Root Clusters in St'át'imcets and the Organization of the Mental Lexicon*

Henry Davis
University of British Columbia

Bailey Trotter
University of British Columbia

Abstract: This paper constitutes the first investigation of *root clusters* in a Salish language, where a root cluster is defined as a set of derivationally distinct roots which are nevertheless related by systematic associations between sound and meaning. We explore both the phonology and semantics of root clusters in St'át'imcets (Lillooet Salish), before going on to investigate the diachronic relations between root clusters in St'át'imcets and their counterparts in Proto-Salish.

Keywords: St'át'imcets, Proto-Salish, roots, lexicon, phonaesthesia, dictionaries

1 Introduction

In this paper we investigate a phenomenon in St'át'imcets (a.k.a. Lillooet Salish: ISO 639-3: lil) that as far as we know has never systematically been examined in Salish before: the existence of *root clusters*, by which we mean sets of derivationally distinct roots which are nonetheless related by both sound and meaning, with an identifiable common phonological substring and a common lexical-conceptual core.

The immediate motivation for our investigation is work we have been doing on the St'át'imcets-English index to an English-Upper St'át'imcets dictionary we have been compiling in collaboration with the Upper St'át'imc Language, Culture and Education Society (Davis et al. in prep.). Unlike the main English-St'át'imcets part of the dictionary, the index is root-based; during the process of compiling it, we were struck over and over again by what we now recognize as root clusters. However, we were unsure of whether and how to incorporate them into the dictionary. This led to wider questions of what defines a root cluster, why they exist, and what purposes they might serve in the mental lexicon. This paper is a preliminary attempt to tackle these questions.

2 Background

Root clusters exemplify the phenomenon of *phon(a)esthesia*, a term first coined by the British linguist J. R. Firth in 1930 from the Greek φωνή *phone*, 'sound', and αἴσθημα *aisthema*, 'perception' (from αίσθάνομαι *aisthanomai*, 'I perceive') to label the systematic pairing of form and meaning in a language. Here's a better description, from a 1998 Linguist List post by Philip Grew:

> Clusters of similar-sounding words with semantic affinities are the basis of phonaesthesia. In this phenomenon, associations arise among groups of words, which

* Acknowledgements. We thank all the speakers of St'át'imcets, past and present, who have contributed to ongoing dictionary compilation on their language, to the Upper St'át'imc Language, Culture and Education Society, sponsor of the English-Upper St'át'imcets dictionary that has provided much of the empirical foundation for this research, and to members of the Salish Working Group, who contributed many useful comments at an earlier presentation of this work. Research on St'át'imcets has been supported by SSHRC Insight Grant 435-2020-1293 to the first author.
Authors' e-mails: henry.davis@ubc.ca, bailey.trotter@gmail.com

Papers for the International Conference on Salish and Neighbouring Languages 58.
D. K. E. Reisinger, Laura Griffin, Gloria Mellesmoen, Sander Nederveen, Julia Schillo, Bailey Trotter (eds.). Vancouver, BC: UBCWPL, 2023.

> may have close, distant or no etymological relations. These associations may then transfer to a sequence of phonemes shared by the words with some perceived common element of meaning, creating phonaesthemes. The presence of that same string of phonemes may then in turn lend a shade of the meaning felt to characterize the phonaestheme to another expression, simply because the latter contains that string. (Grew 1998)

However, even if you have never heard the term before, you have almost certainly wondered about the similarity between the members of the following sets of English words, whose bolded substrings are phonaesthemes (there are others, but these are amongst the best-known):

(1) a. ***gl****ass* ***gl****isten* ***gl****ow* ***gl****are* ***gl****aze* ***gl****int* ***gl****itter* ***gl****eam* ***gl****immer* ***gl****oss* ***gl****amour* ***gl****abrous*

b. ***sn****out* ***sn****ot* ***sn****eeze* ***sn****iff* ***sn****uffle* ***sn****ort* ***sn****ub* ***sn****ooty* ***sn****ore* ***sn****ooze* ***sn****ivel* ***sn****ob* ***sn****eer*

c. *c****lash*** *cr****ash*** *sm****ash*** *b****ash*** *m****ash*** *h****ash*** *g****ash*** *gn****ash*** *sl****ash*** *spl****ash***

It is important to emphasize a couple of points before we go further. First, as noted by Grew, words in a phonaesthetic relationship are not necessarily diachronically related, though a set of phonaesthemes often contains a core of etymologically connected forms. For example, amongst the set of *gl-* words above, *glisten*, *glint*, *glitter*, *gleam*, and *glimmer* all derive from the Proto Indo-European (PIE) root **ǵʰley*, but *glow*, *glass*, *glare*, and *gloss* come from **ǵʰelh* (also the root for *gold* and *yellow*). Of course, these two PIE roots are also close in both sound and meaning, suggesting that phonaesthetic relations themselves may well be inherited diachronically: we return to this question in a Salish context in Section 6 below.

Second, it is necessary to draw a distinction between phonaesthesia and the more general term *sound symbolism*. The latter is used for a much wider range of sound-meaning correspondences: for example, Hinton et al. (1994:5) describe sound symbolism very broadly as "the direct linkage between sound and meaning". Much of the literature on the topic is — not coincidentally — devoted to establishing a typology that attempts to draw principled distinctions between different kinds of sound symbolism. A basic distinction that emerges is between *iconic* and *non-iconic (conventional)* types.

Iconic types involve a non-arbitrary relation between particular sounds and real-world referents. They include: (i) *imitative sound symbolism* (i.e., onomatopoeia), including animal noises ('woof', 'bowwow', 'grrr' (cf. 'growl')) as well as more general sounds ('squeak', 'creak', 'crack', 'boom', 'whistle');[3] (ii) *vocal gestures* (e.g., paralinguistic exclamations of surprise, fear, pain,

[3] There is little doubt that onomatopoeia plays a similar role in Salish to English. Here are some iconic (non-arbitrary) sound-symbolic correspondences between English and St'át'imcets words describing particular noises:

	SOUND	ST'ÁT'IMCETS	ENGLISH
(i)	floorboards, stairs	√***ʕik***, √**kik**, √**ki**ƛ̓	***cr****eak*, ***sq****ueak*
(ii)	whistling (e.g., wind)	√***xʷit***	***whi****stle*
(iii)	thunder	√***k****n****ap***	*thunder****cl****ap*
(iv)	rumbling (herd of cattle, big truck)	√***ʕəm-p***	***rumb****le*
(v)	rustling (leaves, creek)	√***ʕəx, ʕəxʷ***	***ru****stle*
(vi)	snuffling noise	√***x̌nu̓ƛ̓***	***snuffle***

etc.); (iii) *diagrammatic* sound symbolism, including *gestalt iconicity*, where features of the utterance directly reflect properties of the intended meaning (e.g., reduplication for event repetition or plurality, vowel lengthening for increased time or distance); and *relational (synesthetic) iconicity*, where for example languages employ high vowels for small things and low vowels for large things, reflecting the size of the vocal tract aperture or the frequency of the resulting acoustic signal.[4]

Non-iconic or conventional sound symbolism, on the other hand, involves associations between *arbitrary* (sets of) sounds and (sets of) meanings. Hinton et al. (1994:5) define it as "the analogical association of certain phonemes and clusters with certain meanings"; Johansson et al. (2020) define it as "complex (associative) … language-internal mappings which emerge through analogy"; and most eloquently, Bloomfield (1914/1983:410, cited in Hinton et al. 1994) captures it as follows: "The signification of any word is arbitrarily attached to some sound contained in it, and then co-generic names are created by means of this infused or one might say irradiated or inspired element."

It is this second, non-iconic type of sound symbolism which covers phonaesthesia, and which concerns us here.

3 Phonaesthetic patterns in St'át'imcets

As mentioned in the introduction, our investigation of phonaesthesia in St'át'imcets was prompted by our discovery of *root clusters* during dictionary compilation. By hypothesis, these clusters exemplify phonaesthetic relations, meaning that they involve non-derivational correspondences between sets of phonologically and semantically related roots. Accordingly, we now turn directly to an examination of their properties.

We begin with an obvious question: *How do we know when a set of roots form a root cluster?* There are two parts to the question, one based on sound and one on meaning:

(i) *When do roots count as phonologically close enough to count as a cluster?*
(ii) *When do roots count as semantically close enough to count as a cluster?*

We address these sub-questions in Sections 4 and 5, respectively.

4 The phonology of root clusters

At least at first glance, the phonological sub-question might appear to be more tractable than the semantic one. After all, it is possible to devise metrics for phonological closeness based on the number of phonemes/features shared by a given set of roots.

We begin with a simple case, which we will refer to as the **Stop-Still** cluster.

Linguistically, however, these are less interesting for us than cases where there *isn't* an iconic correspondence, so we set them aside here: see Thomason (1999) on Montana Salish for more on iconic sound symbolism in Salish.

[4] The latter also relates to a strand of psycholinguistic research around the *Bouba-Kiki effect*: the basic idea is that certain sounds (labials and back vowels) are associated universally with soft, round objects, and others (coronals, dorsals, and short high vowels) with spiky, angular objects.

(2) a. √*ƛ̓al* ‘stop’ (as in *ƛ̓ál-ləx* ‘stop (oneself)’)
b. √*ƛ̓əl* ‘still’ (as in *ƛ̓ál-p* ‘settle down’)
c. √*ƛ̓il* ‘calm, still’ (as in *kʷa-ƛ̓íl-a* ‘calm down’)
d. √*ƛ̓ul̓* ‘calm’ (as in *ƛ̓əl̓-ƛ̓úl̓* ‘calm, of water’)

These four roots count as a phonological cluster on the following basis:

(i) *They are not derivationally related*. There is no systematic synchronic (nor, for that matter, diachronic) process of ablaut/apophony which could lead to any of them being derived from the others.

(ii) *They share common phonological content:* in this case, the consonantal skeleton *ƛ̓__l/l̓*.

(iii) *They differ along some phonologically identifiable parameter*: in this case, the vowel melody.

Things get more complicated when we consider what we might call *cluster continua*. Consider the following **Rumble-Rustle** cluster:

(3) a. √*mə̓m*, √*məm* ‘sound of tires on a road’
b. √*ʕəm* ‘make a rumbling noise (like a herd of animals or a rockslide)’
c. √*ʕəx*, √*ʕəxʷ* ‘make a rustling noise (like a creek)’
d. √*ʕəlx̓ʷ* ‘make a noise (not human or animal)’
e. √*qəlx̓ʷ* ‘make a lot of noise’
f. √*q̓əlx̓ʷ*, √*q̓ʷəlx̓ʷ* ‘make a rumbling noise (like stomach)’

Notice that aside from the schwa, there is no phonological correspondence between the root in (3a), √*mə̓m*, √*məm* and that in (3f), √*q̓əlx̓ʷ*, √*q̓ʷəlx̓ʷ*. Yet each root in between shares at least one consonant with the preceding and following roots, forming a continuum that links the peripheries of the cluster.

Still more complex examples exist in which two or more clusters overlap, forming “macro-clusters”. Take for example the **Yellow-Green** cluster in (4), based around the retracted root √*kʷl̓* ‘yellow(ish), green(ish)’:

(4) a. √*kʷl̓* ‘yellowish, greenish’
b. √*kʷl̓iʔ* ‘yellow, green’
c. √*kʷl̓it* ‘brass, pig iron’
d. √*kʷl̓im* ‘pus under skin’
e. √*kʷə̓l̓ən̓* ‘gall bladder’
f. √*nakʷúl̓* ‘butterfly (coloured)’[5]

[5] Van Eijk (2013:235) groups (4a), (4b), (4c), (4e), and (4f) under the same root, √*kʷə̓l̓*. While historically this is surely the case (see Kuipers 2002:44), synchronically it depends on positing a number of otherwise unmotivated affixes (*-iʔ, -it, -ən̓, na-)* which van Eijk labels simply as “formatives”. We prefer to treat these cases as involving closely related roots with a common diachronic source, though this cluster, unlike those in (2) and (3), clearly does contain a core root (√*kʷ(ə)l̓*), extended by various pieces of fossilized morphology. There are a number of other related stems (e.g., *kʷl̓-m-q̓akst* ‘wolf lichen, yellow tree moss’, *s-kʷl̓-q̓al̓st* ‘jade’) which we have treated as synchronically derived from √*kʷ(ə)l̓*, but which could also be treated as lexicalized

This cluster overlaps with the **Sunshine** cluster, (loosely) based around the non-retracted root √*kʷəl(ʼ)*.[6]

(5) a. √*kʷəlʼ* ‘sunshine’
b. √*kʷəl(ʼ)tapís* ‘palomino horse’
c. √*kʷluʔ* ‘red alder’[7]
d. √*kʷalʼá* ‘dark brown horse’

This is in turn overlaps with a **Ripen-Cook-Blacken** cluster centred around the root √*qʼʷəl* ‘cook, ripen’:

(6) a. √*kʷəl* ‘cook on an open fire’
b. √*qʼʷəl* ‘cook, ripen’
c. √*qʼʷəɬ* ‘scald, burn’
d. √*qʼʷuɬ* ‘burn’
e. √*qʼʷaz* ~ *qʼʷ*əz ‘burnt, blackened, painted’
f. √*qʼaɬ* ‘soot, heat’
g. √*qʷuɬ* ‘blistered (from heat)’

And this in turn overlaps with a **Burn-Shine** cluster centred around √*ʕʷəl* ‘burn’ (7):

(7) a. √*ʕʷəl* ‘burn’
b. √*ʕʷáyʼəp* ‘northern lights’
c. √*wəl̓* ‘shine
d. √*wəl* ‘flash’
e. √*wəqʼ* ‘shine’

In view of the data above, let us now re-assess the three criteria we hypothesized above for two or more roots to count as a cluster.

(i) *They are not derivationally related*.

This is trivially true in order for a root cluster to exist in the first place (otherwise we’d have a single root), but it is not true diachronically, since root clusters can certainly contain roots with a common core, differentiated by fossilized morphology: see (4) above for a case in point.

(ii) *They share common phonological content.*

---

and therefore as part of the root cluster in (4); obviously, which of these stems to treat as derived and which as roots is to a certain extent a judgment call.

[6] Following Kuipers (2002), we ignore glottalization on resonants for the purposes of root-to-root comparison, since it is unreliable diachronically and variable synchronically. Likewise, the presence of schwa in a root is generally linked to syllable structure and doesn’t tell us anything about the underlying representation (which is probably schwa-less: see Matthewson 1994).

[7] The semantic connection here is probably to the orange inner bark of the red alder (see Kuipers 2002:44).

This is clearly not true for all members of root cluster continua, since as observed above, forms at opposite edges of a continuum need not share any common phonological content. However, differences are clinal, in the sense that each member of the continuum must share phonological content with at least one other adjacent member.

(iii) *They differ along some phonologically identifiable parameter.*

As with (ii), each pair of adjacent forms in a continuum differ on one parameter, but non-adjacent forms may differ along multiple parameters, as long as the differences are connected step-wise.

What this indicates is that root clusters are not simply sets, but structured continua with edges shared by other clusters. In this sense, they can be thought of in terms of *associative networks*, rather than discrete categories. It remains to be determined, however, whether they form simple chains of adjacent related pairs, or (we suspect more likely) "clouds" with core and peripheral members (or possibly both types). This is a direction for future research; see Section 6 below for a first attempt at a graphic representation of relations within and between clusters.

We now turn to a different phonological question, namely: *To what extent do the phonological differences between members of root clusters mirror synchronic and/or diachronic phonological alternations?*

In terms of synchronic phonological processes, the answer is: surprisingly little. The commonest phonological differences between members of root clusters are the following (not in order of frequency):

(8) *Ejective vs. plain stops:*

a.	√*qəlʼxʷ* 'make a lot of noise'	vs.	√*qʼəlʼxʷ ~ qʼʷəlʼxʷ* 'make a rumbling noise (like stomach)'
b.	√*x̌əkʷ* 'get jammed into something.'	vs.	√*x̌əkʼʷ* 'get stuck onto something'
c.	√*səq* 'split'	vs.	√*səqʼ* 'crack'
d.	√*pək* 'get partially flattened'	vs.	√*pʼək* 'get flattened or dented'

(9) *Velar vs. uvular stops:*

a.	√*pʼəkʼ* 'get flattened or dented'	vs.	√*pʼəqʼ* 'get flattened completely'
b.	√*λʼikʷ* 'popping sensation'	vs.	√*λʼiqʷ* 'cracking or crackling sound'
c.	√*łukʷ* 'bail a boat'	vs.	√*łuqʷ* 'scoop, ladle, bail'

(10) *Lateral resonant vs. lateral fricative:*

a.	√*luq̓ʷ* 'bark gets peeled'	vs.	√*łuq̓ʷ* 'get taken of, stripped'
b.	√*lə̓s* 'cave in'	vs.	√*łə̓c̓* 'collapse'
c.	√*q̓ʷəl* 'cooked, ripe'	vs.	√*q̓ʷəł* 'burned, scalded'

(11) *Ablaut/apophony (cf. (2) above):*

a.	√łam̓ 'get put in'	vs.	√łum(') 'get put on, attached'
b.	√łəxʷ 'get put on (e.g. clothes)'	vs.	√łuxʷ 'get inserted into something hollow'
c.	√ɣəp 'upright'	vs.	√ɣap 'tree' vs. √ɣip 'grow'

(12) *Retracted vs. non-retracted root:*

a.	√*kʷə̓l* 'yellowish'	vs.	√*kʷəl'* 'sunshine'
b.	√*łə̓k* 'plop or flop down'	vs.	√*łək* 'deflate'
c.	√*pə̓lxʷ* 'pop up'	vs.	√*pəlx̓ʷ* 'appear all of a sudden'
d.	√*łu̓ƛ̓* 'slurp, sip'	vs.	√*łuƛ̓* 'sip or swallow something whole'

None of these alternations have any synchronic reflex in St'át'imcets.

However, though less easy to pinpoint, it does seem that most of them have historical reflexes: at least (10) to (12) show up quite frequently in diachronic changes (Kuipers 2002), though this question merits a great deal more attention, preferably on the basis of more detailed comparison of the entire set of roots of multiple languages.

It is also worth pointing out that some processes which are likely important historically *don't* show up (at least, with any regularity) in St'át'imcets root clusters. Prominent amongst these (though it is always possible we've missed some cases) is *metathesis* (which Kuipers 2002 refers to as *inversion*). So far, we have only found, one case which probably actually involves dialect variations of the *same* root:

(13)	√*kʷəm* 'blunt, dull' ("west side" dialect)	vs.	√*makʷ* 'blunt, dull' ("east side" dialect)

5 What about the semantics of root clusters?

We now turn to the lexical semantics of root clusters. So far, we have been making a tacit appeal to what we might call *trans-lexical semantic similarity*, referring to resemblances that hold between independent lexical entries. But what can we say explicitly about these relations?

As perhaps might be expected, what we find is that root clusters correspond to lexical semantic fields organized around common conceptual cores. Here are some prominent cases.

(i) *Colours*. See (4) above for the Yellow-Green cluster, but 'white', 'red', and 'black' also form clusters. Here's the **Red-Bleed** cluster:

(14) a. √*ciqʷ* 'red'
b. √*ciʕʷ* 'bleed'
c. √*cáqʷəm* 'saskatoon berry'
d. √*cəqʷlíṁ* 'chestnut horse'

And here's the **White-Faded** cluster:

(15) a. √*pəq, paq* 'white'
b. √*ṗaqm* 'flower'
c. √*ṗaʔkʷz* 'white moss'
d. √*páqəm* 'mouldy'
e. √*pəʕ, paʕ* 'faded, pale, greyish'

(ii) *Manners of motion*. One of the larger clusters we have identified involves circular motion: we call it the **Circle-Spiral** cluster:

(16) a. √*zal* 'twist, whirl'
b. √*zalk̓ʷ* 'wrap around'
c. √*zulkʷ* 'knead, rub, or massage'[8]
d. √*zəl* 'ripple'
e. √*zaw* 'swerve around'
f. √*zánm* 'go around'
g. √*zənk* 'circle'
h. √*zánuxʷ* 'year'
i. √*zənṗ* 'wind around'

A related cluster involves rotation, including movement through the air: we call it the **Drill-Whirl** cluster.

[8] Several roots involving circular motion (including √*za**lk̓ʷ***, √*zu**lkʷ***, √*xʷu**lk̓ʷ***, √*xʷəlá**kʷəm̓***, and also √*mə**lk̓ʷ*** 'wrap or roll up') have an accretion consisting *of -kʷ ~ -kʷ*. Van Eijk (2013:480) treats *-kʷ* as a suffix, but it is not productive, nor does it fall into any existing class of suffixes in the language: it does not show the behavior of a lexical suffix, for example, though it has a lexical rather than a functional meaning. In fact, it looks rather like a classical 'phonaestheme' — a stretch of phonological material with a semantic value but no morphological status, like the sequence *gl-* in 'gleam', 'glow', etc.: see (1).

(17) a. √*xʷəlp* ‘turn, spin’
b. √*xʷul̓* ‘drill or bore’
c. √*xʷulk̓ʷ* ‘roll (e.g., a cigarette)’
d. √*xʷəlákʷəm* ‘gust, whirlwind’
e. √*xʷə̓l̓* ‘breeze, draft’
f. √*xʷəlátən* ‘flying squirrel’
g. √*xʷálap̓* ‘fan’

Another large cluster (or more likely, cluster complex) involves peeling, stripping, skinning, splitting, and cracking: we call it the **Peel-Strip-Crack** complex.

(18) a. √*siʕʷ* ‘loosen’
b. √*səʕʷ* ‘peel’
c. √*səx̌ʷ* ‘peel, split’
d. √*sux̌ʷ* ‘strip’
e. √*suq̓ʷ* ‘skin (e.g., an animal)’
f. √*sáq̓ʷuɬ* ‘(split in) half’
g. √*səq* ‘split’
h. √*səq̓* ‘crack’

(iii) *Spatial configuration*. The clusters under this heading are related to manner-of-motion clusters and could possibly be either assimilated to them or grouped together with them as macro-clusters. They include the **Bent-Crooked** cluster in (19) and the **Stretch-Stand** cluster in (20).

(19) a. √*k̓ʷə̓lc̓* ‘bend (road or river)’
b. √*k̓ʷə̓l̓* ‘hooked (beak)’
c. √*k̓ʷuc̓* ‘crooked’
d. √*k̓ʷult, k̓ʷəlt* ‘go downhill diagonally’

(20) a. √*taɬ* ‘stand’
b. √*taɬáʔ* ‘other side’
c. √*təɬ* ‘stretch, extend’
d. √*tə̓ɬ* ‘over-exert’
e. √*tə̓l̓* ‘unwind (of string or thread)’

(iv) *Sound*: see (3).

(v) *Light*: see (7).

(vi) *Heat*: see (6).

These last three categories possibly fall under a larger macro-category of *sensation.*

It is important to emphasize that, just as with the phonological resemblances that characterize root clusters, the semantic associations outlined above do not define discrete and mutually exclusive clusters, but rather a network of associations centered around one or more conceptual “nuclei”. This is most obvious when we consider macro-clusters such as that described in (4) to (7)

above, for example: note the associative links that take us from the colour 'yellow' to the colour 'black': *yellow* ↔ *sunshine* ↔ *heat* ↔ *cook* ↔ *burn* ↔ *black(en)*. Here, we are clearly not dealing with a set of isolated clusters, but rather a network of connected nuclei. It is even possible that the entire lexicon can be defined as a single network, with all conceptual nuclei connected in this way via associative links.

Big questions arise as to the universality of these trans-lexical conceptual relations (and potentially, their basis in primitives of human cognition and/or culture). These questions cannot be answered on the basis of a single language family, let alone a single language. Rather, they set a research agenda which begins with individual languages, then looks at the proto-languages from which they descend, and then compares root clusters across unrelated families.

In the following section, we take a tentative step towards addressing the second (historical) part of this agenda, by comparing the root clusters we have discovered in St'át'imcets with those we have been able to detect in Proto-Salish (or in some cases, Proto-Interior Salish), based on the reconstructions in Kuipers (2002).

6 Representing Salish root clusters across time

Each point of root cluster comparison between Proto-Salish and its daughter languages has three potential outcomes. First, a root cluster in a daughter language could be directly inherited from the proto-language, in which case all (or at least, most) of its members will have counterparts in Proto-Salish. Second, a cluster in a daughter language could have no (or very few) counterparts in the proto-language, in which case we would have to conclude it was an innovation. And third, a Proto-Salish cluster could have no or few counterparts in a daughter language, which would constitute a case of cluster loss.

Here we confine ourselves mostly to the first two possibilities, for practical reasons: in order to properly investigate the third, we would have to identify all the root clusters in Proto-Salish, and then check whether each one corresponded to a cluster in St'át'imcets. While beyond the scope of this preliminary inquiry, that is certainly a feasible project, and would furthermore be useful for a parallel exercise involving any of the other 22 daughter languages in the Salish family.

As suggested above, we think the best way to represent root clusters is as associative networks: we have therefore begun to explore these relations graphically. In the diagrams that follow, we represent associations at the synchronic level by horizontal lines and historical connections by vertical lines. This allows for several kinds of diachronic relation: one-to-one correspondences, in which each member of cluster in a daughter language is directly inherited from a cognate of the same cluster in the mother language; one-to-many correspondences, in which a daughter expands a cluster inherited from the mother language; many-to-one correspondences, in which not all members of a cluster in the mother language are inherited by the daughter language; and non-correspondences, where either a root cluster in the mother language has no correspondents in the daughter language, or vice versa. Note that we do not deal here with many-to-many correspondences: however, we suspect that a more accurate picture of root clusters both in the synchronic and diachronic dimensions might well involve such correspondences, which would require a more sophisticated multi-dimensional graphic representation.

We begin with a relatively simple case: the **Still-Stop cluster** featured in (2) above.

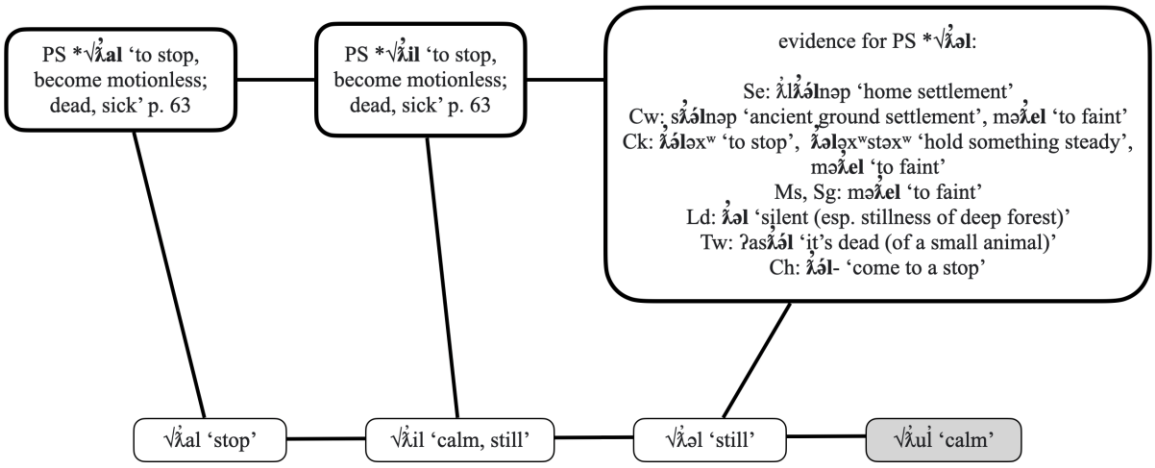

Figure 1: The Still-Stop cluster[9]

The diagram shows that all except one of the members of the St’át’imcets cluster are inherited from Proto-Salish, as indicated by the vertical lines. The odd member out is *√ƛ̓uɬ̓*, which appears to be a St’át’imcets innovation; we mark non-corresponding cluster members in grey. In the large box, we have provided more detailed evidence from Kuipers (2002) for the Proto-Salish reconstruction; language abbreviations are the standard ones used in Salish historical linguistics.

In Figure 2, we give the cluster centered around the conceptual core ‘bent, crooked’: see (19).

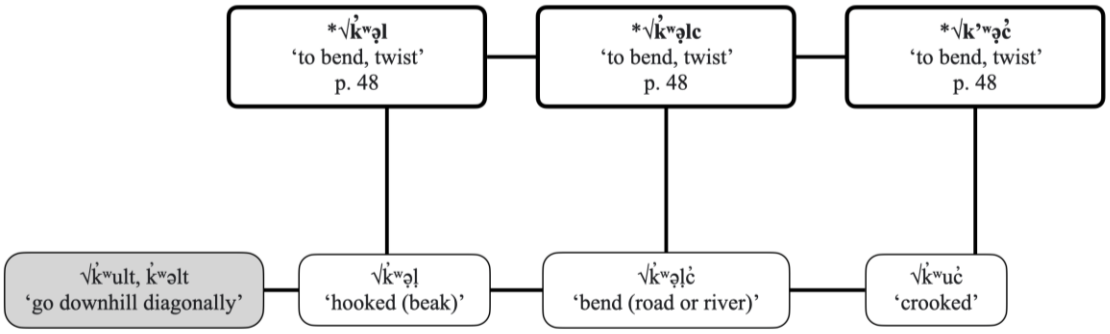

Figure 2: The Bent-Crooked Cluster

Once again, three members out of the four roots in the St’át’imcets cluster are inherited. The fourth, *√k̓ʷult*, differs from the other three in that (i) it is not retracted, and (ii) it has a frozen version of the “immediate” suffix *-t* attached to it (*-t* is not productive in St’át’imcets, but appears on a number of adjectives and intransitive verbs.)

Figure 3 features the cluster centered around the colour ‘red’ (see 14).

[9] The page numbers included in figures refer to Kuipers (2002).

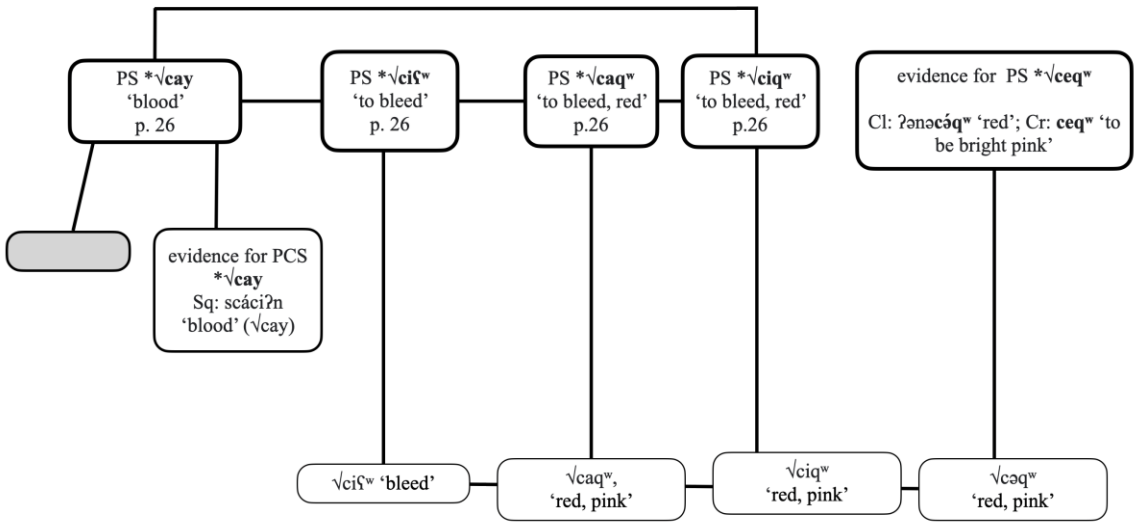

Figure 3: The Red-Bleed Cluster

Here, we see the inverse situation from the first two clusters: every member of the St'át'imcets cluster has a Proto-Salish counterpart, but not every member of the Proto-Salish cluster has a counterpart in St'át'imcets. In particular, the Proto-Salish root *√cay* 'blood' has no reflex in St'át'imcets (which uses *√ptə́laʔ* for 'blood'); the same appears to be true more generally of Interior Salish (Kuipers 2002:26), suggesting the form was lost in the descent from Proto-Salish to Proto-Interior Salish. The non-descending member of the Proto-Salish cluster is marked in grey.

Figure 4 has 'white' as its semantic core: see (15) above.

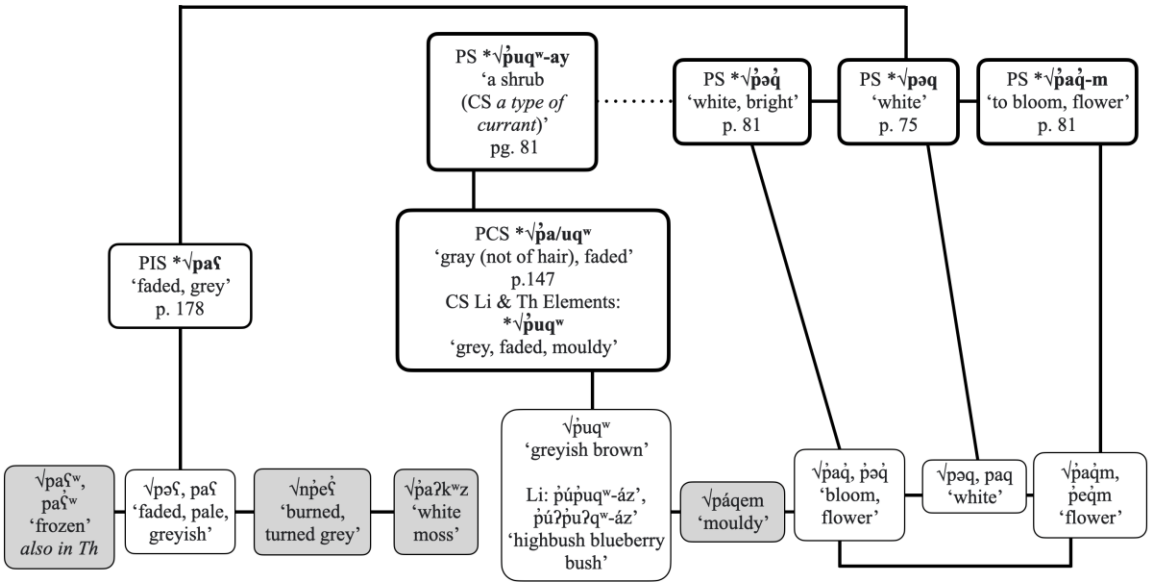

Figure 4: The White-Faded Cluster

This is a more complex case, consisting of at least two sub-clusters, one dating back to Proto-Salish and centered around ‘white’ and ‘flower’, the other confined to Interior Salish and centered around ‘grey’ and ‘faded’. Note that the Proto-Interior Salish root *√*paʕ* features a pharyngeal; pharyngeals are absent elsewhere in the family, though Kuipers (2002:3) claims that they have uvular reflexes in Coast Salish languages, thus accounting for the possible historical link between *√*paʕ* and *√*pəq*.

Figure 5 features the **Peel-Split-Crack** complex in (18).

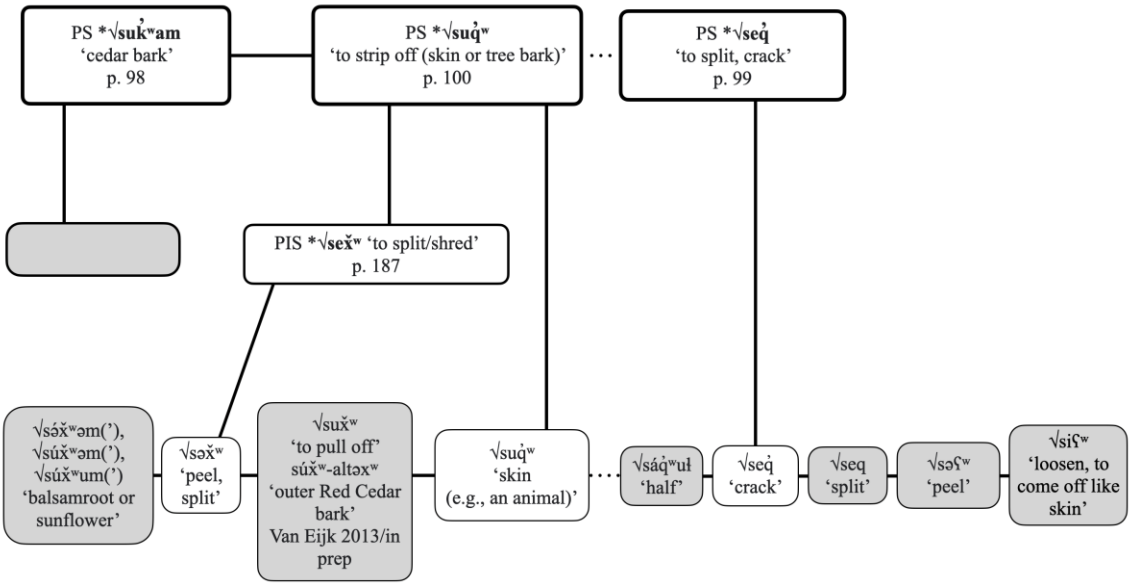

Figure 5: The Peel-Split-Crack complex

There are clearly two clusters here, one linked to PS √*suqʷ* ‘strip, peel, skin’, the other to PS √*səq* ‘split, crack’. The connection between them is more robust in St’át’imcets than in PS, as evidenced particularly by √*səq* ‘split’ and √*səʕʷ* ‘peel’. Overall, this complex shows more diversity in the daughter language than in the proto-language, a situation which is more common than the converse, and we suspect may be more an artifact of reconstruction than a reflection of the languages themselves (it is easier to reconstruct multiple forms in a daughter language to a single proto-form than vice-versa).

Figure 6 features a macro-cluster including the **Circle-Spiral** and **Drill-Whirl** clusters, as shown in (16) and (17) above, respectively. We refer to it as the **Rotation** complex.

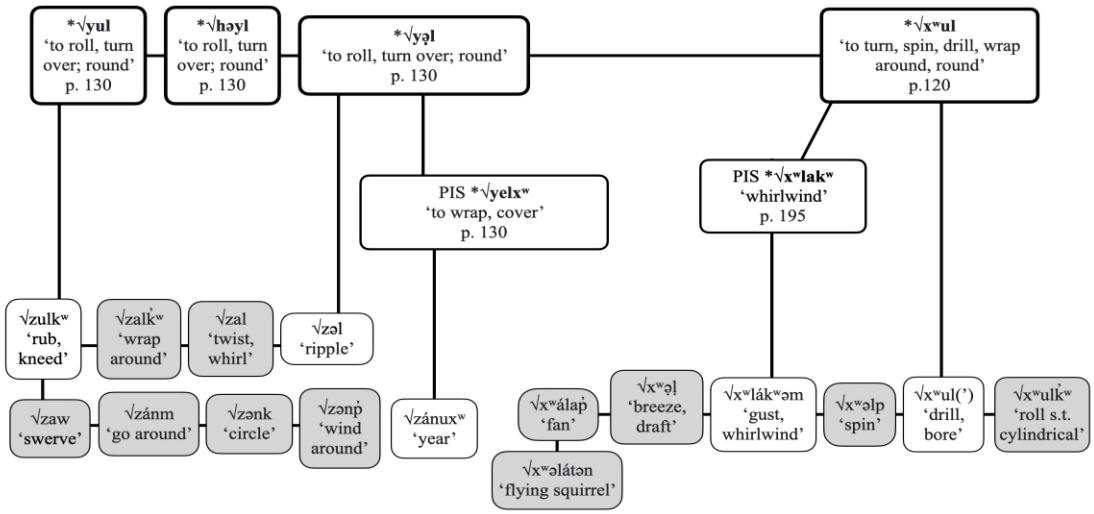

Figure 6: The Rotation complex

The two clusters in this complex (**Circle-Spiral** and **Drill-Whirl**) are semantically close but differ phonologically in their onset consonant ([z] < PS *[y] versus [xʷ]); however, they are phonologically linked by an [l] in their codas, as well as by the *-k̓ʷ ~ -kʷ* increment which appears on members of both clusters: see footnote 12. There has been considerable expansion of this complex in St'át'imcets, as shown by the number of grey boxes at the bottom of the diagram.

Figure 7 also features a complex of two related clusters, the **Yellow-Green** cluster in (4) and the **Sunshine** cluster in (5).

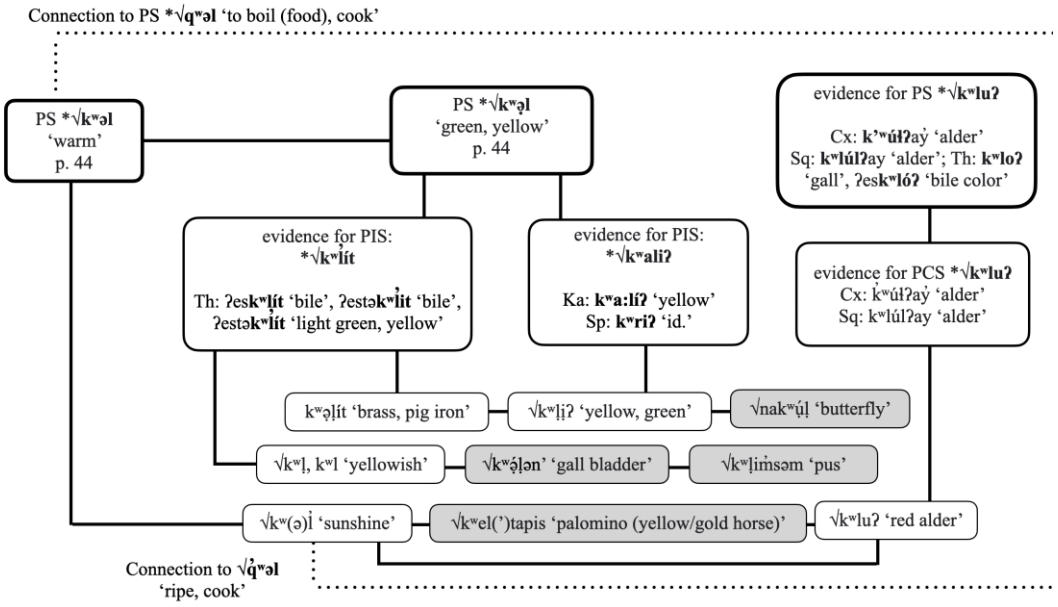

Figure 7: The Yellow-Sunshine Complex

Diachronic relations in this complex are complicated by the fact that the core root for 'yellow/green' in PIS contains a lexically retracted vowel, but retracted vowels are absent in other branches of the family. Kuipers (2002:44) reconstructs the retracted root *√*kʷẹ̓l* back to Proto-Salish, but St'át'imcets must have inherited it via PIS. On the other hand, non-retracted variants could have two sources: they could have either been inherited from PIS and subsequently lost their retraction, or been borrowed from a Central Salish language (as is likely for √*kʷlu?* '(red) alder', whose distribution is largely coastal and which has cognates in many Central Salish languages).

The Yellow-Sunshine complex is likely related to the next complex, the **Ripen-Cook-Blacken** complex, via the associations 'sunshine' ↔ 'ripen' ↔ 'cooked' ↔ 'burn' ↔ 'blacken'.

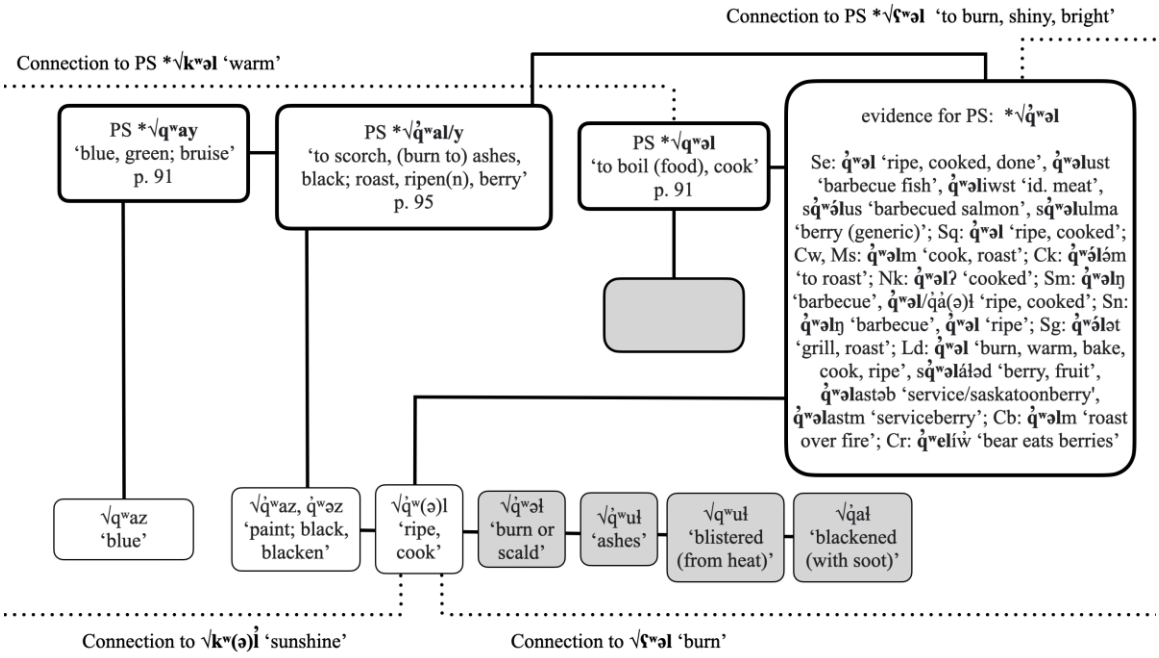

Figure 8: The Ripen-Cook-Blacken Complex

The nucleus of this complex is the root *√*q̓ʷəl*, which is very widely attested across Salish with the two associated meanings 'ripen' and 'cook'. In Kuipers (2002:95), this root is included under *√*q̓ʷal/y* 'scorch, (burn to) ashes, black', but we have separated the two roots here: we believe that they represent an instance of a Proto-Salish root cluster, with closely related but distinct phonological and semantic lexical entries. The two meanings of *√*q̓ʷəl* are extended to 'ripen, heat, sun' on the one hand, which links to the Sunshine-Yellow complex in Figure 7 above, and to 'burn, char, blacken' on the other, which links to the **Burn-Shine** complex featured in Figure 9 below.

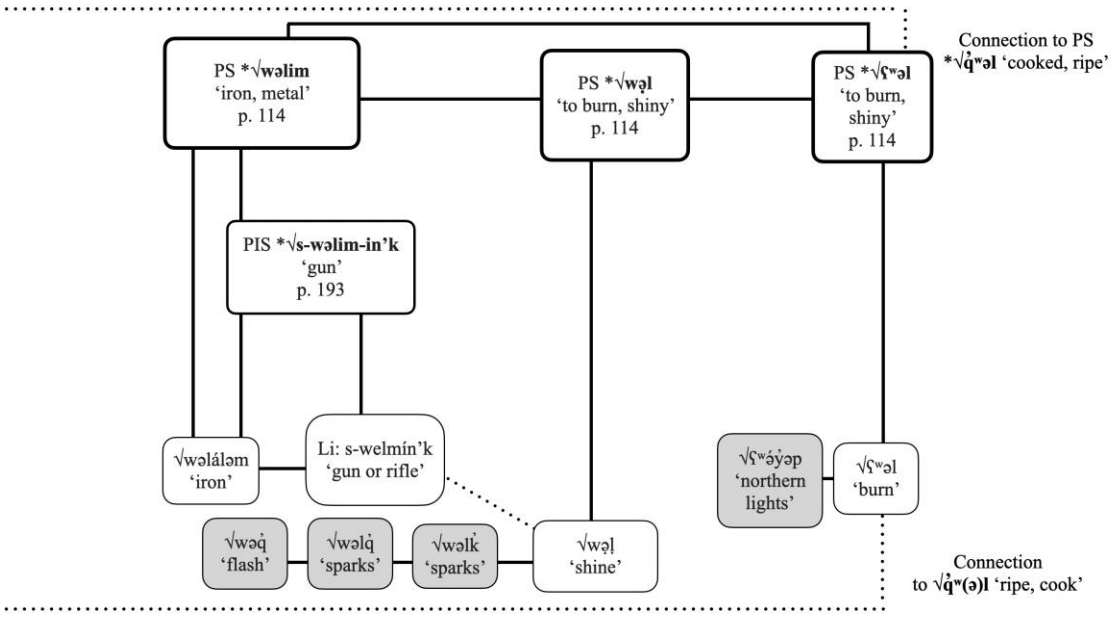

Figure 9: The Burn-Shine Complex

This is an interesting complex because it is more closely linked at the level of PS than in St'át'imcets, where it is split into three closely related clusters, one centered around 'burn', one around 'metal', and one around 'flash, shine, spark'. The whole complex links to the Ripen-Cook-Blacken complex via the association *√*q̓ʷal/y* ↔ *√*ʕʷəl* at the PS level and √*q̓ʷəl* ↔ √*ʕʷəl* at the daughter language level.

What can we conclude from this preliminary investigation of the diachrony of root clusters? First of all, that there is a strong tendency for clusters to be inherited: all of the root clusters we have independently identified in St'át'imcets at least have cluster counterparts in Proto-Interior Salish, and most of them go back to Proto-Salish. This speaks to the putative universality of the clusters and associative networks that connect them.

Second, and complementary to the first point, there is considerable innovation in the daughter languages (at least, judging from St'át'imcets). This is significant, in that it potentially identifies a different diachronic mechanism than the classic Neo-Grammarian model of conditioned and unconditioned sound changes. More specifically, it provides for a type of associative change that comes about through the creation of new lexical items, rather than through changes in existing ones; this both complicates reconstruction and potentially provides a new way of investigating perceived irregularities in standard diachronic accounts.

7 Why do root clusters exist?

Assuming we have identified a real phenomenon and not just an epiphenomenon, let us briefly consider the question of *why* roots cluster in the way they do. (One way of thinking about this is to ask why lexical distribution is "lumpy" rather than "smooth".)

There are a number of ways to answer this question, not all of them mutually exclusive. Perhaps the most obvious is to appeal to mechanisms of lexical innovation. When a language needs to coin

new words at the root level, one option is to extend its existing vocabulary by modifying the phonology of semantically related roots. This process can be conscious, as in, e.g., Lewis Carroll's famous poem 'The Jabberwocky', which makes explicit use of blends ('lithe' and 'slimy' → *slithy*, 'miserable' and 'flimsy' → *mimsy*), archaisms (both 'gyre' and 'beamish' were obsolescent words rescued by Carroll), and sound symbolism (*snicker-snacker* mimics the sound of cutting).[10] However, these are not the same mechanisms we see operating in root clusters, suggesting that conscious coining of neologisms does not operate on the same principles as root cluster extension.

An appeal to earlier and less conscious stages of neologism formation might be more illuminating. Under such an account, mis-learning in language acquisition drives the formation of root clusters: learners innovate roots by misperceiving existing forms, and rather than eliminating earlier errors when the target roots are acquired, then add them to their lexicon along with the original roots (presumably, as long as the innovated roots fill a useful function in their expanded lexicon). This account makes the prediction that the types of phonological relation instantiated in clusters (see Section 4) mirror errors in lexical acquisition, a hypothesis that is certainly in principle testable.

An alternative type of explanation might appeal to lexical retrieval. The idea here is that lexical access is facilitated by clustering roots in the same phonological and semantic space. For example, the Still-Stop cluster in (2) could be accessed all at once, narrowing the lexical choices necessary for production and perception purposes. Again, this type of hypothesis is at least in principle testable, perhaps via artificial language learning.

There might also be a connection between phonaesthesia and cases of language loss such as Wernicke's (receptive) aphasia, which is characterized by extreme difficulty in lexical retrieval, often involving phonological and semantic *paraphasia*, as defined below:

(21) a. *Phonemic* (literal) *paraphasias*: involves the substitution, addition, or rearrangement of sounds so that an error can be defined as sounding like the target word. Often, half of

[10] It is worth noting, however, that Lewis Carrol's neologism formation wasn't entirely conscious. In fact, a lot of his etymology may have been post-hoc, as he indicates in an 1877 letter to Maud Stanton:

> I am afraid I can't explain '*vorpal blade*' for you — nor yet '*tulgey wood*', but I did make an explanation once for '*uffish thought*'! It seemed to suggest a state of mind when the voice is gruffish, the manner roughish, and the temper huffish. Then again, as to '*burble'*, if you take the three verbs 'bleat, murmer, and warble', then select the bits I have underlined, it certainly *makes* 'burble', though I am afraid I can't distinctly remember having made it in that way. (Graham 1981)

He also hints broadly at the same point in Humpty Dumpty's exposition of *Jabberwocky* to Alice in *Alice Through the Looking Glass* (Carroll 1882:126–127). The exposition begins very reasonably with Humpty Dumpty giving a lucid explanation of a linguistic blend (or "portmanteau", as he terms it), but then becomes increasingly ludicrous, veering into elaborate and evidence-free etymologies that perhaps constitute a dig at the excesses of the historical linguistics of the era:

> "I see it now", Alice remarked thoughtfully: "and what are '*toves*'?"
> "Well, '*toves*' are something like badgers — they're something like lizards — and they're something like corkscrews."
> "They must be very curious creatures."
> "They are that", said Humpty Dumpty: "also they make their nests under sun-dials — also they live on cheese."

the word is still intact which allows for easy comparison to the appropriate, original word: e.g. 'bap' for 'map'.

b. *Semantic* (verbal) *paraphasias*: saying a word that is related to the target word in meaning or category; frequently observed in Wernicke's aphasia: e.g., 'jet' for 'airplane' or 'knife' for 'fork'.

(Wikipedia 2023: "Receptive aphasia")

If paraphasias involve the kinds of associative networks we have identified here, we would have further evidence for a cluster-based lexical architecture.

These two types of explanation are not incompatible. One accounts for the *formation* of root clusters via language acquisition, the other for the *persistence* of clusters across time as a means of facilitating lexical access. Together, they suggest that root clusters provide a clue to something rather fundamental about the architecture of the mental lexicon.

Is there a practical application for this kind of model? Again, in principle, yes. Orthodox dictionaries of Salish languages (and, in fact, of virtually all languages) are based on one of two sets of organizing principles: standard alphabetical order, or some kind of thematic approach (as in, e.g., the Classified Word Lists originally devised by Swadesh to elicit basic vocabulary for glottochronological reasons (Swadesh 1971), and which became standard for fieldworkers on Salish from the 1960s on. But suppose we were to design a dictionary instead as an associative network based on root clusters?

There would be practical challenges, not least because dictionaries in book form are organized linearly, but networks are not. However, electronic dictionaries are not subject to these kinds of constraints, and in principle could allow a cluster-based dictionary to be constructed where any root would be associated with multiple neighboring members of its root cluster, following the same principles of phonological and semantic distance which we have outlined above.

Of course, the question then arises as to whether such a dictionary would serve any practical purpose. Again, the answer in principle could be yes: if such a dictionary mimicked human cognitive architecture more closely, it would by hypothesis be easier to use — and enable learners to acquire new lexical items more easily. Again, this is testable, particularly via artificial language experiments: would an artificial language with a "lumpy" lexical distribution be more easily learnable than one with a "smooth" distribution, and would a teaching strategy for vocabulary that exploited clusters be more effective than a conventional one? We leave these questions open here.

Finally, let us return briefly to the issue of iconicity with which we began this paper. Recall that phonaesthesia involves a systematic pairing of sound and meaning which, as Grew (1998) put it "…has been given short shrift because of the heretical nature of any investigation pairing phonological parallels with semantic affinities". We do not think we are guilty of heresy here, however. The reason is that the heresy invoked by Grew is a rejection of the arbitrariness of the sign, as in iconic sound symbolism, which, involves a "natural" affinity between a linguistic sign and its referent. But as we have emphasized, phonaesthetic affinities are arbitrary: there is no reason why, for example, 'red' should be associated with the Proto-Salish root √**ca/i/q*^{w}, but with the Proto-Indo-European root *√*H₁reud*^{h}-. What is non-arbitrary is the semantic network that connects each root to its cluster members, which by hypothesis exploits a universal set of conceptual associations, and the principles of phonological resemble which enable those clusters to be linked via sound as well as via meaning.

References

Bloomfield, Leonard. 1914 [1983]. *An Introduction to the Study of Language.* Philadelphia, PA: John Benjamins.

Carroll, Lewis. 1882. *Through the Looking Glass: And What Alice Found There*. (Second Edition). London, UK: MacMillan.

Davis, Henry, Beverley Frank, Rose Whitley, & the Upper St'át'imc Language Authority. [in preparation]. *Nqwal'uttenlhkálha: An English-Upper St'át'imcets Dictionary.* Lillooet, BC & Vancouver, BC: The Upper St'át'imc Language, Culture and Education Society and PNWLL Press.

Firth, John R., 1930. *Speech*. (Reprinted in *The Tongues of Men & Speech*, 1964). London, UK: Oxford University Press.

Graham, Eleanor. 1981. Lewis Carroll and the Writing of *Through The Looking Glass*. (Introduction to *Through The Looking Glass*). In: *Alice's Adventures in Wonderland / Through The Looking Glass*: London, UK: Puffin Books.

Grew, Philip. 1998. Re: phonological clusters of semantically similar words. *Posted on Linguist List*, 05-08-1998, retrieved 19-01-2023 from https://linguistlist.org/issues/9/9-1106/

Hinton, Leanne, Johanna Nichols, & John Ohala. 1994. Introduction. In Leanne Hinton, Johanna Nichols, & John Ohala (eds.), *Sound Symbolism*: Cambridge: Cambridge University Press, 1–11.

Johansson, Niklas Erben, Andrey Anikin, Gerd Carling, & Arthur Holmer. 2020. The typology of sound symbolism: Defining macro-concepts via their semantic and phonetic features. *Linguistic Typology* 24(2).

Kuipers, Aert. 2002. *Salish Etymological Dictionary*. Missoula, MN: UMOPL.

Matthewson, Lisa. 1994. Syllable structure in St'át'imcets. *Proceedings of the Canadian Linguistics Association*.

Swadesh, Morris. 1971. *The Origin and Diversification of Language*. Chicago, IL: Aldine.

Thomason, Sarah. 1999. Crackle, pop, twinkle: Sound symbolic words in Montana Salish. *Papers for ICSNL* 29: 251–266. Kamloops, BC: Simon Fraser University.

van Eijk, Jan. 2013 [in prep]. *A Dictionary of the Lillooet Language.* Vancouver: UBCOPL and PNWLL Press.