

# The A'ingae Co-Occurrence Constraint

Chiara Repetti-Ludlow

*New York University*

## 1 Introduction

A'ingae, also known as Cofán, is a language isolate spoken by approximately 1,500 people in parts of Ecuador and Colombia. Similar to co-occurrence restrictions in other languages, A'ingae undergoes a co-occurrence process in which two stops or affricates in a root cannot differ only in aspiration; they must also differ in some combination of manner, place, and voice. So, for example, words like [to.to] 'white' and [ts<sup>h</sup>a.ts<sup>h</sup>a] 'to draw' are attested, but words like \*[to.t<sup>h</sup>o] are not. Furthermore, any two consonants in a root that match in manner, place, and voice also tend to be followed by identical vowels, so we find words like [to.to] but not \*[to.ti].

Co-occurrence phenomena in which non-adjacent consonants must match in certain features are typologically well-attested and include such variants as long-distance nasal agreement, liquid agreement, coronal agreement, dorsal agreement, and laryngeal agreement, as we see in A'ingae (see Rose & Walker, 2004 for a more in-depth discussion of the typology of long-distance consonant agreement). In many of the languages that have this type of consonant agreement, once consonants share certain features, other features assimilate as well. This means that constraints exist on similar but not identical consonants. For example, Gallagher & Coon (2009) examine Chol, a Mayan language spoken in Mexico by approximately 150,000 people, which has a co-occurrence phenomenon whereby two non-ejective stridents in a root must be completely identical (1a-b). Put another way, if two stridents in a root are ejective, they must also share all other features. Meanwhile, if a root has one ejective strident and one non-ejective strident, they must only agree in anteriority (2c-d).

- (1) a. [sus] 'scrape'  
b. \*[tsus]  
c. [ts'is] 'sew'  
d. \*[ts'if]

Gallagher & Coon (2009) note that in the vast majority of long-distance consonant agreement cases, the consonants are not only more similar but, in fact, identical (1a-b). In the cases where consonant agreement does not result in identical consonants, the agreement can largely be explained by local articulatory spreading (1c-d). Gallagher & Coon refer to instances where long-distance agreement results in identical consonants as "total identity" and instances where spreading results in more similar but non-identical consonants as "partial identity."

Zuraw (2002) points out another cross-linguistic pattern whereby a consonant and vowel within the same syllable can agree with another consonant-vowel pair in the word. This results in a structure that looks like reduplication, but unlike standard reduplication, has no morphological function. Zuraw calls words that have this pattern "pseudoreduplicated," and proposes that there is a phonological motivation to impose a reduplication-like structure on words. This is formalized with the constraint REDUP, which requires that words have coupled substrings, meaning that some strings of phonemes (usually syllables) are associated and corresponding segments (e.g. onsets, nuclei, or codas) must be identical. This is the proposal I adopt in discussing the A'ingae data, as it is the only proposal that accounts for both consonant and corresponding vowel agreement.

---

\* I would like to thank Chelsea Sanker, Scott AnderBois, members of NYU's PEP Lab, and the audience members and anonymous reviewers at AMP 2020 for invaluable comments and guidance. I would also like to express my gratitude to our wonderful consultant, Hugo Lucitante. All errors are my own.

My goals in this paper are threefold. First, I will present new data on the A'ingae co-occurrence constraint (§2). Second, I will determine how best to formalize an analysis of this phenomenon (§3.) Third, I will explore potential explanations for the existence of this co-occurrence constraint in (§4). And finally, I will conclude in §5.

## 2 Data

The vowels of A'ingae, shown in Figure 1, are [a, e, i, o, i], all of which have both nasal and oral counterparts. Furthermore, A'ingae has the diphthongs [ai], [oe], [oa], [oi], [ii], and [ao]. The consonant inventory, shown in Figure 2, consists of nasals, approximants, a tap, voiceless fricatives, and a three way stop and affricate series consisting of voiceless unaspirated, voiceless aspirated, and prenasalized voiced counterparts (Repetti-Ludlow et al. 2019).

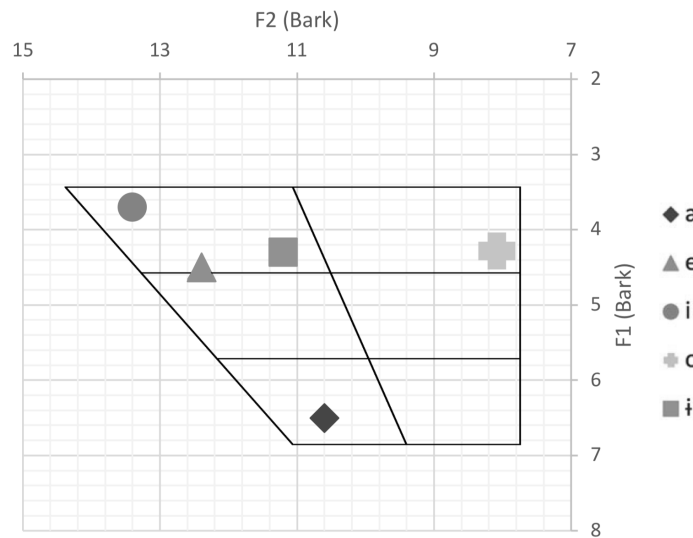


Figure 1: A'ingae vowels (Repetti-Ludlow et al. 2019)

	Bilabial	Labio-dental	Alveolar	Post-alveolar	Palatal	Velar	Glottal
Plosive	p p <sup>h</sup> m <sup>b</sup>		t t <sup>h</sup> n <sup>d</sup>			k k <sup>h</sup> ŋ <sup>g</sup>	ʔ
Affricate			ts ts <sup>h</sup> n <sup>d</sup> dz	tʃ tʃ <sup>h</sup> n <sup>d</sup> dʒ			
Fricative		f	s	ʃ			h
Nasal	m		n		ɲ		
Approximant		v			j	ɥ	
Tap			r				

Figure 2: A'ingae consonants (Repetti-Ludlow et al. 2019)

**2.1 Consonants** As previously mentioned, in A'ingae, when a root has two consonants that match in place, manner, and voice, the consonants will always match in aspiration value (2a-d). This does not apply to segments that differ in manner of articulation (2e), place of articulation (2f), or non-aspiration laryngeal features such as voicing (2g), nor does it apply to segments that are in different morphemes (2h). However, if segments are in the same morpheme, they do not have to be in adjacent syllables to match (2i). These conditions are never violated.

(2)

a. *[to.t <sup>h</sup> o]	f. [a.ta.k <sup>h</sup> u]	‘undergrowth’
b. *[ts <sup>h</sup> a.ts̄a]	g. [ta. <sup>n</sup> dan]	‘to tie’
c. [to.to]	h. [k <sup>h</sup> a-ki]	‘following-day’
d. [ts <sup>h</sup> a.ts̄ <sup>h</sup> a]	i. [ku.ja.ku.pa]	‘coloradilla’
e. [tsi.t <sup>h</sup> a]		‘bone’

It is worth noting that prenasalized stops generally pattern with voiceless unaspirated stops. This means that we find prenasalized voiced stops and unaspirated voiceless stops of the same place of articulation in a root (as with 2g), but we don’t have evidence for prenasalized voiced stops and aspirated voiceless stops of the same place of articulation in a root. This patterning is consistent with some A’ingae phonological reconstructions, which have shown that the prenasalized series of stops and affricates developed from the unaspirated series of stops and affricates historically (Sanker & AnderBois, [forthcoming]). It is therefore possible that the co-occurrence restriction that stops and affricates of the same manner and place of articulation match in aspiration value developed before the voicing contrast emerge. That being said, prenasalized voiced segments are quite rare in the language, so we do not have enough data yet to draw definitive conclusions on this matter.

**2.2 Vowels** Notably, of 135 roots compiled from the A’ingae dictionary (based on Borman 1976) that have matching stops or affricates, 110 show corresponding vowel agreement, as can be seen in the examples in (3).

(3) a. [t̄ja.t̄ja]	‘to loosen dirt’
b. [ts <sup>h</sup> o.ts̄ <sup>h</sup> o <sup>?</sup> .pa]	‘lungs’
c. [te.te.te]	‘wild’
d. [p <sup>h</sup> ɪ̄.p <sup>h</sup> ɪ̄]	‘maize’
e. [ki.ki]	‘cat’

Of the 25 roots that do not have corresponding vowel agreement, the second syllable is always either [a] or [o], as shown in the examples in (4).

(4) a. [te <sup>?</sup> .ta]	‘flower’
b. [t <sup>h</sup> ɪ̄.t <sup>h</sup> a. <sup>m</sup> ba]	‘day’
c. [o.pi.pa <sup>?</sup> .t̄fo]	‘shoulder’
d. [t <sup>h</sup> e.t <sup>h</sup> o <sup>?</sup> .t̄fo]	‘beak’
e. [ti.to]	‘uncle’
f. [t̄ <sup>h</sup> ɪ̄ <sup>?</sup> .ni.t̄ <sup>h</sup> o.k <sup>h</sup> o]	‘small’

This could suggest a rightward direction of feature agreement to which [a] and [o] are resistant, which is consistent with Pulleyblank’s (1998) and Walker’s (2005) findings that back vowels are more sonorous than other vowels and are therefore less likely to assimilate. Given that [a] and [o] are the only [+back] vowels in A’ingae, it makes sense that these two segments are the only two that do not take on the features of other vowels.

In order to document these restrictions, the Observed/Expected (O/E) ratio was calculated. The O/E ratio is calculated by counting the observed instances of a given phenomenon (e.g. the number of times /a/ appears following each of two matching stops) and dividing that by the expected instances of the phenomenon, which is calculated by multiplying the observed instances of each individual aspect (e.g. the number of times /a/ is found in the first syllable and the number of times /a/ is found in the second syllable) divided by the total number of roots. For example, in this database, /a/ is found as both the first and the second vowel following matching stops and affricates 23 times. There are 29 instances of /a/ being the first vowel and 35 instances

of /a/ being the second vowel (the difference is due to the fact that [a] does not change when it is in the rightmost coupled syllable). There are 135 roots that were analyzed in the database. Thus, the calculation is  $23/((29*35)/135) = 3.06$ . An O/E value above 1.5 indicates that a given combination is overrepresented (Pierrehumbert 1993; Frisch et al. 2004). Every instance of a word with matching stops or affricates in the

A'ingae dictionary (Borman 1976) was compiled and analyzed, as was a random sampling of words with mismatching consonants within a root. The O/E ratios provide support to the observations stated above, as when a root has matching stops or affricates, having matching vowels is overrepresented, as shown in Figure 3. Meanwhile, when a root has mismatching consonants, there is no combination that is overrepresented or restricted, as shown in Figure 4. However, it is not the case that if two vowels match, there is a tendency for consonants to match. It seems instead that this is a consonant-driven process.

		V2				
		a	e	i	o	u
V1	a	3.06	0.00	0.00	0.54	0.00
	e	0.81	7.11	0.00	0.55	0.00
	i	0.28	0.00	8.04	0.00	0.00
	o	0.19	0.00	0.00	1.96	0.00
	u	0.60	0.00	0.00	0.16	4.05

Figure 3: O/E ratio for vowels following matching stops/affricates (N=135)

		V2				
		a	e	i	o	u
V1	a	0.96	1.02	1.04	0.96	1.07
	e	1.21	0.63	0.77	0.87	1.27
	i	0.94	1.11	1.15	0.92	0.90
	o	1.03	0.96	0.96	1.29	0.73
	u	1.00	1.08	0.64	0.75	1.23

Figure 4: O/E ratio for vowels following mismatching consonants (N=537)

This agreement is not limited to stops and affricates. Although stops and affricates are the only segments that show a difference in aspiration ([sg] value), any two segments with matching manner, place and voice will also have vowel agreement. Put another way, any two identical consonants will also follow the vowel restrictions defined above. Of 118 words in the A'ingae root database that have two identical non-stop/affricate consonants, all show the same vowel pattern.

### 3 Analysis

In order to formalize an analysis of this phenomenon, I adopt Zuraw's (2002) framework of aggressive reduplication, as it accounts for the association of full syllables, rather than individual segments. This is central to our analysis if we are to account for the vowel agreement as well as aspiration agreement for stops and affricates. Aggressive reduplication relies on two types of constraints; REDUP, and IDENT-KK[F]. REDUP requires that a word contain some substrings that are coupled and IDENT-KK[F] constraints require that corresponding segments in coupled strings match in a given feature. First, let's discuss how to formalize the consonant aspiration agreement. To do this, we first need to posit a constraint IDENT-KK[spread glottis], which states that for a given pair of coupled syllables, the corresponding consonants must agree in aspiration. The constraints outlined in (5) and their ranking can be seen in the tableau in Figure 5.

- (5) **IDENT-*kk*[M/P/V]**: A group of constraints positing that if a word contains coupled substrings (denoted by [...]<sub>α</sub>), coupled segments must have the same values for manner, place, and voicing features
- IDENT-*io*[M/P/V]**: A segment of given values for manner, place, and voice in the input must have the same values in the output
- REDUP**: A word must contain substrings that are coupled

**IDENT-*kk*[sg]**: If a word contains two substrings that are coupled, segments must have the same value for [sg]

kik <sup>h</sup> i	ID-KK[M/P/V]	ID-IO[M/P/V]	REDUP	ID-KK[SG]
☞ a. [ki] <sub>α</sub> [ki] <sub>α</sub>				
b. [ki] <sub>α</sub> [t <sup>h</sup> i] <sub>α</sub>	*!	*		*
c. kit <sup>h</sup> i		*!	*	
d. kik <sup>h</sup> i			*!	
e. [ki] <sub>α</sub> [k <sup>h</sup> i] <sub>α</sub>				*!

Figure 5: Tableau showing how aggressive reduplication can account for consonant agreement

The constraint ID-IO[M/P/V] isn't solely important in helping avoid dissimilation (as seen in examples b and c in Figure 5), it is also key in preventing REDUP from causing all words to develop reduplication-like structure. Take, for instance, the word [t<sup>h</sup>e.si] 'jaguar.' Were it not for the constraint ID-IO[M/P/V], we would expect reduplication-like structure in every morpheme. The importance of the constraint for maintaining the integrity of non-matching features can be seen in Figure 6.

t <sup>h</sup> esi	ID-KK[M/P/V]	ID-IO[M/P/V]	REDUP	ID-KK[SG]
☞ a. t <sup>h</sup> esi			*	
b. [t <sup>h</sup> e] <sub>α</sub> [t <sup>h</sup> i] <sub>α</sub>		*!		

Figure 6: Tableau showing how the system allows for words without reduplication structure

In order to account for vowel agreement, we need to rank the constraints IDENT-IO[+back] and IDENT-IO[+round] quite high to avoid [a] and [o] becoming identical to the previous vowel. Furthermore, we need to account for the fact that agreement seems to be a rightward process, which I do with a new constraint: IDENT-IO[K<sub>L</sub>]. This constraint requires that the leftmost linked syllable remain identical in the input and output. This is consistent with research on clitics, with researchers like Peperkamp (1997) finding that there is asymmetry in coherence between word beginnings, proclitics, and prefixes as opposed to word endings, enclitics, and suffixes. Essentially, proclitics and prefixes are less likely to cohere than enclitics and suffixes, which Peperkamp attributes to constraints on language processing. Finally, we must add IDENT-KK[F] constraints dealing with vowel identity in order to account for total identity of vowels. The constraints are outlined in (6) and their functions can be seen in Figure 7.

- (6) **IDENT-IO[+BACK]**: A [+back] segment in the input must be [+back] in the output
- IDENT-IO[+ROUND]**: A [+round] segment in the input must be [+round] in the output
- IDENT-IO[k<sub>L</sub>]**: Let k<sub>L</sub> and k<sub>R</sub> be corresponding segments such that k<sub>R</sub> follows k<sub>L</sub>. k<sub>L</sub> must be the same in input and output
- IDENT-*kk*[V]**: A group of constraints that posit if a word contains coupled substrings, coupled segments must have the same value for [back], [front], [high], and [low]

teʔt <sup>h</sup> a	ID-IO[+BCK]	ID-IO[RND]	ID-KK[M/P/V]	ID-IO[k <sub>L</sub> ]	REDUP	ID-KK[SG]	ID-KK[V]
☞ a. [teʔ] <sub>α</sub> [ta] <sub>α</sub>							*
b. [teʔ] <sub>α</sub> [te] <sub>α</sub>	!*						
c. [teʔ] <sub>α</sub> [sa] <sub>α</sub>			*!				*
d. [taʔ] <sub>α</sub> [ta] <sub>α</sub>				*!			
e. teʔt <sup>h</sup> a					*!		
f. [teʔ] <sub>α</sub> [t <sup>h</sup> a] <sub>α</sub>						*!	*

Figure 7: Tableau showing how aggressive reduplication can account for the A'ingae vowel patterns

Although the constraint ID-IO[ROUND] does not serve a purpose in the example above, it is central in the analysis, as it helps us account for words like [k<sup>h</sup>a.k<sup>h</sup>o.ni] ‘new moon,’ in which both vowels are [+back] and would otherwise assimilate to [k<sup>h</sup>a.k<sup>h</sup>a.ni], given that the ID-IO[+BACK] constraint would not be violated. The function of this constraint is illustrated by the tableau in Figure 8.

k <sup>h</sup> ak <sup>h</sup> oni	ID-IO[+BCK]	ID-IO[RND]	ID-KK[M/P/V]	ID-IO[k <sub>L</sub> ]	REDUP	ID-KK[SG]	ID-KK[V]
☞ a. [k <sup>h</sup> a] <sub>α</sub> [k <sup>h</sup> o] <sub>α</sub> ni							*
b. [k <sup>h</sup> a] <sub>α</sub> [k <sup>h</sup> a] <sub>α</sub> ni		*!					

Figure 8: Tableau showing the necessity of ID-IO[ROUND]

Thus, we are able to fully account for the A'ingae data using aggressive reduplication and the addition of one constraint, IDENT-IO[k<sub>L</sub>], which is necessary to account for the leftmost linked syllable's resistance to assimilation and the potential rightward directionality of agreement. That being said, we do not have evidence that the inputs I posit in the tableaux are actually the underlying forms of the attested words, yet these invented inputs are nevertheless useful in demonstrating why we do not see certain patterns.

Hayes & Wilson's (2008) Maximum Entropy model (MaxEnt) would potentially be a better method for studying this phenomenon, as it emphasizes gradient phonotactics, but a classic OT framework was adopted here in accordance with the majority of the co-occurrence literature and due to the relatively small data size, which is incompatible with MaxEnt,

#### 4 Pseudoreduplication

The fact that identical consonants tend to be followed by identical vowels results in a word structure that looks very much like reduplication. While there is evidence for a historical reduplication process in A'ingae, as with the words [o.t<sup>h</sup>a.je] ‘hit’ and [o.t<sup>h</sup>a.t<sup>h</sup>a.je] ‘hit repeatedly’, for many of the words involved in this co-occurrence process, there is no evidence of reduplication either current or historical. Rather, these words are in line with what Zuraw (2002) calls “pseudoreduplicated words,” which are words that look like they have undergone reduplication but haven't.

I hypothesize that the REDUP constraint attains a higher ranking in a given language when the language has an active morphological reduplication process or historically had a morphological reduplication process. When this is the case, it is plausible that speakers develop an affinity for a certain word shape whereby syllables are identical or nearly identical, in line with the language's morphological reduplication. This hypothesis is in line with the A'ingae data, as a productive morphological reduplication process historically may have led speakers to prefer a word shape in which similar-sounding syllables were made to be identical. Furthermore, Tagalog, the language that Zuraw (2002) uses to exemplify pseudoreduplication, also has an active morphological reduplication process, which suggests that the same hypothesis could hold there as well. If this is true of languages beyond A'ingae and Tagalog, it would provide solid typological evidence that the REDUP constraint is ranked higher in languages that have either active or historical morphological reduplication.

## 5 Conclusion

To conclude, aggressive reduplication is the best classic OT framework for accounting for the A'ingae data, but the new constraint IDENT-IO[K<sub>L</sub>] is necessary to account for the potential rightward direction of agreement. Furthermore, I propose that this co-occurrence process may exist due to A'ingae's historical morphological reduplication process, which led to an expectation for word shapes in which roots that have matching consonants also have matching vowels. Then, after the morphological reduplication process became fossilized, speakers still maintained an affinity for certain word shapes, leaving us with the lexical pattern we see today. In future work, the formalization of this pattern should make use of MaxEnt, which is well-suited to the A'ingae data.

## References

- Borman, Marlytte B. 1976. *Vocabulario Cofán* [Cofán vocabulary]. Summer Institute of Linguistics [Serie de Vocabularios Indígenas 19].
- Frisch, Stephen, Janet Pierrehumbert, and Michael Broe. 2004. Similarity avoidance and the OC Language and Linguistic Theory 22
- Gallagher, Gillian, and Jessica Coon. 2009. Total and Partial Identity: Evidence from Chol. *Natural Language & Linguistic Theory* 27, 3: 545-582.
- Hayes, Bruce, and Colin Wilson. 2008. A Maximum Entropy Model of Phonotactics and Phonotactic Learning. *Linguistic Inquiry* 39, 3: 379-440.
- Peperkamp, Sharon. 1997. *Prosodic Words*. The Hague: Holland Academic Graphics.
- Pierrehumbert, Janet. 1993. Dissimilarity in Arabic verbal roots. In *NELS 23: Proceedings of the North East Linguistic Society*, ed. Amy Schafer, 367-381. Amherst.
- Pulleyblank, Douglas. 1998. *Yoruba vowel patterns: deriving asymmetries by the tension between opposing constraints*, ms., University of British Columbia.
- Repetti-Ludlow, Chiara, Haoru Zhang, Hugo Lucitante, Scott AnderBois, and Chelsea Sanker. 2019. A'ingae (Cofán). *Journal of the International Phonetic Association*. 1-14.
- Rose, Sharon, and Rachel Walker. 2004. A typology of consonant agreement as correspondence. *Language* 80: 475-532.
- Sanker, Chelsea, and Scott AnderBois. [forthcoming]. *Reconstruction of A'ingae Nasality*.
- Walker, Rachel. 2005. Weak Triggers in Vowel Harmony. *Natural Language & Linguistic Theory*, 23, 4: 917-989.
- Zuraw, Kie. 2002. Aggressive Reduplication. *Phonology* 19 3: 395-439.