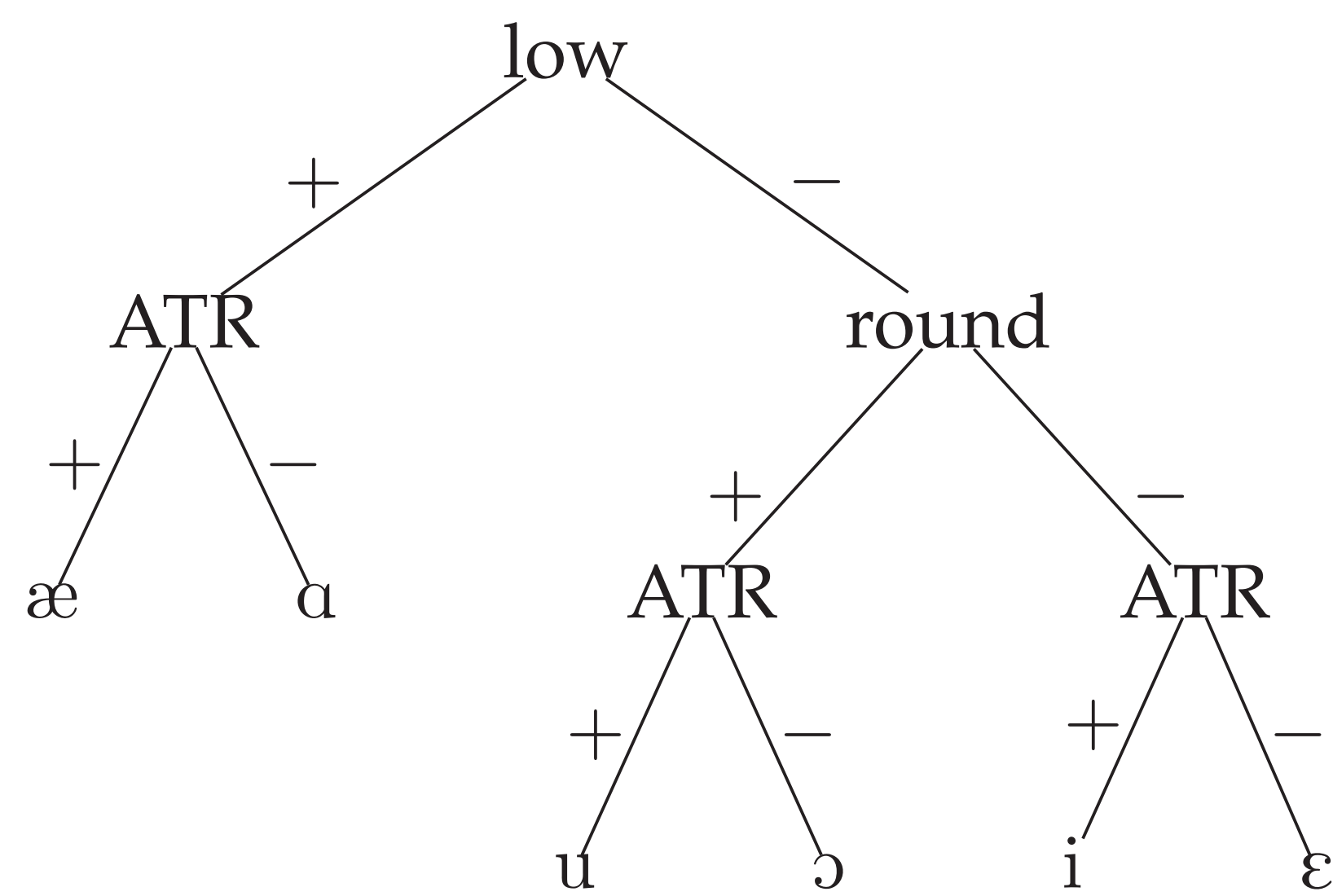


CONTRASTIVE HIERARCHY

- The contrastive features of a language's segment inventory can be specified by recursively dividing the sounds based on a prescribed feature order (Cherry et al. 1953; Dresher 2009).
- For example, Mackenzie and Dresher (2003) propose the feature order low > round > ATR for the vowel inventory of Nez Perce (Sahaptian; Idaho, Washington, Oregon).



- Feature orders are language-particular and based on what features are phonologically active.
- Possible feature orders are not believed to be fully unconstrained. Dresher (2009: 168): 'The limits of this variation remain to be determined... That there are limits is suggested by the fact that certain feature orders produce unnatural-looking inventories.'

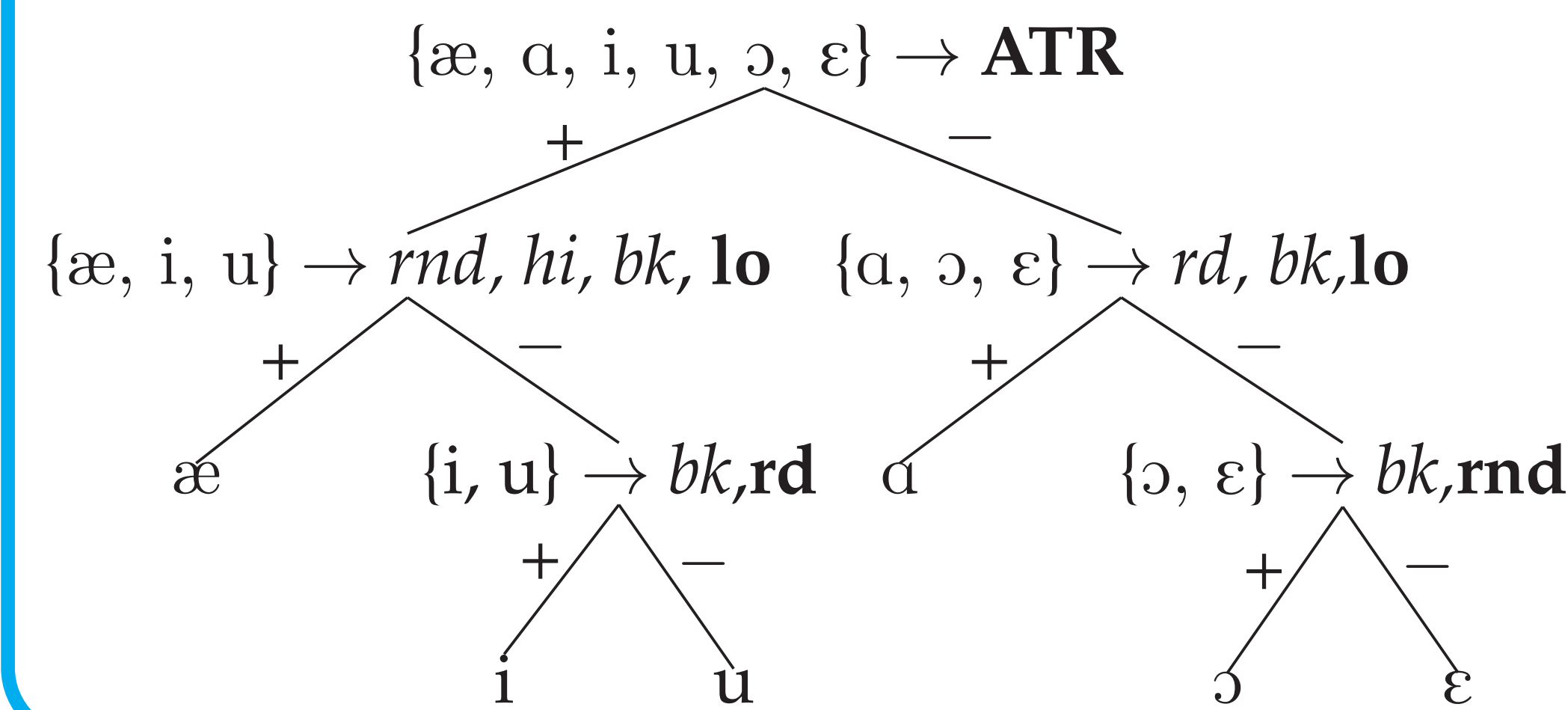
ID3 (QUINLAN 1986)

- Iterative Dichotomiser: decision tree learning algorithm that recursively selects the feature with the largest *information gain*, defined as the greatest reduction in *entropy*.
- Informally, entropy can be thought of as the degree of uncertainty in identifying the correct phoneme based on the information (i.e., feature specifications) available so far.
- Formally, it is $\sum_{i=1}^N -p_i \log_2 p_i$, where p_i is the proportion of decisions/phonemes in a given category; N = number of categories.

PROPOSED MODIFICATION

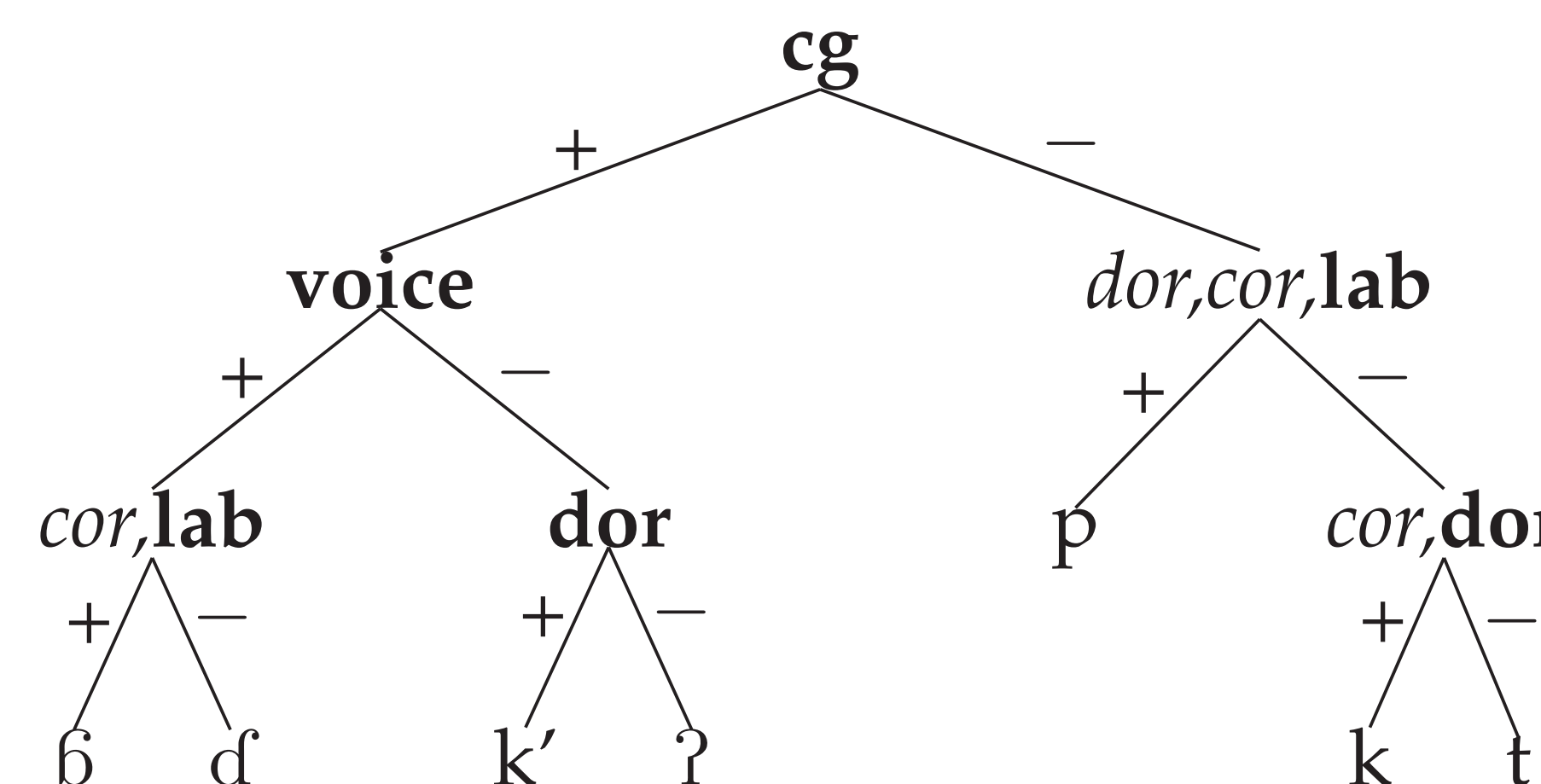
- A choice is made among features that tie for information gain.
- The result is a default partial order, with language-particular variation only possible when information gain is indecisive.

	-back		+back	
	+ATR	-ATR	+ATR	-ATR
+hi, -lo	i		u	
-hi, -lo		ɛ		ɔ
-hi, +lo			æ	ɑ



TZUTUJIL (MAYAN; GUATEMALA)

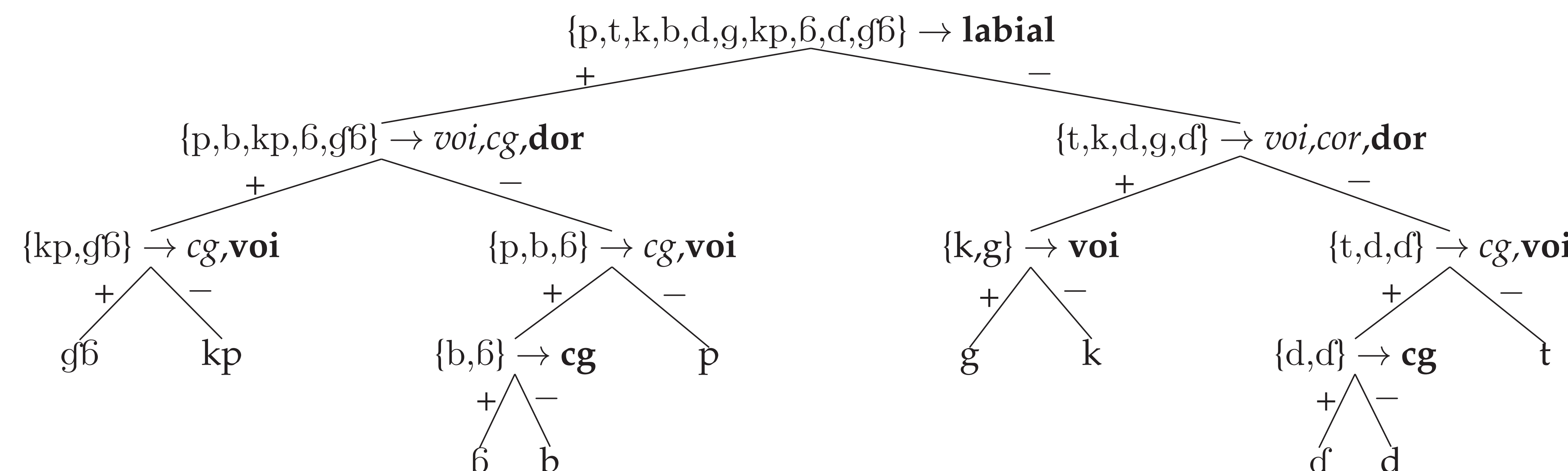
- Co-occurrence restriction on 1) non-identical ejectives and 2) homorganic ejective-plain stop pairs (MacEachern 1999). Implosives do not participate.
- Mackenzie (2009) proposes the order labial > dorsal > coronal > voice > c.g., such that implosives are not specified for c.g.
- Given the same subset of the inventory, ID3 gets this exactly wrong! But things get more interesting when it's given the larger inventory (see manuscript for details).



CONSONANT HARMONY IN BUMO IZON (NIGER-CONGO; NIGERIA)

- Co-occurrence restriction on {ɓ, d} and {d, b} (Efere 2001; Hansson 2001; Harry 2004).
- Mackenzie (2009) proposes a constraint $*[\alpha cg][-\alpha cg]_{Root}$ and the feature order labial > dorsal > voice > c.g.
- Non-participating segments (velars and labiovelars) are not specified for c.g.

	labial	dorsal	coronal
+c.g.	ɓ		d
		gɓ	
-c.g.	p b	k g	t d
		kp	



RESULTS

Language	=	≡	≈
Nez Perce		✓	
Classical Manchu	✓		
Bumo Izon	✓		
Kalabari Ijo	✓		
Dholuo	✓		
Chaha			✓
Anywa	✓		
Hausa			✓
Tzutujil			✓
Aymara	✓		

(Feature orders proposed by Zhang 1996; Dresher 2009; Mackenzie 2009)

= exact tree
 ≡ different tree but same feature specifications
 ≈ same feature specifications for participating features

IMPLICATIONS

- Hypothesis for the limits on variation in feature orders: **a less informative feature will not need to be ordered above a more informative one.**
- If this hypothesis is correct, ID3 serves as a valuable analytical tool for verifying a proposed feature order's well-formedness.
- It also provides a principled means of situating features into the hierarchy when language-internal cues don't dictate an exact position.

SELECTED REFERENCES

- Cherry, E. Colin, Morris Halle, and Roman Jakobson. 1953. Toward the logical description of languages in their phonemic aspect. *Language* 29(1), 34-46.
- Dresher, Elan. 2009. The contrastive hierarchy in phonology. Cambridge: Cambridge UP.
- Mackenzie, Sara. 2009. Contrast and similarity in consonant harmony processes. PhD thesis, University of Toronto.
- Quinlan, J. Ross. 1986. Induction of decision trees. *Machine Learning* 1(1), 81-106.