StressTyp: A database for word accentual patterns in the world's languages

Rob Goedemans and Harry van der Hulst

Introduction

It is possible, although inadvisable to discuss the structure of a linguistic database without saying a few things about the nature and linguistic analyses of the data that the database aims to store and query. In cognitive science terms, this would be like jumping ahead to the implementational level, without taking note of the computational and algorithmic levels (however, one delimits these levels in detail). Here, we take the computational level to involve specifying the nature of the data, and the algorithmic level to refer to the way in which linguists have generally captured regularities in the data. Even though the goal of the present volume is to focus on database structure and use (the implementational level), we supply a discussion of the nature and linguistic analyses of stress in section 1, hoping not to make it a barrier to the discussion of the database.

In 1991, we started working on a database for word stress systems, and it is hard to believe that we have been working on this project, off and on for 15 years now. We called the database StressTyp, but if we had had the perspective we have now we would probably have called it AccentTyp because stress is just one manifestation of the broader phenomenon of accent (see section 1). Since the database is mostly designed to store information about the location of the accent that lies behind stress, it would appear that the focus of StressTyp is, in fact, on accent location. Over the years, StressTyp has developed into a full-fledged typological database that currently contains information on the word accentual systems of 510 languages.

In this chapter we will describe the theoretical underpinnings of StressTyp (section 1), the history and current status of StressTyp (section 2.1), the goals of StressTyp (section 2.2), the limitations (section 2.3), dissemination (section 2.4), future developments (section 2.5), the architecture of StressTyp (section 3) and what one can do with StressTyp (section 4). The prefinal section (section 5) is devoted a comparison of StressTyp with other data collections or databases that store information on word accentual systems. We also provide an appendix with StressTyp fields and codes that

may be useful for reference while reading this chapter (other reference material, like a list of languages contained in StressTyp is available on the web at http://stresstyp.leidenuniv.nl). Section 6 concludes this chapter.

1. The nature and linguistic analyses of our data

A considerable number of languages (including English) display a phenomenon known as **word stress**. Word stress is one manifestation of a general characteristic of human languages, which is that linguistic expressions appear to have a 'prominence structure'. Linguistic prominence can be studied with reference to various domains such as 'words', 'phrases' or complete 'sentences' and even though there is no undisputed clear-cut definition of these domains, it does seem clear that the prominence patterns are neither universal nor randomly diverse. In other words, even though there are differences among the languages of the world, there are, at the same time, recurrent patterns. These patterns are typically (but apparently not always) grounded in general principles of rhythm, according to which 'beats' are spaced apart by a recurrent small number of non-beats, as well as being grammaticalized and lexicalized, by being put to use, among others, as markers of morphological and syntactic structure, in particular, but not exclusively, structural **edges**.

Most prominence patterns, then, have at least two general characteristics. Firstly, there tends to be the above-mentioned regular alternation of beats and non-beats; this is their rhythmical aspect. Secondly, as markers of domains, there must be one unit in these domains that stands out over all others. This unit marks the domain in opposition to other, adjacent domains, and ideally one of its edges (either the left edge or the right edge) by being either at the edge or close to it. Domain and edge marking as well as rhythmical alternation, are two entirely different sides of the prominence pattern, the first elevating one unit within the domain to a unique status, the other causing a regular strong – weak alternation among the units within the domain.

Rhythm and domain marking are independent, and we will see that neither is crucially dependent on the other. However, rhythm and domain/edge marking, when co-occurring, are interrelated in that it so happens that the unique unit, which we will call the **head** of the domain, at least typically, is a rhythmically strong unit. Classical metrical theory (Liberman & Prince 1977) was founded on making this connection by seeing rhythm as necessarily feeding head-location. A further unification of rhythm and domain

marking was then achieved by construing rhythm itself as the marking of domains called feet (prototypically consisting of two syllables or morae), so that the strong beat of the foot effectively became the head of the foot, while the head of the word would be the head of one of the feet within the word. Thus, **head-marking** became the unifying device for constructing (or analyzing) linguistic prominence patterns. As such, metrical theory embodied a theory of phonological dependency which, in fact, had already independently been established in Dependency Phonology (Anderson and Jones 1974, 1977; Anderson and Ewen 1987).

In the linguistic literature on word stress, a distinction is commonly made between primary stress and non-primary stress. Primary stress corresponds to head marking at the word level, whereas non-primary stresses refer to the rhythmic alternation. The latter is sometimes further divided into secondary, tertiary etc. stress and, at this point in our story, it is not clear how such finer distinctions can result from 'pure' rhythm, which has so far been depicted as a 'flat' regular alternation of 'strong' and 'weak'. In addition to the term 'stress', we also find the term 'accent'. Why this difference? When we say that one unit (let us say, a syllable) within the word stands out as the head, we say nothing about the manner in which it stands out. Using the pre-theoretical term 'prominence' also implies little, if anything in this respect. It turns out that heads can be manifested in a variety of ways. Hyman (1977) made a distinction between stress-accent languages and pitch-accent languages. The generalizing notion for him was **accent**, which we will take to be an alternative for the term head in the formal characterization of prominence patterns. (We also speak of heads of syntactic phrases or morphological complex words, where the term accent is not used. The notion head, in our view, is relevant in all components of the grammar, which sometimes goes unnoticed precisely because people use different terms for it.) In Hyman's view accents do not have an inherent manifestation. In a *pitch*-accent language, the accent is cued by a pitch property (an elevated pitch or a pitch rise, typically). In a stress-accent language, the manifestation is 'stress' which he took to be the kind of properties that are typically associated with 'stress' in languages such as English (extra duration, extra loudness, hyper-articulation etc.). However, there is no reason (and Hyman would, we are sure, agree) to limit the manifestation possibilities of accent to these two cases. Accent could be manifested by duration alone (a duration-accent language), or by full vowel quality (stressless vowels being reduced), etc. In addition, the head may distinguish itself from non-heads by a greater array of phonotactic possibilities, or by being the locus of tonal distinctions, or by being the anchor point for intonational

tones.[1] It would now appear that the study of 'prominence' (with a Eurocentric focus on stress-accent languages) must merely be seen as one way of getting to the deeper notion of accent (i.e. headedness).

In analyzing accents, the two dimensions of relevance are the specification of the domain of the accent and the determination of the location of the accent within the domain. Subsequent, or parallel to that inquiry we can ask how the accent is manifested within the domain, i.e. what cues are available to determining the location of the accent such that the accent can function as the marker of the domain (edges). Manifestations can broadly be grouped into phonetic cues (pitch, duration, loudness, fortition, hyperarticulation, and/or their reverse in unaccented syllables), phonological cues (phonotactic complexity, including both segmental and tonal distributional patterns), or any role in other kinds of regularities, be they phonological, morphological or pertaining to intonation. A comprehensive typology of accent manifestation remains to be developed, but given the broad area of cues and functions it is likely that many more languages may have word accent than just those in which accent is manifested as 'pitch' or 'stress'. As a working hypothesis, we might assume that all languages have accent. Pulgram (1970) has argued that marking of the *word* domain is not perhaps a universal fact, referring among other to the notorious case of French in which he argued that only a phrase final accent can be *observed* in the form of cues such as extra loudness and being an anchor for intonational tones. However, if we acknowledge a broader array of accentual manifestations such claims might well turn out to be misdirected.

An argument for the universality of word accent could follow from recognizing word headedness not only as serving the parsing of linguistic expressions into domains, but also as serving the mental storage of words. If words are not stored as linear strings of syllables, but rather as hierarchically structured objects (with perhaps no linear order as such), it might be that the nature of this hierarchy already involves the notion of head. If this is so, the ultimate motivation of heads would not lie in parsing, but in necessary properties of mental representations. This being the case, we could still argue that the edge-biased location of word accent is grounded in their

[1] The literature on intonational units refers to tones or tone combinations that anchor to word accents as *pitch accents*. This notion of pitch accent is different from the notion of pitch accent as one type of word accentual system, but the two are clearly related. In both cases pitch units are linked to heads of domain. Intonational pitch accents are tone units that link to phrasal heads (which, lower down, are also word heads) while word level pitch accents link to word heads.

role in being parsing cues. Thus, we are making a distinction between motivating the very existence of word head and motivating the location of the head. The fact that heads exist in other grammatical (and most likely linguistic external) domains where there is no linkage to parsing cues suggests a deeper motivation of heads. Since heads of phrases are not necessarily peripheral to their domain, edge-bias may not be intrinsic to the notion head, but follow, in the area of phonology, from their role in determining parsing cues.

We have revealed thus far that the location of accent within the 'word' domain may be dependent on rhythmic alternation (specifically seeking out strong syllables) and the domain itself (specifically seeking out its edges); cf. (1a). The preference for rhythmically strong syllables can be seen as a specific instance of the tendency for head location to be parasitic on a differentiation between syllables that is independently present. When a syllable is rhythmically strong it stands out (along with the other strong syllables) by virtue of the *externally* imposed rhythm. However, syllables are also differentiated from each other in terms of their internal properties. Thus a syllable containing a long vowel, or high tone or a fully articulated, non-reduced vowel may stand out in comparison to syllables that lack such properties. Syllable-intrinsic 'weight' may thus determine location of word accent (1b), but it can also determine the distribution of rhythmically strong syllables ('foot accents') which in turn determine the location of word accents (1c). If neither rhythm nor syllable weight feeds the location of accent, only domain edges provide a guide to its location (1d):[2]

(1) a. Rhythm-based system: Rhythm ⇒ Edge ⇒ Word Accent
b. Weight-based system: Weight ⇒ Edge ⇒ Word Accent
c. Weight and rhythm-based system:
Weight ⇒ Rhythm ⇒ Edge ⇒ Word Accent
d. Minimal system: Edge ⇒ Word Accent

We are assuming, then, that the notion of word accent is independent, in principle, from both weight and rhythm, *its location* being primarily motivated by being a parsing cue (although *its existence* may lie in the nature of mental representations) for which reason its location is always dependent

[2] The logical possibility Rhythm ⇒ Weight Edge ⇒ Word Accent exists, but in this case weight is a phonetic exponent of rhythm, for example when rhythmically strong vowels have a longer duration than weak vowels.

on edges. However, weight and/or rhythm may play a role in determining the location of word accent. Note that systems of type b and d show that rhythm is not a crucial aspect of all accentual systems, while domain marking (headedness) *is*.

The external and internal differentiation between syllables can exist independently from head location such that head location can be parasitic on these properties. However, as we showed earlier, when speaking about the manifestation of accent, such properties can also be the result of accent location, *a reversal of dependency*, so to speak (see footnote 2).

Focusing on systems in which rhythm seems relevant to the location of word accent, Metrical Theory made the entirely reasonable move to design layered algorithms in which, for type (1a) languages, firstly words are 'parsed' into left- or right-headed feet (a choice that was thought to differentiate languages) after which a second rule picks out the head of the rightmost or leftmost foot to be the head of the word; see (2a). (This second step was initially formalized as a tree-building procedure, constructing a left-branching or right-branching tree taking the feet as terminal elements. This idea was later abandoned by a device that simply elects a peripheral foot (head) as the head of the string. Type (1b) languages would not have rhythmic feet. Heavy syllables would stand out and the rightmost or leftmost of these syllables would be promoted to primary word accent; see (2b) Initially, heavy syllables were thought to be heads of 'unbounded' feet (unbounded meaning the foot domain is often larger than the prototypical two syllables, maximally comprising the whole word), an idea that some gave up and others maintained. In type (1c) languages, the parsing into foot domains was made dependent on a procedure that would designate heavy syllables as necessary foot heads. The location of these heavy syllable heads would then take priority over the default (left- or right-oriented) procedure of locating the heads of feet by imposing a constraint that bars heavy syllables from a weak position in the foot (hence they will always be heads); see (2c). Finally, type (1d) languages in which accent could be located purely with reference to an edge were in practice also seen as having a foot layer (see 2d-ii). The fact that no rhythm could be 'perceived' would be consistent with the idea that heads of feet can, but need not have an audible phonetic cue:

(2) a. Rhythm-based system: Rhythm ⇒ Edge ⇒ Word Accent

(i)
```
          *
  *   *   *   *
((σσ)(σσ)(σσ)(σσ))
```

- Feet can be left or right-headed, and the primary accent can be left or right-oriented.)
- Primary accent can even be located on the third syllable from the edge if a peripheral is marked as extrametrical:[3]

(ii)
```
        *
  *  *  *  *
((σσ)(σσ)(σσ)(σσ) <σ>)
```

- Extrametricality creates ambiguity in that, for example, a system with penultimate stress can be derived with left-headed feet as in (i) or right-headed feet plus extrametricality.
- To differentiate between the various non-primary accents, additional structure would have to be postulated, as was the case in the original versions of Metrical Phonology.

b. Weight-based system: Weight ⇒ Edge ⇒ Word Accent

```
    *
*   *
(σσσσσσσ)
```
(a bold sigma indicates a heavy syllable)

- Systems of this sort need a default rule for words that lack heavy syllables. This default appears to be independent in its edge orientation from the rule that promotes a peripheral heavy syllable. For example, a system that promotes the rightmost heavy syllable can have the initial or final syllable as its default location.

c. Weight and rhythm-based system:
Weight ⇒ Rhythm ⇒ Edge ⇒ Word Accent

```
      *
*  *  *  *
((σσ)σ(σ)(σσ)σ(σ))
```

- The same remarks as for (2a) apply here. Note that heavy syllables must be heads of feet, a factor that disturbs the rhythm which may trigger de-accenting rules to avoid accent clashes.

d. Minimal system: Edge ⇒ Word Accent

(i)
```
     *
(σσσσσσ)
```

[3] The device of extrametricality stipulates that a peripheral syllable can be ignored.

- Even penultimate accent can be derived if it is allowed to make a peripheral syllable 'extrametrical')
- As mentioned, this kind of system *can* be derived via an inaudible foot layer:

(ii)
```
          *
    *  *  *  *
((σσ)(σσ)(σσ)(σσ))
```

Variants and notational issues aside (see van der Hulst 1999), all these systems have a predictable location of the word accent. There are, as it turns out, also languages in which the accent location must be lexically specified. In practice, these systems can all be analyzed as involving 'diacritic weight': we simply mark the syllable that unpredictably has primary accent as heavy and then apply the above weight-sensitive algorithms (cf. 2b,c) to derive different types of lexically marked systems.

Systems of type (1b, 2b) have been called *unbounded* because the location of primary accent can be anywhere in the word. This is in sharp contrast with all other systems, called *bounded*, in which the primary accent is (a) strictly peripheral (final, initial), (b) near-peripheral (post-initial, prefinal) or (c) 'third-in' (critically due to extrametricality).

The metrical approach is committed to the idea that primary accent is always dependent on prior foot assignment. (Liberman and Prince 1977; Vergnaud and Halle 1978; Idsardi 1992; Halle and Idsardi 1994; Kager 1993; for surveys see van der Hulst 1999, 2000a, 2000b, 2002, 2006). Given that one needs to account for rhythmic structure, the location of primary accent almost comes for free. Van der Hulst (1984) first noted that the location of primary accent (putting unbounded systems aside) does not always follow from the principles that determine the rhythmic structure of the word. (see also van der Hulst 1990, 1992, 2002, 2006, to appear; van der Hulst and Kooij 1994; van der Hulst and Lahiri 1988; for similar views see Harms 1981; Roca 1986; Hurch 1995; McGarrity 2003). The location of primary accent and the distribution of rhythmic beats can display subtle differences. For example, one can be weight-sensitive while the other is not. Also, while primary accent can be lexically determined, rhythmic beats never are. These and other reasons suggested that even though primary accent has rhythm-like distributions, it would seem that the rhythmic grounding of primary accent has been grammaticalized by becoming a separate algorithm that thus can be divorced from the rhythmic principles that continue to account for the overall rhythmic patterns of words. Being grammaticalized, the algorithm can become sensitive to purely lexical factors such as diacritic weight,

dependence on word class and stratal layers in the lexicon, which suggests that the algorithm for primary word accent is a lexical procedure. Rhythmic structure on the other hand has all the properties of post-lexical or rather implementational processes. In the emerging view the dependency between rhythm and primary accent is reversed in comparison to metrical theory. Rhythmic structure is now dependent on the prior location of primary accent in the sense that this primary accent location must be properly integrated into the rhythmic structure which must be built around it.

Why would primary accent location and not rhythm tend to grammaticalize and lexicalize? We suspect that this may be related to the above-mentioned idea that mental representations of words need a headed structure of some sort.

The idea of separating primary and non-primary accent structure leads to the following approach. Some variant of the part of metrical theory that distributes rhythmic beats can be maintained as a procedure that applies to actual utterances. For primary accent location, van der Hulst (1996, 1999, to appear) suggests the following approach. For all systems in which accent is bound (a, c and d in 1 and 2) we say that a bisyllabic domain is selected at the left or right edge of the word. However, to accommodate unbounded systems (b in 1 and 2), we also allow the option that the domain for accentuation is the whole word. With respect to both options, extrametricality can apply. If the system is weight-insensitive (cf. a in 1 and 2), this produces only one case, (3a). We only need to say whether the left or right edge is selected for primary accent. If the system is weight-sensitive, the heavy syllables (phonologically or diacritically) stand out. If the domain contains only one heavy syllable, this will be the only syllable that is available for primary accent selection. If there is more than one heavy syllable, we need to say which one wins. If there is none, we need a default rule (cf. Prince 1983). These procedures apply in (3b) where the domain is the whole word and in (3c) where the domain is a two-syllable window. Finally, we can select the whole word and not have weight as a relevant factor (see 3d):

(3) a. Rhythm-based system: Rhythm ⇒ Edge ⇒ Word Accent

$$\overset{*}{(\sigma\sigma\sigma(\sigma\sigma))}$$

b. Weight-based system: Weight ⇒ Edge ⇒ Word Accent

(i) $\overset{*}{(\sigma\sigma\sigma\sigma\sigma\sigma)}$

(ii)
```
    *
  *   *
(σσσσσσσσ)
```

(iii)
```
  *
(σσσσσσσσ)
```

c. Weight and rhythm-based system:
Weight ⇒ Rhythm ⇒ Edge ⇒ Word Accent

(i)
```
    *
    *
(σσσ(σσ))
```

(ii)
```
    *
    *
(σσσ(σσ))
```

(iii)
```
    *
  * *
(σσσ(σσ))
```

(iv)
```
      *
(σσσ(σσ))
```

- This is a system in which primary accent lies on the final syllable if it is heavy, otherwise the pre-final syllable is accented.
- The separate edge orientation for the heavy – heavy and light – light case predicts four types of bounded weight-sensitive systems which are attested, both on the right side and the left side of the word.

d. Minimal system: Edge ⇒ Word Accent

(i)
```
  *
(σσσσσσ)
```

(ii)
```
      *
(σσσσ(σσ))
```

We need to add extrametricality to the mix to derive third-in systems, as well as all bounded weight-sensitive systems in which a peripheral heavy syllable is ignored (such as Classical Latin). This creates several instances of structural ambiguity that, apparently, cannot be avoided. In particular, systems of type a and d can be difficult to differentiate. For example, a weight-insensitive penultimate system can be of type a (locating the head on the left in a right-edge bisyllabic window) or of type d (locating the domain on the right in a unbounded window subject to extrametricality). For pen-peripheral systems (second syllable accent, or penultimate accent), this ambiguity is caused by having the option of extrametricality. However, in peripheral systems (initial or final accent), the ambiguity exists regardless (cf. 3d-i and 3d-ii). Often, the exceptional locations of accents will reveal the nature of the system. Turkish, which has regular final accent, allows exceptional locations of accents far inside the word. This system is thus unbounded. Polish, on the other hand, having regular penultimate accent,

only allows exceptions on the final or antepenultimate syllable. This system is therefore bounded.

In this approach, primary accent location can be analyzed in terms of seven parameters, two of which (4a-i and 4b-i) are dependent on the setting of another parameter:

(4) a. Domain size: bounded/unbounded
   (i) Edge of bounded domain: left/right
   b. Extrametricality: yes/no
   (i) Edge of extrametricality: left/right
   c. Project weight: yes/no[4]
   d. If two (or more) heavies: leftmost/rightmost
   e. If no heavies: leftmost/rightmost

One type of system remains to be accounted for. This is a system (termed a 'count system') in which the location of primary accent is apparently *necessarily* dependent on the prior exhaustive rhythmification of the entire word. Consider the following primary accent rule:

(5) a. In a word with an even number of syllables, primary accent is prefinal
   b. In a word with an odd number of syllables, primary accent is preprefinal

It would seem that we have to establish a left-to-right, left-headed rhythmic pattern and then select the rightmost beat as the primary accent, (6a). At first sight, it might, however, also be possible to assume a right-edge bounded domain and extrametricality and the projection of 'rhythmic' weight, (6b)

(6) a. (i)
```
          *
  *   *   *   *
((σσ)(σσ)(σσ)(σσ))
```
(ii)
```
  *
  *   *   *   *
((σσ)(σσ)(σσ)(σσ)σ)
```
b. (i)
```
  *   *   *   *
(σσσσσ(σσ)<σ>)
```
(ii)
```
  *   *   *   *
(σσσσσσ(σσ)<σ>)
```

[4] Note that the weight parameter could be replaced by two constraints that can enter into a dependency relation ('ranking'): if weight is on: weight > rhythm; if weight if off: rhythm > weight. This is possible, as any parameter can be replaced by two constraints. It is not clear that anything is gained by this alternative.

The oddity of the alternative in (6b) (in which we have suppressed the foot boundaries for clarity) is that rhythm, which we so far attributed to the utterance level, must feed the lexical procedure for primary accent assignment. What is the solution to this paradox? One solution is to assume that the primary accent algorithm can be post-lexical as well as lexical. Being post-lexical, the procedure can be sensitive to rhythmic weight, or not. If it is, we get the 'count system' type. Allowing the primary accent procedure to be post-lexical creates additional ambiguity, however. It essentially allows a standard metrical ('rhythm first') treatment of all systems (not just count systems) in which primary accent can harmlessly be said to fully depend on independently needed principles for rhythmic structure. This would leave the cases in which primary accent cannot be derived from rhythmic feet as oddities that require some special procedure, a route taken in Hayes (1995). It could be that this is just how matters are. Some systems simply would be open to both a fully post-lexical analysis (in which case we expect no lexical exceptions at all) and a lexical analysis (in which lexical exceptions are possible).

In an attempt to reduce ambiguity, van der Hulst (1997) explores the position which holds that primary accent location can only be lexical (which makes sense if it is a lexical requirement for reasons of necessarily needing a head in the mental representation of words). To maintain this position he claims that all count systems are systems in which the primary accent lacks any overt manifestation and/or is in diachronic transition. An additional speculation was that polysynthetic languages, which rank high among the count systems, simply lack a lexical notion of word altogether. It was furthermore suggested that the apparent rhythm-based primary accent was an enhancing effect that results from phrasal accentuation and/or intonation. These ideas call for considerable empirical confirmation and must remain, at this point, theoretically-driven speculations.

A different approach is to forget about a lexical – postlexical divide and derive the different types of systems in terms of different dependency relations between the constraints that govern rhythm and primary accent:

(7) Rhythm ⇒ Primary accent
Primary accent ⇒ Rhythm

This is the approach taken in Optimality Theory (Prince and Smolensky 1993) where 'dependency' is called 'ranking'. This approach, however, fails to explain why rhythm is never lexically determined, while primary accent is, either mostly, or in the form of exceptions (that are almost always

present). We must leave the definite solution for count systems for future research.

This concludes our theoretical preamble concerning the nature of word accentual systems and available theories in this area. In the development of our database on accentual systems, we were inevitably inspired by the theoretical considerations presented in this section, which are based on considerable study of both accentual systems and available theories. However, we did make an effort to design the record structure in as theoretically neutral a manner as possible. The resulting record structure is described in section 2.4. Before we look at that structure, however, let us first present the development of StressTyp from its inception to the present day in which it has become the leading typological database on stress systems.

2. StressTyp – an overview

2.1. History and current status of StressTyp

Work on StressTyp was initiated by van der Hulst in 1991 as a pilot project of EUROTYP (1990–1994), a project on the typology of European languages, financed by the European Science Foundation (ESF). EUROTYP consisted of 9 Theme Groups, each studying an aspect of European languages from a comparative and typological point of view. The topic of Theme Group 9 (coordinated by van der Hulst) was Word Prosodic Systems.[5] In the course of the EUROTYP project the question regarding storing language data (both original and from written sources) received special attention and in 1991 it was decided to start two pilot projects, one of which was StressTyp. The idea was to develop an intelligent filing system for data (i.e. rules, generalizations, patterns) on word prosodic systems.

The structure of the records was developed by Harry van der Hulst (then at HIL, Leiden), in collaboration with Aditi Lahiri (then at the Max Planck Institute, Nijmegen). Some relevant equipment was made available by a grant from the EUROTYP project and further support of the Faculty of Arts of Leiden University. Kees van der Veer (Max Planck Institute, Nijmegen) implemented the record structure in 4th Dimension for MacIntosh. Since then, Rob Goedemans has controlled all aspects of the implementation side of the database.

[5] The results of this EUROTYP project have been published in van der Hulst (1999) (ed.).

The first data for StressTyp were extracted from typological studies, or theoretical works that refer to a lot of languages, such as Hyman (1977), Greenberg and Kashube (1976), Hayes (1980/95), Lockwood (1983), Halle and Vergnaud (1987) and so on. Additional data came from the Masters theses of Aglaia Cornelisse (Australian languages) and Bernadette Hendriks (Papuan languages), both supervised by van der Hulst. These data were first combined in so-called Data Entry Sheets (basically a paper-and-pencil version of the record structure) and subsequently Rob Goedemans and Ellis Visch transferred the data into the 4th Dimension database. In this process they checked the information for consistency and correctness by going back to the original sources, and often also to additional theoretical or descriptive studies. At the end of this phase, StressTyp contained 154 languages.

After the Eurotyp phase, work on StressTyp was continued by Ellis Visch, Ruben van de Vijver and Rob Goedemans. Other people who have contributed their time in this early phase were Simone Langeweg, Bernadette Hendriks and Paulus-Jan Kieviet.[6] The combined efforts of these people resulted in more complete coverage of the accentual systems of the individual languages, thoroughly checked records, and the addition of accentual information for 116 new languages, bringing the total to 270.

From 1997–2001, StressTyp was included in the Prosody of Indonesian Languages (PIL) project coordinated by Vincent van Heuven (Leiden University), during which time the database implementation was improved and the number of languages went up from 270 to 510. Goedemans checked the content of the old records for errors, checked the primary sources of languages for which the entry was only based on remarks in secondary sources, added examples where these were missing, and updated the language names and affiliations according to the SIL Ethnologue 13th edition standard (Grimes 1996). At this point, only a handful of records for languages in StressTyp are based on secondary sources only.

During the PIL project we were approached by the editors of the World Atlas of Language Structures (WALS), a cooperative effort of the Max Planck institute for Evolutionary Anthropology in Leipzig and linguists with typological databases from all over the world. Specifically, we were invited to produce a number of maps that would show the distribution of

---

[6] Some of these people worked on StressTyp in the context of other projects that were funded by the Netherlands Foundation of Scientific Research, NWO, the Holland Institute for Generative Linguistics (HIL), the Department of General Linguistics of Leiden University and the Department of General Linguistics of the Free University of Amsterdam.

various kinds or aspects of word accentual systems. We produced four such maps (see Goedemans and van der Hulst 2005a–d). StressTyp has benefited greatly from the cooperation with the WALS editors. The WALS project compiled a representative list of the world's languages and we ensured that all languages in this list were present in our database. To this end we used the list of descriptive sources that the WALS editors provided and added a significant number of the 240 languages with which StressTyp grew during the PIL phase. Finally, StressTyp was expanded with 2 fields for geographical location and a procedure was developed to draw distribution maps of StressTyp data with the help of the mapping programme AGIS.

StressTyp is now also included in the Typological Database System (TDS), a joint venture of the Universities of Amsterdam, Leiden, Nijmegen, and Utrecht, which aims at development of a common query interface for several typological databases. A prototype of the system is up and running (http://languagelink.let.uu.nl/tds see Dimitriadis et al., this volume).[7] In the first phase of the TDS project, Rob Goedemans ported StressTyp to an MS-Access implementation that follows the original database design, so that we can now serve a much bigger user group. To facilitate a smooth integration in the TDS, examples in IPA were converted to Unicode and the Ethnologue codes were updated to the 15th Edition (Gordon 2005).

In section 2.4. we will report additional information on past and current activities involving StressTyp. First we will say a few words about the goals and limitations of our database project.

2.2. Goals

One of the main goals 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 first-hand descriptions of language data, including examples, generalizations and the like. By secondary sources we refer to theoretical works on stress which themselves draw on such primary sources. Critically, by using the word 'primary sources', we do not imply that the data stored in StressTyp are collected first hand from, or, checked with native speakers by us. Nothing in the idea behind StressTyp would preclude collecting and storing first hand data, but we have simply not had the means to do this.

[7] The TDS also contains SyllTyp, another database designed by Harry van der Hulst and Rob Goedemans.

There was no intent to include only a representative sample of the languages of the world (but see below). We recorded information for whatever language we could find accentual information for. We have included all languages for which clear statements on accent location were present in the sources. The record structure allows for much more (see section 2.4), but additional information was only added if it was readily available in the source, in the hope that the record could be made more complete later on (as was often the case, although all records are still incomplete).

As a matter of course, one of the goals of StressTyp is typological in nature. A sufficiently rich database allows for quantitative research and checking of implicational relationships. We can use StressTyp to expose common and uncommon traits of stress patterns, to check the validity of certain claims made in theoretical works and to discover new dependencies between various stress (and perhaps even other) parameters.

2.3. Limitations

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. Every record, of course, specifies the sources on which we have based the coding.

Specifying values in database fields necessitates interpreting sometimes very limited information. Although we do not wish to criticise the hard and important work that has been done to obtain first-hand descriptions of languages, and without which an enterprise like StressTyp would not be possible, we do note that the data and generalizations that are provided are often insufficiently precise to conclusively determine the exact nature of the accentual system. This is not surprising given the amazing variety of accentual systems that often differ in very subtle details pertaining to factors that determine syllable weight, rhythm, word length and so on (even ignoring the role of morphology, where this appears to be relevant). 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 for information on syllable struc-

ture and morphological structure as well. The former is often present in an elementary form, the latter is mostly absent.

Misinterpretations on our part are also possible. The coding system requires interpretation of the sources. In addition, our records cannot always be 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.

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 (see section 3). Besides, collecting information on as many languages as possible is simply the only manner to proceed if one wishes to develop general theories. In the 'old days', every student of stress would keep records like this in note books, or on filing cards. Clearly, with the availability of computers, those efforts are more likely to result in digital storage.

We emphasize that StressTyp cannot be held responsible for providing incorrect, or incomplete information. We always encourage those who use StressTyp in publications to not only acknowledge the use of our database, but also 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. Dissemination

To promote the use of StressTyp we have published a collection of articles in 1996, of which some are based on StressTyp information, while others describe the database structure and ways to go to from descriptions in sources to the StressTyp coding. In addition, this volume describes some direct numerical results and examples of queries.[8] A second volume based on StressTyp data is underway. In that book we present part of the data in several geographically oriented appendices, while chapters written by experts on those respective areas comment on generalizations and patterning, provide new insights into the phenomenon of accent, and supplement the

[8] Goedemans, van der Hulst and Visch (1996a).

StressTyp data with additional languages.[9] Also in 1996, in a short article in Glot International we offer basic information about StressTyp.[10]

In addition, we have promoted StressTyp on the web. On the StressTyp website we describe in which ways others can use the database, either directly on the web, or by obtaining a copy of the application. Several versions of the database are available:

- Legacy versions: a full 4th Dimension version which allows you to use all the standard facilities of the 4th Dimension database package. For PC and MacIntosh. Also available for users who do not own 4th Dimension, but with reduced functionality.
- Access version: All the original data, and a new user interface. Examples in Unicode.
- Online version (at http://stresstyp.leidenuniv.nl/).
- TDS version (at http://languagelink.let.uu.nl/tds/, incorporated in a larger system).

The legacy versions are no longer updated and we thus strongly advise users to obtain the free Access version.

The core fields from StressTyp are represented in the on-line version. One can use this version for relatively simple queries. However, for more advanced work it will be necessary to obtain the stand-alone (Access) application. The same core fields are represented in the TDS. You can query StressTyp fields in the TDS in combination with fields from other databases. Also, the TDS system will guide users who are not too familiar with accentual phonology with ample explanatory notes, links to related fields and the like.

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

---

[9] Goedemans, van der Hulst and van Zanten (to appear).

[10] Goedemans, van der Hulst and Visch (1996b).

2.5. Future developments

StressTyp started with very little funding, and has, since its inception, piggy-backed on other projects or on people's 'free' time. At the moment, the two authors of this chapter ensure the continuation of the StressTyp project. We continue to plan extending the content of the database by systematically trying to add information on language families or linguistic areas that are now underrepresented. To do this systematically and rigorously we need a grant that is exclusively dedicated to the development of StressTyp. Efforts to acquire such funding are currently under way.

Long ago, we also planned to extend the scope of the database by means of a questionnaire. This questionnaire was designed and is intended to be filled in by linguists who are familiar with a particular language. If appropriate means come our way, we will develop this questionnaire further and start distributing it. In addition, we would try to systematically solicit (references to) books or articles that contain useful information on stress systems, especially of languages that are not yet contained in the database and thus build an archive of primary and secondary sources (preferably in machine-readable form).

Our most recent attempt (mentioned above) has been to invite a number of phonologists to write a survey of stress systems in various parts of the world to be published in Goedemans, van der Hulst and van Zanten (to appear). These surveys will be used to add new information to StressTyp.

It is inevitable that others have developed databases on word accentual systems (perhaps with less detailed record structures) or that information on stress systems is part of less specialized databases. Such databases could be 'old fashioned' paper-and-pencil collections or actual digital collections. We would like to be made aware of such systems and, more importantly, of the availability of the information contained in them. We are interested in collaborating with others if such systems are still accessible. In section 5.3 we discuss two other databases that were constructed while StressTyp was already in existence. Such duplication is unfortunate and we plan to establish collaborations with these projects if they are still active.

It is possible to be more ambitious. Originally we aimed at embedding StressTyp in a network of related databases that would 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 global indicator for this imaginary network was StressEx (Stress Expert System). Meanwhile we have also developed a data-

base for phonotactic information (SylTyp), so a more general 'umbrella' would be Word Prosody Database. This could include a database initiative started by Larry Hyman (XTone: http://xtone.linguistics.berkeley.edu/) which collects information on word tonal patterns. Ultimately, we could then establish a database for Word Phonology if the work by Ian Maddieson on segmental inventories would be included.[11] Finally, one could imagine combining all these pattern-oriented databases (of which, we are sure, there are more around) with process-oriented databases, i.e. databases that collect information on phonological processes, such as NasDat (http://acvu.nl/staf/wlm.wetzels/pwp_en.htm), ATR/Vowel Harmony (an old, now dormant project of Maarten Mous, Norval Smith and Harry van der Hulst) and others. All this work is, of course, strongly reminiscent of the considerable efforts that were developed by Joseph Greenberg and his collaborators in the universals project which, for the most part, led to pencil-and-paper databases, or to digital collections that are now perhaps no longer accessible or useable. (If more general initiatives on a word phonology database could be developed, it would be advisable to try and incorporate this earlier work. Time permitting, the authors of this chapter will make an effort to develop such an initiative.)

3. The record structure

It is well known that the designer of a database faces a paradoxical problem (which we call 'The Database Paradox'). To make the perfect database one needs to know exactly what the extent, structure and nature of the data is, what one is going to be interested in and how the data can be best represented to serve that goal. However, there is a need for a database because all these things are not known in detail. The design of StressTyp, as described in this section, is therefore nothing more than an attempt to capture all the properties of stress (or rather: word accentual) systems in a set of parameters that we now consider to be complete, but which is "open" in the sense that future discoveries may force us to change parameters or add new ones. So far, extraordinary systems that we have encoded have occasionally forced us to expand the set of possible values for some of the fields, but to date we have never encountered a language whose stress system necessitated expansion of the set of fields. This is largely due to the separation of the encodings for primary and secondary accent, which allows us to encode

[11] His work, as reported in WALS, in fact also includes syllable structure.

even the most unyielding of the exotic stress patterns. We must emphasize that, although the parameters used in this database embody a particular view on word accentual structure (as developed in van der Hulst 1984, 1996, to appear), they are presented as purely 'descriptive'. It is entirely possible to interpret the information without making a commitment to any theory that assumes separation of information on primary accent and rhythmical structure (as the one that is described in section 1).

It is unavoidable that the translation of properties of certain systems (as described in the sources) into the format of the database is not always entirely straightforward, because these systems have special or sometimes even conflicting features. Most typically, as we stated in section 2.3., we find that descriptions in primary sources are insufficiently explicit, while examples given leave open different interpretations. As a result, there are often different possible ways of storing properties of systems. We have consistently stored similar ambiguous systems in a similar fashion. When more detailed information becomes available for such languages, reinterpretation may be necessary, as has proven to be the case for some languages.

As far as the "we cannot know what one is going to be interested in"-part of the paradox is concerned, we encoded every aspect of stress systems that we could think of separately and sometimes redundantly. The result is that single parameters of stress systems can be queried straightforwardly, but that combinations of parameters, which typically arise when one wants to find certain prototypical stress systems, sometimes result in complicated multiple queries. In our experience, we have always been able to query StressTyp on every aspect of stress that held our interest. Sometimes it was cumbersome, but never impossible. With the incorporation of StressTyp in the TDS, the complicated queries for prototypical systems have been incorporated as part of the knowledge-base. Many sets that needed multiple queries in StressTyp can be generated with a few clicks when one uses the TDS. Future wishes can, of course, not be foreseen, but we are confident that StressTyp can accommodate them as is, or with only a few minor modifications and/or additions (apart from expansion of StressTyp coverage to other fields, of course).

We will now present the record structure of StressTyp (of which the database structure is a straightforward 'flat' table with one record per language). Below, the fields are presented with their respective definitions. Where necessary, a more detailed explanation is given. The presentation exactly follows the order of the fields in the user interface of the StressTyp application. Here we present only the core fields of StressTyp, since only these are relevant for the discussion that follows. The rest of the fields can be found in Appendix B (without description). A much more detailed explanation of

all the fields can be found in the manual (http://stresstyp.leidenuniv.nl) and Goedemans, van der Hulst and Visch (1996c)

Language: The name of the language or dialect in the naming-conventions of SIL Ethnologue 15th ed. If the language is referred to by other names in the literature, these are added to prevent double records in the database.
Dialect of: The name of the mother language of which the language specified in the record field language is a dialect.
Genetic affiliation: The family tree, root first (as in SIL Ethnologue).
Region: All geographical areas in which the language, or dialect of the language in question, is spoken.
Latitude and Longitude: Exact geographical location of the (centre of) the area in which the language is spoken.
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 from a list of items, or a combination of abbreviations and one (or more) connective(s). The most common items include I(nitial), S(econd), T(hird), A(ntepenultimate), P(enultimate) and U(ltimate) for stress systems that place primary accent on a fixed syllable in every word of the language. Combinations may take the following shape: U/P for a language that places stress on the final (ultimate) syllable if it is heavy, else on the penult; I;S for a language that generally has fixed initial stress but has a significant group of exceptions that all have stress on the second syllable; F/L for a unbounded stress language that places stress on the first heavy syllable in the word, and on the last syllable if there are no heavies. More codes and examples can be found in appendix A and a more detailed explanation is given in the manual (http://stresstyp.leidenuniv.nl).
Quote: Description of the stress pattern in words, usually taken from the primary source.
Descriptive source: The primary source (in most cases) in which the description of the stress pattern was found.
Theoretical Source: Reference to authors that have analyzed the pattern in a certain theoretical framework.
Examples: Illustrate all aspects of the stress pattern (if available).
Stress Domain: Indicates whether stress is assigned at the Left or Right edge of the word in bounded systems, or whether it can be assigned anywhere in the word in Unbounded systems.
Extrametricality: Specifies whether any phonological unit is ignored in the selection of the domain at the Left or Right edge of the word, or whether Extrametricality does not play a role.

Extrametrical Unit: Specifies what unit is ignored, if any. Possible values: consonant, vowel, mora, syllable, heavy syllable and foot.
Weight: Does syllable weight play a role in the assignment of primary accent, Yes or No?
Stress if Both Heavy: If weight plays a role, heavy syllables (prototypically those with long vowels or codas) in the domain attract stress. If the bisyllabic domain contains only one heavy syllable, it is clear where the stress must go. If the domain contains two heavy syllables we need to specify what happens in this case. This field does exactly that. The options are, naturally, Right or Left.
Stress if Both Light: In languages that use weight this field describes what happens when both syllables in the domain are light. In languages that do not use weight this field describes what happens in all cases. The options are Trochaic (i.e. left headed) and Iambic (i.e. right headed). (see section 4 for motivation of the choice for trochaic and iambic instead of left and right).
Stress Repair: Yes/No field that indicates whether there is a shift outside the two-syllable stress window, if both syllables inside the window are light.
Degenerate feet: Incomplete feet (monosyllabic in languages that do not use weight, monomoraic in languages that do) can (Yes) or cannot (No) be used in the analysis. If they can be used, single syllables that are left over when the rest of the word is parsed into binary feet (bisyllabic or bimoraic) do get a secondary stress by virtue of this degenerate foot that may be used to parse the syllable. In languages that do not use such feet, these syllables remain unparsed, hence stressless.
Subminimal words: Words that are smaller than a foot exist (Yes) or are prohibited (No) in the language. In quantity-insensitive (QI) languages subminimal words are all monosyllabic words. In quantity-sensitive (QS) languages these are only monosyllabic words than consist of a light syllable.
Rhythm: The language employs a pattern of secondary stresses, Yes or No.
Starting edge: Specifies at which end of the word rhythmic patterning starts. Left, Right, Edge-in (i.e. patterning starts at both edges) or Centrifugal (i.e. Rhythm echoes away from the primary stress that is assigned somewhere in the middle of the word.
Extrametricality, Extrametrical Unit and **Weight**: see above.
Type: Specifies whether Trochaic, Iambic or both types of feet are used in the rhythmic pattern.
Repair: Yes/No field that indicates whether the rhythmic surface patterning may deviate from that specified by the fields.
Iterativity: Field that specifies whether secondary stress is assigned only once, at the starting edge (No), or as many times as possible given the num-

ber of syllables in the word that can be parsed into rhythmic feet (Yes).
Rhythm ternary: Specifies whether ternary (trisyllabic) feet with Trochaic or Iambic heads are used in the analysis of secondary stress. Default setting is No ternary rhythm.
Template: The full set of possible syllable types in CV notation.
Obligatory Onsets: All syllables in the language must have onsets (Boolean).
Branching Onsets: Onsets with more than one segment are allowed (Boolean).
Long Vowels: Long vowels occur in the language (Boolean).
Closed Syllables: Closed syllables occur in the language (Boolean).
Geminates: The language uses geminates (Boolean).
Heavy for stress: Specifies exactly which syllables count as heavy in the assignment of primary stress.
Heavy for rhythm: Specifies exactly which syllables count as heavy in the assignment of secondary stress.
Repair: In full text what happens and when, in languages for which either one of the repair fields above has the value Yes.
Remarks: Any remaining remarks about the stress pattern or its encoding.

4. What can one do with StressTyp?

The goal of the coding system has been to make it possible to search through the database for the occurrence of quite specific properties. With the search facilities of Access or the web interface, StressTyp can be instrumental in testing and developing hypotheses (given that the limatitions of StressTyp have been taken into account; cf. section 2.3).

Goedemans and van der Hulst (2005a–d), van Zanten and Goedemans (2007) and Goedemans (to appear) all use StressTyp to present data on stress patterns in various ways: primary data on the occurrence of the most common types of stress systems, exponents of syllable weight, secondary stress types in the languages of the world, the geographical distribution of languages that have certain specified characteristics. Moreover, these articles, especially Goedemans (to appear), contain examples of phonological claims that can be put to the test, using StressTyp to quantify exactly for what percentage of the world's languages supposedly universal claims hold true.

We will not repeat those exercises here, but rather present four new examples in the same fashion as the ones presented in the aforementioned ar-

ticles.[12] The most direct way in which we can present data from StressTyp is in simple graphs that show how many languages in the sample of 510 exhibit certain stress patterns. In the past, we have generally had a rather global outlook when we presented such data. An article in which one presents complete overviews of the types of stress patterns that occur in the world's languages is necessarily coarse, simply because there are too many possibilities. The sheer number of possible types has prevented us from revealing some of the finer grained distinctions in earlier publications. A subset of languages that has suffered from this contains the so-called *unbounded* languages. As explained above, unbounded languages are those in which primary accent can, in principle, be assigned anywhere in the word, no matter how long it is. Languages which are traditionally called unbounded are always quantity-sensitive (see section 1 for discussion of applying this notion to quantity-insensitive languages in so-called minimal systems, 3d), and the decision where to place main stress in the word is usually based on syllable weight (and in some cases diacritic weight or prominence due to pitch). This type comes in four basic flavours, depending on which of the heavy syllables in the word receives primary accent and what happens when there are no heavy syllables:

(8) F/F stress the first heavy syllable, and in case there are none, stress the first syllable

F/L stress the first heavy syllable, and in case there are none, stress the last syllable

L/F stress the last heavy syllable, and in case there are none, stress the first syllable

L/L stress the last heavy syllable, and in case there are none, stress the last syllable

To this we add languages in which stress is lexically marked, can occur anywhere in the word, but does not adhere to one of the four rules above (usually that means there is only one syllable in the word that is lexically marked). The category, containing languages like Kewa, is labelled 'Lex'. Languages like Russian, in which the stress rule is also sensitive to lexical marking, but which display patterns like 'stress the first lexically marked

[12] Readers who are interested in more examples, quantifying basic patterns and claims are referred to Goedemans and van der Hulst (2005a–d), van Zanten and Goedemans (to appear) and Goedemans (to appear).

syllable or else the first' are incorporated in the appropriate category (lexical marking being, in our view, just another instance of weight). A final category of unbounded languages groups unique, fairly exotic, stress patterns, usually variations on one of the four types listed above under the label 'Irr' (for Irregular).[13] In previous publications we have lumped the unbounded systems together in one group to be able to show all possible quantity-sensitive stress systems in one simple graph. We will now present the subsets in detail. When we consult StressTyp to get the actual numbers we arrive at the following result.

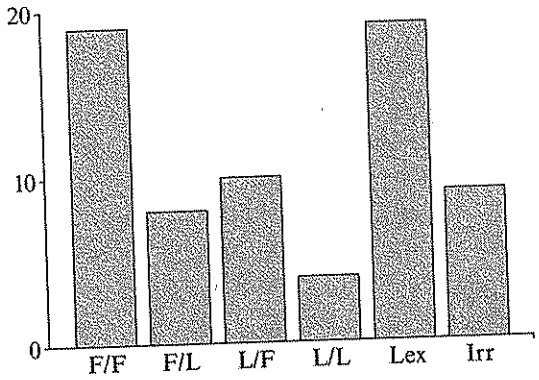

Figure 1. Number of languages for each of the six categories of unbounded systems.

Not counting the 'Lex' systems, for which it is anyone's guess what their preferred location for stress is, we observe that the general preference for bounded stress systems, namely to use left headed constituents (see Goedemans, to appear), is also reflected within the unbounded category. 27 of the 41 unbounded languages in the groups F/F, F/L, L/F and L/L place stress on the leftmost heavy syllable, and 29 of the same 41 languages place stress on the first syllable when no heavy syllable is present in the word.

A second usage of StressTyp is to look directly at the values of the parameters we have incorporated and draw a map showing the way these values are spread among the languages around the globe. Let us continue on the note set in above in our second exercise. We will look at bounded systems

[13] In earlier publications we have commented on the fact that we also consider the count systems to be of the unbounded variety. We leave them aside here, and present only those systems that are uncontroversially unbounded, whatever theory one adheres to.

and determine the default value for stress placement, i.e. the value that indicates the edge choice when the domain does not contain a heavy syllable, or when the language is not quantity-sensitive to begin with. In other words, we will determine at what edge of the domain primary accent ends up when no weight is in play and see if any interesting geographical clustering appears. As was noted above, the flavours we have are trochaic (head on the left) and iambic (head on the right). We can query StressTyp for systems that are iambic or trochaic in this respect and feed the results, together with the geographical coordinates of the languages in the result set, to a mapping program. When we do that we arrive at the following map.

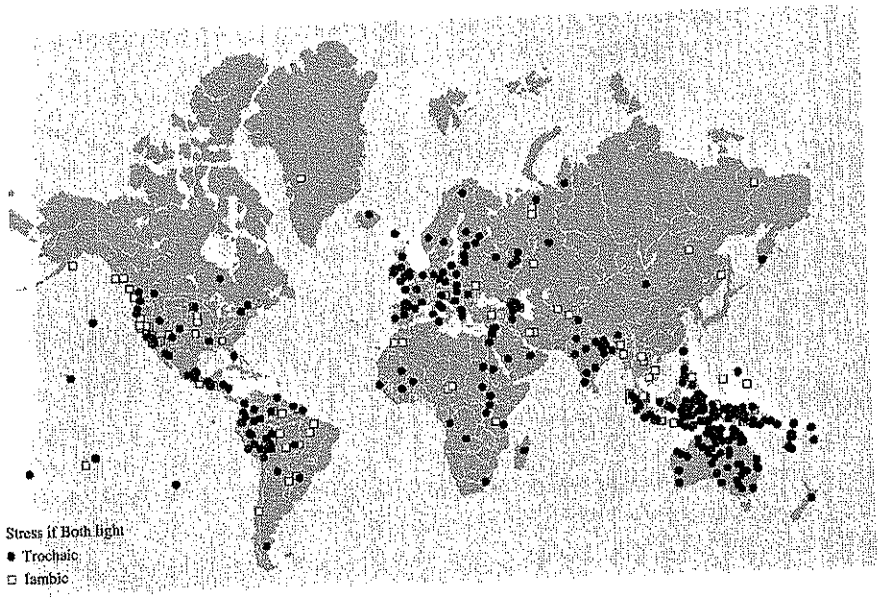

Figure 2. Geographical spread of languages that show Trochaic (black dots) or Iambic (white squares) primary accent patterning in the default case.

We have produced maps like the one in Figure 2 before (WALS maps 14–17, see Goedemans and van der Hulst 2005a–d) and on one of these (map 17) we presented the languages in StressTyp that show trochaic or iambic patterning for secondary accent. With respect to that map we noted the following tendencies:

(9) (i) Iambic rhythm occurs mostly in North and South America.
(ii) South America and Australia always seem to have clear rhythmic patterns.

(iii) Africa, on the other hand, shows little evidence for rhythmic patterns.

The map presented here shows that with respect to default edges for primary accent:

(10) (i) Left is the default edge for most languages, on a par with the preference for trochaic rhythm that most languages show.

(ii) In Australian, Austronesian, Indian and European languages the trochaic option is chosen in an overwhelming majority of languages.

When we compare the observations, we note that:

(11) (i) For lack of African languages in StressTyp that use rhythm, no preference for either iambic or trochaic feet could be detected on the rhythm map. The current map clearly shows that African languages, in keeping with the general trend, have a clear preference for the trochaic default for main stress.

(ii) Where iambic rhythm was largely confined to the Americas, the usage of the right edge of the domain as the default location for primary accent is more widespread.

We take the languages' preference for trochaic or iambic rhythm, as presented on the WALS map, to reflect *default* preference for either left or right headed constituents, an assumption we think is hard to contest. In StressTyp, we have made clear the affinity between rhythmic feet and the default location of primary stress through the values for the *Stress if Both Light* parameter, which are not Left and Right as one might expect but rather Trochaic and Iambic (see section 3). In most other theories the two parameters are even inseparable. In this light we can state some extended generalizations based on a combination of the data on these two maps.

(12) (i) Left-headed constituents are the ones most commonly used in word accentual systems (on average, about 81% of the languages use them either as rhythmic feet or as the default for primary accent).

(ii) Australian, Austronesian, Indian, African and European languages clearly prefer left-headed constituents.

(iii) North-American languages show no preference. In South-America the split is roughly between Andean, left-headed, non-Andean (perhaps Amazonian), right-headed preference.

A question that comes to mind immediately when we consider these maps is whether there are any mismatches. Are there languages that use trochaic feet in the assignment of secondary accent, and yet assign primary accent with an iamb in the default case? Do languages exist that use iambic feet for rhythm but which use a trochee to place default primary accent on the left-hand side in the domain? If these mismatches exist, then what percentage of the total sample do they constitute? An example of the third type of query that we can execute in StressTyp yields the answer. As in most databases, we can use combined queries, designed to list for which records (languages) in the database combinations of two or more parameter values hold true, and thus, to discover whether dependencies between these parameters exist. In our case, the dependency is quite clear. Prototypical trochaic languages use trochaic feet for rhythm and assign primary accent at the left [within the accentual domain, not necessarily the left side of the word], whereas iambic languages do the opposite. This fact is exploited in standard metrical theory, which regards the foot on which main stress falls as a rhythmic foot, either on the left or right side of the word, instead of constructing a separate domain for primary stress with its own rules. In StressTyp parameters, the dependency is found by comparing the fields *Stress if Both Light* and *Rhythm Type*. Standard metrical theory predicts that these two fields should always have the same value, either trochaic or iambic. What we are looking for now is whether languages exist that defy this common pattern. The table below shows the results.

Table 1. Default location of primary accent, broken down by rhythmic foot type.

		Rhythm Type	
		Trochaic	Iambic
Stress if Both Light	Trochaic	133	2
	Iambic	9	15

With respect to the total number of languages for which we have clear information on both the foot type for rhythm and the default edge for primary accent, the number of mismatches is not impressive (only 7%), but cannot be dismissed as insignificant either. As noted in section 1, such mismatches are problematic in standard metrical theory (and its offshoots). In a theory that separates the treatment of primary and secondary stress, these are logical possibilities. As such the mismatches provide support for the validity of the separation between primary and secondary accent assignment. One

might object that a theory that separates primary and secondary accent predicts a random correlation between the values for both fields queried here and, strictly speaking, that is true. We believe that there are two reasons for why most primary accent patterns mirror rhythm (or vice versa). Firstly, it is reasonable to assume that primary accent locations are grounded in rhythmic patterns historically; thus they would, in principle, start out mirroring rhythm. Secondly, the correlation will remain stable, even after the two aspects of word accentuation patterns have been separated into a lexical (primary accent) and a post-lexical part (rhythm) because we expect rhythmic patterns to be constructed avoiding clashes with the primary accent; thus we expect that, for example, a penultimate primary accent will cause a trochaic rhythm to the extent that rhythmic patterns tend to 'echo' away from the primary accent site. In this sense, rhythm will tend to mirror primary accent.

As a final exercise, let us try a cross database query. For that, we need to turn to the TDS, which allows us to define integrated queries using fields from multiple databases (a second example of how to use the TDS for such a query is given in Dimitriadis et al., this volume). Let us stay in the realm of heads and edges and compare the *Stress if Both Light* parameter to a field from another database that has nothing to do with stress, but which does relate to headedness. The Typological Database Nijmegen (http://www.hum.uva.nl/tds) contains fields in which information on basic word order is stored. It has been claimed that headedness in syntax may be correlated with headedness in phonology. If we query the TDS for the full matrix of possible combinations of default location for primary stress and word order, we obtain the following result.

Table 2. Default location of primary accent, broken down by basic word order types

		Stress if Both Light	
		Trochaic	Iambic
Word order	SVO	19	8
	OVS	0	0
	VSO	11	2
	VOS	0	0
	SOV	24	5

Concentrating on the order of Object and Verb we conflate the values of VSO and SVO (ignoring the empty OVS and VOS categories). Unfortu-

nately, there is no clear correlation to be discovered, though the percentage of trochaic languages that have a verb-final predicate is higher than that same percentage for the iambic languages. We hasten to add, however, that we must not draw iron-clad conclusions from such a small sample. We merely present this little exercise as an example of the things one can do with cross database queries. This concludes our tour of the typological possibilities that StressTyp offers. Other types of queries will be possible, but these will most likely be similar to one of the four types presented above. One such type involves cross-database queries of multiple databases that all store data on stress. These queries are potentially very interesting, since they hold the key to confirmation of conclusions drawn from StressTyp data through independent sources, but also to expansion of the sample on which such typological conclusions can be drawn. Therefore we devote the next section to a brief comparison of three databases on stress.

5. Comparison to other databases

Although (as far as we know) StressTyp was the first initiative to systematically store information on stress patterns electronically, and is at present by far the most complete one, both in coverage of aspects of the phenomenon and in number of languages, it is not the only one. There are two other computerized databases on stress that we are aware of. We discuss them below.

5.1. Bailey's Stress System Database

The first stress database to appear beside StressTyp was developed by Todd Mark Bailey, who derived his database from the data he collected for his dissertation: 'Nonmetrical constraints on stress' (Bailey 1995). Information about his "Stress System Database", can be found at http://www.cf.ac.uk/psych/subsites/ssdb/. It is interesting to note that Bailey's theoretical perspective on stress systems incorporates the idea of viewing primary and secondary accent as separate phenomena, deserving different theoretical mechanisms. In particular, he shares with van der Hulst (1984 and later publications) the idea that primary accent assignment is not based on rhythm, with the possible exception of count systems. In his database, each record consists of five fields:

- **Long Word SPC** (Syllable Priority Code)
  A code that captures the basic pattern of the language
- **Short Word SPC**
  A code that captures the pattern of words that are shorter than the maximal size of the 'stress window'
- **Language**
  One or more language names
- **References**
  The source(s) of the code
- **Comments**
  Miscellaneous comments about the nature of syllable weight, or the foot type of 'bottom-up stress' systems (when rhythm crucially feeds primary accent; cf. section 1), lexical exceptions, or the presence of special features such as ternary rhythm.

Bailey's coding system is based on the idea that in locating primary stress certain syllables get priority over others. For example, in Hopi primary stress falls on the first syllable if heavy, otherwise on the second syllable. Bailey encodes this as follows:

(13) Hopi: 12/2

Here there are two priority codes separated by '/'. The first code specifies the relative priority among heavy syllables saying that syllable 1 (the first syllable from the left; cf. below) takes priority over syllable 2, i.e. a heavy second syllable is stressed only if the first is not heavy. The second term specifies the case in which neither the first nor the second syllable is heavy. In dealing with a system in which the same calculation occurs on the right side of the word (stress the last if heavy, otherwise the penult), Bailey uses the same coding, differentiating the two by adding 'R' (right side of the word) or 'L' (left side):

(14)

	Bailey Code	StressTyp Code
Hopi:	12/2L	I/S
Hawaiian:	12/2R	U/P

(Read: "1 if heavy else 2 if heavy else 2", StressTyp codes given for reference)

We wonder why Bailey did not use '1/2' instead ("1 if heavy else 2"). This would be almost equivalent to the StressTyp code which builds the L versus R option into the syllable count: 1/2L = I/S. If stress is weight-insensitive the codes are:

(15) Latvian: 1L ST: I
Cambodian: 1R ST: U

Stress systems with multi valued (non-binary) weight distinctions challenge any coding system. In Pirahã, for example, there are five weight levels. The heaviest syllable in a right-edge three-syllable window receives primary stress. In case of a tie, the rightmost heaviest syllable wins. If there are no heavies, the rightmost (final) syllable is stressed. For cases like Pirahã Bailey's coding assumes the following form:

(16) Pirahã: 123/123/123/123/1R ST: Pirahã A-U/U
(Read: 1, else 2 else 3 in the heaviest weight class / 1, else 2 else 3 in the next heaviest weight class, 1, else 2 else 3 in the next heaviest weight class / 1, else 2 else 3 in the next heaviest weight class / 1 if no heavies. StressTyp code read: Antepenult if heavier than the two syllables in the window, else Ultimate if it is equal in weight to the penult, else select the heaviest of the two syllables in the domain, Ultimate if all syllables are light.)

Bailey's code seems a bit redundant, since the principle is the same whatever the weight class is. However, if the code is meant to be an algorithm we would need Bailey's level of explicitness. StressTyp does not aim at this level of explicitness in the 'stress type' field because a fully explicit coding is provided in another set of fields.

For unbounded systems, Bailey uses the numbers 1 through 4 to refer to the edge at which the accent is assigned and 9 through 6 for the syllables on the other side of the word, irrespective of the number of syllables:

(17) σσ σσσσσσσσ σσ
9 8 7 6 ..... 4 3 2 1

Thus he assumes (for practical purposes, not as a theoretical claim) that even in unbounded systems, primary accent is assigned within a four syllable window at one of the edges.

(18) First/First system: 12..89/1L ST: F/F
First/Last system: 12..89/9L ST: F/L

(read: 1 if heavy, else 2 if heavy ...else penult if heavy, else final if heavy / else first)

(read: 1 if heavy, else 2 if heavy ...else penult if heavy, else final if heavy / else last)

The stress system of Hindi poses a challenge: according to Bailey, primary stress is on the rightmost superheavy (excluding the final syllable, except when this is the only superheavy in the word), else on the rightmost heavy (excluding the final syllable), else on the initial syllable:

(19) Hindi: 23..891/23..899/9R ST: Hindi U%A

(Read: penult, else antepenult, ...else third, else second, else last in heaviest weight class / penult, else antepenult, ...else third, else second in next heaviest weight class / else first)

Here we witness a difference in interpretation of the sources or use of different sources. Whereas Bailey's characterization of the system is that it is unbounded, StressTyp analyzes the Hindi system as a bounded system. This is not the place to resolve the true nature of the system, which, as is widely admitted (cf. Hayes 1995), is both complex and possibly dialectally heterogeneous. What we learn here is that a comparison of different databases reveals areas that call for a closer look.

Finally, we discuss Wargamay as an example of a count system in which the first syllable is stressed in words with an even number of syllables, and the second in words with an odd number of syllables. Bailey assumes that the word is parsed in trochaic feet from right to left and that the final foot head receives primary accent (single leftover syllables are not parsed). The priority code must then take rhythmic strength to be a type of weight. '@s' means 'weight is secondary stress' (this approach is rather similar to the one we promoted in Goedemans, van der Hulst and Visch 1996).

(20) Wargamay: 12@sL ST: F(CNT)

(Read: "the first if a foot head, the second if a foot head")

In conclusion, we see here how two approaches that differ primarily in the depth of coding. Whereas Bailey tries to capture the nature of stress systems

in a single code, StressTyp, in addition to using a comparable code, provides a more detailed and fractured analysis of the system. In addition, StressTyp offers much greater overall detail in many other areas. This is not a criticism of Bailey's work which, as we assume, served his goals quite well. StressTyp is more ambitious in that it tries to be a tool for all researchers working on stress phenomena. It thus must be richer, more redundant and more explicit.

5.2. Gordon's database of QI stress systems

The second database that we discuss here was created by Matthew Gordon in the context of an inquiry into weight-insensitive systems (Gordon 2002) Hence, his database specifically only contains information about systems of this type. His codes are:

(21) Initial (+ antepenult, + penult)
Peninitial
Antepenult (+ initial)
Penult (+ initial)
Final (+ initial)

The information between parentheses refers to the location of secondary stress, which is encoded if mentioned in his sources. Presumably, any combination of a primary accent and secondary accent code is, in principle, possible. In addition, he uses the following code for systems that have no primary accent and only rhythm.

(22) (e,l) even syllables from left to right (first beat is second syllable = S)
(e,r) even syllables from right to left (first beat is penultimate syllable = P)
(o,l) odd syllables from left to right (first beat is initial syllable = I)
(o,r) odd syllables from right to left (first beat is ultimate syllable = U)

In parentheses we added the location of the first beat that is assigned, and 'translated' this into the primary accent position to facilitate comparison to the coding in StressTyp.

Gordon's database was constructed for an even more specific purpose than Bailey's database. It is only to be expected, then, that his dataset and his coding system is much narrower than that of StressTyp, and Bailey's

database. Again, this is not a criticism, it merely is a consequence of the different goals for which databases are built.

5.3. Comparison

In this section, we will make some explicit comparisons between the three databases. Let us first present a few numbers, determining the overlap between the three databases. Investigation on the basis of automated name comparison with reference to SIL code and sources in cases of doubt allowed us to create a relational combined database in which language names together with elementary stress type encoding for the three databases are stored. Languages that occur in more than one database are linked.[14] Simple queries now tell us that 160 of the 197 languages in the Stress System Database (SSD) also occur in StressTyp, while the SSD contains 37 languages that are not present in StressTyp. Gordon's database contains 273 languages. The overlap with StressTyp is 123 languages, so that there are no less than 150 languages that do not occur in StressTyp.[15]

The overlap between Gordon and Bailey is 62 languages, while the overlap between these two and StressTyp is 51 languages. All of these are weight-insensitive (because no other type is represented in Gordon's database).

We now look at the ways in which these 51 languages are encoded in the three databases. In the Gordon column we added the equivalent StressTyp code for the '(x,y)' codes:

---

[14] We thank Menzo Windhouwer who wrote the scripts that generated our comparison lists.

[15] We have to be careful not to count languages twice in these comparisons. In four cases for instance (Sierra Miwok, Turkish, Mam, Dakota) a single SSD record corresponds to two StressTyp records (if two varieties [e.g. dialects] of a language appear in StressTyp and we cannot decide to which of these the SSD record relates). In this case a comparison list will show 164 rows, but in these rows the SSD languages above appear twice. When we compare StressTyp and SSD these should, of course, be counted only once. In the comparison of StressTyp to Gordon's database a similar reduction has been applied with respect to the records for Tsaxur, Aramaic, Hebrew and Basque (the latter even corresponds to nine StressTyp records).

	TOTAL CODE INFO			
StressTyp_Name	**StressTyp**	**Gordon**	**Bailey long word**	**Bailey short word**
Armenian	L/F (L but one)	U	1R	
Cavineña	P	(e,r) P	2R	
Cayubaba; Cayuvava	A (NMS)	A	3R (3+)	1L (2–)
Chamorro	P/A	P	2R	
Apoze	A	A	3R (3+)	1L (2–)
Baadi; Bardi; Badimaya	I	(o,l) I	1L	
Banggarla; Parnkalla	A	A	3R (3+)	1L (2–)
Cahuilla, (Desert and Mountain Dialects)	I	I	1L	
Czech	I	(o,l) I	1L	
Dakota; Sioux	S	S	2L	
Dehu; Lifu	I	(o,l) I	1L	
Dieri; Diyari	I	(o,l) I	1L	
Djingili; Tjingili	P	(e,r) P	2R	
Emae; Mae	A	A	3R (3+)	1L (2–)
French	U/P	U	1R	
French	U/P	U	1R	
Garawa	I	I	1L	
Georgian	A;I (NMS)	A	1L	
Mansi; Vogul	I	(o,l) I	1L	
Paiute, Southern	S	P	2L	
Tübatulabal	U (NMS)	U	1R	
Hebrew, Tiberian	U/P	U	12/21/1R	
Hungarian	I	(o,l) I	1L	
Icelandic	I	(o,l) I	1L	
Karelian	I	(o,l) I	1L	
Mullukmulluk; MalakMalak	F (CNT)	(e,r) F	12@s L (3+)	1L (3–)
Kuku-Yalanji	I	I	1L	
Latvian	I	I	1L	
Lezgi; Lezgian; Kiurintsy	I/1 (IRR)	S	1R	
Liv; Livonian	I	(o,l) I	1L	

TOTAL CODE INFO				
StressTyp_Name	**StressTyp**	**Gordon**	**Bailey long word**	**Bailey short word**
Macedonian	A	A	3R (3+)	1L (2–)
Mapuche; Araucanian; Aucan	S	(e,l) S	2L (3+)	1L (2)
Maranunggu	I	(o,l) I	1L	
Meso Grande Diegueño (*)	U/P	U	12/2R	
Nengone	P	(e,r) P	2R	
Ngalkbun; Dalabon; Boun	P;I	(o,l) I	1L	
Ono	I	(o,l) I	1L	
Pintupi-Luritja	I	(o,l) I	1L	
Piro; Yine	P	P	2R	
Pitta pitta; Bidhbidha	I	(o,l) I	1L	
Polish	P	P	2R	
Ruija	I	(o,l) I	1L	
Selepet	I	(o,l) I	1L	
Sorbian	I	I	1L	
Swahili	P	P	2R	
Tajik	U	U	1R	
Uzbek, Northern	U	U	1R	
Vod; Votic	I	(o,l) I	1L	
Warao; Guarao	P	(e,r) P	2R	
Weri; Were	U	(o,r) U	1R	
Wongkumara; Wankumara	I	(o,l) I	1L	

We observe that the codes presented here match to a high degree, which means that all three databases either made the same kinds of mistakes or, a more likely scenario, that there are very few, if any, mistakes in this sample.

Alternatively we can look at the relative number of QI systems per type in the three databases and see whether there are significant differences in the percentages.

Weight-Insensitive Systems	**Bailey**	**Gordon**	**StressTyp**
I	40 (=43,5%)	103 (=38%)	92 (=33%)
S	3 (= 3%)	15 (= 6%)	16 (= 5,5%)
T	0	0	1 (= 0,5%)
A	8 (= 9%)	8 (= 3%)	12 (= 4%)
P	16 (=17,5%)	77 (=28%)	110 (=39%)
U	25 (=27%)	69 (=25%)	50 (=18%)
Total	**92** (=47% of total 197)	**272**	**281** (=55% of total 510)

The global patterns are comparable. However, a few striking differences need an explanation. Since the sample sizes in Gordon's database and StressTyp are almost identical and quite sizeable, let us concentrate on these two first. Sample size should help here to reveal real tendencies. The one major difference between Gordon's database and StressTyp is the size of the P category. In StressTyp it is much too large, a feature we have commented on before. It more than likely is due to the fact that during the Prosody of Indonesian Languages project, we have added many Austronesian languages. Since these have predominantly quantity insensitive stress systems with primary stress on the penultimate syllable (see van Zanten, Stoel and Remijssen, to appear), this category is overrepresented in StressTyp. Should we reduce the number of Austronesian languages in the sample, we are sure that the overall percentages will eventually quite closely resemble those we found for the Gordon database, since the percentage for P will go down and that of the other categories will go up.

The smaller database in this overview is Bailey's, and it is therefore more susceptible to influence of imbalances in the sample (like the one we noted above for StressTyp). With respect to the other two, I, A and U are overrepresented, and P is underrepresented. Careful analysis of the Bailey sample may reveal what causes this. If we assume that this database is a 'little off' because of the relatively low sample size, and that the other two reflect more accurately what is going on in the languages of the world, we may, in any case, conclude that languages prefer initial stress, prefinal stress being a good second, with final stress not far behind. Antepenultimate stress and stress on the second syllable are relatively uncommon (only 1 out of 20 languages for both categories), while stress on the third syllable is virtually non existent.

Finally, we can compare the numbers of weight-sensitive systems in the SSD and StressTyp:

Weight-Sensitive Systems	**Bailey**	**StressTyp**
I or S	14 (= 14%)	37 (= 20%)
I, S or T – S or T	0	2 (= 1%)
P or U	22 (= 21%)	65 (= 35%)
U or P – P or A (or pre-A)	23 (= 22%)	27 (= 15%)
Unbounded	44 (= 43%)	54 (= 29%)
Total	**103** (= 52% of total)	**185** (= 36% of total)

Striking differences here are the relatively high number of unbounded languages in the SSD, and the fact that in the SSD there are as many systems that have stress on one of the last three syllables as systems that have stress on one of the final two syllables only. In the latter case, StressTyp languages clearly prefer stress to occur on one of the final *two* syllables. The cause of these differences eludes us, but we suggest that it may again be due to imbalances in the samples. We have no way of telling which column of percentages more closely reflects the objective truth, but we tend to place more trust in the one with the larger sample size.

In conclusion we note that this comparison supports the StressTyp data. We have seen that the codes for QI languages closely match the codes for the same languages in two other databases. We have also seen that one of these two other databases contains an almost equally large sample of QI languages and that the percentages of QI languages in each of the possible categories in this database are similar to those we find in StressTyp, especially if we reduce the number of Austronesian languages in StressTyp (since these are overrepresented). Finally, we note that it seems imperative for quantitative research on stress systems to work with rather large sample sizes. Without further research, we cannot be sure which of the three databases we compared here comes closest to accurately describing the tendencies in the languages of the world, but we do think it is a tell-tale sign that the smaller one of the three seems to be the odd man out in the large comparison of QI systems in all databases, and shows some unexpected patterns in the comparison of the QS systems. We suggest that the StressTyp sample has enough critical mass to do quantitative research, but that it could benefit greatly from an increase to about 1500 records, if the additional languages are carefully selected to make the whole sample genetically and areally more balanced.

6. Concluding remarks

In this chapter we have provided a detailed description of StressTyp, a database for word accentual systems in the languages of the world, discussing both the history, current state and intended future developments. We have indicated the record structure of the database and shown how the stored information can be used for queries of various kinds.

It is our intention to continue the development of Stresstyp both regarding its structure and content and we welcome any kind of comment based on reading this chapter or using the database. Finally, let us repeat that we also welcome any kind of collaboration either in the area of word accentual systems or, more broadly in other areas of word phonology toward establishing larger and more ambitious projects.

Appendix: Additional StressTyp fields and Codes for the Type field

A. Fields not mentioned in section 3.

Exceptional Patterns
Exceptional Patterns
Examples
Source

Unaccented Words
Category
Monosyllables Only Y/N

Prefixes	***Suffixes***
Stress Neutral Y/N	Stress Neutral Y/N
Stress Sensitive Y/N	Stress Sensitive Y/N
Stressed Inherently Y/N	Stressed Inherently Y/N
Cyclic Effects Y/N	Cyclic Effects Y/N
	Pre-stress Y/N
Comments	

Compounds
Category 1: sw/ws
Category 2: sw/ws
Category 3: sw/ws
Category 4: sw/ws

Clitics
Comments
Examples

Phonetic Realization
Lexical Pitch Y/N
Tone Classes Y/N

Processes
Processes
Examples

B. StressTyp codes

Fixed Stress Patterns

I Primary stress always occurs on the initial syllable.
S Primary stress always occurs on the second syllable.
T Primary stress always occurs on the third syllable.
A Primary stress always occurs on the antepenultimate syllable.
P Primary stress always occurs on the penultimate syllable.
U Primary stress always occurs on the final syllable.

Variable stress patterns

I/I Place stress on the initial syllable if it is heavy (even if the second syllable is also heavy), otherwise place stress on the second syllable if it is heavy, if neither first nor second syllables are heavy, then place stress on the first syllable.

I/S Place stress on the initial syllable if it is heavy (even if the second syllable is also heavy), otherwise place stress on the second syllable if it is heavy, if neither first nor second syllables are heavy, then place stress on the second syllable.

S/I Place stress on the second syllable if it is heavy (even if the first syllable is also heavy), otherwise place stress on the first syllable if it is heavy, if neither first nor second syllables are heavy, then place stress on the first syllable.

S/T Place stress on the second syllable if it is heavy (even if the third syllable is also heavy), otherwise place stress on the third syllable if it is heavy, if neither second nor third syllables are heavy, then place stress on the third syllable.

U/U Place stress on the ultimate syllable if heavy (even if the penultimate syllable is also heavy), otherwise place stress on the penultimate syllable if it is heavy, if neither are heavy, place stress on the ultimate syllable.

U/P Place stress on the ultimate syllable if heavy (even if the penultimate syllable is also heavy), otherwise place stress on the penultimate syllable if it is heavy, if neither are heavy, place stress on the penultimate syllable.

P/U Place stress on the penultimate syllable if heavy (even if the ultimate syllable is also heavy), otherwise place stress on the ultimate syllable if it is heavy, if neither are heavy, place stress on the ultimate syllable.

P/P Place stress on the penultimate syllable if heavy (even if the ultimate syllable is also heavy), otherwise place stress on the ultimate syllable if it is heavy, if neither are heavy, place stress on the penultimate syllable.

Or: Place stress on the penultimate syllable if heavy (even if the antepenultimate syllable is also heavy), otherwise place stress on the antepenultimate syllable if it is heavy, if neither are heavy, place stress on the penultimate syllable. The code for this type is also P/P with the note that EM=right.

P/A Place stress on the penultimate syllable if heavy (even if the antepenultimate syllable is also heavy), otherwise place stress on the antepenultimate syllable if it is heavy, if neither are heavy, place stress on the antepenultimate syllable.

A/A Place stress on the antepenultimate syllable if heavy (even if the penultimate syllable is also heavy), otherwise place stress on the penultimate syllable if it is heavy, if neither are heavy, place stress on the antepenultimate syllable.

F/F Place stress on the first heavy syllable in the word. If there is no heavy syllable present, place stress on the first syllable.

F/L Place stress on the first heavy syllable in the word. If there is no heavy syllable present, place stress on the last syllable.

L/F Place stress on the last heavy syllable in the word. If there is no heavy syllable present, place stress on the first syllable.

L/L Place stress on the last heavy syllable in the word. If there is no heavy syllable present, place stress on the last syllable.

Other codes and connectives

Lex The locations of either main or secondary stresses are specified in the lexicon for the majority of the words in the language. This means that stress

can be phonemic, because two non-monosyllabic words that are identical in segmental make up may differ in stress location and meaning.

NMS Stands for No Main Stress. All stresses are reported to be equally prominent.

L(CNT) This is a so-called "count system". Primary stress is assigned to the head of the last foot in the word. Stress is assigned from left-to-right. This leads to different stress locations for words with an odd and an even number of syllables.

F(CNT) This is a so-called "count system". Primary stress is assigned to the head of the first foot in the word. Stress is assigned from right-to-left. This leads to different stress locations for words with an odd and an even number of syllables, usually Initial stress in the even case and Second stress in the odd case.

IRR is used to indicate that stress varies unpredictably within the domain.

Pitch and **Tone** are added between parentheses to indicate interaction between pitch or tone assignment and metrical structure.

; This connective indicates that there is some degree of variation between two (or more) patterns for main stress. The dominant pattern comes before the semicolon.

- This connective indicates that "superheavy" syllables are involved in the computation of stress. If such a syllable occurs in the position indicated before the hyphen, it bears stress. Otherwise a standard rule (placed after the hyphen) comes into operation.

% This connective indicates a stress shift outside the bounded stress domain under special circumstances. Stress shifts to the location after the % sign under these circumstances, and stays in the bounded domain otherwise.

References

Anderson, John M. and Charles Jones

1974 Three theses concerning phonological representations. *Journal of Linguistics* 10: 1–26.

1977 *Phonological Structure and the History of English*. Amsterdam: North-Holland.

Anderson, John M. and Colin J. Ewen

1987 *Principles of Dependency Phonology*. Cambridge: Cambridge University Press.

Bailey, Todd Mark

1995 Nonmetrical constraints on stress. Ph.D. diss., University of Minnesota.

Garde, Paul

1968 *L'Accent*. Paris: Presses universitaires de France.

Goedemans, Rob

to appear Stress Typology. In *Stress Patterns of the World: The Data*, R. Goedemans, H. van der Hulst and E. A. van Zanten (eds.). Berlin/New York: Mouton de Gruyter.

Goedemans, Rob and Harry van der Hulst

2005a Fixed stress locations. In *The World Atlas of Linguistic Structures*, Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie (eds.), 62–65. Oxford: Oxford University Press.

2005b Weight-sensitive stress. In *The World Atlas of Linguistic Structures*, Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie (eds.), 66–69. Oxford: Oxford University Press.

2005c Weight factors in weight-sensitive stress systems. In *The World Atlas of Linguistic Structures*, Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie (eds.), 70–73. Oxford: Oxford University Press.

2005d Rhythm types. In *The World Atlas of Linguistic Structures*, Martin Haspelmath, Matthew Dryer, David Gil, and Bernard Comrie (eds), 74–77. Oxford: Oxford University Press.

Goedemans, Rob, Harry van der Hulst, and Ellis Visch

1996a The organization of StressTyp. In *Stress Patterns of the World*, Rob Goedemans, Harry van der Hulst, and Ellis Visch (eds.), 27–68. (Holland Institute of Generative Linguistics Publications 2.) The Hague: Holland Academic Graphics.

1996b *StressTyp Manual*. Leiden: Holland Institute of Generative Linguistics.

1996c StressTyp: A database for prosodic systems in the world's languages. *Glot International* 2 (1/2): 21–23.

Goedemans, Rob, Harry van der Hulst, and Ellen van Zanten
to appear *Stress Patterns of the World: The Data.* John Benjamins: Amsterdam.
Gordon, Matthew
2002 A factorial typology of quantity insensitive stress. *Natural Language and Linguistic Theory* 20: 491–552.
Greenberg, Joseph H. and Dorothy Kashube
1976 Word prosodic systems: A preliminary report. *Working Papers on Language Universals* 20: 1–18.
Halle, Morris and William J. Idsardi
1994 General properties of stress and metrical structure. In *A Handbook of Phonological Theory*, John A. Goldsmith (ed.), 403–443. Oxford: Basil Blackwell.
Halle, Morris and Jean-Roger Vergnaud
1987 *An Essay on Stress.* Cambridge, MA: MIT Press.
Harms, Robert T.
1981 A Backwards Metrical Approach to Cairo Arabic Stress. *Linguistic Analysis* 7: 429–451.
Hayes, Bruce
1980 A Metrical Theory of Stress. Ph.D. diss., Massachusetts Institute of Technology. [Distributed in 1981 by the Indiana University Linguistics Club, Bloomington, Indiana.]
1995 *A Metrical Theory of Stress: Principles and Case Studies.* Chicago, Illinois: University of Chicago Press.
Hulst, Harry van der
1984 *Syllable Structure and Stress in Dutch.* Dordrecht: Foris Publications.
1996 Separating primary accent and secondary accent. In *Stress Patterns of the World*, Rob Goedemans, Harry van der Hulst, and Ellis Visch (eds.), 1–26. (Holland Institute of Generative Linguistics Publications 2.) The Hague: Holland Academic Graphics.
1990 *The book of stress.* Unpublished manuscript, Department of General Linguistics, Leiden University.
1992 The independence of main stress and rhythm. Paper presented at the Krems Phonology Workshop.
1997 Primary accent is non-metrical. *Italian Journal of Linguistics/Rivista di Linguistica* 9 (1): 99–127.
1999 Word accent. In *Word Prosodic Systems in the Languages of Europe*, H. van der Hulst (ed.), 3–116. Berlin/New York: Mouton de Gruyter.
2000a Issues in foot typology. In *Issues in Phonological Structure*, Mike Davenport and Stephen J. Hannahs (eds.), 95–127. Amsterdam: John Benjamins. [Also appeared in *Toronto Working Papers in Linguistics* 16 (2007): 77–102.]
2000b Metrical phonology. In *The First Glot International State-of-the-Article Book: The Latest in Linguistics*, Lisa Chen and Rint Sybesma

(eds.), 307–326. (Studies in Generative Grammar 48.) Berlin/New York: Mouton de Gruyter. [Originally published in *Glot International* (1995) 1(1): 3–6.]

2002 Stress and accent. In *Encyclopedia of Cognitive Science*, Vol. 4, Lynn Nadel (ed.), 246–254. London: Nature Publishing Group.

2006 Word stress. In *The Encyclopedia of Language and Linguistics.* 2nd Edition, Vol. 13, Keith Brown (ed.), 655–665. Oxford: Elsevier.

to appear Brackets and grid marks or theories of primary accent and rhythm. In *Representations and Architecture in Phonological Theory*, Charles Cairns and Eric Raimy (eds.), Cambridge, MA: MIT Press.

Hulst, Harry van der (ed.)

1999 *Word Prosodic Systems in the Languages of Europe.* Berlin/New York: Mouton de Gruyter.

Hulst, Harry van der and Jan Kooij

1994 Two modes of stress assignment. In *Phonologica 1992*, Wolfgang Dressler and John Rennison (eds.), 107–114. Torino: Rosenberg and Sellier.

Hulst, Harry van der and Aditi Lahiri

1988 On foot typology. *NELS* 18: 286–209.

Hurch, Bernhard

1995 Accentuations. In *Natural Phonology: The state of the Art on Natural Phonology*, Bernhard Hurch and Richard A. Rhodes (eds.), 73–96. Berlin/New York: Mouton de Gruyter.

Hyman, Larry M.

1977 On the nature of linguistic stress. In *Studies in Stress and Accent*, Larry M. Hyman (ed.), 37–82. (Southern California Occasional Papers in Linguistics 4.) Los Angeles: Department of Linguistics, University of Southern California.

Idsardi, William J.

1992 The computation of prosody. Ph.D. diss., Massachusetts Institute of Technology.

Kager, René

1993 Alternatives to the iambic-trochaic law. *Natural Language and Linguistic Theory* 11: 381–432.

Liberman, Mark and Alan Prince

1977 On stress and linguistic rhythm. *Linguistic inquiry* 8: 249–336.

Lockwood, David G.

1983 Parameters for a typology of stress. In *The Ninth LACUS Forum 1982*, John Morreall (ed.), 231–241. Columbia, South Carolina: Hornbeam Press.

McGarrity, Laura W.

2003 Constraints on patterns of primary and secondary stress. Ph.D. diss., Department of Linguistics, Indiana University.

Prince, Alan
1983 Relating to the grid. *Linguistic Inquiry* 14: 19–100.

Prince, Alan and Paul Smolensky
1993 Optimality Theory: Constraint Interaction in Generative Grammar. (Technical Report #2 of the Rutgers University Center for Cognitive Science and Computer Science Department, University of Colorado at Boulder.) Piscataway, New Jersey: Rutgers University.

Pulgram, Ernst
1970 *Syllable, word, nexus, cursus.* The Hague: Mouton.

Roca, Iggy
1986 Secondary stress and metrical rhythm. *Phonology Yearbook* 3: 330–341.

Vergnaud, Jean-Roger and Morris Halle
1978 Metrical structures in phonology. Unpublished Ms., Massachusetts Institute of Technology.

Zanten, Ellen van and Rob Goedemans
2007 A functional typology of Austronesian and Papuan stress systems. In *Prosody in Indonesian Languages.* LOT: Occasional Series 9, Vincent van Heuven and Ellen van Zanten (eds.), 63–88, Utrecht: Igitur, Utrecht Publishing and Archiving Services.