

Cluster splittability in Tagalog: corpus and survey evidence

C_1C_2 onsets differ in “splittability”. Cross-linguistically, in loan adaptation, epenthesis into stop-glide/liquid onsets is more likely than into *s*-nasal or *s*-stop onsets (Singh 1985, Broselow 1987, Fleischhacker 2002, and others); within sibilant- C_2 clusters, the more sonorant C_2 , the more likely that the cluster is split by epenthesis (Fleischhacker 2002). In a Pig-Latin task, English-speaking subjects had more difficulty splitting *sC(C)* onsets than stop-liquid onsets (Barlow 2001). And English speakers taught a VC infix were more likely to place the infix inside *pl* or *sl* than *st* (Pierrehumbert & Nair 1995). These results are consistent with the scale below, though not all support all distinctions (*S*=sibilant fricative):

most splittable S-glide → S-rhotic → S-l → S-nasal → S-stop *least splittable*
 stop-glide/liquid

Fleischhacker explains the scale perceptually: when C_2 is more sonorant, the C_1 - C_2 transition is more like a C-V transition, so C_1VC_2 is more similar to C_1C_2 . If epenthesis preferably produces a result that is perceptually similar to the original, it should, therefore, more likely split C_1C_2 when C_2 is more sonorant. In a C_1C_2V sequence, *V* is also more likely to overlap a sonorant C_2 and create an excrescent vowel in the C_1 - C_2 transition (Hall 2003). This may further contribute to the C_1 - C_2 transition’s similarity to a C-V transition. The similarity explanation extends to other types of splitting in which C_1 becomes prevocalic, such as split-onset Pig Latin (where $C_1C_2X \rightarrow C_2XC_1V$) and -VC- infixation (where $C_1C_2X \rightarrow C_1VC_2$).

This talk examines the splittability of CC clusters in Tagalog infixation using corpus and survey data. It argues that infix placement is conditioned by cluster splittability, not cluster markedness, and provides further support for the splittability scale above.

Loanwords in Tagalog can have initial clusters, which the infixes *um* and *in* may optionally split: *g-um-radwet* ~ *gr-um-adwet* ‘graduated’. In a written corpus drawn from the web, infixes split stop-glide clusters more often (387/494=78%) than stop-liquid clusters (593/1759=34%), adding a further sonority-based distinction to the splittability scale. This result is unexpected if infixes are deployed to break up marked onset clusters, because C-glide clusters appear *less* marked than C-liquid. First, C-glide onsets can be created in native words by optional syncope (*pijak* ~ *pjak* ‘squawk’), but C-liquid onsets cannot. Second, stress data indicate that word-internal CC clusters are normally syllabified $C_1.C_2$, avoiding a complex onset; in the rare exceptions, C_2 is a glide (*di.li.ɲjo* ‘delirium’).

Because Tagalog rarely allows SC onsets (except *s*-glide), prothesizing most (*um-iskor*, ‘scored’), the corpus provides little evidence about them. A survey was designed to probe both rare and common clusters. Participants choose one of two infixed forms to complete a sentence, and rate each option. The survey, administered over the web to participants around the world, is in progress, with 565 usable choice-rating-rating triples from 44 participants so far. For attested clusters, survey results reflect corpus statistics: splitting is chosen more often for stop-glide clusters than stop-liquid ($p < .0001$, Fisher’s Exact Test), and the difference between split and unsplit ratings is greater for stop-glide than stop-liquid ($p < .0001$, Scheffé’s *F*). Stop-glide and stop-liquid clusters both pattern as more splittable than *s*-stop.

Within SC clusters (excluding orthographic *sy*, which often represents [ʃ]), results are consistent with the scale above, though not every distinction is supported. The table in (1) shows, for each pair, whether (i) the rates at which participants chose splitting and (ii) the split-minus-unsplit ratings are significantly different ($p < .05$). All significant differences are in the direction predicted by the splittability scale. It is remarkable that Tagalog speakers make distinctions among onsets that are, except for *sw*, so rare in the language, suggesting that speakers can access non-language-specific information about clusters.

(1)

difference significant?	<i>sw</i>	<i>fr</i>	<i>sl</i>	<i>s-nasal</i>	<i>s-stop</i>
<i>sw</i>		n, n	y, y	y, y	y, y
<i>fr</i>			y, n	y, y	y, y
<i>sl</i>				n, n	y, y
<i>s-nasal</i>					y, y

References

- Barlow, Jessica (2001). Individual differences in production of initial consonant sequences in Pig Latin. *Lingua* 111, 667-696.
- Broselow, Ellen (1987). Non-obvious transfer: on predicting epenthesis errors. In Georgette Ioup and Steven Weinberger, eds., *Interlanguage Phonology: the acquisition of a second language sound system*. Cambridge: Newbury. 292-304.
- Fleischhacker, Heidi (2002) Cluster-dependent epenthesis asymmetries. *UCLA Working Papers in Linguistics* 7, *Papers in Phonology* 5, 71-116.
- Hall, Nancy (2003). Gestures and segments: vowel intrusion as overlap. UMass dissertation.
- Pierrehumbert, Janet and Rami Nair (1995). Word Games and Syllable Structure. *Language and Speech* 38, 77-114.
- Singh, Rajendra (1985). Prosodic adaptation in interphonology. *Lingua* 67, 269-282.