

On the comparability of prosodic categories: why ‘stress’ is difficult

Nikolaus P. Himmelmann
Universität zu Köln

July 2022

To appear in a special issue on phonological typology

For the time being, use <https://ling.auf.net/lingbuzz/006684> to refer to this ms.

Abstract

This article argues that the prosodic category *stress* in West Germanic languages, which implicitly underlies practically all work on stress, is a complex cluster concept consisting of at least six dimensions which in turn involve a number of subdimensions. Because of its complexity, this concept is not useful for cross-linguistic comparison. A promising starting point for further typological inquiry is one of the six dimensions, i.e. acoustic and auditory prominence. However, identifying acoustic and auditory prominence distinctions cross-linguistically is also not straightforward and requires considerable empirical effort. Nevertheless, cross-linguistic comparison is still possible in the case of ‘difficult’ cluster concepts such as *stress* and does not require the use of arbitrary comparative concepts.

Keywords: stress, duration, syllable, prosody, cross-linguistic comparison

1. Introduction

In the call for contributions to this volume, the editors ask “whether formal descriptive categories can be equated and made comparable across languages”? With this, I assume, they are referring to the current debate on comparability in typology, specifically to the claim that descriptive categories need to be distinguished from comparative concepts, as proposed by Haspelmath (2018 *inter alia*). In Haspelmath’s approach, language-specific categories are considered to be incommensurable because they are to be exclusively based on language-internal evidence. Comparative concepts, on the other hand, are considered to be essentially arbitrary and should be defined in such a way that the properties defining a particular concept can be easily identified across different languages.

The arbitrary comparative concepts-approach assumes that the cross-linguistically identifiable properties used to define comparative concepts have no role to play in defining language-specific categories (and vice versa). It is this assumption where many other approaches to cross-linguistic comparability (e.g., Auwera & Sahoo 2015, Beck 2016, Dahl 2016, Himmelmann 2021, Lander & Arkadiev 2016, Lehmann 2018, Moravcsik 2016, Round & Corbett 2020, Spike 2020) clearly disagree. Some of the properties used to define language-specific categories are cross-linguistically identifiable and thus may serve as a basis for cross-linguistic comparison, as further detailed in sections 2 and 3 below.

In this regard, it bears emphasizing that the differences between the arbitrary comparative concepts-approach and other approaches should not be exaggerated. As David Beck (2016) notes, quite a few linguistic categories are defined in such a way that they can be used in both language description and language comparison without major problems (and hence without

causing major controversies). He calls them *portable concepts* (Beck 2016: 393).¹ Typical examples include *personal pronouns*, *comitative*, *dual*, or *bilabial stop*.

But there are also other categories which cannot be applied as easily across languages. Usually, these are complex cluster categories involving a number of different dimensions that do not necessarily correlate with each other. A primary example is the notion *subject* which is well known for extended controversies regarding its cross-linguistic identifiability. An instructive example is the volume edited by Li (1976) which includes much-quoted early proposals for a cross-linguistically applicable definition of *subject*.

The current contribution is concerned with the notion of *stress* which has been used for a long time as if it were a portable concept, not requiring careful and explicit definition in language description or cross-linguistic comparison.² Only in the last two decades or so, has it become clearer that *stress* is also a complex cluster concept, associated with a range of different phenomena in different dimensions of linguistic structure, ranging from lexical entries to intonational units. Therefore, it is necessary to carefully distinguish the different phenomena associated with this notion and to identify the ones that are likely to be cross-linguistically identifiable properties and thus can serve as *tertia comparationis* in cross-linguistic comparisons. While the status of stress as a cluster concept has been recognized in specialist circles to varying degrees, a fully explicit statement of this view is still missing. It is one of the two aims of this contribution to make explicit what exactly is involved when stress is conceived of as a cluster concept and to make this view more widely known within the typological community at large.

The other goal of the present contribution is to use *stress* as an illustration of the kind of problems that cluster concepts typically pose for cross-linguistic comparison, and the kind of solutions that are available for such problems. Although cluster concepts typically cause some difficulties for cross-linguistic comparison, it is not the case that cross-linguistic comparisons involving cluster concepts require a methodology or a conceptual framework that differs from the one used in comparing portable concepts. The difference between comparing cluster concepts and portable concepts is actually an almost trivial one: the former requires more work (and time), conceptually and empirically.

The exposition proceeds in three steps. We begin with a few general notes on cross-linguistic comparability, identifying three different basic types of properties of language-specific categories that can also be used in cross-linguistic comparisons. Section 3 briefly looks at ‘the syllable’ as an instructive example for cross-linguistic comparison in the prosodic domain. Section 4 is an attempt to sort out some of the problems involved in comparing word-level prosodic prominence (‘stress’) across languages. Section 5 concludes.

2. On cross-linguistic comparability in general

As already pointed out in the introduction, the current contribution is based on the view that cross-linguistic comparison is possible and empirically valid, because language-specific

¹ Actually, Beck uses *term* in roughly the same sense as *concept* is being used here and in much of the recent literature on the comparability of linguistic categories.

² Of course, this does not mean that there are no publications from earlier decades which provide explicit and detailed definitions for their use of the concept *stress*. Work by Hyman and van Heuven dating back to the 1970s, for examples, includes such explicit definitions, and it is thus not by chance that their work plays a major role in the present argument.

categories have properties in common which can be reliably identified across unrelated languages, in this regard differing from an approach that makes use of essentially arbitrary comparative concepts (cp. Himmelmann 2021: 17-24). More specifically, it is assumed that language-specific categories may have the following three types of properties in common: phonetic and semantic substance, methodological and theoretical assumptions, and shared diachronic trajectories. The first two are well-known and widely referred to both in the actual practice of cross-linguistic research and in the theoretical literature on linguistic comparability (cp. the contributions in Hyman and Plank 2018; Hyman 2006, 2008, 2017, van der Hulst 2017; and the literature already referred to in section 1). The third one is less often invoked in general discussions of comparability but is foundational for most work on grammaticization. The summary here closely follows Himmelmann (2021) where fuller argumentation and references to the literature can be found.

Phonetic substance refers to the well-known articulatory, acoustic, and auditory properties of phonological categories investigated in phonetics, which include tongue movements, air flow mechanisms, frequency bands, perceptual normalization processes, etc. On the content side, phonetic substance is matched by semantic substance, a “shared universe of meaning”, as Harder (2016) puts it.

Regarding shared methods and theoretical assumptions used in establishing language-specific categories, a very general example is the fact that all language-specific categories are based on distributions. When voiced and voiceless stops are said to be in an allophonic relation in language X, they share this distributional property with the voiceless and voiced stops in other languages where they are also in an allophonic relation, though the language-specific details may vary (and thus make for an interesting object of comparison).

The third type of properties that language-specific categories may have in common pertains to shared trajectories of historical evolution such as the well-known paths of grammaticization. To give but one standard example, inasmuch as article-like elements in different languages can be safely assumed to derive from deictic elements, they can be profitably compared with regard to the degree of their grammaticization (range of uses, prosodic dependence, phrasal position, paradigmatic oppositions entered into, etc.; cp. Greenberg 1978, Himmelmann 1997, among many others), which in turn provides the basis for a typology of article systems (e.g. Himmelmann 2001, Becker 2021).

For phonological categories, shared diachronic trajectories include, for example, common types of sound change such as lenition, palatalization, etc. Shared diachronic trajectories are, however, rarely mentioned in discussions of the cross-linguistic comparability of phonological categories. In part, this is probably simply due to the fact that little is known about the historical development of the categories to be compared in a cross-linguistic project, especially when a large number of languages from all around the world is being compared. Still, there are indirect ways to diagnose the degree to which a given category is grammaticized (or activated) in a particular system. For phonological categories, the number of rules and constraints associated with a particular unit may provide some evidence in this regard, as further illustrated in the next section (see also Blevins 2004 for pertinent discussion).

All three property types may play a role for the possibility and the success of cross-linguistic comparisons of language-specific categories. Strictly speaking, none of them is absolutely necessary for successful comparison, but they are also not of equal importance. In many ways,

comparability in terms of substance, be it phonetic or semantic, is of particular importance (but also not absolutely necessary, as shown in the next section). Similarities in methods and theoretical assumptions all by themselves are usually too abstract to render a comparison meaningful. To compare all elements in allophonic relation across different languages, for example, will not be particularly fruitful unless the comparison also includes aspects of their phonetic substance, allowing for the comparison of segments with similar place or manner features. Evidence from shared diachronic trajectories may often not be available.

The following section will illustrate the use of these three property types in cross-linguistic comparison using the syllable as its main example.

3. An instructive example: the syllable

Being able to define the objects of cross-linguistic comparison via their phonetic substance is strictly speaking not always necessary, as shown by the example of the syllable. As Ladefoged (1982: 220) aptly noted decades ago: “Although nearly everyone can identify syllables, almost nobody can define them”. That is, even though it was unclear for a long time what exactly defines syllables in phonetic terms, linguists (and native speakers) rarely disagree on what to identify as a syllable, and many useful and important descriptive as well as comparative observations have been made with regard to the structure and function of syllables before it became clearer what its phonetic substance is.³

Nowadays, following the seminal work by Browman & Goldstein (1989), it seems to be widely agreed that syllables basically have a temporal coordination function and can be defined by the synchronization of articulatory movements. The movements themselves and their temporal synchronization can be empirically investigated with the help of electromagnetic articulography (EMA). See Xu (2020) for a recent survey of the relevant literature and its implications for our understanding of the phonetics of syllables.

The syllable also makes for a good example of the relevance of diachronic evolution as an important factor in language comparison. If the view that the syllable provides a major synchronization mechanism at the very core of the human capacity for speaking is correct, syllables in the sense of temporal coordination points are necessarily a universal characteristic of spoken language. This, however, does not necessarily imply that syllables are *structural* units in all languages.

Hyman, for example, has argued that the African language Gokana lacks syllables as phonological units (Hyman 1983, 2011, 2015). The argument essentially rests on the demonstration that all major phonotactic generalizations are best stated in reference to what Hyman calls a “prosodic stem”, consisting of a root plus optional suffixes (Hyman 2011: 62). Importantly, however, this claim does not mean that Gokana speakers do not make use of a mechanism to synchronize their articulatory gestures. As Hyman puts it:

... some languages show extensive activation of syllable properties, while others show fewer. The activation is so slight in Gokana that all we have is the ambiguous interpretation of the CVV(C)V [PROSODIC STEM, NPH] maximum. ...
... As these and many other such examples amply demonstrate, phonological typologists should be concerned with characterising and explaining these interesting

³ This is not to deny that controversies may arise, for example, about syllable boundaries in the case of consonant gemination. But such cases do not put into question the observation that in most instances identifying syllables is unproblematic.

variations in how phonetic substance is phonologised in different languages – but we needn't claim that every available phonetic feature or structure will be exploited to the same extent in every language. (Hyman 2011: 82)

The notions of 'activation' and being 'phonologized' used in this quote may be interpreted to involve the kind of shared diachronic trajectories mentioned in the previous section.⁴ That is, one could hypothesize that languages may follow similar diachronic trajectories in their phonologization of a universal phonetic property of speech sounds and then become comparable with regard to the degree of phonologization they have reached.

How can the degree of activation or grammaticization of a linguistic unit or category be measured, especially in the absence of historical records that allow for a reconstruction of the diachronic trajectory? Important evidence in this regard may come from the number of regularities and constraints that the unit is involved in, *inter alia*. In the case of the syllable, for example, Hyman (2011: 62) mentions “distributional constraints conditioned by syllable structure”, “phonological rules conditioned by syllable structure”, “morphological rules or allomorphy conditioned by syllable structure”, “prosodies or word-stress targeting the syllable as a feature-bearing unit”, and “prosodic grouping of syllables into higher-order constituents, e.g. feet”.

However, evidence for the degree of activation of phonological categories is not of a single type which fits them all. Rather, such evidence can also be category specific. In the case of the intonation unit as a basic unit of prosodic organization, for example, Himmelmann et al. (2018: 240) argue that the occurrence of a limited set of boundary tones provides major evidence for the language-specific phonologization of the intonation unit in distinction to the universally available phonetic intonation unit (which is defined by universally available phonetic cues; Himmelmann et al. 2018: 213). Furthermore, it should be emphasized that to date very little is known about the actual historical trajectories of specific categories such as syllables or intonation units. Therefore, the proposal made here that such trajectories may be of relevance for cross-linguistic comparison in the prosodic domain is a hypothesis, not a well-established fact. Note also that this hypothesis does not involve the claim that stronger phonologization is necessarily incremental in each and every instance. It may very well be the case that a number of additional regularities strengthening a particular category arise at a single instant, which is, in fact, to be expected in as much as language-specific categories form a system of interdependent units.

Finally, as also noted by Hyman in a part of the quote omitted above, the idea of different degrees of activation or grammaticization also applies to 'stress'. Here distinctiveness, but also the number of rules and constraints that refer to it are seen as correlates of activation.

... some languages, like English, provide extensive activation of stress-accent properties, such that there can be no question that stress both exists and constitutes a major organisational property of English phonology. In other languages, stress accent is identifiable, but less activated. Thus, in Hungarian, 'stress does not play a significant role in the word level phonology' (Kenesei et al. 1998: 428). Similarly, in Turkish, stress can be identified mostly on the basis of F0 (Levi 2005), but not a single phonological constraint or rule refers to stress. In fact, some languages have so little

⁴ As one of the reviewers correctly points out, Hyman remains non-committal with regard to how exactly differences in activation come about. That is, Hyman may, in fact, not agree with the processual, diachronic interpretation of the terms 'activation' and 'phonologization' employed here.

activation that they have been claimed not to have stress at all, e.g. Bella Coola ... (Hyman 2011: 82)

4. Why *stress* causes problems (and hard work) for cross-linguistic comparison

Typological studies of stress are strongly influenced by our understanding of stress in European, and in particular, in West Germanic languages; for two reasons. On the one hand, these are the most thoroughly investigated languages with regard to almost any aspect of linguistic structure, and certainly with regard to prosody in general and stress in particular. On the other hand, most researchers who have written on stress are native speakers of a European language, with a clear majority having a West Germanic background. This fact is relevant because much of the published work on stress is based primarily, if not exclusively, on auditory impressions, a point we will return to in section 4.1.

Stress in West Germanic languages happens to be a highly complex phenomenon which is prone to lead to confusions when studying stress in cross-linguistic perspective. The goal of the present section, therefore, is to show that stress in West Germanic is a highly complex cluster category and that the different dimensions combined in this notion need to be clearly distinguished when engaging in cross-linguistic comparison.

4.1 The *stress* conundrum

In the descriptive and theoretical literature on West Germanic languages, and in particular the very thoroughly investigated Dutch, English and German languages, the notion of stress applies to a complex cluster notion, involving at least the following six different dimensions of linguistic structure:

1. acoustic and auditory prominence
2. phonotactic structure, phonological alternations
3. metrical structure (foot structure)
4. lexical structure
5. function in text-tune alignment (intonational anchoring)
6. function in conveying information-structural distinctions

These dimensions themselves each involve a number of different subdimensions, not all of which can be discussed here. The following only provides a few terse remarks on what each dimension pertains to and which problems it poses to cross-linguistic comparison. The major goal here is to make clear that stress as found in West Germanic languages is a multi-dimensional cluster concept, with some dimensions themselves being cluster concepts (e.g. acoustic prominence). West Germanic stress is thus not a suitable object for cross-linguistic comparison simply because it is a complex composite notion which involves a large number of correlations that are unlikely to be widely attested across the languages of the world.

Acoustic and auditory prominence refers to the fact that stressed syllables are heard as being more prominent in comparison to adjacent syllables. Acoustically, a range of correlates are considered to contribute to the prominence of stressed syllables, including in particular duration, various measures of intensity and F0. However, as recent surveys by Gordon & Röttger (2017) and Röttger & Gordon (2017) make clear, there are considerable differences in how these acoustic parameters are actually measured and evaluated in studies investigating stress in individual languages. Apart from the lack of an agreed-upon set of properties that need to be measured and the significant differences in the way a given correlate is measured,

there is one issue in particular that renders available acoustic stress analyses incommensurable: many studies fail to properly separate effects of sentence-level (postlexical) prominence (also known as ‘sentence stress’) from effects of word-level prominences (see also Gordon 2014 for the probably most thorough discussion of the issues involved here to date). That is, acoustic properties are, in principle, well-suited to provide the basis for cross-linguistic comparison of putatively stressed syllables, but considerable care has to be taken to ensure that the measures used are indeed comparable, as further detailed in section 4.2.

As for the auditory prominence of stressed syllables, most statements on this topic in the literature are impressionistic. That is, the analysts hear one syllable in a word as more prominent. But a number of studies has made amply clear that hearers are strongly influenced by their native language(s) with regard to hearing prominence. Peperkamp and Dupoux (2002) and Peperkamp et al. (2010), for example, show that speakers of languages where word-level prominences do not exist or have no major functional role tend to be ‘stress-deaf’. Furthermore, Tabain et al. (2014: 64) note:

It is possible that the expectation that stress will be regular and foot-based (i.e. rhythmic) has a cognitive basis in the case of field linguists who are native speakers of prototypical stress languages, such as English or German. Because of the largely regular and rhythmic structure of lexical stress in their languages, field linguists may expect to hear a similarly regular and rhythmic structure in the languages they are studying – and they hear this stress whether the acoustic cues are present or not. This would then be a case of stress ghosting comparable to stress deafness for speakers whose native language does not have lexical stress.

Inasmuch as these observations are taken seriously,⁵ proper assessment of stress as a perceptual phenomenon needs proper perceptual testing, which is rarely done. Where it is done, the results are often not straightforward to interpret, as a range of studies on the putatively stressless varieties of Malay in Indonesia show.⁶ Consequently, more work needs to be done to clarify how stressed syllables as perceptual units can be productively compared across languages, an issue also taken up in section 4.2.

Regarding the phonotactic dimension, in West Germanic languages stressed syllables allow for larger inventories of segments than unstressed syllables. Furthermore, stress shifts often correlate with segmental reductions in the formerly stressed syllable (for example, the vowel becoming schwa or at least more centralized). The question of whether the syllables in a given language differ systematically with regard to the inventory of segments they allow for is a question that appears to be relatively easy to address in cross-linguistic perspective. It is not clear, however, that differences with regard to the inventory of admissible segments systematically correlate with the other properties of stress mentioned above. Hyman (2018), for one, makes a strong argument for clearly separating what he calls positional prominence from other kinds of word-level prosodic prominence.

⁵ Mühlbauer (2006) discusses another interesting case where three different authors, one of them being Leonard Bloomfield, hear stress differently in Plains Cree. Many thanks to one of the reviewers for bringing this paper to my attention.

⁶ Cp. Odé (1994) and the literature building on this study, conveniently summarized in van Heuven & van Zanten (2007). Later work includes van Heuven & Faust (2009), Riesberg et al. (2020), and Kaland (2020).

Turning to the third dimension mentioned above, i.e. metrical structure, it is fairly clear that the notion of secondary (word-level) stress and rhythm that is at the center of this structure is not linked to primary stress in the same way in all languages (cp. González 2008, van der Hulst 2012; Goedemans & van der Hulst 2014). What is more, cross-linguistic comparison in this dimension is severely hampered by the lack of a cross-linguistically applicable definition of ‘foot’. In many works, the existence of feet is simply presupposed rather than established empirically. Özçelik (2017) provides an instructive illustration of the problems arising here for cross-linguistic comparison with data from Turkish, French and English child language.

The fourth dimension in the above list of stress properties concerns the role of stress in the organization of the (mental) lexicon and more specifically, its role in spoken-word recognition. As work by Cutler and colleagues (e.g. Cutler 2005, 2012; cp. also Kaland et al 2021 for a recent investigation of a non-European language) shows, stress may facilitate spoken-word recognition in a number of different ways, with classic minimal stress pairs such as English *'permit* vs *per'mit* playing a rather marginal role. What appears to be more relevant, in some languages at least, is the fact that stress may be of use in disambiguating a word form like *music* from a word form like *museum* in that a recognizably stressed first syllable leads to the exclusion of the candidate form *museum* early on in incremental word recognition (Cutler 2005: 276-279).

With regard to the fifth dimension, stressed syllables in West Germanic languages have a special role to play in associating the segmental chain with intonational targets. In intonational phonology (Gussenhoven 2004, Ladd 2008), a major distinction is made between intonational targets that are associated with the edges of prosodic constituents such as intonation units, and those that are associated with metrically prominent syllables, regardless of their position in a phrase. The former are known as edge (or boundary) tones, the latter as (postlexical) pitch accents. Cross-linguistic comparison in this domain is very complex as it presupposes reasonably comprehensive analyses of the intonational systems of the languages involved. Most importantly, it requires a cross-linguistically applicable definition of what it means for an intonational target *not* to be associated with the edge of a prosodic unit, but rather with a metrically prominent syllable.

There is, however, increasing evidence for the view that many languages do not make use of (postlexical) pitch accents and thus are edge-marking only languages. Some of these languages have been argued to lack word-level prominences altogether. Examples include French (Gussenhoven 2004, Féry 2017), Ambon Malay (Maskikit-Essed & Gussenhoven 2016), and Totoli (Bracks 2021). Others have been argued to “ignore stress” in aligning their intonational targets, as claimed for the Papuan language Kuot by Lindström & Remijsen (2005) and for Hindi by Féry (2017). In short, the important role that stress plays in text-tune association in West Germanic and many other European languages does not appear to be found universally.

Finally, the role of stress in conveying information-structural distinctions is closely related to the immediately preceding point. The pitch accent that is associated with a stressed syllable may convey two meanings. On the one hand, the accent may signal that the word highlighted by it (or the larger constituent the word is a part of) is the focus constituent of the clause. On the other hand, it may also indicate the information status of the referent of a noun phrase bearing a pitch accent, with expressions referring to more highly activated referents being less prominently marked than ones which newly activate a participant (cp. Baumann 2006). So far,

a similar combination of two separate information structural functions, i.e. focus marking and activation status marking, has, to the best of my knowledge, not been reported for languages outside of Europe. While the functions of pitch accents are, of course, a matter of intonation and not of stress, these functions are indirectly linked to stress as the association to a stressed syllable is definitional for intonational pitch accents.

This concludes our brief overview of (some of) the complexities involved in each of the six dimensions of linguistic structure that are associated in one way or the other with stress in West Germanic languages. It should be clear that this language-specific category needs to be taken apart and individual component properties have to be carefully assessed as to whether or not they may serve as a useful basis for cross-linguistic comparison. Major problems for cross-linguistic comparison arise when phenomena from different dimensions are arbitrarily chosen and included in a comparison set of examples for stress. That is, for language X, phonotactic constraints are counted as evidence for stressed syllables, while in language Y it is the association with intonational pitch accents. Such a procedure will not provide the basis for a sound and productive comparison.

This is a major problem regarding all works concerned with the cross-linguistic comparison of stress that have appeared to date (e.g. Hayes 1995, Van der Hulst et al. 2010, Gordon 2016, chapter 7). The sources these works are based on are often not very clear as to how they define stress. To be sure, more often than not auditory prominence of some sort will be part of the evidence used to identify stressed syllables, but what else has played a role is rarely commented on. It thus remains unclear to what extent the language-specific categories compared under this label really have enough in common to warrant cross-linguistic comparison.

In this regard, it will be instructive to note that Rob Goedemans, who has been the driving force in compiling the most comprehensive database for stress systems available to date (i.e. *StressTyp2*),⁷ has voiced concerns regarding the validity of some of the stress systems that are included in the database. In Goedemans & van Zanten (2014), the authors list the following four heuristics that render reports on stress suspect to their minds:⁸

- a. Stress is reported to vary in different utterances of the same word
- b. Stress is reported to be a phenomenon related to phrases
- c. Stress is reported to be very weak and unstable
- d. Fundamental differences in stress use of various speakers reported

(Goedemans & van Zanten 2014: 88)

Recall also the possibility raised in the quote by Tabain et al. above that researchers ‘hear’ their own stress systems into the language(s) they work. Both Goedemans & van Zanten’s as

⁷ From the website for *StressTyp2* (<http://st2.uliet.net/?content=page&id=14>): “... *StressTyp2* (ST2) merges the information from the *StressTyp* (ST) database, constructed de novo by Harry van der Hulst and Rob Goedemans in the early 1990s and continually updated since (Goedemans et al. 1996; Goedemans and van der Hulst 2009, 2010), and the *Stress Pattern Database* (SPD), constructed by Jeffrey Heinz (Heinz 2007) which is based on the *Stress System Database* (SSD), developed by Todd Bailey (Bailey 1995) as well as additional materials, notably data collections compiled by Larry Hyman (Hyman 1977) and Matt Gordon (Gordon 2002). *StressTyp2* contains information on over 750 distinct languages, with nearly every language family represented.”

⁸ The authors, in fact, list a fifth heuristic, i.e. “[t]he reported stress rule makes no sense in any current metrical theory” (Goedemans & van Zanten 2014: 88), which is of a somewhat different kind. The four other heuristics relate to reports casting doubt on the analysis which are found in the sources themselves. The fifth heuristic involves a theoretical evaluation by the compilers of the data themselves, which by itself will typically engender controversy.

well as Tabain et al.’s observations strongly suggest that reports of linguistically salient acoustic prominence differences not accompanied by proper acoustic and perceptual experimentation are ill-founded.

The identification of the major component parts of the West Germanic cluster category *stress* is only the first step in preparing a proper cross-linguistic investigation of this category. As already hinted at above, the next step is to scrutinize each component part in more detail with regard to properties that can be reliably identified across different languages. In this regard, it should be obvious that such an investigation is particularly important for the first two component parts, i.e. acoustic and auditory prominence or phonotactic structure (positional prominence). The other four component parts actually presuppose that stressed syllables can be identified.

Unfortunately (for the typologist), the identification of the acoustic, auditory or phonotactic properties that render syllables prominent is also not a straightforward task, as will be shown in the next section, using acoustic and auditory prominence as its main example.

4.2 On identifying acoustic and auditory prominence across languages

In a recent paper, van Heuven (2018) provides a blueprint for the investigations necessary to determine whether a given language has stress in the sense of an acoustically and perceptually prominent syllable. There are three essential assumptions that characterize this program.

1. Word-level prominence has to be strictly separated from phrase-level prominence, as exemplified in the classic work by Sluijter & van Heuven (1996) and thoroughly discussed in Gordon (2014). When this is properly done, it is very likely that fundamental frequency does not play a major role in cueing (word-level) stress (van Heuven 2018: 31).⁹
2. Word-level prominence should, in principle, be detectable on every phonological word (the classic obligatoriness criterion, cp. Hyman 2006: 231). This assumption is not without problems, especially with regard to the level of detail that can be achieved in its empirical validation. But, to my mind, there are sensible standard procedures that are sufficiently strong to support the claim that in the large majority of phonological words in a particular language one syllable is more prominent than other syllables. Such a claim could, for example, be supported by showing that for a sample of say 200 lexemes, each with a minimum of 8 attestations in a corpus of spontaneous speech, the same syllable in each token is consistently longer by at least 30% than neighboring syllables (cp. Kaland 2019 for an investigation along these lines).
3. Both acoustic and perceptual experiments need to be carried out. That is, it is not enough to show that there are systematic acoustic differences between stressed and unstressed syllables, but it needs to be shown that native listeners are able to discern these differences, regardless of whether or not much use is made of the distinction in the phonological system.

⁹ As this observation is rarely found in the literature, it will be useful to quote it in full: “Typically, as long as there is no sentence stress on a word, the speaker makes no effort to change the vocal pitch. To be true, there may well be a small rise-fall contour on any vowel (with or without word stress) but this is due to an involuntary response of the glottal mechanism to the greater transglottal pressure that comes about when the oral tract opens during the articulation of the vowel sound; during the articulation of consonants the oral tract is fully or partially closed so that intraoral impedance yields a transglottal pressure drop causing the vocal folds to vibrate more slowly. It has been estimated that the involuntary effect of mouth opening on the rate of vocal fold vibration does not normally exceed a threshold of four semitones (a frequency rise and subsequent fall of less than 25%).”

This is just a rough outline of a program for collecting the empirical evidence needed for a proper comparison of language-specific stress categories across languages. Several issues need to be further scrutinized and worked out in detail before it can actually be put into practice. One issue, for example, is the question to what extent time-consuming experimental investigations are actually feasible as a prerequisite for cross-linguistic comparison and whether there are ways to expediate the data collection process. Thus, one may ask whether it would be sufficient to focus the analysis on duration rather than attempting to test the full range of potential stress correlates, which would include a number of complex intensity measures (e.g., spectral balance, spectral expansion; cp. van Heuven 2018: 24-30). Such a focus on duration could be supported by the observation that, once the distinction between the word and the phrase levels are properly controlled for, duration has been found to be the major correlate of stress, as noted also by Gordon & Röttger (2017: 8), *inter alia*:

Statistically, duration was the most reliable exponent of stress across languages, although all of the measured parameters succeeded in differentiating stress in the majority of languages for which they were assessed.

That the investigations required here are practically feasible in principle is shown by the series of studies Kaland has recently undertaken for Papuan Malay, a Malayic trade variety spoken in eastern Indonesia. Kaland (2019) provides a careful investigation of acoustic parameters, finding duration to be the major correlate of stress, but vowel quality and spectral tilt also play a role. Kaland (2020) reports a number of perception experiments which support the claim that speakers actually discern the difference between stressed and unstressed syllables. Finally, Kaland et al. (2021) show that stress distinctions may have a (minor) role to play in word recognition, following up on the leads in the processing literature, in particular Cutler (2012).¹⁰

Identifying systematic and salient prominence distinctions is of course only a first step in a comprehensive typology of stress systems. Once a number of language-specific manifestations of stress has been assembled on the basis of proper acoustic and perceptual experimentation, one can then ask what the range of variation with regard to acoustic cues is and whether it holds true that duration is the most stable cue for stress, the other cues having more a kind of enhancing function. Furthermore, one can explore which other cross-linguistically identifiable properties often or perhaps even always correlate with stress defined in this way. Does it play a role in word recognition? How does it relate to foot structure? And so on. Last but not least, regarding the hypothesis about shared diachronic trajectories proposed in the preceding section, one may ask whether it is possible to distinguish different degrees to which stress has been activated in a particular prosodic system and what kind of evidence could be used to support claims to this regard.

4.3 One further complication: stress vs. accent

There is one major additional issue that needs to be mentioned in order to provide a reasonably comprehensive idea of what is involved in proper stress typology. The preceding remarks only concern the kind of highlighting of a syllable well-known from Germanic and Romance languages where the highlighting is due to the “greater physiological effort” (van Heuven 2018: 17) involved in the production of stressed syllables. This kind of highlighting is to be distinguished from highlighting which is purely pitch-based. Inasmuch as highlighting

¹⁰ The Praat scripts used in these investigations are available at <https://doi.org/10.17605/OSF.IO/CR4VN>.

by pitch also pertains to the word level, the question arises whether it should also be included in cross-linguistic comparisons investigating stress. Beckman (1986), building on a longer tradition, proposes to use *accent* as the superordinate term for word-level prominence regardless of its phonetic substance, with the two major subtypes *stress-accent* and *pitch-accent* (see Beckman & Venditti 2010 for a more recent statement of this proposal). This usage is found in a large segment of the literature as documented in an extended review by van der Hulst (2010). Hyman (2006, 2009, 2012) strongly opposes this usage and argues that stress and tone (pitch) should be kept strictly separate, so-called lexical pitch accent systems being but a reduced type of tonal system. Kuznetsova (2018) argues that a simple opposition of stress-accent vs. pitch-accent is insufficient and that a more fine-grained typological grid is needed to provide a proper space, for example, for the Danish and Estonian prosodic systems which do not easily fit stress-accent vs. pitch-accent opposition.

At this point, I do not have anything of relevance to offer with regard to this debate except that it is my impression that here as well further language-specific investigations are needed before the competing proposals can be properly evaluated. Excepting Japanese, which is Beckman's primary example for a (lexical) pitch-accent language, and possibly Basque, both the phonetics and the phonology of (lexical) pitch accents in linguistic areas such as Papua New Guinea, for example, are still in need of considerable further empirical scrutiny. One issue that has arisen in ongoing work on Papuan highland languages in West Papua is the question whether so-called lexical pitch accents are really lexical. Alternatively, they are better analyzed as post-lexical boundary tones delimiting phonological words in a similar way as the higher level phonological and intonational phrases are analyzed as being delimited by boundary tones.

5. Conclusion

The main purpose of this contribution was to show that and how cross-linguistic comparison is very well possible also in the case of complex cluster categories such as *stress*. More specifically, the cross-linguistic comparison of cluster categories does not require a special framework and methodology which makes use of a set of concepts exclusively defined for the purpose of language comparison, as proposed by Haspelmath (2018). Instead, cross-linguistic comparison can make use of the cross-linguistically identifiable properties of language-specific categories. Such properties typically relate to the phonetic or semantic substance of a category, the methods and theoretical assumptions used to establish it, or the fact that it shares the trajectory of its diachronic evolution with similar categories in other languages.

The fact that *stress* is a highly complex cluster concept has not been explicitly and widely recognized in the typological community. Therefore, a second major goal of this contribution was to show that the phenomenon called *stress* in the analysis of West Germanic languages, which implicitly underlies practically all work on stress, manifests itself in at least six different dimensions of linguistics structure, each of which involving its own complexities. Proper cross-linguistic comparison therefore needs to scrutinize these dimensions carefully and identify those properties that can be identified cross-linguistically and thus make the cross-linguistic comparison of language-specific categories possible (cp. also Round & Corbett 2020). In the case of stress, acoustic and auditory prominence is one promising starting point for further typological inquiry, but the identification and comparison of acoustic and auditory prominence distinctions cross-linguistically is also not straightforward and requires considerable empirical effort (see section section 4.2).

The category *stress* thus provides an example for categories where substantial further empirical work is required before proper progress can be made in cross-linguistic comparison. While stress may be particularly demanding in this regard, it is probably true for many current and future typological projects that the empirical and methodological workloads and benchmarks have become significantly higher than they used to be. This includes the need to control for areal and genetic sampling biases, the use of state-of-the-art statistic techniques, as well as the significantly increased amount of descriptive and documentary materials available today. The times when it was possible to make important cross-linguistic observations on the basis of a convenience sample of a few dozen languages, using loosely defined semantic categories as *tertia comparationis* as in Greenberg (1963/66) have definitely passed.

ACKNOWLEDGEMENTS

A first presentation of the core claims of this article took place at *CoEDL Fest 2017* in Brisbane (*Centre of Excellence for the Dynamics of Language*). It has also been presented in evolving versions at the Universität zu Köln and the Universität Zürich. I am very grateful for the comments received on these occasions. Many thanks to Constantijn Kaland for reading the first draft and making numerous useful suggestions. I gratefully acknowledge the superb feedback provided by the reviewers for LINGUISTIC TYPOLOGY and the editors of the volume. Their very detailed and extensive comments helped me to clarify important aspects of the ideas presented here and considerably improved their presentation. Many thanks also to Janet Bachmann, Jonas Heinen and Jonas Frings for checking references.

Writing has been supported by the German Research Foundation (DFG) through the Collaborative Research Centre 1252 *Prominence in Language* (Project-ID 281511265, Project A03 *Prosodic prominence in cross-linguistic perspective*) and the *Leo Spitzer Prize* of the Universität zu Köln.

References

- Bailey, Todd M. 1995. *Nonmetrical constraints on stress*. Ann Arbor: University of Minnesota dissertation.
- Baumann, Stefan. 2006. *The intonation of givenness: evidence from German*. Tübingen: Niemeyer.
- Beck, David. 2016. Some language-particular terms are comparative concepts. *Linguistic Typology* 20. 395–402.
- Becker, Laura. 2021. *Articles in the world's languages* (Linguistische Arbeiten 577). Berlin & Boston: De Gruyter.
- Beckman, Mary E. 1986. *Stress and non-stress accent*. Dordrecht: Foris.
- Beckman, Mary E. & Jennifer J. Venditti. 2010. Tone and intonation. In William J. Hardcastle & John Laver (eds.), *The handbook of phonetic sciences*, 603–652. Oxford: Blackwell.
- Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.
- Bracks, Christoph. 2021. The intonation unit in Totoli. *Oceanic Linguistics* 60. 103–132.
- Browman, Catherine P. & Louis Goldstein, L. 1989. Articulatory gestures as phonological units. *Phonology* 6. 201–251.

- Cutler, Anne. 2005. Lexical stress. In David B. Pisoni & Robert E. Remez (eds.), *The handbook of speech perception*, 264–289. Malden: Blackwell.
- Cutler, Anne. 2012. *Native listening: Language experience and the recognition of spoken words*. Cambridge: MIT Press.
- Dahl, Östen. 2016. Thoughts on language-specific and crosslinguistic entities. *Linguistic Typology* 20. 427–437.
- Goedemans, Rob W.N. & Ellen van Zanten. 2014. No stress typology. In Johanneke Caspers, Yiya Chen, Willemijn F.L. Heeren, Jos J.A. Pacilly, Niels O. Schiller & Ellen van Zanten (eds.), *Above and beyond the segments: Experimental linguistics and phonetics*, 83–95. Amsterdam: Benjamins.
- Goedemans, Rob W.N. & Harry van der Hulst. 2009. *StressTyp: A database for word accentual patterns in the world's languages*. In Martin Everaert, Simon Musgrave & Alexis Dimitriadis (eds.), *The use of databases in cross-linguistic studies* (Empirical Approaches to Language Typology 41), 235–282. Berlin & New York: De Gruyter Mouton.
- Goedemans, Rob W.N. & Harry van der Hulst. 2010. *StressTyp data*. In: Harry van der Hulst, Rob Goedemans and Ellen van Zanten (eds.), *A survey of word accentual patterns in the languages of the world*, 669–846. Berlin & New York: De Gruyter Mouton.
- Goedemans, Rob W.N. & Harry van der Hulst. 2014. The separation of accent and rhythm: Evidence from StressTyp. In Harry van der Hulst (ed.), *Word Stress: Theoretical and typological issues*, 119–148. Cambridge: Cambridge University Press.
- Goedemans, Rob W.N., Harry van der Hulst & Ellis A.M. Visch. 1996. *Stress Patterns of the World. Part 1: Background* (HIL Publications 2). The Hague: Holland Academic Graphics.
- González, Carolina. 2008. Typological evidence for the separation between stress and foot structure. In Matti Miestamo & Bernhard Wälchli (eds.), *New challenges in typology* (Trends in Linguistics. Studies and Monographs 189), 55–76. Berlin & New York: De Gruyter Mouton.
- Gordon, Matthew K. 2002. A factorial typology of quantity insensitive stress. *Natural Language and Linguistic Theory* 20. 491–552.
- Gordon, Matthew K. 2014. Disentangling stress and pitch-accent: A typology of prominence at different prosodic levels. In Harry van der Hulst (ed.), *Word stress*, 83–118. Cambridge: Cambridge University Press.
- Gordon, Matthew K. 2016. *Phonological typology*. Oxford: Oxford University Press.
- Gordon, Matthew K. & Timo Roettger. 2017. Acoustic correlates of word stress: A cross-linguistic survey. *Linguistics Vanguard* 3(1).
- Greenberg, Joseph H. 1963/66. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Language*, 73–113. Cambridge: MIT Press.
- Greenberg, Joseph H. 1978. How does a language acquire gender markers? In Joseph H. Greenberg, Charles A. Ferguson & Edith Moravcsik (eds.), *Universals of Human Language. Vol. 3: Word Structure*, 47–82. Stanford: Stanford University Press.
- Gussenhoven, Carlos. 2004. *The phonology of tone and intonation*. Cambridge: Cambridge University Press.
- Harder, Peter. 2016. Substance(s) and the rise and imposition of structure(s). *Acta Linguistica Hafniensia* 48. 7–34.

- Haspelmath, Martin. 2018. How comparative concepts and descriptive linguistic categories are different. In Daniël Van Olmen, Tanja Mortelmans & Frank Brisard (eds.), *Aspects of linguistic variation* (Trends in Linguistics. Studies and Monographs 324), 83–114. Berlin & Boston: De Gruyter Mouton.
- Hayes, Bruce. 1995. *Metrical stress theory*. Chicago: The University of Chicago Press.
- Heinz, Jeffrey. 2007. *The inductive learning of phonotactic patterns*. Los Angeles: University of California dissertation.
- Himmelman, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Tübingen: Niemeyer.
- Himmelman, Nikolaus P. 2001. Articles. In Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals. Vol. 1* (Handbooks of Linguistics and Communication Science 20/1), 831–841. Berlin & New York: De Gruyter.
- Himmelman, Nikolaus P. 2022. Against trivializing language description (and comparison). *Studies in Language* 46(1).
- Himmelman, Nikolaus P., Meytal Sandler, Jan Strunk & Volker Unterladstetter. 2018. On the universality of intonational phrases in spontaneous speech – a cross-linguistic interrater study. *Phonology* 35. 207–245.
- Hyman, Larry M. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed.), *Studies in stress and accent* (Southern California Occasional Papers in Linguistics 4), 37–82. Los Angeles: Dept. of Linguistics, University of Southern California.
- Hyman, Larry M. 1983. Are there syllables in Gokana? In Jonathan Kaye, Hilda Koopman, Dominique Sportiche, & André Dugas (eds.), *Current approaches to African linguistics. Vol. 2*, 171–179. Dordrecht: Foris.
- Hyman, Larry M. 2006. Word-prosodic typology. *Phonology* 23(2). 225–257.
- Hyman, Larry M. 2008. Universals in phonology. *The Linguistic Review* 25(1–2). 83–137.
- Hyman, Larry M. 2009. How (not) to do phonological typology: the case of pitch-accent. *Language Sciences* 31. 213–238.
- Hyman, Larry M. 2011. Does Gokana really have no syllables? Or: what’s so great about being universal? *Phonology* 28(1). 55–85.
- Hyman, Larry M. 2012. In defense of prosodic typology: A response to Beckman & Venditti. *Linguistic Typology* 16. 341–385.
- Hyman, Larry M. 2015. Does Gokana really have syllables? A postscript. *Phonology* 32. 303–306.
- Hyman, Larry M. 2017. What (else) depends on phonology? In Nicholas Enfield (ed.), *Dependencies in language*, 141–58. Berlin: Language Science Press.
- Hyman, Larry M. 2018. Positional prominence versus word accent: Is there a difference? In Robert Goedemans, Jeffrey Heinz & Harry van der Hulst (eds.), *The study of word stress and accent: Theories, methods and data*, 60–75. Cambridge: Cambridge University Press.
- Hyman, Larry M. & Frans Plank. 2018. *Phonological Typology* (Phonology and Phonetics 23). Berlin & Boston: De Gruyter Mouton.
- Kaland, Constantijn. 2019. Acoustic correlates of word stress in Papuan Malay. *Journal of Phonetics* 74. 55–74.

- Kaland, Constantijn. 2020. [Offline and online processing of acoustic cues to word stress in Papuan Malay](#). *The Journal of the Acoustical Society of America* 147(2). 731–747.
- Kaland, Constantijn, Angela Kluge & Vincent J. van Heuven. 2021. [Lexical analyses of the function and phonology of Papuan Malay word stress](#). *Phonetica* 78(2). 141–168.
- Kenesei, István, Robert M. Vago & Anna Fenyvesi. 1998. *Hungarian*. London & New York: Routledge.
- Kuznetsova, Natalia. 2018. What Danish and Estonian can show to a modern word-prosodic typology. In Rob Goedemans, Jeffrey Heinz, and Harry van der Hulst (eds.), *The study of word stress and accent: Theories, methods and data*, 102–43. Cambridge: Cambridge University Press.
- Ladd, D. Robert. 2008. *Intonational phonology*, 2nd edn. Cambridge: Cambridge University Press.
- Ladefoged, Peter. 1982. *A course in phonetics*, 2nd edn. Los Angeles: University of California.
- Lander, Yury & Peter Arkadiev. 2016. On the right of being a comparative concept. *Linguistic Typology* 20. 403–416.
- Lehmann, Christian. 2018. Linguistic concepts and categories in language description and comparison. In Marina Chini & Pierluigi Cuzzolin (eds.), *Typology, acquisition, grammaticalization studies*, 27–50. Milano: Franco Angeli.
- Levi, Susannah V. 2005. Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association* 35. 73–97.
- Li, Charles N. (ed.). 1976. *Subject and topic*. New York: Academic Press.
- Lindström, Eva & Bert Remijsen. 2005. Aspects of the prosody of Kuot, a language where intonation ignores stress. *Linguistics* 43. 839–870.
- Maskikit-Essed, Raechel & Gussenhoven, Carlos. 2016. No stress, no pitch accent, no prosodic focus: The case of Ambonese Malay. *Phonology* 33. 353–389.
- Moravcsik, Edith A. 2016. On linguistic categories. *Linguistic Typology* 20. 417–425.
- Mühlbauer, Jeff. 2006. Pitch as accent in Plains Cree nominals. In H.C. Wolfart (ed.), *Actes du 37e Congrès des Algonquinistes*, 229–268. Winnipeg: Université du Manitoba.
- Odé, Cecilia. 1994. On the perception of prominence in Indonesian. In Cecilia Odé & Vincent J. van Heuven (eds.), *Experimental studies of Indonesian prosody* (Semaian 9), 27–107. Leiden: Vakgroep Talen en Culturen van Zuidoost-Azië en Oceanië, Leiden University.
- Özçelik, Öner. 2017. The foot is not an obligatory constituent of the prosodic hierarchy: “stress” in Turkish, French and child English. *The Linguistic Review* 34(1). 157–213.
- Peperkamp, Sharon & Emmanuel Dupoux. 2002. A typological study of stress ‘deafness’. In Carlos Gussenhoven & Natasha Warner (eds.), *Laboratory phonology 7* (Phonology and Phonetics 4–1), 203–240. Berlin & New York: De Gruyter Mouton.
- Peperkamp, Sharon, Inga Vendelin & Emmanuel Dupoux. 2010. Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics* 38(3). 422–430.
- Riesberg, Sonja, Janina Kalbertodt, Stefan Baumann & Nikolaus P. Himmelmann. 2020. Using Rapid Prosody Transcription to probe little-known prosodic systems: The case of Papuan Malay. *Laboratory Phonology* 11(1). 1–35.
- Roettger, Timo & Matthew K. Gordon. 2017. Methodological issues in the study of word stress correlates. *Linguistics Vanguard* 3(1).

- Round, Erich R. & Greville G. Corbett. 2020. Comparability and measurement in typological science: The bright future for linguistics. *Linguistic Typology* 24. 489–525.
- Sluijter, Agaath M. C. & Vincent J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America* 100. 2471–2485.
- Spike, Matthew. 2020. Fifty shades of grue: Indeterminate categories and induction in and out of the language sciences. *Linguistic Typology* 24. 465–488.
- Tabain, Marija, Janet Fletcher & Andrew Butcher. 2014. Lexical stress in Pitjatjantjara. *Journal of Phonetics* 42. 52–66.
- van der Auwera, Johan & Kalyanamalini Sahoo. 2015. On comparative concepts and descriptive categories, such as they are. *Acta Linguistica Hafniensia* 47. 136–173.
- van der Hulst, Harry. 2010. Word accent: Terms, typologies and theories. In: Harry van der Hulst, Rob Goedemans and Ellen van Zanten (eds.), *A survey of word accentual patterns in the languages of the world*, 3–54. Berlin & New York: De Gruyter Mouton.
- van der Hulst, Harry. 2012. Deconstructing stress. *Lingua* 122. 1494–1521.
- van der Hulst, Harry. 2017. Phonological typology. In Alexandra Y. Aikhenvald & Robert M.W. Dixon (eds.), *The Cambridge handbook of linguistic typology*, 39–77. Cambridge: Cambridge University Press.
- van der Hulst, Harry, Rob W.N. Goedemans & Ellen van Zanten (eds.). 2010. *A survey of word accentual patterns in the languages of the world*. Berlin & New York: De Gruyter Mouton.
- van Heuven, Vincent J. & Ellen van Zanten (eds.). 2007. *Prosody in Indonesian languages* (LOT Occasional Series 9). Utrecht: LOT.
- van Heuven, Vincent J. & Vera Faust. 2009. Are Indonesians sensitive to contrastive accentuation below the word level? *Wacana, Jurnal Ilmu Pengetahuan Budaya* 11. 226–240.
- van Heuven, Vincent J. 2018. Acoustic correlates and perceptual cues of word and sentence stress: Towards a cross-linguistic perspective. In Rob Goedemans, Jeffrey Heinz & Harry van der Hulst (eds.), *The study of word stress and accent: Theories, methods and data*, 15–59. Cambridge: Cambridge University Press.
- Xu, Yi. 2020. *Syllable is a synchronization mechanism that makes human speech possible*. PsyArXiv. <https://doi.org/10.31234/osf.io/9v4hr> (accessed 21 Jun 2022).