To encode information about the hierarchical structure of syllables, our word models consist of:

- a set of positions $D = \{1, 2, 3, \ldots\}$
- an alphabet of symbols denoting relations, $\Sigma = \{\text{ons, nuc, cod, a, b, c, \ldots}\}$
- unary labeling relations, e.g., $\text{ons}(x)$ means "position $x$ is labeled 'ons'."
- binary relations of precedence (e.g., $x < y$), dominance (\(\delta\)), and sonority ($\sigma$).

Berber-Specific Principles

Exemplary data with syllabic consonants in boldface uppercase letters, as in [1, 2]:

- $\text{radkt} \quad [\text{ra.d.k.t} \quad \text{she will remember}]
- \text{biddl} \quad [\text{bD.d.L} \quad \text{exchange}]
- \text{tzntl} \quad [\text{tX.nN} \quad \text{you stored}]

- Any vowel or consonant can be a syllabic nucleus.
- Non-initial syllables must have an onset.
- We still write constraints in CNF, but the EQCs we refer to are word-model-theoretic sub-structures, like $\psi_{\text{nuc}}$ in Figure 2.

The Sonority Sequencing Principle (SSP)

A particular version of the SSP is obeyed in Berber. Specifically:

- Sonority must not fall rightward of an onset consonant.
- Sonority must not fall leftward of a coda consonant.

Universal Structural Well-formedness Principles

All syllables must obey certain universal constraints.

- We use $\Phi_{\text{set}}$ to denote the set of negative literals that correspond to universally banned substructures (e.g., a syllable with more than one nucleus).
- We use $\Phi_{\text{req}}$ to denote the set of positive literals that correspond to universally required substructures (e.g., every syllable must have a nucleus).
- Figure 5 illustrates the sub-structure $\psi_{\text{nuc}}$, which corresponds to a syllable with multiple nuclei.
- $\psi_{\text{nuc}}$ is universally banned, so it is a member of the set $\Phi_{\text{set}}$.

Characterizing Well-formedness in Berber

The 1-CNF expression for Berber is:

$\Phi_{\text{Berber}} = (\neg \psi_{\text{ons}}) \land (\neg \psi_{\text{nuc}}) \land \psi_{\text{req}} \land \neg \psi_{\text{set}}$

- A word model is ill-formed if it contains any banned sub-structure or if it lacks any required sub-structure.
- Banned sub-structures are highlighted in the two ill-formed models below.

Why It Matters

- We add syllable well-formedness to a long list of phonological phenomena shown to be computationally simple [3, 4, 5, 6].
- This provides further evidence that phonology is restricted to a small subclass of patterns and functions, explaining many typological gaps.
- Computational simplicity also allows for easier learning.
- Valiant’s [7] Probably Approximately Correct (PAC) learner for 4-CNF formulas requires less time and data than the 1-CNF learner.
- 1-CNF formulas are efficiently learnable from positive data [5].

References