Assessing Similarity Avoidance in the Evaluation of PIE Root Structure Constraints

Adam I. Cooper, Cornell University

The Proto-Indo-European (PIE) root has served as the domain of a number of posited constraints on well-formedness, particularly restrictions on root shape and the co-occurrence of certain consonants (Benveniste 1935, Szemerényi 1996, Weiss in press, et al.). Such restrictions, a sampling of which is presented in (1), are generally captured in skeletal schemata which, though descriptively adequate, do not account for the factors potentially motivating this state of affairs.

This paper explores and assesses a means to provide an analysis of PIE root structure constraints in a way which does identify such motivations, by adopting a framework found in current phonological theory, namely the similarity avoidance (SA) approach developed by Frisch, Pierrehumbert and Broe (2004) to account for co-occurrence restrictions in the Arabic triliteral root system. Frisch et al.'s account basically consists of two components: first, the identification of co-occurrence restrictions through calculation of a ratio of observed / expected forms (O/E) (2), whereby an O/E value of less than one suggests a restriction; and second, the calculation of similarity between segments through the natural classes similarity metric, given in (3).

I have constructed a database comprising 630 confidently-reconstructed PIE verbal roots collected in Rix (2001), and have computed the O/E values of consonants in prevocalic and postvocalic positions (the **-CVC-** core) in these roots. The values attained support the general validity of PIE root structure constraints such as those in (1), but reveal that indeed the situation is much more fine-grained than can be conveyed by these individual statements alone. Similarity across multiple dimensions, involving both place and manner of articulation, can influence the compatibility of two consonants to co-occur within a root. The nexus of these two influences is seen in the absolute lack of co-occurrence in any verbal root in Rix of the two PIE liquids /l, r/ (O/E = 0.0000), which of course share both place and manner features. Save for Ringe (1998), the restriction suggested by this finding has generally gone unmentioned in current work on PIE.

The picture is complicated, though, when looking more closely at certain sets of consonants. On the one hand, SA can indeed provide an explanation for the restricted co-occurrence between identical consonants (1a.), based on their absolute identity (similarity = 1.0000). Also, the relatively higher similarity values holding within the class of labials, due to its smaller size as compared to, for instance, the coronals, can account for the greater degree of restricted co-occurrence between these consonants (1b.). Both of these results are similar to those obtained by Frisch et al. for Arabic. Finally, the newly-recognized constraint on co-occurrence of the PIE liquids can be accounted for as well, given the strength of their similarity (0.8333).

On the other hand, the constraint on voiced unaspirated stops (1c.) is not satisfactorily dealt with by SA, given that similarity values holding between members of this series of stops are practically no different than those holding between members of the other two series of PIE stops, voiceless unaspirated and voiced aspirated (4). Further, the free co-occurrence of voiced aspirated stops (O/E = 2.4561) shows that more than similarity avoidance must be relevant in the morpho-phonotactics of the PIE root. Such findings can perhaps corroborate the claims made by Iverson and Salmons (1992) and Barrack (2002) about stop co-occurrence in PIE and the invalidity of the purported constraint *DVD-; their explanation for the relevant phenomena is, however, problematic itself – the claim that stops are disfavored in postvocalic position is not borne out by a more careful consideration of the data – and thus this issue lacks satisfactory treatment.

In sum, these mixed results may raise concerns about the SA approach. Still, in subjecting PIE to a means of analysis afforded by current approaches in phonological theory – an exercise which has been attempted previously, though not unproblematically, in, for instance, the statistical study of PIE root forms by Jucquois (1966) and work promoting the Glottalic Theory (Hopper 1973, Gamkrelidze & Ivanov 1973, et al., but see Barrack 2002 for refutation) – this study not only sheds new light on the phonotactic restrictions of PIE, but it also seeks to bridge the methodological gap between traditional PIE scholarship and current phonological theory.

(1) **Some PIE root consonantal co-occurrence restrictions**

    a.   *C_iVC_i-   in which the consonants are identical.
    b.   *mVP-/PVm-   in which *m* is the labial nasal and *P* is any labial oral stop.
    c.   *DVD-   in which *D* represents a voiced unaspirated stop.
    d.   *TVD^h-/D^hVT-   in which *T* represents a voiceless stop and D^h represents a voiced
                                  aspirated stop.

(2) **O(bserved)/E(xpected** = $\dfrac{\text{Observed } \{C_1, C_2\} \text{ co-occurrence in roots}}{\text{Total roots}}$

$$\frac{\text{Observed } /C_1/ \text{ occurrence in roots}}{\text{Total roots}} \times \frac{\text{Observed } /C_2/ \text{ occurrence in roots}}{\text{Total roots}}$$

(Khan (to appear): 5)

(3) **Similarity**   = $\dfrac{\text{Shared natural classes}}{\text{Shared natural classes + Non-shared natural classes}}$

(Frisch et al. (2004): 198)

(4) **Similarity values for the three PIE stop series**

    a.   Voiceless unaspirated (O/E = 1.3315)

	p	*t*	*ḱ*	*k*	*kʷ*
p	1.0000	0.3684	0.3684	0.3684	0.3500
t		1.0000	0.3333	0.3333	0.3182
ḱ			1.0000	0.6471	0.6111
k				1.0000	0.6111
kʷ					1.0000

    b.   Voiced unaspirated (O/E = 0.0000)

	b	*d*	*ǵ*	*g*	*gʷ*
b	1.0000	0.3462	0.3462	0.3462	0.3333
d		1.0000	0.3333	0.3333	0.3214
ǵ			1.0000	0.6364	0.6087
g				1.0000	0.6087
gʷ					1.0000

    c.   Voiced aspirated (O/E = 2.4561)

	bʰ	*dʰ*	*ǵʰ*	*gʰ*	*gʷʰ*
bʰ	1.0000	0.3750	0.3913	0.3913	0.3750
dʰ		1.0000	0.3750	0.3750	0.3600
ǵʰ			1.0000	0.6842	0.6500
gʰ				1.0000	0.6500
gʷʰ					1.0000