

Application of the comparative method to vocoid sequences in Nivkh

Robert Halm*

Bloomington, IN, USA

The Nivkh language family of Sakhalin Island and the adjacent mainland in Northeast Asia is generally considered to be without known external relatives. Since its internal diversity is relatively shallow – leading some authors to treat it as a single ‘language’ divisible only into ‘dialect’-level varieties – comparative linguistics internal to the family has been neglected. The internal diversity of Nivkh is not, however, as trivial as has been portrayed, and involves at least two (Gruzdeva, 1998) and possibly three Fortescue (2016) mutually unintelligible varieties, indicating fertile ground for the application of the Standard Comparative Method within the family. In the present paper, the correspondences of vocoid sequences among six attested varieties are examined, allowing an important sound change affecting one major variety group (Proto-Nivkh /*a, *i, *u/ > Amur Nivkh, West Sakhalin Nivkh, and North Sakhalin Nivkh /ə/ when followed by a glide) to be reconstructed, as well as the applicable environment for this change to be precisely circumscribed, and furthermore allowing for an important phonological contrast for the proto-language to be reconstructed which is not well documented in the living varieties; namely, a contrast between sequences of vowel-glide and similar diphthongs, /*aw, *iw, *aj, *uj/ ≠ /*au, *iu, *ai, *ui/.

Keywords: Nivkh, Gilyak, comparative method, Proto-Nivkh

1. Introduction

The Nivkh family of languages is or was spoken across most of the Island of Sakhalin, along the facing coast of the mainland, and along the lower course and some lower tributaries of the Amur River in Northeast Asia (Gruzdeva, 1998). Although it has sometimes been referred to as a single language, divergence among the attested varieties is sufficient to nearly or entirely prohibit mutual comprehensibility between at least some different varieties; based on the criterion of mutual intelligibility, either two (Gruzdeva, 1998) or three (Fortescue, 2016) ‘language’-level subdivisions of the family are recognized. However a total of five or six identifiable varieties might be recognized at a level of differentiation sufficient to speak of separation by regular sound changes: Amur Nivkh (hereafter, AN), spoken on the mainland; West Sakhalin Nivkh (WSN) – historically spoken in the northwestern part of Sakhalin across the narrowest part of the Gulf of Tartary from the mainland (although Soviet-era resettlement policies have relocated most speakers of this and the other Sakhalin varieties) – is very close to AN, but recognized as distinct by at least Fortescue (2016) and Shiraishi (2007), the body of whose work constitutes the primary documentation of this variety; North Sakhalin Nivkh (NSN), historically spoken on the Schmidt Peninsula at the northern tip of the island; East Sakhalin Nivkh (ESN), historically spoken along the northeast coast of the island; and South Sakhalin Nivkh (SSN), historically spoken in isolated pockets in the south and southeast of Sakhalin, separated from the other varieties by areas of primarily Tungusic Uilta-speaking populations. Amur Nivkh and West Sakhalin Nivkh, forming a very tight cluster, may be collectively referred to as Western Nivkh (WN).

To this might be added the 21st century variety of the city of Nogliki (NgN), as documented in Tangiku et al. (2008); Nogliki is within the historical sprachraum of ESN, and the variety documented in that source shows the greatest similarity to earlier sources documenting ESN; however, it seems to also show some signs of koineization on at least the lexical level (including etyma traditionally noted as WN shibboleths) as well as possibly on the phonological level, showing the application of sound changes which were earlier

*Address correspondence to rthalm@gmail.com

documented as applying to WN and not to ESN, either uniformly (such as Proto-Nivkh /*#w/ > /#β/)¹ or in a lexically restricted distribution.

It has often been reported on an impressionistic basis that the greatest divide is between ESN and SSN on the one hand, and AN (and WSN, for those authorities who recognize this variety as distinct) on the other, with NSN taking an intermediate position. However, this impressionistic assessment has never been supported as a classificatory hypothesis in the strict sense with evidence from shared phonological innovations. Good grammatical and phonological or morphophonological descriptions of attested varieties include Mattissen (2003) and Nedjalkov & Otaina (2013) for AN, Shiraishi (2007) for WSN, and Gruzdeva (1998) for a juxtaposition of these subsystems across the family, especially between AN and ESN.

1.1. Note on presentation of data. In order to conserve space in the body of this paper, a table with the bulk of the forms which are compared as evidence has been included in the Appendix, organized by variety and source (columns) and cognacy (rows), with rows grouped together according to which Proto-Nivkh sound correspondence they evidently reflect.

For the sake of greater consistency and clarity, all Nivkh forms will be standardized to a single transcriptional scheme, although the transcriptions used in the sources vary. The (maximal) consonant inventory of all varieties is shown in the following table as it will be transcribed. WN and probably NSN differ in lacking a /w/ in syllable onset, and at least the major sources for these varieties also fail to distinguish /Vu/ from /Vw/ when another vowel does not immediately follow, although it is not perfectly clear whether this is a phonologically real merger or merely an orthographic shortcoming, and the same may apply to some sources for other varieties. Where this occurs, we will follow our sources in transcribing /u/ for the undifferentiated segment (usually Cyrillic **y**). SSN differs from this inventory by lacking a contrast between the voiced and the voiceless unaspirated plosive series in initial position; the SSN fortis series is transcribed as voiceless unaspirated, and the lenis as voiced, following our sources. Note that although /r/ and /ɾ/ are generally described as alveolar trills, either with or without some frication, they are considered fricatives in the sense that they alternate with the alveolar stops in the same way as the true fricatives at the bilabial, velar, and postvelar loci alternate with their homorganic stops in productive synchronic processes. Likewise, /ś/ and /ž/ are often described as alveolar or alveopalatal sibilants, with or without some frication, but are considered palatal in that they similarly alternate with the palatal stops. All varieties have a six-vowel inventory, which will be transcribed /a, e, i, o, u, ə/, although the phonetic realization of these may be closer to [æ, ɛ, ɪ, ɔ, ʊ, ɤ] in at least some varieties.

2. Previous comparative work

Previous authors have performed some internal reconstruction (notably, Austerlitz, 1956, 1982, 1983, 1990), and Janhunen (2016) has done some interesting reconstruction through the examination of loan words

¹The following orthographic conventions will be used throughout this paper: an asterisk is used to indicate reconstructions by the present author (e.g., /*form/); the degree symbol to indicate Fortescue's (2016) typical or canonical forms (/°form/); the double asterisk to indicate unattested or disallowed forms (/**form/); single tilde to represent predictable phonological or grammatical variation due to known processes (/form phorm/); double tilde to represent free, unpredictable, or inexplicable variation or doublets (/form ≈ βform/); hyphen to mark a synchronically productive or otherwise well-understood and uncontroversial morpheme boundary (/form-i/); double hyphen or equals to indicate a conjectural, perhaps purely etymological morpheme boundary (/*φ=orm/); square brackets to indicate an uncertain reconstruction or a doubtful transcription ([φ]orm/); parentheses to indicate that a form can or does appear both with and without the enclosed segment or segments (/([φ]orm/); the plus /+/, dollar sign /\$/ , and pound sign /#/ , respectively to indicate morpheme juncture, the syllable boundary, and pausal word boundary in specifying the environment of sound laws; and subscript 'W' and 'S', respectively, to indicate the weak or alternating versus the strong or invariant final fricatives or nasals (/form_W/ ≠ /form_S/); and the sign /θ_W/ is used to indicate a phonetically elided but morphophonemically detectable weak nasal. The sign /V/ is used to indicate any vowel; the sign /C/ to indicate any consonant; and the sign /W/ to indicate any glide. Outside of slashes, the equals sign is used to indicate that two forms are identical or non-contrastive, or exist in free variation (e.g., /φo:rm/ = /φoβrm/); a not-equals or slashed equals sign is used to indicate that two forms are contrastive and non-identical, /φorm/ ≠ /φoβrm/; the question-marked equals is used to indicate that the identity or contrastiveness of two forms is uncertain (e.g., /φorm/ =?= /φo:rm/); single chevron or shaftless arrow is used to indicate diachronic change (/*φorm_W/ > /φoθ_W/); bidirectional chevron or bidirectional shaftless arrow is used to indicate cognacy (WSN /φorm/ <> AN /φorm/); bidirectional chevron with a question mark is used to indicate uncertain cognacy (WSN /φorm/ <?> AN /φoβrm/); double chevron or double shaftless arrow is used to mark synchronic processes (/φorm_W-kun/ >> [φormgun]); and shafted arrow is used to indicate borrowings (Uilta /form/ → WSN /φorm/). In glosses, (v.) indicates a verb, (v.tr.) a transitive verb, (v.intr.) an intransitive verb, (n.) a noun, and the abbreviation 't.o.' to shorten 'type of.'

	Bilabial	Alveolar	Palatal	Velar	Post-Velar
Fortis Stop	p ^h	t ^h	c ^h	k ^h	q ^h
Lenis Stop	p	t	c	k	q
Voiced Stop (<i>non-contrastive in SSN</i>)	b	d	ɟ	g	ɢ
Nasal	m	n	ɲ	ŋ	
Voiceless Fricative	ɸ	ɾ̥	ç	x	χ
Voiced Fricative	β	r	ʒ	ɣ	ʁ
Approximant	w		j		
Lateral		l			
Voiceless Onset	← - - - - - h - - - - - →				

Table 1: Transcription of the (maximal) Nivkh consonant inventory.

shared with neighboring but presumably unrelated languages (what the author terms *external reconstruction*), but the only work published to date which takes a comparative approach to reconstructing Proto-Nivkh (hereafter PN) is Fortescue’s comparative dictionary for the family (2016). Fortescue, however, does not seem to use the Standard Comparative Method in the strictest sense of the term; while he very briefly notes sound correspondences within the family, he generally does not propose what proto-phoneme or sequence might underlie these correspondences diachronically. Nor (with a few exceptions) does he propose specific sound changes, and he generally does not identify specific conditioning environments for these changes. Although he provides a “Proto-Nivkh” form for each etymon and bound morpheme which he documents, the resulting sound correspondences between his proto-forms and the attested forms are often irregular,² and it would seem that these forms are better considered as typical or canonical forms, rather than reconstructed, diachronic proto-forms *sensu strictissimo*. This is very much in line with the thrust of that volume, which (as the title itself indicates) is comparative rather than reconstructive in bent; in this vein Fortescue is an excellent resource of the highest quality, and without it the present work, or any of a similar purpose, would be possible at best only with enormous difficulty and would be of much lower quality. The situation, then, is that no previous work strictly applying the Standard Comparative Method to the members of the Nivkh family in order to reconstruct Proto-Nivkh has been published.

3. Trivial and nontrivial sound correspondences among vocoid sequences

Sources for the Nivkh varieties uniformly indicate in transcription a distinction between a glide and a full vowel when followed by another vocoid; i.e., /wa/ ≠ /ua/, /ja/ ≠ /ia/ (and likewise for the other five full vowels in place of /a/). However, only some sources transcribe a distinction between a glide and a full vowel when preceded by a full vowel; i.e., either /aw/ = /au/ or /aw/ ≠ /au/, and either /aj/ = /ai/ or /aj/ ≠ /ai/, depending on the source (and again likewise for the other five full vowels in place of /a/).

3.1. Sequences of /\$WV/. Where the sequence of vocoids is limited to a glide in the onset, followed by a full vowel (excluding, apparently, glides in intervocalic position), the sound correspondences across varieties are trivial, with an identical vowel in essentially all cases of clear cognacy; the glide also remains unchanged except for the shift of PN /*w/ to WN and probably also NSN /β/ in the onset, which is already well known and well documented (Gruzdeva, 1998; Mattissen, 2003; Shiraishi, 2007; Fortescue, 2016). This appears to

²To take a few examples, PN /°r/ is reconstructed in many etyma where a majority or even all varieties attest /s/ or /z/, while in other etyma with the same patterns of attestation, he puts forward PN /°z/ instead, without any pattern of conditioning being mentioned or discernable. Similarly, while PN /°w/ in the syllable onset in his forms usually reflects the correspondence AN, ESN (NgN) /β/ <> NSN, SSN /w/, there are etyma showing exceptions to this reflex for all four varieties, without any obvious conditioning factors. Where a /ə/ in some varieties corresponds to a different vowel in one or more other varieties, the criteria for when the schwa versus the other vowel should be reconstructed do not seem to follow any particularly regular pattern, especially in the cases where a glide or a second vowel follows, in which case the presence or absence of the following vocoid in his protoform is likewise not evidently predictable. Irregularity also flows in the opposite chronological direction, with a single PN shape or sequence reconstructed behind differing patterns of modern reflexes. For instance, PN /°#Cəw-/ is reconstructed as yielding modern AN /Cəu/ <> ESN /Cəw/ ≈ Cuw/ in the etymon for ‘take off’, but AN, NSN, ESN /Cə/ <> SSN /Cu/ in the etymon for ‘swell’.

hold true for all ten possible combinations of glide and full vowel. Examples to support the analysis that intervocalic glides are treated as belonging to the coda of the preceding syllable rather than the onset of the following one in at least some and perhaps all cases (although this may be unusual cross-linguistically) will appear in our data below. The two exceptions to this of which the present author is aware (viz. AN, NSN, NgN /βaβu-/ <> ESN, SSN /wawu-/ ‘chew, bite’; AN /βuβu-/ <> ESN /βuβu-/ <> SSN /wuβu-/ ‘to hoot’, all Fortescue, 2016) appear to be onomatopoeic in nature, and may have been subject to irregular phonological development for this reason.

3.2. Sequences of /əj/ and /əw/. It must also be mentioned preliminarily that PN appears to have had a vowel phoneme /*ə/, as do all its daughters, and that sequences of /*əj/ appear retained in identical shape in all varieties, e.g. PN /*həjk/ ‘hare’; /*ḡəjrəḡ/ ‘right (side)’; /*təj- ≈ *t^həj-/ ‘blue, green’; /*t^həj-/ ‘dream, sleep’; and /*məjk/ ‘wooden idol’.

Sequences of /əw/, contrarily, are much less clearly attested. Although one example of this sequence is found, which is reflected in two varieties (PN /*mlə[ɣ]wo/ ‘abode of the dead’ > AN /mləβo/ <> NgN /mləβo ≈ mləɣβo/), this attestation has two strikes against it: first, that historically there may have been an intervening consonant, and second, that the sequence /əw/ clearly straddles a morpheme boundary, since this term is a certain compound of /mlə/ ‘wooden representation of the dead’ and /wo/ ‘village’ (also perhaps explaining how it escapes root-internal vowel harmony; see Shiraishi & Botma, 2012; Botma et al., 2015). It is unclear whether this absence of /əw/ is merely an incidental gap in the documentation (conceivable, especially since /ə/ seems to have been a rather rare phoneme in PN to begin with), or whether this sequence may have been banned in PN, or whether perhaps the reflexes of PN /*əw/ have merged with some other sequence in most or all varieties.

Regardless of the unclear status of PN /*əw/, the fact that PN /*əj/ > AN, WSN, NSN, NgN, ESN, SSN /əj/ is attested with a reasonable degree of robustness is important because it allows us to be certain that PN /*əj/ is not the antecedent of any of the sound correspondences which will be discussed below in which ESN and SSN show a full vowel other than /ə/, and the high degree of symmetry which seems to apply to the outcomes of /*j/ versus /*w/ in all environments suggests that the same can probably be said of /*əw/ (i.e., that it is not the antecedent of any of the sound correspondences discussed below).

4. AN, WSN, NSN /ə/ <> ESN, SSN /a, i, u/

The first nontrivial correspondences among vocoid sequences which will be discussed are those in which the vowel /ə/ in AN, WSN, and NSN corresponds to a vowel /a, i, u/ in NgN, ESN, and SSN (with the latter two regularly showing the same one of these vowels in any given etymon) when followed by another vocoid. These are taken as evidence of the posited historical merger of PN /*a, i, u/ > AN, WSN, NSN /ə/ immediately before a glide. Examples of this include the reflexes of PN /*ajɸ/ ‘always’; /*ajx-/ ‘to flow (of a river)’; /*caj/ ‘further, yet, else’; /*haj-/ ‘to notice, observe’; /*hajm(u)-/ ‘to be old, grow old’; /*hajm-/ ‘to know’; /*ḡaj ≈ *ḡaju/ ‘gulch, valley, stream, backwater, spring’; /*taj/ ‘tundra, plain, wasteland’; /*t^hajr[ə]ḡ/ ‘left (side)’; /*waj, *-waj/ ‘bottom, under, down’; /*k^hunu-jβ[u]-/ ‘to dawn, be about to dawn’; /*hujβ(V)-/ ‘to remember’; /*puj/ ‘to fly’; /*aw/ ‘voice’; /*ḡaw/ ‘inside, within’; /*ḡaw[ʒork]-/ ‘to sorrow, grieve, be sad’; /*taw(u)-/ ‘to accustom, teach, learn’; /*c^huw(u)-/ ‘to take off, throw off’; /*muw-/ ‘to spin, turn’; and /*ciw/ ‘sound (n.)’ (see Appendix for the attested reflexes of all of these forms, as well as those of the other examples cited in the present paper). In each of these examples, the vowel preceding the glide is reflected unchanged in NgN, ESN, and SSN, but is reflected instead as /ə/ in AN, WSN, and NSN. The only documented irregularities defying this posited sound change appear to be limited to three etyma.

First, the reflexes of /*hajm-/ ‘to know’, show irregular doublets AN /hajm-, him-/ and SSN /[h]əjm-/ beside the expected forms AN /[h]əjm-/ and SSN /hajm-/, and in NSN is reflected only by irregular /[h]im/. Secondly, the reflexes of /*hujβ-/ ‘to remember’ and /*cuw(u)-/ ‘to take off, throw off’, seem to show doublets resulting from internecine borrowing. In the case of /*hujβ-/, this borrowing appears to have been from ESN or SSN into AN, yielding AN /hujβi-, hujβ-/ beside expected /həjβi-/. Similar borrowing of the reflexes of /*cuw(u)-/ from WN into NgN and ESN seems to have yielded NgN /səwu-/ and ESN /səu-/ beside expected ESN /suwu-/. Some of these examples are shown in Table 2.

Gloss	Posited PN Root	AN Reflex	ESN Reflex
'always'	/*ajɸ/	/əjɸ/	/ajɸ/
'further, yet, else'	/*caj/	/cəj/	/caj/
'under, down'	/*waj/	/βəj/	/waj/
'to dawn'	/*k ^h unu-jβ[u]-/	/[q]unə-jβi-/	/*k ^h nu-jβu-/
'to fly'	/*puj-/	/pəj-/	/puj-/
'voice (n.)'	/*aw/	/əw/	/aw/
't.o. duck'	/*awɪq/	/əwɪk/	/awɪq/
'to teach, learn'	/*taw(u)-/	/rəw-/	/raw-, rawu-/
'to cast off, throw off'	/*c ^h uw(u)-/	/səw-/	/suwu-/
'to spin, turn, dance'	/*muw-/	/məw-/	/muw-/
'sound (n.)'	/*ciw/	/cəw/	/ciw/

Table 2: Examples of the correspondence AN, WSN, NSN /əW/ <> NgN, ESN, SSN /[a,i,u]W/.

While some of these etyma have sequences of /VWV/, it seems that all three segments are phonemic, and that the glide in this case is treated as a coda of the preceding vowel rather than the onset of the following one, at least for the purposes of this sound change. We can be relatively secure in the conclusion that these etyma reflect sequences of /*VW/, and not /*VV/; the reflexes of the latter will be treated below, and it will be found that they have developed differently. We may also note that /*aj/ and /*aw/ are fairly well attested (ten and four etyma each) while /*uj/, /*iw/, and /*uw/ are much rarer, appearing in three, two, and one etyma, respectively. This may be due simply to the fact that /a/ has a much higher lexical frequency than /i/ and /u/ in Nivkh – rough counts suggest that /a/ is at least half again more frequent than /u/, and more than twice as frequent as /i/. The PN sequence /*ij/ does not appear to be attested at all in the data available to the present author, which could either be because this sequence was actually banned, or, again, simply by chance due to the low frequency of /*i/ overall. Another possibility is that the sequence is in fact present in Nivkh, but has been mistranscribed as simply /i/, due to the perceptual difficulty in distinguishing /ij/, especially when the glide falls in the strict syllable coda, rather than intervocalically.³

4.1. PN /oW/, /eW/ retained in all varieties. Regarding the first vocoid in such sequences, it appears that NgN, ESN, SSN /a,i,u/ can correspond to AN, WSN, NSN /ə/, but that NgN, ESN, SSN /e,o/ cannot; the latter always correspond to an identical vowel in AN, WSN, and NSN, as attested in cognate sets such as the reflexes of PN /*hoj]/ 'loach?'; /*poj]/ 'be visible, appear'; /*c^hɔoj]/ 'to wash (of clothes, v.tr.)'; /*hoj]ɪ/ 'blackcurrant?'; /*ojd[a]m/ 'infant'; /*ɲoj/ 'penis'; /ɲoj(e)q/ 'egg'; /q^ho-[i]nə-/ 'be sleepy, about to sleep'; /*he[w]ɪ/ 'alder?'; /*ple[w]-/ 'stroll'; /*we[w]-/ 'be deep'; and /*c^he-u-/ 'to dry (v.tr.)'. Some of these etyma, it must be noted, are slightly problematic in one way or another (in particular, the etyma for 'penis' and 'egg' have become entangled over the shared sense of 'testicle' in at least NgN, and the cluster /*[w]ɪ/ in the etymon for 'alder' shows seemingly irregular reflexes). Also, Savel'eva & Taksami (1979) seem to fail to distinguish /i/ ≠ /j/ following /o/. But overall there seems to be sufficient material here to have confidence that PN /*oj/ and /*ew/ are maintained unchanged in all attested varieties. It may be tentatively assumed that PN did not feature tautomorphic sequences /eu/, /ej/, /ei/, /oi/, /ow/, or /ou/,⁴ since these seem to be unattested in all Nivkh varieties, and hence we may tentatively infer that they were not extant in PN.

Regarding the second vocoid in such sequences, sources do not generally distinguish in transcription between the full vowels /i, u/ and the glides /j, w/, with /i ≠ j/ in this position being distinguished only in SSN by Takahashi (1942, non vid., reported in Fortescue, 2016), and in AN and ESN by Savel'eva

³Thank you to an anonymous reviewer for pointing out this possibility.

⁴The only apparent exception in the data available to the present author is WSN /βow ≈ poβu/ 'chew' (Shiraishi, 2007), but it is not clear whether or not this is etymologically independent of the forms in /waw/ or /βaw/ of similar meaning in other varieties, perhaps having undergone an irregular sound change related to onomatopoeia. If it is independent, then it may indeed attest PN /*ow/ (although it would still be problematic that no cognates from other varieties are attested within our sources).

& Taksami (1965, 1979), and /u ≠ w/ being distinguished only in NSN by Peiros and Starostin (1986, non vid., reported in Fortescue, 2016), and in the sources for SSN. Based on the forms attested in the sources which do transcribe these distinctions, however, it appears that this correspondence of AN, WSN, NSN /ə/ <> NgN, ESN, SSN /a, i, u/ applies only when the second vocoid is a glide, and not when it is a full vowel.

4.2. PN /Vi*/, /**Vu*/ retained in all varieties (PN /**Vj*/ ≠ /**Vi*/; /**Vw*/ ≠ /**Vu*/).** Besides the forms just mentioned, we have many others in which a vowel followed by another vocoid seems to yield essentially identical sequences across all varieties (notwithstanding transcriptional variability in distinguishing between glides and full vowels in post-vocalic position), with the full vowel retained unaltered. Examples include the reflexes of PN /**ai*-/ ‘to do, do so, make, build’; /**ŋai*/ ‘hawk’; /**phai*-/ ‘to heal, correct, mend’; /**uiy*(ə)/ ‘sin, taboo’; /**uiyi*-/ ‘to disappear, vanish’; /**kuiβ*[aŋw]/ ‘ring’; /**kuiŋ*(i)/ ‘willow’; /**mu*-inə/ ‘to become sick, be about to die’; /**tau*-/ ‘to spend’; /**ŋau*[c]u-/ ‘to order, tell, inform’; /**ŋaur*(k)/ ‘brain’; /**hiu*-/ ‘to kill a bear’; /**kiu*/ ‘grass for lining shoes or boots’ (from /**ki*-uγ-/ |shoe-put.into|); /**k^hiu*/ ‘single cell of a net’; and /**k^hiu*[n-ŋ]/ ‘brother (brother of female ego? younger brother?)’. Although even within the sources which transcribe the distinctions /i/ ≠ /j/ or /u/ ≠ /w/ in postvocalic position it is not uniformly the case, the reflexes of these etyma are generally transcribed with the second vocoid in sequence as a full vowel, rather than a glide. By contrast, this is nearly never the case for the correspondences reported just above, in which AN, WSN, NSN /ə/ <> ESN, NgN, SSN /a, i, u/. Moreover, we can be certain that the second element is etymologically a vowel and not a glide in some of these etyma because they are transparent compounds (particularly /**mu*-inə/ ‘become sick’, lit. |die-INCEPTIVE|, and /**kiu*/ < /**ki*-uγ-/ ‘grass for stuffing shoes; to stuff shoes with grass’, lit. |shoe-put.into|). We can also be fairly confident that no other conditioning is at work to separate these reflexes from those given above with AN, WSN, NSN /ə/, because there are several minimal or near-minimal pairs, such as: PN /**taw*(u)-/ ‘to accustom, learn, teach’ versus PN /**tau*-/ ‘to spend’ (which in fact are documented perfect homophones in ESN); /**ajφ*/ ‘always’ and /**ajx*-/ ‘to flow’ beside /**ai*-/ ‘to do’; and /**ŋaw*-/ ‘inside’ with /**ŋaw*[zork]-/ ‘grieve’ on the one hand beside /**ŋaur*(k)/ ‘brain’ and /**ŋau*[c]u-/ ‘inform’ on the other. Hence, the protoforms of this latter group are reconstructed as containing sequences of /VV/, probably as tautosyllabic diphthongs, although the possibility of a disyllabic /V.V/ sequence cannot yet be conclusively ruled out.

One additional observation helps to support the hypothesis that these sequences should be interpreted as diphthongs specifically. A regular sound change in Western Nivkh is the raising of PN /**a*/ > AN, WSN /ə/ when this vowel is adjacent to or tautosyllabic with a velar of any manner. However this change has failed to act in several etyma of the second set (viz.: PN /**ŋai*/ > AN /ŋa[j]š-/; PN /**ŋau*[c]u-/ > AN /ŋaucu-ʃ/; and PN /**ŋaur*(k)/ > AN /ŋaur/; as well as possibly PN /**k^hau*-/ > WSN /xaw/ ‘to dry’, although this form appears to be entangled with a doublet with a postvelar initial consonant in place of the velar, making it unreliable). This failure is much easier to explain if PN /**ai*/ and /**au*/ are considered to be diphthongs, possibly best analyzed as each constituting a single phoneme, rather than as sequences of consecutive and separately syllabified vowels. Furthermore, whereas the PN sequence /**uw*/ seems to be attested, the absence of a diphthong /***uu*/, /***ii*/, or /***aa*/ is quite natural; a sequence of two identical vowels cannot form a diphthong, and, if required to be tautosyllabic, must instead take some other form such as a long monophthong. The absence of the sequences PN /***ia*/, /***ua*/ is expected, as these would violate Nivkh’s vowel harmony system, in which a low vowel /a, e, o/ cannot follow a high vowel /ə, i, u/ either directly or indirectly within the same root (Shiraishi & Botma, 2012; Botma et al., 2015).

5. Ambiguous cases

Some etyma are less easy to develop a secure Proto-Nivkh reconstruction for. These fall primarily into two classes: etyma in which the WN and NSN sound change /**a*/ > /ə/ has been blocked or would be blocked by the presence of a postvelar consonant, and probable post-Proto-Nivkh loans.

5.1. Postvelar blocking. There are certain phonological environments in which it is difficult to discern whether PN /**i*/ or /**j*/ should be reconstructed, or likewise in other cases whether /**u*/ or /**w*/. It has been repeatedly observed (e.g., Austerlitz, 1990; Gruzdeva, 1998; Shiraishi, 2007) that in all attested varieties of Nivkh, there is a prohibition against the co-occurrence of a postvelar consonant and a vowel of the high

set /ə,i,u/ when the consonant and the vowel would be both tautomorphemic and adjacent or tautosyllabic. This synchronic prohibition appears to have been maintained diachronically in part through the blocking of the sound shift of PN /*a/ > AN, WSN, NSN /ə/ before a glide when the vowel in question stands in the environment with a postvelar just described. We can infer this first of all by the apparent absence of sequences /əw/ or /əj/ when proximate to a postvelar consonant (or, for that matter, of any transcriptions of /əu/ or /əi/ in this environment) in any of our sources. There are, however (notwithstanding transcriptional uncertainty) several apparent incidences of /aw/ or /aj/ in AN, WSN, or NSN, in which the shift /*a/ > /ə/ appears to have been blocked by a postvelar consonant, such as the reflexes of PN /*a[w]ɨq/ 't.o. duck'; /*ɨalɨq/ 'puppy'; /*qa[w]-/ 'to gulp, swallow'; and /*q^ha[w]-/ 'to not be, be absent'. In the case of /*ɨalɨq/, /*a[w]ɨq/, and /*qa[w]-/, Hattori (as reported in Fortescue, 2016) transcribes a glide rather than a full vowel, which is a contrast he seems to recognize in this position, and the same seems to be true of /*qa[w]-/ and /*q^ha[w]-/, vis-à-vis Tangiku et al. (2008). Finally, the reflexes of /*q^ha[w]-/ are actually transcribed with a fricative /β/ apparently as the reflex of /*w/ in ESN according to Nakagawa, Sato and Saito, and in SSN according to Takahashi (both as reported in Fortescue, 2016) in the forms where the root is immediately followed by a segment /r/, which would be an unsurprising allophone (or possibly mistranscription) of /w/, but a less likely allophone of /u/. All of this suggests that these roots originally contained a sequence of /*VW/ rather than a sequence of /*VV/. However we must allow a degree of uncertainty here, since the evidence is less clear than for those etyma in which we can refer to the more reliably transcribed contrast /ə/ ≠ /a/ for our reconstruction of the PN form.

5.2. Probable loans. A few remaining items which seem to violate the regularity of the sound correspondences or the diachronic developments which have been hypothesized above are probable or certain loans, which, judging by their phonological irregularity, are presumably assignable to independent borrowing into one or more Nivkh varieties after the breakup of PN. Some of these include forms similar to WN /kəj/ 'sail (n.)', which seems to be a loan from Ainu /kaja/ 'sail' (Vovin, 1993, pp. 24, 100), which may in turn be a borrowing from Japanese /ko:kai/ 'to sail, go on a sail voyage (v.); sail voyage (n.)', and therefore ultimately from Chinese (and the same goes, of course, for its compound derivative WN /kəj-r̥e/ 'door of a tent', lit. 'sail-door'). As Shiraishi (2007, p. 97) has already observed, forms similar to WSN /taj/ 'tobacco pipe' are loans from Chinese, either directly or indirectly, and so too seem to be the forms similar to AN /p^haj/ 'playing card' ← Mandarin /p^haɿ/ 'playing card' or some related form. Also clear loans are the forms similar to AN /matauz/ 'twine', presumably from some Russian form similar to Belarusian /matuz/ 'string'; Ukrainian /motuzka/ 'string'; and Czech /motouz/ 'twine'.

Forms similar to AN /tui-/ 'dust; be dusty' belong to a very widespread family of wanderwords, represented by etyma in Turkic (e.g. Turkish /toz/ 'dust'; Tatar /tuzan/ 'dust'; Tuvan /dozuun/ 'dust'), Mongolic (e.g. Mongolian /toos/ 'dust'), Tungusic (e.g. Oroqen /t:rag/ 'dust'), and even in Uralic Udmurt (/tuzon/ 'dust') and Tocharian (Tocharian A /tor/ 'dust'; Tocharian B /taur/ 'dust') (Bauer & Pinault, 2003, p. 259). The attestation in Turkic especially is complete enough to allow reconstruction of Proto-Oghuz-Turkic /*toz/ and Proto-Turkic /*to:ɾ/, and the comparison of forms also allows us to suppose with some reasonable confidence that all the forms with a sibilant reflect either direct inheritances from Proto-Oghuz-Turkic or borrowings therefrom, while the forms with a rhotic represent earlier diffusion. Debate still exists as to whether this etymon originated in IE, and was borrowed into Proto-Turkic and thence more widely, or conversely originated in Turkic and was borrowed into both Tocharian and other language families; but in any case, it certainly cannot have originated in Nivkh. Unfortunately, the scarcity of lexical resources available to us covering the Tungusic languages makes it difficult to assess what the exact path of transmission to Nivkh might have been.

Another widely traveled loan may be the forms similar to SSN /ajɞaŋ/ 'wild boar'. It may be suspected that this is ultimately a loan from Arabic /ħajawa:n/ 'beast' which has passed through Turkic (e.g. Kazakh /xaywan/ 'animal'; Kyrgyz /ayban/ 'animal'; Tuvan /k^havan/ 'pig') acquiring the sense of 'pig' or 'boar' specifically in some lects, from which it has passed into Slavic (e.g. Russian /kaban/ 'wild boar'), and also by various paths into a plethora of other lects such as Malay, Armenian, Urdu, Punjabi, and Swahili. That such a wanderword could have reached Nivkh is confirmed beyond doubt by the parallel case of Nivkh /mem/ 'monkey, ape', which has a very similar global dispersal, originating with Arabic /majmu:n/ 'monkey', and could not, after all, be a native Nivkh etymon by virtue of its semantics, since non-human

primates have not lived on Sakhalin or in the Amur region at any point in human history or prehistory. The elision of the initial consonant /h/ or /x/ in the form for ‘boar’ is not cross-linguistically unusual, and is even attested in Kyrgyz. The shift of the medial consonant from /b/, /β/, /w/, or /v/ to the back fricative /ɸ/ or /χ/ attested in Nivkh may suggest that the path of transmission has passed through Tungusic, and specifically through Evenki (spoken along the Amur River), which shows alternation /w ≈ g ≈ θ/ in some forms, such as /togo ≈ toβo ≈ to:/ ‘fire’.

Fortescue (2016) helpfully identifies the forms similar to AN /xaulus/ ‘paper’ as loans from Tungusic, offering the comparandum of Ul’ch /xausuli/ ‘paper’. The present author has been unable to find any similar comparandum for the Nivkh forms similar to AN /taju-/ ‘to write’, but we may suspect on the basis of its semantics as well as its phonology that it too is a loan, since the Nivkh did not have a native writing system prior to the twentieth century.

6. Conclusion

Evidence has been provided in the argumentation above in Section 4 (with supporting data in the appendix below), showing that a synchronic sound correspondence exists between AN, WSN, NSN /ə/ and NgN, ESN, SSN /a, i, u/ when immediately followed by a glide, /w, j/. Since there also exists a correspondence AN, WSN, NSN /ə/ <> NgN, ESN, SSN /ə/ in the same environment (as shown in Section 3), but no regular correspondence of AN, WSN, NSN /a, i, u/ <> NgN, ESN, SSN /a, i, u/ in this environment is attested,⁵ we conclude that a historical merger PN /*a, i, u/ > AN, WSN, NSN /ə/ — /_|w,j|/ has occurred.

The scope of this merger is circumscribed by evidence that PN /*o, e/ were not subject to the same merger (Section 4.1), and that the merger only applied when the vowel directly preceded /*w/ or /*j/, and not /*u/ or /*i/, which contrasted with the glides postvocally, as shown in Section 4.2; this in itself is a notable finding, as this contrast is poorly attested in sources describing the living Nivkh varieties. Some cases which are problematic at first glance have been clarified in Section 5, through attribution to either borrowing from a non-Nivkh source after the date of this change, or to other phonological rules or changes.

It is hoped that, beyond making the case for this specific sound change, the present paper will serve to demonstrate that both sufficient data in the documentation and sufficient divergence in the Nivkh varieties themselves exist to permit the fruitful application of the comparative method for reconstructing Proto-Nivkh.

References

- Austerlitz, R. (1956). Gilyak nursery words. *Word*, 12(2), 260–279.
- Austerlitz, R. (1982). Gilyak internal reconstruction, 1: Seven etyma. *Folia Slavica*, 5, 81–88.
- Austerlitz, R. (1983). On the penumbra of questions surrounding the internal reconstruction of Gilyak. *Research on Language & Social Interaction*, 16(3–4), 75–87.
- Austerlitz, R. (1990). Typology in the service of internal reconstruction: Saxalin Nivx. *Language Typology 1987: Systematic Balance in Language*, 17–33.
- Bauer, B. L., & Pinault, G. J. (Eds.). (2003). *Language in time and space: A festschrift for Werner Winter on the occasion of his 80th birthday*. Berlin: Walter de Gruyter.
- Botma, B., Iosad, P., & Shiraishi, H. (2015). Phonetic (non-)explanation in historical phonology. Presented at *The Second Edinburgh Symposium on Historical Phonology*. Edinburgh: The University of Edinburgh.
- Fortescue, M. D. (2016). *Comparative Nivkh dictionary*. Munich: Lincom GmbH.
- Gruzdeva, E. (1998). *Nivkh*. Munich: Lincom GmbH.
- Janhunen, J. (2016). Reconstructio externa linguae ghiliacorum. *Studia Orientalia Electronica*, 117, 3–27.
- Mattissen, J. (2003). *Dependent-head synthesis in Nivkh: A contribution to a typology of polysynthesis*. Amsterdam: John Benjamins Publishing.
- Nedjalkov, V. P., & Otaina, G. A. (2013). *A syntax of the Nivkh language: The Amur dialect*. Amsterdam: John

⁵Except the regular correspondence of AN, WSN, NSN /a/ <> NgN, ESN, SSN /a/ where a consyllabic or adjacent postvelar consonant has blocked the shift, as discussed above in Section 5.1.

- Benjamins Publishing.
- Savel'eva, V., & Taksami, C. (1965). *Russko-nivkhsij slovar' [Russian-Nivkh dictionary]*. Moskva: Sovetskaja entsiklopedija.
- Savel'eva, V., & Taksami, C. (1970). *Russko-nivkhsij slovar' [Russian-Nivkh dictionary]*. Moskva: Sovetskaja entsiklopedija.
- Shiraishi, H. (2007). *Topics in Nivkh phonology*. Ph.D. dissertation, Rijksuniversiteit Groningen.
- Shiraishi, H., & Botma, B. (2012). Asymmetries and attractors in Nivkh vowel sequences. Presented at the *Workshop on Segmental Architecture, 20th Manchester Phonology Meeting*. Manchester.
- Tangiku, I., Tanzina, N. Y., & Nitkuk, N. V. (2008). *Basic vocabulary of the Sakhalin dialect of Nivkh language (Nogliki dialect)*. Inst. of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Vovin, A. (1993). *A reconstruction of Proto-Ainu*. Leiden: Brill.

gloss	PN (ours)	PN (Fort. 2016)	AN (Sav. & Taks. 1970)	AN (Fort. 2016)	WSN (Shir. 2007)	NSN (Peiros) NSN (Nak.)	Ngn (Tangku & Li. 2008)	ESN (Sav. 1970)	ESN (Nak.)	ESN (Taks.)	SSN (Takahashi)	SSN (Hatt)	SSN (Taks.)	SSN (Aust.)
Byma Refl. cng Proto-Nivk. /VW/ Sequences														
always	/ʔajp/		/ʔajp/	/ʔajp-ʔ/	/ʔak-c/ (Fort. 2016)	/ʔak-t/	/ʔajp/	/ʔajp-d/						
to flow (river)	/ʔajkʔ-/		/ʔajkʔ/	/ʔajkʔ-ʔ/										
further, yet, else	/ʔaj/		/ʔaj/											
note,	/ʔajp/		/ʔajp/											
observe	/ʔajm/		/ʔajm/	/ʔajm-ʔ/										
be, grow old	/ʔajm/		/ʔajm/	/ʔajm-ʔ/										
know	/ʔajmʔ/		/ʔajmʔ/	/ʔajmʔ-ʔ/	/ʔajmʔ/	/ʔajm-ʔ/		/ʔajm-d/	/ʔajm-d/	/ʔajm-d/	/ʔajm-d/	/ʔajm-d/	/ʔajm-d/	/ʔajm-d/
valley, ravine,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
backwater,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
spring	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
puppy	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
tundra	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
left(side)	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
under, down	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
today	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
remember	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to fly	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to fly	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to, duck	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
inside, within,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
girls,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
intestines	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
today	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
sorrow, grief	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
swallow, gulp	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
not be	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
teeth, beam	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
take off (throw)	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
off	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to spin, turn,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
off	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
dance	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to dance	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
Byma Refl. cng Proto-Nivk. /VW/ Sequences														
to, do so,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
build, make	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
bank	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
near, correct,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
fl.	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
sin, taboo	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
wash	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
ring	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
willow	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to become sick	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to spend	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
order, tell,	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
inform	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
brain	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
dry	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
to name, call	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
brake	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
kill a bear	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
grass for lining	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
shoes	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
single cell of a	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
net	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
younger	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
brother(?)	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
brother of	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										
female egg(?)	/ʔajpʔ/		/ʔajpʔ/	/ʔajpʔ-ʔ/										

gloss	PH (ours)	PH (Fort. 2016)	AN (Sav. & Taks. 1965)	AN (Sav. & Taks. 1970)	AN (Fort. 2016)	NSN (Shir. 2007)	NSN (Peiros)	NSN (Nak.)	NgP (Tangku Sad. 2008)	ESN (Sav. 1970)	ESN (Nak.)	ESN (Taks.)	SN (Takahashi)	SSN (Hatt)	SSN (Taks.)	SSN (Aust.)
Evya reflecting Proto-NWH /s/~/sequences																
light	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/	/hɔj/			/hɔj/							
right (side)	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/	/hɔj/			/hɔj/							
blue, green	/ʔʰɔj-ʔ/	/ʔʰɔj-ʔ/		/tɔ-ɔj-ʔ/ /tɔj/					/tɔj/							
dream, sleep	/ʔʰɔj-ʔ/	/ʔʰɔj-ʔ/		/ʔʰɔ-ʔʰɔj-ʔ/					/ʔʰɔj-ʔ/							
wooden bowl	/ʔhɔj/	/ʔhɔj/		/hɔj/					/hɔj/							
abode of the dead	/ʔhɔj-ʔ/	/ʔhɔj-ʔ/		/hɔj/					/hɔj/							
Evya reflecting Proto-NWH /ʔ/ or /ʔʰ/ followed by a vowel																
leach	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/	/hɔj-ʔ/ (Fort. 2016)			/hɔj-ʔ/							
be visible	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
apple	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
to wash (v.t.)	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
blackcurrant	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
infant	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
male reproductive organ, penis	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
sea	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
be sleepy, about to sleep	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
elder	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
to stroll	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
be steep	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
to dip (v.t.)	/ʔhɔj/	/ʔhɔj/		/hɔj/	/hɔj/				/hɔj-ʔ/							
possible post-Proto-NWH loans																
sail (n.)		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							
colored pipe		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							
playing card		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							
write		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							
wild boar		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							
paper		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							
dust, dusty		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							
vine		/ʔhɔj/		/hɔj/	/hɔj/				/hɔj/							

Figure 1: In the chart above, the sources are abbreviated for reasons of space. (Sav. & Taks. 1965) is (Savel'eva & Taksami, 1965), and (Sav. & Taks. 1970) is (Savel'eva & Taksami, 1979), but in general the latter is relied upon and only forms absent from (1970) are cited from (1965); (Shir. 2007) is (Shiraishi, 2007); (Peiros) is for data originally from Peiros & Starostin (1986, non vidimus) as reported in (Fortescue, 2016); (Nak.) is data originally from Nakagawa, Sato, & Saito (1993, non vidimus) as reported in (Fortescue, 2016); (Tangku & al. 2008) is (Tangku et al., 2008); (Taks.) is Taksami (1983, non vidimus) as reported in (Fortescue, 2016); (Takahashi) is Takahashi (1942, non vidimus) as reported in (Fortescue, 2016); (Hatt) is Hattori (1962a,b, non vidimus) as reported in (Fortescue, 2016); (Aust.) is any of the papers by Austerlitz, with the year and page given in the same chart cell with the form; and (Fort. 2016) is other data reported in (Fortescue, 2016), not originally taken from any of the original sources already mentioned.