Interpreting the spreadsheets:

* First column: tag (e.g. left side of colon in stem list)
* Second column: initial variant (“initial”)
* Third column: variant after person prefix (“person”)
* Fourth column: variant after everything else (“elsewhere”)
  + Sometimes, these will be blank
* Fifth column: next continuation lexicon
* Sixth column: flag diacritics required (not super common)
  + Flags are given in quotation marks because excel doesn’t like @ so the quotation marks will need to be deleted
* (Sixth column: other notes)

To create FST stems for the “initial” list”:

* [column 6][ column 1 ]:[column 6][column 2]\t[column 5] ;

To create FST stems for the “person” list:

* [column 6][ column 1 ]:[column 6][column 3]\t[column 5] ;

To create FST stems for the “elsewhere” list:

* [column 6][ column 1 ]:[column 6][column 4]\t[column 5] ;

e.g. lines 189-190 of the animate nouns

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| iihtáíhksoohpattsaakio'pa | iihtáíhksoohpattsaakio'p |  | ohtáíhksoohpattsaakio'p | NA | "@R.person.NULL@" |
| iihtáíhksoohpattsaakio'pa |  | %^SPomohtáíhksoohpattsaakihp | ohtáíhksoohpattsaakihp | NA | "@D.person.NULL@" |

* These will result in four entries in the stem lists, one on the initial list, one on the person list, and two on the elsewhere list
* Initial:

@R.person.NULL@iihtáíhksoohpattsaakio'pa:@R.person.NULL@iihtáíhksoohpattsaakio'p NA ;

* Person:

@D.person.NULL@iihtáíhksoohpattsaakio'pa:@D.person.NULL@%^SPomohtáíhksoohpattsaakihp NA ;

* Elsewhere:

@R.person.NULL@iihtáíhksoohpattsaakio'pa:@R.person.NULL@ohtáíhksoohpattsaakio'p NA ;

@D.person.NULL@iihtáíhksoohpattsaakio'pa:@D.person.NULL@ohtáíhksoohpattsaakihp NA ;

Adding new stems to FST spreadsheets:

* These will often be identical
* If the diagnostic forms for a particular lemma do not give enough information, try to find related entries (e.g. another verb with the same initial morpheme)
  + If there are multiple allomorphs for the same position, create two lines for that form
  + If no other examples exist, assume that the person and elsewhere forms are the same, and make a note in the notes column that an assumption was made
    - E.g. could be explored in fieldwork
* Only encode irregular allomorphy
  + E.g. sometimes I deletes initially, sometimes it doesn’t; sometimes a form has different vowels depending on whether it’s initial or not; encode these types of allomorphy on a stem by stem basis
  + Do not encode glide deletion, etc. – e.g. yaak ‘future’ is always yaak, even though it essentially always surfaces as aak initially and following person prefixes
  + Be consistent with special characters: e.g. i2 is only relevant in the “elsewhere” column, but encode throughout
    - Exception being triggers like %^SP, which can occur only in the person column

Tag:

* The lemma/stem form in the dictionary (for easy comparison)
* If more than one entry, choose the one that seems to have the most information associated with it
* For prestems, use PS/[lemma]+

Initial:

* The form the morpheme takes at the beginning of a word

Person:

* The form the morpheme takes immediately following a person prefix

Elsewhere:

* The form the morpheme takes following any other morpheme

Lexicon:

* The next continuation lexicon
  + Will usually be the same as others of the same category, except for nouns that require -im

Flag

* If any flag diacritics are needed, put them here with quotation marks around them

Special characters:

* i2 – breaking i
  + diagnostic forms: check future tense or prefix ending in k, if k>ks, i2
* m2, n2, s2 – non-permanent consonants at the end of noun stems
  + diagnostic forms: check the plural, if the consonant deletes, it is a non-permanent consonant
* G – underspecified glide, used for prestems ending in (w/y) in the online dictionary or . in the paper dictionary (this also occurs for the final -oo, but this is managed using the diphthongization trigger for simplicity)
* y2 – a weird one, currently only used in yamoohk, which becomes i2amoohk in certain contexts (so far only know as after t); this will hopefully change with more research and analysis

Triggers: (only those relevant to the stem lists, not those that occur only in affixes)

* %^SP – short prefix
  + Use when person prefix surfaces as n, k, w, m
  + Stem should always begin in a vowel (not the best interpretation of Blackfoot morphology, but efficient)
* %^AIM – aa + im > aim (rather than aam)
  + E.g. ninaa vs imitaa
* %^DIPH – diphthongization
  + For specific stems, especially those ending in the final -oo ‘go’
  + In the dictionary backend, these can be determined by searching for ~ in the admin notes
  + Sometimes, diphthongization is assumed, e.g. when a verb derived from verb for which this is documented, include a note like “diph cf [other verb]”
* %^3MM – 3mm verbs
  + Used when verbs surface with mm in the third person
  + Include the mm in the underlying forms, the rules delete rather than insert
* %^IC – initial change
  + Use when sV, ICV > sayV, CayV in initial position for past tense, imperative
* %^IMP – imperative it > (i)st
  + Use when initial it becomes (i)st in the imperative
  + (this trigger also occurs at the beginning of all imperatives in the lexc, to interact with IC or IMP as needed)

Katie’s system of assumptions/questions

* “assumed”
  + pers assumed (based on example for elsewhere) or vice versa, if both are assumed, the assumption is that the initial vowel holds across all three forms
  + for stems that begin in consonants that say i2 assumed, this is based on the frequency of i2 deleting initially
  + n2, s2 assumed: no plural example, but assumed based on the morphemes involved
* “unknown”
  + Could be i2, n2 but no evidence/related forms to check against, so not coded
* “cf [other entry]”
  + Allomorphy based on another stem, where form and meaning are clearly related
  + If followed by a question mark, the relationship between forms is less certain
* “VLV?”
  + Could be a case of predictable variable length in vowels, but needs more research and hasn’t been modelled yet; where examples clearly show length variation this is hard coded
* “short prefix?”
  + Uncertain if possession takes short prefix
* “?” (just a question mark)
  + No related forms to check against or examples to show, especially used for stems that start in consonants, so their non-initial forms could do a number of things