

Fragmented Languages Scale for WIELD

1. Summary

The table below provides a rough guide to the fragmented languages scale value (FLV) system.

Table 1. Summary of Fragmented Languages Scale

	PUBLICATIONS & THESES	UNPUBLISHED WRITTEN DATA	EXTANT AUDIO (& VIDEO)	SIZE & STRENGTH OF SPEECH COMMUNITY
VERY GOOD	A	4	4	4
GOOD	B	3	3	3
AVERAGE	C	2	2	2
POOR	D	1	1	1
VERY POOR	F	0	0	0

2. Detailed Explanation of the Fragmented Languages Scale

The fragmented languages scale is composed of four categories:

- (1) quantity & quality of theses and published resources
- (2) quantity & quality of known archived written material
- (3) quantity & quality of known audio (and video) documentation
- (4) number (and quality) of living speakers

Each category is assigned a value. For the first category, the scale is A B C D F (A = highest); for all other categories (those we typically cannot change), the scale is 4 3 2 1 0 (4 = highest). The first and last categories are to be separated from the second and third by dashes, thus A-44-4 would be the rating of English (highest value for all categories). A guide to the meanings of each value is given below.

With regard to category (1), on the quantity and quality of theses and published resources, the scale of A B C D F has the following values:

- A = [Quantity] one grammar (published or thesis), one dictionary (published or thesis), one published paper, transcribed text(s) (in publication or thesis)
 [Quality] accurate phonetic description and transcription throughout a majority of published works or theses; accurate glossing and morpheme parsing (as appropriate) throughout published works or theses; data drawn from all extant sources, where practicable (e.g. for small languages)
- B = [Quantity] same as A save one publication/thesis is missing (e.g. there exist a grammar, a dictionary, and a peer-reviewed journal article but none of these contain a glossed text)
 [Quality] same as A, though the publications might be of slightly lower quality.
- C = [Quantity] same as A save two publications/theses are missing (e.g. there exist a grammar, a dictionary, but there is not a peer-reviewed journal article and neither the grammar nor the dictionary contain a glossed text)
 [Quality] same as A, though the publications might be of slightly lower quality.
- D = [Quantity] Only one publication or thesis
 [Quality] Potentially the same as A, though the publications might be of slightly lower quality.
- F = Nothing published and not the sole focus of any thesis

With regard to category (2), on the quantity and quality of unpublished (non-thesis) materials, the scale of 0-4 has the following values:

- 4 = [Quantity] Roughly 1,000 pages of archived materials, including a mixture of data (elicited word lists/slip cards, transcribed texts, explanatory notes, parsed phrases, etc.)
 [Quality] From two or more researchers who worked with a combined total of two or more speakers, at least one male and one female; accurate

phonetic description and transcription by at least one scholar who worked with two or more speakers; accurate glossing and morpheme parsing (as appropriate) throughout

- 3 = [Quantity] Same as 4, but missing a substantial amount of critical information (e.g. connected speech in the form of a narrative text); might have examples of a wide variety of phenomena but have less than 500 pages and/or slip cards of extant data
[Quality] Same as 4, though it might be a slightly inferior quality; there should still be a significant amount of phonologically correct transcriptions
- 2 = [Quantity] More than 50 pages but fewer than 500 pages/slip cards of extant data; there should be enough data to make linguistic claims about the language on several parameters
[Quality] Data might not be in reliable transcription and might have significant gaps or, even if the data are extensive, they may come from only one source
- 1 = [Quantity] Under 50 pages of data
[Quality] Little (to none) of the extant data collected by linguists; transcriptions might be unreliable, and might come from too few researchers or too few speakers or both
- 0 = Nothing or very little (e.g. loose place names and one or two questionable vocabulary items, as in the Ais language of Florida or Pictish, etc.)

With regard to category (3), on the quantity and quality of known audio (and or video) documentation, the scale of 0-4 has the following values:

- 4 = [Quantity] Roughly 1+ hours of archived audio, including a mixture of data (elicited words & phrases, monologic and/or dialogic narratives) recorded from more than one speaker (at least one male and one female)
[Quality] Clear recording(s) stored on magnetic tape or digital formats. Sampling is at a rate of 41,000kHz (no less than 16-bit) for digital and

20kHz for magnetic tape media. Digital files should be stored in a non-compressed format. All recordings should be free of distracting ambient noise (animal noises, running water, etc.) and distracting noise from the recording device itself

- 3 = [Quantity] Recordings might run less than an hour but should be more than 30 minutes
[Quality] Audio should be of high-quality (broadly similar to the type expected for 4). There might be more ambient noise than would be the case for 4, and data might come from only one person; data might not be as useful for acoustic analysis in the same way as 4
- 2 = [Quantity] Less than an hour of audio (potentially less than 30 minutes)
[Quality] Audio might be of inferior quality, though sampling rate might be known; recording might be made by only one researcher or of only one speaker or both
- 1 = [Quantity] Any amount; quality is the deciding factor
[Quality] Audio is of objective inferior quality and sampling specifications are entirely unknown (or only extractable from an examination of the physical recording medium); little acoustic data (e.g. stress or pitch) can be extracted from the recordings, though individual lexical items can be identified.
- 0 = Nothing or very little (e.g. a single song with little to no obvious lexical content)

With regard to category (4), on the number (and quality) of living speakers, the scale of 0-4 has the following values:

- 4 = A major national language with millions of speakers (e.g. English, Spanish, Khmer, Korean, Amharic, etc.) or with fewer than a million speakers but no linguistic or territorial threat from another language (e.g. Icelandic, Samoan, etc.)

- 3 = A major language with hundreds of thousands to millions of speakers that is not the functioning language of a modern government. Speakers have little, nominal, or no political standing (e.g. Welsh, Hmong languages, Mon, Yucatec, Guarani, Quechua, Nahuatl, Afrikaans, etc.)
- 2 = A smaller language community with under 250,000 speakers but more than 1,000 speakers. Speakers have little political power and another larger language (either adjacent to it or surrounding it) is spreading into the community (e.g. Navajo, Breton, Yaqui, Greenlandic); language is possibly endangered
- 1 = A small language community with fewer than 25,000 speakers and few if any child speakers; clearly moribund (e.g. most North American languages)
- 0 = No living speakers

3. Examples

In Table 2, eight languages have been given a fragmented scale rating. These eight languages do not represent all of the permutations possible within the scale, obviously.

Table 2. Sample ratings for various languages

	PUBLICATIONS & THESES	UNPUBLISHED WRITTEN DATA	EXTANT AUDIO (& VIDEO)	SIZE & STRENGTH OF SPEECH COMMUNITY
Spanish	A	4	4	4
Yucatec	A	4	4	3
Southern Pomo	B	4	4	1*
NE Pomo	D	4	2	0
Purisimeño	F	3	0	0
Timucua	A	2	0	0
Karkin	D	1	0	0
Ais (a.k.a. Ays)	F	0	0	0

*This might now be 0; the sole living speaker was in his late 90s at the time this table was created.

4. Standard FLV Display for WIELD Projects

Each language project sponsored by WIELD will eventually include an FLV for the language in a display that combines the aforementioned alphanumeric values with a quick-reference color code. The colors which correspond with the alphanumeric values are as follows:

Table 3. FLV color code

COLOR					
ALPHABETIC VALUE	A	B	C	D	F
NUMERIC VALUE	4	3	2	1	0

As a quick reminder (or, for first-time visitors to WIELD’s site, a very rough guide), the basic meaning of each alphanumeric value in the FLV will be summarized above each cell. The target FLV (i.e. the maximum FLV WIELD currently believes is attainable) is also provided below the first cell within parentheses. An example FLV display for Northeastern Pomo is given below in Table 3.

Table 4. FLV for Northeastern Pomo

ONE PUBLICATION		GREAT MATERIALS	MODERATE AUDIO		NO SPEAKERS
D	-	4	2	-	0
(A-42-0)					