Szemerényi's Law and Stang's Law in Nonlinear Phonology

Bert Vaux, Harvard University

Unpublished manuscript, 1994

1. Introduction

Szemerényi's Law and Stang's Law describe the development of long vowels in the nominative and accusative singular respectively of animate athematic sonorant stems in Proto-Indo-European. I present a standard example of each of these laws in (1) and (2):

(1) Szemerényi's Law (Sz)
Greek paté:r 'father-nom' < PIE *ph$_2$té:r < Pre-PIE **ph$_2$tér-s

(2) Stang's Law (St)
Greek zê:n 'Zeus-acc' < PIE *dyé:m < PPIE **dyéw-m

Assuming for the sake of argument that the insights of Szemerényi and Stang were correct, these laws raise several interesting theoretical and historical questions. First, is it possible that both laws are results of a single phonological process? Second, did these laws involve assimilation or deletion? And finally, what do these processes tell us about Indo-European syllabification, specifically concerning continuants and nasals? I propose that both Szemerényi's Law and Stang's Law are manifestations of an Indo-European constraint forbidding adjacent identical place specifications within consonant clusters; violations of this constraint produced by adding the nominative *-*s* to coronal stems and the accusative *-*m* to labial stems are repaired by delinking one member of the offending cluster, and spreading the features of the other member to the vacated timing slot. At a later stage of Proto-Indo-European all geminates are simplified, causing the compensatory lengthening of the preceding vowel reflected in the two laws.

This analysis raises several problems concerning Indo-European syllabification, the selection of the segment to be assimilated, and the behavior of degemination in later PIE, which I consider in section 3 of this paper. I first present in section 2 a systematic survey of the basic data involved in Sz and St, which so far have not received a comprehensive treatment in the literature. Finally, in section 4 I offer some conclusions and final thoughts.

2. Basic data

Though Mayrhofer 1986 discusses Stang's Law to a limited extent, there is still no systematic presentation of the facts pertaining to Sz and St. In this section I attempt to clarify the exact environments in which the two processes occur and do not occur, in order to get a better idea of

how they might be related to each other and how they might be accounted for in phonological terms.

2.1. Lex Szemerényi

Szemerényi's Law (Szemerényi 1956; cf. Szemerényi 1996:181-2) receives a superficial treatment in Collinge 1985:237-8, where for spurious reasons it is classed as a "minor law"; Mayrhofer 1986:121 (and references cited there) grants it only passing mention. I have summarized in (3) what appears to be the basic distribution of Sz:

(3)	**Sanskrit**	**PIE**	**Pre-PIE**
(a) showing Sz			
**-rs	Gk. paté:r 'father'	*ph$_2$té:r	**ph$_2$tér-s
**-ls	Lat. sa:l 'salt'[?]	*sa:l	**sal-s
**-ns	śvá: 'dog'	*k^{j}wó:n	**k^{j}wón-s
**-ms	Gk. xθó:n 'earth'[1]	*g^{hj}d^{h}ó:m	**g^{hj}d^{h}óm-s
**-ys	sákha: 'friend'	*sók^{w}h$_2$o:y	**sók^{w}-h$_2$-oy-s
**-ss	apá:s 'active'	*h$_1$epó:s	**h$_1$epós-s
**-ts	pá:t 'foot'	*pó:t[2]	**pót-s
(b) not showing Sz			
**-ws	Gk. Zeús 'Zeus'	*dyéws[3]	**dyéw-s
**-h$_1$s	pántha:s 'path'	*pónth$_2$eh$_1$s	**pónt-h$_2$-eh$_1$-s
	Av. mazda:	*mn̥s-d^{h}éh$_1$s	**mn̥s-d^{h}éh$_1$-s
**-h$_2$s	sthá: 'standing'		
**-h$_3$s	??		

[1] If the proposal that I make later in this paper that **-ms assimilated to **-ns at some point before the application of Sz is correct, then the cited PIE form for 'earth' would more properly be *g^{hj}d^{h}ó:n. But NB. Skt. kṣá:s < *g^{hj}d^{h}ó:s, and also Avestan zyå: 'winter' < IIr *g^{hj}hya:s < PIE *g^{hj}hyems, which indicate that the final m of the stem assimilated to the suffix -s. One could argue that the nominative n was restored by analogy to the rest of the paradigm.

[2] Though the stem of 'foot' ends in *-d, it normally devoiced in PIE before *s. Consequently, I assume the proto-form immediately before the loss of final *-s was **pó:ts. Once the *-s was lost, I do not know whether the *-t remained, or reverted to a d by analogy to the rest of the paradigm. We cannot tell from Sanskrit, which has final devoicing. This point is not important for my analysis.

[3] In light of Sanskrit dyá:uṣ we might prefer to reconstruct *dyé:ws and invoke Osthoff's Law to get the Greek short vowel. For the sake of argument I have employed the short vowel form in my chart, following Szemerényi 1956.

The basic distribution, if we suppose that **-ms sequences assimilated to *-ns at an early stage, is that stems ending in coronal sonorants undergo Sz, and all other types of stems do not. I assume that the case of coronal obstruents, which appear to show Sz, is actually different than that of coronal sonorants.

Note that Skt. *pántha:s* also shows lengthening with preservation of the *-s*. Given that 'path' does not show Stang, we might want to assume that it does undergo Sz, for reasons to be discussed below. I argue to the contrary that 'path' and laryngeal stems in general pattern with obstruent stems in not undergoing Sz or St, but get their lengthened nominative from laryngeal loss. The feminines in *-h_2- must then receive an alternate explanation.

Palatal stems in Sanskrit such as *tvác* 'skin', *bhiṣaj* 'physician' do not appear to undergo lengthening, though palatals belong to the class of coronals. The Sanskrit palatals develop from IE dorsals, though, and were dorsals at the point where Sz applied. Laryngeal-final stems are problematic: some seem to show Sz, and others do not. We must examine the Stang facts before considering this problem further.

2.2. Lex Stang

Stang's Law (Stang 1965, Mayrhofer 1986:163 and references cited there) is claimed by Mayrhofer to occur only in roots ending in *-*w* and *-h_2 (he does not take a position on the other laryngeals). To these I suggest we add *-*m*, as in the word for 'earth', which shows Stang in both Greek and Sanskrit. Other types of stems do not appear to show Stang. I represent the overall distribution in (4):

(4)	**daughter language**	**PIE**	**Pre-PIE**
(a) not showing St			
**-rm	Gk. patér-a 'father-acc'	*ph$_2$térm̥	**ph$_2$tér-m
**-lm	Lat. salem 'salt-acc'[?]	*salm̥	**sal-m
**-nm	Skt. śvá:nam[4] 'dog-acc'	*k$^{\text{j}}$wónm̥	**k$^{\text{j}}$wón-m
**-ym	Skt. sákha:yam 'friend-acc'[5]	*sók$^{\text{w}}$h$_2$oym̥	**sók$^{\text{w}}$-h$_2$oy-m
**-sm	Skt. apásam 'active-acc'	*h$_1$epésm̥[6]	**h$_1$ep-es-m
**-dm	Skt. pá:dam 'foot-acc'[7]	*pódm̥	**pód-m

[4] With long a: by Brugmann's Law.

[5] With long a: by Brugmann's Law, paralleled in Av. kauuae:m 'poet-acc' < IIr. *kawa:yam < IE *kewoym̥.

[6] I assume for this word ablaut of the type common with *-es- suffixes: nom.sg. o-grade, elsewhere e-grade.

[7] With a: by Brugmann's Law.

**-h_1m	Skt. pántha:m ‘path-acc’	*pónth$_2$eh$_1$m	**pónt-h$_2$-eh$_1$-m
	Av. mazdãm[8]	*mn̥s-d^{h}éh$_1$m	**mn̥s-d^{h}éh$_1$-s
**-h_2m	Skt. sthá:??		
(b) showing St			
**-wm	Gk. zê:n ‘Zeus-acc’	*dyé:m	**dyéw-m
**-mm	Skt. kṣá:m ‘earth-acc’	*g^{hj}d^{h}ó:m	**g^{hj}d^{h}óm-m

The basic generalization here is that labial stems have an accusative singular with non-syllabic *-m, and all other stems take syllabic *-m̥ (note, however, that the generally problematic *newm̥ ‘9’ does not show St, though we might expect it to). Obstruent stems never show Stang, except perhaps for the problematic feminine *-h$_2$- stems of the devi: class. Outside of the feminine laryngeal stems, which will be discussed below, the facts are relatively straightforward: sonorant-final stems which show Sz do not show St, and vice versa. Consequently we can extrapolate that forms which do not show St in the accusative most likely undergo Sz in the nominative, even though the nominative forms on their own might not indicate one way or the other, as in the cases of pántha:m and mazdãm. This complementary distribution is quite striking, and makes us suspect that the two processes might be interrelated, and perhaps even result from a single process. In order to begin investigating this possibility, however, we must first address the problem of laryngeal stems, in particular feminines.

Laryngeal stems are problematic: feminine stems in *-h$_2$ behave differently from root nouns in *-h$_1$ and *-h$_2$. The accusative ending of pántha:m and mazdãm scans disyllabically, indicating that the laryngeal did not assimilate to the following m, so that St had not applied. Conversely, feminine accusatives in -a:m never scan as two syllables, indicating that they did undergo St; in addition, the intonation of the Balto-Slavic accusative points to IE *-a:m, with no laryngeal, indicating that Stang applied. According to the principle described above we should then expect that *-h$_2$ feminines did not undergo Sz, but this leaves us without an explanation of their asigmaticity. By the same token, we should expect that devi: feminines would not show St, whereas vr̥ki: feminines would, but we get the opposite result. Some have suggested an alternate principle, “sigmatic nominative → no Stang” [ref?]. This accounts for feminines, but not for other stem types: ‘sky’, for example, has a sigmatic nominative in Greek and Sanskrit but shows St; ‘earth’ behaves similarly in Sanskrit and Avestan, though it is asigmatic in Greek. In short, the feminine facts as a whole go exactly counter to the rest of the data; nevertheless, it would be nice to correlate the opposition in sigmaticity between the devi: and vr̥ki: classes with their Stang behavior.

[8] ã is always long in Avestan.

We should bear in mind during all of this discussion, however, that laryngeal stems might not behave like sonorant stems. We initially assumed that they did out of a desire to collapse the complementary -s/Stang distribution of the devi: and vṛki: feminines with the structurally-similar behavior of sonorant stems; it could be, however, that root nouns such as 'path' and 'Mazda' show the regular outcomes of laryngeal stems with respect to Sz and St, and feminines require another explanation. I explore this possibility in the next section.

3. Analysis

For the purposes of this analysis I assume that Sz applies to coronal sonorant stems (if we assume that **-ms assimilated to **-ns before Sz applied), and St applies to labial sonorant stems; I leave the problem of laryngeals for later in the discussion. It is important to notice that these place features are exactly complementary within the class of sonorants in Indo-European: {r l n y} are coronal, and {w m} are labial. If we wish to include *s* and the laryngeals in our rules, which is not necessarily required by the data, we have to bring in the feature [+continuant], which would exclude *m* and *n*. Descriptively speaking, then, Sz and St apply within the set of consonants that are not [-sonorant, -continuant]. This suggests that our two laws are somehow sensitive to syllabification, which tends to distinguish oral stops from other consonants. In addition, both processes seem to be sensitive to place features: the coronal *s* of the nominative singular is sensitive to coronal stems, and the labial *m* of the accusative singular is sensitive to labial stems. In this section I attempt to determine exactly what roles syllabification and place constraints are playing in the workings of Sz and St, and consider whether the two rules can be collapsed into one.

As mentioned above, oral obstruents behave differently than other consonants with respect to Sz and St. I believe that this fact follows from the nature of IE syllabification, which is sensitive to the difference in sonority between oral obstruents and other consonants. This can be seen in the process of nucleus assignment, which cannot target [-sonorant, -continuant] segments. Interestingly, among the class of [-sonorant, +continuant] segments IE syllabification distinguishes between laryngeals, which can be syllabified as nuclei, and s, which cannot. The phoneme /s/ is also distinctive in being able to serve as a syllable appendix, notably in word-initial position (s-mobile). Given that *-s is preserved after obstruents (e.g. Lat. re:x 'king' < *re:g^{hj}-s), we are forced by the sonority sequencing principle (SSP) to assume that IE had an appendix position available at the end of a word as well, and that this position could be occupied by s. In cases where -s followed a sonorant, however, the SSP would not be violated, and the s could be syllabified in the coda with the sonorant.

Based on the fact that Sanskrit does not allow word-final obstruent+s clusters, and given the existence of forms such as Greek poús (Doric pó:s), Latin pe:s 'foot', we should postulate an immediate pre-breakup form *pe:/o:ts, which does not show the s-loss that generally characterizes

Sz. I propose that these sequences behaved differently from sonorant-s sequences because IE converted **-ts sequences into *t^{s} affricates, which could trigger a compensatory lengthening process when the timing slot of the s was lost. Dental affricates are postulated for IE in cases where two dentals come together across a morpheme boundary; in these cases **-t#t- > *-t^{s}t-, which yields -zt- (/t^{s}t/) in Hittite, -st- in Avestan, Greek, and Balto-Slavic, -tt- in Sanskrit, and -ss- in Italic, Celtic, and Germanic. Note that the nominative of 'foot' shows (in the branches where it is not lost or thematized) the very outcome we would expect from an original *t^{s}:

(4)	*t^{s}	'foot'
Sanskrit	t	t
Greek	s	s
Latin	s	s
Hittite	t^{s}	thematized
Avestan	s	thematized
Germanic	s	thematized
Celtic	s	replaced
Balto-Slavic	s	replaced

It would be difficult to motivate compensatory lengthening resulting from monosegmentalization of a **ts sequence. We would either have to assume that IE was trimoraic, and employed double flopping (cf. Hayes 1989, and discussion below), or that there was an intermediate long affricate, parallel to the long consonants we propose as an intermediate stage for Sz and St, but both of these alternatives are problematic.

This suggests a natural environment for the application of Sz, namely the coda. Given that coronal sequences which would be syllabified within a coda (rs ls ns ys ss) undergo Sz, whereas those which would be coda-appendix sequences (ts ds d^{h}s) do not, we can say that the OCP applied on the Place tier in IE codas, i.e. identical Place specifications were not allowed to surface within single coda positions. Assimilation within codas but not between codas and appendices is common among the world's languages (cf. Reiss 1994 on Old Norse, Calabrese 1994 on Gothic). For Indo-European we might want to state more generally that appendices do not participate in Place restrictions, because PIE was apparently sensitive to coda-onset sequences as well—labial stems did not take *-mo- suffixes, for example.

Place cooccurrence restrictions are well-known in the Semitic languages, where the first two consonants of roots are not allowed to contain the same place feature (McCarthy 1991:15). Indo-European shows signs of similar place restrictions in its root structure: TeT, DeD, D^{h}eD^{h}, and D^{h}eT, and TeD^{h} roots are not allowed (Watkins 1992:31), and though they are allowed in

principle, I do not know of any D^heD or DeD^h roots. Nasal roots of the type *d^hen do appear to be allowed, however. I propose that Indo-European had active OCP constraints on the Place tier very similar to those proposed by McCarthy 1979 for Semitic; in addition to applying to root structures, however, the IE place constraints also applied within codas, resulting in Sz and St, and between codas and onsets, producing the asymmetry with *-mo- suffixes mentioned above, and perhaps also cases of dissimilation such as Sanskrit adbhís 'water-instr.pl' < *ap-bhis. We will consider how these OCP violations were repaired below.

The next question then is whether the accusative cases can be described in the same way as the nominatives. The accusative is more problematic because it involves *m, whose level of sonority relative to the other sonorants is ambiguous. In initial position, m+sonorant onsets are allowed: *mneh$_2$- 'remember', *mlewH- 'speak', *mreg^{hj}wih$_2$ 'short'. In final position, however, the mirror-image sequences are not allowed, and the *m is syllabified as a nucleus: *k^{j}wónm̥ 'dog-acc', *salm̥ 'salt-acc', *ph$_2$térm̥ 'father-acc'. I believe that this asymmetric behavior results from the fact that IE allowed different levels of complexity in its onsets and codas. Clements (1992) has established a universal complexity hierarchy of syllable onsets and codas, reproduced in (5):

(5) onsets OL >> OG >> [NG] >> [LG] >> ON >> NL [OL least complex, NL most complex]
codas GL >> GO >> NO >> GN >> LN >> LO [GL least complex, LO most complex]

It appears that PIE allowed onsets up to and including NL, though I know of no NG or LG cases. Interestingly, underlying initial {ur ul ui} sequences before vowels were syllabified as{wr- wl-wy-}; I assume that the w- in these cases must have been syllabified as an appendix. By the same token one could argue that IE actually allowed only onsets up to OG, and the ON, NL, GL and GG cases were all syllabified as appendix-onset sequences. This would account for the lack of NG and LG onsets, but would require the stipulation that IE could appendicize segments that by the SSP might be expected to syllabify within the onset.

On the other hand PIE only allowed codas up to NO; in sequences of GN, LN, and ON, the nasal is syllabified as a nucleus, and in LO sequences, if they existed, the obstruent would have been syllabified as an appendix.

This analysis faces two problems, however: the negative *n- becomes syllabic before initial sonorants, and the accusative *-m apparently syllabifies in the coda of feminine laryngeal stems. In the case of the negative prefix we can say that the syllabic form was extended by analogy, as may have been the case before vowels, where we would expect *n- but instead get reflexes of *n̥. The laryngeal case is more difficult: not only do we get non-syllabic m in feminines, but we also get the expected syllabic m̥ in non-feminine laryngeal stems such as 'path' and 'Mazda'. Given that non-feminine laryngeal stems show the expected Sz and St outcomes and the expected

syllabification, I assume that it is the feminines which are aberrant. One possibility is that feminines (except for the vṛki: type, which shows the expected syllabic ṃ in the accusative) did not have a laryngeal in the accusative at the time St applied; this idea has many ramifications which I cannot explore in this paper[9].

If we disregard feminines for the moment we can say that St, like Sz, applies to OCP violations within codas--in this case, the rule targets adjacent labial specifications. It is not possible to test whether appendix labials undergo Stang or not, because labials are not allowed in IE appendices (cf. English, which only allows coronal segments to be syllabified as appendices). Consequently, it seems that the two laws are actually manifestations of a single process, which acts to repair OCP violations within IE syllable codas. The next question then is what exactly the nature of this process is.

There are basically two possible ways that the OCP violations produced in Sz and St cases could be repaired: either one of the segments could be deleted, or one could assimilate to the other; we would also want this to trigger compensatory lengthening of the preceding vowel. I have represented these possibilities in (6) and (7):

(6) deletion

(a)	Szemerényi	**VRs	>	*V:R
(b)	Stang	**VRm	>	*V:m

(7) assimilation

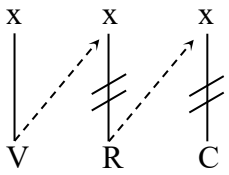
(a)	Szemerényi	***VRs	>	**VR:	>	*V:R
(b)	Stang	***VRm	>	**Vm:	>	*V:m

Although the deletion analysis superficially appears to be simpler, it involves invoking the mechanism known as “double flopping” (Hayes 1989), whereby segments shift over timing slots or moras, as in (8):

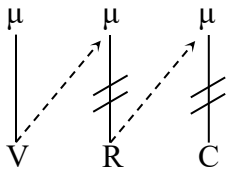
[9] For example, we would no longer be able to motivate the long vowel in the weak cases of deví: feminines.

(8) double flopping

(a) in X-theory



(b) in moraic theory



This mechanism crucially involves assuming that IE was trimoraic, which is not supported by any other evidence. Consequently, double flopping should be avoided if any viable alternatives exist.

The assimilation analysis works well for the most part: we simply assume that the OCP violations are repaired by delinking one member of the offending cluster and spreading the features of the surviving segment to the vacated timing slot. The problem lies in determining the principle governing which segment is assimilated: in the Sz cases the second and less sonorous segment assimilates, and in the St cases the first and more sonorous segment assimilates. The *m-m cases are not a problem in this regard, because linear order and relative sonority do not need to be invoked. In order to get the Stang facts we must assume that IE *m was less sonorous than *w, but this has been suggested for independent reasons by Halle and Calabrese (forthcoming).

We must then explain the simplification of the geminate and the compensatory lengthening of the vowel. Given that PIE, like modern English, did not allow geminates (except in some expressive formations; see Watkins 1992:15), it is a relatively simple matter to propose that pre-PIE geminates produced by Sz and St were simplified at the PIE stage, triggering compensatory lengthening of the preceding vowel. This type of process is also attested in several Middle Indic languages, where reduction of geminates produces lengthening of the preceding vowel.

4. Conclusions

I have argued that Szemerényi's Law and Stang's Law, which heretofore have been treated as independent processes, are both actually manifestations of a collaboration between two constraints well-attested in other areas of Indo-European phonology, namely OCP constraints on the Place tier, which disallow adjacent identical place feature specifications in IE roots, codas, and syllable junctures, and a constraint against geminates, which reduces assimilated geminate structures

produced by Sz and St and triggers compensatory lengthening of the preceding vowel. In addition, I have shown that the exact complementarity of Sz and St is not an inherent characteristic of these rules, but results from the fact that all sonorants in IE were either coronal or labial. I have also argued that the feminines in *-h_2-, which formerly were thought to be standard cases of Sz and St, actually show the opposite of Sz and St outcomes, and should receive an alternative explanation.

References

Calabrese, Andrea (1994) Sievers' Law in Gothic: A synchronic analysis and some speculations on its historical development. *The Linguistic Review* 11:149-194.

Clements, G. Nick (1992) The Sonority Cycle and Syllable Organization. Phonologica, W. Dressler, ed.

Collinge, N. (1985) The Laws of Indo-European. Philadelphia: Benjamins.

Dunkel, George (1992) Sprache 34:2.

Halle, Morris and Andrea Calabrese (forthcoming) Theoretical Indo-European Phonology. Manuscript, Harvard and MIT.

Hayes, Bruce (1989) Compensatory Lengthening in Moraic Phonology. Linguistic Inquiry.

Kuryłowicz, Jerzy. Indogermanische Grammatik. Heidelberg: Winter. [p. 220]

Manaster-Ramer, Alexis (1996) The Szemerényi-Stang Laws and PIE *-Hs Clusters. Historische Sprachforschung 109:163-165.

Mayrhofer, Manfred (1986) Indogermanische Grammatik I: Phonologie. Wiesbaden: Harrassowitz.

McCarthy, John (1991) The Phonology of Semitic Pharyngeals. Manuscript, UMass Amherst.

Reiss, Charles (1994) The outcome of /lr/ and /nr/ in Old Icelandic. *The Linguistic Review* 11:329-350.

Stang, C. (1965) Indo-europeen *gwōm, *d(iē)m. Symbolae linguisticae in honorem Georgii Kuryłowicz = Prace komisii jazykoznawczej 5: 292-296.

Strunk, Klaus. (1988). [Review of Indogermanische Grammatik. Band I, by J. Kurytowicz, M. Mayrhofer, W. Cowgill, A. Bammesberger, & M. Peters]. Historische Sprachforschung / Historical Linguistics, 101(2), 302–312.

Szemerényi, Oswald (1956) Latin rēs and the Indo-European long-diphthong stem nouns. Zeitschrift für vergleichende Sprachforschung 73:167-202.

Szemerényi, Oswald (1962) Trends and tasks in comparative philology. London: University College.

Szemerényi, Oswald (1996) Introduction to Indo-European Linguistics. Oxford: Oxford University Press.

Watkins, Calvert. 1998 [1993]. Proto-Indo-European: Comparison and Reconstruction. In: *The Indo-European Languages,* ed. by Anna Giacalone Ramat & Paolo Ramat, 25-73. Translated from the Italian. London & NY: Routledge.