

## Online dictionaries for language revitalization

Andrew Garrett

Words and word choice are the most salient aspects of language for many people. Therefore, among all the various kinds of materials that may present information about endangered languages in support of language revitalization, dictionaries may be the most highly valued and frequently used in communities, by learners and teachers. Yet since lexicography is rarely taught in linguistics graduate programs, linguists who contribute to language revitalization programs typically receive no training in dictionary writing. Like language teaching itself, then, dictionary creation is a key part of language revitalization that linguists learn about only from experience. As we piece together what we know from the experiences we and our collaborators have had, we may be rediscovering what our colleagues and predecessors already knew. It does not help that the many fine books on lexicography are not oriented toward the special circumstances of endangered languages (ELs) and language revitalization.<sup>1</sup>

Generally speaking, dictionaries are of two broad types. Monolingual dictionaries like Merriam-Webster ([merriam-webster.com](http://merriam-webster.com)) or the Oxford English Dictionary ([oed.com](http://oed.com)) are for language users (typically fluent native speakers) who seek information about words they do not know, or meanings or usage patterns they seek to understand better. Bilingual dictionaries like the *Collins Robert French Unabridged Dictionary* (Atkins et al. 2006) are for second-language users, and help them translate or understand the second language (called the *source language*) in terms of their first language. EL dictionaries, including those used in language revitalization, are almost always bilingual dictionaries because of the relatively small number of EL speakers, and because people who are not fluent speakers may wish to learn the EL. Here I will use the term *matrix language* for what lexicographers call the *target language*, since it is usually the language of the broader society to which the EL community belongs (such as English, French, or Spanish in North America).

While every dictionary for language revitalization is inevitably also an EL dictionary, there is a crucial difference in goals.<sup>2</sup> As an aid to research scholars, or to help people read texts or a grammar, an EL dictionary is helpful if it merely glosses words, giving a sense of their meaning and usage or their translation in the matrix language.<sup>3</sup> The bar is higher for

---

<sup>1</sup> Good general books on lexicography include Atkins and Rundell (2008), Durkin (2016), Landau (2001), Svensén (2009), and Zgusta (1971); they all address the general issues discussed in this chapter. Papers in the *International Journal of Lexicography* also often address concerns of interest not only in general lexicography but for EL dictionaries.

<sup>2</sup> For fine discussions of EL dictionary writing, see Frawley et al. (2002), Haviland (2006), Kroskrity (2015), Mosel (2011), and Thieberger (2015).

<sup>3</sup> Obviously some areas of meaning require additional detail, such as scientific names for plant and animal terms. But greater semantic precision is often warranted in many other areas of meaning, such as deixis, direction and location, motion, and time among others.

language revitalization: a dictionary needs to help people *use* the EL. For instance, when several EL words have similar meanings or identical translations in the matrix language, a dictionary for language revitalization must help users understand the implications of their word choices. Which word is appropriate in which contexts of use? How do they differ in semantics, pragmatics, and syntax? Learners who rely on a dictionary that lacks detail about such parameters of usage may tend to combine EL vocabulary with syntactic and semantic patterns from the matrix language, yielding a "relexified" matrix language.<sup>4</sup>

My experience in this field comes mainly from four online dictionary projects, listed in (1), that involve Native languages of California. These languages are severely endangered: Hupa and Karuk have only a very few fluent traditional first-language speakers, Yurok has no active traditional first-language speakers, and while Northern Paiute has dozens of fluent speakers they are elderly and dispersed in communities of the Great Basin.

- (1) (a) **Yurok** (Algic): The Yurok Language Project (Garrett 2011) has integrated a text corpus, a lexical database (adapting the printed dictionaries of Robins 1958 and Conathan et al. 2005), and other information about the language and its context; see <http://linguistics.berkeley.edu/~yurok/>. The Yurok Language Project was begun in 2001 with Juliette Blevins and is now maintained at Berkeley with the collaboration of the Yurok Tribe's language program.
- (b) **Hupa** (Athabaskan): The *Hupa Online Dictionary and Texts* is maintained by Kayla Begay (Humboldt State University) and Justin Spence (UC Davis); see <http://nalc.ucdavis.edu/hupa/hupa-lexicon.html>. It originated as a way to put a community-oriented dictionary online (Golla 1996), and now incorporates a text corpus as well as other improvements.
- (c) **Northern Paiute** (Uto-Aztecan): The Northern Paiute Language Project is maintained by Maziar Toosarvandani (UC Santa Cruz) in collaboration with Northern Paiute community members. It includes a lexicon and text database; see <http://paiute.ucsc.edu/>.
- (d) **Karuk** (Hokan): *Ararahih'urípih* ("the people's language net") is a dictionary and text corpus of the Karuk language developed in collaboration with Susan Gehr, Line Mikkelsen, and others in the Karuk community and at Berkeley (Garrett et al. in preparation). It started as a way to put a dictionary online (Bright and Gehr 2005, partly based on Bright 1957), but now includes a substantial text corpus and other resources; see <http://linguistics.berkeley.edu/~karuk/index.php>.

All four projects in (1) originated (and two are still housed) at the University of California, Berkeley, under the sponsorship of the Survey of California and Other Indian Languages. Three were initially motivated by a wish to make print dictionaries accessible

---

<sup>4</sup> There is nothing wrong with this: all languages change all the time, including via language contact. Even setting aside prototypical creoles, many of the world's languages seem to have been shaped by relexification; examples include Angloromani (Hancock 1984) and Media Lengua (Muysken 1997). Yet some language learners or communities will choose to avoid this outcome insofar as possible.

online. Some examples discussed in this chapter will be drawn from work on the projects in (1).

Below, I will discuss problems that arise in the construction of dictionaries under four rubrics: *lemma* choice (section 1); variation in the *form* of words (section 2); the selection and use of example sentences (section 3); and word *meaning* and *usage* (section 4). In each section I will note problems for EL dictionaries used in language revitalization, and I will write about the particular circumstances of online dictionaries.

## 1. Lemma choice

Every dictionary is organized around headwords or lemmas, with closely related words (typically, inflectional forms) included in the entry for a given lemma. For example, in a dictionary of English, the words *sings*, *sang*, *sung*, and *singing* can all be found under the lemma *sing*. For English it is relatively easy to decide which related word-forms should be associated with a lemma (what linguists might call a *lexeme*) and which others should be given their own lemmas. In other languages, aspects of the morphology or morphosyntax may make this decision rather difficult. Different lexicographers then choose differently.

Thus, in Rhodes's (1993) compendious Ojibwa dictionary, related morphosyntactically distinct words are assigned to different lemmas. Given on page 3, for instance, are separate lemmas for the verbs in (2), each with its own set of inflected forms.

- (2) *aabnaabdamaadzod* "look back" (animate intransitive)  
*aabnaabid* "look back" (animate intransitive)  
*aabnaabmaad* "look back at s.o." (transitive animate)  
*aabnaabdamaadzod* "look back" (animate intransitive)  
*aabnaabdamaad* "look back for someone" (transitive animate)  
*aabnaabdang* "look back at s.t." (transitive inanimate)

On the other hand, in Beavert and Hargus's (2009) Sahaptin dictionary, a different approach is taken in which all derivatives of a root are assigned to the same lemma. The noun *p̄xwí* "thought" heads a lemma that contains example sentences and inflected forms as well as the sub-entry *p̄xwí-* "think, wonder", which itself has the sub-entries in (3), some with additional (third-level!) sub-entries (not shown here) and example sentences.

- (3) *páyu p̄xwí-* (v) "be sad, dejected"  
*pináp̄xwí-* (v) "wonder, think to oneself, think about oneself, be conceited"  
*p̄xwinútpa* (adv) "apparently, it seems, it looks like"  
*p̄xwít* (n) "feeling, emotion, opinion, thought"  
*p̄xwíp̄xwí* (n) "worry"

Both choices create difficulties for users. In the Ojibwa dictionary, it is not always easy to find closely related words unless they happen to be adjacent (via accidents of prefixation). In the Sahaptin dictionary it is not always easy to find words, since they are alphabetized according to the lemma's headword. An index like the thorough one provided by Beavert and Hargus reduces these problems, but users must remember to use the index. My experience has been that often, when users use an index, they may not check the main entry and may consequently miss details explained only there.

In an online dictionary, some of these difficulties can be minimized. In *Ararahih'urípih* (the Karuk dictionary and text corpus), for example, morphology is part of the representation of dictionary entries. A typical example is the verb *ikyiv* "to fall" (with conditioned allomorphs that include a final *-m* or lengthening of the second vowel). This has at least 20 derivatives, including *ikyíim-namnih* "to fall in" (with a suffix form *-namnih* "in") and *ikyíimnamnih-math* "to drop (something) in" (with a suffix *-math* "causative"). These three words would constitute separate lemmas in an Ojibwa-like dictionary, and would be listed under a single lemma in a Sahaptin-like dictionary.

In *Ararahih'urípih*, it is possible to show information in both styles. Figure 1 shows part of the display for *ikyiv* "to fall"; included is an optional list of all morphological derivatives, each of which is a link to the appropriate lemma. As seen near the bottom of Figure 1, sentence examples (from the *Ararahih'urípih* text corpus) can be set to include or exclude derivatives; adjusting this setting here would change the display from 6 to 58 sentence examples.

<p><b>Search Index</b></p> <p><a href="#">ikyiv / ikyim-</a> to fall; to be propelled through the air</p>	<p><b>Dictionary Entry</b></p> <p>lexicon ID #2375   revised Dec 14 2014</p> <p><b>ikyiv / ikyim-</b> • v • to fall; to be propelled through the air</p> <p><i>Derivatives (20; <a href="#">hide derivatives</a>)</i></p> <p><a href="#">ikyíima</a> "(a thrown or shot object) to reach (a certain distance)"</p> <p><a href="#">ikyíimka</a> "(of a disease) to strike a person"</p> <p><a href="#">ikyíimkar</a> "to fall into the river"</p> <p><a href="#">ikyíimkurih</a> "to fall into water"</p> <p><a href="#">ikyíimnam</a> "to fall into"</p> <p><a href="#">ikyíimnamnih</a> "to fall in"</p> <p><a href="#">ikyíimnamnihmath</a> "to drop something in"</p> <p><a href="#">ikyíimnav</a> "to fall into"</p> <p><a href="#">ikyíimnishuk</a> "to fall out"</p> <p><a href="#">ikyíimnupaa</a> "be propelled upward"</p> <p><a href="#">ikyíimnuprih</a> "to be propelled in through"</p> <p><a href="#">ikyíimshur</a> "to fall off"</p> <p><a href="#">ikyíimvar</a> "be propelled inward through something"</p> <p><a href="#">ikyíivúur</a> "to fall for a long time"</p> <p><a href="#">ikyíivish</a> "to fall down"</p> <p><a href="#">ikyíivivraa</a> "to fall over backward"</p> <p><a href="#">ikyíivivrath</a> "to fall into a sweathouse; (tossel in stick game) to fall over the goal-line"</p> <p><a href="#">ikyíivivruk</a> "to fall down over"</p> <p><a href="#">ikyíivunih</a> "to fall downward"</p> <p><a href="#">ikyíivuraa</a> "to be propelled up (into the air)"</p> <p>Source: WB 560, p.343</p> <p>Note: Cf. <i>ithyiv</i> (v/m) '(du.) to fall', <i>ivrara-</i> (pl.) to fall'.</p> <ul style="list-style-type: none"> <li>• <b>kári xás úkyiv, xás vúra xára póokyiv.</b> And he fell, and it was a long time that he fell. [Reference: WB T9. Coyote Goes to the Sky 055]</li> </ul> <hr/> <p><b>Sentence examples (6)</b></p> <p>Include derivatives: <a href="#">yes</a>   no</p> <p>Display mode: sentence   <a href="#">word</a>   <a href="#">word components</a></p> <ol style="list-style-type: none"> <li>1. <b>axmá yúra yítha ukyíivrath.</b> Suddenly one fell in.   Source: Nettie Ruben, "Coyote's Homecoming" (second telling) (WB_KL-02a)   <a href="#">read full text</a> Spoken by Nettie Ruben   <a href="#">Download</a>   <a href="#">Play</a></li> </ol>
---	--

Figure 1: *ikyiv* "to fall" in *Ararahih'urípih* (featuring derivative words)

## 2. Orthography, dialect, and pronunciation

Orthography in ELs can be contentious (Hinton 2014): there are community-internal disputes about multiple writing systems, disputes between linguists and communities about the best writing system for a language, and variant spellings of individual words in a

single coherent writing system (along the lines of English *honor* vs. *honour*). As Fishman (1972) put it, "the creation of writing systems is significant only insofar as it leads to the acceptance and implementation of writing systems." The best writing system, in other words, is one that people actually use. Acceptance may in some cases be an iterative process, but when a writing system is settled a useful dictionary will take it as given.

Where multiple spellings do remain (whether for individual words or for the language as a whole due to competing systems), words can be cited in different ways. Unfortunately, a print dictionary almost inevitably gives greater weight to the spelling chosen as the lemma headword (or for alphabetization). This is an area where online dictionaries can have a helpful leveling function. For example, different Northern Paiute communities use writing systems that differ in how vowel sounds are spelled. Because different systems are each unambiguous, automatic translation between them is feasible. The spelling used in the underlying database in the Northern Paiute Language Project allows users to choose a spelling system and see dictionary and text examples only in that writing system. Information can thus be presented without seeming to prioritize another orthography.

Similarly, an online dictionary in which the underlying database encodes information about dialect (or idiolect), it is possible to display only words or examples from the speech variety of interest to the user. The Northern Paiute Language Project again illustrates this feature, showing data from four California, Nevada, and Oregon dialects (Bridgeport, Burns, Mono Lake, Walker River) and allowing users to choose the dialect of interest to them. Such a resource can be of broad interest throughout a language area but of value as well to those whose goals are focused on only one dialect area.

In a print dictionary, it is possible and sometimes useful to flag pronunciation variants, but often the differences are not easily reducible to writing (without using representations whose phonetic detail would deter typical users). Audio can be useful for this purpose, and is virtually essential in any case for almost all language learners. A good EL dictionary may thus at least be created in connection with a broader revitalization program that includes personal contact with fluent speakers and other teachers. With some print dictionaries are distributed with CDs with recordings of words and sentences; an example is the Sahaptin lexicon (Beavert and Hargus 2009). But CDs are easy to lose, and in any case it requires work to associate specific audio files on a disk with entries in a print dictionary.

For obvious reasons, online dictionaries allow very effective integration of media files (including audio and video) with lexicographic information. This is extremely helpful for revitalization in EL contexts where learners may not often get to work with fluent first-language speakers (or, worse, where no first-language speakers remain). In the online Yurok dictionary, we have tried to include audio examples of as many words and short phrases as possible, spoken by as many fluent speakers as possible. Users report greatly appreciating a chance to hear the range of variation that would have been present in the speech community when Yurok was still used as a first language in many households. An example from the Yurok dictionary website is shown in Figure 2, showing part of the entry for *meweehl* "elk". Users can hear recordings as spoken by six fluent speakers recorded in the 2000s; by selecting other options they can see a photo, or examples of *meweehl* in the text corpus. (Note that the search index in Figure 2 is ordered by text frequency.)

## Search index

[ch'eeshah](#) dog  
[puuek](#) deer, deer meat  
[segep](#) coyote  
[megokw](#) dog, barking dog  
[chyer'erw'](#) black bear  
[neqeneech](#) mouse, bastard child  
[twegoñ](#) raccoon  
[meweehl](#)  
[kwegeruer'](#) pig, hog  
[nepe'weeshneg](#) otter  
[puesee](#) cat, house cat  
[wergers](#) fox  
['echkwoh](#) sea lion  
[muelah](#) horse  
[neekwech](#) grizzly bear  
[hevomues](#) skunk  
[keget](#) panther, puma, mountain lion  
[pemey](#) grease, cooking fat, oil, blubber, lard, liquid fat  
[chekas](#) donkey  
[hikeleekra'](#) shrew, mole  
[herk'werh'](#) cottontail rabbit, bunny rabbit  
[skovom](#) bat (animal)  
['wahchehl](#) civet cat  
[chmuuek](#) wild cat, bobcat  
[le'go'](#) fisher  
[s'echoh](#) horn, horns, wedge  
[tegeen'](#) porcupine, porcupine quill, yellow moss for dying basket materials, anything yellow, canary  
[chkweges](#) hair seal  
[hegovek](#) pine squirrel, flying squirrel  
[hekwsa'](#) whale  
[karkehl](#) ringtail, ring-tailed cat  
[kegor](#) porpoise  
[kepcheeesh](#) fawn  
[kwechovekw](#) ground squirrel  
[meqeseek](#) weasel  
[merueh chego'onep'](#) five-point buck  
[muesmues](#) cow, bull, cattle

## Dictionary entry

**meweehl** • *n* • elk (*Cervus elaphus*)

Lexicon record # 1759 | Source references: [WQ\(WEM264\)](#) [R222](#) [JE48](#)  
Semantic domain: [hoofed mammals](#)

Derived phrases or compounds

- [Meweehl 'O Legehl](#) "Cann's Prairie (on the ridge of the Bald Hills above the mouth of Redwood Creek)"
- [meweehl 'ue-segon](#) "canvasback"
- [meweehl 'we-chpegar'](#) "fat solomon"
- [meweehl 'we-s'ech](#) "elk horn"

Short recordings (6) | [Sentence examples \(11\)](#) | [Photo](#)

1. [meweehl](#) "elk" (spoken by Aileen Figueroa)  
 | [Download](#)
2. [meweehl](#) "elk" (spoken by Archie Thompson)  
 | [Download](#)
3. [meweehl](#) "elk" (spoken by Georgiana Trull)  
 | [Download](#)
4. [meweehl](#) "elk" (spoken by Jimmie James)  
 | [Download](#)
5. [meweehl](#) "elk" (spoken by Jessie Van Pelt)  
 | [Download](#)
6. [meweehl](#) "elk" (spoken by Glenn Moore)  
 | [Download](#)

Figure 2: *meweehl* "elk" in the Yurok Language Project website (featuring recordings)

## 3. Examples

It is uncontroversial that a thorough dictionary presents example sentences to show how words are used. These are typically either culled from a corpus or created in work with fluent speakers. In some dictionaries, partly for reasons of space, the practice is to include only one example for most lemmas. The dictionaries of Bright and Gehr (2005) for Karuk and of Beavert and Hargus (2009) for Sahaptin have this profile. In such a dictionary the choice of illustrative sentences can make a big difference for language learners. To choose an example nearly at random, the Sahaptin adjective *p'ísláni* "dented, flattened" is illustrated with a sentence translated as "Don't buy a dented car." This is a good choice, showing a specific noun that the adjective can modify; a word like "flattened" can refer to many other situations, and learners might not think that it could apply to dents in metal. Had *p'ísláni* been illustrated with a sentence translated simply as "it is dented", language learners would be no wiser.

In some dictionaries, there is space to give a large number of examples for some (or even many) lemmas. This benefits language learners by way of showing them the range of contexts in which words can be used. For example, in the nearly 1000-page Klallam dictionary of Montler (2012), the lemma *lúyəs* "abandon" includes examples with the translations in (4), among many others. This gives users a sense of the range of contexts in which the verb can be used. Assiduous researchers can classify examples by meaning or verify whether an intended sense is indeed documented for this verb.

- (4) "Leave it! Give it up!"  
 "I left him."  
 "I left Port Angeles."  
 "She got out of what she had been in."  
 "I left my dog in Port Angeles."  
 "His wife left him."  
 "I'm going to leave my coffee pot behind."  
 "I left the beaver that was crying."  
 "Then I left him and walked to Pysht."

An even more complete presentation of examples is possible in an online dictionary linked to a text corpus. For example, in the Karuk and Yurok websites, each of which includes a lexicon linked to a lexically tagged corpus (with thousands of words), many dictionary searches yield substantial sets of text examples. Shown in Figure 3 is a part of the *Ararahih'urípih* entry for the Karuk verb *mah* "to see, to find"; the text corpus has 163 examples (or 179 if derivative words are included). In the Karuk and Yurok communities, language learners and teachers make use of text examples to fill in gaps in lexicographic explication — for example, as often, where semantic information or lexical cooccurrence patterns are not well described.

#### Search Index

1-50 of 2965 result(s)  
 Previous 50 | [Next 50](#)

**piip** (variant **ipi**) to say  
**xus / xu-** to think, to know, to realize; to feel  
**av / am-** to eat  
**mah** to see; to find  
**ikyav / ikyáa-** to make, make into; to do; to fix; to prepare; to gather, to acquire (things other than food)  
**ikrii / ikriv-** to live, sit, stay, be; to be at home  
**ipeer / ipeen-** to say to, to tell; (rarely) to call  
**uum / uumu-** to arrive (there); to go (to a place)  
**kúupha / kúuphi-** (variant **kupi**) to do  
**váaram / váaramu-** to go, to go away, to leave  
**áapunma / aapúnmu-** to know, to know about (a person or thing), to find out  
**áhoo** to go, walk, travel, arrive (here)  
**pakúriihva / pakuríihvu-** to sing songs  
**thítiv / thítim-** to hear  
**ipak** to come back, to arrive (here) again, to return  
**imus- / imus-** to look at  
**if** to grow; (person) to grow up; (water) to boil; (bread) to rise  
**ihyárih** (animate) to stand, be standing  
**iykar / iykára-** to beat; to kill; to catch (fish)  
**tháaniv / tháanim-** (inanimate object) to sit, be; (animal or person) to lie, to be dead  
**imnish** to cook  
**ish** to drink  
**vik** to weave (esp. basketry; more specifically, to twine with two strands)  
**xúriha / xúrihi-** to be hungry  
**ikvít-ha / ikvít-hi-** to sleep; to fall asleep; to be sleepy  
 ..

#### Dictionary Entry

lexicon ID #4155 | revised Aug 22 2005

**mah** • v • to see; to find

#### Derivatives (4)

**fatamáhaan** "lucky person"  
**ipmah** "to see again"  
**ishkítmáhaan** "lucky person"  
**ishkítmahar** "lucky"

Source: WB 956, p.366

- **vaa páy pishiichyáach yáas nimáhat.** That's the very first one I ever saw. [Reference: KV]
- **kíri nu'áapunma hōoy kunmáahti.** Let's find out where they were finding it (the salmon). [Reference: DeA&F 1 Salmon 004]

#### Sentence examples (163)

Include derivatives: [yes](#) | no

Display mode: sentence | [word](#) | [word components](#)

1. **xas yítha pamuxuunak ifuni umah.**  
 Then one of the boys found a hair in his mush.  
 | Source: Margaret Harrie, "The Ten Young Men who Became the Pleiades" (DAF\_KT-02) | [read full text](#)
2. **yaas'ára iin kinmáahtiheesh."**  
 And the people who live will see them."  
 | Source: Margaret Harrie, "The Ten Young Men who Became the Pleiades" (DAF\_KT-02) | [read full text](#)
3. **xas kaan nimáhat pi'ēep va'iripraam.**  
 There I saw an old mine.  
 | Source: Benonie Harrie, "How I Found Gold" (DAF\_KT-05b) | [read full text](#)
4. **kāan xas kēechas péeshpuk cháavura áxxak nimma.**  
 And right there I found two big gold nuggets.  
 | Source: Benonie Harrie, "How I Found Gold" (DAF\_KT-05b) | [read full text](#)
5. **karu káakum kumatupichas-háyaachas xakinivki'itráhyar chávura nimma koovura.**  
 Then there were a lot of smaller ones, so that in the end I had seventy dollars in all.  
 | Source: Benonie Harrie, "How I Found Gold" (DAF\_KT-05b) | [read full text](#)

Figure 3: *mah* "to see" in *Ararahih'urípih* (featuring sentence examples)

## 4. Meaning and usage

The most important defects of EL dictionaries have to do with meaning and usage. These stem from inherent limitations in EL research: for any one language relatively few people work on lexicographic questions; and text corpora in which answers can be found are far more limited than for global languages like English or French.

### 4.1. Translation vs. meaning

In bilingual dictionaries, including EL dictionaries used for language revitalization, a word in one language is defined or glossed in terms of another language. In EL dictionaries, EL vocabulary is defined or glossed in the matrix language. It is extremely common for the glosses to be short (even single-word) translations. This is the single biggest problem in EL lexicography for language revitalization; it may unwittingly encourage learners to assume a simple one-to-one mapping between matrix-language and EL vocabulary.

Examples of unsatisfactory translation renderings are all too easy to find in almost any EL dictionary. For example, the Yurok dictionary shows that the most common verb whose meaning is translated "run" is *ro'opek'* "I run". The lexicon itself simply glosses the verb as "run" and does not say whether it is used figuratively for fish, water, machines, or in other such contexts.<sup>5</sup> Likewise, verbs whose approximate meanings are "lie", "sit", and "stand" are simply glossed as such, with no information about the range of contexts in which they may be used — even though there is considerable cross-linguistic variation in such verbs' denotations (Ameka and Levinson 2007, Newman 2002). For example, does a spider or a thumbtack "sit" on the wall? Does it "hang"? Simple questions like these, whose answers are essential for idiomatic language use, are hard to answer in almost all EL dictionaries.

Relatedly, it is uncommon that EL dictionaries explicitly distinguish subsenses in their description of meaning. Compare (admittedly unfairly) the Oxford English Dictionary and the Karuk and Yurok dictionaries in their treatment of "put", as given in (5).

- (5)
- a. English: *put* has 29 senses identified in the Oxford English Dictionary as currently in use (not counting those deemed obsolete, rare, or dialectal)
  - b. Karuk: *thar-* "to put, to lay" (thus translated; modified by directional suffixes but otherwise no other senses or sub-senses explicitly identified)
  - b. Yurok: *nekek'* "I put" (no other senses or sub-senses explicitly identified)

Perhaps English *put* is unusually complex in its meaning, but the general trend of most EL dictionaries is to give basic translation equivalents rather than semantic paraphrases, let alone an inventory of subsenses. In the absence of rich lexicographic information about semantics, language learners may tend to fill in the details with transfer effects from their first languages, enhancing any trend toward matrix-language relexification.

---

<sup>5</sup> From the text corpus displayed with the lemma, it emerges that *ro'opek'* has been used for fish and machines, but there is no evidence for other contexts. The point in any case is that the dictionary itself is silent.

## 4.2. Multiword units (MWUs)

Multiword units — phrases with noncompositional semantics — pose a problem for any dictionary. The details differ from language to language according to syntax. In English, for example, any dictionary must find a way to explain the semantics and usage of phrasal verbs like *put aside* "disregard", *put down* "anesthetize", *put off* "postpone", *put on* "deceive", and the like. In the Oxford English Dictionary, such MWUs are included in the lemmas for their verbs. Another example is constructions with *light verbs* (Jespersen 1942) like *take (a picture, a test, a shower, a walk)*; these are again associated with verbs in the Oxford English Dictionary.

As long as dictionary users know where to look (under *put*, not *down*; under *take*, not *picture*), they can find the information they need if it is somehow included in a lexicon. Many EL dictionaries do not include MWU information, however, perhaps because the patterns can require a larger corpus to identify. If MWUs are included in a dictionary, online presentation has the benefit of permitting double listing without expanding the physical size of a book. Information about an MWU can be linked with all its component words so that it appears in each lemma display. In Yurok, for example, the preposition and preverb *mehl* can be translated generally as "about, because of, by, for, from, of, with", but the actual interpretation in any context is dependent on the verb it is construed with. Some specific collocations are grouped in (6) according to broad semantic patterns.

- (6)
- a. *kooychkwook* "buy (it)" + *mehl* = "buy (it) for (someone)"  
*muehlkochek* "sell" (intr.) + *mehl* = "sell (something)"  
*nee'nowok* "watch" + *mehl* = "watch for (something or someone)"  
*pewomek* "cook" (intr.) + *mehl* = "cook for (someone)"
  - b. *hlmeyowok* "be angry" + *mehl* = "be angry about (something)"  
*kemolochek* "be jealous" + *mehl* = "be jealous of (someone)"  
*pelomoyek* "fight" + *mehl* = "fight about (something)"  
*soosek* "think" + *mehl* = "think about (something)"  
*tohkow* "talk" (collective) + *mehl* = "talk about (something)"  
*t'p'ohlkwek* "be sensible" + *mehl* = "understand (something)"
  - c. *'e'gah* "eat, have a meal" (collective intr.) + *mehl* = "eat (something)"  
*hegook* "walk" + *mehl* = "walk with (an instrument)"  
*hohkuemek* "make (it)" + *mehl* = "make (it) from (some material)"  
*myop* "it's crowded" + *mehl* = "it's crowded with (something)"  
*reeek* "be full" + *mehl* = "be full of (something)"  
*ruerowok* "sing" (intransitive) + *mehl* = "sing (some song)"

For each verb, a dictionary should describe the meaning and syntax of the construction with *mehl*. In an online dictionary, these patterns can easily be presented twice, once for the verb and once in the *mehl* lemma. A single database entry for each MWU will thus be drawn on as needed for online display.

## 5. Conclusion

I have looked at a few elements found in dictionaries, identifying a few challenges associated with EL lexicography. These flow from the fact that less is known about

underdocumented languages, for which, it should go without saying, there is usually also no lexicographic tradition.<sup>6</sup> The absence of large text corpora makes rich documentation of semantics and MWU patterns more difficult; sense discrimination and other work on facets of meaning (e.g. hyponymy) also requires time that is often absent in the under-resourced context of EL work. Yet these aspects of lexicographic knowledge are crucial in dictionaries that serve the goal of language revitalization: learners who do not have access (whether via fluent speakers or written sources) to lexically specific patterns that are described in a good dictionary may fill in the gaps with matrix-language patterns.

Online dictionaries in particular pose significant problems. They require internet access, by definition, smartphones or computer access for users, and technological infrastructure for those who create and host them. Even where they are socially and technically feasible, they are no panacea (they cannot replace actual language use), but they can make it easier to present some important lexicographic information. Relational information in particular — morphological and semantic relationships, examples in text corpora, and the like — are well suited to presentation online. If information in a database has been tagged suitably and an effective interface designed, users can use links and tags to explore a structured lexicon with far less effort than it would take to move between sections of a print book, or among a dictionary, grammar, and texts. In this sense a well-designed online dictionary may assist in the project of repatriating indigenous knowledge from the academy and the archive, into the communities that created it.

---

<sup>6</sup> A notable exception has to do with languages belonging to larger language families, like Algonquian or Athabaskan, that have considerable grammatical uniformity and therefore have developed a lexicographic tradition. But other larger families, like Uto-Aztecan, the grammatical differences across sub-branches make a shared lexicographic tradition less plausible.

## References

- Ameka, Felix K., and Stephen C. Levinson. 2007. The typology and semantics of locative predicates: Posturals, positionals, and other beasts. *Linguistics* 45: 847-871.
- Atkins, B. T., et al. 2006. *Collins Robert French dictionary*. 8th edn. New York: HarperCollins.
- Atkins, B. T., and Michael Rundell. 2008. *The Oxford guide to practical lexicography*. New York: Oxford University Press.
- Beavert, Virginia, and Sharon Hargus. 2009. *Ichishkíin Sinwit Yakama / Yakima Sahaptin dictionary*. Seattle: University of Washington Press.
- Bright, William. 1957. *The Karok language*. Berkeley: University of California Press.
- Bright, William, and Susan Gehr. 2005. *Karuk dictionary*. Happy Camp, Calif.: Karuk Tribe of California.
- Conathan, Lisa, Andrew Garrett, and Juliette Blevins. 2005. *Preliminary Yurok dictionary*. Berkeley: Department of Linguistics, University of California, Berkeley.
- Durkin, Philip, ed. 2016. *The Oxford handbook of lexicography*. Oxford: Oxford University Press.
- Fishman, Joshua A. 1971. The uses of sociolinguistics. *Applications of linguistics: Selected papers of the Second International Congress of Applied Linguistics*, ed. by G. E. Perren and J. L. M. Trim, pp. 19-40. Cambridge: Cambridge University Press.
- Frawley, William, Kenneth C. Hill, and Pamela Munro, eds. 2002. *Making dictionaries: Preserving indigenous languages of the Americas*. Berkeley: University of California Press.
- Garrett, Andrew. 2011. An online dictionary with texts and pedagogical tools: The Yurok Language Project at Berkeley. *International Journal of Lexicography* 24: 405-419.
- Garrett, Andrew, Susan Gehr, Line Mikkelsen, Erik H. Maier, and Clare Sandy. In preparation. *Ararahih'urípih: An online Karuk language resource for learners, scholars, and teachers*.
- Golla, Victor. 1996. *Hupa language dictionary*. 2nd edn. [Arcata, Calif.: Center for Indian Community Development.]
- Hancock, Ian. 1984. The social and linguistic development of Angloromani. *Romani rokkeripen to-divvus: The English Romani dialect and its contemporary social, educational and linguistic standing*, ed. by Thomas A. Acton and Donald Kenrick, pp. 89-122. London: Romanestan Publications.
- Haviland, John B. 2006. Documenting lexical knowledge. *Essentials of language documentation*, ed. by Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, pp. 129-162. Berlin: Mouton de Gruyter.
- Hinton, Leanne. 2014. Orthography wars. *Developing orthographies for unwritten languages*, ed. by Michael Cahill and Keren Rice, pp. 139-168. Dallas, Texas: SIL International.
- Jespersen, Otto. *A modern English grammar on historical principles*, part 6: *Morphology*. Heidelberg: Carl Winter.

- Kroskrity, Paul V. 2015. Designing a dictionary for an endangered language community: Lexicographical deliberations, language ideological clarifications. *Language Documentation and Conservation* 9: 140–157.
- Landau, Sidney I. 2001. *Dictionaries: The art and craft of lexicography*. 2nd edn. Cambridge: Cambridge University Press.
- Montler, Timothy. 2012. *Klallam dictionary*. Seattle: University of Washington Press.
- Mosel, Ulrike. 2011. Lexicography in endangered language communities. *The Handbook of endangered languages*, ed. by Peter Austin and Julia Sallabank, pp. 337-353. Cambridge: Cambridge University Press.
- Muysken, Pieter. 1997. Media Lengua. *Contact languages: A wider perspective*, ed. by Sarah G. Thomason, pp. 365-426. Amsterdam: John Benjamins.
- Newman, John, ed. 2002. *The linguistics of sitting, standing, and lying*. Amsterdam: Benjamins.
- Rhodes, Richard A. 1993. *Eastern Ojibwa-Chippewa-Ottawa dictionary*. Berlin: Mouton de Gruyter.
- Robins, R. H. 1958. *The Yurok language: grammar, texts, lexicon*. Berkeley: University of California Press.
- Svensén, Bo. 2009. *A handbook of lexicography: The theory and practice of dictionary making*. Cambridge: Cambridge University Press.
- Thieberger, Nicholas. 2015. The lexicography of indigenous languages in Australia and the Pacific. *International handbook of modern lexis and lexicography*, ed. by Patrick Hanks and Gilles-Maurice de Schryver. Berlin: Springer (DOI 10.1007/978-3-642-45369-4\_92-1).
- Zgusta, Ladislav. 1971. *Manual of lexicography*. The Hague: Mouton.