Abstractness in phonology: the case of virtual geminates

Philippe Ségéral and Tobias Scheer

1. Introduction

Paul Kiparsky's (1968) question *how abstract is phonology?* initiated a discussion about the dissimilarity that phonological theory should allow for between phonetic reality and phonological representations. One aspect of the debate concerned abstract segments. These are objects which occur in the underlying structure of a morpheme, but appear in none of its phonetic manifestations, or, worse, are even absent from the segmental inventory of the language.[1]

In Generative Phonology, the discussion on abstractness led to the view that abstract structures should be dispensed with in the presence of solutions that are bound to the surface in a more direct fashion. This position is expressed in subsequent work by Kiparsky (Kiparsky 1982 contains a number of relevant papers) and, among others, by Tranel (1981). It was the cornerstone of Natural Generative Phonology (Hooper 1976, Vennemann 1974a,b) and Natural Phonology (e.g. Stampe 1973; Dressler 1974). However, the debate never came to a conclusive end. In particular, no satisfactory definition of what is natural, simple, elegant, phonetically plausible, psychologically real or typologically invariant could be achieved (Hellberg 1977; Koutsoudas 1980; Dinnsen 1980). Neither could an evaluation metric be derived that would be able to measure different degrees of abstractness (Kiparsky 1974; Campbell 1981; Goyvaerts 1981). Proponents of abstract analyses with no other limitation than the learnability of rules (c.f. Skousen 1981) are, among others, Schane (1968,1974), Dell (1973), Selkirk and Vergnaud (1973), Dinnsen (1980) and, taking abstractness to an extreme point, Lightner (1981). Kenstowicz and Kisseberth (1977:1-62)

provide an extensive discussion of the topic. They argue in Kenstowicz and Kisseberth (1979:204ff.) that in some cases, there is no alternative to abstract analyses that are motivated by compelling internal evidence.

Abstractness as a topic of phonological debate has played little or no role in the 80's. Focus was switched on the internal structure of segments, as well as on issues regarding autosegmental representations. In the 90's, theories that are concerned with constraints are involved in neither of the former issues; abstractness, the internal structure of segments and phonological representations play little or no role in the discussion.

However, we believe that the questions related to abstractness, though raised in the 70's, are still relevant for phonological theory. As a fact in point, another typical topic of the rule-ordering period has come back into the spotlight recently: Optimality Theory is seeking to cope with opacity.
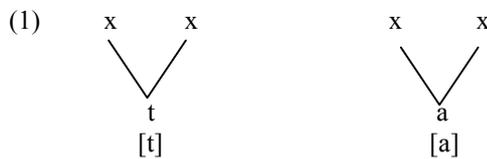
In this paper, we discuss a type of abstract segment that seems to be recurrently required for the purpose of various analyses in genetically unrelated languages. Attention is drawn on underlying geminates that are always realised as simple consonants. Such an object, which we call a *virtual geminate*, never betrays its geminate identity by a phonetic clue related to length, but by other properties that can be read off the phonetic environment. A situation of this kind has been identified for example in Danish, cf. Larsen (1994), Lowenstamm (1996).

Virtual geminates are not exactly abstract segments in the sense of the above discussion. Assuming autosegmental representations, their melodic structure is identical to the one of the corresponding simplex consonant, which is part of the phonemic inventory of the language in all cases we are aware of. Their virtual character is due to the double association they enjoy in regard of the skeleton, and which does not lead to a long pronunciation.

The very same situation is actually encountered for vowels, too. In certain Afro-Asiatic languages, peripheral vowels behave as long, even though they are phonetically as short as central vowels, which behave like true simplex segments. Lowenstamm (1991) has

formalised this observation, and a number of generalisations in various languages crucially rely on this view (Ségéral 1995, 1996; Bendjaballah 1999).

If there are reasons to believe that vowels may enjoy a double association and yet be phonetically simplex, not only is it possible, but expected to also find virtual geminates. For there is no difference in representing virtual long vowels and virtual geminates in autosegmental structures. (1) below illustrates this situation for a virtual /tt/ and a virtual /aa/.

(1)

```
     x     x              x     x
      \   /                \   /
       \ /                  \ /
        t                    a
       [t]                  [a]
```

Hence, autosegmental representations, that were not available (or not well enough established) when the debate on abstractness took place, invite to differentiate two kinds of abstract segments. Phonemes whose underlying identity is never visible on the surface (or not even present in the inventory of the language) on the one hand, and an underlying relationship between two or more autosegmental tiers that has not the expected surface-effect on the other. Although it is not advisable to make a formal prediction on the grounds of the empirical evidence of one article, it may turn out that "abstract segments" in the pre-autosegmental sense are always (or mostly) of the kind that we are dealing with in the next few sections, that is as under (1).

In any event, we submit that the two case studies which we present conform to the minimal requirement imposed on abstract underlying forms: they must be retrievable by children during the process of acquisition. Objects that do not occur on the surface may be incorporated into underlying representations as long as their identity can be recovered on the grounds of phonetic information provided by the context.

We adduce evidence for the existence of virtual geminates in Cologne German (section 2) and Somali (section 3), arguing that a proper analysis of these languages cannot be carried out unless the existence of virtual geminates is recognised.

2. Cologne German

Cologne German (henceforth CG) is peculiar in that it features a velarisation of dental consonants. The reverse movement illustrated by palatalisations is extremely common, and usually triggered by an adjacent palatal segment. The few velarisations on record are either of dissimilatory nature or obey syllabic configurations, cf. English "l-darkening" in Codas. The Cologne German velarisation is worthwhile because it does not instantiate a dissimilation, nor is it conditioned by any syllabic pattern. Compare the Cologne forms below with their Standard New High German (henceforth NHG) cognates.

(2)	correspon-dence	NHG	Cologne German	spelling	gloss
a.	n - ŋ	bʁawn	bʁuŋ	braun	brown
		t͡sawn	t͡suŋ	Zaun	fence
		ʁajn	ʁiŋ	Rhein	Rhine
		lajnə	liŋ	Leine	rope
		nɔjn	nyŋ	neun	nine
b.	t – k	bʁawt	bʁuk	Braut	bride
		ʁajtən	ʁigə	reiten	ride
		hɔjtə	hyk	heute	today
	d – g	ʃawderliç	ʃugiʃ	schauderlich	frightening
		ʃnajdən	ʃnigə	schneiden	cut
		lajdən	ligə	leiden	suffer
c.	nt - ŋk	bunt	buŋk	bunt	colorful
		pintə	piŋk	Pinte	pint
	nd - ŋg[2]	ʁundən	ʁyŋə	runden	round
		bindən	biŋə	binden	tie
		byndəl	pyŋəl	Bündel	bundle

The correspondence shown is regular, and is present in the entire Cologne lexicon. A complete description is given in Scheer (forthc.); the phenomenon is also discussed in the traditional comparative and dialectological literature, e.g. Hönig (1877), Münch (1904), Lessiak (1933), Frings and Schmitt (1942), Bertram (1935), Müller (1942).

Synchronically, there is no way to predict the context in which this velarisation occurs. The only generalisation that may be drawn is of comparative nature: New High German simple dentals as under (2a,b) occur as velars in Cologne German words when preceded by a diphthong [aj,aw,ɔj] in New High German, and only in this environment.

However, there are numerous Cologne German words that do not possess velars where New High German shows diphthongs, e.g. both NHG and CG [ʃtajn, bajn, bʁajt, lawt, fχɔjd(ə)] *Stein, Bein, breit, laut, Freude* "stone, leg, large, noisy, joy", CG *[ʃtiŋ, biŋ, bʁik, luk, fχyk].

Diachronic evidence clears up the situation: the common ancestor of both New High German and Cologne German is Middle-High German (henceforth MHG, about 1050-1350 A.D.). As shown under (3), NHG diphthongs have two different MHG sources (e.g., Paul et al. 1989:68ff., 105ff.).

(3)
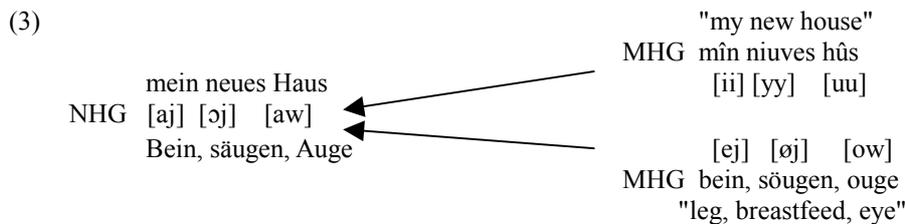
Cologne German contains 97 velar roots that have a New High German dental cognate, 42 of which are simplex consonants as under (2a,b)[3]. All of those 42 words show a root-vowel that goes back to a MHG long high vowel, none originates in a MHG diphthong.

We may thus identify Cologne German velarisation as the following diachronic process: MHG simplex dentals become Cologne German velars when preceded by MHG long high vowels.

(4) MHG [n,t,d] > CG [ŋ,k,g] / MHG {[ii,uu,yy]}__

The velarisation discussed is to be regarded as a synchronically inactive process that has occurred in the evolution from MHG to Cologne German. It cannot be part of the synchronic phonology of Cologne German since the triggering context "a preceding long high vowel" has been altered since MHG.

Another piece of evidence concerns the triggering context: the vowel preceding velarised consonants has encountered an abnormal evolution. In non-velarised words, MHG long high vowels come out unaltered in CG, e.g., MHG *hûs, wîp, tiuvel* [h**uu**s, v**ii**p, t**yy**vəl] > CG [h**uu**s, v**ii**f, d**yy**vəl] "house, woman, devil".[4] Contrary to this regular evolution, all velar Cologne German forms that possess a New High German dental cognate come along with a short vowel, e.g., MHD *brûn, hiute, snîden* [bʁ**uu**n, h**yy**tə, sn**ii**dən] > CG [bʁ**u**ŋ, h**y**k, ʃ**ni**gə] *braun, heute, schneiden* etc., cf. (2). Vowel-shortening in velarised words is thus to be interpreted as a consequence of velarisation.

When looking at the part of the corpus that involves dental clusters becoming velar in Cologne German (2c), it turns out that the generalisation made under (4) also holds true as far as vowel-quality is concerned. MHG [nt,nd] become CG [ŋk,ŋg] if preceded by a high vowel. The dental cluster of all items on record in the 97-root list mentioned earlier is preceded by a high vowel. Only is this vowel short, and not long as before.

We are thus facing a general process whereby velarisation is triggered for all candidates by preceding high vowels. This condition is sufficient in case of dental clusters. It is only necessary for the velarisation of simplex dental consonants, however. The sufficient condition here is the length of the preceding vowel. Hence, the quality of the triggering vowel is uniform for all kinds of velarised items. Its quantity depends on the quantity of the target: if the candidate is short, the trigger must be long. If it is long, the trigger is short.

For the sake of clarity, all generalisations that have been made so far are summed up below.

(5) generalisations
  a. only high vowels that precede dentals trigger velarisation
  b. simplex dentals velarise only if preceded by a long vowel
  c. when velarisation of simplex dentals takes place, the preceding long vowel shortens
  d. dental clusters [nt,nd] velarise after short vowels[5]

The length of the trigger and the target is in complementary distribution. This is a strong indication to the effect that the overall volume involved in the process is constant. Namely, it must be of three timing units: in all cases, trigger and target occupy three skeletal slots. This is illustrated in (6) below (T=dentals, K=velars).

(6) VNT > VŋK   MHG *bunt* [bunt] > CG [buŋk]
     VVT > VKK   MHG *hiute* [hyytə] > CG [hyk]

The crucial feature of (6) is the gemination of simplex dentals when velarisation occurs: VVT > VKK. Indeed, this is the only way of accounting for vowel shortening (5c) while maintaining the overall constant volume of three skeletal slots.

Indeed, if simplex dentals are not only affected by velarisation, but also by gemination, the overall picture makes sense. Only under this proviso can the Cologne German velarisation be described as a uniform and unique process. The intricate alternations concerning the quantity of the segments involved may then be explained as under (7).

(7) a. velarisation of simplex consonants takes place only after long vowels because only long vowels may provide the additional skeletal slot which is required for the gemination of the velar.
     b. velarisation of dental clusters [nt,nd] is triggered by short vowels because clusters do not need to acquire an additional skeletal position in order to be geminate: they are already double.

c. long vowels preceding velarised consonants undergo shortening because their second skeletal position is occupied by the velar.

If velarised simplex dentals are to be regarded as geminates at the end of the diachronic process at stake, are they still underlyingly geminate in present-day Cologne German? On the surface, they never appear as such, as was stated before.

The following diachronic analysis could be argued for: MHG simplex dentals were velarised and geminated in the mentioned context. The stage of the language where this result occurred in is unrecorded. In further development, an independent process of degemination took place. These three steps appear in (8) below.

(8) a. MHG   [VVT]
     b. intermediate stage:
        after velarisation and gemination   *[VKK]
     c. CG: after degemination   [VK]

There is, however, evidence to the effect that phonetically simplex velars which come from dentals are still underlyingly geminate in actual Cologne German. Consider the distribution of Cologne German underlying /g/, which enjoys six different surface manifestations (so-called *spirantische Kontaktregel*, see Frings 1955 on this issue). MHG [g] appears in Cologne German as shown under (9).

(9)

		CG	spelling	gloss
[j] /	#__	jynstiʃ, jʁoos	günstig, groß	favourable, big
	V__V	ɛjaal	egal	doesn't matter
	VV$_{front}$__V	nææjəl	Nägel	nails
	C__	balje	balgen	scuffle

[ɣ] / VV$_{non\ front}$__V   zaaɣə, naaɣəl   sagen, Nagel   say inf, nail

ø / __C   zææt, ʁææn   er sagt, regnen   he says, rain

[ʃ] / V$_{front}$__#   œvʁiʃ, fuftsiʃ   übrig, fünfzig   be left, fifty

[χ] / V$_{non\ front}$__#   zaaχ !   sage !   say ! 2sg

[k] /	__#C	zaaks	sag es	say it! 2sg
	__-(C)	zaak	sagt !	say! 2pl

/g/

This alternation is synchronically active, as witnessed by the fact that the entire Cologne German lexicon is affected, including contemporary loans[6]. For instance, words such as *Gage* ("fee", from French)*, Galerie* ("galery", from French)*, Gig* ("concert of a band", from English) and *Glacéhandschuhe* ("transparent gloves", from French) are pronounced [jaaʃ, jalᵃʁii, jik, jlasantᵃ]. If a Cologne German speaker is asked to repeat a nonsense word such as *Guntil*, the result is [juntil].

As may be seen under (9), none of the allophones is identical to the underlying identity of the phoneme. Hence, if the allophony is part of the synchronic grammar of Cologne German, one should expect that there is no [g] at all in this language.

This indeed is a correct prediction – except for voiced velar stops that are the result of velarisation: [g]s that are phonetically present in Cologne German have come into being through velarisation. Lexical items containing [g] are given under (10) below.[7]

(10)

NHG	CG	NHG	CG	NHG	CG
krauten	kʁugə	reiten	ʁigə	streiten	ʃtχigə
lauten	lugə	schneiden	ʃʁigə	Beutel	bygəl
schauderlich	ʃugiʃ	weiten	vigə	zeitig	t͡sigiʃ
leiden	ligə	schreiten	ʃʁigə	deuten	dygə
Trude	dʁygəlʃə	läuten	lygə	häuten	hygə

If the synchronic status of velar voiced stops which come from dentals were /g/ as suggested in (8), they would of course be expected to behave in the same way as regular /g/. Since velars that have a dental source are not affected by the allophony described under (9), their underlying identity must be different from /g/. The only solution that is consistent with both the velarisation and the allophony grants a *synchronic* geminate identity /gg/ to velars that originate in MHG dentals.

These virtual velar geminates may be easily identified through the process of acquisition. Assuming that learners have successfully incorporated into their grammar the process shown under (9), they know that any time they come across a [g], its underlying representation must be different from /g/. If one does not wish to introduce an otherwise non-existing phoneme into underlying representations, and if absolute neutralisations are disqualified (cf. Kiparsky 1968), then the geminate identity is the only straightforward way for the learner to implement [g] into his/ her grammar. In fact, learnability may be used as an argument to prefer non-phonemic abstract segments in the sense of (1) over abstract phonemes: virtual length is the only way to achieve a picture where the structure arrived at in the process of acquisition is consistent with diachronic and synchronic analysis.

In sum, we are thus left with an underlying geminate that does never appear on the surface as such. Only its degeminated version [g] is observable. Note that there are no geminate consonants on the surface in Cologne German at all. It could thus be argued that degemination is a feature that is enforced upon /gg/ by the general structure of the language. In the next section, however, we discuss a parallel case of virtual geminates in Somali, a language where phonetic geminate consonants do occur.

3. Somali

In Somali (Cushitic), intervocalic [t] such as in [daqnataa] "you/ she SG feel(s) pain" must be analysed as a phonological geminate. That is, the melodic identity of "t" is associated to two consonantal positions, while being invariably simplex on the surface. This interpretation is enforced by the interplay of three facts, that we present hereafter.

3.1. Morphological analysis

In order to identify the morphological structure of the quoted [daqnataa], let us consider the conjugational paradigm of an unmarked verb such as keen "bring", which represents the class of verbs that is traditionally termed "weak".[8]

(11)

	root	SUBJ. AGR.	Tense marker		surface
1 SG	keen	ø	aa		keenaa
2 SG	keen	t	aa		keentaa
3 SG MASC	keen	ø	aa		keenaa
3 SG FEM	keen	t	aa		keentaa
1 PL	keen	n	aa		keennaa
2 PL	keen	t	aa	n	keentaan
3 PL	keen	ø	aa	n	keenaan

Table (11) shows that the subject agreement marker [-t-] characterises 2SG, 3SG FEM and 2PL forms, while [-n-] indicates 1PL. All other persons enjoy a zero marker. Subject agreement is located between the root and the tense marker. The latter is [-aa] in the above examples illustrating the "Habitual Present" paradigm, as opposed to "Past" forms that are achieved by [-ay]: [keen-ay] "I/ he/ they have brought", [keen-t-ay] "you SG, she has brought". In addition, a final [-n] marks 2PL and 3PL. The set and the distribution of Subject Agreement markers shown is constant throughout all weak verbal paradigms in Somali.[9]

Somali weak verbs fall into different classes, two of which bear on the present demonstration. The above *keen* illustrates the unmarked verb class (so-called first conjugation). The verb [daqnataa] quoted earlier, on the other hand, belongs to the third conjugation, which contains autobenefactive and reflexive items. The characteristic marker of verbs belonging to this class is the suffix /-(a)t-/ that is directly attached to the stem, and whose vowel alternates with zero. For instance, the first person singular of the first conjugation verb *fir* "comb TRANSITIVE" is [fir-aa] "I comb TRANSITIVE", with the Subject Agreement marker [-ø-] evidenced under (11). The autobenefactive version of the same verbs is [fir-t-aa] "I comb my hair", where the class-marker /-(a)t-/ occurs in zero grade.

Hence, the morphological composition of the autobenefactive form [daqnataa] "you SG/ she feel(s) pain" must be as under (12a). (12b) shows the related non-autobenefactive counterpart [daqantaa] "you SG/ she hurt(s)".[10]

(12)	stem	AUTOBEN	2SG	HAB. PRES.		
a.	daqn	-at	-t	-aa	[daqnataa]	"you SG/ she feel(s) pain"
b.	daqan		-t	-aa	[daqantaa]	"you SG/ she hurt(s)"

In other words, the simplex [-t-] in [daqnataa] is the phonetic reflex of two underlying /t/s that belong to two different morphemes.

3.2. Resistance to voicing

The simplex [t] in [daqnataa] is also unexpected because dental stops voice and spirantise in intervocalic position in Somali. This process may be observed for the feminine definite article /-ta/, which surfaces as [-ða] if preceded by a vowel-final stem. Compare /laan-ta/ = [laanta] "the branch", /deg-ta/ = [deg-ta] "the peak" vs. /mindi-ta/ = [mindiða] "the knife", /hoojo-ta/ = [hoojaða] "the mother".[11]

Illustration of the same process is also given by the autobenefactive /-(a)t-/ at stake in [daqnataa]. Its /-t-/ appears in voiced and spirantised form on the surface when it comes to stand in intervocalic

position. For instance, the first person singular of the autobenefactive verb *kabbo* "sip" is [kabbaðaa] from /kabb-at-aa/, just as the first person signular of *daqno* "feel pain" [daqnaðaa] from /daqn-at-aa/.[12] In autobenefactive verbs, 1SG that bears a zero Subject Agreement marker and 2SG where agreement is achieved through /-t-/ are thus opposed by the sole contrast [-ð-] vs. [-t-].

(13)

daqno "feel pain"	kabbo "sip"	t͡ʃoogso "stop o.s."	
[daqnaðaa]	[kabbaðaa]	[t͡ʃoogsaðaa]	1SG = STEM + (a)t + ø + aa
[daqnataa]	[kabbataa]	[t͡ʃoogsataa]	2SG = STEM + (a)t + t + aa

It is a striking fact that intervocalic [t] resists voicing precisely in the cases where the morphological analysis reveals the presence of two dentals. On the other hand, whenever morphology provides one single /-t-/, voicing applies and a [-ð-] is observed on the surface.

3.3. Vowel – zero alternations

The attentive reader has noticed that vowel-zero alternations are frequent in the examples mentioned. They occur in two different sites. In a CV_1CV_2C-stem, V_2 may alternate with zero (*daqn- / daqan-* "feel pain", *ɖerg- / ɖereg-* "feel full", *kidf- / kidif-* "shop", *boqr- / boqor-* "make king", *gudb- / gudub-* "cross"), in which case its vocalic manifestation is always a copy of V_1. The other instance concerns the vowel of the autobenefactive suffix, which appears as either [-øt-] (/fir-(a)t-aa/ = [fir-t-aa]) or [-að-] (/daqn-(a)t-aa/ = [daqn-að-aa]). In this case, the alternating vowel is invariably [a].

These alternations are controlled by the phonotactics of the consonantal environment: iff the alternation-site is preceded or followed (or both preceded and followed) by more than one consonant, the vowel appears.[13] In any other configuration, no vowel may be observed. Somali thereby instantiates the cross-linguistically stable pattern of vowel-zero alternations found in many other languages

such as Slavic, French, Moroccan Arabic or Hungarian (Scheer 1997,1999).

Another way of describing the situation is to say that three consonants in a row (or CC#) are prohibited: $*C_1C_2C_3$. If the alternation-site is stem-internal, a copy of the first stem-vowel breaks up the cluster between C_1 and C_2 like in first conjugation verbs /daqn-t 2SG -aa/ = [daq**a**ntaa] "you SG hurt". In case the alternation-site is located in the autobenefactive suffix, potential $*C_1C_2C_3$ clusters are vocalised between C_2 and C_3, e.g. /daqn-(a)t-ø 1SG -aa/ = [daqn**a**ðaa] "I feel pain".

Accordingly, no vowel surfaces in first person singular forms of first conjugation verbs. In /gudb-ø-aa/ = [gudbaa] "I cross", /kidf-ø-aa/ = [kidfaa] "I shop", no *CCC cluster arises since the 1SG marker is zero. No *CCC cluster is encountered in 1SG forms of autobenefactive verbs either despite the presence of the suffixal *-t-* if the stem of these verbs ends in one single consonant: /fir-(a)t-ø-aa/ = [firtaa] "I comb my hair", /tuur-(a)t-ø-aa/ = [tuurtaa] "I pile up".

Contrastingly in second person forms of the first conjugation where the Subject Agreement marker is *-t-*, a three-membered cluster arises, which is broken up by a copy of the first stem-vowel, cf. /gudb-t-aa/ = [gud**u**btaa] "you SG cross". In the same way, first person singular forms of autobenefactive verbs whose stems end in two consonants are vocalised since the autobenefactive /-(a)t-/ adds a third consonant: /kabb-(a)t-ø-aa/ = [kabb**a**ðaa], /t͡ʃoogs-(a)t-ø-aa/ = [t͡ʃoogs**a**ðaa], /daqn-(a)t-ø-aa/ = [daqn**a**ðaa].

In second person forms of autobenefactive verbs, it may come as a surprise that the suffix does vocalise in [fir**a**taa, tuur**a**taa, qab**a**taa] even though no *CCC seems to occur. If the above analysis is correct, only two reasons may be invoked for this inordinate behaviour: either the alternating vowel is preceded, or it is followed by a consonant cluster. In other words, either the stem-final [-r,-b] in the above examples are to be viewed as clusters, or the following [-t-] is a geminate. As a point in fact, Somali has overt voiced geminates, among which [rr] and [bb], while voiceless consonants never occur phonetically geminated. Hence, there is no way to analyse the stem-final consonants in the examples mentioned as underlying geminates,

since they would appear as such on the surface. The [-t-] following the alternating vowel, however, would not betray its identity if it were an underlying geminate.

It is hard to believe in coincidence when considering that the [-t-] which triggers the abnormal presence of the alternating vowel is precisely the same [-t-] that irregularly resists voicing in intervocalic position. Its status as a phonological geminate would explain both issues. Moreover, its geminate identity is consistent with the morphological structure of the forms at stake: [fir**a**taa] is /fir-(a)t-t-aa/.

In sum, we face phonetically simplex [t]s in intervocalic position that 1) resist voicing and 2) trigger the vocalisation of a preceding alternation-site in the same way as CC clusters do. In other words, these [t]s behave like geminates. Furthermore, morphological analysis shows that they do indeed represent two distinct /t/s.

We believe that this set of facts cannot be accounted for unless it is admitted that intervocalic [t]s in Somali are phonological geminates /tt/. In absence of this analysis, the autobenefactive paradigm remains mysteriously irregular, no rule or intervocalic voicing may be formulated, nor do vowel-zero alternations obey any recoverable pattern.

3.4. Benefits

The assumption of virtual geminates sheds also light on other intriguing situations elsewhere in the grammar.

In some first conjugation verbs such as *matag* "vomit", *hitiq* "walk slowly" or *mutuħ* "speak without reserve", a medial intervocalic voiceless [t] occurs. Curiously, these verbs also disregard the regular vowel-zero alternations that are normally found in CV$_i$CV$_i$C-verbs (Orwin 1995:75; Saeed 1998:23). Compare the 1SG form [gudbaa] from /gudb/ with 1SG forms [mat**a**ɣaa, hit**i**qaa, mut**u**ħaa], *[matgaa, hitqaa, mutħaa], from *matag, hitiq, mutuħ* respectively. If [t] in these verbs is a geminate, both its resistance to voicing and its triggering quality of the stem-vocalisation are regular. These verbs do not need to be regarded as "exceptions" any longer.

In Somali compounds, the initial consonant of the second member is regularly geminated if the first member ends in a vowel. Compare for instance *ɖeɣo* "ears" attached to *beel* "lack of", yielding [ɖeɣabbeel] "deafness". However, when the same word *ɖeɣo* "ears" precedes *tir* "cancel", the result is [ɖeɣatir] "ignore, refuse to listen", bearing a phonetically simplex [t]. This [t] obviously instantiates a geminate.

Another piece of evidence comes from the free variation between geminates and homorganic nasal-stop clusters that is found in Somali (Keenadiid 1976:XXII). Consider for example the doublets *ʕaɖɖo / ʕanɖo* "scabies", *middi / mindi* "knife", *higgo / hiŋgo* "hiccough". Other cases of this free variation with homorganic nasal-stop clusters, however, involve simplex [t]s: *mitid / mintid* "persevere", *ʕutub / ʕuntub* "small group", *sitaaʕso / sintaaʕso* "put ornamental trappings on a horse", *matag / mantag* "vomit". Again here, the phonetically simplex [t] clearly has the phonological value of a geminate.

Finally, the autobenefactive paradigm of verbs whose stem ends in *–ad* adduces striking evidence for the geminated character of intervocalic [t]. The second person singular forms of *gaɖo* "buy", *qoɖo* "dig" and *quuɖo* "eat, feed on" are [gaɖataa], [qoɖataa] and [quuɖataa], respectively. The presence of the boldfaced [-a-] does not come as a surprise anymore since the [t] on its righthand side is an underyling geminate (/STEM + (a)t (autob.)+ t (2SG) + aa/). When looking at the corresponding first person singular forms, a geminate [dd] could be expected. Indeed, the stem-final /-d/ together with the autobenefactive suffix /-(a)t-/ (which occurs in zero grade because the personal marker is zero) form an underlying /-d-øt-/ cluster, which could assimilate to /-dd-/ = [-dd-]. The actual result, however, is a simplex [-t-], as shown under (14).

(14)

	stem	AUTOBEN.	PERS.	HAB. PRES.	
1SG	/…Vt	-(a)t	-ø	-aa/	[gataa, qotaa, quutaa]
2SG	/…Vt	-(a)t	-t	-aa/	[gaɖataa, qoɖataa, quuɖataa]

This realisation could be the result of a regressive assimilation /dt/ —> /tt/ = [t]. Or rather, the stems could be underlyingly voiceless and

undergo intervocalic voicing in quotation forms /gat-o/ = [gaðo]. In this case, the underlying cluster in first person singular forms is /tt/ = [t] without assimilation. Whatever the correct analysis, under (14), in yet another morphological environment, we encounter one more case of simplex [t]s that necessarily enjoy a geminate phonological identity.

In sum, a very easy pattern, that learners may discover without any pain, holds true for all instances of dental stops in Somali: any intervocalic [t] is an underlying geminate /tt/, while all voiced interdental fricatives [ð] represent an underlying simplex /t/ or /d/.

3.5. *Velar virtual geminates*

The demonstration so far concerned only dentals. However, other segments behave alike in Somali. We shall examine the case of the velar stop hereafter in some more detail, place restrictions do not allow for the discussion of other segments.[14]

Underlyingly simplex /k/ is voiced and spirantised in intervocalic position just as /t/. This is evidenced by the masculine definite article –*ka*: /dab-ka/ = [dabka] "the fire", /beer-ka/ = "the liver" vs. /guri-ka/ = [guriɣa] "the house", /qoraa-ka/ = [qoraaɣa] "the writer". In the same way, clusters of two preverbal prepositions (/u+ka/ = [uɣa] "to+from") as well as sequences of object pronouns and prepositions (/i+ku/ = [iɣu] "me+in") yield the same result. There are, however, intervocalic [k]s in Somali. Their behaviour is strictly parallel to the one that has been reviewed for [t], and a geminate underlying identity /kk/ must be concluded on.

Just as [t], [k] is in free variation with homorganic nasal-stop clusters, e.g. *sokor / soŋkor* "sugar", *maakaal / maaŋkaal* "residue left after the pressing of sesame".

The geminated character of intervocalic [k] is also witnessed by the same apparently deviant behaviour with respect to vowel-zero alternations. For instance, first conjugation CV_iCV_iC verbs with medial [k] behave like those with medial [t], i.e. their second vowel never alternates with zero. Examples are 1SG [fekeraa], *[fekraa]

from *feker* "think", 1SG [mak**a**laa], *[maklaa] from *makal* "put in danger", 1SG [ħuk**u**maa], *[ħukmaa] from *ħukun* "govern, rule". In these forms, the simplex [k] has the same effect as a consonant cluster in that it prohibits absence of the second stem vowel (again, if the stem-final consonant were a geminate, it would surface as such).

Finally, stems that end in [-Vk] should never tolerate the autobenefactive suffix /-(a)t-/ to occur in zero-grade if [k] is an underlying geminate. In first person singular forms, the morphological structure is /…Vkk-(a)t-ø-aa/. The corresponding verbs confirm this prediction: *buko* "become sick", *ɗako* "hide oneself" and *tuko* "pray" seem to display the same structure as *firo* "comb (autoben)", but in fact pattern with *kabbo* "sip", cf. 1SG [tuk**a**ðaa] from *tuko* as [kabb**a**ðaa] from *kabbo*, against [firtaa] from *firto*.

The existence of a voicing process in intervocalic position in Somali, as well as the perfect regularity of vowel-zero alternations allow phonological geminates /tt, kk/ to enjoy a simplex phonetic interpretation.[15] The recoverability of the geminate status of intervocalic [t,k] is guaranteed by these external clues. As before in Cologne German, learners may unambiguously reconstruct the geminate whose identity is not betrayed by its length, but by the "irregular" behaviour of its environment.

A parallel situation obtains for voiced stops. Underlying voiced geminates /bb,dd,gg/ are usually realised as such [bb,dd,gg]. Optionally, however, a Somali speaker may pronounce simplex [b,d,g] only. This possibility is due to that fact that underlying simplex voiced stops, as was mentioned earlier, are spirantised in intervocalic position: /b,d,g/ surface as [β,ð,ɣ], respectively. The non-spirantisation of a simplex phonetic voiced stop in intervocalic position therefore is sufficient in order to unambiguously betray its geminate identity. The difference between geminate and non-geminate voiced stops is that the latter always spirantise, while the former never do.

4. Conclusion

In the preceding pages, we have argued that the phonology of two particular languages, Cologne German and Somali, may not be understood unless the existence of underlying geminates that have a simplex phonetic value is admitted. The geminate identity of these segments is contained in the acoustic signal. But instead of being encoded in the length or melody of the segment itself, it may be read off its environment.

As stated in the introduction, the need of assuming virtual geminates seems to be recurrent in various analyses of different languages. Moreover, the exact vocalic counterpart, virtual long vowels, have also been identified (Lowenstamm 1991).

The discussion on abstractness in the 70s, and namely the debate concerning abstract segments, relates in the following way to the issues discussed in this paper. According to current autosegmental representations, geminates and long vowels are melodically identical to their short counterparts. The only difference is the number of timing-units this melodic content is attached to.

It is interesting to note that virtual geminates and virtual long vowels are virtual precisely because of the relation that two autosegments (the segmental and the timing tier) contract, and because of nothing else. Abstract segments in the sense of the early generative debate thus fall into two different species if autosegmental representations are assumed: segments that are abstract because 1) their melodic identity or a part of it does never occur on the surface, or 2) their association to another autosegment has not the usual phonetic consequences.

By the time the discussion on abstractness took place, autosegmental representations were not part of phonological theory, or were just emerging. Hence, the formal distinction between different abstract segments could not be made, and indeed is absent from the literature. We would like to suggest that the majority, if not all cases where analyses recurring to abstract segments are compelling involve an association between two autosegments which is not phonetically executed.

In epistemological terms, this analysis is cumulative in the sense of Encrevé (1997) and Durand and Laks (1996). When new theories arise and science moves on, open questions that are related to and were debated in the old framework are often left behind. This would not be harmful if people expected the new theory to provide tools that allow to view the unresolved questions in a new light after a while. New theories are cumulative if they do not forget about old problems that still beg the question. They are non-cumulative if they ignore old insights and questions that have been raised prior to their birth and with no relation to them. Unfortunately, generative phonology in the past 30 years bears quite a bit of a non-cumulative character. Rule-based theories such as SPE were mainly concerned with procedural issues, i.e. rule-ordering. They had their typical set of data, which was different from the one that gave rise to autosegmental representations. These emerged through the analysis of tone and syllabic phenomena. In the 80s, mainstream phonology investigated the internal structure of segments, leaving behind the questions that were of interest before. The relevant data were e.g., vowel harmony and all kinds of adjacent and distant assimilation that could reveal natural classes of sounds. In a subsequent step, syntagmatic relations among segments such as Licensing were focused on, leading to more restrictive views of syllable structure. Relevant data involved for example lenition. And finally, the 90s are back to procedural issues and the general organisation of grammar (Optimality Theory, Declarative Phonology). It is true that part of the moves are of cumulative nature, cf. the recent debate on counter-bleeding within OT. Most frequently, however, new theories do not consider old questions and old data, cf. for instance the absence of issues related to phonological representations in OT.

Our contribution to the old question of abstractness is of cumulative nature. It takes advantage on a formalism that did not or poorly exist at the time abstractness was discussed (autosegmental representations) in order to propose a generalisation that may make the old question appear in a different light: abstract segments are abstract because of an association between two autosegments that is not phonetically executed.

Finally, the old problem of abstractness referring to the dissimilarity that grammar ought to allow between underlying and surface representations is mirrored in more recent discussion by the so-called family of faithfulness constraints that are designed to minimise the contrast between deep and surface structure. Violation of faithfulness is viewed as a consequence of the action of another set of constraints. In principle, the balance of both "forces" is not predictable, it may go way to one side, as well as far on the other. In this paper, we have intended to show how this problem may be handled in a principled way. Instead of allowing for a principle to be violated in undetermined and a priori unlimited fashion, we submit that the area in which it is contradicted is well defined and predictable. Circumscribing the violation in place, nature and function allows to capture the phenomenon by setting a parameter, rather than by dispensing with the principle, or allowing for its violation. Moreover, it adds a generalisation to the grammar that does not emerge when the violation of principles is thought of as the unmarked case.

Notes

[1] The discussion on abstractness in synchronic grammar has a parallel in diachronics. Reconstructed forms that indo-europeanists arrived at by the end of the last century as displayed in Brugmann and Delbrück (1886-1990) did not take into account the fact that indo-european was a real language pronounced by real speakers. Highly abstract constructions bearing impressive consonant clusters remained unquestioned. In later times, these were cast doubt on as for instance by Benveniste (1935:1f). More recently, typological considerations of living languages have substantially transformed the classic Brugmannian picture through Gamkrelidze and Ivanov's (1995) glottalic theory. However, Saussure's (1879) Laryngeal theory has become a consensual grounding of indo-european reconstruction. Laryngeals are abstract segments in the sense that their existence was inferred by Saussure on the basis of an empirical record they were absent from. In this view, the relation holding between the two levels of synchronic analysis (surface – underlying) has the same status as the one relating earlier and subsequent forms in diachronic analysis. In both cases, the existence of abstract segments may be evidenced by advances in

"technical" sophistication: Anatolian brought Laryngeals to light, and the reality of underlying abstract segments may be expected to appear when we will be able to identify neuronal structures.

[2] As in Standard German and a number of English varieties, Cologne /ŋg/ is realised [ŋ], explaining the absence of the alleged /g/ in CG [ʁyŋə, biŋə, pyŋəl]. The underlying identity of [ŋ] is established by the same arguments as for New High German and English such as, for instance, the absence of word-initial [ŋ]. In addition, it may be observed on the surface when devoiced by a following voiceless obstruent as e.g., in the second person singular: /biŋg-s/ = [biŋks].

[3] The exhaustive list is available in Scheer (forthc.).

[4] See Hönig (1877:15) for more illustration.

[5] The behaviour of dental clusters following long vowels may not be tested since configurations such as [VVNC] do not exist in MHG. The few apparent counter-examples that may be come by in New High German are the result of recent evolutions: *Mond* "moon" [moond] < *Monat* "month" [moonat], *fahnden* [faandən] is an intensive of *finden* "find" and was short until MHG: *vanden*.

[6] See Wrede (1958, 1:265) and Hönig (1877: 17ff.) on this issue.

[7] There are two other sources of [g]. Loans from Low Germanic such as NHG *Roggen* [ʁɔgən] "rye" that show a geminated -gg- in German spelling appear in Cologne German with a simplex [g], i.e. CG *Rögge* [ʁœgə]. Words that contain a -ck- in German spelling, which is the graphic indication of its former geminated character, may also be represented by [g] in Cologne German. In NHG *wackeln, einnicken, Rücken* [vakəln, ajnikən, ʁykən] = CG *waggele, nügge, Rögge* [vagələ, nygə, ʁœgə] "be shaky, nod off, back (part of the body)", an underlying /k/, or even /kk/, is voiced intervocalically. In any event, no CG [g] is a manifestation of /g/.

[8] As opposed to "strong" verbs whose personal markers are prefixed. The five verbs at stake are *yilil* "be in a place", *yimi(d)* "come", *yiqiin* "know", *yidhi* "say" and *yahay* "be".

[9] Except for imperatives and reduced forms.

[10] The stem-internal vowel-zero alternation that is observed (*daqn-* / *daqan-*) will be discussed in a subsequent section.

[11] In the latter example, an independent process controls the stem-final o-a alternation, cf. Saeed (1998: 26).

[12] Somali verbs are traditionally given in their quotation form, that is the imperative 2SG. For autobenefactive verbs, IMP 2SG is [stem-o]. Strangely enough, this is the only form in the entire paradigm that does not show the autobenefactive suffix.

[13] The right margin of the word behaves like a consonant: alternation-sites followed by a word-final consonant are vocalised, e.g., imperative 2SG /kidf/ = [kidif], *[kidf] "shop!".

[14] Parallel behaviour is observed for [t͡ʃ,ʃ,q,w,j] (there is no [p] in Somali). See Barillot and Ségéral (forthc. a,b) for a complete record of the facts.

[15] A similar case can be made for Modern Hebrew. Berman (1997: 315) points out that the otherwise regular syncope of non-high vowels in nouns and adjectives before stressed suffixes "is blocked by historical medial geminates". Compare for instance SG [matos] – PL [mtos-im] "air-plane" vs. SG [ganav] – PL [ganav-im] "thief". Clearly, the "historical" geminate in [ganav-im] < Tiberian Hebrew [ga**nn**a<u>b</u>iim] is still present in the underlying form of modern Hebrew. It is not phonetically expressed since Modern Hebrew does not geminate anymore, but the resistance to the otherwise regular (and synchronically active) syncope guarantees its underlying presence, and makes it recoverable. If no virtual geminate is assumed here, there is no way of viewing the Modern Hebrew syncope as a regular phonological process.

References

Barillot, Xavier & Philippe Ségéral forthc. a. On phonological processes in the '3rd conjugation' in Somali. To appear in Afro-Asiatic Grammar 3, edited by Denise Perett.

Barillot, Xavier & Philippe Ségéral forthc. b. La syncope en somali. To appear in Linguistique Africaine.

Bendjaballah, Sabrina 1999. Trois figures de la structure interne des gabarits. PhD. dissertation, Université Paris 7.

Benveniste, Emile 1935. Origine de la formation des noms en indo-européen. Paris: Maisonneuve.

Berman, Ruth 1997. Modern Hebrew. In: Robert Hetzron (ed), The Semitic Languages, 312-333. London: Routledge.

Bertram, Otto 1935. Der Wandel nd zu ng am Oberrhein. Zeitschrift für Mundartforschung **11**, 6-12.

Brugmann, Karl & Berthold Delbrück 1886-1900. Grundriss der vergleichenden Grammatik der indogermanischen Sprachen. Straßburg: Trübner.

Campbell, Lyle 1981. Generative phonology vs. Finnish phonology: retrospect and prospect. Phonology in the 80's, edited by Didier Goyvaerts, 147-182. Ghent: Story-Scientia.

Dell, François 1973. Les règles et les sons. Paris: Hermann.

Dinnsen, Daniel 1980. Phonological rules and phonetic explanation. Journal of Linguistics **16**, 171-191.

Dressler, Wolfgang 1974. Diachronic Puzzles for Natural Phonology. Papers from the parasession on Natural Phonology, edited by A.Bruck, R.Fox & M.La Galy, 95-102. Chicago: Chicago Linguistic Society.

Durand, Jacques & Laks, Bernard 1996. Why Phonology is one. Current trends in Phonology: Models and Methods, edited by Jacques Durand & Bernard Laks, 3-13. Salford: ESRI.

Encrevé, Pierre 1997. L'ancien et le nouveau. Quelques remarques sur la phonologie et son histoire. Langages **125**, 100-123.

Frings, Theodor 1955. Vom g, von seinen Lautwerten und von germanischen Sprachlandschaften. Rheinische Vierteljahresblätter **20**, 170-191.

Frings, Theodor & L.E. Schmitt 1942. Gutturalisierung. Zeitschrift für Mundartforschung **18**, 49-58.

Gamkrelidze, Thomas & Vjačeslav Ivanov 1995. Indo-European and Indo-Europeans. Berlin, New York: Mouton de Gruyter.

Goyvaerts, Didier 1981. Introduction. Phonology in the 80's, edited by Didier Goyvaerts, 1-26. Ghent: Story-Scientia.

Hellberg, Stefan 1978. Unnatural phonology. Journal of Linguistics **14**, 157-177.

Hönig, Fritz 1877. Wörterbuch der Kölner Mundart. Köln [2]1952: Bachem.

Hooper, Joan 1976. An Introduction to Natural Generative Phonology. New York: Academic Press.

Keenadiid 1976. Qaamuuska af-Soomaaliga. Florence: Le Monnier.

Kenstowicz, Michael & Charles Kisseberth 1977. Topics in phonological theory. New York: Academic Press.

Kenstowicz, Michael & Charles Kisseberth 1979. Generative Phonology. Description and Theory. San Diego: Academic Press.

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

Kiparsky, Paul 1974. On the Evaluation Measure. Papers from the parasession on Natural Phonology, edited by A.Bruck, R.Fox & M.La Galy, 328-337. Chicago: Chicago Linguistic Society.

Kiparsky, Paul 1982. Explanation in Phonology. Dordrecht: Foris.

Koutsoudas, Andreas 1980. The question of rule-ordering: some common fallacies. Journal of Linguistics **16**, 19-35.

Lessiak, Primus 1933. Beiträge zur Geschichte des deutschen Konsonantismus. Brünn, Prag, Leipzig und Wien: Rohrer.

Larsen, Bergeton 1994. Some aspects of vowel length and stød in modern Danish. MA thesis, Université Paris 7.

Lightner, Theodore 1981. New explorations into Derivational Morphology. Phonology in the 80's, edited by Didier Goyvaerts, 93-99. Ghent: Story-Scientia.

Lowenstamm, Jean 1991. Vocalic length and syllable structure in Semitic. Semitic Studies in Honor of Wolf Leslau on the occasion of his 85th birthday, edited by A.S. Kaye, 949-965. Wiesbaden: Harrassowitz.

Lowenstamm, Jean 1996. CV as the only syllable type. Current trends in Phonology: Models and Methods, edited by Jacques Durand & Bernard Laks, 419-441. Salford: ESRI.

Müller, Josef 1942. Einige Bemerkungen zur rheinischen Gutturalisierung, bei der Wörterbucharbeit aufgelesen. Zeitschrift für Mundartforschung **18**, 58-59.

Münch, Ferdinand 1904. Grammatik der ripuarisch-fränkischen Mundart. Bonn: Cohen.

Orwin, M. 1995. Colloquial Somali. A Complete Language Course. New York, London: Routledge.

Paul, Hermann, Peter Wiehl, Siegfried Grosse 1989. Mittelhochdeutsche Grammatik. Tübingen: Niemeyer, 23th edition.

444

Vennemann, Theo 1974b. Words and syllables in natural generative grammar. Papers from the parasession on Natural Phonology, edited by A.Bruck, R.Fox & M.La Galy, 346-374. Chicago: Chicago Linguistic Society.
Wrede, Adam 1958. Neuer Kölnischer Sprachschatz. 3 Vol. Köln: Greven.