Chapter 2
The Organisation of StressTyp

Rob Goedemans
Harry van der Hulst
Ellis Visch

2.0 Introduction

StressTyp stands for Stress Typology. StressTyp is a database containing information about word-level stress patterns of (some of) the languages of the world. The database has been implemented in 4th Dimension, a database program for Apple Macintosh. Work on StressTyp began in 1992 as a pilot project of EUROTYP (1990-1994), a project on the typology of European languages, financed by the European Science Foundation (ESF). StressTyp is not intended to be limited to European languages, however, and work on StressTyp has continued after 1994.[1]

The primary aim of this chapter is to describe the (record) structure of StressTyp. Section 2.1 provides information about the *goal* of StressTyp, and section 2.2 and 2.3 specify the *sources* that have been used and indicate the (current) *limitations* of StressTyp. Section 2.4 gives a detailed description of the *record structure*. It explains the names of the fields and the values that may appear in each field. This chapter takes the coding system as the point of departure and relates it to the terminology used in theoretical stress research. It thus serves as a legend on the coding system that we have used.

---

[1] We thank the ESF and the Faculty of Arts of Leiden University for financial support for a MacIntosh computer and printer. Through those who work on StressTyp additional financial support comes from the Netherlands Foundation of Scientific Research, NWO, the Holland Institute of Generative Linguistics (HIL), the Department of General Linguistics of Leiden University and the Department of General Linguistics of the Free University of Amsterdam.

The structure of the record was developed by Harry van der Hulst (HIL, Leiden), in collaboration with Aditi Lahiri (then at the Max Planck Institute, Nijmegen). The relevant equipment was made available by a grant from the EUROTYP project and further support of the Faculty of Letters of Leiden University. Kees van der Veer (Max Planck Institute, Nijmegen) implemented the record structure in 4th Dimension. After a period of using the database Rob Goedemans made a number of structural modifications.

For the relation between coding system and *linguistic terminology* from the viewpoint of the latter we refer to chapter 3. Taking linguistic terminology as the point of departure, this chapter explains how various types of stress systems are coded in the database.

2.1 Goal of the project

The goal of StressTyp is to offer a quick entry to the primary and secondary literature on stress systems of the languages of the world. By primary literature we mean grammars and articles that provide descriptions of stress patterns, examples and the like. By secondary sources we refer to theoretical works on stress. The coding system makes it possible to search through the database for the occurrence of quite specific properties. With the search facilities of 4th Dimension, StressTyp can be instrumental in testing and developing hypotheses (after the current limitations of StressTyp have been dealt with); examples of searches are provided in the appendix to this chapter.

In StressTyp we have tried to encode properties of stress systems without making a commitment to a specific theoretical framework, although it is necessarily the case that a coding system implicitly embodies theoretical decisions. We have tried, however, to encode the relevant properties without avoiding redundancy, allowing for duplication of information and fields for informal statements and comments.

In the more distant future we aim at embedding StressTyp in a network of related databases that provide information on various aspects of stress research, such as an annotated bibliography of stress (StressBib, currently in progress), a terminological database (StressTer, currently in progress), addresses of linguists who do research on stress (StressRes) and so on. The present global indicator for this imaginary network is StressEx (Stress Expert System).

By making the database available to other researchers we hope to benefit from their knowledge (or personal databases in whatever form) and co-operation in adding more languages to the system, and improving the quality of information presently contained in StressTyp.

2.2 StressTyp sources

Although the work on StressTyp started in 1992, StressTyp is still in its first phase, due to the fact that we have had only limited means to develop it more rapidly.

The first phase involves extracting (often elementary) information about stress systems from (cross)linguistic studies. Both strictly typological and theoretical sources have been consulted.

The collection of material was started by Aglaia Cornelisse. She did not enter material into the database. The first sources consisted of the 'usual typological studies' such as Hyman (1977), Greenberg and Kashube (1978), Hayes (1981/95), Lockwood (1982), Halle and Vergnaud (1987) and so on, i.e. studies that provide the empirical basis for much of the present-day theoretical discussion. Later, Simone Langeweg assisted in adding data from more sources. The collected data were transferred into the database format (that had meanwhile been implemented) by Rob Goedemans and Ellis Visch. They then started checking the records for consistency and correctness and also added new languages. Checking is done by consulting the original sources, and often additional theoretical or descriptive studies. In this phase Ruben van de Vijver provided valuable assistance.

We plan to further check the data that are now contained in StressTyp and to add new information, both of languages already included and of new languages. We also plan to extend the scope of the database by means of a questionnaire.

2.3 Limitations

Our first goal has been to store information in the database on as many languages as possible. This means that the languages included now do not form a well defined sample by any criterion. Some language families or linguistic areas are clearly overrepresented (e.g. Australian languages) or underrepresented (Native American and South American languages).

The data that StressTyp contains are as trustworthy as the information we found in the sources. If that information is wrong, StressTyp has copied that wrong information. (Of course, whenever we had any reason to believe that the information was wrong we did not copy it.) We have tried to trace the information back to the original descriptive source wherever this was possible.

Specifying values in database fields necessitates interpreting sometimes very limited information. This means that the information in StressTyp is very often rather incomplete. The information stored for each language ranges from very elementary statements (like "initial stress", all further fields unspecified) to fairly detailed specifications for a number of fields. The record allows information on morphological structure, but this is mostly unspecified.

Misinterpretations on our part have undoubtedly been made. The coding system requires interpretation of the sources. Records are not always faithful to any particular source. Where we have consulted more than one source for one language an attempt has been made to reconcile the sources. In doing so we may have come up with a coding that does not correspond to an actually existing dialect or language variety. Another factor that may have attributed to inconsistencies is that various people have been involved in the coding. Every record specifies the sources on which we have based the coding. It is unavoidable that the translation of properties of certain systems (as described in our sources) into the format of the database is not always entirely straightforward, because they have special or sometimes even conflicting features. Often, there are different possible ways of storing properties of systems. In those cases, choices have been made to fit the description into the format of the database. In chapter 3, our choices are clarified. It exemplifies all types of stress systems identified in the theoretical literature with example languages.

Despite its limitations, our own experience is that StressTyp can be helpful in developing and testing hypotheses by offering data and properties of different languages in an identical format.

We emphasize that StressTyp cannot be held responsible for providing incorrect, or incomplete information. We encourage those who use StressTyp in publications to check crucial information in the original descriptive sources or with native speakers. We welcome all corrections and additions both regarding specific languages and the overall organization.

2.4 Description of the record fields

2.4.0 Introduction/general remarks

In this section, all the record fields of the stress database are described in detail. All descriptions of the record fields are accompanied by sample languages from the database to provide as much clarification to the user as possible. The bibliography at the end of this chapter does not contain references to all these languages, but such references are available through StressTyp.

StressTyp is based on a theoretical perspective that competes with other metrical theories. Its starting point is the separation of main stress and rhythm (cf. chapter 1). The separation of information on main stress and rhythmical structure is not commonly assumed, but we emphasize, that it is possible to interpret the information without making a commitment

to this particular view. The parameters used in this database embody a particular view on stress, but they are presented as purely 'descriptive'.

Since this chapter follows the structure of StressTyp rather closely, some information on databases might be in order. A database contains *records* (this chapter basically describes the contents of a StressTyp record). These are packages of information relating to one item. In our case the records are related to languages. One record contains all the information we were able to find for a certain language. The information in a record is organised into *fields*. A field contains one piece of information or a set of related pieces. In our database, for instance, we have fields for *language name, affiliation, examples, stress type* etc.

In general, all abbreviations (indicating a field value or parameter setting) are exemplified in this chapter by a reference to one or more example language(s). Sometimes, no examples of an abbreviation occur in the database, which we indicate by "no example in StressTyp". This could be merely a coincidence or a systematic gap. No claims are made with respect to the relative markedness of a parameter setting.

Finally, if a field is completely irrelevant to a particular stress system, this is indicated by a hyphen. If information is not available for a field that is relevant, this may be indicated by a blank or question mark. In general, blanks indicate that no effort has so far been made to uncover the information elsewhere. A question mark indicates that the relevance of the record field is uncertain or that there is doubt about the interpretation of the information in the source. Sometimes a field does contain information but it is still accompanied by a question mark. In those cases, the information may be correct, but conclusive evidence is not available.[2]

2.4.1 Language information

LANGUAGE

The name of the language or dialect in the naming-conventions of Voegelin & Voegelin (1977). If the language is referred to by other names in the literature, these names are also indicated.

[2] A general remark concerning the notation used in the database is in order. In a large subset of record fields, simple abbreviations are used to express information, such as **Y**(es) or **N**(o). In some cases, the interpretation of an abbreviation is dependent on the contents of a specific record field. For example, **L** can stand for *left*, expressing for instance at which edge footing starts, or for *last* in which case a specific syllable or position in the word is singled out. This practice is not unusual in metrical phonology either. The StressTyp user must bear these ambiguities, which we hope to eliminate from future versions, in mind.

DIALECT OF

The name of the mother language of which the language specified in the record field language is a dialect.

PATH

The family tree, root first (following Voegelin & Voegelin 1977).

AREA

All geographical areas in which the language, or dialect of the language in question, is spoken.

Example:

Language	Bidyara-Gungabula
Dialect of	Mandandanji
Path	Australian - Pama-Nyungan - Pama-Maric - Mari
Area	South Queensland (Australia)

2.4.2 STRESS TYPE

2.4.2.1. List of items/ abbreviations and connectives

The record field **STRESS TYPE** indicates the main stress type of the language by means of a code, identifying the position(s) of main stress. It can either be a simple abbreviation, or a combination of abbreviations and one (or more) connective(s).

It is general practice to describe so called unbounded systems by statements like "last heavy, or first". "Last heavy" could identify any syllable in the word, even the first one if this is the only heavy syllable present. The first part of this statement, therefore, singles out only specific syllables, i.e. heavy ones, and identifies which one of these syllables is promoted. It identifies a specific syllable in a domain, rather than a specific position. The second part of the statement, however, identifies a specific position, ~~in this~~ case the first syllable in a domain.

The same practice is used in this database. An abbreviation not followed by a connective identifies a specific position in a word; when followed by a connective its interpretation slightly changes and identifies a specific syllable **within** a domain. The following combinations occur (**x** and **y** are variables representing a set of values that are actually used in the coding):

x	= **position** x
x/	= x **if heavy**
x;	= x **and**
% x	= **shift to** x
x-	= x is **special case**
(x)	= **additional information** x to clarify and extend a preceding code

Thus, if it is followed by a connective, **x** is never the only stress position. Moreover, a code can contain more than one connective. The following complex codes can be identified in StressTyp (to be exemplified later):

x/y	=	**x** if heavy, otherwise **y**
x;y	=	**x** and **y** both occur in free variation, but **x** is dominant
x/y;z	=	**x** if heavy, otherwise **y**; in free variation with **z**, but **x/ y** is dominant
x%y	=	**x** if heavy, otherwise shift to **y**
x%y/z	=	**x** if heavy, otherwise shift to **y** if **y** is heavy, otherwise **z**
x-y/z	=	**x** is special case; if there is no **x** in domain, **y** if heavy, otherwise **z**

In general, the first abbreviation in a code identifies the special case, the final one the default case.

The following items can occur in a code. For convenience, we have indicated the specific interpretation of the codes when followed by the connective /, which is the connective that is used most often in StressTyp.

UNBOUNDED SYSTEMS

L	= last	**L/**	= last heavy
F	= first	**F/**	= first heavy

BOUNDED SYSTEMS

I	= initial	**I/**	= initial heavy
S	= second	**S/**	= second heavy
T	= third	**T/**	= third heavy
U	= ultimate (final)	**U/**	= ultimate heavy
P	= penultimate	**P/**	= penultimate heavy
A	= antepenultimate	**A/**	= antepenultimate heavy

MISCELLANEOUS

NMS = no main stress, all stresses are perceived equally prominent
CNT = main stress depends on counting of the syllables in a word
LEX = main stress is lexical
IRR = main stress is irregular
Pitch = main stress interacts with pitch-accent assignment
Tone = main stress interacts with tone-assignment

The code is deduced from information stored in record fields in the **formal analysis** of section 2.4.3 below. For many substitutions of **x** and **y** by the values above there are no actual examples in the database. An interesting question is whether these are coincidences or systematic gaps.

2.4.2.2. Guidelines and exemplification of codes

Single abbreviations are used to identify **quantity-insensitive** systems. In systems where main stress is always found on a fixed syllable with reference to one of the edges of words, irrespective of the form of this syllable. Such systems are usually described as bounded, but we include systems that have been described, or could be described, as unbounded (see chapter 3 section 3.1.2 for more details). The code-character identifies a syllable in a specific position in the stress domain.

I main stress is on the first syllable
cf. Maranunggu

S main stress is on the second syllable
cf. Araucanian

T main stress is on the third syllable
cf. no example in StressTyp

U main stress is on the ultimate or final syllable
cf. Weri

P main stress is on the penultimate syllable
cf. Warao

A main stress is on the antepenultimate syllable
cf. Macedonian

Usage of the connective "/"

The slash is used to identify systems in which the position of main stress is influenced by weight, i.e. for **quantity-sensitive** systems (see also section 2.4.3.1. **stress weight** and 2.4.4. **heavy for stress**). Their code contains two symbols **x** and **y**, separated by the connective slash: **x / y**. The first symbol 'x' identifies the position of main stress if the stress domain contains heavy syllables. The second symbol 'y' identifies the default case, i.e. only light syllables are present in the stress domain. Systems that are identified with more than two symbols are special and other connectives are used. They will be exemplified below.

QUANTITY-SENSITIVE UNBOUNDED systems are divided into the four (familiar) types using **L** and/or **F** and the connective /. The stress domain comprises the entire word (cf. chapter 3 section 3.1.1).

L/L main stress on **last heavy** syllable in domain;
if no heavy syllables (i.e. only light syllables in domain) on **last** syllable cf. Golin

F/F main stress on **first heavy** syllable in domain;
if no heavy syllables (i.e. only light syllables in domain) on **first** syllable cf. Amele, Lhasa Tibetan

L/F main stress on **last heavy** syllable in domain;
if no heavy syllables (i.e. only light syllables in domain) on **first** syllable cf. Chuvash

F/L main stress on **first heavy** syllable in domain;
if no heavy syllables (i.e. only light syllables in domain) on **last** syllable cf. Komi-Permyak

In **QUANTITY-SENSITIVE BOUNDED** systems (see also chapter 3 section 3.2.2), the stress domain, which is called a window, is limited to two syllables and located at or nearby one of the two edges of a word. Four situations can occur, where square brackets indicate the edges of the stress domain, on either the left or the right side of the word:

(1) a. [h h] c. [h l]
b. [l l] d. [l h]

The code that StressTyp gives specifies **only** what happens in cases (1a)

and (1b). The symbol 'x' before the slash identifies the position of main stress when **both** syllables are **heavy**, i.e. (1a). The symbol 'y' after the slash identifies the position of main stress when **both** syllables are **light**, i.e. (1b).

When the two syllables in the stress domain are not equal in weight, cf. cases (1c) and (1d), it is always the heavy syllable that bears main stress. This information is considered to be redundant; therefore, (1c) and (1d) are left out of consideration in determining the code. In the code, we use the following terminology:

(2) # [I S] [P U] #
| | | |
first last first last[3]

Four main stress codes will be mentioned here that have a stress domain aligned to the **left** edge of the word. The stress type contains a choice from the symbols **I** (initial) and **S** (second), identifying 'first' and 'last' in the domain and the connective /. Schematized:

[3] The informal use of a 'first' or 'last' terminology reflects the idea that unbounded and bounded systems really are very much alike, differing only with respect to the domain that is relevant (cf. Chapter 1). On the other hand, in unbounded systems, 'first' or 'last' does not necessarily correspond to a specific position, whereas in bounded systems 'first' or 'last' uniquely identifies a specific position, i.e. a specific syllable, in the two syllable domain. This, and an attempt to stay with common terminology, has been the reason for using different abbreviations for bounded and unbounded systems in StressTyp. In the formal analysis we do not use the traditional labels, but rather strictly l(eft) and r(ight).

It is our intention, however, to eventually replace all abbreviations by 'last' and 'first'. As a consequence of this decision, domains in which the syllables are unequal in weight will be covered implicitly by the code too. Consider as an illustration an unbounded system like (i) and a bounded system like (ii):

(i) a. * b. * c. * d. *
[l l h l h l] [h l l l] [l h l l] [l l l l]

(ii) a. * b. * c. * d. *
#[h h] #[h l] #[l h] #[l l]

Both systems can be described as 'first/first'-systems: if there is more than one heavy syllable, the first heavy one catches the accent; if there are no heavy syllables the first syllable catches the accent. However, if there is only one heavy syllable that syllable receives the accent, independently of its position in the word. Since there is only one heavy syllable it can be captured by the statement: first heavy in domain. With respect to bounded systems, the information at which edge of the word the two-syllable domain is located is not expressed in the **stress type** code any more. However, since this information is stored elsewhere too, i.e. in record field **stress domain**, this is not seen as a loss.

(3) # [I S]
| |
first last

I/I main stress is on **first heavy** syllable in domain;
if no heavy syllables, stress is on the **first syllable**

* *
#[h h] #[l l] cf. Southeastern Tepehuan

I/S main stress is on **first heavy** syllable in domain;
if no heavy syllables, stress is on the **last syllable**

* *
#[h h] #[l l] cf. Hopi

S/I main stress is on **last heavy** syllable in domain;
if no heavy syllables, stress is on the **first syllable**

* *
#[h h] #[l l] cf. Capanahua

S/S main stress is on **last heavy** syllable in domain;
if no heavy syllables, stress is on the **last syllable**

* *
#[h h] #[l l] cf. no example in StressTyp

The following stress codes can be identified when the stress domain is aligned to the **right** edge of the word. The stress code contains a choice from the symbols **P** and **U**, identifying 'first' and 'last' in domain, and the connective /. Schematized:

(4) # [P U]
| |
first last

U/U main stress is on **last heavy** syllable in domain:
if no heavy syllables, stress is on the **last syllable**

* *
[h h]# [l l]# cf. Yapese

U/P main stress is on **last heavy** syllable in domain;
if no heavy syllables, stress is on the **first syllable**

```
  *     *
[h h]#  [l l]#        cf. Jicaque
```

P/U main stress is on **first heavy** syllable in domain;
if no heavy syllables, stress is on the **last syllable**

```
 *      *
[h h]#  [l l]#        cf. Aklan
```

P/P main stress is on **first heavy** syllable in domain;
if no heavy syllables, stress is on the **first syllable**

```
 *     *
[h h]#  [l l]#        cf. Awadhi
```

More complex are systems in which **extrametricality** interferes. The two-syllable stress domain is shifted one syllable from the edge. These systems are identified by the use of the symbols **T** and **A** in their code. Schematized:

(5) #σ[**S** **T**] [**A** **P**] σ#
first last first last

As an example, consider:

P/A stress domain comprises the antepenultimate and penultimate syllable.
main stress is on the **last heavy** syllable in domain;
if no heavy syllables, stress is on the **first syllable**

```
  *     *
[h h] σ#  [l l] σ#    cf. Latin
```

Usage of the connective ";"

The semicolon is used in those cases where there is a certain degree of unpredictable free variation in the position of main stress, though restricted to some specific positions. Therefore, it is used typically for

bounded systems. In general, there is one dominant and regular pattern, but there are exceptions, found at (a) specific position(s). Therefore, main stress is not to be considered irregular. For example:

P;A main stress is penultimate in most cases; in a few cases, stress occurs on the antepenultimate syllable. These antepenultimate cases have to be specified lexically. cf. Tacana

See also, the section on IRR below, for the contrast with the use of **IRR**. The semicolon can also be preceded by more specific information, eg:

P/U;P main stress is penultimate if heavy, otherwise main stress is ultimate. Exceptions to this dominant pattern are always found at the penultimate syllable. Therefore, it is not an irregular system. cf. Aklan

Usage of the connective "-"

The hyphen occurs in the code for systems in which there is a not a simple contrast in syllable-quantity or syllable-prominence, but syllable types must be represented on a scale. (e.g. many languages make a three-way contrast between light, heavy and superheavy syllables). In many of the languages that have more than two weight distinctions, main stress can be found on more than two different positions. The special case (i.e. 'superheavy') is indicated in front of the stress code, followed by the hyphen. After the hyphen, the code contains two symbols and the connective / which covers the predictable variation of stress within a two-syllable domain (cf. chapter 3 section 3.2.2). For example:

U-P/A main stress is on the **final** syllable if it is '**superheavy**'; if such a syllable is absent, stress is on the **last heavy** syllable in domain, i.e. the penultimate syllable; if no heavy syllables in domain, main stress is on the **first** (=antepenultimate) syllable. cf. Damascene Arabic

L-P/A main stress is on the **last 'superheavy'** syllable in the entire word; if such a syllable is absent, main stress is on the **last heavy** syllable in a two-syllable domain located one syllable away from the right edge of the word; if no heavy syllables are present in the two-syllable domain stress is on the **first syllable** in that domain. cf. Klamath

Usage of the connective "%"

The percentage mark is used to identify what we call 'broken-window' systems (see also chapter 3 section 3.3.1). These are quantity-sensitive systems that show a shift outside the two-syllable main stress window in case both syllables inside the window are *light*. The syllabic position to which main stress will be moved is identified and preceded by the percentage mark:

P%A/P main stress is on **first heavy** (=penultimate) syllable in (right edge) domain; otherwise, i.e. if (l l), there is a shift to the antepenultimate syllable, under the condition that this syllable is **heavy**; otherwise, i.e. the antepenultimate syllable is **not heavy**, main stress is on **first** syllable in domain (=penultimate).

```
 *       *     * ←
[h h]#  l [l l]#  h [l l]#      cf. Maithili
```

U%A main stress is on **last heavy** syllable in domain; otherwise, i.e. if two light syllables, main stress is shifted to the antepenultimate syllable, irrespective of its make-up.

```
  *     * ←     * ←
[h h]#  l [l l]#  h [l l]#      cf. Hindi
```

Other categories

NMS

NMS there is no distinction between primary and non-primary stresses, i.e. they are all perceived as equally prominent according to a descriptive source. cf. Central Siberian Yupik

X (NMS) there is no distinction between primary and non-primary stresses, but X (i.e. a specific position is always stressed. cf. Gidabal I (NMS): stress occurs on the first syllable, regardless of weight and on every syllable which surfaces with a long vowel

If **NMS** is used without any further specifications, the record fields on main stress contain hyphens, indicating that these fields are irrelevant.

The relevant information is stored in the record fields on rhythm.

However, in X(**NMS**) cases, the record fields of the formal analysis are filled as if there were a distinction between main stress and rhythm. We suspect that most systems requiring NMS fall in this category because we believe that in most NMS systems, a distinction between main stress and secondary stress can nevertheless be made. The fact that a certain syllable is always stressed may strengthen this hypothesis.

CNT

CNT is always used between parentheses and preceded by either **L** or **F**. It is used to identify what we call 'count systems' (cf. Van der Hulst 1992). In these systems the position of main stress depends on counting syllables in terms of feet, which, in most cases, leads to a difference in the position of main stress between odd-numbered words and even-numbered words. These systems are analyzed as unbounded systems because the whole word is relevant for the computation of main stress. Therefore, we use **L** or **F**.

L(CNT) **last foot** in word carries main stress, i.e. stress is found at the right edge of the word. cf. Cayuga

F(CNT) first foot in word carries main stress, i.e. stress is found at the left edge of the word. cf. MalakMalak

Notice, that last/first must not be interpreted as last/first assigned foot. It is always the last assigned foot which carries main stress. Last/first identifies at which edge main stress is located, as in the unbounded systems described in the section on the connective "/".

LEX

LEX is used to identify systems that use lexical accent marking. This can mean that main stress is unpredictable and can be found on any syllable in the word, although each word has a predictable stress position once the locations of the lexical accents are given. These positions are diacritically marked for each word. Most typically, additional information must be specified, **LEX** is added in those cases, sometimes between parentheses.

LEX main stress is completely unpredictable and can be found anywhere in the word. Main stress has to be specified in the lexicon by means of a diacritic accent or marker. cf. Spokane

F/F (LEX) main stress is on the first syllable that has a diacritic marker; otherwise, i.e. if there are no diacritic accents, main stress is found on the first syllable cf. Russian

L/LEX main stress is on the last heavy syllable; if there are no heavy syllables, i.e. only light syllables, main stress is unpredictable and can be found anywhere in the word. In those cases, stress has to be specified in the lexicon by means of a diacritic accent. cf. Mountain Cheremis

IRR

IRR is used if main stress varies unpredictable in a disyllabic domain in a word. They are bounded. Often, there is one (or more) identifiable pattern(s). Exceptions, however, cannot be captured by specifying a specific position. Therefore, the use of **IRR** is additional information, often added between parentheses or preceded by the connective semicolon (indicating free variation, see discussion on ";" above).

Consider as an example Sangil, in which the dominant pattern can be described by **P/P**. There are exceptions to this pattern in both ways, i.e. a final light or heavy syllable can receive main stress, or, if both syllables are unequal in weight, the light syllable can catch main stress instead of the heavy one. Its code will be **P/P;IRR**. This code expresses the fact that there are exceptions which cannot be specified more precisely.

Now consider Dutch. Part of the Dutch stress system is that words ending in three light syllables have either stress on the antepenultimate syllable or on the penultimate syllable, the latter being dominant: *página / pyjáma*. Stress on the final syllable is truly exceptional. Although the stress system of Dutch is more complicated, this part of the system could be captured by the code: **P;A (IRR)**. The use of the semicolon says that there is variation between **P** and **A**, with **P** as dominant pattern. The addition of **IRR** implies that the second position is regarded as exceptional.

We are aware of the somewhat vague use of **IRR** in general, although its interpretation in specific cases will be clear.

Pitch and tone

Pitch and **Tone** are used incidentally and sometimes between parentheses to indicate that there is an interaction between pitch or tone assignment and metrical structure. 'Pitch' or 'tone' are added if these terms are used as such in the record field **quote** which is based upon a theoretical or descriptive source. Specifics about interaction between pitch/tone and metrical structure assignment are provided in the record fields stress repair and remarks.

stress type: pitch
cf. Chocktaw

stress type: L/- (CNT: tone)
cf. Seneca

2.4.2.3 Stress type information continued

SOURCE QUOTATION

A quotation that describes the stress pattern of the language in words (not necessarily literally quoted from the literature) with reference (author, year and page numbers).

EXAMPLES

Examples that illustrate the various stress patterns in the language. Examples are glossed and represented with the help of IPA phonetic symbols (not nessecarily IPA translations). Sounds for which this IPA set gives no symbol may be represented by any standard typographic symbol not otherwise used in the description.

DESCRIPTIVE SOURCE

Complete reference to authors that have done field work on the language. If a reference is preceded by an asterisk, the source has been consulted by us.

THEORETICAL SOURCE

Reference to authors that have analyzed the stress pattern in a certain framework. Only author and year are given. Complete references are provided in a library file (StressBib). If no analysis exists but a name is given anyhow, address and E-mail address of the person who is responsible for the information of the language entry can be found in the address file (StressAdr). The information provided by this person is translated by us into the database framework. If no names are given at all, StressTyp is responsible for incorporating the information on the basis of a descriptive source.

2.4.3 Formal analysis

2.4.3.1 Main stress

In this database we have made a distinction between main stress and rhythm.

The (main) stress domain contains exactly two syllables, irrespective of their weight in case of a bounded system, and it comprises the entire word in case of an unbounded system. If extrametricality occurs, the stress domain is shifted one syllable inside the word.

There are seven record fields for information on main stress. On the basis of these seven fields on main stress, the **stress type** code of the language is determined (see section 2.4.2. above).

STRESS DOMAIN

L domain is bounded and located at the left (L) edge of the word. cf. Djirbal, Capanahua

R domain is bounded and located at the right (R) edge of the word cf. Polish, Yapese

U domain is unbounded (U), i.e. the entire word (modulo extrametricality). cf. Chuvash, Khalka Mongolian

- domain is irrelevant, i.e. there is no distinction between primary/non-primary stresses. cf. Central Siberian Yupik (but see also Gidabal: **I (NMS)**

STRESS EXTRAMETRICALITY

L extrametricality occurs on the left (L) edge of the word
cf. Iliaura/Alyawarra, Western Aranda

R extrametricality occurs on the right (R) edge of the word
cf. Macedonian, Chamorro

N extrametricality is not (N) involved
cf. Yapese, Capanahua

STRESS EXTRAMETRICALITY UNIT

seg segment is extrametrical
cf. Alyawarra (vowel)

mora mora is extrametrical
cf. Hindi

syll syllable is extrametrical
cf. Chamorro, Macedonian

foot foot is extrametrical
cf. Unami

STRESS WEIGHT

Y the position of main stress is determined on the basis of syllable-quantity, syllable-prominence, rhythmically strong positions or diacritically marked syllables: weight is Yes (Y)
cf. Capanahua (syllable weight): quantity-sensitive system
cf. Pirahã (syllable-prominence): prominence system
cf. MalakMalak (rhythmically strong positions): count system
cf. Russian (diacritically marked syllables): lexical accent

N the position of main stress is determined without reference to syllable-quantity, syllable-prominence, or rhythmically strong positions, i.e. main stress assignment is quantity-insensitive: weight is No (N). cf. Garawa, Cavinena

Specific information about what counts as heavy in a system is stored in the record field **heavy for stress**.

STRESS HEAVY SYLLABLE

L stress is located on the leftmost (L) **heavy** syllable in domain
cf. Southeastern Tepehuan, Aklan (bounded)
cf. Amele, Komi-Permyak (unbounded)

R stress is located on the rightmost (R) **heavy** syllable in domain. cf. Capanahua, Jicaque (bounded)
cf. Chuvash, Golin (unbounded)

- system is quantity-**in**sensitive
cf. Araucanian, Garawa

In a **bounded** quantity-sensitive system this field specifies the position of stress in a domain with two heavy syllables, i.e. [h h]. If the two syllables in the stress domain are unequal in weight, i.e. [h l] or [l h], it is always the heavy syllable which bears stress (disregarding exceptions).

STRESS if no HEAVY SYLLABLE

tr in two syllable-domain, stress is on lefthand syllable, i.e. trochaic: [* .]. cf. Garawa, Piro (QI-bounded)
cf. Capanahua, Jicaque (QS-bounded)

ia in two syllable-domain, stress is on righthand syllable, i.e. iambic: [. *]. cf. Araucanian, Weri (QI-bounded)
cf. Hopi, Yapese (QS-bounded)

L in unbounded domain, stress is on first syllable, i.e. located at left (L) edge of word. cf. Khalka Mongolian, Chuvash

R in unbounded domain, stress is on last syllable, i.e. located at right (R) edge of word. cf. Golin, Komi-Permyak

For a **bounded** quantity-**in**sensitive system, this is the only field that provides information about the position of main stress. For a **bounded** quantity-**sensitive** system, this field specifies the position of stress in a domain with two light syllables, i.e. [l l][4].

STRESS REPAIR

Y there is a **shift** outside the two-syllable stress window, if both syllables inside the window are **light**.
cf. Manam, Maithili

N no repair, stress falls always inside the two-syllable window
cf. Maranunggu

In chapter 3 (section 3.3.1), we refer to systems in which there is a shift outside the two-syllable window, as 'broken-window' systems. The **stress code** of this type of system contains the connective **%**. Information about the specific conditions under which (main) stress shift occurs, can be found in the record field **Repair**.

[4] For bounded systems, the terms 'iambic' and 'trochaic' are used: within the two-syllable domain one of the two syllables is stressed. In unbounded systems, **L** or **R** are used to express at which edge stress is located since it makes no sense to speak of iambic/trochaic here, but we could have used L and R for bounded systems as well.

2.4.3.2 Rhythm

There are nine fields providing information about non-primary stress, headed under RHYTHM. The rhythm domain consists of all syllables of a word, i.e. including the syllables of the stress domain. Rhythmic feet are assigned in such a way that the syllable that is stressed in the stress domain is also stressed in the rhythm domain (see also chapter 3 section 3.2.1.2). In other words, rhythm is sensitive to the presence of main stress, rather than the other way around, as is mostly assumed in the metrical literature.

RHYTHM

Y there is a distinction between main stress and non-primary stress (Yes) cf. Garawa, Maranunggu, Seri

N only main stress is present, i.e. there are no non-primary stresses (No). cf. Chamorro

RHYTHM DIRECTION

L rhythm, i.e. non-primary stress assignment, starts at the left (L) edge of the word. cf. Maranunggu

R rhythm, i.e. non-primary stress assignment, starts at the right (R) edge of the word. cf. Garawa, Warao

E rhythm is bi-directional and echoes away from both edges (E) ("edge-in") of the word. cf. Auca

C rhythm is bi-directional and echoes away from main stress in both directions towards the edges (C=Centrifugal).
cf. no example in StressTyp

RHYTHM EXTRAMETRICALITY

L extrametricality occurs on the left (L) edge of the word
cf. Iliaura, Western Aranda

R extrametricality occurs on the right (R) edge of the word
cf. Macedonian, Chamorro

N extrametricality is not (N) involved
cf. Yapese, Capanahua

RHYTHM EXTRAMETRICALITY UNIT

seg segment is extrametrical
cf. Western Aranda

mora mora is extrametrical
cf. Hindi

syll syllable is extrametrical
cf. Palestinian Arabic

RHYTHM WEIGHT

Y the position of non-primary stress is determined on the basis of syllable-quantity, i.e. weight is Yes (Y)
cf. Palestinian Arabic, Aklan, Seri

N the position of non-primary stress is determined without reference to syllable-quantity, i.e. the system is quantity-insensitive, weight is No (N). cf. Weri, Capanahua

Specific information about what counts as heavy in the system is stored in the record field **heavy for rhythm**.

RHYTHM TYPE

tr the general rhythmic pattern is trochaic, i.e. regular alternation of a stressed and an unstressed syllable, i.e. a sequence of (* .). cf. Maranunggu, Garawa

ia the general rhythmic pattern is iambic, i.e. regular alternation of an unstressed and a stressed syllable, i.e. a sequence of (. *). cf. Aklan, Cayuga

bo both (bo) "iambs" and "trochees" are found
cf. no example in StressTyp

\- no regular **alternation** of non-primary stresses
cf. Gidabal in which only long vowels are stressed

The regular alternating pattern can be disturbed by weight. In some quantity-sensitive cases, it is crucial that trochaic is interpreted as a moraic trochee, i.e. either a sequence of two light syllables (l l) or one heavy syllable (h) (cf. Hayes 1995); this is specified in the record field **remarks**, cf. Cahuila, Cairene Arabic. In all other cases, the trochee may

be taken to be unbalanced (i.e. just avoiding (lh) and (hh)).

RHYTHM REPAIR

Y the rhythmic surface patterning deviates in some cases from the one specified by the fields **Rhythm Type** and/or **Rhythm Ternary** and must be repaired (Yes). cf. Chugach

N the rhythmic pattern does not need repair (No).
cf. Maranunggu

This field is specified as Yes, for instance in cases in which rhythm is both ternary and quantity-sensitive; in this case, a situation could arise where syllables are skipped or left over in parsing. In some cases these unparsed syllables are supplied with structure or incorporated. Specific information about the conditions on repair is stored in the record field **Repair** (see also chapter 3 section 3.3.1).

RHYTHM ITERATIVE

Y there is more than one non-primary stress; rhythm is assigned iteratively (Yes). cf. Maranunggu, Garawa

N there is only one non-primary stress; rhythm is assigned non-iteratively (No). cf. Jabulajabula, Nengone

RHYTHM TERNARY

N. the rhythmic pattern is binary, not ternary (No)
cf. Aklan, Maranunggu, Capanahua

ia the rhythmic pattern is ternary; the rhythmic foot comprises a complex head of two syllables and a non-head of one syllable. the relation between the two syllables within the complex head is iambic; the relation between complex head and non-head is specified in the field **Rhythm Type**

cf.	**Rhythm Type**	**Rhythm Ternary**
Chugach	([. *].) trochaic	iambic (=amphibrach)
Not in StressTyp	(.[. *]) iambic	iambic (=anapest)

tr the rhythmic pattern is ternary; the rhythmic foot comprises a complex head of two syllables and a non-head of one syllable. the relation between the two syllables within the complex head is trochaic; the relation between complex

head and non-head is specified in the field **Rhythm Type**

cf.	**Rhythm Type**	**Rhythm Ternary**
Cayuvava	([* .] .) trochaic	trochaic (=dactylus)
Sentani	(. [* .]) iambic	trochaic (=amphibrach)

For details, see chapter 3.

2.4.3.3 Degenerate feet and subminimal words

DEGENERATE FEET

Y there are incomplete feet (i.e. monosyllabic or monomoraic) to express **non-primary** stresses at left-over syllables at the edge of a word. cf. Maranunggu

N there are no incomplete feet (i.e. monosyllabic or monomoraic), i.e. there are no **non-primary** stresses at left-over syllables at the edge of a word. cf. Warao

Since in this stress database, stress and rhythm are separated, **degenerate feet** =Yes covers rhythmic properties only. This field does not provide any information about the occurrence of monosyllabic words. This is specified in a separate field **Subminimal Words**. The settings of these two fields are determined independently.

SUBMINIMAL WORDS

Y there are subminimal words
cf. Araucanian, Maranunggu

N there are no subminimal words
cf. Cayuga, Maithili

In quantity-sensitive systems, subminimal words are monomoraic, i.e. light syllables; in quantity-insensitive systems subminimal words are monosyllables, irrespective of their content[5]. Specific information about monosyllables, i.e. whether they are stressed or not, can be found somewhere else in the record, most likely in **source quotation** or **remarks**.

[5] Sometimes it is claimed that subminimal words in Quantity-Insensitive systems are heavy, as for example in Pintupi. This information is stored in the **remarks** field.

2.4.4 Basic syllable types

TEMPLATE

Specification of the types of syllables that occur in the system, e.g. CV, CVC etc.

OBLIGATORY ONSETS

Y onsets are obligatory
N onsets are not obligatory

BRANCHING ONSETS

Y branching onsets are allowed
N branching onsets are not allowed

LONG VOWELS

Y long vowels are present underlyingly
N long vowels are not present underlyingly

CLOSED SYLLABLES

Y there are closed syllables
N there are no closed syllables

GEMINATES

Y there are geminates
N there are no geminates

CODA RESTRICTIONS

Specification of coda restriction if any. If not: -

HEAVY FOR STRESS

Indication of what counts as heavy with respect to main stress in the system: syllable or segment types (possibly in the form of a scale) (i.e. syllable-quantity or syllable-prominence), diacritically marked syllables or rhythmically strong positions (see chapter 4).

HEAVY FOR RHYTHM

Indication of what counts as heavy with respect to non-primary stress(es) in the system: syllable or segment types that are heavy for rhythm, i.e. syllable- quantity.

2.4.5 Repair strategies

REPAIR

This field specifies under which conditions **Stress Repair** and/or **Rhythm Repair** occur. With respect to so-called 'broken-window' systems, containing the connective % in their **Stress Type** and with **Stress Repair** set to **Y**, this fields specifies whether shift outside the two-syllable stress window, containing (l l), takes place always or only if the target syllable is heavy, see the use of the connective %. With respect to so-called 'ternary' systems, for which **Rhythm Repair** is set on **Y**, this field specifies under which circumstances unparsed syllables are grouped into a foot after all or under which circumstances reparsing occurs (see also chapter 3 section 3.3.1). Finally, **Repair** can contain information about destressing of syllables.

2.4.6. Exceptional patterns

EXCEPTIONAL PATTERN

Description of the exceptional stress patterns that have to be specified in the lexicon.

EXAMPLE PATTERNS

Examples of exceptional stress patterns. Examples are glossed and represented with the help of IPA phonetic symbols. Sounds for which this IPA set gives no symbol may be represented by any standard typographic symbol not otherwise used in the description.

EXCEPTION SOURCE

Reference to source of exceptions (author + year+ pagenumber(s)).

2.4.7 Processes

PROCESSES

Description of (phonological) processes that occur in the language (especially the ones related to stress).

EXAMPLE PROCESSES

Examples of the processes stated above. Examples are glossed and are represented with IPA phonetic symbols in the LaserIPA TM version. Sounds for which the IPA set has no symbol are presented by any typographic symbol not otherwise used in the description.

2.4.8 Remarks

REMARKS

This field can contain any (additional) information relevant with respect to the system, especially if certain choices have been made to fit a more or less problematic system into the format of the database.

It contains information about the (un)stressedness of monosyllables if subminimal words occur in the language.

It can specify the relation between metrical structure and tone or pitch assignment if the **Stress Type** contains an indication for Tone or Pitch.

And, finally, it states explicitly that all stresses are perceived as equally prominent if the **Stress Type** contains **NMS** between parentheses.

2.4.9 Morphology and stress

In these entries information about the role of morphology is stored, i.e. the behaviour of prefixes and suffixes, clitics, the stress patterns of compounds, etcetera. This part of the stress database is in a very preliminary state. Record fields are incorporated but only in a few cases information has been stored.

2.5 Conclusion

In this chapter we have described the record structure of StressTyp. We realize that the coding system can be improved in a number of ways, but since it is our experience that this is always the case (even after changes have been made), we felt that the time had come to present the system to its potential users. We welcome comments and help in any form. The following chapter in this book approaches the structure of StressTyp from the view point of current (metrical) theory. In a volume that will form a sequel to this one, we plan to present a selection of the record of StressTyp and outcomes of searches through the database. (Some example searches are already presented in the Appendix to this chapter.) In this way we hope to make StressTyp information available to those who (for whatever reason) cannot consult the database at any desired moment.

References

Dresher, E. & A. Lahiri (1991). The Germanic Foot: Metrical Coherence in Old English. *Linguistic Inquiry* 22: 251-286.

Goedemans, R., H. van der Hulst & E. Visch (1996). *The StressTyp Manual*. Ms, Leiden University.

Greenberg, J. & D. Kashube (1978) Word prosodic systems: a preliminary report. *Working Papers on Language Universals* 20, 1-18

Halle, M. & J.-R. Vergnaud (1987). *An Essay on Stress*. Cambridge, Massachusetts: MIT Press.

Hayes, B.P. (1981). *A Metrical Theory of Stress Rules*. Doctoral Dissertation, MIT, Cambridge, Massachusetts (distributed by the Indiana University Linguistics Club, Bloomington, Indiana).

Hayes, B.P. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.

Hulst, H.G. van der (1992). The independence of stress and rhythm. Discussion paper, *Phonologica 7*. Ms, Leiden University.

Hyman, L. (1977). On the nature of linguistic stress. In: L. Hyman (ed.) *Studies in Stress and Accent*, 37-81. University of Southern California, Los Angeles.

Lockwood, D.G. (1982). Parameters for a typology of word accent. *Lacus Forum* 9: 231-240.

Prince, A.S. (1983). Relating to the grid. *Linguistic Inquiry* 14: 19-100.

Rice, C. (1992). Binarity and ternarity in metrical theory: parametric extensions. PhD Dissertation, University of Texas, Austin.

Voegelin, C.F. & F.M. Voegelin (1977). *Classification and Index of the World's Languages*. New York, Oxford, Amsterdam: Elsevier.

Appendix A: Searches

1 Introduction

One of the most important advantages of storing information in a database is the fact that a database programme allows you to extract records that share a certain characteristic quickly, methodically and (usually) without errors. We call this operation a *search*.

It will be clear that the possibility to search the database for records that have the contents of one of the fields in common might be very useful to a theoretical linguist who works on stress theory. This would allow the phonologist to make lists of languages that have say, initial stress and no degenerate feet, or stress on final superheavy syllables. When, in the future, the database is extended to contain more languages in a subset that is more evenly distributed over the language families, searches might be used for quantitative studies that tell the linguist what percentage of the world's languages have a certain characteristic "y" in common. However, we must warn the user that the results of a search should always be interpreted in light of the structure of StressTyp itself, and must never be taken for granted. As pointed out earlier, choices have been made to incorporate systems. It is the task of the linguist to decide whether or not the results of a search have more general validity by inspecting the contents of the parameter setting(s) of (a) record field(s).

This appendix describes how searches can be executed in StressTyp. Some examples are provided to illustrate the usefulness of such searches.

2 Searches in StressTyp

2.1 The logics

On certain occasions one might want to select a set of records that have a certain feature in common. To do this we need to use the *search editor*, which is an integral part of the 4th Dimension programme. We can best illustrate how this works theoretically by dealing with a few examples (a practical description of how to execute searches is given in Goedemans, Van der Hulst & Visch 1996). This manual describes in detail the characteristics (theoretical and practical) of StressTyp.

Suppose one wants to make a list of all the languages that have the stress type I (stress on the first syllable). We must then search the database using a logical filter that says something like: select all the records for which the field *stress type* is filled with I, and nothing but I.

To execute our example we must make sure that we select only the languages for which *stress type* is "I". We can do this by entering a filter expression into the search editor. These have the following format:

(1)	*field*	*logical expression*	*content*
eg.	stress type	is equal to	I

If we had wanted the list to contain the languages that have some other stress position besides "I" we would have selected a different logical expression like "is equal to or larger than" (i.e. "contains").

If we execute this search on the computer, a list of languages should appear that only contains languages that have "I" in the Stress column.

Other, more complicated questions we may put to the database may involve more difficult queries. In these queries the operators *And, Or, Except* may be used. A few example queries are given below.

If we want to know which non-Indo-European languages are in the database we type:

(2) affiliation / does not contain / Indo-European

When we are interested in languages that have degenerate feet but no subminimal words:

(3) deg feet / is equal to / Y AND
subminimal words / is equal to / N

And; which languages use syllable weight for either stress or rhythm:

(4) str weight / is equal to / Y OR rhy weight / is equal to / Y

In principle these searches can be expanded to contain three or even four sub-expressions, but be careful. Searches with only *And* or only *Or* between expressions can be safely used, but searches which use combinations of the operators may yield unexpected answers. Consider the following example: we want to get a list of languages that have a domain at the right edge and either stress or rhythm weight. The search expression would then be:

(5) str domain / is equal to / R AND str weight / is equal to / Y OR rhy weight / is equal to / Y

If we execute this query we get a list of languages that have either a right edge domain and stress weight or **any domain type** and rhythm weight. Clearly not what we expected. The snag here is that 4th Dimension cannot use brackets in search expressions, it just processes the commands in linear order. The dominance relation, which should look like: x and (y or z), cannot be expressed. We can circumvent this problem by rephrasing the search expression such that linear processing produces the right results. It should look like this:

(6) str weight / is equal to / Y OR rhy weight / is equal to / Y AND
str domain / is equal to / R

Another way to go about it makes use of the "search in selection" option provided by the search editor. Normally, a search that is carried out after an earlier search, uses the **entire list** of languages as its input, and not just the subset that was the result of the earlier search (this is a point to keep in mind). When you click the "search in selection" box, however, the subsequent search does take the subset as its input. We can use this to our advantage when solving the bracketing problem. We must first execute the query "str domain - is equal to - R". The result is a list of right edged domain languages in which we can search for those that have either stress or rhythm weight by creating a new expression with the rest of the old expression (str weight = y OR rhy weight = y). Especially when expressions contain more than two operators, this latter method is more surveyable and far less cumbersome.

2.2 Examples

2.2.1 Introduction

We conclude this appendix with a few examples. These may illustrate why database searches are useful. The example searches in this section will have the following format. We express in words which subset of languages we wish to extract. Then an example follows in which we put the search expression and the quantitative results. This example is followed by a discussion of expected and unexpected results of the search. Currently (March, 1996), 268 languages are stored in StressTyp, which by no means form a well-defined sample by any criterion. Some language families or linguistic areas are clearly overrepresented (e.g. Australian languages) or underrepresented (e.g. languages spoken in South America). Since StressTyp is in progress, the results presented in this appendix are preliminary and only meant as illustrations of how

searches can be executed. No quantitative conclusions can be drawn on the basis of these results.

2.2.2 Weight

Let us first try to make a list of the languages that have quantity-sensitive main stress, but in which rhythm is quantity-insensitive. Notice that the parameter values are opposite. Cases like these are interesting in view of the hypothesis that main stress and rhythm are separate entities. In a foot-based analysis we expect that all feet obey the same restrictions.

(7) stress weight / is equal to / Y AND
rhythm weight / is equal to / N 11 out of 268

Language	Stress Type	Area
Capanahua	S/I	Peru
Djaru	F/F	Western Australia
English	U/P;A	World
Inga	U/P	Colombia
Kutenai	P/U	Alberta
Madimadi	S/I	Australia (NSW)
Maithili	P%A/P	India/Nepal
Malakmalak	F (CNT)	Australia (NT)
Murawari	F/F	Australia (NSW,Q)
Seneca	L/- (CNT:tone)	Ontario, New York
Yidiɲ	F/F	Australia (Q)

(area names have been adapted to fit the table: NSW = New South Wales, NT = Northern Territory, Q = Queensland)

All these languages have main stress on a specific marked syllable. There is a great diversity of systems in the table. As noted in the introduction, we must be careful in taking the result for granted. In the bounded systems, syllable weight determines the position of main stress like in English, Capanahua. The assignment of non-primary stress however is by

no means influenced by the syllable make-up. We also see two count systems in the table, cf. MalakMalak and Seneca. In these two languages, rhythm is quantity-insensitive, but main stress is computed on the basis of foot heads (although Seneca is a very special case, see discussion in chapter 3, section 3.5.3). In StressTyp, we have incorporated this property by saying that main stress is sensitive to rhythmically strong positions. The fact that these systems show up in the table is therefore caused by the way we have set up StressTyp. There are also three unbounded systems. Main stress can lie on any syllable in the word that has a specific syllable make-up. However, in all three systems there is at least one non-primary bounded stress foot. These cases are very interesting in the light of standard metrical theory, in which the presence of a bounded rhythmic foot is unexpected in an unbounded system (see also chapter 3, fn 3). The existence of these systems seems to support the underlying theory of StressTyp. Again, we have to add that the fact that these systems show up in the search might be caused by the way in which we have incorporated the specific properties of these systems. This is not the place to go into details, but it is certainly true that these systems qualify for closer inspection.

Finally, we find Maithili. The Stress Type contains the percentage mark, indicating that there is something special in the system. However, it is undoubtedly an example of opposite specifications of weight, although rhythmic footing is non-iterative. A detailed discussion of this languages can be found in chapter 3, section 3.3.1.

In a similar fashion we can also make a table for languages that have quantity-insensitive main stress and quantity-sensitive rhythm. We expect this option to be very marked, but in the perspective of the underlying theory of StressTyp nevertheless existing.

(8) stress weight / is equal to / N AND
rhythm weight / is equal to / Y 5 out of 268

Language	Stress Type	Area
Estonian	I	Estonia
Gidabal	I;S (NMS)	Australia (NSW)
Irish Gaelic	I	Ireland
Koi, Koya	I	India
West Greenlandic	U (NMS)	Greenland

We see mainly initial systems. Non-primary stresses occur on syllables with a specific syllable make up. Two cases contain the indication NMS between parentheses. According to the descriptive sources there is no perceptually perceived distinction between stresses. However, in addition to stressed syllables with a certain make up, there is always one syllable that is stressed independently of what that syllable looks like. Consider Gidabal. In addition to all syllables that contain a long vowel, the first syllable is also stressed. To be able to express this ambiguous behaviour, we have specified the record field for stress as if there were a distinction between stresses. Therefore, they show up in the list.

The following exercise is also very useful to illustrate that we have to handle the results with care (from now on we will refrain from printing result tables, but we will indicate stress types that are relevant). Suppose we searched for the languages that have main stress weight:

(9) stress weight / is equal to / Y — 97 out of 268

In the light of the previous two search jobs, we could conclude that the number of languages in the database that have both rhythm weight *and* stress weight is 86, since we know there are 11 languages that have stress weight but no rhythm weight and 97 - 11 = 86. This is not the case however. Some extra searches are needed to reveal why.

(10) stress weight / is equal to / Y AND
rhythm weight / is equal to / Y — 28! out of 268

but stress weight / is equal to / Y AND
rhythm weight / is equal to / — 38 out of 268

stress weight / is equal to / Y AND
rhythm weight / is equal to / - — 20 out of 268

Together these account for the 97 languages that have stress weight. Thus, we have to remember that implication relations do not work both ways because of the fields that have blanks (we have no information), question marks (we are in doubt about the sources) or intended hyphens (we are sure there is no rhythm and fill the rhythm fields with hyphens).

The same argumentation holds if we ask to represent all the languages that have rhythm weight. StressTyp produces a list of 42 cases out of 268. In (8), we already have seen 5 systems that have no stress weight. Again we might (wrongly) conclude that there are therefore 37 cases which have both stress weight and rhythm weight, but as we have seen in (10) above this is not the case: there are only 28 cases. Again we have to realize that there could be systems in StressTyp for which we do not have information about stress weight (11a) or in which stress weight is irrelevant because there is no main stress at all (11b).

(11) a. stress weight / is equal to / AND
rhythm weight / is equal to / Y — 2 out of 268

b. stress weight / is equal to / - AND
rhythm weight / is equal to / Y — 7 out of 268

We can conclude on the basis of these illustrative searches that there are many cases for which the relevant information is lacking at present. In drawing conclusions one must be aware of these misleading effects.

2.2.3 Rhythm type and footing direction

A different test might involve checking whether there is a correlation between footing direction and foot type. It has often been observed that in most languages that have iambic rhythm, footing starts at the left edge of the word. We can execute the following searches to find out (again bearing in mind that there are cases in which there is rhythm reported but for we cannot decide whether the foot type is iambic or trochaic on the basis of the available information):

(12) rhythm type / is equal to / ia AND
rhythm direction / is equal to / L — 18 out of 128[6]

rhythm type / is equal to / ia AND
rhythm direction / is equal to / R — 5 out of 128

There are 23 cases in which the rhythm type is specified as iambic in StressTyp. 18 of them have left-to-right footing and qualify as genuine iambic languages. They have either main stress on the second syllable (except when the first syllable is heavy) or are count systems with iambic feet, i.e. stress type: L (CNT), or they have no main stress but only rhythmic feet, like the Yupik dialects (of course, no value should be attributed to the qualitative number because StressTyp does not contain a representative sample of languages of the world at present).

5 cases seem to contradict the observation mentioned above, i.e. iambic footing is not from left-to-right but from right-to-left. Again, the result must not be taken for granted. A closer inspection of the languages on the list, reveals that all of them are more or less suspect, or can be reanalysed as having moraic or syllabic trochees according to the theoretical

[6] We first selected the languages that have rhythm to begin with.

literature. Our list contains Aklan, Big Nambas, Cyrenaican Bedouin Arabic, Sentani, Tübatulabal and Weri. For Aklan, it is difficult to exclude an alternative with right-to-left moraic trochees (cf. Hayes 1995:265; see also chapter 3, section 3.5.1). Kager (1989) reanalysed Tübatulabal by arguing that main stress is placed in final position prior to foot construction, followed by top-down assignment of moraic trochees. For Big Nambas, the footing direction could be left-to-right, but data needed to decide on this issue are missing. Sentani is a special case since it has a ternary rhythm in which the complex head itself is trochaic.

Finally, Weri has main stress on the final syllable and non-primary stress is found on every other preceding syllable. In StressTyp, the system is described as having a stress domain at the right edge with stress if no heavy: R and with rhythm type iambic. This is the most direct and simple description that covers the observed pattern. However, it is not impossible to reanalyse the system as trochaic, since we adopted a correspondence rule that demands that main stress is always respected by rhythmic footing. StressTyp is ambiguous in this respect. The latter description would adhere more closely to views that deny the existence of quantity-insensitive iambic feet (note that we do not exclude this option).

This example shows us that the results of a search must be carefully considered. However, it offers a quick way of finding some interesting cases with respect to a specific hypothesis. This is what StressTyp is intended for. It does not replace theoretical argumentation however.

We can also make a list of languages with a trochaic foot type:

(13) rhythm type / is equal to / tr AND rhythm direction / is equal to / R — 26 out of 128
rhythm type / is equal to / tr AND rhythm direction / is equal to / L — 46 out of 128

As we can see there are many systems with right-to-left trochaic footing (not unexpected) as well as systems with left-to-right trochaic footing. The latter number, however, might be somewhat flattered since StressTyp contains a relative large number of Australian languages, which are often quantity-insensitive and have initial main stress. Also a relative large number of Arabic dialects are incorporated in StressTyp. Many of these dialects can be qualified as count systems, i.e. main stress is found somewhere at the right edge of the word but footing starts at the left edge (see chapter 5, for a detailed discussion of Arabic dialects). No quantitative conclusions can be drawn yet.

The searches in (12) and (13) cover 95 out of 128 cases for which rhythm is specified as Yes. For the missing 33 languages information

about the rhythmic foot type is either not available or inconclusive, or the system is unbounded and has non-alternating, secondary stresses on specific (heavy) syllables, in which case the foot type is "-".

2.2.3. Degenerate feet and subminimal words

StressTyp is (eventually) meant to be a tool for the testing of hypotheses and the quick retrieval of possible counterexamples. Of course, we have to reformulate theoretical questions into the format of StressTyp. In this section, we will show how this can be done without drawing any real conclusions for the theories under consideration. This would be too premature, because StressTyp is still in a too preliminary state at present.

It is claimed in the literature by Hayes (1995) that monosyllabic (or monomoraic) feet are only allowed in some languages under main stress (the 'weak ban' on degenerate feet). This covers languages which start the parse at one side of the word and have main stress at the opposite side. In words with an odd-number of syllables this would leave us with an incomplete foot.

(14) a. L-->R trochaic footing:
```
  *          *
(* .) (* .) (*)  and  (* .) (* .)
```

b. R-->L trochaic footing:
```
  *          *
(*) (* .) (* .)  and  (* .) (* .)
```

c. L-->R iambic footing:
```
    *          *
(. *) (. *) (*)  and  (. *) (. *)
```

d. R-->L iambic footing:
```
  *          *
(*) (. *) (. *)  and  (. *) (. *)
```

In StressTyp, there is a record field degenerate feet: Y/N. As a consequence of our decision to describe properties of stress and rhythm independently, this field expresses rhythmic properties only. Whenever we observe a non-primary stress in a position that can only form an incomplete foot, we describe this situation by saying that there are degenerate feet (see for a more detailed discussion chapter 3, section 3.2.1.2). Therefore, before we make a list of languages that have degenerate feet, we must realize how the systems in (14) are incorporated into the database format, to be able to interpret the results correctly.

Systems like (14a) would be analyzed as count systems, in which the position of main stress depends on a count of syllables in terms of feet.

The rhythmically strong positions count as heavy with respect to main stress. The record field degenerate feet is specified as Yes, since we clearly need incomplete feet to predict the position of main stress in (14a). In systems like (14b), the initial syllable is always stressed. In StressTyp, these systems are analyzed different from those in (14a). They have a stress domain at the left edge of the word with initial stress. Rhythmic footing starts at the opposite edge and respects the presence of main stress. Although the initial foot on the rhythm plane can consist of a single syllable, we have decided not to consider these feet as degenerate. In some sense, they are heavy for rhythm. The record field degenerate feet is set to No (an example of this type is Garawa, see 3.2.1.2). Systems like (14c) are analyzed in the same way as (14b): they always have final main stress. Of course, from a theoretical point of view, it would be interesting to reconsider both these types of systems and reanalyze them as count systems. Finally, systems with iambic footing from right-to-left are rare as we have seen in the preceding section. For systems like (14d), degenerate feet is specified as Yes.

Suppose that we want to make a list of all the languages for which degenerate feet = Yes. What we expect to find is a list which contains at least examples of systems like (14a) and (14d), but also cases in which we need incomplete feet to cover non-primary stresses. These languages are potentially problematic for Hayes' claim, that degenerate feet are only allowed under main stress. We get the following result. There are 15 cases out of 268 that have degenerate feet specified positively (as far as information is available of course).

Interestingly enough, there are no examples of systems like (14a) and (14d) in StressTyp. Hayes presents Auca as an example of (14a). However, the source does not indicate any particular stress as the strongest. It is probably a NMS system in which an incomplete rhythmic foot is needed and therefore, it is a potential counterexample. What we call count systems do not seem to allow incomplete feet. The position of main stress varies between two positions near one edge, which can be calculated on the basis of proper feet only (cf. 3.1.3).

All 15 cases have incomplete feet to express non-primary stress. Most of these cases are initial trochaic systems, like Maranunggu, in which final syllables in odd-numbered words have secondary stress:

(15) *
(* .) ?
mérepèt

Hayes claims that the observed secondary stress in these systems should not be attributed to the presence of incomplete feet, but to other, more phonetic factors, like final lengthening.

In the literature, it has also been argued that there is a correlation between the occurrence of incomplete feet and subminimal stressed words. However, the range of this relation differs radically in different proposals. Hayes (1995) claims that if a language allows for degenerate feet under main stress, it will have subminimal words. Also, the occurrence of subminimal words indicate that the weak ban on degenerate feet is operative. However, it is not necessary that we will find degenerate feet in these languages. Consider Warao, in which subminimal words occur. Therefore, we expect degenerate feet to be allowed. However, the system operates in such a way that degenerate feet cannot be created in the position of main stress. It has penultimate main stress and trochaic footing from right-to-left:

(16) *
. (* .) (* .)
yi wa ra na e 'he finished it'

Therefore, the occurrence of subminimal words does not necessarily imply that degenerate feet are present in the language, only that it is possible. The reverse claim is stronger: if we need degenerate feet for main stress, then there must be subminimal words.

Another proposal concerning the relation between subminimal words and degenerate feet is made in catalexis theory (cf. Kiparsky 1992, Kager 1995). In this theory, the relation between these two is much stronger than in Hayes' proposal. Catalexis is the mirror-image of extrametricality. Whereas extrametricality explains cases in which certain elements are not seen by the stress rule, catalexis covers cases in which there seems to be more material than can be seen at the surface. In fact, degenerate feet do not exist at all, but are really complete feet with an additional empty element. The head of this foot can be a main stressed syllable or a non-primary stressed syllable. As an illustration, consider again the case of Maranunggu, in which the catalectic empty syllable is indicated between brackets:

(17) *
(* .)(* .)
merepet[σ]

Both final main stress and final secondary stress signal a catalectic element. It is claimed by proponents of catalexis theory that in these

cases subminimal words occur, which in the case of Maranunggu seems to be true. A potential counter example for catalexis theory would be a system that behaves like Maranunggu, but has no subminimal words.

Now let us see how the claimed correlations can be tested. Recall that StressTyp contains a separate record field subminimal words, which has no connection at all with the record field degenerate feet (they are specified independently). Four queries can be executed:

(18) a. degenerate feet /is equal to / N
subminimal words / is equal to/ N — 31 out of 268
b. degenerate feet / is equal to / Y
subminimal words / is equal to / Y — 8 out of 268
c. degenerate feet / is equal to / Y
subminimal words / is equal to / N — 5 out of 268
d. degenerate feet / is equal to/ N
subminimal words / is equal to / Y — 27 out of 268

The first query (18a) covers those cases that Hayes terms 'the strong ban on degenerate feet'. No incomplete feet are needed at all: left-over syllables at edges are not stressed. Subminimal words do not occur. These cases are unrevealing with respect to the hypotheses we are examining.

The second query (18b) is more interesting. Recall from the first query in this subsection that all cases in which degenerate feet = Yes, are cases in which incomplete feet are needed for non-primary stress.[7]

Almost all these systems are initial systems. These cases confirm the claim made by catalexis theory, that if there are non-primary stresses at edges, subminimal words occur in the language. As indicated above, the fact that there are subminimal words is seen by Hayes as an indication that degenerate feet can occur under main stress. That in these initial systems this situation does not arise, is in itself not problematic as was explained above for Warao. The only questionable aspect of these systems is that we find a non-primary stressed final syllable, for which Hayes has to call on other factors, like final lengthening.

The third query (18c) gives us a list of languages that are at least problematic for catalexis theory, but not necessarily for Hayes' theory. There are five cases in which degenerate feet are needed to describe non-primary stress, but no subminimal words occur.

[7] Notice that the total number of (18b) and (18c) is not equal to the result of the query 'degenerate feet/ is equal to / yes', which was 15 out of 268. For two cases no information is available on the occurrence of subminimal words.

Wangkumura is an initial stress system and behaves like Maranunggu, except that there are no subminimal words. Wangkumura is therefore problematic for catalexis theory, but not necessarily for Hayes, since he would say that the final stressing is caused by final lengthening. Moreover, according to Hayes this final stress disappears in connected speech.

Maithili has a very complicated stress system with main stress on the penultimate or antepenultimate syllable (cf. chapter 3, section 3.3.1). It has also one initial secondary stress, which can be of degenerate size in words with penultimate stress. This case clearly indicates that catalexis theory is on the wrong track here, because the degenerate size of the foot can only be explained by assuming an iambic foot here with a left-edge catalectic element, which is suspect because the main stress system is trochaic. Hayes explains the presence of an initial secondary stress by assuming a kind of beat addition: an intermediate layer is inserted in the grid and additional grid marks are assigned. (Notice that it is not possible to appeal on the mirror image of final lengthening in these cases, because there is no evidence for such a phenomenon.)

Aklan has a secondary stress on the final syllable following prefinal main stress. Since there are no subminimal words, degenerate feet are disallowed. The perceived final stress must be caused by final lengthening. Catalexis theory, however, has no explanation for the non-occurrence of subminimal words.

The fourth case is Southern Paiute, which has an iambic system with main stress on the second syllable. The rhythmic pattern is alternating, except that the final syllable is never stressed. This indicates that the final syllable is extrametrical. However, in a four syllable word, a non-primary stress is adjacent to the main stress. This case is problematic for both theories: here we need an incomplete foot to express a non-primary stress not located at an edge or under main stress. Both theories would deny the existence of degenerate feet (and thereby predict that there are no subminimal words in the language). The observed non-primary stress can be explaned by assuming that words in Southern Paiute must be parsed exhaustively. In a four-syllable word this can only be done by a reversal of foot type under pressure of the final extrametrical syllable.

The fifth case on this list is Banawa, which is strange in many respects (for instance the skipping of an initial vowel, cf. chapter 4) and maybe has to be reanalyzed if we have more information about the system.

The fourth query (18d) gives us a list of 27 languages. 4 cases are unbounded, including two count systems (see chapter 3, section 3.1.3). That they have no degenerate feet does not have to surpise us, because, although these systems can have non-primary stressed (heavy) syllables,

alternation is absent. These systems do not contradict the claims made by the two theories, since unbounded systems do not fall within the scope of the predictions. There are 2 systems which have final main stress. Catalexis theory would say that there is a final foot which contains a catalectic element predicting the occurrence of subminimal words. Within Hayes' theory, these systems might be analyzed by assuming that there is top-down stressing with an End Rule right, followed by footing which respects the main stress which is of degenerate size. The occurrence of subminimal words is predicted then.

In principle, all other 21 systems on the list are unproblematic for Hayes. The occurrence of subminimal words implies that incomplete feet might occur in the language, but only in the position of main stress. There is no correlation between subminimal words and rhythmic stresses.

However, these 21 systems are problematic from the perspective of catalexis theory. We find several different types on the list, i.e. main stress is initial, second, penultimate or a combination of these. The occurrence of subminimal words suggests that there is catalexis. However, for the assignment of main stress we do not need incomplete feet and there are also no secondary stresses at the end of the parse.

As we have indicated, we cannot yet draw any firm conclusions for adopting or rejecting one of the two proposals, although it seems at least, as if catalexis theory is faced with more problematic systems than Hayes' theory. In any case, the sample of queries quickly makes available some interesting cases to (re)consider.

Hopefully, the discussion in this appendix has made clear that using StressTyp in searches, always implies a careful consideration of how it is designed. For some queries, the results will be straightforward, because the contents and use of a term used in a record field is identical to its use in the theoretical literature. In other cases, we will have to translate terms into the format of the database, as was illustrated in this section by the discussion on degenerate feet. We must be aware that the use of a term can be different from that in the theoretical literature. Besides, the results of a query must never be taken for granted. In storing systems, we have made choices how to incorporate the information. It is possible that there are more ways to treat the same information, which would give a different result for the query. StressTyp is a tool and it can never take over the job of the linguist.

Chapter 3
The Linguistic Structure c
StressTyp: Stress system

Ellis Visch

3.0 Introduction

The basic concept underlying StressTyp is the separation of main stre and rhythm. This separation of metrical structure and rhythmical structu is not commonly assumed. StressTyp is thereby based on a theoretic perspective that competes with other metrical theories (cf. also chapter 1

In standard metrical theory (cf. Halle & Vergnaud 1987, Hayes 198 1995, Prince 1983 among others), it is commonly assumed that footin, i.e. rhythm precedes the assignment of main stress. In other words, fe are assigned starting at one side of the word and one of these assigne feet is promoted to main stress by a so-called End Rule. As has bee emphasized by Van der Hulst (1984; 1992; forthc. and chapter 1) ar Hammond (1984), it is striking that in most languages main stress located at the same edge of the word as where footing starts, i.e. the fir foot that is assigned carries main stress. Therefore, main stress does n necessarily depend on exhaustive footing. This fact is neglected in mo variants of metrical theory. Van der Hulst (1984; 1992) takes th observation seriously by making it the cornerstone of a 'main stress first theory. Main stress and non-primary stresses are assigned by separa algorithms: main stress is assigned first, followed by footing. In mo cases, rhythm moves away from the main stress foot ('echo rhythm'), bu in a minority of cases, rhythm can move towards it from the opposi edge of the word ('polar rhythm').

In this theory, main stress is assigned to a special syllable (heavy c marked), or, to a default syllable in absence of a special syllable in domain. In bounded systems, the stress domain comprises exactly tw syllables, located at one of the word edges. In unbounded systems, th stress domain comprises the entire word. In the stress domain, a gri mark is assigned or projected to all special syllables. Main stress i assigned by an ordinary End Rule which picks out the leftmost/rightmo syllable, respecting the presence of grid marks in the stress domain. A an illustration of this algorithm, consider the following two cases (cf. Va der Hulst forthc., chapter 1):