Acta Linguistica Academica Vol. **66** (2019) 1, 115–135
DOI: 10.1556/2062.2019.66.1.5

Where now with Optimality Theory?

Jeroen van de Weijer
Shenzhen University
jeroen.van.de.weijer@gmail.com

Abstract: In this article I discuss the various parts of the OT architecture in phonology: its basic structure as an input-output system, Gen, Con and Eval. These aspects are considered in the light of the question what is innate in language (i.e., part of Universal Grammar) and what could be based on acquisition (i.e., experience with language). We will see that all parts of OT can be related to general cognitive learning strategies, obviating the need for any assumption of innateness. It is hoped that future research in this area will bring OT closer to the field of general cognitive science.

Keywords: Optimality Theory; cognition; Universal Grammar; innateness; acquisition

1. Introduction

The advent of Optimality Theory represented one of the milestones in linguistics in the twentieth century, perhaps comparable with de Saussure's lectures on the nature of language (see de Saussure 1916; Harris 1983) and the start of Chomsky's research programme in generative linguistics (Chomsky 1959; Chomsky & Halle 1968, henceforth SPE). The circulation of Technical Report #2 of the Rutgers Center for Cognitive Science in the early 1990s, which resulted in the publication by Prince and Smolensky (2004 [1993]), shook the foundations of the linguistic world, especially with respect to the issue of the architecture of phonology (and the mental grammar as a whole). The Optimality way of doing phonology had (and continues to have) tremendous implications for the analysis of linguistic phenomena, our view of the language faculty, and how this functions in human cognition as a whole.

Now, a quarter of a century later, it is time to take stock. Optimality Theory (OT) has definitely proven its value in analyses of a wide range of phonological phenomena. It has inspired tremendously successful work, especially in typology and acquisition, two fields which are of central interest in OT. It has brought deeper understanding of interface phenomena, such as the interaction between morphology and phonology, and between prosody and segmental structure. It is perhaps not surprising that OT is ideally suited in areas where conflict arises (between morphemic structure and phonetic pronouncability, between syllable structure and the sounds that constitute syllables), because the very architecture of OT is designed to resolve such conflicts. And it is no surprise that OT has inspired a clear research programme for the study of L1 (and other kinds of) acquisition, since it lays out its premises, predictions and questions so very clearly. Supported by this great empirical success, it is time to revisit the foundational principles of OT, and position this theory in a wider context of phonological, linguistic and general cognition. This is especially important with respect to questions of innate knowledge and knowledge gained from experience ("nature vs. nurture"), which have become increasingly prominent in recent years, and for which language is an ideal testing ground (see e.g., Hauser et al. 2002; Smolensky & Legendre 2006; Prinz 2012; van der Hulst in preparation).

The present contribution therefore approaches OT from the vantage point of Universal Grammar (UG), with the question what part(s) of OT would be innate (part of UG) and what could be acquired on the basis of experience with language (or perhaps other cognitive domains). We will see that OT is rather "UG-rich", i.e., depends on quite a bit of supposedly innate machinery for its architecture and functioning. In my view, this is unfortunate. In the twenty-first century, cognitive science in general seems to be going into a direction where knowledge (e.g., of language) is more and more rooted in experience and general learning strategies, and away from module-specific innate specifications (such as UG) (comparable to the Emergent Phonology approach, e.g., Archangeli & Pulleyblank 2015 and references cited there). We shall therefore ask where OT could perhaps benefit from these new developments, i.e., where bridges could be built between "theoretical" linguistics (i.e., OT) and experience-based learning. We will see that it is in fact easy to find areas where such bridges could be successfully laid, although details of course remain to be worked out.

The organisation of this article is as follows. In section 2, I will first discuss the role of UG in OT, identifying four areas of interest, which are discussed in the following sections: the overall architecture of OT as input-output system (section 3), the Generator (section 4), the constraint set (Con; section 5) and the Evaluation procedure (Eval, section 6). Section 7 briefly concludes and looks ahead to future research.

2. OT and UG

Universal Grammar is most closely connected to the Language Acquisition Device (LAD), posited by Chomsky in early work, motivated especially by arguments like Poverty of the Stimulus (Chomsky 1980). According to this proposal, the LAD would contain a number of principles and parameters, which would assist the language-discovering child to construct the grammar of his or her language (see e.g., Dresher & Hornstein 1976). In this way, the LAD is also responsible for language universals (consistent properties across languages) and parametric variation (where languages show only a few options for a given property of grammar) (see e.g., Thomas 2004). The exact content of the LAD, or UG, has been a matter of intense debate ever since it was first proposed. There are two approaches to this issue, which coincide with approaches to knowledge in general: the first is a nativist viewpoint and liberally postulates innate knowledge, in this case linguistic principles and parameters as part of UG ("nature"). The second is an empiricist approach, which takes the viewpoint that all knowledge (including linguistic knowledge) is derived from experience ("nurture"). The first approach adopts the LAD, the second rejects it or seeks to limit its scope as much as possible.

There is an important methodological point to make here. With respect to the nature-nurture debate, all sides agree that at least **some** of our phonological knowledge is acquired on the basis of experience. Empiricists will say that this in fact is the only basis of phonological knowledge (or take the approach that any phonological knowledge should be explained as such, where a non-understood residue might be imputed to UG; this is also the approach advocated by Culicover 2016). Nativists argue that something more is required, referred to as Universal Grammar. Since empiricism is the leaner theory (not requiring the postulation of an extra, innate element), it is up to the nativists to support this, since nativism postulates something extra. Questions of universalism are very hard to prove (who knows all the facts of all the languages?), so the empiricist approach is also methodologically sound.

Let us focus on what is innate (or derived from experience) in OT. I will discuss four aspects of OT (or rather parts of its architecture) for which this question arises. These are (i) the overall architecture which converts inputs into outputs, (ii) the Generator, i.e., the part of the model that generates an infinite number of candidates from a given input, (iii) Con, the putatively universal constraint set, and (iv) the Evaluator, the mechanism by which the ranked constraint hierarchy in a particular language selects a winning candidate from the candidate set. All vital parts are displayed in Figure 1:

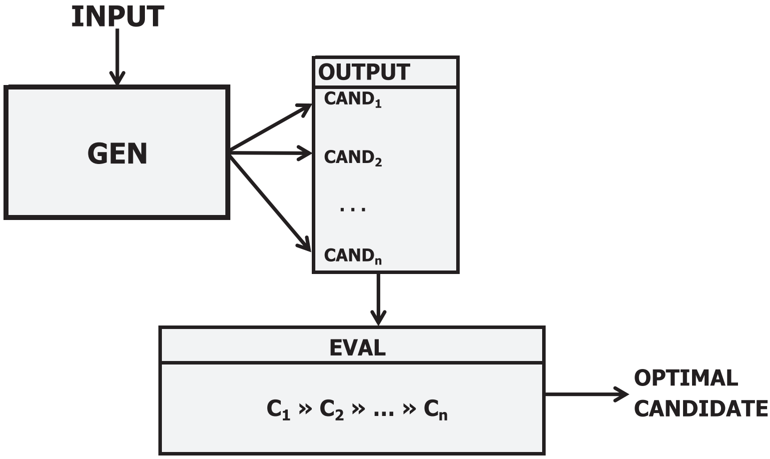

Figure 1: The broad architecture of OT; the constraints C_1, C_2, ..., C_n together make up Con

Let us divide OT into **architecture** (the design of the theory) and **content** (the particular constraints and linguistic alphabets a researcher chooses). The first is more likely to be part of UG, while the second may cause more disagreement. If UG consists of "the linguistic properties shared by all languages" (Kager 1999, 1–4; Thomas 2004), then the architecture (i) should definitely be part of UG, including the Generator (ii) . But OT (unlike SPE, for instance) assumes also that all languages share the same constraint set (iii) , so this should also be innate. Different languages rank these constraints differently. This happens in the course of acquisition on the basis of the data to which the language-acquiring child is exposed. So the ranking within the Evaluator (iv) is not innate, but based on experience.

All these parts are well described in sources like Kager (1999) and many others. In what follows, I will discuss all parts in the light of the question whether they are postulated as part of innate UG. I will then show that alternatives are possible: either the kind of knowledge that is postulated is not unique to language (so that it might be part of "universal cognition" but not of "universal grammar") or could be acquired on the basis of experience. The goal should be clear: to minimize and thus clarify the role of innateness, taking seriously the approach advocated by Culicover (2016). In Culicover's view, linguists should make a serious effort to explain facts of language from a non-universalist point of view, if only to isolate the aspects that cannot be so explained – and in the process we will always learn more about language. This is also consistent with the methodological point made above. So this is my view on the future of Optimality Theory: to bring diverse approaches closer together by establishing common ground and identifying the general, cognitive mechanisms that may underlie this theory. If the results indicate some aspects of language are universal, either specific to language or as part of general cognition, we make real progress.

3. Inputs and outputs

First, it is important to establish that OT is a generative theory, where "generative" is meant in the sense of derivational, i.e., having (at least) two levels: an underlying "abstract", "lexical" level, consisting of inputs (which I will take to be roughly equivalent to "underlying forms" in generative phonology in general), which by a specific mechanism are converted into output ("surface", "phonetic") forms. In this respect, OT falls squarely within the SPE tradition (Chomsky & Halle 1968), although of course OT uses a selection mechanism (based on Con and Eval) instead of rules that change inputs into outputs directly (perhaps for some thorny phenomena rules might still be invoked, see McCarthy (1993), while newer variants of OT, such as harmonic serialism (McCarthy 2011; Wolf 2011), seem to re-institute rule (ordering) in a roundabout way). Since OT does not employ rules, it could be regarded as a non-derivational generative theory.

It is a basic assumption in linguistics that all languages share the same general architecture; that is, it is not the case that some languages use an SPE-style rule grammar, while others employ OT. Thus this basic architecture holds for all languages and is therefore, by assumption, part of Universal Grammar. Is it also innate? And is the two-level structure specific to language? Why are there two levels and not more (or fewer)?

In my view, the basic structure of OT is not innate, but is the result of more general cognitive strategies which shape our knowledge and processing of language. It is also not specific to language. This general cognitive strategy is **categorization**, a mental process by which concepts and things are recognized, differentiated and understood on the basis of sensory data (e.g., see Eysenck & Keane 2000; Cohen & Lefebvre 2005 and, recently, Feldman Barrett 2017). Thus, categorization refers to the process by which we recognize a bicycle as a bicycle when we see one, the process of recognizing by which we recognize two shades of red as the "same" colour, or blue and green as different colours. It also refers to the process by which we recognize one person as the same person (or anything else) regardless of the time of day, which side of their face we see, what clothes they wear, etc. Once categories are in place, they can be used to recognise new sensory input. Although the exact mechanism of categorization is subject to debate, it is vital in all mental processes. Lakoff (1987, 5) wrote: "Categorization is not a matter to be taken lightly. There is nothing more basic than categorization to our thought, perception, action, and speech" and Harnad (2005) famously quipped "cognition **is** categorization". (Note that categorization is not a uniquely human ability, see e.g., Zentall et al. 2002; Smith et al. 2012). The importance of categorization for phonology, and our task at hand of understanding the relation between inputs and outputs, should be obvious. Categorization is involved when we recognize a newly heard word as one that we have heard before and know. It is involved when we correctly recognize a word regardless of the time of day, the speed and style at which it is pronounced, whether it is pronounced by a male or a female, etc. And it is involved when we hear a new word, or a particular pronunciation that we cannot recognize. By the same token, categorization can also be applied to the level of the segment: speakers of English categorize all manners of phonetic [t] as a "phoneme" /t/: aspirated [t^h] in *towel*, unaspirated in *Stowell*, with inaudible release in *Atkins* and perhaps flapped (depending on the dialect) in *mighty*. The process is completely parallel to that by which we recognize, or classify, different kinds of bicycles as all belonging to the "category" of bicycle.

How does categorization take place? This is a major issue in psychology, and two theories have been developed (see again the references above): Prototype Theory and Exemplar Theory. Both operate on the basis of experienced data, and the differences need not concern us here (the existence of two competing theories in this area suggests that they might be fruitfully combined, for a proposal with respect to phonology in this respect, see Sloos 2013). In Prototype Theory, a typical image is constructed, on the basis of experienced tokens (let's say, of bicycles). New tokens of experience (new bicycles that are seen) are compared to such mental images and categorized as such if the new token matches the mental image sufficiently closely. This is quite comparable to the way in which words or sounds could be recognized. In Exemplar Theory, category judgements are made by comparing a new instance with stored memories for other instances of the category. That is, previous experiences of bicycles are stored in memory, and new instances are compared to the whole collection of tokens (sometimes referred to as "clouds") at once. The memories of previously stored tokens are subject to decay (you will not typically remember whether you have seen a particular bicycle before or not), but every experience strengthens its particular category (the more bicycles you see, the more familiar they are – in other words, the stronger the category is). There is an obvious relation with frequency here, which also plays an important role in language: stronger categories (e.g., words that are used very frequently) are more firmly entrenched in memory than words that are seldom used. Exemplar Theory is rooted in cognitive psychology (Goldinger et al. 1991; Medin & Schaffer 1978; Nosofsky 1988) and has been applied to language by Johnson (1997); Pierrehumbert (2001); Bybee (2006; 2012); Wedel (2007), and others. Apparently, Exemplar Theory allows for faster category judgements than Prototype Theory (Eysenck & Keane 2000), but the debate between the two theories (or rather: two families of theories, because there are different variants of both Prototype Theory and Exemplar Theory) is by no means settled.

The relation between category (a unique mental image, either as a prototype or as a collection of experienced tokens in memory) and instances in the "real world" (existing tokens, e.g., bicycles) is quite the same as the relation between inputs (underlying forms) and phonetic realisations of those inputs. The latter is illustrated in Figure 2, where different phonetic outputs (instances in the real world) are mapped onto an input (the "mental image"):

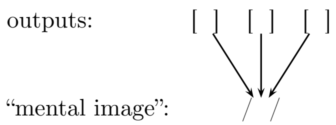

Figure 2: Outputs and a mental image

Thus, the idea of 'something abstract in the mind' (in cognition: a concept or a category) (in OT: inputs) which corresponds to 'real instances in the real-world' (tokens, pronounced words; in OT: outputs) is not limited to language. The OT architecture (or in general the concept in generative linguistics of 'two levels') applies much more widely than just in phonology (or in language in general), where we should note that this distinction is much clearer for the relation between 'deep structure' and 'surface structure' in phonology than between the relation between the two in syntax: after all, in syntax sentences are not understood on the basis of previously-heard tokens (also this might well play a role in early acquisition); unlike in phonology, syntactic parsing takes place.

Also relevant here is the common assumption that is made in (generative) linguistics that each morpheme (meaning-bearing unit) has a unique underlying form (we might add: unless they don't, such as in cases of suppletive allomorphy (e.g., Nevins 2011). Thus, one mental unit corresponds to (i.e., is recognized as, or is realized as) a multitude of (phonetic) events in the real world. Note that the mental unit here is quite distinct from an underlying representation envisaged in SPE or Prince & Smolensky (2004 [1993]): instead of a plain underlying form (from which all predictable information is preferably stripped, as in underspecification theories), the mental image may have (rich, i.e., varied, including lots of information that would be considered redundant in a strict generative approach) internal structure (especially in an Exemplar approach), and is also related to other forms in the mental lexicon (structured like a network).

The idea of two simple levels: an abstract category and concrete instances, must be one of the reasons why OT appealed so much as an alternative to SPE and related models: instead of recognizing numerous intermediate (abstract) levels (as in Lexical Phonology (Kiparsky 1982), Derivational OT (Rubach 2000) or Functional Phonology (Boersma 1998)), outputs are directly mapped onto inputs (in perception) and vice versa (in speech production). It is also related to the abstractness debate (see e.g., Kiparsky 1968), which has raged in generative linguistics since underlying forms were first proposed. The two theories of categorization mentioned above could help to steer this debate, if underlying forms are either prototypes or collections of experienced tokens. Of course, this observation alone does not settle this discussion: the question remains whether underlying forms in phonology could be profitably understood as prototypes or "average" kind of pronunciations (Prototype Theory), or as collections of heard forms (Exemplar Theory) (see above). The work cited above in which Exemplar Theory is applied to linguistics suggests that these approaches can shed new light on many linguistic phenomena. For instance, they give an excellent perspective on variation in language (different variants will be recorded and stored in an Exemplar approach), which has been a thorny issue ever since Chomsky assumed (for good methodological reasons) a "homogeneous speech community" and "an ideal speaker-hearer" (Chomsky 1965). Secondly, such richer representations provide a useful approach to frequency effects in language, which abound in phonology, morphology (both synchronic and diachronic) and sociolinguistics (see e.g., the work by Bybee such as that cited above, Haspelmath 2002; Sloos 2013, and references cited there). Finally, the enrichment provides a sorely needed handle on perception: the standard OT model is first and foremost a speech production mechanism, a theory about how a particular input form will be realized (on the basis of the constraint hierarchy): it tells us nothing about how forms are perceived (whether correctly or not). This has instigated important research such as that by Boersma (1998) (et seq.). The inclusion of exemplar-type representations could be an alternative or complementary way of approaching this issue. Turning this around, Exemplar Theory does not say how a particular form will be pronounced: there is no algorithm for picking out a particular token in an exemplar cloud to utter in a particular speech context. This is where OT could provide an explicit mechanism. We will return to this possibility in the next section.

Let's end this section by asking how we should assess the clear parallel between the input-output model of OT and the cognitive concept of categorization. Three possibilities come to mind. First, we could take this parallel as an external piece of motivation for OT, and nothing else. The OT architecture might still be a language-specific module, which may have been motivated by categorization some time during evolution, but which has now become innate, with an architecture specific or language. This is a possible viewpoint, but it would ignore the parallels between language processing and other kinds of general cognitive abilities humans have. Second, we could regard these ideas an invitation to combine OT with theories like categorization such as Exemplar Theory (as advocated in van de Weijer 2009; 2012), which will have benefits for both. Finally, and most definitely, it should be taken as an encouragement to study categorization further, as a general psychological ability, and add to our knowledge of this phenomenon using our specific expertise as linguists.

With this in mind, let us turn to the second piece of architecture of OT, Gen.

4. The Generator

"Gen is a function that, when applied to some input, produces a set of candidates, all of which are logically possible analyses of this input", (Kager 1999, 19), see Figure 3. The only true restriction on Gen is that output candidates should consist of legitimate elements from the alphabets of linguistic representation such as segmental structure (e.g., features, or Dependency Phonology style elements) and prosodic structure (syllable structure and higher prosodic structure). OT is neutral as to which theories of phonological representation (features, elements, syllable structure, moras) are selected, although some might work better than others (see e.g., van Oostendorp & van de Weijer 2005). This means that, for all intents

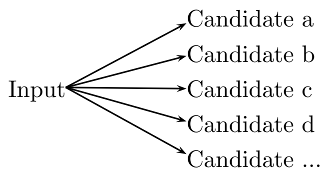

Figure 3: Gen (Kager 1999, 22)

and purposes, in classic Optimality Theory the candidate set is infinite, so it is also identical for any given input, even across languages (assuming that all languages make use of the same representational alphabets). More recent developments in OT have tried to improve on this point (e.g., harmonic serialism; see references above).

Gen is quite easily the most striking technical innovation in OT. Fears have been raised that choosing from an infinite set of possible outputs might be computationally intractable, but this fear is unfounded (Kager 1999, 25–27). Still, from a psycholinguistic viewpoint, a mechanism like Gen seems wildly implausible. When pronouncing a word, it might make sense to select from a number of possible pronunciations, but it would not make sense to consider an infinite set. Since Gen is part of the basic OT architecture, it would also be considered innate.

In my view, Gen is not innate, but the result of general cognitive strategies. The output of Gen (a rich array of candidate word forms) can be compared to the memory representations of categorization theories like Exemplar Theory (see above), which are clouds of previously heard tokens (subject to memory decay). This means that all representations of variation, contexts, pragmatic circumstances are stored when a particular word is heard, strengthening words that are heard often. Although these clouds are units, they specify much information on variation. That such rich levels of information are indeed stored is apparent from experiments and from common knowledge: speakers of a language have no difficulty recognizing dozens or hundreds of voices (or e.g., popular songs) that are familiar to them, including ones that they may not have heard for decades (Bartlett & Snelus 1980; Schulkind et al. 1999). This means that such information is available to the speaker, so it must be part of memory.

Before returning to Gen, let us examine the Exemplar view on memory representations a bit further, to identify more advantages that this theory might bring. Eysenck and Keane (2000, 320) provide the following key points:

(1) – Categories are made up of a collection of instances or exemplars rather than any abstract description of these instances.
– Instances are grouped relative to one another by some similarity metric.
– Categorization and other phenomena are explained by a mechanism that retrieves instances from memory given a particular cue.
– When exact matches are not found in memory the nearest neighbour to the cue is usually retrieved.

While these points leave a number of questions unresolved (e.g., what counts as 'similar' according to the "similarity metric"? How does the "mechanism" of categorization work exactly?), it is obvious that these key points apply forcefully in language, too. Above the advantages for accounts of variation and frequency effects were already mentioned. We can add to this that the exemplar clouds are arranged in a network, so that clusters of clouds may evolve. And clearly, words in the mental lexicon are related. Here we can think of rhyme, for instance (words that rhyme will be stored close together, allowing for faster retrieval (and better memory)), semantic fields (words that pertain to a particular topic will be stored close together, allowing for faster retrieval in e.g., psycholinguistic priming experiments) and morphological paradigms (words that are inflected forms of another word will be stored, if they are frequent enough, close to uninflected words, giving rise to paradigm uniformity effects (see e.g., Hall 2005; Kenstowicz 2005).

If such detail-rich clouds are part of memory, Gen becomes superfluous. The infinite candidate set does not have to be generated: relevant candidates are already there! In other words, we can use the tokens of experience as possible pronunciations for a given concept. This view is more in line with a traditional view on speech processing: a speaker intends to convey a particular meaning, so exemplar clouds are first and foremost units of meaning. In this cloud, the speaker will have at their disposal many possible pronunciations. Exemplar Theory does not specify how a particular pronunciation form is selected: this could still be achieved by an OT-type grammar. At any rate, it is not likely that an elaborate rule system (with intermediate levels) is still necessary: words that are pronounced are more or less ready-made, and must only be adapted to the particular style or linguistic and pragmatic circumstances (sentence position, style) that the speaker is faced with. It is striking that a move like this is foreshadowed in OT itself, witness the almost complete disappearance of underspecification (Archangeli 1984; Steriade 1987) in OT input forms (see some discussion in Itô et al. 1995; Artstein 1998): OT input forms are also more or less ready-made forms, ready for pronunciation.

Finally, one might object at this point that the idea of exemplars is quite similar to the infinite candidate set of OT. But this is not the case: the candidate set is theoretically infinite, while tokens in an exemplar cloud are based on experience and subject to decay. Moreover, the exemplar model has a wider application than just language (it is part of a general cognitive theory), and has been supported by empirical evidence.

Let us now turn from architecture to substance and consider the constraint set.

5. Con

While the above two parts of OT were closely related to the architecture of the theory, we now turn to the substantive part, Con and Eval. Con is the set of constraints, which is putatively universal as well. From one point of view, this is not surprising: speakers all have the same speech organs (including mouths and ears), so it makes sense that a concept like markedness ("easier structures are preferred over more complex ones") plays a role in all languages. Different authors have therefore tried to derive markedness constraints, e.g., by grounding them in phonetic concepts like ease of articulation (Hayes 1999). This raises the question whether it makes sense to include such constraints in an innate capacity (which would more likely include things that are difficult to learn), or whether it represents a step on the way to an experience-based account of constraints. It then raises the question whether other constraints (faithfulness, alignment) could be similarly derived.

With respect to the universality/innateness of the constraint set, consider the telling quote below from Tesar & Smolensky (2000, Ch. 2, fn. 1):

> "The set Con of constraints is universal: the same constraints are present in all languages. The simplest interpretation of this is that the constraints are innately specified, but that is not required by the theory itself: OT only requires that Con be universal. Neither the authors nor their research programs are committed to the innateness of Con; however, all the work in this book presumes that the constraints in Con are available to the learner as early as could possibly be useful. Obviously, any alternative theory would have to say more about how the constraints are developed, as well as how universality is to be maintained."

Here we question whether the idea that the constraints are "innately specified" is the simplest interpretation (we will also see below that constraints are not necessarily universal). Keeping in mind that innate mechanisms should not be postulated unless an experience-based interpretation has turned out not to be feasible (cf. above), it would be proper to investigate whether constraints could or could not be acquired during acquisition. Adding to this, assuming an innate set of constraints would seem to obviate the need to ask important questions about the constraint set: the crude assumption of innateness bars deeper empirical investigation (see again Culicover 2016; Davis et al. 2002).

It turns out not to be difficult at all to derive constraints from data. If constraints are not universal, or “innately specified”, then they must somehow be derived from the language a child is exposed to in the course of acquisition (on the assumption that the grammar of a child’s L1 arises during this period, which seems reasonable enough). In this respect therefore child-directed speech plays an important role (as well as any other speech the child is exposed to, such as that between others). Here I draw on an argument also made in van de Weijer (2014; 2017) (see also van de Weijer & Sloos 2013; van de Weijer & Tzakosta 2017). Children can derive most constraints from the speech data they are exposed to. Consider the constraint *Complex, which clearly characterizes child speech production, in which many clusters are reduced or simplified in other ways. This is true in many languages; the examples in (2) are taken from two English-learning children, Amahl and Gitanjali, data is taken from Johnson & Reimers (2010):

(2)	Amahl (Smith 1973)		Gitanjali (Gnanadesikan 2004)	
	[b̥ei]	play	[kin]	clean
	[b̥u]	blue	[piz]	please
	[g̊in]	clean	[fen]	friend
	[g̊ai]	sky	[dɔ]	straw
	[b̥ɔt]	sport	[gin]	skin
	[b̥un]	spoon	[bun]	spoon

Let us, for the sake of the argument, assume that realizations like those above are (at least partly) due to the grammar that the child is constructing. In such a grammar the markedness constraint *Complex will be relatively high-ranked (above a faithfulness constraint forbidding deletion of segments). Gnanadesikan 2004) and others have argued that facts like these argue for the innateness of a constraint like *Complex, because adult language violates this constraint (witness words like *play, blue*, etc.) so the question is how children could have learned this constraint on the basis of ambient data. It turns out that children could have learned this constraint on the basis of data they are most frequently exposed to. Consider the 150 most frequent words in child-directed speech, taken from the

Childes data collection (MacWhinney 2000) (note that some words here might not be counted as words by other researchers; this point is irrelevant here):

(3) you, the, it, a, and, to, I, that, oh, on, what, in, we, is, do, are, no, there, that's, your, it's, don't, of, one, going, have, this, well, can, not, right, he, yes, like, now, got, think, go, with, look, put, then, for, you're, they, just, was, all, want, up, some, she, see, Thomas, at, get, be, me, them, because, what's, isn't, did, know, out, but, come, here, little, so, her, if, very, when, you've, there's, I'm, down, didn't, shall, yeah, okay, Mummy, big, nice, good, back, bit, he's, can't, about, where, off, would, say, had, my, were, or, his, two, does, they're, more, him, doing, *please*, I'll, doesn't, we've, these, haven't, aren't, has, as, dear, why, let's, again, have_to, over, we'll, those, make, she's, *from*, gone, will, *play*, need, take, really, an, how, I've, hmm, car, other, another, Daddy, round, been, who, sit, eat, where's, time, something, alright, too

The data show that **none** of the 100 most frequent words have an onset cluster, and only three of the 150 most frequent ones: *please, from*, and *play*. We know that children pay attention to statistical regularities in the (linguistic and non-linguistic) data they are exposed to (e.g., Saffran 2001; Dawson & Gerken 2011, etc.); this is, for instance, how they learn word boundaries in the first place. Faced with frequent data like that in (3), the conclusion is warranted that children will generalize across them, resulting in a "constraint" which says exactly the same as the OT constraint *Complex: No onset clusters (or, positively formulated: consonant clusters are rare). Of course, as children are exposed to and acquire more and more words (including less frequent ones), they will gradually find out that clusters are not so rare (see the next section for discussion).

Several points can be made here. Frequency is not a source of information that is generally used in generative accounts (like OT), but for this argument to work, such kind of information is of course crucial. Research in the past decades has shown that there are many linguistic areas, both in grammar and use of language, where frequency is of vital importance (see e.g., references above). The omission to give frequency a role in all aspects of grammar should be regarded as a grave error. We have seen above that frequency is a natural consequence of adopting an Exemplar approach to categorization. Secondly, the objection might be raised that the "real" question is **why** English apparently has so few words with clusters among its frequent words (like those in (3)). I would say that this is indeed an interesting question, but that it is fundamentally a different question than we are trying to answer here, which is how a child could learn an OT constraint during acquisition. The question why a language is shaped as it is, has many answers, in which certainly markedness factors (such as ease of articulation) but also other factors (borrowing, language contact, etc.) play a role. Thirdly, it should be pointed out that under this account, constraints will not be universal, in spite of Tesar and Smolensky's (2000) demand (see above): constraints will only be posited in cases where the data allow a generalization. In a language without, say, prenasalized stops (like [^{m}b], a "marked" kind of segment), a child will not discover a generalization that prenasalized stops are rare. I claim this is an advantage, rather than a disadvantage, in that it disallows spurious, inactive constraints in the child's grammar. Finally, this account relies on the child's ability to keep track of frequent and infrequent data in its environment, a skill for which there is ample support (both with respect to language and non-language data) and to generalize across such data. Thus, here again, a language-specific, putatively innate part of OT (the constraint set Con) can be replaced by a very general cognitive mechanism.

In a paper on the acquisition of French, van de Weijer and Sloos (2013) show that a number of markedness constraints that play a role in L1 child French can be similarly derived. Two questions remain: how about other constraints, especially OT constraints of the faithfulness and alignment family (Kager 1999, chapters 1–3)? Secondly, does ranking of constraints (Eval) still play a role, and if so, how is it derived? Before turning to the second question in the next section, let's consider the first. Faithfulness constraints militate against deleting or inserting segments, or changing feature values. Such constraints have a lesser role to play in an approach to grammar that makes use of Exemplar-style representations, because words that are related morphologically are all included in the mental lexicon (if they are used frequently enough). Thus, alternations are simply encoded in lexical forms (e.g., Dutch [hɔnt] 'dog' vs. [hɔndə] 'dogs'): the role of grammar is diminished (see also Ernestus & Baayen 2006). Morphology, as well as alignment constraints, can be learned as well: these are learned in exactly the same way as constraints: if there is a sufficiently large number of forms with a specific meaning and a specific sequence of sounds, such a sequence will be recognized as a morpheme. A morpheme boundary may be posited in such a case and if morphemes appear in particular, sufficiently frequent, contexts, patterns captured by alignment constraints may be deduced (van de Weijer 2012).

Let us finally look at constraint ranking in the next section.

6. Eval

Eval here refers to the mechanism by which constraints are ranked. In classic OT, ranking takes place according to a ranking algorithm, the Constraint Demotion Algorithm (Kager 1999; Tesar & Smolensky 2000), for which some alternatives have been proposed (see especially Boersma 2000; Boersma & Hayes 2001). For our purpose it is important that the ranking is established during acquisition, on the basis of the ambient data. So although the initial ranking may be innate (Kager 1999, 298), the ranking is brought about by linguistic experience.

If, as we suggested above, the constraint set is acquired, the question how these constraints become ranked, gains in significance. Several approaches come to mind here. First, we might simply follow standard OT here and assume that constraints, once they are formed based on generalization as described in the previous section, may become reranked in accordance with OT's Constraint Demotion Algorithm or a related mechanism (Tesar & Smolensky 2000). There are at least two alternatives, however. The first is that constraints may appear and disappear as a larger portion of the language is acquired by the language-learning child. We saw above that onset clusters are extremely uncommon in the most frequent words in English. When a larger number of words are considered, however, the frequency of clusters rises (as expected). Consider the following data for Greek (taken from van de Weijer & Tzakosta 2017, based on data from Protopapas et al. 2012):

Table 1: The relation between word frequency and clusters in Greek

Most frequent words	Words with clusters	Percentage
100	21	21%
200	47	24%
300	93	31%
400	128	32%
500	177	35%
600	214	36%
700	257	37%
800	309	39%
900	348	39%
1000	395	40%

Table 1 shows that, as more and more of the language is known, the constraint against clusters becomes weaker and weaker, up to a point that words with clusters are almost as common as words without (see also Table 1 in van de Weijer & Sloos 2013, 194, which shows the same tendency for some, but, interestingly, not all of the seven markedness constraints examined there). On the assumption that there is a rough correlation between words that are frequent and words that are acquired early on in acquisition, this means that a constraint against clusters will be strong in the initial stage of acquisition and lose force later on. Thus, the constraint might remove itself from grammar in a natural way (or play a smaller role, accounting for "emergence of the unmarked" effects (Kager 1999)).

A second possible approach is that, at any stage of acquisition, some constraints are stronger than other ones, where "strong" is directly related to the number of counterexamples that there are against it. This is the approach taken by van de Weijer and Sloos (2013) for French. For instance, when comparing two markedness constraints, there may be 5 or 15 words (out of the 100 most frequent) that do not conform to these two markedness generalizations, respectively. Both generalizations may be discovered by the child (since both 5 and 15 are significantly less than 50), but the former generalization is obviously stronger than the latter. We could relate strength of generalization to constraint position in the hierarchy. A point to make here, however, is that two markedness constraints (such as the ones discussed by van de Weijer and Sloos 2013) do not conflict. Recognizing this, van de Weijer and Sloos (2013) relate the strength of generalizations (i.e., constraints) to age of acquisition (constraints which are less strong, i.e., have more counterexamples in the most frequent words, are expected to be lost earlier, i.e., the marked segments or structures against which they militate are expected to be acquired relatively early), and find a very good fit between the two (for more details, cf. van de Weijer & Sloos 2013). This means that constraints may not only be acquired (rather than stipulated to be innate), an approach that derives them from input data makes extra, interesting predictions that seem to be borne out. Of course, the approach in question should be tested with data from more languages and in different stages. An approach to L2 acquisition might also be developed in the same manner, as well as an approach to L1–L2 interference.

To conclude, the idea of deriving constraints not only looks feasible, but also promising as an approach to order of acquisition.

7. Conclusion

In this article I have outlined a cognitive approach to Optimality Theory. By assessing the innateness of different parts of the theory, we found alternatives that made it possible to rely on general cognitive mechanisms rather than on stipulations of innateness. It is hoped that this approach inspires further research on this topic. OT has had tremendous success in dealing with scores of linguistic phenomena (mostly, but not exclusively, in phonology); if we can understand its cognitive basis better, our understanding of the human language capacity will definitely benefit.

Acknowledgements

Many thanks to Marjoleine Sloos as well as the students of the SISU "Optimality and Beyond" class, who commented generously on previous versions and presentations of this paper. I'm also very grateful to two anonymous *Acta Linguistica Academica* reviewers who helped me to express a number of points more clearly. The usual disclaimers apply.

References

Archangeli, Diana. 1984. Underspecification in Yawelmani phonology. Doctoral dissertation. MIT.

Archangeli, Diana and Douglas Pulleyblank. 2015. Phonology without Universal Grammar. Frontiers of Psychology 6. 12–29.

Artstein, Ron. 1998. The incompatibility of underspecification and markedness in Optimality Theory. In R. Artstein and M. Holler (eds.) Ruling papers 1: Working papers from Rutgers University. New Brunswick, NJ: Rutgers University. 7–13.

Bartlett, James C. and Paul Snelus. 1980. Lifespan memory for popular songs. The American Journal of Psychology 93. 551–560.

Boersma, Paul. 1998. Functional Phonology: Formalizing the interactions between articulatory and perceptual drives. Doctoral dissertation. University of Amsterdam. (Published by Holland Academic Graphics, The Hague).

Boersma, Paul. 2000. Learning a grammar in Functional Phonology. In J. Dekkers, F. R. H. van der Leeuw and J. van de Weijer (eds.) Optimality theory: Phonology, syntax, and acquisition. Oxford: Oxford University Press. 465–523.

Boersma, Paul and Bruce Hayes. 2001. Empirical tests of the Gradual Learning Algorithm. Linguistic Inquiry 32. 45–86.

Bybee, Joan L. 2006. From usage to grammar: The mind's response to repetition. Language 82. 711–733.

Bybee, Joan L. 2012. Usage-based theory and exemplar representations of constructions. In G. Trousdale and T. Hoffmann (eds.) The Oxford handbook of construction grammar. Oxford: Oxford University Press. 49–69.

Chomsky, Noam. 1959. A review of B. F. Skinner's *Verbal behavior*. Language 35. 26–58.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Chomsky, Noam. 1980. Rules and representations. New York: Columbia University Press.

Chomsky, Noam and Morris Halle. 1968. The sound pattern of English. New York: Harper & Row.

Cohen, Henri and Claire Lefebvre (eds.). 2005. Handbook of categorization in cognitive science. Oxford: Elsevier.

Culicover, Peter W. 2016. Foreword. In M. H. Christiansen and N. Chater (eds.) Creating language – Integrating evolution, acquisition, and processing. Cambridge, MA: MIT Press. vii–x.

Davis, Barbara L., Peter F. MacNeilage and Christine L. Matyear. 2002. Acquisition of serial complexity in speech production: A comparison of phonetic and phonological approaches to first word production. Phonetica 59. 75–107.

Dawson, Colin and LouAnn Gerken. 2011. When global structure "explains away" local grammar: A Bayesian account of rule-induction in tone sequences. Cognition 120. 350–359.

Dresher, B. Elan and Norbert Hornstein. 1976. On some supposed contributions of artificial intelligence to the scientific study of language. Cognition 4. 321–378.

Ernestus, Mirjam and R. Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: The case of past-tense formation. In L. Goldstein, D. H. Whalen and C. T. Best (eds.) Laboratory phonology 8. Berlin & New York: Mouton de Gruyter. 27–49.

Eysenck, Michael W. and Mark T. Keane. 2000. Cognitive psychology: A student's handbook (Fourth ed.). Hove: Psychology Press.

Feldman Barrett, Lisa. 2017. How emotions are made: The secret life of the brain. Boston & New York: Houghton Mifflin Harcourt.

Gnanadesikan, Amalia E. 2004. Markedness and faithfulness constraints in child phonology. In R. Kager, J. Pater and W. Zonneveld (eds.) Constraints in phonological acquisition. Cambridge: Cambridge University Press. 73–108.

Goldinger, Stephen D., David B. Pisoni and John S. Logan. 1991. On the nature of talker variability effects on recall of spoken word lists. Journal of Experimental Psychology: Learning, Memory, and Cognition 17. 152–162.

Hall, Tracy A. 2005. Paradigm uniformity effects in German phonology. Journal of Germanic Linguistics 17. 225–264.

Harnad, Stevan. 2005. To cognize is to categorize: Cognition is categorization. In Cohen & Lefebvre (2005, 20–45).

Harris, Roy. 1983. F. de Saussure: Course in general linguistics. London: Duckworth.

Haspelmath, Martin. 2002. Understanding morphology. London: Arnold.

Hauser, Marc D., Noam Chomsky and W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? Science 298. 1569–1579.

Hayes, Bruce. 1999. Phonetically-driven phonology: The role of optimality theory and inductive grounding. In M. Darnell, E. Moravcsik, M. Noonan, F. Newmeyer and K. Wheatly (eds.) Functionalism and formalism in linguistics volume I: General papers. Amsterdam & Philadelphia: John Benjamins. 243–285.

Hulst, Harry van der. in preparation. An introduction to the nature and nurture of language. Cambridge: Cambridge University Press.

Itô, Junko, Armin Mester and Jaye Padgett. 1995. Licensing and underspecification in Optimality Theory. Linguistic Inquiry 26. 571–613.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In K. Johnson and J. W. Mullennix (eds.) Talker variability in speech perception. San Diego, CA: Academic Press. 145–166.

Johnson, Wyn and Paula Reimers. 2010. Patterns in child phonology. Edinburgh: Edinburgh University Press.

Kager, René. 1999. Optimality theory. Cambridge: Cambridge University Press.

Kenstowicz, Michael. 2005. Paradigmatic uniformity and contrast. In L. Downing, T. A. Hall and R. Raffelsiefen (eds.) Paradigms in phonological theory. Oxford: Oxford University Press. 145–69.

Kiparsky, Paul. 1968. How abstract is phonology? Bloomington, IN: Bloomington University Linguistics Club.

Kiparsky, Paul. 1982. Lexical morphology and phonology. In I.-S. Yang (ed.) Linguistics in the morning calm: Selected papers from SICOL-1981. Seoul: Hanshin Publishing Company. 3–91.

Lakoff, George. 1987. Women, fire, and dangerous things: What categories reveal about the mind. Chicago: The University of Chicago Press.

MacWhinney, Brian. 2000. The CHILDES Project: Tools for analyzing talk (Third edition). Mahwah, NJ: Lawrence Erlbaum.

McCarthy, John J. 1993. A case of surface constraint violation. Canadian Journal of Linguistics/La revue canadienne de linguistique 38. 169–195.

McCarthy, John J. 2011. Perceptually grounded faithfulness in Harmonic Serialism. Linguistic Inquiry 42. 171–183.

Medin, Douglas L. and Marguerite M. Schaffer. 1978. A context theory of classification learning. Psychological Review 85. 207–238.

Nevins, Andrew. 2011. Multiple agree with clitics: Person complementarity vs. omnivorous number. Natural Language & Linguistic Theory 29. 939–971.

Nosofsky, Robert M. 1988. Similarity, frequency, and category representations. Journal of Experimental Psychology: Learning, Memory, and Cognition 14. 54–65.

Oostendorp, Marc van and Jeroen van de Weijer. 2005. Phonological alphabets and the structure of the segment. In M. van Oostendorp and J. van de Weijer (eds.) The internal organization of phonological segments. Berlin & New York: Mouton de Gruyter. 1–23.

Pierrehumbert, Janet. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In J. L. Bybee and P. Hopper (eds.) Frequency and the emergence of linguistic structure. Amsterdam & Philadelphia: John Benjamins. 137–157.

Prince, Alan and Paul Smolensky. 2004 [1993]. Optimality Theory: Constraint interaction in Generative Grammar. Malden, MA & Oxford: Blackwell.

Prinz, Jesse J. 2012. Beyond human nature: How culture and experience shape the human mind. New York: Norton.

Protopapas, Athanassios, Marina Tzakosta, Aimilios Chalamandaris and Pirros Tsiakoulis. 2012. IPLR: An online resource for Greek word-level and sublexical information. Language Resources and Evaluation 46. 449–459.

Rubach, Jerzy. 2000. Glide and glottal stop insertion in Slavic languages: A DOT analysis. Linguistic Inquiry 31. 271–317.

Saffran, Jenny R. 2001. Words in a sea of sounds: The output of statistical learning. Cognition 81. 149–69.

Saussure, Ferdinand de. 1916. Cours de linguistique générale. Paris: Payot.

Schulkind, Matthew D., Laura Kate Hennis and David C. Rubin. 1999. Music, emotion, and autobiographical memory: They're playing your song. Memory & Cognition 27. 948–955.

Sloos, Marjoleine. 2013. Phonological grammar and frequency: An integrated approach. Evidence from German, Indonesian, and Japanese. Doctoral dissertation. University of Groningen.

Smith, J. David, Mark E. Berg, Robert G. Cook, Matthew S. Murphy, Matthew J. Crossley, Joseph Boomer, Brian Spiering, Michael J. Beran, Barbara A. Church, F. Gregory Ashby and Randolph C. Grace. 2012. Implicit and explicit categorization: A tale of four species. Neuroscience & Biobehavioral Reviews 36. 2355–2369.

Smith, Neil V. 1973. The acquisition of phonology: A case study. Cambridge: Cambridge University Press.

Smolensky, Paul and Géraldine Legendre (eds.). 2006. The harmonic mind: From neural computation to Optimality-Theoretic grammar. Cambridge, MA: MIT Press.

Steriade, Donca. 1987. Redundant values. In A. Bosch, B. Need and E. Schiller (eds.) CLS 23-II: Papers from the Parasession on Autosegmental and Metrical Phonology. Chicago: Chicago Linguistic Society. 339–362.

Tesar, Bruce and Paul Smolensky. 2000. Learnability in Optimality Theory. Cambridge, MA: MIT Press.

Thomas, Margaret. 2004. Universal Grammar in second language acquisition. London & New York: Routledge.

Wedel, Andrew B. 2007. Feedback and regularity in the lexicon. Phonology 24. 147–185.

Weijer, Jeroen van de. 2009. Optimality Theory and Exemplar Theory. Phonological Studies 12. 117–124.

Weijer, Jeroen van de. 2012. Grammar as selection: Combining Optimality Theory and Exemplar Theory. Nagoya: Kougaku Shuppan.

Weijer, Jeroen van de. 2014. The origin of OT constraints. Lingua 142. 66–75.

Weijer, Jeroen van de. 2017. Emergent phonological constraints: The acquisition of *Complex in English. Acta Linguistica Academica 64. 153–165.

Weijer, Jeroen van de and Marjoleine Sloos. 2013. Learning markedness constraints: The case of French. In S. Aalberse and A. Auer (eds.) Linguistics in the Netherlands 2013. Amsterdam & Philadelphia: John Benjamins. 188–200.

Weijer, Jeroen van de and Marina Tzakosta. 2017. The status of *Complex in Greek. In T. Georgakopoulos (ed.) Proceedings of the Twelfth International Conference on Greek Linguistics, vol. 2. Berlin: Edition Romiosini & Center for Modern Greece. 1259–1274.

Wolf, Matthew. 2011. Limits on global rules in Optimality Theory with candidate chains. Phonology 28. 87–128.

Zentall, Thomas R., Mark Galizio and Thomas S. Critchfield. 2002. Categorization, concept learning, and behavior analysis: An introduction. Journal of the Experimental Analysis of Behavior 78. 237–248.