Mutual intelligibility of closely related languages

LOT-winterschool

Exercise Levenshtein-distance

In the following, you will find transcriptions of Dutch and German cognate words. The transcription system used is X-SAMPA, an alphabet which is equivalent to the IPA but uses only ASCII-symbols, so it's machine-readable. You can find an overview of the X-SAMPA-symbols and their IPA-equivalents on the IPA-sheet. In this data-set, diphthongs are symbolized by a combination of a vowel and _w, _j, or _6. Treat diphthongs as one segment in the following tasks.

Get together in small groups and perform the following tasks:

1.) **Align the word pairs.** At first, align vowels with vowels and consonants with consonants. Prefer matching sounds when aligning (e.g. prefer to align [a] with [a] in favor of [a] with [E]). Align [w] and [j] as well as [u] and [i] either with vowels or with consonants. Align [ə] (@) with any sonorant (i.e., any vowels, liquids, and nasals). Fill empty spaces (i.e., when a sound has to be inserted/deleted) with the symbol '0'.

Example:

```
Dutch       s  x  a: d  @
            |  |  |  |  |
German      S  0  a: d  n
```

2.) **Calculate the Levenshtein distance.** For each substitution and for each insertion/deletion, increase the distance by 1. Normalize the distance by dividing it with the number of aligned sounds.

Example

```
Dutch       s  x  a: d  @
            |  |  |  |  |
German      S  0  a: d  n
            1  1  0  0  1        = 3/5, distance of 60 %
```

3.) **Discuss** where you find the alignment or the calculation problematic. Also consider calculating in different ways, e.g. counting diphthongs as two segments or increasing the distance by only 0.5 if the difference depents solely on length ([u] vs. [u:]). Considering that we have to make decisions on such points when calculating Levenshtein distances, can we state which manner of calculation is the most objective or the best-suited for our task (i.e., relating linguistic distances to intelligibility)?

a) stul	Stu:l	sto:l
b) v\Ex	v\e:k	v\EC
c) pErspEktif	pE_6spEkti:v\@	pE_6spekti:f
d) a:nvurIN	anfy:RUN	anf2:rUnk
e) prosEs	pRotsEs	prOtsEs
f) kOntEkst	kOntEkst	kOntEkst

g) IndrYk	a_jndRUk	IndrUk
h) fiGyr	figu:_6	figu_6
i) Int@rnEt	Int6nEt	IntE_6nEt
j) ku	ku:	ko_w
k) bOrst	bRUst	bOst
l) EkspE:r	EkspE_6t@	EkspE_6t

m) v\Il	v\Il@	v\Iln
n) mikrofo:n	mikRofo:n	mikrOfO_wn
o) n2:s	na:z@	ne:s
p) vlYxt@lIN	flYCtlIN	flYCtlIN
q) sEntrYm	tsEntRUm	tsEntrUm