The Areal & Genealogical Dimensions of Hiatus
Kristine A. Hildebrandt, The University of Manchester
kristine.hildebrandt@manchester.ac.uk

I. What is Hiatus & Why Consider it Here?¹

- Two vowels become adjacent via morphological (or syntactic) concatenation:

(1) Persian (Indo-European) (Mahootian 1997)
mi-xa=æm > [mi.xam] DUR-want=1SG ‘I want’

- Across languages, hiatus environments subject to different ‘responses’:

(2) Possible Responses

"NOTHING"

a. Heterosyllabification
Digueño (Yuman)

"RESOLUTIONS"

b. Vowel elision/deletion (Note: in different lgs, different vowels may delete under different circumstances)
Cayuvava (Isolate)
ki-učarahi > [kučarahi] SUB-he.is.coming

c. Consonant insertion (Note: in different lgs, different consonants under different circumstances)
Kusunda (Isolate)
tui-i > [twi:jii] bee-GEN (Watters 2005)

d. Glide formation
Luganda (Benue-Congo)
ki-aa > [kjaa] 3P-GEN.LKR (Hyman & Katamba 2003)

e. Morpheme coalescence/merger
Dolakha Newar (Sino-Tibetan)
ye-u > (jeu)σ come-NR1 (Genetti 1994)

f. Allomorphy
Nepali (Indo-European)

- Autosegmental approaches (e.g. Sanskrit & Chicano Spanish; Schane 1987)
- Domain-oriented approaches (e.g. Isoko; Donwa-Ifode 1985)
- Optimality Theory & alignment constraints (e.g. vowel elision in Casali 1995; 1997)
  o Positionally Sensitive Faithfulness:
  o MAXLEX: maintain phonological material in elements encoding greater semantic content (root/lexical morphemes).
  o MAXWD: maintain phonological material in word-initial position

¹ The data & inspiration for this work come from the Word Domains component of the Autotyp Project (www.uni-leipzig.de/~autotyp). Thanks to Franziska Crell and Thomas Goldammer for helping me collect data from Indo-European & Austro-Asiatic languages, in particular, but any errors are my own responsibility. Ultimately, my representation of phenomena here is dependent on exhaustivity/comprehensiveness of grammatical descriptions & feedback from researchers. If you see a case here that is mis-represented or if something is omitted, feedback is welcome!
II. Goals & Methods

A. Explore 'hiatus' in cross-linguistic perspective
- Findings mixed significance
- Exploratory endeavour & matching of observations with other, related findings in the Word-Domains project
- Critical appraisal of past analyses in languages with 'mixed' responses & resolution types

B. Language Sample
- 45 languages (Appendix)
- "Balanced" representation of 3 families:
  - Indo-European, Sino-Tibetan, Austro-Asiatic, plus a mixture of 'other'
- "Balanced" representation of 3 areas:
  - Europe, Indic/S. Asia, Southeast Asia, plus 'other' areas
- Final sub-sample: Indosphere, Buffer Zone & Sinosphere sub-areas for Sino-Tibetan only
- All languages coded for:
  - Whether a hiatus environment is possible (i.e. is syllable canon only min. /CV/?)
  - Whether or not, if an environment is possible, there is a resolution or heterosyllabification
  - Resolution type
  - The morphological domain (i.e. stem-suffix; prefix-stem, affix-affix, etc.)
  - What is affected (e.g. if deletion in stem-suffix domain, what is deleted)
  - Multiple different resolutions in a single domain/edge or 'mixed' resolutions (i.e. a mixture of heterosyllabification and resolutions in same domain)

III. Selected Findings

A. Resolutions: Presence/Absence/Mixed
- Languages with minimal syllable template of /CV/ excluded
  
**Question 1:** Of languages that display hiatus environments, how often is it “resolved”?
- From OT Perspective, this is not really a question to be asked
- From “empirical” perspective (e.g. Blevins 1995; Maddieson 2005), [CV] a preferable structure

Fig. 1. Response across all languages (n = 40)

Answer 1: If we do not expect a preference for [CV], then resolution is a significantly preferred response to hiatus across all languages in the database (Chi-Square; p = .017)
Question 2: Is this preference similar in languages of the three sampled families?
Fig. 2 Response across three families

Answer 2: In Indo-European, resolution is preferred; in Sino-Tibetan all responses are attested; in Austro-Asiatic, languages only have mixed responses; These patterns significantly differentiate the 3 families. (Chi-Square; p = .040)

Question 3: Is this preference similar in languages of the three sampled areas?
Fig. 3. Response across three areas

Answer 3: In Europe & Indic all responses are attested; In SEA languages either always resolve hiatus or never resolve it (heterosyllabification). These patterns do not significantly differentiate the 3 areas.
**Question 4: Is this preference similar in the languages of the three (proposed) sub-areas for S-T?**

Fig. 4. Response across three spheres

![Resolution by Sphere (Sino-Tibetan only)](image)

**Answer 4:** In Indospheric & Buffer Zone languages all responses are attested; in Sinospheric languages heterosyllabification is the only response. These patterns do not significantly differentiate the 3 spheres.

- The ‘global’ trend is resolution (all things assumed equal) & this trend is supported in Indo-European languages
- In the three areas and three spheres responses are more mixed
- Interesting to note that Sinospheric languages permit ONLY heterosyllabification

**Question 5:** Given the cross-linguistic tendency for hiatus resolution, can we expect languages without resolution to have an overall lower “coherence” profile?

**Coherence (COH)** (Bickel & Hildebrandt 2005): A ratio of a language’s available morphology to number of these morphemes included in phonological words in that language.

(3) Belhare (Sino-Tibetan; data from Bickel & Hildebrandt 2005)

- Available Morphology: root, prefix, suffix, circumfix, enclitic
- Big P-word: 4 of 5 (Final Velar Drop)
- Small P-word: 2 of 5 (Nasal P.O.A. Assimilation)

a. ka-ak-lu-kak=phu>(kaaklugakphu) 1sP-OPT-tell-2A=REP ‘You may tell me, they say.’

b. mi-N-kai-chi-n > (miŋkai)ch‘in 3nsS-NEG-come.up-d-NEG’

Overall COH value for Belhare is .25 (about middle so far)

- A typological variable based on exhaustive surveys of phonological word types in a similar sample of languages (Austro-Asiatic languages excluded here due to lack of COH values)
- Smaller COH value (approaching 0) = smaller P-words; Bigger value = bigger P-words
Answer 5: A significant ($p = .045$) correlation between the PRESENCE of hiatus resolution and SMALLER COH values (& between the LACK of resolution and LARGER COH values). Suggests that we may not be able to rely on a particular process to indicate the overall coherence profile of a language. Thus, certain, highly salient processes may give a false impression of a language’s COH profile. This trend is repeated in the family, area and sphere samples, but is not significant.

B. “Conflicted” Hiatus
- Two types
  i. Different Resolutions, Same Domain
     - Languages resolve hiatus, but we see unique resolution types happening in the exact same domain/edge
     - Suggests that lexically specified details are necessary to describe the processes

(4) Persian
different resolution types with \(\text{æm}\) 1.SG pronominal enclitic, depending on domain type
a. ketab-\(\text{haæm}\) > ketab.[\(\text{haæm}\)] book-PL=1.SG (glide insertion)
b. \(\text{baæm}\) > [\(\text{ba.hæm}\)] with-1.SG (h-insertion)
c. mi-\(\text{xaæm}\) > [\(\text{mi.xaæm}\)] DUR-want=1SG (vowel deletion)

(5) Kinnauri (Sino-Tibetan; www.linfil.uu.se/personal/anjusaxena/kintext07.html)
a. byo-\(\text{o}\) > [\(\text{bi.jo}\)] or [\(\text{bo}\):] go-PROG (glide-insertion or coalescence)
b. ba-\(\text{o}\) > [\(\text{ba.dø}\)] come-PROG (consonant insertion)

**Question 6:** Given that there are several languages with mixed strategies in a single domain, would we see a similar pattern/degree of non-isomorphism (Diversity) of P-words in the same language?
- DIV (Bickel & Hildebrandt 2005): The number of non-isomorphic P-words in a language; a value approaching .1 indicates higher degree of non-isomorphism; a value approaching .00 indicates a lower degree

(6) Chukchi (Chukotko-Kamchutkan) (Partial Representation)  (DIV = .095)

<table>
<thead>
<tr>
<th>Vowel-Harmony</th>
<th>CF1</th>
<th>PF</th>
<th>Σ</th>
<th>CF2</th>
<th>SF</th>
<th>ENCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vowel Glottalization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasal Coda P.O.A.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Answer 6: Languages with mixed resolution types within a single domain have a generally LOWER DIV value, while languages with non-mixed types have a HIGHER DIV value; however, these findings are not significant.

- Although we cannot say with certainty, the suggestion is that mixed resolution strategies are not necessarily indicative of more non-isomorphism.

ii. ± Resolution, Same Domain

- A language may or may not resolve hiatus cases within the same domain type.
- Dolakha Newar: (C)(Cg)V(V)(C) (Genetti 1994); A dedicated section on “vowels in combination”

(7) Stem-Suffix in Dolakha Newar

a. Vowels of identical quality
   bi-i > [bi] give-INF (vowel elision)

b. Stem final high vowel & suffix high vowel
   bi-u > [bju] give-NR1 (glide formation)

c. Exception: nominalizer NR2 -e; the result of this hiatus is “a disyllabic word of two distinct pulses” (p. 30)
   ye-e > [je.e] come-NR2 (heterosyllabification)²

d. Exception: -eu 3sFUT or -en PART following stem-final /e/, also two distinct pulses, but vowel quality change
   ye-eu > [ye.ɛu] come-3s.FUT (heterosyllabification)

- Cases like these not dealt with in OT or Autosegmental or domains literature
- We must turn to both phonology & item-based information to account for ± resolution & type

² I don’t have data, but I assume that other -e suffixes like the locative and the linker are subject to deletion when following stems with identical vowel quality
IV. Concluding Remarks

- Resolution of hiatus is cross-linguistically preferred, seems to characterize stocks
- A focus on resolution types within theoretical framework overlooks language-specific particularities & complexities
- And at this point it is not likely that the facts about a particular process in a language are necessarily indicative of the overall prosodic (word) profile of that language
- Additional support for individual historical sources of observed phenomena

References


Hyman, L. M. & Francis X. Katamba. The Word in Luganda, Ms.


<table>
<thead>
<tr>
<th>LANGUAGE</th>
<th>HIATUS POSSIBLE?</th>
<th>HIATUS RESOLVED?</th>
<th>GEN. AFFIL.</th>
<th>AREA</th>
<th>SPHERE (if S-T)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belhare</td>
<td>Sino-Tibetan</td>
<td>Indic</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Boumaa Fijian</td>
<td>Austronesian</td>
<td></td>
<td>Oceania</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cantonese</td>
<td>Sino-Tibetan</td>
<td>SE Asia</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Cayuvava</td>
<td>Isolate</td>
<td></td>
<td>NE S America</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chukchi</td>
<td>Chukotko-Kamchutkan</td>
<td></td>
<td>N Coast Asia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diegueño</td>
<td>Yuman</td>
<td></td>
<td>California</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diyari</td>
<td>Pama-Nyungan</td>
<td>S. Australia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dolakha Newar</td>
<td>Sino-Tibetan</td>
<td>Indic</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Egyptian Arabic</td>
<td>Semitic</td>
<td></td>
<td>N. Africa</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finnish</td>
<td>Uralic</td>
<td></td>
<td>Europe</td>
<td></td>
<td></td>
</tr>
<tr>
<td>German</td>
<td>Indo-European</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greek (Modern)</td>
<td>Indo-European</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hixkaryana</td>
<td>Cariban</td>
<td></td>
<td>NE S. America</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Irish</td>
<td>Indo-European</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jahi</td>
<td>Austro-Asiatic</td>
<td>SE Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kayah Li</td>
<td>Sino-Tibetan</td>
<td>SE Asia</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Kewa</td>
<td>Engand</td>
<td>S. New Guinea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kinnauri</td>
<td>Sino-Tibetan</td>
<td>Indosphere</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kalkha Mong.</td>
<td>Mongolian</td>
<td>Inner Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kham</td>
<td>Sino-Tibetan</td>
<td>Indic</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Kusunda</td>
<td>Isolate</td>
<td>Indic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kyirong Tibetan</td>
<td>Sino-Tibetan</td>
<td>Indic</td>
<td>Buffer Zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lahu</td>
<td>Sino-Tibetan</td>
<td>SE Asia</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Limbu</td>
<td>Sino-Tibetan</td>
<td>Indic</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Lithuanian</td>
<td>Indo-European</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luganda</td>
<td>Benue-Congo</td>
<td>S. Africa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manange</td>
<td>Sino-Tibetan</td>
<td>Indic</td>
<td>Buffer Zone</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mandarin</td>
<td>Sino-Tibetan</td>
<td>SE Asia</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Mapudungun</td>
<td>Isolate</td>
<td>Andean</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Martuthunira</td>
<td>Pama-Nyungan</td>
<td>S. Australia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Meithi</td>
<td>Sino-Tibetan</td>
<td>Buffer Zone</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mundari</td>
<td>Austro-Asiatic</td>
<td>Indic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nama</td>
<td>Kwadi-Khoe</td>
<td>S. Africa</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nepali</td>
<td>Indo-European</td>
<td>Indic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Persian</td>
<td>Indo-European</td>
<td>Cauc.-Mesop.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Polish</td>
<td>Indo-European</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sko</td>
<td>Macro-Sko</td>
<td>W. N. Guinea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Santali</td>
<td>Austro-Asiatic</td>
<td>Indic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spanish</td>
<td>Indo-European</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td>Indo-European</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Turkish</td>
<td>Turkic</td>
<td>Europe</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vietnamese</td>
<td>Austro-Asiatic</td>
<td>SE Asia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu (Chinese)</td>
<td>Sino-Tibetan</td>
<td>SE Asia</td>
<td>Indic</td>
<td>Indosphere</td>
<td></td>
</tr>
<tr>
<td>Yidiny</td>
<td>Pama-Nyungan</td>
<td>N. Australia</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yimas</td>
<td>L. Sepik</td>
<td>Inn N. Guinea</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>