Parameters in phonological analysis: stress

Van Oostendorp, M.

published in
Contemporary Linguistic Parameters
2015

document version
Early version, also known as pre-print

Parametric and nonparametric approaches in the analysis of stress typologies

Marc van Oostendorp

July 22, 2014

1 Introduction

Phonological stress might be one of the most successful testing beds for any theory of linguistic variation and language acquisition. One reason is that stress systems — on which syllables in a word do languages put primary and secondary stress — seem relatively easy to isolate from other properties of language. We can approximate the stress pattern of a word as a string of numbers (e.g. 2001 is a four-syllable word with primary stress on the last syllable and secondary stres on the first), and thus view an individual language as a set of such strings, and the grammar as an intensional description of such sets. It seems fairly easy to write grammars of such sets of strings; and then to generalize over such grammars.

A second, related, reason is that we have a relatively good overview of stress systems of the worlds languages, for instance, in the stresstyp database[1] created by Harry van der Hulst and Rob Goedemans. Also the *World Atlas of Linguistic Structures*[2] features several chapters on stress.

And thirdly, the parametric space seems relatively simple; it has turned out to be possible to determine a relatively small number of parameters which in interaction can describe virtually the whole space.

It was probably for this reason that some of the first relatively complete, and relatively fruitful applications of Chomsky (1981)'s idea of parameters were found in the phonological domain: Halle and Vergnaud (1987) and Dresher and Kaye (1990). The first offered a way of showing how a vast number of languages could be described by using only a small set of parameters. The second furthermore showed that it was possible to implement these parameters in a

[1] `http://www.unileiden.net/stresstyp/`

[2] `http://wals.info/feature`

computer programme that could then simulate the acquisition of a language based on these parameters.

Around the same time, however, the parametric model suffered two important blows in metrical theory, one empirical and one theoretical. The empirical blow was that a serious gap in the predicted typology was shown to exist: some of the predicted languages are very rare, or maybe even absent from the database Hayes (1987); McCarthy and Prince (1986). This led to an approach in which a universal foot inventory took over the central role from parameters; languages would pick their feet from this non-binary inventory.

The theoretical blow came from Optimality Theory (OT Prince and Smolensky, 1993), which replaced the view on binary parameters switched 'on' and 'off' with the idea of universal, but conflicting constraints that would get a different ranking in different languages. Most of the typological as well as most of the acquisition modeling work within formal phonology since then has been set within OT. Parameters formally survived in some non-mainstream models such as Government Phonology; but between 1995 and 2005 constraint ranking seemed pervasive in the stress theory - also because GP practicioners for some reason did not seem very interested in stress. Meanwhile, OT evolved, as did some of the dominant ideas on what the proper theoretical account of language variation is. One important direction this evolution took was in the direction of more gradient distinctions between languages and small, stochastic influences on the location of stress within individual languages. This seems now also to slowly be applied to stress.

This chapter aims to describe the developments towards and away from parametric models in some more detail. What motivated the various moves? What have we learned about the structure of language variation from this discussion? And where are we now? I follow the course of history which I just sketched. It is *not* my aim to provide a complete picture of the course of history; I merely want to sketch the main arguments which played a role in the discussion, focusing in particular on the theoretical aspects of the discussion, and what this can teach us about language variation, not just for phonology but from a more general linguistic perspective. My main goal is to explain the relevance of the phonological discussion also for syntacticians and morphologists interested in the formal study of linguistic variation.

2 Parametrizing stress

The idea that language variation may be reduced to a set of binary parameters, first proposed in 1979 in Chomsky's so-called *Pisa Lectures* (and published as Chomsky (1981)) caught on almost immediately in phonology, and in particular in metrical phonolology. Conceptually, this is not very strange, since many of the choices that languages make in this typological space are, obviously, binary. There are two types of (binary) feet: iambs and trochees;

languages can be quantity sensitive of not; primary stress can be assigned at the left edge of the word, or at its right edge; etc.

Very important in this respect was Hayes (1980)' dissertation, published as Hayes (1985), which proposed that the stress system of languages can be described using parameters such as the following, on binary foot types:

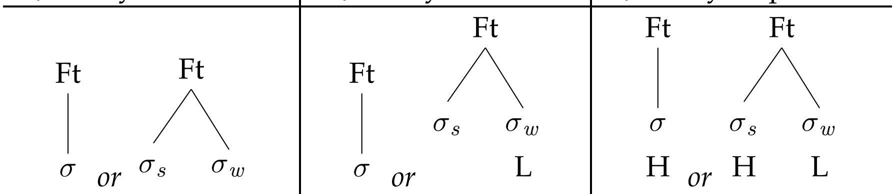
(1) a. **Dominance** [Right/Left] (The head of the foot is on the left / on the right)
b. **Quantity-sensitive** [Yes/No] (Does the language distinguish between 'light' and 'heavy' syllables?)
c. **Obligatory Branching** [Yes/No] (Should the head be heavy?)
d. **Labeling Based On Branching** [Yes/No] (does headedness switch from a light syllable?)

Taken together, these parameters cover quite some typological ground. The first distinguishes between languages with left-headed feet (−∪) and those with right-headed feet (∪−).

The other three parameters are related to each other in an interesting way. The Quantity-sentitivity Parameter distinguishes those languages the stress of an individual syllable is only dependent on its numerical position in the string (quantity-insensitive languages) and those in which the internal structure — the difference between 'heavy' and 'light' syllables matters – does play a role (quantity-sensitive languages in which heavy syllables are disallowed from weak postions). The other two parameters can be seen as dependent on the first: they only apply to quantity sensitive-languages. Languages in which Obligatory Branching is switched on demand that strong positions are filled by heavy syllables (the logical inverse of Quantity-sensitivity). Languages of this type are sometimes called Quantity-Dependent. We thus end up with a typology of three possible left-headed feet (and three right-headed feet mirroring them):

(2)

Quantity-Insensitive	Quantity-Sensitive	Quantity-Dependent
- Quantity-Sensitive	+Quantity-Sensitive -OB	+Quantity-Sensitive +OB



The last parameter, LBOB, makes a further division in the set of quantity-sensitive languages. It decides that within a 'left-dominant' foot the left syllable is strong if and only if it is heavy; otherwise the right syllable is strong (also if it is light). The resulting foot inventory is a bit complex, and has been disregarded in later work since Hammond (1986).

As far as I am able to see, the parameters idea of stress typology met with almost general acceptance in the years that followed. One difference between a lot of phonological (metrical) literature and its syntactic counterpart (within Government-Binding Theory) was that the latter had at the time abandoned the notion of rules almost completely. Grammatical processes were supposed to be universal and exceptionless, and basically reducible to Move α. The paramaters mostly were statements about what kinds of representations were permissble.

In phonology, on the other hand, it was often supposed that *rules* could be parametrized. This could happen in two ways, both of them exemplified in Prince (1983) influential paper, in which basically all parameters are defined on the formulation of rules and not on representation. A parameter would specify whether the assignment of metrical structure starts from the left or from the right, whether it starts with a 'peak' or with a 'trough'. He also discusses an alternative option, viz. of making the presence or absence of a rule (the 'End Rule') parametrizable.

The discussion that followed concentrated on a different aspect, viz. the right way to represent stress, e.g. in terms of trees, of grids (which was Prince (1983)' proposal) or of 'bracketed grids'. The latter was the point of departure for Halle and Vergnaud (1987). Like Prince, these authors assumed a primarily parametrization of phonological rules rather than of representations.

This is true for instance for the headedness parameters. In both syntax and phonology, we find variation in which element is the head in linearly ordered constituents [XY]. In syntax, there are roughly two ways to represent this. One is to assume that left-headed structures are well-formed in some languages, whereas in others only right-headed structures are acceptable. In principle, all languages can construct both – but the unwellformed ones will be filtered out. Alternatively, one can assume that the structure-building rule itself is parametrized: languages choose either (3) or (4) (Travis, 1983, 1988, 1989):

(3) $X' \rightarrow X\ (YP)$

(4) $X' \rightarrow (YP)\ X$

Since Stowell (1981), the standard conception in syntax on these matters seems to be, however, that the parameters are restrictions on structures, and not properties of rules.

That idea did not seem to be accepted very generally, however, in metrical analysis. Hayes (1995), arguably the cumulative standard work on parametric metrical stress theory, 7 basic parameters are proposed (along with a rather large number of accessory parameters regulating e.g. extrametricality). Hayes states explicitly that (p. 54) '[b]y setting all the relevant parameters, one derives a stress rule'.

An exception to this derivational view on rules was Dresher and Kaye (1990), which presented the following list of parameters:[3]

(5)
a. The word tree is strong on the [Left/Right]
b. Feet are [Binary/Unbounded]
c. Feet are built from the [Left/Right]
d. Feet are strong on the [Left/Right]
e. Feet are quantity sensitive (QS) [Yes/No]
f. Feet are QS to the [Rime/Nucleus]
g. A strong branch of a foot must itself branch [No/Yes]
h. i. There is an extrametrical syllable [NoYes]
   ii. It is extrametrical on the [Left/Right]
i. A weak foot is defooted in clash [No/Yes]
j. Feet are noniterative [No/Yes]

Of these parameters, only two or possibly three, are stated in derivational terms ('built from the Left/Right', 'is defooted' and maybe 'are noniterative')

Dresher and Kaye (1990) thus provide a rather vast typological space of languages: 10 binary parameters gives $2^{10} = 1024$ possible languages, although the parameters are not all independent. For instance if feet are Unbounded, it is no longer relevant whether they are built from the left or the right or whether or not they are assigned iteratively; and if languages are not quantity-sensitive, we do not need to know whether or not they look at the level of the nucleus or that of the rime.

Dresher and Kaye (1990) is an important paper for parameter theory also because it provides an explicit learning algorithm and shows how a rather complex set of parameters, that is close to complete in its empirical domain, can be set by observing the stress patterns of a language.

Following Lightfoot (1989), Dresher and Kaye propose that every parameter has a default status and is accompanied with a cue. The learner pays attention to the cue — some piece of positive evidence — in order to be able to set the parameter rightly. (See also Dresher (2012).)

Consider the quantity-sensitivity parameter. The default setting here will be quantity-insensitivity (QI): there is no difference between heavy and light syllables. The learner may set this to quantity-sensitivity (QS) if she finds

[3] Government Phonology (Kaye et al., 1985, 1990, GP) is a framework which is modelled to a large extent on Government and Binding Theory in syntax, including representational parameters. There is not a lot of work on stress within GP. An exception is (Scheer and Szigetvari, 2005), which present purely representational parameters to account for the weight of different kinds of 'heavy' syllables.

that there are words which have the same number of syllables but a different stress pattern (e.g. *páta, patán*. Notice that this cue is *not* an individual word: the learner thus needs to have stored a few patterns, and possibly apply statistics to them to avoid being misled by one misinterpretation, in order to be able to decide.

Another property of the parametric system is that constraints are ordered. For instance, the learner can only decide about the parameter on the word tree once several decisions about feet have been made (since the word stress by necessity is built on top of a foot). Because constraint settings are not completely independent from each other, Dresher and Kaye (1990) calculate that their system generates 216 possible languages.

Some parameters are left out of consideration; an important one is lexical stress, which actually causes some problems, because lexical stress systems (in which the location of stress is specified for each lexical item separately) are by definition unpredictable. This means that given a limited amount of input, the learner might be misled to believe that she is confronted with a certain language with predictable stress, just because the words she has heard so far fit within a certain pattern. For instance, it is possible that she has by some accident heard only those words which happen to have stress on the final syllable, even though in some other words of the language, stress can fall on any other syllable in the word.

On the other hand, setting the 'lexical stress' parameter too quickly will never lead to a resetting of that parameter, since there is no possible word pattern which violates the resulting grammar. In other words, we can only confidently set the lexical stress parameter once we have explored the full parametric space of predictable stress languages and concluded that our language does not fit into that space. That should make the lexical stress languages the most difficult to learn. At the same time, there is never a guarantee for any learner that she is not actually learning a lexical stress language, especially given the fact that many languages seem to have 'exceptions', e.g. on loanwords.

Gillis et al. (1995) give an exhaustive review of Dresher and Kaye (1990). They judge that the learning procedure is more complicated than it would seem at first sight:

> we found that in formulating the cues, much more was needed than positive evidence; some cues rely heavily on indirect negative evidence, whereas others involve extensive cross-word comparisons and multiple ways of recoding the input material. Furthermore, a rather elaborate ordering of the parameters was set up so that parameter settings could be made dependent upon the value of previously set parameters. Finally, exhaustive search was used in restricted parts of the solution space. (...) Both the use of indirect negative evidence and extensive cross-word compar-

isons in the form of consistency checks require all relevant data to be present when learning is initiated. In the context of a deterministic learner, oversights in the data collection phase may lead to wrong initial decisions.

Gillis et al. (1995) also did an empirical test in which they made up (artificial) examples of each of the 216 languages predicted by the parameter set, and submitted them to the learner. About 75-80% of them was assigned a correct analysis; which means that Dresher and Kaye (1990) learning algorithm would be less than perfect.

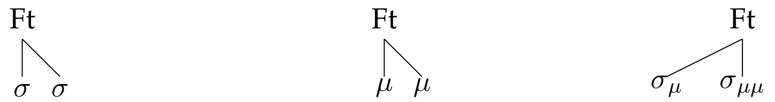
3 Inventories

The arguments given in Gillis et al. (1995) did not have a strong impact on phonological theory, at least at first. Around the same time, however, it seems that many ‘mainstream’ phonologists were abandoning the notion of parameters for the explanation of stress patterns.

The first step in this development was taken by Hayes (1987, 1995); McCarthy and Prince (1996). In parallel, but independently from each other, these authors made the same discovery, viz. that there is a gap in the typological space predicted by binary parameter theory: while trochaic languages can be either quantity-sensitive or quantity-insensitive, iambic languages always seem to be of the former type. Obviously, a parametric system containing the parameters QS/QI and trochee/iamb predicts four types of languages.

The solution suggested by these authors was that Universal Grammar provides an inventory of three different foot types:

(6) Syllabic (QI) trochee: $(\sigma\sigma)$ Moraic (QS) trochee: $(\mu\mu)$ Iamb: $(L\sigma)$ or (H)



Trochees are ‘even’ (there is no asymmetry between head and dependent in terms of syllable structure), but come in two types: QI trochees consist of two syllables, regardless of their weight. QS trochees consist of two mora’s, which means that they either consist of one heavy syllable or of two light syllables. Iambs are always weight-sensitive, and this in a slightly different way than trochees: they consist of a light syllable, followed by a heavy one, or only of a heavy one, if no light syllable is avaialble in the word. This implies that iambic languages always need to have some way to express weight on syllables, and typically they will have a form of iambic lengthening.

It is clear that this system cannot be expressed by binary parameters, although it could obviously be expressed in a system with a parameter with three possible values. Such a system would not have a lot of explanatory value, however: where do exactly these three values come from? Why do we not find the mirror images of the foot types in (6)?

Hayes (1987) suggested that the answer to this question was to be found in what he called the Iambic-Trochaic Law (ITL, see Hyde, 2011, for an overview):

(7) *The Iambic-Trochaic Law*

a. Elements contrasting in intensity naturally form groupings with initial prominence.

b. Elements contrasting in duration naturally form groupings with final prominence.

The ITL is assumed to have a wider psychological applicability. It is based on psychological experiments (Bolton, 1894; Woodrow, 1909) in which informants listened to sequences of (non-linguistic) sounds, and were asked to group these. Some sequences alternated in intensity and others in duration. The results seemed to conform the ITL. (Later studies, such as Iversen et al. (2008) question some aspects of this universality; for instance, Japanese listeners seem to react differently than English participants.)

This development was in several ways representative of certain changes which were in the air towards the end of the 1980s, and in the early 1990s within phonology. The most prominent of these probably was a move away from purely formal accounts towards accounts that were grounded in extralinguistic factors, cognitive or otherwise. In this case, this concerned a psychological law which obviously itself is somewhat mysterious: there is no immediate explanation why the human mind would have a preference for grouping elements contrasting in duration in a different way from other kinds of auditory signal.

To this extent, then, the ITL does not really 'explain' the typological gaps we find, but it rather suggests that there is a correlation between linguistic typology and the results of a psychological experiment.

Another way in which this theoretical turn was typical for phonological research was the interest in gaps in the typology. One could argue that the fact that a certain class of languages — those with QI iambs — has not been found, does not mean that they not exist, and even that *if* they do not exist, this does not necessarily mean that they could not be generated by the language faculty. Both of these facts could be accidents of history, and have a non-linguistic explanation, as was correctly observed by Hale and Reiss (2000, 2008). (The typological gap argument has played a role in the literature also after Hale and Reiss made their observations.)

The issue is an interesting one: what are the responsibilities of typological theory? On the one hand, the theory should definitely account for all systems that are attested. At the same time, the theory should also be relatively tight, in order to make predictions and to be falsifiable.

Now there seems to be consensus that some systems are definitely to be excluded from linguistic theory. An example that is sometimes cited would be a language in which stress is put on every prime-numbered syllable of the word (so syllables number 2, 3, 5, 7, 11, etc.). Such languages do not exist and it may be assumed that they would be outside of ordinary human linguistic capacities.

In a thoughtful paper, Blaho and Rice (2013) discuss the issue of falsification at great length. There basic argument is as follows. Suppose we dispose of two theories Θ_1, which predicts that there are four types of languages based on two binary parameters (languages A, B, C and D), and Θ_2, which predicts that there are only three types of languages. Inherently, Θ_2 seems preferable; however, it is not clear that this is the case if Θ_2 has all of the machinery of Θ_1 plus the extra statement 'language D cannot exist'. One might argue that in the case at hand, setting up an inventory of three different feet is not necessarily really more restrictive than a theory which is conceptually simpler, but makes the extra prediction.

As a matter of fact, it could be argued that Θ_2 is not really more restrictive in any deep sense: if it would turn out that language D would exist, Θ_2 would be very easy to repair, viz. by lifting the restriction, or adding the extra foot to the typology. The prediction thus is not very deep in any way. Θ_1, in turn could relegate the explanation of why language D is so rare to other factors, such as learnability or (other) diachronic considerations.

Another objection against the introduction of inventories, is that the same typological gap has also been explained without reference to the ITL. An influential contribution was (Kager, 1993), arguing that the inventory of feet was indeed symmetric, consisting of syllabic and moraic trochees and iambs, and that the asymmetries which are observed in the typology can be derived from independent requirements, such as the avoidance of 'lapses' and 'clashes' (sequences of consecutive unstressed or stressed syllables, respectively) on the grid. Take for instance the fact that there are no 'uneven trochees', consisting of a heavy syllable, followed by a light syllable, even though there are 'uneven iambs'. If we admit that the first mora in a heavy syllable is the head, an uneven trochee would consist of a head, followed by two non-head, causing a lapse within the foot. An uneven iamb, on the other hand, consists of a weak-strong-weak sequence, so that there is no lapse.

A different approach was taken by (van de Vijver, 1998), who argued that there is no specific parameter setting (or constraint) in favour of iambs: languages prefer trochees by default, and iambs emerge only when this can avoid a conflict. This approach is embedded within Optimality Theory, which brings us to the next blow which was brought to the parametric vision of metrical structure.

4 Constraint ranking

As is well-known, Optimality Theory has dominated phonological theory for the past two decades. As a matter of fact, OT is, if anything, an alternative to Principles and Parameters Theory, that is to say a theory of the way in which languages can vary: it presents a radically different view of language variation, which is seen in this case not as a matter of binary choices within an otherwise unchangeable frame, but of different rankings of surface constraints. As a matter of fact, within classical OT the claim is that those rankings are the *only* possible systematic phonological differences between languages: there are no differences in underlying representations, and languages also do not differ in the absence or presence of certain phonological processes. Everything is universal, except for the interaction of constraints on surface representations.

One difference between constraint ranking systems and parametric systems is in the numbers. Assuming that parameters are binary and independent from each other, a set of n parameters gives 2^n possible languages; a system with n constraints, on the other hand, making the same assumptions, gives $n!$ possibilities, a number that is growing much more rapidly. This means that constraint-based systems need to be set up more sparsely: all else being equal, the number of constraints should be smaller than the number of parameters.

Another important empirical difference between a parameter theory and OT is that the latter predicts so-called *the emergence of the unmarked* (TETU) effects. Prince and Smolensky (1993) give the following example. One difference between languages is that some allow closed syllables, whereas others do not. Parameter theory could analyse this as a parameter: 'Allow closed syllables yes/no'. OT provides for a constraint NOCODA which is undominated in some languages and dominated by faithfulness constraints against deletion of consonants or insertions of vowels in others.

Presumably, in all languages monomorphemic *VCV* sequences are syllabified as *V.CV*. This is true even in languages which otherwise allow for closed syllables — in other words in which *VC* is a well-formed syllable. Within parameter theory, this fact needs a separate explanation: some languages apparently have the parameter ALLOW CODA'S switched to YES, so why would they still disprefer them? Within OT, it can be explained by the constraint NOCODA. Even in languages in which it is low ranked, it is still present. If the higher-order constraint are irrelevant — like here, where we do not have to be unfaithful in order to satisfy the constraint — we will still see the effects of the constraint. In cases such as these, OT can thus boast an advantage of theoretical elegance: two phenomena which have to be analysed as separate under a different set of assumptions get a uniform explanation here.

I am not aware of explicit TETU arguments against parameter theory from the domain of stress. They seem probably also be difficult to provide, since TETU arguments, like the one above, usually involve some interaction with faithfulness to underlying representations: a markedness constraint is usually marked by a faithfulness constraint, but when such a faithfulness constraint does not play a role, the markedness constraint shows its working again. In the syllable structure example, NOCODA is ranked below faithfulness constraints preventing the insertion of vowels or the deletion of coda consonants (two ways to satisfy the constraint), but in *VCV* words we can satifsfy the constraint without deleting or inserting anything. However, in the analysis of stress, underlying representations do not usually play a role: it is assumed that feet are not underlyingly present, so that faithfulness is never applicable to them. At the same time, underlying at least some of the theoretical work on stress typologies within OT, the TETU idea does play a role, as I will show shortly.

At the same time, OT does not fit very well intuitively with the kind of binary options that are typical of parametric systems and, possibly, of the typology of stress. This is a potential disadvantage, as I will now illustrate.

Consider, for example, the issue of iambs vs. trochees. Some languages have iambs, while others have trochees. How are we going to represent this typological fact? One option would be to have two constraints IAMB and TROCHEE, of which the requirements are exactly the opposite. Trivial constraint-ranking will then give the required result:

(8) a. IAMB: Feet are right-headed
b. TROCHEE: Feet are left-headed.
c. IAMB$\gg$TROCHEE: Iambic feet
d. TROCHEE$\gg$IAMB: Trochaic feet

Although it is obviously technically possible to do it like this, it does not seem to be a very insightful way to pursue the analysis. Furthermore, the system seems to predict that there can be languages which mix foot types: by default there would be iambs, but in case some other higher-order constraint disallows those, trochees show up. Such mixed language types do not seem to exist.

The same problem arises at other levels of prosodic analysis as well. E.g. some languages have main stress on the right-most foot, and others have it on the left-most foot. There are no languages which have stress exactly in the middle of the word, so to some extent this is again a binary choice: stress is drawn to the edges of the word, the only issue is which edge.

The problem is traditionally treated in OT by reference to so-called constraint schemata, of which the first instance was the so-called AGREE family of constraints McCarthy and Prince (1993), which is defined as follows:

(9) ALIGN(X, L/R, Y, L/R): Align the left/right edge of every X to the left/right edge of a Y.

Constraint schemata are thus constraints with built-in parameters; by setting the parameters, we get constraints:

(10) a. ALIGN(Ft, R, Hdσ, R): Feet are right-headed (align the right edge of every foot to the right edge of a stressed syllable.
b. ALIGN(Ft, L, Hdσ, L): Feet are left-headed.
c. ALIGN(Wd, R, HdFt, R): Main stress is at the right.
d. ALIGN(Wd, L, HdFt, L): Main stress is at the left.

Two interpretations are possible of the relation between a constraint schema and its instantiations. One can either believe that all individual constraints are universally present, and they are ranked in individual languages such that the requirement for iambs obscures that for trochees, or the other way around. This is the interpretation which McCarthy and Prince (1993) give, but still does not solve the problems just mentioned.

Alternatively, one might assume that the parameters are really set on a language-specific basis. In my view, that would solve the problems just mentioned: a language would choose once and for all whether it likes right-headed or left-headed feet, and stick to that choice.

In van Oostendorp (1995) I argue for such a position also on other grounds. I observe that there is a typological it split between languages such as French in which many phonological processes (like stress and schwa epenthesis) seem sensitive to the phonological phrase, whereas in other languages such as Dutch those same processes seem sensitive primarily to the phonological word. These processes are independent from each other and probably to be accounted for by independent constraints. Yet those constraints seem to refer to a different parameter setting in French than they do in Dutch.

If this were to be accepted — there are no indications that the idea has gained a very wide following, but I would like to bring it again to the reader's attention —, it would mean that languages differ in two kinds of ways: in the parameter settings which are chosen in individual constraint schemata, and in the rankings of the resulting constraints. This goes against a central tenet of classical OT, viz. that constraints are universal and that the only difference between languages can be their respective rankings.

It has to be noted, however, that the concept of one universal set of constraints seems to have been abandoned in a lot of OT literature, where it is often proposed that constraints are constructed by the language learning child in the course of acquisition. Moreton (2008); Hayes and Boersma (2001). In such theories, there obviously has to be some model on which new constraints are based, and these can thus be seen as parametrized constraints. It thus seems that the parameter is playing a role also in OT theories.

In any case, the discussion on typologies within OT of the last twenty years seems to have concentrated mostly on issues of representation, although it has been argued that certain representational options work better under the assumption of violable constraints. In particular, it has been argued that the technology of surface-based constraints seems to lead in particular to a certain type of representational assumption. Gordon (2002), for instance, argues that the typology of quantity-insensitive languages can be generated with only bracketed grids, and constraints against lapses and clashes:

(11) a. CLASH: A string of more than one consecutive stressless syllable may not occur.
b. LAPSE: A string of more than one consecutive stressed syllable may not occur.

The point is, of course, that languages sometimes do allow for lapses or clashes. For instance, in a word consisting of an uneven number of syllables and binary feet, there is no way to avoid them. In such cases, one of the two constraints can be violated, but this will happen only when it is necessary. This is thus a TETU argument: even though a language sometimes shows clashes, it does not mean that some parameter CLASH is switched 'on': they will still be avoided where possible.

Similarly Brett Hyde has argued in a number of publications (Hyde, 2002, 2007a,b, e.g.) that OT allows us to have overlapping feet (so syllables which are in the weak branch of one foot and at the same time in the strong branch of another foot). The reason why we need violable constraints for this, is that we need to assume that such overlapping feet are usually avoided; they only arise under certain very specific circumstances. This can of course be realized by assuming that the constraint against overlapping feet is universal, but violable. In a parametric approach, one would be forced to assume that overlapping feet are allowed in some languages, but this would not explain why also in those languages they are marked.

Both Gordon and Hyde thus rely on an effect that can be understood as a form of TETU after all: some structures are disallowed in some languages and demonstrably dispreferred in languages which have them. This correlation is (part of) the classical definition of markedness, but parameter theory cannot express it.

An interesting result in this respect is in the analysis of ternary rhythms by Elenbaas and Kager (1999). In some languages, stress does not fall on every second, but on every third syllable. A well-known example is Cayuvava (Key, 1961):

(12) a. ˈda.pa 'canoe'
b. ˈto.mo.ho 'small water container'
c. a.ˈro.po.ro 'he already turned around'

d. a.ri.ˈpi.ri.to ‘already plantes’
e. ˌa.ri.hiˈhi.be.e ‘I have already put the top on’
f. ma ˌra.ha.ha.ˈe.i.ki ‘their blankets’
g. i.ki ˌta.pa.ra.ˈre.pe.ha ‘the water is clean’

Parametric theories will typically need to introduce some special device in order to describe such languages, such as ternary feet or (parametric) ‘weak local parsing’ (Hammond, 1984; Hayes, 1980, 1995). This implies that the parametric system somehow needs to be extended to incorporate languages of this type.

Elenbaas and Kager (1999) however propose that nothing special has to be posited in order to derive such languages. They show how the Cayvava pattern can be derived from the interaction between the constraints LAPSE (above), ALLFT-R and PARSE which all have an independent justification:

(13) a. ALLFT-R: Every foot should be at the right edge of the word (give a violation for every syllable between the right edge of a foot and the right edge of a word)
b. PARSE: Every syllable should be inside a foot.

(14)

$\sigma\sigma\sigma\sigma\sigma\sigma\sigma$	*Lapse	AllFt-R	Parse
a. ☞ $\sigma(\sigma'\sigma)\sigma(\sigma'\sigma)$		***	**
b. $(\sigma'\sigma)(\sigma'\sigma)(\sigma'\sigma)$		***!**	
c. $(\sigma'\sigma)\sigma(\sigma'\sigma)\sigma$		***!*	**
d. $\sigma\sigma(\sigma'\sigma)(\sigma'\sigma)$	*!	**	**

Since ALLFT-R gives a violation for every foot that is not at the right edge, it will be violated by any word that has more than one foot. In this sense, then, it can also do the work of the unboundedness parameter, giving us languages with only one foot per word. The constraint PARSE, on the other hand, prefers all syllables to be parsed in a foot structure, something we need for languages that have exhaustive parsing anyway. The independent use of LAPSE we have already seen. When they are ranked in the order LAPSE≫ALLFT-R≫PARSE, these constraints turn out to ‘predict’ weak local parsing.[4]

We thus do need an extra parameter to account for ternary stress, but not an extra constraint. I have pointed out above that adding one constraint

[4] A slightly different approach is taken by Martinez-Paricioinez-Paricio (2013), who argues in favour that ternary feet are actually ‘recursive’, branching into a weak and a strong branch twice. Also in this case, the recursive feet are presented as ‘last-resort devices’, “to avoid degenerate feet and ensure exhaustive parsing of syllables”. The idea of recursive feet was also proposed in parametric frameworks, however, as Martinez-Paricioinez-Paricio (2013) aknowledges. See for instance Hulst (2010) for an overview.

to an existing hierarchy of constraints gives much more power than adding a binary parameter (at least if the existing hierarchy has at least two constraints). Together, these two factors imply that it is not so clear which of the approaches is better, because the counting of theoretical parsimony becomes rather difficult. Parameter theory would need an extra device, but can add this in a relatively cheap way; constraint ranking theory does not need an extra device, but already contains a rather vaster typological space.

The structure of this typological space is a topic that has been studied quite widely in the literature on stress. Tt seems to me that, as in the days of parameter theory, stress still is one of the best explored topics in phonological variation under OT, and maybe in language variation generally. It is a typical requirement — although not a standard practice — for any OT paper to explore the factorial typology which is predicted by its constraints. There is also software to do so: two packages which can calculate the typological predictions of a given constraint set are OTSoft[5] and OTWorkplace[6]. It is clear that the authors of at least the latter have stress typology in mind: their example constraint set contains a number of constraints describing stress.

Also learning theory has developed quite extensively within OT, and profited from computer implementations; OTSoft for instance, provides a model for learning OT grammars, as does Praat[7]. Different from the parametric studies we observed, learning models are not usually cue-based, but more often error-driven: they assume a grammar and compare it to the actual data of the language, fixing the assumptions when things go wrong, i.e. when the outputs of the grammar are different from the observed data.

Optimality Theorists have been rather successful in their typological work, as well as the acquisition models; more work has been performed in this model than in any other of these topics than in any alternative theory, it seems. An interesting recent development is in unpublished work by Birgit Alber and Alan Prince, who study the structure of OT typologies based on a few very simple stress-related constraints, and show how we can learn things from those typologies quite easily.

5 Stochastic ranking

The last decade of phonological research in the phonological mainstream has been characterized by a gradual shift away from categorical models towards more stochastically oriented views of language. This is also starts having its effect on the study of stress. As far as I am able to tell not a lot of work has been done along these lines, but we may expect the importance of this line of work to grow.

[5] http://www.linguistics.ucla.edu/people/hayes/otsoft/
[6] https://sites.google.com/site/otworkplace/
[7] http://www.praat.org/

An important focus point of this line of work is its emphasis on *weak generalizations*: languages do not altogether classify completely within a particular typological niche. An example of this can be found in Italian. Cei and Hayes (2012) note that there is controversy whether or not this language has quantity-sensitive stress. For most words with a penultimate heavy syllable, stress is on that syllable. If the penultimate syllable is light, however, stress is (usually) on the antepenultimate syllable.

However, there is "half a dozen" words which have an heavy penultimate syllable, yet stress on the antepenult: *mandorla* 'almond', *acanto* 'acanthus', *polizza* 'little note' and the place names Otranto, Lepanto and Taranto. How should we evaluate these 'exceptions'? Den Os and Kager (1986) argue that these examples show that Italian is not quantity-sensitive after all. D'Imperio and Rosenthal (1999), on the other hand, argue that these are truly exceptions which fall outside of the system proper, which therefore *is* quantity-sensitive.

Cei and Hayes (2012) say that their opinion is "somewhere in between": "heavy penult is an exceptionally strong factor for semi-predicting Italian stress". They work with a computational model that can deduce surface-based constraints from a set of real Italian words with a stress marking. Those surface-based constraints get some weight assigned to them which is represented as a positive real number. In the evaluation of a new word, the different values for each of the constraints is calculated according to some formula which gives a set of new real numbers, which are each to be interpreted as the likelihood that stress is on some syllable of the word.

There is already quite a long-standing tradition in generative phonology studying the way in which lexical exceptions can be fitted into the overall formal theory (see Zonneveld, 1978, for an early example). In previous theories, the solution was typically a representational one: the exceptional forms would have some kind of 'hidden' structure that would make them behave differently. Formal markedness would correspond to frequency: the more hidden structure, the less likely a form is. Ironically, it is not completely clear how exceptionality works out in the newer approaches: stress would still have to be marked on all forms, but how would that relate to the relevant frequency of forms?

However this may be, the constraints can refer to many different aspects of the phonological shape of the word. There is no assumption that this is constrained beforehand, e.g. by foot structure. Instead, anything that could possibly be categorized is taken into consideration, such as the complexity of the rhyme, but also of the onset, as well as the quality of the vowel. These latter factors play an important role in this type of work; see for instance also Ryan (2011) for a more extensive discussion of the role of onset complexity in the role of stress. The important point is that such complexity rarely (if ever) plays a decisive role in the assignment of stress. For traditional parametric theories, this means that we have to assume that they do not play a role in the typology and therefore also not in the grammar of the languages in question. It is therefore the goal of the recent types of work to show that native speakers do indeed show knowledge of these patterns in experimental situations.

It may be presently too early at this point to evaluate this evidence, but the development is definitely an interesting one. We could characterize it as a development in the direction of micro- or even nanotypology: stress in a language is not just characterized by a small set of universal parameters of constraints, but by a relatively large number of partly language-specific constraints that can have their origin in universal preferences as well as in the accidents of the history of the individual language.

One question that may arise is why some of such constraints (such as those about the structure of the rhyme) make it to the macrotypology, whereas others (such as those about the structure of the onset) do not. The argument would typically be that this is grounded in extragrammatical factors, such as phonetics (in this case the fact that rhymal complexity may be more salient and add more to the length than onset complexity).

One could wonder, however, *why* people would have this kind of stochastic knowledge at all, and why it would be represented in this way. Given that the knowledge is stochastic, people are not completely sure where to put stress when confronted with a new word – if they do not hear that stress to begin with. If they *do* hear a new word, on the other hand, they will already know where the stress is located.

Notice that is partially different under more standard accounts, in which the child learns a stress pattern which then becomes redundant and does not have to be lexically stored. Learning stress is then thus a matter of economy. Under both approaches the stress system seems most useful in parsing, because it makes it predictable where word boundaries are (e.g. if a language has initial stress on every word, you know that a new word has started once you hear a stress). However, if stress is lexical or only stochastic, it also becomes less useful as a marking device.

6 Conclusion

We can thus conclude that the fact that languages seem to have stress systems at all, and that speakers have knowledge about those systems, is not completely understood. We can add to this an interesting observation made by Hyman (2008): there are languages which do not seem to have word stress at all; but when a language *does* have stress, then *all* lexical words will have it (clitics and other function words of course do not have to have it).

Hyman (2008) notes that this is in itself a remarkable fact, given that there are languages with rather complicated stress systems and default rules: stress the rightmost heavy syllable in the word, and if there is no heavy syllable, stress the leftmost light syllable. However, there is no language which would have as a rule 'stress the rightmost heavy syllable in the word, and if there is no such syllable, do not assign stress'.

This seems to point then in the direction of at least one big macroparameter: LEXICALSTRESS (YES/NO). This parameter also makes clear sense from a learning perspective: when switched on, the child knows that she will have to look for stress in every word she encounters. This cannot really be expressed in terms of violable constraints, let alone in terms of stochastics. This widens the gap between the macrolevel and the microlevel of variation even more.

It seems fair to see that formal theory has brought a lot of insight to our understanding in particular of the macrotypology of stress systems in human language. Although obviously, many questions and problems remain, existing theories have gained a broad empirical coverage, and this has definitely contributed to the success of parameter theory first and Optimality Theory afterwards.

I submit that one important goal for phonological theory is to try to find a coherent framework which can express the different kinds of insights of the different approaches that have been tried in a natural way: micro- and macrovariation, the fact that most variation seems to fit in a system of simple binary choices and the the fact that some of these choices seem to be seldom or even never chosen.

Bibliography

Blaho, S., Rice, C., 2013. Overgeneration and falsifiability in phonological theory. URL `http://www.blahosylvia.info/papers/falsifiability.pdf`.

Bolton, T., 1894. Rhythm. American Journal of Psychology 6, 145–238.

Cei, E., Hayes, B., 2012. Multiple factors affecting italian word stress. URL `http://www.linguistics.ucla.edu/people/hayes/ItalianStress/Text/Phonc` handout Phonology Seminar UCLA.

Chomsky, N., 1981. Lectures on Government and Binding. Foris, Dordrecht.

Den Os, E., Kager, R., 1986. Extrametricality and stress in Spanish and Italian. Lingua 69, 23–48.

D'Imperio, M., Rosenthal, S., 1999. Phonetics and phonology of main stress in italian. Phonology 16, 1–28.

Dresher, B. E., 2012. Covert representations and the acquisition of lexical accent. Slides available at `http://homes.chass.utoronto.ca/~dresher/talks/Dresher%20UDCSA%20slide`

Dresher, B. E., Kaye, J., 1990. A computational learning model for metrical phonology. Cognition 34, 137–195.

Elenbaas, N., Kager, R., 1999. Ternary rhythm and the Lapse constraint. Phonology 16, 273–330.

Gillis, S., Durieux, G., Daelemans, W., 1995. A computational model of p&p: Dresher & kaye (1990) revisited. In: Verrips, M., Wijnen, F. (Eds.), Approaches to Parameter Setting. Amsterdam Series in Child Language Development, pp. 135–173. URL `http://www.cnts.ua.ac.be/~walter/papers/1995/gdd95.ps`.

Gordon, M., 2002. A factorial typology of quantity-insensitive stress. Natural Language and Linguistic Theory 20, 491–552.

Hale, M., Reiss, C., 2000. 'substance abuse' and 'dysfunctionalism': Current trends in phonology. Linguistic Inquiry 31, 157–169.

Hale, M., Reiss, C., 2008. The Phonological Enterprise. Oxford University Press, Oxford.

Halle, M., Vergnaud, J.-R., 1987. An Essay on Stress. MIT Press, Cambridge, Mass.

Hammond, M., 1984. Constraining Metrical Theory: A Modular Theory of Rhythm and Destressing. Ph.D. thesis, UCLA, Los Angeles. Published 1988, Outstanding Dissertations in Linguistics series, Garland Press, New York.

Hammond, M., 1986. The obligatory-branching parameter in metrical theory. Natural Language and Linguistic Theory 4, 185–228.

Hayes, B., 1980. A Metrical Theory of Stress Rules. Ph.D. thesis, MIT, Cambridge, Mass. Published by Garland Press, New York, 1985.

Hayes, B., 1985. Iambic and trochaic rhythm in stress rules. In: Niepokuj, M., VanClay, M., Nikiforidou, V., Jeder, D. (Eds.), Proceedings of BLS 11: Parasession on Poetics, Metrics, and Prosody. Berkeley Linguistic Society, Berkeley, pp. 429–46.

Hayes, B., 1987. A revised parametric metrical theory. In: McDonough, J., Plunkett, B. (Eds.), Proceedings of the North East Linguistic Society 17. GLSA Publications, Amherst, pp. 274–289.

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

Hayes, B., Boersma, P., 2001. Empirical tests of the Gradual Learning Algorithm. Linguistic Inquiry 32, 45–86.

Hulst, H. v., 2010. A note on recursion in phonology. In: Hulst, H. v. (Ed.), Recursion and Human Language. Mouton de Gruyter, Berlin/New York.

Hyde, B., 2002. A restrictive theory of metrical stress. Phonology 19, 313–359.

Hyde, B., 2007a. An alignment approach to stress windows. Colloquium presentation, Rutgers University (available at http://www.artsci.wustl.edu/ bhyde/windowshandout.pdf).

Hyde, B., 2007b. Nonfinality and weight sensitivity. Phonology To appear.

Hyde, B., 2011. The iambic-trochaic law. In: van Oostendorp, M., Ewen, C., Hume, E., Rice, K. (Eds.), The Blackwell Companion to Phonology. Blackwell-Wiley, Oxford and New York, pp. 1052–1977.

Hyman, L., 2008. Universals in phonology. The Linguistic Review 25, 85–137.

Iversen, J., Patel, A., Ohgusi, K., 2008. Perception of rhythmic grouping depends on auditory experience. JASA , 2263–2271.

Kager, R., 1993. Alternatives to the iambic-trochaic law. Natural Language and Linguistic Theory 11, 381–432.
Kaye, J., Lowenstamm, J., Vergnaud, J.-R., 1985. The internal structure of phonological elements. a theory of charm and government. Phonology Yearbook 2, 305–328.
Kaye, J., Lowenstamm, J., Vergnaud, J.-R., 1990. Constituent structure and government in phonology. Phonology 7.
Key, H., 1961. The phonotactics of Cayuvava. International Journal of American Linguistics 27, 143–150.
Lightfoot, D., 1989. The child's trigger experience: Degree-0 learnability (with commentaries). Behavioral and Brain Sciences 12, 321—375.
Martinez-Paricioinez-Paricio, V., 2013. An Exploration of Minimal and Maximal Metrical Feet. Ph.D. thesis, University of Tromsø.
McCarthy, J., Prince, A., 1986. Prosodic morphology. Manuscript.
McCarthy, J., Prince, A., 1996. Prosodic Morphology. Tech. rep., Rutgers University Center for Cognitive Science.
McCarthy, J. J., Prince, A., 1993. Generalized Alignment. In: Booij, G., van Marle, J. (Eds.), Yearbook of Morphology. Kluwer, Dordrecht, pp. 79–153. Excerpts appear in John Goldsmith, ed., Essential Readings in Phonology. Oxford: Blackwell. Pp. 102?136, 1999.
Moreton, E., 2008. Analytic bias and phonological typology. Phonology 25, 83–127.
Prince, A., 1983. Relating to the grid. Linguistic Inquiry 14, 19–100.
Prince, A., Smolensky, P., 1993. Optimality theory: Constraint interaction in generative grammar. Manuscript, Rutgers University and University of Colorado at Boulder. Available at ROA and published 2004 at Blackwell.
Ryan, K. M., 2011. Gradient syllable weight and weight universals in quantitative metrics. Phonology 28, 413–454.
Scheer, T., Szigetvari, P., 2005. Unified representations for stress and the syllable. Phonology 22.
Stowell, T., 1981. Origins of Phrase Structure. Doctoral dissertation, MIT.
Travis, L., 1983. Word order change and parameters. MIT Working Papers in Linguistics 5, 277–289.
Travis, L., 1988. Parameters of phrase structure. McGill Working Papers in Linguistics 4, 121–140.
Travis, L., 1989. Parameters of phrase structure. In: Baltin, M. (Ed.), Alternative conceptions of phrase structure. University of Chicago Press, Chicago, pp. 263–279.
van Oostendorp, M., 1995. Vowel quality and phonological projection. Ph.D. thesis, Katholieke Universiteit Brabant, Tilburg.
van de Vijver, R., 1998. The Iambic Issue. Iambs as a Result of Constraint Interaction. Ph.D. thesis. Vrije Universiteit te Amsterdam.
Woodrow, H., 1909. A quantitative study of rhythm. Archives of Psychology 14, 1–66.
Zonneveld, W., 1978. A formal theory of exceptions in generative phonology. Peter de Ridder Press, Lisse.