ON PROSODIC WORD OF COMPOUNDS IN ROMANIAN

CARMEN RĂRIȘ-FIERBINȚEANU[1]

Abstract. The present paper aims to highlight the prosodic word features in Romanian, focusing on its properties, the Prosodic Hierarchy and its components, while examining various perspectives and interpretations. Another part of this paper illustrates the characteristics of the Prosodic Hierarchy, which form the basis for constraints in Optimality Theory. Additionally, the paper provides an overview of compound words in Romanian, presenting their complex status and structure, which pose challenges for prosodic analysis. The prosodic structure of these compounds is further analyzed using a constraint hierarchy.

Keywords: Prosodic Word, Prosodic Hierarchy, Optimality Theory, Compound Words.

1. INTRODUCTION

The present article focuses on the concept of the prosodic word, applied to the Romanian language. Within the field of the Prosodic Phonology, the prosodic word represents one of the most examined categories, especially related to the Prosodic Hierarchy. The goal here is to define, classify and exemplify this concept, in relation to a distinct class of Romanian compound words.

Compound words, alongside derived words, have undergone extensive analysis in various languages (see Itô and Mester 2007b on Japanese), therefore the application of the prosodic word to compound words in Romanian reveals interesting insights.

The first part illustrates definitions, features and different interpretations of the prosodic word as found in studies. The second part focuses on the compound words in Romanian, their characteristics and classifications, their diverse prosodic structure analyzed within the framework of the Optimality Theory (OT). The article concludes with final remarks.

2. PROSODIC WORD. THE THEORETICAL OVERVIEW

In phonology studies, the category of *prosodic word* (also referred to as the phonological word, p-word or Pw) stands as one of the main constituents of the Prosodic Hierarchy. It is a crucial concept, used in several types of phonological developments: metrical phonology, prosody, prosodic phonology, morphophonology and the interface between syntax and phonology.

[1] University of Bucharest, carmen_raris@yahoo.com.

2.1. How many words?

Generally, words are understood as "a combination of orthographic, grammatical, phonological and semantic/conceptual criteria; at times, these criteria neatly align to converge on a notion of words that embodies all of these senses at once, but at other time they do not" (Hildebrandt 2018: 255). These criteria lead to lexical units that can be nouns, verbs, adjectives, adverbs.

However, despite the clarity these criteria provide regarding what constitutes a word, mismatches/misalignments can occur. For example, relying more on pronunciation, speakers conceptualize differently the phoneme /n/ and its allophone, which could lead to spelling mistakes (*înv͡ăț* – /ɨn͡m.ˈvəʦ/ – **îmvăț*). This discrepancy is explainable through the inherent complexity of natural languages, where the multiple layers of a word may not always align perfectly, leading to challenges for segmentation.

Another example is represented by function words (such as preposition, conjunctions) and discourse markers, which deviate from the prototypical characteristics of the Pw. While some function words may exhibit primary stress and be monosyllabic or disyllabic, very similar to a Pw, they lack substantial lexical meaning. Instead, their main role lies in facilitating grammatical and pragmatic linkage (Hildebrandt 2018: 255), maintaining a fixed order, coherence and established meaning (Dixon and Aikhenvald 2002: 19).

For these reasons, specialists increasingly have recognized the need to consider the concept of a word across various linguistic domains, including phonology, morphology, syntax and other areas.

Consequently, several terminological distinctions have been proposed, to differentiate between the underlying form of the word, its root – word, and its inflected forms – lexeme (Dixon and Aikhenvald 2002: 7).

2.2. Pw – definition(s)

The term Pw has had a long history, starting to be used in 1970's and subsequently becoming a fundamental concept in phonological research. It has been utilized by numerous scholars to organize phonological categories and to account for facts relevant for both phonology and syntax.

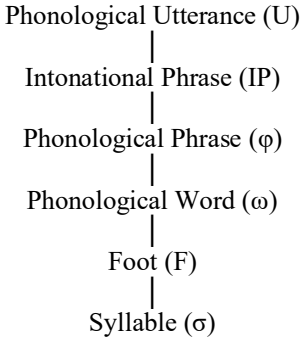
Hyman (2008: 335-336) illustrates the different kinds of Pw, by highlighting the many roles of the Pw:

(i) the demarcative word – a property that marks the beginning or end of the word;
(ii) the culminative word – a feature that occurs only once per word;
(iii) the harmonic word – a feature realized throughout the word;
(iv) the metrical word – a word consists of hierarchically arranged moras or syllables;
(v) the minimal word – a word must consist of a minimum of moras or syllables;
(vi) the maximal word – a word can consist of a maximum of moras or syllables;
(vii) the phonotactic word – a word permitting only certain output segments/sequences;
(viii) the morphophonotactic word – a word permitting only certain input segments/sequences.

Being a multifaceted concept, used in analysis from different theoretical perspective, the Pw incorporates various meanings. There are, however, several features on which most specialists agree on (Selkirk 1978, Nespor and Vogel 2007): a Pw is typically understood as

a phonological unit greater than a syllable (although in some languages, it may consist of only one syllable) that exhibits at least one phonological feature (a segmental one, related to its syllabic and segmental structure, a prosodic feature, illustrated stress or tone assignation and phonological rules, that can apply within the Pw or across its boundaries) (Dixon and Aikhenvald 2002: 13). Additionally, the Pw represents the domain of prosodic minimality constraints, referring to minimum size restrictions imposed on prosodically free units in some languages (Hildebrandt 2018: 258).

Non-linear phonologies, developed as a response to traditional approaches to stress at the word-level, treat the Pw as a category within a hierarchical representation of phonological units – the Prosodic Hierarchy (Hildebrandt 2018: 259, Itô and Mester 2007a):



Other hierarchies do not include the smaller units, subordinated to the Pw, that only have phonological relevance (Dresher 1996: 42).

Nevertheless, this hierarchy illustrates the fact that the Pw typically aligns with the grammatical word, highlighting a strong connection between phonology and morphology. The categories above the Pw are relevant for syntactic analysis, containing more than one word.

The Prosodic Hierarchy is governed by several principles, that reflect the relationship between the Pw and other constituents:

(1) Generality Assumption (Nespor and Vogel 2007: 18, Schiering et al. 2010) states that lexically specified processes and patterns are not included in this theoretical account.

(2) Clustering predicts the fact that the hierarchy is universal, alongside its constituent categories. Accordingly, the Pw is considered to be universal, present in all natural languages. However, there is evidence that certain languages, such as Vietnamese lack this category, while others have multiple domains for the Pw. For instance, Limbu, a Sino-Tibetan language, exhibits multiple Pw domains, which cannot be explained by reduction, redistribution or sub-categorization (see Schiering et al. 2010 for a comprehensive discussion on this topic).

(3) Strict Layer Hypothesis refers at the hierarchy construction. It states that a given prosodic domain *n* should have a dominant category *n+1*. It also forbids reduplication of domains. For instance, a Pw must dominate at least one foot and must be dominated, on a superior level, by a prosodic phrase (Itô and Mester 1992, Hildebrandt 2018: 260).

The Strict Layer Hypothesis translates in Metrical Phonology as a Continuous Column Constraint, related to metrical grids construction (Gussenhoven and Jacobs 2011: 248).

```
(                   )  U
(     )(            )  IP
( )( )(     )(      )  φ
( )( )( )( )( )     )  ω
```

These principles have been further developed as a violable set of constraints in Optimality Theory (Selkirk 1986; 2000). For example, the prosodic alignment theory developed in 1986 (EDGE-BASED PROSODIFICATION: The right (left) boundary of a prosodic constituent C corresponds to the right (left) boundary of a morphosyntactic category X.) is a precursor of Alignment Constraint family in OT.

However, it is important to note that these principles are not absolute, as they are applied to natural languages, whose behavior can be unpredictable. Recursion is frequently observed in relation to compound words, across many languages, being referred to as Weak Layering hypothesis (Itô and Mester 1992, 2009) which allows for multiple layers of identical prosodic structures.

2.3. Pw – developments and potential variables

It is widely accepted that the fundamental principles governing the Pw are still in use and have proven their explanatory power. However, in-depth studies have revealed the necessity to refine certain concepts and introduce new ones, in order to deal with language-specific phenomena. One such issue concerns the alignment between the Pw and the word boundaries. Another aspect refers to the significance of the Clitic Group and its inclusion into the Prosodic Hierarchy. These developments reflect ongoing efforts to adapt theoretical framework to accommodate diverse linguistic phenomena and to increase the understanding of prosodic structure.

2.3.1. *Pw – alignment*

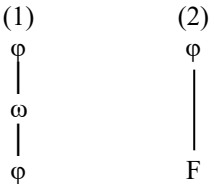
The Pw alignment with morphological boundaries is a topic often mentioned in phonological studies (Hildebrandt 2018: 268, Gussenhoven and Jakobs 2011: 249-251), alongside possible restrictions. Several assumptions have been made: (a) the Pw cannot contain more than one grammatical word; (b) the Pw should contain less than two grammatical words; therefore, it is sometimes possible to contain two grammatical words.

The phenomenon of non-isomorphism has been noted in relation to this alignment between the Pw and a grammatical word, when the two misalign in different ways: (a) the Pw is smaller than a grammatical word or (b) the Pw is larger than a grammatical word or it contains two of them. The case in (b) is often analyzed in relation to compound words or clitics.

Non-isomorphism represents a justification for the existence of two sets of constraints, a prosodic one and a syntactic one, to explain the two types of processes involved.

2.3.2. *Odd prosodic alignments*

It has been illustrated above, in (3), that the Prosodic Hierarchy is governed by the Strict Layer Hypothesis, which prohibits skipping and recursive domains. However, there is evidence of representations that go against the Strict Layer Hypothesis (Hildebrandt 2018: 270):



The first representation illustrates recursion (prosodic adjunction), a common phenomenon in compound words, whereas (2) shows a missing level, the Pw, between F and φ. *RECURSION, considered one of the constraints of the Strict Layer Hypothesis, is considered violable, with implications for the shaping and acceptance of different prosodic representations (grids or trees) (Vigário 2010: 489).

The main factors that influence the prosodic phrasing are (i) the morphosyntactic structure, (ii) the constituent length and (iii) the information structure of the sentence, referring to the focus constituents. These factors have been translated into OT constraints in later studies: ALIGNXP: Align the right edge of an XP with the right edge of a φ; BINARY: The first φ of the sentence must contain minimally two Pws; ALIGNFOC: Align the left edge of a FOC-constituent with the left edge of a φ (Gussenhoven and Jakobs 2011: 250-251).

2.3.3. *The Clitic Group (CG) and the Prosodic Word Group (PWG)*

The inclusion of clitics (broadly defined as a hybrid class of small, function words) into the Pw has been extensively debated in linguistic studies. Scholars have made assumptions, based on cases from various languages. For example, in the case of Italian, Nespor and Vogel (2007) have brought arguments for the introduction of a new category within the Prosodic Hierarchy, the Clitic Group, positioned between the Pw and φ, since clitics are included in several processes at word-level (Vigário 2010: 486). The motivation for this addition is to accommodate function words that are neither elements which can be included in the same Pw as their host word, nor elements that are able to form a Pw on their own (Gussenhoven and Jacobs 2011: 258). Studies following the initial proposal for the introduction of the CG have argued for a distinction between maximal and minimal projections for adjunction structures. Therefore, there is a need to introduce a distinct category, separate from the CG, in order to explain diverse phonological facts, as the CG appears to be insufficient. For example, preverbal clitics in Italian have received different interpretations: (1) Vogel (2009: 78-79) considers that a clitic like *gli* stands outside the domain of word stress assignment, arguing for the following representation: [$_{CG}$gli[$_{\omega}$perdono]] 'I forgive him', where the Pw is the domain of the word stress, not the CG; (2) Bennett (2018: 23) considers that preverbal clitics are, in fact, incorporated into a recursive Pw structure, considering it a different projection levels of the Pw (non-minimal Pw); (3) Peperkamp (1997: 100-204) places these clitics

inside the domain of the φ, not necessarily forming a recursive Pw with the verb[2]. These different approaches highlight the varying interpretations of language facts and require further investigations.

Vigário (2010) argues for the introduction of a new category in the Prosodic Hierarchy, placed between the Pw and φ, to replace the CG, in order to account for facts found in many languages, which could not be included in the CG. The author opts for this label because this constituent is a Pw, but it does not necessarily include clitics. Therefore, the domain of the PWG is broader than that of the CG. Other attempts to include a new category were mentioned in previous studies, but under different labels (phonological cluster, composite group) (Vogel 2009).

3. COMPOUND WORDS IN ROMANIAN

Word compounding represents a word formation process, having as a result a new word, created from multiple independent already existing words in the language (*floarea-soarelui*, *doisprezece*), from compounding elements that cannot be independently used in the language (*geolog*) or from abbreviated words (*Plafar* etc.) (Ciobanu and Hasan 1970: 7).

Compound words represent a special category with respect to the Pw, because they represent words, being included in different grammatical categories (nouns, verbs, adjectives). However, each constituent of a compound may form a distinct phonological domain, impacting various phonological processes involved, such as word stress, syllabification (Gussenhoven and Jakobs 2011: 255). The first term of a compound begins the structure, as a lexical category.

In Romanian, compound words are relevant for the Pw determination if we take into account certain aspects, such as their cohesion degree. Coteanu (2007: 71) and Groza (2004: 140-141) point out that the cohesion degree of a structure's components varies depending on the age of the entire structure in the language. Sometimes, the overall meaning of the structure is closer to its components' meaning, while in other cases, there is no semantic link between them, particularly when word compounding is based on a metaphor (e.g., in botany, it is the case for many plant names: *traista-ciobanului, ochiul-boului* etc.). Spelling may not always reflect the in-use cohesion degree between the structure's components. There can be open, closed or hyphenated compound words: *precum că*, *locțiitor, câine-lup*, which can sometimes be modified in usage: *câinele său lup* (see below). However, in most cases, when the components are more fused, they form a single unit (e.g., *rozalb*), regardless of the grammatical category of the components. Conversely, less fused compounds may be hyphenated or open (*nou-născut, încât să* etc.) (DOOM[3]: 125-138).

In order to determine if a structure represents a compound word or not, several criteria have been taken into account: (a) the semantic cohesion of the compound structure, which is supposed to express a new meaning, to have a different referent than its components; (b) the morphological criteria examines the compound behaviour in relation to its components' internal position (genitive case markers); (c) the syntactic

[2] We mention that the Italian example mirrors a Romanian structure. Anca Chereches (2014) analyzes the prosody of Romanian pronominal clitics using an OT framework and reaches the conclusion that they are free prosodic words in all cases, except when they combine with each other.

perspective concerns both the internal syntactic structure of the compound (coordination or subordination way of word compounding), and the syntactic usage for the entire compound (Coteanu 2007: 72).

A morphological classification illustrates the fact that compound words include words from all grammatical categories, which is an argument for the compounding high productivity in Romanian (DOOM[3]: 125-141).

The component elements of a compound structure tend to combine, to lose their autonomy, merging into a new unit, which represents the final step in getting a new compound word (Groza 2004: 139-140).

Compound words represent a difficult category to determine the Pw for, as their structure can be (i) recognized (*mâine-seară*), (ii) semi-recognized, if one of the components modifies its form (***pai****sprezece*) or (iii) its original components merged completely (*mujdei* < *must de ai* juice of garlic 'garlic juice'). Therefore, we attempt to illustrate the prosodic structure of compounds and analyze them from an OT perspective.

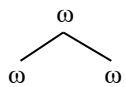
3.1. Diversity of compounds' internal structure

It is important to note that only closed compound words have been selected for this demonstration. This decision is based on the fact that hyphenated and open compounds could have, in some cases, an ambiguous status, between compounds and word combinations. Another argument for this selection refers to the different morphological behavior of the same type of compound, which influences the projection of the Pw. For instance, consider the compounds *câine-lup* and *floarea-soarelui* – both are compound nouns, yet they appear to differ in their cohesiveness degree. While *câine-lup* allows for a possessor between its components *câinele său lup*, *floarea-soarelui* does not – **floarea sa soarelui*).

Normative dictionaries indicate the number of stresses in compound words. According to the principle "one word, one stress", the entire compound receives only the compound's primary stress, not for each component, regardless of the number of components or word's length. Secondary stress is typically marked only in polysyllabic words containing compounding elements.

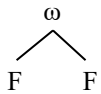
Following Itô and Mester (2021) perspective, we consider that closed compound words in Romanian may exhibit one of the following structures:

(a) Coordinative recursion



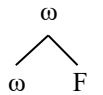
[altə. ˈdatə]

(b) Strictly layered

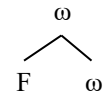


[a.ʃa.ˈdar]

(c) Adjunctive recursion



[un.de.ˈva]



[cu.ˈmin.te]

We notice that the superior intersection of the compound corresponds to a Pw, but not all the component elements constitute Pw (some represent prosodic categories subordinated to the Pw, such as the F).

The predominant structure among closed compounds is illustrated in (a), where two Pw are formed ([bunəˈwarə]). The second model in (c) present compounds with a non-Pw attached to the left ([atoʧtiuˈtor]).

3.2. OT constraints

In section 2 we have introduced several general OT constraints, relevant for the analysis of the Pw and accounting for its different forms. Here, we present them in more detail, in order to establish their ranking in Romanian, specifically for the closed class of compounds.

Making use of syntactic structures, the role of the constraint MATCH (X^0, ω) is to map morphological/syntactic words to Pw (Itô and Mester 2011: 5): MATCH (X^0, ω) – Assign one violation mark for every terminal node X^0 in the syntax such that the segments belonging to X^0 are not all dominated by the same prosodic ω in the output.

Another constraint is WORDBINARITY: Prosodic words must be binary. It is violated by words formed of no more than a single foot (Itô and Mester 2011: 9). Binarity is a requirement observed at other levels in the Prosodic Hierarchy (F binarity, φ binarity).

In order to account for the short compound components that violate the WORDBINARITY, constraint MATCHHEAD is introduced: Assign one violation mark for every terminal node X^0 in the syntax that is the head such that the segments belonging to X^0 are not all dominated by the same Pw in the output.

The interaction of these constraints can be illustrated through examples of compounds. WORDBINARITY outranks MATCHHEAD and MATCHX^0, allowing a compound component consisting of one foot not to be parsed as a Pw. Therefore, in (b), the first candidate is the winner, despite the fact is violates MATCH-H and MATCHX^0, but it respects WORDBIN.

In (c), the winning candidate is selected by the MATCH, since the third and the fourth candidates violated WORDBIN, one of the compound components being a single foot. When the two compound members respect WORDBIN, the result can be a winning candidate, as in (a), where the last candidate does not violate any constraint.

From these examples, it can be concluded that the constraint hierarchy is the following: WORDBIN>> MATCH-H>> MATCHX^0.

(a)

ω[ωω]	WORDBIN	MATCH-H	MATCHX^0
[ω(altə)(datə)]		*	**!
[ω[ω(altə)](datə)]]		*	*!
[ω(altə)[ω(datə)]]			*!
☞ [ω[ω(altə)][ω(datə)]]			

(b)

ω[FF]	WORDBIN	MATCH-H	MATCHX^0
☞ [ω(aʃa)(dar)]		*	**
[ω[ω(aʃa)](dar)]]	*!	*	*
[ω(aʃa)[ω(dar)]]	*!		*
[ω[ω(aʃa)][ω(dar)]]	**!		

(c)

ω[ωF]	WORDBIN	MATCH-H	MATCHX^0
[ω(unde)(va)]		*	**!
☞ [ω[ω(unde)](va)]]		*	*
[ω(unde)[ω(va)]]	*!		*
[ω[ω(unde)][ω(va)]]	*!		*

(d)

ω[Fω]	WORDBIN	MATCH-H	MATCHX^0
[ω(cu)(minte)]		*	**!
[ω[ω(cu)](minte)]]	*!	*	*
☞ [ω(cu)[ω(minte)]]			*
[ω[ω(cu)][ω(minte)]]	*!		

It is worth noting that, as other compound words are taken into account, the hierarchy could be subject to modification, with the introduction of another constraint, PARSE-F-into-ω. This constraint could account for compounds whose members are lexical words consisting of only one F (for example, the compound [roz.ˈalb], which follows the structure in (a), would violate WORDBIN and no candidate could win. Therefore, the introduction of PARSE-F-into-ω would alter the ranking of constraints in the hierarchy.

4. CONCLUSIONS

In Romanian, compound words represent a productive class, with a complex internal structure, from a prosodic perspective. Their classification into fully compound words or word combinations is relevant for both the prosodic hierarchy and the Pw.

By limiting the analysis to closed compound words only, it becomes evident that WORDBIN is a highly ranked constraint for the Pw, similar to how binarity is important at other levels of the Prosodic Hierarchy, like in the F (FOOTBINARITY).

Another important aspect refers to the Pw mapping their corresponding syntactic words as well, since the Pw represents the minimal relevant constituent for syntax. Therefore, MATCH-type constraints are part of the hierarchy.

Further comprehensive analysis of this issue is required, considering other types of compound structures and introducing prosodic categories above the Pw. Such investigations will contribute to a deeper understanding of the prosodic analysis of compound words in Romanian.

REFERENCES

Bennett, R., 2018, "Recursive prosodic words in Kaqchikel (Mayan)", in *Glossa: a journal of general linguistics*, 3, 1, 67, 1–33.

Chereches, A., 2014, "A prosodic analysis of Romanian pronominal clitics", *U. Penn Working Papers in Linguistics*, 20, 1, 1–10.

Ciobanu, F., L., Sfîrlea, 1970, *Cum scriem, cum pronunțăm corect. Norme și exerciții*, București, Editura Științifică.
Coteanu, I., 2007, *Formarea cuvintelor în limba română*, București, Editura Universității din București.
Dixon, R. M. W., A. Y. Aikhenvald, 2002, "Word: a typological framework", in *Word: A Cross-Linguistic Typology*, Cambridge, Cambridge University Press, 1–41.
DOOM^3 = I., Vintilă-Rădulescu (coord.), 2021, *Dicționarul ortografic, ortoepic și morfologic al limbii române*, ediția a II-a revăzută și adăugită, București, Editura Univers Enciclopedic Gold.
Dresher, B. E., 1996, "Introduction to Metrical and Prosodic Phonology", in L. Morgan, K. Demuth (eds), *Signal to Syntax*, Lawrence Erlbaum, 41–54.
Groza, L., 2004, *Elemente de lexicologie*, Humanitas, București.
Gussenhoven, C., H. Jakobs, 2011, *Understanding phonology*, third edition, Hodder Education.
Hildebrandt, K. A., 2018, "The Prosodic Word", in J. R. Taylor (ed.) *The Oxford Handbook of the Word*, Oxford, Oxford University Press, 254–278.
Hyman, L. M, 2008, "Directional asymmetries in the morphology and phonology of words, with special reference to Bantu", *Linguistics*, 46, 2, 309-350.
Itô, J., A., Mester, 1992, "Weak layering and word binarity", in T. Honma, M. Okazaki, T. Tabata, S. Tanaka (eds), *A New Century of Phonology and Phonological Theory. A Festschrift for Professor Shosuke Haraguchi on the Occasion of his Sixtieth Birthday*, Tokyo, Kaitakusha, 26–65.
Itô, J., A. Mester, 2007a, "The onset of the prosodic words", in S. Parker (ed.), *Phonological Argumentation. Essays on Evidence and Motivation*, Equinox, London.
Itô, J., A. Mester, 2007b, "Prosodic Adjunction in Japanese compounds", *MIT Working Papers in Linguistics*, 55, 97–111.
Itô, J., A., Mester, 2009, "The extended prosodic word", in B. Kabak, J. Grijzenhout (eds) *Phonological Domains. Universals and Derivations*, Berin and New York, 135–194.
Itô, J., A., Mester, 2021, "Recursive Prosody and the Prosodic Form of Compounds", *Languages*, 6, 65, 1–24.
Nespor, M., I. Vogel 2007, *Prosodic phonology: with a new forward*, Berlin, Mouton de Gruyter.
Peperkamp, S., 1997, *Prosodic Words*, The Hague, Holland Academic Graphics.
Schiering, R., B., Bickel, K., Hildebrandt, 2010, "The prosodic word is not universal, but emergent", *Journal of Linguistics*, 46, 657–709.
Selkirk, E., 1986, "On derived domains in sentence phonology", *Phonology Yearbook*, 3, 371–405.
Selkirk, E., 1978, *Phonology and Syntax: The Relation between Sound and Structure*, Cambridge, MA, MIT Press.
Selkirk, E., 1996, "The prosodic structure of function words", in *Signal to Syntax*, Lawrence Erlbaum, 187–213.
Selkirk, E., 2000, "The interaction of constraints on prosodic phrasing", in M. Horne (ed.), *Prosody: Theory and experiment*, Dordrecht, Kluwer Academic Publisher, 231–261.
Vigário, M., 2010, "Prosodic structure between the prosodic word and the phonological phrase: Recursive nodes or an independent domain?", *The Linguistics Review*, 27, 485-530.
Vogel, I., 2009, "Universals of Prosodic Structure", in S. Scalise, E. Magni, A., Bisetto (eds), *Universals of Language Today*, Springer, 59–82.