br.*		*back field*

Before the *vorfeld*, another – marked and very restricted – position (*vorvorfeld,* 'pre-prefield') can be introduced.

Each of these 'fields' has special properties:

- The verbal elements all stand in the s*atzklammern*. In main clauses the finite part of the verb is in the *left satzklammer*, the remainder of the verbal material in the *right* one. In subordinate clauses all verbal material is in the *right satzklammer*, the complementizer is in the *left* one.
- The *nachfeld* is usually filled with subordinate clauses or otherwise 'heavy' elements.
- Most of the non-verbal sentence material stands in the *mittelfeld*. There are no constraints whatsoever on what can stand in the *mittelfeld*, as long as it is not verbal. There are certain constraints on the order of elements, however (see e.g. Hoberg [1997], as summarizing representative of an abundant research literature).
- The *vorvorfeld* can only contain main clause connectives and material which can be shown to be left dislocated.

We are mostly interested in the *vorfeld*. The *vorfeld* in Modern German can contain exactly one constituent. There are some exceptions to that, and the further back in history we go the more frequent these exceptions become, so that we are forced to assume that the one-constituent-only constraint of Modern German is a recent development, and that originally more than one constituent could stand before the left sentence bracket. This is going to be of immediate importance for Early German and English. We will return to this question in section 5.3.

It can easily be seen that the *Feldermodell* translates directly into modern generative terms (cf. den Besten 1981; Vikner 1995; slightly differently Sabel 2000): the *vorfeld* corresponds to SpecCP, the *left satzklammer* to the C-head, the *mittelfeld* to everything under C' save for the – in German right-peripheral – V-head(s) and the I-head, which form the *right satzklammer*. The *nachfeld* contains IP adjuncts to the right.

For Modern English, using the field model does not make much sense and does not offer great insights, although it could be done (the left sentence bracket contains all verbal material, the default filler of the *vorfeld* is the subject, although more than one phrase can stand in the *vorfeld*, and the distinction between *mittelfeld* and *nachfeld* is hard to draw as there is never overt material in the right sentence bracket). The positions of the field model would not correspond however to generative entities in Modern English. This is different for earlier stages of English in which the sentential structure shared some properties with Modern German. Therefore, terms of

the field model will occasionally be used for Old and Middle English in the course of this study.

1.3. Further concepts

1.3.1. Verb second

In this study I will frequently make use of the term *verb second* (V2). The usage of the label V2 tended to be rather imprecise in the past, and therefore it is perhaps useful to dwell a bit on this subject. V2 can be used in a more typological manner to express the property a language can have of putting the verb in the second position in the sentence, that is, the position after the first constituent. Note that whoever uses V2 in this sense does not have to commit oneself to a specific analysis: he or she simply states that at the surface we have the verb in second position, no matter what the underlying analysis is that takes care of having the verb at exactly that spot.

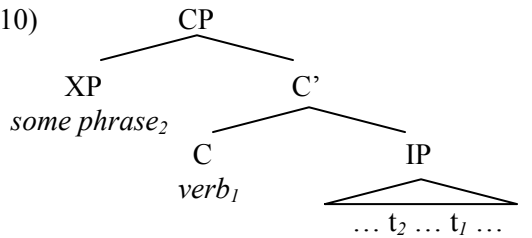
A related notion is that of the *verb second constraint* which on a descriptive level says not much more than the following: some languages (among which are the Germanic languages) show a tendency to build their sentences in such a way that the verb is in second position. The reasons for this tendency are unknown. Brandt et al. (1992) assume the presence of sentence type features that have to be saturated by movement of the verb to C and in some cases (with wh-questions and declarative sentences) also another phrase to SpecCP. Erteschik-Shir (2005) sees it as a phonological process. Lately the hypothesis has been put forward that verb-seconding (and by that the creation of a 'vorfeld') serves to establish a topic-comment structure. Under this view, the verb serves as marker which divides the sentence into these two parts (Hinterhölzl 2009). But this is of no concern for us here. The only thing to mention is that again, if one uses 'verb second constraint' on this descriptive level, nothing is said about the underlying structure.

There is however a less non-committal usage of the term. At least since Vikner (1995), 'V2' is often used to denote a special syntactic configuration, in which there is one functional projection above IP (which is usually referred to as CP). The V2-effect is derived by moving the verb into the head of that projection and some other constituent into the specifier projection of it (10). This corresponds closely to the analysis of the Modern German declarative sentence by den Besten (1981). When the term V2 is used,

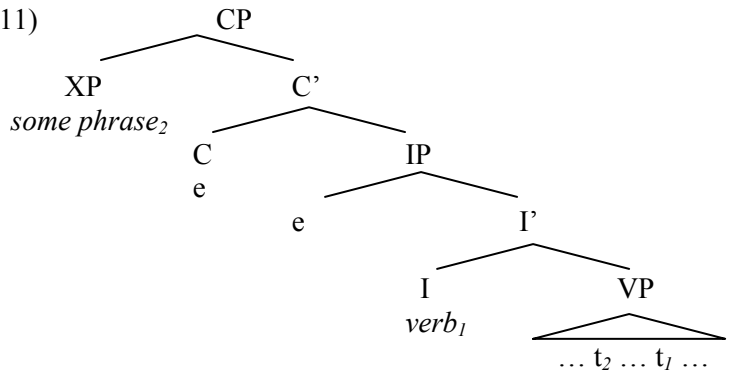
it is often implied that something like the structure in (10) is necessarily the underlying structure of any V2-sentence.

The problem is now, of course, that a surface V2 order can be the outcome of a variety of analyses, of which the one outlined under (10) is only one. For instance, a verb second order can also be the result of a structure as in (11).

(10)



(11)



It turns out that in Old English we have both kinds of V2: V2 by movement of the verb to C and of some phrase to SpecCP (I will hitherto refer to this kind of V2 as CP-V2) and V2 by movement of some phrase to SpecCP, but no movement of the verb from I to C and no element in the specifier position of the projection in whose head the verb has landed (e.g. Kroch and Taylor 1997; Haeberli 2002). I denote it here as IP-V2 for the ease of the exposition. We will get back to that question more precisely in part 5.

When I use V2 in this study I do not mean V2 by movement of the verb to C. For this special usage I use the term CP-V2. The structure of V2 I am mostly concerned with is the version of V2 outlined in (11). It is important to note that this sentence structure is optional throughout the history of English (quite in contrast to CP-V2 in languages which have this structure, where it tends to be compulsory), and therefore it makes sense to speak of