Chris Golston and Richard Wiese[1]

The structure of the German root

> ‘Nur das Geheimnis der Wurzeln oder des Benennungsgrundes der Urbegriffe lassen wir unangetastet.’ (Bopp 1833, III)

0. Goals

The goal of the present study is to make the notion of the root more accessible to morphology and phonology by looking at a large corpus of roots and describing the frequency of different types of root in a constraint-based grammar.

Our research is based on a computer-readable data base of German roots compiled by Ortmann (1993) from ten different sources. We identify a number of phonological properties of roots and demonstrate that unmarked root shapes are much more common than marked ones. To base the research in a theoretical framework we use Direct OT (Golston 1996b), according to which underlying forms are represented solely in terms of markedness, as sets of constraint violations. Our goal is to show that phonologically unmarked roots are more common than marked ones.

1. Morphemes and roots

Every serious theory of language is based on morphemes. This is true even of theories that explicitly deny them. We know that aside from a grammar a language learner must learn a large number of arbitrary pairings of signals-to-meanings. The overwhelming majority of such pairings are of the familiar type in (1):

(1) *Typical morpheme*

+<=> [kæt]

[1] The present paper has benefitted from discussions with many colleagues, in particular at the universities of Arizona, Düsseldorf, Jena, and Marburg. Wolfgang Kehrein’s comments on the paper have been very useful. The invaluable help of Carmen Scherer, Marburg, in the work with the data base deserves a special mention.

The classic definition runs something like this: a morpheme is 'eine Gruppe von ein bis mehreren Lauten mit einem bestimmten Bedeutungskern' [a grouping of one to several sounds with a specific semantic core] (Liebich 1899, 5). We have a more refined notion of all this now, with morpheme tiers and floating features, but the essential insight is the same: some signal is related to a meaning. We can quibble over whether people store roots or stems or processes or constraints or whatever, but the central fact remains unchanged: the basic pairing of signal to meaning in any language is completely idiosyncratic and memorized. To allay all doubts, though, let's look at claims in the literature that roots and morphemes do not exist.

Anderson (1992) proposes a completely processual morphology that abjures the use of morphemes altogether. Roots are thrown out with the morphemic bathwater: 'a theory that questions the status of morphemes in linguistic analysis has little room for such a construct' (1992, 4). But the evidence that any morphology must be processual is scant: the clearest case is subtractive morphology of the type found in Tohono O'odham and even that can be reanalyzed non-processually (Benua 1995, Golston 1996b, Fitzgerald & Fountain 1995). And even Anderson does not propose that *roots* are processes, so he too has fairly traditional morphemes at the base of his grammar.

Other approaches which downplay the usefulness of the root or which reject the notion altogether include the work of Matthews (1974, 1993), Becker (1990), Lieber (1992), Raffelsiefen (1995) and Neef (1996). Bat-El argues 'that there is no consonantal root in the morphology' of Hebrew either (1994, 594). But none of these authors would deny that there is an arbitrary pairing of signal and meaning that must be memorized in learning a language or that there's a difference between affixes and the things they attach to.

Aronoff (1976, 1992, 1994) has proposed that morphology is stem-based, rather than morpheme-based; but his stems are essentially of the type in (1). For Aronoff roots are somewhat derivative, but nonetheless real: 'A root is what is left when all morphological structure has been wrung out of a form' (1992, 15). Unanalyzable stems (roots) are fairly easy to come by in a language like German.

A number of researchers have proposed that morphemes are really *constraints* that regulate the sound-meaning relation. This is the claim of Declarative Phonology (Bird 1995, Coleman 1991, Scobbie 1991), for instance, which treats *all* of grammar as constraints. Within OT it has been proposed that all morphemes be treated as constraints as well (Hammond 1995; Russell 1995). And this is also the central claim of *Wortdesign* (Neef 1996), but only for affixes, not for roots. None of these proposals circumvents the need for morphemes, only the formal means by which they are encoded.

One often reads in the standard OT literature that there are no constraints on underlying representations (which is true), but this has nothing to do with the existence of morphemes. A language learner must posit some underlying form for *cat* in OT as in any other theory of grammar. The standard assumption in OT is that morphemes are encoded as strings of segments, as brief perusal of any OT article will verify. Golston (1996b) has proposed that morphemes aren't encoded as strings of segments (standard OT) or as constraints (Declarative Phonology, etc.) but as constraint violations. This doesn't do away with morphemes either, it just recasts how they are encoded in a grammar.

We can take the synactic road as well and deny any independent existence to morphology. But even if morphology is just word-syntax (Selkirk 1982, Höhle 1982, diSciullo & Williams 1987, Lieber 1992), we still have morphemes. More than ever, in fact, because now all terminal nodes in a syntactic tree are morphemes.

So let's not worry about claims that language makes no use of morphemes: the concept exists in any explicit theory of grammar and the details of analysis need not blind us to the necessity of the concept. This is particularly clear for roots. Roots are called unanalyzable stems in stem-based morphology and we'll be happy to call them that too if it helps avoid confusion. They are called morpheme constraints in Declarative Phonology and some corners of OT, and we are comfortable with that terminology as well. Likewise with treating roots as sets of constraint violations, as proposed in Direct OT or *even* as strings of segments, the path well-trodden from Saussure to Smolensky.

We want to look at what generalizations arise when we look at the type of thing in (1) in German and we want to see if these generalizations can be related to what we know about phonology generally. The question we want to look at here is not 'What is a root?,' but 'What properties do roots have?' Using a large computer corpus of roots collected independently by others, we'll propose that roots in a given language tend to have a basic shape and that such a shape is inherently related to markedness universals in phonology. The claim will be that less marked shapes have greater numbers of roots than more marked shapes. For English this would mean that there are less roots like *strict* than like *tea*.

2. A data base for German roots

Any large scale study of roots faces the problem of finding an adequate empirical domain over which all generalisations and claims are supposed to range. We see two problems here. The first is the decision of what to treat as native or borrowed. Native speakers usually don't know the difference, so it's not clear that the distinction is psychologically real or in any way important for grammar. Given that the number of foreign words can be huge and that words are usually assimilated to the host language, the issue is not trivial. There is good evidence that the stock of morphemes used in a language is not uniform. Itô & Mester (1995), e.g., have convincingly argued that the over-all set of Japanese morphemes consists of various subsets, normally called strata, which differ on a number phonological properties: a set of native morphemes (*Yamato*), a set of Sino-Japanese roots, a set of foreign roots, and a set of mimetic morphemes.

Second, the division into morphemes (affixes and roots) can itself be difficult. Following Aronoff's dictum that roots are what remains when affixes are stripped off can still leave a large number of dubious cases. To take up just one case from German, many nouns display a final schwa, as in [augə] 'eye', [naːzə] 'nose', [vaŋə] 'cheek'. The data base of frequent nouns by Ruoff (1981) yields 1484 such nouns, 19% of the total of 7863. For various reasons, it is unclear whether this final schwa should be regarded as a suffix or not. If so, it's not clear what the suffix marks. If not, it could be seen either as part of the root or as epenthetic material, not belonging to any morpheme. Deciding one way or the other makes a big difference for many of the statistics below: removing schwa from the root morpheme removes one syllable and makes the preceding consonant root-final. Wiese (1986, 1996a) and others have argued that schwa is generally a vowel of epenthesis and not an underlying segment and we will incorporate this into our analysis of German roots. The database introduced below does not treat final schwa as part of the root; it contains roots such as /aug/, /naːz/ and /vaŋ/.

We think it best that the selection of roots and their analysis be made independently of each other. Fortunately, there already exists for German a good deal of work in this area. Our starting-point is a computer-readable data base of *Kernmorpheme* (core morphemes) of German, compiled by Ortmann (1993), who based it on several extant root dictionaries.[2] The dictionaries used are of various sorts and span 80 years, from about 1890 to 1970. Together the resulting dictionary of Ortmann's can be regarded as a fairly complete set of German roots.

The work by Augst (1975) constitutes Ortmann's major source. This large-scale study is a three-volume dictionary of German morphemes; it is strictly synchronic, structuralist in orientation, and based mainly on two recent dictionaries of Standard German, Wahrig (1968) and Duden (1968). Augst notes variants of morphemes, related morphemes and borrowings, and marks them as such. In general, decisions made in Augst (1975) for morphemic segmentation were taken over by Ortmann. However, a fairly large number of simplex morphemes postulated by Augst were segmented further by Ortmann (1993: XLI). Among these are many words with final schwa, with a prefix *be-* or *ge-* (/bəgɪn/ 'beginning', /gəʃtalt/ 'gestalt'), and with various other types of formal changes. While these decisions could be regarded as problematic, we did not reverse them.

In setting up the data base, Ortmann made a number of further non-trivial decisions on the organization of his material. The four decisions which are quantitatively and qualitatively most important are the following:

1. A large number (several thousand) of words classified as foreign in Augst (1975) are simply omitted (*Abbé* 'abbot', *Schafott* 'scaffold', *Tohuwabohu* 'id.'…).
2. Words from a second list are singled out and listed in an appendix of Ortmann's data base (*Abenteuer* 'adventure', *add-* 'add', *Akt* 'act', *Scharlach* 'scarlet fever'). This list comprises 792 items. We interpret these as nativized loans in contrast to the foreign words mentioned in 1.
3. Of words such as *Gemälde* 'painting', which are listed in August under *mal* as being related to the latter, Ortmann isolates a 'core' *mäld*, and includes these in his list. Also, the endings *-e*

[2] Wolf Dieter Ortmann, Munich, has kindly made available his computer-readable database of the German „Kernmorpheme". On the basis of this database, Carmen Scherer, Marburg, has provided an Excel datasheet of the complete database and helped with the present calculations. At a later stage, we plan to make the database available through the internet.

and *-er* are truncated by Ortmann in a number of cases (*Schaffner* 'conductor' in Augst, *Schaffn* in Ortmann).
4. For more than 600 morphemes which are marked as capable of being umlauted in Augst, Ortmann generates a separate umlauted entry. That is, for *Buch* 'book', the data base contains both /bu:x/ and umlauted /by:ç/ as separate entries.

The last three decisions, the exclusion of assimilated loan words, the inclusion of roots such as *mäld* or *schaffn*, and the treatment of umlauted variants as separate roots, we regard as problematic. First, putative *mäld* appears in *Gemälde* only, *schaffn* is obligatorily followed by *-er* in the word *Schaffner*, and has no other status in the language. For umlauted forms, the inclusion of two forms leads to a duplication of forms which are otherwise completely identical. Second, Wiese (1996b) has argued that umlaut is a (lexical) phonological rule. In consequence, we have retraced the final three steps in the data base, but have remained faithful to Ortmann's first step. In this case, there was no choice here, anyway. Augst's several thousand *Fremdwörter* (3771, according to Ortmann 1993, p. XX) could have been added by hand only. The resulting list in our data base comprises 6512 entries.[3] Not a single entry has been altered, added or removed, even if there seemed to be good reasons for doing so, except for the correction of a few obvious spelling errors. The list of roots is thus (almost) completely based on the independent proposals made by Ortmann and by the dictionaries his work is based upon. Entries exist in orthographic and surface-phonemic form, as part of a (rather large) standard data base, at present an Excel sheet.

We take this data base to be the best current list of German roots. The process of arriving at this guess involves three steps. First, dictionary makers made their own decisions of classification and segmentation, which may or may not have been independent of each other.[4] Second, Ortmann (1993) made a number of carefully documented decisions in setting up his data base, as described above. Finally, we have changed it slightly as detailed above. Through this three-way filtering process, we arrive at a reasonable set of German roots. To see the amount of filtering which has occurred here, consider the fact that a good-sized German dictionary has 120,000 *words* (Duden 1989), while our list yields a mere 6512 *roots*. Etymologically, this set largely derives from the native Germanic vocabulary; else, an item is nativized to such an extent that it can be included in the same class of items. The class we have extensionally defined by our procedure may be compared to the *Yamato* items discussed by Itô & Mester (1995).

[3] A number of forms (about 600) have multiple entries in Ortmann's morpheme list; this is because they had been assigned to several „word families“ in Liebich's (1899) dictionary. As these families are etymological in nature, and as the shapes of the morphemes with multiple entries are always identical, we have removed this multiplication. Otherwise, the number of entries would be 7,134.

[4] It is of course well-known that new dictionaries are normally compiled on the basis of existing dictionaries.

3. (Direct) Optimality Theory

Both traditional accounts and those informed by recent phonological theory tend to describe roots as strings of segments. We will depart from this view in a way to be described now. We start with the central claims of Optimality Theory (Prince & Smolensky 1993). While we do not want to provide here an introduction to Optimality Theory (OT), we summarize its major claims in (2) and provide an example for its application immediately below. For published introductions to OT, we refer the reader to McCarthy & Prince 1993b, and to three edited volumes, Beckman et al 1996, Archangeli & Langendoen 1997 and Roca 1997.

(2) *Principles of Optimality Theory:*

- Constraints express universal grammatical preferences.
- Constraints evaluate surface forms as to the fulfillment of these constraints.
- Grammars differ in constraint ranking alone, not in the constraints they contain.

Grammaticality of a linguistic form is defined as best-fulfillment of a given constraint hierarchy: constraints and constraint violations are the stuff of which grammar is built. No rules, principles or parameters play any role in OT.

Here, we illustrate the concept by means of an example from German. Like all languages, German prefers syllables that begin with a consonant (CV) to those that begin with a vowel (V). Also like other languages, German avoids epenthesis whenever possible. Following work in OT, let us call the constraints responsible ONSET and FILL-C:

(3) ONSET 'Syllables begin with consonants.'

(4) FILL-C 'No epenthesis of consonants.'

The two constraints conflict with a vowel-initial word like ɑχt 'eight'. If we respect ONSET, [ʔɑχt], we violate FILL-C (by inserting the glottal stop) but if we respect FILL-C we violate ONSET: [ɑχt]. The issue is resolved by weighting or 'ranking' ONSET above FILL-C:

(5) ONSET >> FILL-C

The correct output will be generated in an OT grammar by penalizing [ɑχt] more than [ʔɑχt], as follows:

(6) *Glottal stop epenthesis*

ɑχt	ONSET	FILL-C
☞ ʔɑχt		*
ɑχt	*!	

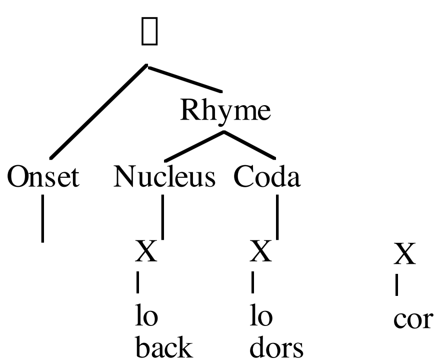
The first candidate isn't perfect but it's better-formed than the second. Both candidates violate a constraint but the violation of ONSET is fatal (!) to the second because it is worse than the violation of FILL-C incurred by the first.

It turns out that the constraints used in standard OT are so varied and so many that violations of them can be used to specify morphemes: linguistic forms can be characterized as specific sets of constraint violations. This makes standard representations redundant since constraint violations can be used both to evaluate forms and to represent them. This is the central claim of Direct OT:

(7) *Direct OT* (Golston 1996b)
Morphemes are represented by the constraints they violate.

This equivalence between standard representations and sets of constraint violations is illustrated for the German word [ɑχt] '*8*' below. Consider a standard autosegmental representation of the surface form of [ɑχt]:

(8) *Autosegmental representation of* ɑχt:



There are a bunch of things that are marked in (8) including the empty onset, the coda [χ] and the extrasyllabic stop [t] (Wiese 1996a: § 3.2.3). In addition, the tongue must be lower and further back in the mouth for the vowel than it is for quiet breathing and the dorsum must be low in the mouth to contact the uvula. The representation in (8) will be evaluated as violating ONSET, NOCODA, NODORSAL, etc. in the course of the grammar. If we freeze the grammar at this point and note just how marked our form is we will have something like (9) which says what is bad about [aχt] purely in terms of markedness. These distinctive violations or desiderata constitute a grammatical address for the meaning *8*:

(9) *Desiderata for [aχt] '8'*

ONSET	NO CODA	NO DORSAL	NO CONT	NO STOP	NO CORONAL	NO LO
8	8	8	8	8	8	8

The idea is that the listener will select the meaning *8* just in case the utterance she hears has no onset, does have a coda, has a dorsal articulation, etc. Note that (9) says nothing about linear order (the ranking of the constraints is irrelevant): the fact that the consonants all come after the vowel has to be inferred from the violation of ONSET and NOCODA.

Empirical evidence for encoding morphemes in terms of constraint violations (9) instead of autosegmental representations (8) comes from various types of morphology generally regarded as problematic for sign-based models. Such types of morphology include truncation, zero-affixation and reduplication. These receive a natural interpretation in Direct OT. Let's look at two examples to see how this is done. Tohono O'odham has a well-studied alternation involving perfective verb stems in which the perfective is marked by deletion of the stem-final syllable (Zepeda 1983).

(10)	Stem	Perfective		
	cikapana	cikapa	'worked'	(surface [cik^{h}p^h])
	bisiceka	bisice	'sneezed'	(surface [bisc^h])

Subtractive morphology of this sort can be characterized as distinctive violation of the faithfulness constraint PARSESYLLABLE, part of a family of faithfulness constraints in standard OT (Prince & Smolensky 1993; PARSE is recast as MAX in McCarthy & Prince 1995). The desideratum for the Tohono O'odham perfective (pf) is then simply:

(11) *Tohono O'odham perfective*

PARSE SYLL
pf

The hearer accesses the meaning PERFECTIVE just in case she *cannot* locate one syllable of the verbal stem in the utterance. The fact that it is the *final* syllable of the word that is underparsed is attributed to a constraint that aligns distinctive features with the beginning of a word: ALIGN-L (*f*, WD): underparsing any non-final syllable would alter the destinctive (mis-)alignment of features.

Alignment constraints (McCarthy & Prince 1993b) are distinctively violable as well. Recall that German has roots that surface with a final schwa when they stand alone but without schwa when suffixed or compounded: [liːbə] ‘love’, but [liːbloːs] ‘loveless’ or [liːbçən] ‘darling’; [augə] ‘eye’, but [augapfl̩] ‘eye-ball’; [naːzə] ‘nose’ but [naːshorn] ‘rhinoceros’, and so on. The surface peculiarity of the unaffixed forms is that the root is misaligned from the end of the prosodic word by what appears to be an epenthetic vowel [ə]: roots like [liːbə] and [naːzə] *never* occur at the end of the word in German while roots like [siːb] ‘sieve’ and [gʀaːs] ‘grass’ do. Assuming a constraint ALIGN-R (ROOT, WD), we may specify that roots of the [liːbə] and [naːzə] type violate it distinctively: they never surface word-finally. In the event that they are suffixed or compounded, the constraint is satisfied trivially (the root is not word final in [liːbloːs] or [naːshorn]). In the event that they are not suffixed, an epenthetic schwa is inserted to meet the desideratum. In this way we can capture the insight that the vowel is epenthetic but lexically idiosyncratic (Wiese 1986, 1996a), an otherwise difficult thing to do:

(12) *Desideratum for final schwa in [haːzə] ‘bunny’:*

ALIGN-R (RT, WD)
Ÿ

The hearer accesses the meaning BUNNY just in case (inter alia) she locates a misaligned root.[5]

Floating features that mark specific morphological categories are handled analogously. Ladefoged & Maddieson (1996, 69) report on a causative morpheme in Burmese whose sole phonetic reflex is aspiration of word-initial consonants:

[5] Golston & Wiese (1996) argue that the German plural suffix -ə is epenthetic as well, the result of an alignment constraint affecting inflected forms.

(13) *Burmese causative*

state		*causative*	
páuʔ	‘be pierced’	p^{h}áuʔ	‘pierce’
céʔ	‘be cooked’	c^{h}éʔ	‘cook’
nôu	‘be awake’	n̥ôu	‘waken’
láʔ	‘be bare’ l̥áʔ	‘uncover’	

The causative morpheme violates the constraint against a spread glottis, a marked state for the vocal tract. Direct OT represents the morpheme as follows:

(14) *Desideratum for Burmese causative*

NOSPREAD GLOTTIS
`caus`

Here the hearer accesses the meaning CAUSATIVE if she notices a spread glottis configuration. The location of the spread glottis feature is independently given by alignment constraints.

In general, DOT represents every morpheme in terms of a non-null set of constraint violations. This opens a new view on what properties must be categorically true for some item: it will be the expected situation that a particular linguistic item violates a constraint holding for many other items of the same type. For example, if syllables are expected to be open in some language, a closed syllable will simply be represented as one violating the constraint NOCODA.

4. Root markedness

In this section we’ll try and substantiate one of the predictions of Direct OT, viz. that less marked roots are more common within a given language than more marked shapes. We’ll cover both constraints that are never violated and those that are rarely violated, basing our frequency on the corpus described above.

4.1 Minimality

Like the lexical roots of Sanskrit (Steriade 1988), Greek, Latin, English (Golston 1991) and Proto-Indo-European (Golston 1996a), lexical roots in German are minimally bimoraic, CVC or CVː. Examples include the following:

(15)	...VC		...VV		*...V
	ʀɛt	'save'	ʀeː	'deer'	*ʀɛ
	kʊs	'kiss'	kuː	'cow'	*kʊ
	ʃtat	'city'	ʃtau	'traffic jam'	*ʃta

As has often been discussed in the literature on German phonology, stressed syllables cannot end in one of the short full vowels [ɪ, ʏ, ɛ, œ, ʊ, ɔ, a]. Given that lexical roots must be stressed, they cannot consist of **bɪ*, **bɛ*, etc or of **b*, **k*, etc. Assuming that the mora is the unit of which a light syllable contains one, and a heavy syllable two, the relevant constraint is one which requires every root to contain a bimoraic foot:

(16) CONTAIN (RT, FT) 'Every root contains a bimoraic foot.'

CONTAIN (RT, FT) is an absolute constraint in German desiderata (no subminimal roots exist in the lexicon) and in the grammar (roots never become subminimal as a result of deletion).

An absolute constraint is never violated in the grammar or by the desiderata for morphemes. The result is an exceptionless skewing away from subminimal roots:

(17) *No subminimal roots in German*

root-shape	#	%	ALIGN-L (RT, FOOT)
huː, etc.	6512	100	
hʊ	0	0	R

This constraint is not distinctively violated ('R') by any root in German.

4.2 Maximality

German roots tend to be short, usually maxing out at one vowel or two syllables. Vowel and syllable count diverge in German because of large numbers of what we might call syllable-and-a-half roots, roots with a full syllable followed by a schwa (as discussed above) or syllabic sonorant:

(18) *Short roots*

Monosyllable		*Schwa*		*Syllabic sonorants*	
huːt	‘hat’	liːb(ə)	‘love’	beːzn̩	‘bowl’
man	‘man’	viːz(ə)	‘meadow’	laɪtʀ̩	‘ladder’
zeː	‘sea’	bak(ə)	‘cheek’	eːzl̩	‘donkey’

The precise distribution and characterization of the relevant syllables and vowels has been a matter of some debate ever since the phonemic status of these vowels was questioned by Moulton (1947); for a recent line of discussion see Giegerich (1985), Wiese (1986, 1996a) and Neef (1996). All we are saying at this point is that syllables of this type are best regarded as syllables without vocalic features like [hi] [lo] and [rd].

We capture all of this with the constraints in (19) and (20):

(19) ALIGN-L (V_, RT) ‘Every vocalic feature is root-initial.’

(20) ALIGN-R (V_, RT) ‘Every vocalic feature is root-final.’

A vocalic feature in the second syllable of a root is one syllable peak away from the left edge of the root in violation of ALIGN-L (V_, RT); a vocalic feature in the first syllable of a two syllable root is one peak away from the right edge in violation of ALIGN-R (V_, RT). Roots that respect both alignment constraints above are monosyllabic: [ɡuːt] ‘good’, [maχ] ‘make’, [zeːʀ] ‘very’. Those that violate the second are syllable-and-a-half roots: [zeːɡl̩] ‘sail’, [ʔaːtm̩] ‘breath’, [ʔaːbn̩d] ‘evening’—most of these ends in orthographic-*el* or *-er* such as *Hagel* [haːɡl̩] ‘hail’ or *Ruder* [ʀuːdʀ̩] ‘oar’. Roots that violate the first are quite rare: [ʔaʀbaɪt] ‘work,’ [ʔalaɪn] ‘alone.’ Figures for the frequency of each type are given below, alongside a tableau that indicates the markedness of each type with respect to the alignment constraints discussed above:

(21) *Short roots in German*

root-shape	#	%	ALIGN-L (V_, RT)	ALIGN-R (V_, RT)
huːt	5149	79		
zeːgl̩	1227	19		R
ʔaʀbaɪt	131	2	R	R
ʔaleːgʀo	5	0	RR	RR
kalifɔʀnjən	0	0	RRR	

The five roots in our corpus that violate the alignment constraints twice are all loans and felt to be such by most speakers: [ʔaleːgʀo] 'allegro,' [baldʀiaːn] (name), [ʔɛnziaːn] (name), [feːbʀuaʀ] 'February,' and [januaʀ] 'January.' Longer roots that do not occur in the data base are obvious loans and are interpreted as such [kalifɔʀnjən], [ʔalabaːma], etc. These are possible German words (they can be pronounced) but their shapes are so marked in the language that they only occur as loans.

Similar restrictions on maximal length are found in many language families. Here we mention but two. Many Mon Khmer languages also have both syllable and syllable-and-a-half roots, but with exactly the opposite alignment effects of German: syllable-and-a-half roots begin with the half-syllable and end with the full one. Consider root shapes in Sre (Manley 1972) and German:

(22) *Root shapes in Sre and German*

	Sre	German
CVX	√	√
Cə	*	*
CəCVX	√	*
CVCə	*	√

Both languages have mostly bimoraic monosyllabic roots and neither has any monomoraic roots. Sre allows half-syllables initially but not word-finally; German allows them finally but not initially. The general picture in (21) is very clear: monosyllabic forms account for a large majority of roots; monovocalic polysyllabic forms are in a large minority, whereas the number of polyvocalic forms is very small. A similar case is found in the Mayan language Mam (England 1983, 93):

> Most Mam roots are of the shape CVC, and in fact only noun, adjective and particle roots can have more than one vowel, and that rarely. Verbs, positionals, and affect roots are quite restricted as to shape.

So the type of maximal size restriction we find in German is not uncommon in the languages of the world. This adds a measure of support to our analysis, which represents German roots by distinctive violation of universal constraints.

Alignment constraints like these restrict the size of the lexicon considerably. Given *n* possible syllables, there are *n* monosyllabic potential roots, but *n* to the power of *m* possible roots containing *m* vowels. Assuming 2000 syllables, there are 8 billion trisyllabic (and smaller) roots!

4.3 Root alignment

A different type of alignment constraint makes roots begin crisply with a syllable boundary:

(23) ALIGN-L (RT, □) ‘Every root begins with a syllable.’

This seems trivial, but it is necessary for ruling out roots that begin with an unsyllabifiable consonant, like *mhuː or *tkuː:[6]

(24) *No misaligned roots in German*

root-shape	#	%	ALIGN-L (RT, FOOT)
kuː, etc.	6512	100	
tkuː	0	0	R

The constraint is also necessary for correctly deriving surface forms of morphologically complex words. Consider a vowel-initial root like [aʀbaɪt] ‘work’ which takes a prefix [g] and a suffix [t] in the preterite. We might expect the verb to surface as *[gaʀbaɪtət] but it does not. Rather, the root stays aligned with a syllable boundary and the prefix is forced into a degenerate foot:

[6] We assume here that initial [ʃ] or [s] in roots such as *Stadt* ‘town’, *Spiel* ‘play’, *Skelett* ‘skeleton’ are not extrasyllabic, but are part of a complex segment (Wiese 1991).

(25) *[gəaʀbaɪtət] 'worked'*[7]

	ALIGN-L (RT, □)	FILL-V	ONSET
g[aʀ.baɪ.t]ət	*!	*	
☞ gə.[aʀ.baɪ.t]ət		**	*

The winning candidate is prosodically awful, with two epenthetic vowels and an unfilled onset; but the only serious competitor misaligns the root with the edge of the first syllable and so forfeits the competition.

A similar type of constraint explains an otherwise puzzling fact about German: despite the presence of many polymorphemic words like [gəaʀbaɪtət] 'worked' and [bəʃʀɪtn̩] 'stepped on', there are no monomorphemic words (roots) that begin with a syllable headed by schwa or a syllabic consonant.

(26) *Roots begin with a full syllable*

*[ʔəlaɪn]
*l̩zeːg
*m̩taː
*bn̩daː

We may attribute the lack of such roots to a constraint which prohibits roots from beginning with degenerate syllables:

(27) ALIGN-L (RT, V*f*) 'Every root begins with a vowel feature.'

Violations of this constraint are assessed in terms of syllable peaks that intervene between the edge of the root and a given vowel feature. This makes [ʔɑχt] '8' and [mɑχt] 'power' equally well aligned in terms of vocalic features.

There is a sizable class of *stems* that begin with orthographic *ge-* (gənau 'precise', gəzund 'healthy', gəʃtalt 'gestalt') or *be-* (bəʀaɪχ 'scope', bəʀaɪt 'ready'); all of these have schwa in the first syllable but none of them turned up in our database. This is because *ge-* and *be-* are prefixes in German; the forms that contain them are thus not roots, properly speaking, but stems. The *ge-* and *be-* prefixes no longer contribute to the meaning of the stem, but native speakers recognize them as prefixes.

[7] **Note that the brackets used for candidates in (25) denote root boundaries.**

4.4 Initial consonants

One of the clearest generalizations about prosodic structure is that syllables tend to begin with consonants. The constraint has been introduced above and needs little discussion. What is interesting to note is just how many German roots respect it. Fully 94% of our corpus consists of consonant-initial roots.

(28) *Vowel-initial roots in German*

root-shape	#	%	ONSET
mɑχ	6147	94	
ɑχt	365	6	R

The marked status of vowel-initial roots couldn't be more clear. Direct OT marks this directly in lexical representation as violation of Onset. Consider the desiderata for '8' again, repeated below:

(29) *Desiderata for /aχt/ 'eight'*

ONSET	NO CODA	NO UVULAR	NO CONT	NO STOP	NO CORONAL	NO LO
8	8	8	8	8	8	8

(The violation of NOCODA is for the uvular fricative, which surfaces in all forms of the root, even when suffixed, as in [ʔaχ.tsiç] 'eighty'.)

4.5 Final consonants

At the other end of the root we find the same story repeated: fully 96% of German roots are consonant-final. This cannot be attributed to syllable structure, of course, since syllables prefer to be open cross-linguistically. McCarthy (1995) **(Should this be McCarthy 1993?)** has proposed a different type of constraint, however, according to which prosodic words want to end in a consonant in certain dialects of English. Golston (1996a) extends this to roots for Indo-European and we adopt his constraint here:

(30) FINAL-C 'Every root ends with a consonant.'

FINAL-C accounts for the skewing shown in the following table, where significantly more roots respect the constraint than violate it:

(31) *Vowel-final roots in German*

root-shape	#	%	FINAL-C
maχ	6268	96	
drɑɪ	244	4	R

Roots that violate this constraint bear a desideratum to that effect, as the following example makes clear:

(32) Desiderata for drɑi '3'

NOCOMPLEX ONSET	NOCOMPLEX NUCLEUS	FINAL C	NO COR	NO STOP	NO CONT	NO UVU	NO LO	NO HI
3	3	3	3	3	3	3	3	3

(We exclude from consideration vowel-final roots that end in schwa, like [naːzə] 'nose'. As discussed above, we analyze the vowel here as epenthetic, since it fails to appear inside suffixed or compounded forms.)

4.6 Schwa syllables

Let's look a bit more closely at root-final schwa syllables. As we have just seen, the most common of these by far are those whose syllable peak is occupied by a schwa rather than a syllabic sonorant.[8] Next come those whose peak is a syllabic liquid, followed by a syllabic nasal and so on, as shown in the chart below.

[8] **The frequency of such roots cannot be determined from our root data base because the schwa is always stripped off. But we may take the figure 1484 (Ruoff 1981) as a reasonable estimate.**

(33) *Root-final schwa syllables in German*

peak	#	%	NOCODA	*P/NAS	*P/LIQ
ə	1484	53.0			
l̩	620	22.0			*
ʀ̩	593	21.0			*
n̩	50	2.0		*	
n̩d	10	.5	*	*	
n̩t	7	.5	*	*	
ət	5	.4	*		
ʀ̩t	4	.3	*		
əm	3	.3	*		

We can understand the drop in frequency in terms of three constraints: one that prohibits codas entirely, two that prohibit nasals and liquids from forming syllable peaks. The ranking of the last two is not open to cross-linguistic variation: liquids are more sonorous (and thus form better peaks) than nasals in all languages. All three constraints are familiar in OT (Prince & Smolensky 1993):

(34) NOCODA ‘Syllables end in a nucleus.’

(35) *P/NAS ‘Syllable peaks do not contain nasals.’

(36) *P/LIQ ‘Syllable peaks do not contain liquids.’

If we rank NOCODA above the constraints on what may occupy a peak, we predict the exact pattern found in German roots: the most common schwa syllables have (featureless) vowels, followed by syllabic liquids, syllabic nasals and finally syllabic sonorants with codas.

4.7 Stress

As a further case of an asymmetrical distribution of root-shapes, we look at stress. Obviously, this can only be done for polyvocalic ones, as the schwa syllables discussed above are always stressless. There are 136 polyvocalic roots in the corpus, 27 of which have a final schwa (foˈʀɛl(ə) ‘trout’, ʀoˈziːn(ə) ‘raisin’) which we do not treat as part of the root here.[9] This leaves us with 109 polyvocalic roots, out of which more than 80% are stressed non-finally:

[9] **The reason is that there is a strong stress principle in German which says that the vowel preceding a schwa must be stressed. That is, ˈʀoːzin(ə) is not a possible stress pattern for *Rosine* ‘raisin’.**

(37) *Stress-final roots in German*

root-shape	#	%	NON-FINAL
aʀbaɪt	89	82	
aˈpʀɪl	20	18	R

We assume that a constraint directly disallowing final stress (see Prince & Smolensky 1993: 40) is in charge here:[10]

(38) NONFINALITY 'Final syllables are unstressed.'

To state the result more dramatically, we could say that not more than .3 % of the roots in the data-base clearly bears final stress, presupposing that monosyllabic roots (the vast majority) are not distinctively marked for stress. It is also the case that almost all of the 20 stress-final roots are represented in only one of the dictionaries. (Some may even be in the list by mistake, like *Benzol* and *Burgund*.)

We note, however, that a number of common words are not found in our data base: balˈkɔŋ 'balcony', byˈʀoː 'office', idˈjoːt 'idiot', muˈziːk 'music', paˈkeːt 'package', naˈtuːʀ 'nature', kaˈpʊt 'broken' etc. These are not items of Germanic origin, to be sure, but we would have included them in the database simply because of their commonness. It may well be that these words were not included in the database because of their peculiar final stress, but this is something we can only speculate on. We have chosen not to tamper with the corpus for the present.

4.8 Other root features

In principle, Direct OT predicts that every active constraint in a language should skew the morphemes in a lexicon in their favor: for any constraint C there should be more roots that respect C than violate it. We are not in a position to show this yet, though we hope to have shown that it is true for the constraints we consider above.

Let us briefly spell out, however, what we would expect to find. First, we've seen that vocalic features align leftwards in the root in German and we expect consonantal features to do the same. Given a configuration C_1VC_2, we expect a feature like NASAL or LABIAL will show up more often on C_1 than on C_2. Second, given constraints like NONASAL and NOLABIAL, we expect more roots to have oral stops than nasals and more roots to have plain stops than aspirates. Finally, given constraints against complex onsets, nuclei and codas, we expect more roots

[10] Given that roots are predominantly monosyllabic or at least monovocalic, this system will become more apparent on the levels of stems and words, rather than roots; see, e.g., Wiese (1996b; ch. 8), Alber (1998).

like [fɪʃ] ‘fish’ than like [fʀɪʃ] ‘fresh’, [kiːʃ] ‘quiche’ or [zɪçt] ‘view’. Future research will have to determine whether our predictions in these areas are supported or not.

One other comment is in order: some constraints will appear to be violated quite freely, as a result of constraint ranking. Consider the case of NOCODA. We know from cross-linguistic study that syllables prefer to be open (respecting NOCODA), but German *roots* prefer to be closed (respecting FINAL-C) and they prefer to right-align with a prosodic wordIn German the roots win out over the preferred syllable structure as the figures in (31) demonstrate. The grammar does of course provide various means—in particular the addition of final schwa in inflected forms and in uninflected words like naːzə ’nose’—which conspire to give words a shape that is more in accordance with syllabic well-formedness.

5. Conclusions

The full set of constraints we’ve looked at is given below:

(39)	ONSET	‘Syllables begin with consonants.’	(3)
	FILL-C	‘No epenthesis of consonants.’	(4)
	FILL-V	‘No epenthesis of vowels.’	(12)
	CONTAIN (RT, FT)	‘Every root contains a bimoraic foot.’	(16)
	ALIGN-L (V_, RT)	‘Every vocalic feature is root-initial.’	(19)
	ALIGN-R (V_, RT)	‘Every vocalic feature is root-final.’	(20)
	ALIGN-L (RT, □)	‘Every root begins with a syllable.’	(23)
	ALIGN-L (RT, V*f*)	‘Every root begins with a vowel feature.’	(27)
	FINAL-C	‘Every root ends with a consonant.’	(30)
	NOCODA	‘Syllables end in a nucleus.’	(35)
	*P/NAS	‘Syllable peaks do not contain nasals.’	(35)
	*P/LIQ	‘Syllable peaks do not contain liquids.’	(36)
	NONFINALITY	‘Final syllables are unstressed.’	(38)

For all of these constraints we have seen that the number of roots that obeys it is much greater than the number of roots that violates it.

A few of the constraints are exceptionless in our database: ALIGN-L (RT, □), CONTAIN (RT, FOOT) and ALIGN-L (RT, V*f*). These are naturally accounted for in any grammar that countenances morpheme structure constraints. The rest of the constraints admit of exceptions. These are most naturally accounted for with violable constraints, specifically OT. Relating the frequency of shapes to the constraints that they violate is most naturally accounted for in Direct OT, where representation is done purely in terms of markedness. And so we take our results as supporting constraint-based grammars generally, OT more specifically and Direct OT in particular.

Morpheme structure constraints have been problematic throughout the history of generative phonology, because many such constraints are similar or even identical to phonological rules, where the latter seemed to be needed independently (Kiparsky 1982) and because well-formedness constraints are somewhat alien in a rule-based grammar. But as the appeal of rule-based grammars has given way to more surface-oriented models the idea of morpheme structure constraints has become more tractable. This is true despite the fact that OT at least is incapable of putting constraints on underlying forms. Constraints in OT regulate only surface forms, but this is perfectly compatible with the regulation of the forms of roots on the surface as we have seen. Morpheme structure constraints are part and parcel of the grammatical machinery regulating phonology and morphology. The fact that of all possible morpheme structure constraints we are readily identifying those which hold for roots is probably not accidental either: among others, McCarthy & Prince (1995, 364) observe that faithfulness to roots is ranked higher than faithfulness to affixes.

We feel that our results argue for the reality of the root as a linguistic category. Emphasizing that at least some of the properties are not those which typically hold for the syllable, stem or word, we feel justified in concluding that our study provides evidence for the reality of the root in the phonology and morphology of German. The root is not something abstract; it is the core of the signal-to-meaning relation and has a number of phonological properties, some exceptionless, some merely strong tendencies.

Finally, we note that our study of roots in German adds substance to an old hypothesis of the Jakobsonian school of linguistics: markedness can be related formally to frequency. This is to be expected if representation is carried out purely in terms of markedness (Direct OT) but is somewhat surprising if representation and markedness are formally unrelated.

6. References

Alber 1998. ?? In: Wolfgang Kehrein & Richard Wiese (eds.) *Phonology and Morphology of the Germanic languages*. Tübingen, Niemeyer: ??

Anderson, Stephen R. (1992) *A-morphous morphology*. Cambridge: Cambridge University Press.

Archangeli, Diana & D. Terence Langendoen (1997) *Optimality Theory: an overview*. Oxford: Blackwell.

Aronoff, Mark (1976) *Morphology in generative Grammar*. Cambridge Mass.: MIT Press.

Aronoff, Mark (1992) Stem and Root. In: William Bright (ed.): *International Encyclopedia of Linguistics, vol 4*. New York, Oxford: Oxford University Press, 76-77.

Aronoff, Mark (1994) *Morphology by Itself. Stems and Inflectional Classes*. Cambridge, Mass.: MIT Press.

Augst, Gerhard (1975) *Lexikon zur Wortbildung. Band 1 - 3*. Tübingen: Narr Verlag.

Bat-El, Outi (1994) Stem modification and cluster transfer in Modern Hebrew. *NLLT* 12.4, 571-598.

Becker, Thomas (1990) *Analogie und morphologische Theorie*. München: Fink.

Beckman, Jill N., Laura Walsh Dickey and Suzanne Urbanczyk, eds. (1996) *Papers in Optimality Theory. University of Massachusetts Occasional Working Papers 18*. UMass, Amherst: GLSA.

Benua, Laura (1995) Identity effects in morphological truncation. In J. Beckman, S. Urbanczyck and L. Walsh (eds.), University of Massachusetts Occasional Papers in Linguistics 18: Papers in Optimality Theory. Graduate Linguistics Student Association, Amherst, Massachusetts.

Bird, Steven (1995) *Computational phonology: a constraint-based approach*. Cambridge University Press.

Bopp, Franz (1833) *Vergleichende Grammatik des Sanskrit, Send, Armenischen*. Vol. 1, p. III.

Coleman, John (1991) *Phonological representations*. Doctoral dissertation, University of York.

Di Sciullo, A.M. and E. Williams (1987) *On the Definition of Word*. Linguistic Inquiry Monograph, no. 14. Cambridge: MIT Press.

Duden (1968) Duden. Rechtschreibung der deutschen Sprache und der Fremdwörter. Mannheim: ??.

Duden (1989) *Duden Deutsches Universalwörterbuch*. Mannheim, Wien, Zürich: Bibliographisches Institut.

England, Nora C. (1983) *A grammar of Mam, a Mayan language*. Austin: University of Texas Press.

Fitzgerald, Colleen and Amy Fountain (1995) The optimal account of Tohono O'odham truncation. Talk presented at 1995 LSA.

Giegerich, Heinz (1985) *Metrical Phonology and Phonological Structure. German and English*. Cambridge: Cambridge University Press.

Golston, Chris (1991) *Both Lexicons*. UCLA doctoral dissertation.

Golston, Chris (1996a) Prosodic constraints on roots and words. In Ursula Kleinhenz (ed.), *Interfaces in Phonology*. Berlin: Akademie Verlag, 172-193.

Golston, Chris (1996b) Representation as pure markedness. *Language* 72, 713-748.

Golston, Chris and Richard Wiese (1996) Zero morphology and constraint interaction: subtraction and epenthesis in German dialects. *Yearbook of Morphology 1995*, 143-159.

Hammond, Michael (1995) There is no lexicon! Ms. University of Arizona.

Höhle, Tilman (1982) Über Komposition und Derivation: zur Konstituentenstruktur von Wortbildungsprodukten im Deutschen. Zeitschrift für Sprachwissenschaft 1, 76-112.

Itô, Junko & R. Armin Mester (1995) In: Japanese Phonology. In: John Goldsmith (ed.) *Handbook of phonological* theory, 817-838.

Kiparsky, Paul (1982) Lexical Morphology and Phonology. In: The Linguistic Society of Korea (ed.): *Linguistics in the Morning Calm: Selected Papers from SICOL-1981*. Seoul: Hanshin, 3-91.

Ladefoged, Peter & Ian Maddieson (1996) *The sounds of the world's languages*. Oxford: Blackwell.

Lieber, Rochelle (1992) *Deconstructing morphology. Word formation in syntactic theory*. Chicago: University of Chicago Press.

Liebich, B. (1899) *Die Wortfamilien der lebenden hochdeutschen Sprache als Grundlage für ein System der Bedeutungslehre*. Breslau.

Manley, Timothy M. (1972) *Outlihne of Sre structure*. Oceanic Linguistics Special Publication No. 12. University of Hawaii Press.

Matthews, Peter H. (1974/1993)*Morphology: an introduction to the theory of word-structure*. Cambridge: Cambridge University Press.

McCarthy, John J. (1993) A case of surface constraint violation. *Canadian Journal of Lingusitics* 38, 169-95.

McCarthy, John J. and Alan S. Prince (1995) Faithfulness and reduplicative identity. In: Jill Beckman et al. (eds.): *University of Massachusetts Occasional Papers in Linguistics 18*. Amherst, Mass.: GLSA,249-384.

McCarthy, John J. and Alan S. Prince. (1993b) Generalized alignment. *Yearbook of Morphology* 1993, 79-153.

Mester, Ralf A. (1994) The quantitative trochee in Latin. *Natural Language and Linguistic Theory* 12, 1-61.

Moulton, William G. (1947) Juncture in Modern Standard German. *Language* 23, 212-26.

Neef, Martin (1996) *Wortdesign. Eine deklarative Analyse der deutschen Verbflexion*. Tübingen: Stauffenburg (Studien zur deutschen Grammatik 52).

Ortmann, Wolf Dieter (1993). *Kernmorpheme im Deutschen. 4561 allgemeinsprachliche deutsche Morphemgestalten im Vergleich zwischen zehn Quellen.* München: Goethe-Institut.

Prince, Alan and Paul Smolensky (1993) *Optimality theory: Constraint interaction in generative grammar*. Ms, Rutgers University and University of Colorado, Boulder.

Raffelsiefen, Renate (1995) *Conditions for stability: the case of schwa in German*. Düsseldorf, Theorie des Lexikons. Arbeiten des Sonderforschungsbereichs 282, Nr. 69.

Roca, Iggy, ed. (1997) *Derivations and constraints in phonology*. Oxford: Clarendon Press.

Ruoff, Arno (1981) *Häufigkeitswörterbuch gesprochener Sprache: gesondert nach Wortarten alphabetisch, rückläufig-alphabetisch und nach Häufigkeit geordnet*. Tübingen: Niemeyer.

Russell, Kevin (1995) Morphemes and candidates in Optimality Theory. Ms. University of Manitoba.

Scobbie, James. (1991) *Attributive value phonology*. Doctoral dissertation, University of Edinburgh.

Selkirk, Elizabeth O. (1982) The syllable. In Harry van der Hulst and Norval Smith (eds.), The Structure of Phonological Representations, Part 2. Dordrecht: Foris. 337-383.

Steiner

,

Richard (1992) Ancient Hebrew. In: William Bright (ed.): *International Encyclopedia of Linguistics, vol 2*. New York, Oxford: Oxford University Press, 110-118.
Steriade, Donca (1988) Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5, 73-155.
Wahrig, Gerhard (1968) *Deutsches Wörterbuch.* Gütersloh.
Wiese, Richard (1986) Schwa and the structure of words in German. *Linguistics* 24, 697-724.
Wiese, Richard (1991) Was ist extrasilbisch im Deutschen und warum? *Zeitschrift für Sprachwissenschaft* 10, 112-133.
Wiese, Richard (1996a) *The Phonology of German*. Oxford: Clarendon Press.
Wiese, Richard (1996b) Morphological vs. phonological rules: on German umlaut and ablaut. *Journal of Linguistics* 32, 113-135.
Zepeda, Ofelia. 1983. *A Papago Grammar*. Tucson: University of Arizona Press.