Back to the Base.
On the Richness of the Base Hypothesis

Marc van Oostendorp, Meertens Institute, Amsterdam

1. Four dialects in the West-Germanic dialect continuum

One of the central assumptions of Optimality Theory which did not hitherto receive a lot of attention in the literature is *Richness of the Base*, stipulating that there are no constraints on the input structure of words, and that all linguistic constraints are statements on surface structure only. Richness of the Base is a strong hypothesis about natural language, and it has proven quite successful over the past few years.

Itô and Mester (1999) have shown that it is difficult to reconcile Richness of the Base with the standard conception of Sympathy as it has been proposed in McCarthy (1998). They illustrate this point with a few facts from two variants of German: Standard German and a variety they call Colloquial Northern German (CNG).

(1)		End of word	Before schwa	Before full vowel
	a. Standard German	diŋ	diŋə	diftɔŋgirə
		thing	*things*	*diphtongize*
	b. CNG:	diŋk	diŋə	diftɔŋgirə

This paper considers these same facts from these two German dialects, together with those of the closely related systems of Standard Dutch and Eastern Dutch. It argues that a strict interpretation of Richness of the Base runs into problems regardless whether it is combined with Sympathy or not. The relevant facts from Dutch are given in (2):

(2)		End of word	Before schwa	Before full vowel
	a. Standard Dutch	dɪŋ	dɪŋə	dɪftɔŋɣɪːrə
		thing	*things*	*diphtongize*
	b. Eastern Dutch	dɪŋk	dɪŋə	dɪftɔŋɣɪːrə

The argument runs as follows. Richness of the Base has it that all three of /dɪŋ/, /dɪŋg/ and /dɪŋɣ/ are potential inputs for all four dialect systems. It is plausible that all three of these inputs would give the output patterns in (1) and (2) in at least some of the systems. But in order to get this right, a large number of processes need to be invoked, such as hardening fricatives to voiced stops, spirantizing voiced stops to fricatives, deleting voiced velar obstruents, inserting voiced velar obstruents, etc. Since these processes are independent, we expect there to be much more variation than there actually is (the strange thing is that all of these processes seem to operate together). Furthermore, the constraints that are needed to describe these processes need radically different grammars between the different language systems. If, on the other hand, we would abandon the radical version of Richness of the Base and assume that for all dialects there is only one underlying form —/diŋ/ for Standard German and Standard Dutch, /diŋg/ for CNG and Eastern Dutch — the resulting grammars can be much simpler, and much more alike.

This squib is structured in the following way. In section 2 I will first discuss the dialects of German, the analysis given to these by Itô and Mester (1999) and the

problems these facts still pose for Richness of the Base. In section 3 I show how the Dutch facts complicate these problems even further. Section 4 then deals with the question what all of this means for the theory of grammar, and the role of the Richness of the Base hypothesis within that theory.

2. Two German Dialects

Itô and Mester argue for constraints of the following form (the precise formal definition does not matter to us here, but (3b) will nevertheless be replaced by something else below):

–   a.   *VC: No voiced codas
    b. *VCDC: No voiced dorsal plosives in complex codas
    c. Ident(voice): input specifications for [voice] should be respected in the output.
    d. Max: underlying segments should surface

(3a) is the constraint that is responsible for Final Devoicing: underlying /tag/ turns into [tak] in all varieties of German under discussion. We get this result if we assume the ranking *VC, Max» Ident(voice):

(4)

	/tag/	*VC	Max	Ident(voice)
a.	tag	*!		
b.	ta		*!	
c.	☞ tak			*

Voiced obstruents are not allowed in the coda. In order to solve this problem, we can do one of two things: either we delete the entire obstruent (thereby violating Max as in (4b)), or we only delete the specification for [voice]. Apparently the latter option is chosen; this shows that we need the ranking Max » Ident(voice).

    The constraint *VCDC becomes relevant in the analysis of /dɪŋg/, next to the two faithfulness constraints just mentioned. It is now quite easy to get the CNG pattern, we only need to put the *VCVC constraint in the position of *VC:

(5)

	/diŋg/	*VCDC	Max	Ident(voice)
a.	diŋg	*!		
b.	diŋ		*!	
c.	☞ diŋk			*

According to Itô and Mester, the problem mainly resides with Standard German forms. Standard German displays final devoicing effects as well as CNG, so that we have evidence for the ranking Ident(voice) » Max also here. Yet according to the logic of (4), this should mean that the surface form [diŋk] should be optimal in this variant as well.

    Now we could assume, of course, that in Standard German the input for these forms is /diŋ/, so that no problem with Max or Ident arises. The constraint Dep ('Do not insert material') will always choose the winner, independent of its ranking.

(6)

/diŋ/	*VCDC	Max	Ident(voice)	Dep
a. diŋg	*			*!
b. ☞diŋ				
c. diŋk				*!

Unfortunately, under the Richness of the Base hypothesis, we are not allowed to restrict the focus of our attention to the input /diŋ/ only; we need to have some plausible story about /diŋg/ as well. It does not necessarily have to surface as [diŋ], but it at least has to surface in *some* form, and it is not clear what the alternative would be:

– if it would be [diŋk], we would probably expect this to alternate in the plural with [diŋgə] or [diŋə], but the whole point is that such pairs do not exist in Standard German;

– an alternative would be that /diŋg/ would surface as   , i.e. not as a visible morpheme at all. This is a suggestion that is made for 'paradigmatic gaps' in Prince and Smolensky (1993). The problem with this solution in the current context (and under the assumption of standard OT) is that it requires a lot of faithfulness constraint violations that do not seem warranted; it simply is not clear in what sense   would be 'more optimal' than [diŋ] as an output of /diŋg/.

Furthermore, [g] still surfaces in forms such as *diptho*[ŋg]*ieren.* According to Itô and Mester (1999), this is an argument to assume that /g/ is underlying in words ending in [g], but of course this is not necessarily the case. We could also assume that the voiced velar obstruent has to be inserted in cases such as this. As a matter of fact, we probably have to assume that such a thing is happening in some cases, given that *something* plausible should happen to underlying /diftɔŋ/, under Richness of the Base.[1]

All in all, the most plausible option seems to be that the potential input /diftɔŋ/ and the potential input /diftɔŋg/ yield the same range of outputs, viz. [diftɔŋ - diftɔŋə - diftɔŋgirə] in Standard German, and [diftɔŋk - diftɔŋə - diftɔŋgirə] in CNG. Let us see what is necessary for these changes:

(7)   Standard German   CNG

		Standard German	CNG
Before word boundary	input /diftɔŋ/	do nothing	insert [k]
	input /diftɔŋg/	delete /g/	final devoicing
Before schwa	input /diftɔŋ/	do nothing	do nothing
	input /diftɔŋg/	delete /g/	delete /g/
Before full vowel	input /diftɔŋ/	insert [g]	insert [g]
	input /diftɔŋg/	do nothing	do nothing

It is actually not very difficult to describe these patterns if we allow ourselves to assume that the underlying form in Standard German is /diftɔŋ/ and the underlying form in CNG is /diftɔŋg/. Only two changes are needed in this case (apart from Final Devoicing

---

[1] One possibility is that in these cases underlying /ŋ/ would turn into /n/. An ingenious solution along these lines has been suggested to me by Birgit Alber. See section 4 for some discussion.

3

which is independently necessary as we have seen): in Common Northern German /g/ has to be deleted before a schwa, and in Standard German /g/ has to be inserted before a full vowel. It is not at present clear to me what is the reason behind these preferences, but that does not need to matter to us here. We can simply state the following constraints:

(8)   a.   [ŋə]: [ŋ] should be followed by schwa

      b.   *[ŋV]: [ŋ] should not be followed by a full vowel

With these two constraints, we can actually account for the full range of patterns in both dialects:

(9) Standard German input /…ŋ/	[ŋə]	Max(g)	*[ŋV]	Dep(g)
☞ diŋ	*			
diŋk	*			*!
☞ diŋə				
diŋgə	*!			*!
diftɔɲirən	*		*!	
☞ diftɔŋgirən	*			*

The only ranking that is relevant here is that *[ŋV] » Dep(g). The constraints are not crucially ranked otherwise. For CNG we have the following system:

(10) CNG input /…ŋg/	[ŋə]	Max(g)	*[ŋV]	Dep(g)
diŋ	*	*!		
☞ diŋk	*			
☞ diŋə		*		
diŋgə	*!			
diftɔɲirən	*	*!	*!	
☞ diftɔŋgirən	*			

In this case, the only relevant ranking is [ŋə] » Max(g). The ranking of the other constraints cannot be determined. Since this is so, Standard German and CNG could actually be assumed to have the same grammar (in which *[ŋV] » Dep(g) and [ŋə] » Max(g)) if this would prove useful.

Some questions still remain. For instance, the relatively low-ranking faithfulness constraints above refer specifically to [g]. The question is, of course, what would make this segment so special. One option might be to assume that /g/ somehow is an (almost) empty segment in this context, for instance because it shares many features with [ŋ]: /g/ could be easily deleted, because most of its features could be expressed by the velar nasal. The fact that it is [g] which is inserted in other cases in a similar position, might be

an indication in the same direction, since epenthetic segments are usually empty ones. In other words, if [g] is a close-to-empty segment, we *expect* it to be

Unfortunately, things are not as simple if we have to strictly adhere to Richness of the Base. The patterns before schwa and before a full vowel do not pose any specific problems, since they are the same in the two dialects. We only need to introduce a little more ranking among the constraints to deal with the 'alternative' inputs:

(11)

Standard German input /…ŋg/	[ŋə]	Max(g)	*[ŋV]	Dep(g)
☞ diŋə		*		
diŋgə	*!			
diftɔɲirən	*	*!	*!	
☞ diftɔŋgirən	*			

(12)

CNG input /…ŋ/	[ŋə]	Max(g)	*[ŋV]	Dep(g)
☞ diŋə				
diŋgə	*!			*!
diftɔɲirən	*		*!	
☞ diftɔŋgirən	*			*

The grammars of the two languages can thus still be the same. The pattern before pause is where the real problem is. Look at what happens to the input /diftɔŋ/ in Common Northern German: a [k] is inserted. One possibility is that we have a constraint like (13):

(13)   *[ŋ#]:   [ŋ] should not be followed by a word boundary.

This constraint is responsible for insertion of a [k] (rather than [g] because we are in a Final Devoicing environment), as well as for the prevention of a deletion of [g] or [k] in that same context. In other words, the ranking of CNG is:

(14)   *[ŋ#] » Max(g), Dep(g)

But now look at what happens to the inputs /diftɔŋ/ and /diftɔŋg/ in Standard German. The first is rather simple. It shows that Standard German has a different ranking from CNG. It is *worse* to insert a consonant than to satisfy *[ŋ#].

(15)   Dep(g) » *[ŋ#]

(16)

/diŋ/	Dep(g)	*[ŋ#]
diŋ		*
diŋk	*!	

The second potential input, /diftɔŋg/, should give us a very nice output, according to the constraint *[ŋ#], and furthermore, given that Final Devoicing is unviolated, an output which is attested in all varieties of German [ŋk]. Apparently, in spite of all this, underlying /g/ is still deleted in this environment in this variant of German.

This actually is the problem that is studied by Itô and Mester (1999). As we have seen above, these authors propose that the solution to this riddle is constraint conjunction. I will now translate their approach to the analysis sketched above. We have a faithfulness constraint prohibiting underlyingly voiced segments from losing their voice specification (17a). We also have *VCDC, a constraint which we can render here as (17b), in accordance with the format given for the other constraints:

(17)　a.　　Ident-voice: Underlying specifications for [voice] should be respected.
　　　　b.　　*[ŋC]: [ŋ] should not occur in a consonant cluster.

Both of these constraints are violated quite often in German: (17a) by all forms in which we see final devoicing (which is exceptionless) such as [tak] ('day', from underlying /tag/), and (17b) by all forms in which [ŋ] is followed by [k] such as [bɑŋk]. Itô and Mester (1999) propose, as we have seen above, that Ident-voice and *[ŋC] are conjoined into a new, inviolable, constraint in Standard German. [tak] is allowed to violate Ident-voice, and [bɑŋk] is allowed to violate *[ŋC], but *[dɪŋk] is not allowed to violate both constraints at the same time.

Let us accept this line of reasoning for the sake of the argument — after having noted that the mechanism simply is not necessary if we abandon the orthodox interpertation of Richness of the Base rather than Sympathy theory. We now have the following grammar accounting for the output [dɪŋ] from the input /dɪŋg/:

(18)

/diŋg/	Ident-voice　　*[ŋC]	Max(g)
diŋ		*
diŋk	*!	

This means that the explanation why /diŋg/ surfaces as [diŋ] is very different from the explanation for the input /diŋ/: it invokes different completely different constraints, and it is only an accident that the two processes 'conspire' to give the same output pattern. This would become more plausible if we would be able to find dialects in which there was a reason to really distinguish between the two inputs (one would surface as [diŋk] and the other as [diŋ] whereas they would pattern alike in the other parts of the paradigm.) This, however, is not the case, as far as I am aware. The relevant two types are dialects are Standard German and CNG; there are no 'in between' dialects, as we would expect. Connected to this is the problem that the grammars of Standard German and CNG are by necessity very different, just to account for the facts that words ending in [ŋ] in the one dialect end in [ŋk] in the other. A solution that does not have to strictly adhere to Richness of the Base does not suffer from these problems, as we have seen above.

3. Two Dutch dialects

If we now turn our attention to Dutch dialects, the situation becomes even more complicated. Roughly, we can distinguish two variants here as well: one is Standard

6

Dutch which corresponds to Standard German in most relevant respects, and the other is a variant which we will informally call Eastern Dutch here, and which corresponds to CNG.[2]

(19)		End of word	Before schwa	Before full vowel
	a. Standard Dutch	dɪŋ	dɪŋə	dɪftɔŋɣɪːrə
		thing	*things*	*diphtongize*
	b. Eastern Dutch	dɪŋk	dɪŋə	dɪftɔŋɣɪːrə

The only relevant difference between the Dutch and German dialects can be found in the context before a full vowel. Here, we find a fricative [ɣ] rather than [g] in both variants of Dutch. The reason for this probably is that [g] is not a phoneme of Dutch (even though it can be found in a number of fairly recent loanwords such as *goal* and *gate*). We can thus postulate a high ranking constraint for this language to the following effect:

(20)   *[g]: No voiced velar stops

The fact that [g] does not surface, of course does not necessarily imply that it cannot be underlying as well. As a matter of fact, the fact that in the singular we find a form like [dɪŋk] in Eastern Dutch, seems to point in the direction that this is the underlying form here; the fricative could then be due to a process of spirantisation (which might even be triggered by (20)). If we assume that Standard Dutch, like Standard German, has /dɪŋ/ as the underlying form and Eastern Dutch, like CNG, /dɪŋg/. The analysis can exactly be the same as for the two variants of German, except that we have to include (20):[3]

(21)	Standard Dutch input /...ŋ/	[ŋə]	Max(g)	*[ŋV]	Dep(g)	*[g]
	☞ dɪŋ	*				
	dɪŋk	*			*!	
	☞ dɪŋə					
	dɪŋgə	*!			*!	*!
	dɪftɔŋːrən	*		*!		
	☞ dɪftɔŋɣɪːrən	*			*	
	dɪftɔŋgɪːrən	*			*	*!

---

(22)

Eastern Dutch input /…ŋg/	[ŋə]	Max(ɡ)	*[ŋV]	Dep(ɡ)	*[ɡ]
dɪŋ	*	*!			
☞ dɪŋk	*				
☞ dɪŋə		*			
dɪŋɡə	*!				
diftɔŋɪːrən	*	*!	*!		
☞diftɔŋɣɪːrən					
diftɔŋɡɪːrən	*				*!

The constraint *[ɡ] is only relevant in the prevocalic context; in other contexts [ɡ] does not surface for independent reasons even in German dialects. The reason why it now is a voiced fricative which surfaces, is intuitively quite obvious: it is the segment which is closest to [ɣ].

If we have to strictly adhere to Richness of the Base, the analysis becomes much more complicated. The problems mentioned for the German dialects of course arise again. Also for Standard Dutch we need to assume a set of complicated constraint rankings and constraint conjunctions. It even becomes more obvious that the split into two possible dialects is a real one; there also are no 'in between' dialects in Dutch.

Furthermore it now becomes obvious that we could also assume that the *fricative* is underlying. This would of course not cause a very hard problem for the prevocalic cases, and we would need to assume a constraint to the following effect for the others:

(23)   *ŋF: A nasal should not be followed by a (homorganic) fricative.

This constraint is well known from the literature (cf. Padgett 1994, Van Oostendorp 2000). If the fricative is underlying, we could assume that there are two possible ways to react to *ŋF. Either the fricative disappears altogether (like in all Standard Dutch cases and the plural form in Eastern Dutch) or it hardens to a stop (as in the singular form in Eastern Dutch; Dutch, like German, dialects are subject to final devoicing.) A tableau for Eastern Dutch could look as follows:

(24)

/dɪŋɣ/	*ŋF	Max(ɡ)	Ident-contintuant (Ident-voice)
dɪŋ		*!	
dɪŋx	*!		(*)
☞ dɪŋk			*(*)

The question now arises however why, if *ŋF is so strong, the same ranking does not cause the same effect (in this case, hardening) in other circumstances, such as in the context before schwa:

8

(25)

/dɪftɔŋɣ/ + /ɪːrə/	*[ŋV]	*ŋF	Max(g)	Ident-contintuant (Ident-voice)
dɪftɔŋɪːrə	*!		*	
dɪftɔŋɣɪːrə		*!		(*)
☹ dɪftɔŋkɪːrə				*(*)

We could of course add another constraint (or possibly a conjunction of constraints) to the formal apparatus, but this probably will not add a lot of insight to our analysis. Furthermore, Richness of the Base, does not allow us to consider only inputs with a fricative. Eastern Dutch /dɪftɔŋ/ + /ɪːrə/ should also *epenthesize* a fricative, in spite of the constraints we have just argued for.

More generally, we need to have a story about inputs with a fricative, inputs with a voiced stop and inputs with just a velar nasal under the Richness of the Base hypothesis. Since these putative underlying forms correspond fairly closely to at least one of the output forms, it seems most reasonable to assume that these can surface in some (non-null) way, and if they do, that they would pattern alike. This would give us the following table:[4]

(26)

		Standard German	CNG	Standard Dutch	Eastern Dutch
Before word boundary	/diftɔŋ/	do nothing	insert [k]	do nothing	insert [k]
	/diftɔŋg/	delete /g/	devoice	delete /g/	devoice
	/diftɔŋɣ/	delete /ɣ/	harden, devoice	delete /ɣ/	harden, devoice
Before schwa	/diftɔŋ/	do nothing	do nothing	do nothing	do nothing
	/diftɔŋg/	delete /g/	delete /g/	delete /g/	delete /g/
	/diftɔŋɣ/	delete /ɣ/	delete /ɣ/	delete /ɣ/	delete / ɣ /
Before full vowel	/diftɔŋ/	insert [g]	insert [g]	insert [ɣ]	insert [ɣ]
	/diftɔŋg/	do nothing	do nothing	spirantise	spirantise
	/diftɔŋɣ/	harden	harden	do nothing	do nothing

One of the problems with this is, once again, that the processes that are needed to get a given result are quite diverse. For instance, in order to get the effect that velar nasals and full vowels are always separated by a velar fricative, we need to stipulate a process of fricative insertion and a process of spirantization. It is hard to see that there is a lot of independent motivation for the first process and given the other assumptions it is somewhat curious that we do not find spirantization as a solution in other contexts. Different solutions are chosen to similar problems in different contexts. I think this is an undesirable result of strict adherence to Richness of the Base.

---

[4] Another potential input, which should probably pattern in the same way, is /diftɔŋx/. This is omitted, because the operations needed for it are very similar to those necessary for /diftɔŋɣ/.

9

4. Conclusion: The role of Richness of the Base

We have seen that an analysis of the velar nasal in West Germanic dialects suffers from a few hard problems which could be solved easily if we would allow ourselves to say that the dialects involved all have one designated input form.

What does this mean for the theory? One possibility of course would be to simply abandon Richness of the Base altogether. This is not the most interesting solution: it does not take into account that Richness of the Base has been quite successfully applied in many other cases. Furthermore, not *all* questions are still answered by this move. Some have been just shifted to another level Why for instance do we find underlying /dɪŋ/ in some dialects and /dɪŋg/ in others? And why are there no dialects which show evidence for underlying /dɪŋɣ/?

A second possibility is that the analysis given above is misconstrued. It involves several constraints that are quite *ad hoc*; the reason for this is possibly that we do not know enough about the nature of velar nasals and velar obstruents. In the dialects discussed here, these segments display several other characteristics that have not been taken into account. For instance, velar nasals only occur in codas after short lax vowels in all dialects under discussion; they are not found in onsets or after long tense vowels. Similarly, the velar fricative is very variable in Dutch and German dialects also when it is not adjacent to a velar nasal. All in all, it is clear that we have not construed a complete theory of velars yet. It is possible that this ideal theory will give us such a deep insight into the nature of these segments that the problems noted here will disappear. At this point, I know of no representational theory that has the desired properties.

The third option is suggested to me by Birgit Alber (p.c., 11.8.2000) She notes that it is possible to assume that in all cases the input /ŋ#/ is 'redirected' to [n#] instead. This could be done by stipulating that the dorsal place of articulation can only surface if it belongs to a stop underlyingly. This effect could be attained by stipulating for instance the constraint ranking in (27):

(27)   Ident(dorsal-of-stop)»*ŋ»Ident(dorsal)).

This then solves the problems for CNG and Eastern Dutch: if these constraints are ranked high enough in these languages, we have found a way *around* Richness of the Base: certain structures can still be input (/ŋ#/) but they will be neutralized (in this case to /n#/) by very high ranking and very specific faithfulness constraints.

As for Standard German and Standard Dutch we could now assume that they also have high-ranking (27), and then continue along the lines suggested by Itô and Mester (1999), including all the problems this has. Notice that the grammars of these languages now are really very complicated and cause input-output relations to be extremely abstract, since every morpheme that ends in /ŋg#/ surfaces as [ŋ#] most of the time, while at the same time /ŋ#/ is redirected towards [n#]. Also the problem that the two types of grammars are very different and we would expect all kinds of in-between variants that are not actually attested is not solved under this approach to the problem.

The fourth option I want to consider is to look at the role of Richness of the Base within the theory once again. Intuitively, it is quite clear why all inputs in Standard Dutch and Standard German end in /ŋ/, while those in Eastern Dutch and CNG end in /ŋg/: because these are the forms that correspond to the output of the underived form (modulo final devoicing). It therefore is no wonder that *in the lexicon* these forms have got exactly this underlying form. Even new forms (loanwords and the like) could get a

10

'proper' underlying form, at least as long as there is no evidence that the forms behave differently. If the language user has a choice — either giving a new form an underlying form which resembles that of other words in the language and for which there is a fairly simple grammar, or giving it a completely different form which requires a different grammar — it is safe to assume he will choose the former route.

The fixedness of the underlying forms in the dialects in question thus can be seen more or less as a historical accident. Because /ŋg/, /ŋ/ and /ŋɣ/ are in complementary distribution, all three of them should be able to serve as potential inputs. But since some of them require a very complicated machinery — if they allow to derive the right patterning at all — they will not be used. Richness of the Base predicts however that people will be able to learn the alternative patterns. If we would confront people in some experimental setting with words which pattern in some other way, they should be able to learn these. And language users will be able to use different systems at the same time. For instance, there probably are no pure monolingual speakers of CNG or of Eastern Dutch any more: all these speakers also use some version of the Standard. This means that they should be able to produce both [dɪŋk] and [dɪŋ], depending on social and other circumstances. This is what we predict if both Richness of the Base and the analysis put forward on the previous pages are right. In principle, variation is possible, but in the core of a language system we find only one plausible input form for every attested output pattern.

References

Itô, Junko, and Armin Mester. 1999. On the Sources of Opacity in German. Coda Processes in German. Manuscript, UCSC.

McCarthy, John. 1998. Sympathy and Phonological Opacity. Manuscript, University of Massachusetts at Amherst.