

# Lexical Expansion and Lessons from Lexember

Doug Ball

May 4, 2013

## Table of contents

- 1 Lexember – What? When? Why?
- 2 Why Lexember Became a Daily Research Project
- 3 Using Wiktionary
- 4 Wrapping Up

## Basics About Lexember

- First Lexember: December 2012
- Invented by Pete Bleackley and Mia Soderquist to “combat” egregious lexical gaps
- Lexember words → posted to social media (Twitter, Google+, etc.)
  - Akin to other status-a-day memes like ‘30 Days of Thankful’
  - More remotely related to National Novel Writing Month

## My involvement

- I decided to participate in Lexember with my language Skerre
- My involvement was quite happenstance
  - Saw about it on Facebook
  - The first weekend in December was a light one for me
- Yet, Lexember seemed like an interesting thing to do for several reasons ...

# Allures

- The structure: one word a day for one month
  - Feasible yet regimented and finite (perfect combination for me)
- Expansion Possibilities
  - Even starting with  $\sim 1,450$  words in my lexicon, adding 31 more would be useful
  - Helpful to have an activity to ‘frame’ lexical expansion
- An interesting way to utilize social media
- A way to test recently added morphological constructions

## Results

Dec 1–Dec 13

## The Skerre words for Lexember 2012:

| Date   | Word        | Pronunciation              | POS    | Meaning                       |
|--------|-------------|----------------------------|--------|-------------------------------|
| Dec 1  | toneh       | [tonɛh]                    | n.     | 'lock, tuft (of hair)'        |
| Dec 2  | nera        | [nɛrɑ]                     | v.     | 'sew, knit, stitch'           |
| Dec 3  | tskito      | [tʃskito]                  | n.     | 'goal, target, aim'           |
| Dec 4  | yo'e        | [joʔɛ]                     | v.     | 'cling, stick, adhere'        |
| Dec 5  | koonit      | [ko:nit]                   | v.     | 'tangle, mix up in'           |
| Dec 6  | hiwak       | [çiwak]                    | n.     | 'string, thread'              |
| Dec 7  | titenee     | [titɛnɛ:]                  | n.     | 'job, task'                   |
| Dec 8  | ikoya       | [ikojɑ]                    | n.     | 'an itch, itching'            |
| Dec 9  | tseko       | [tʃɛko]                    | v. st. | 'be harsh, bitter, inclement' |
| Dec 10 | waatoohaqua | [wa:to:jak <sup>w</sup> ɑ] | v. st. | 'be intense, oppressive'      |
| Dec 11 | skantiyir   | [skandijir]                | n.     | 'little dear one'             |
| Dec 12 | ihaka       | [ihɑkɑ]                    | n.     | 'an act of cutting, surgery'  |
| Dec 13 | rita        | [rita]                     | adv.   | 'then, at that moment'        |

## Results

Dec 14–Dec 26

| Date   | Word         | Pronunciation | POS | Meaning                       |
|--------|--------------|---------------|-----|-------------------------------|
| Dec 14 | iriisora     | [iriˈsora]    | n.  | 'sorrow'                      |
| Dec 15 | tsite        | [t͡sɪtɛ]      | n.  | 'candle, torch, light source' |
| Dec 16 | oote'aa      | [oːtɛʔaː]     | n.  | 'completion, finish'          |
| Dec 17 | iraak        | [ɪrɑːk]       | n.  | 'flying, flight'              |
| Dec 18 | ooheyen      | [oːhɛjɛn]     | n.  | 'fatigue, exhaustion'         |
| Dec 19 | nawah        | [nawaħ]       | v.  | 'snore'                       |
| Dec 20 | keeriyanka   | [kɛːriˈjɑŋɡɑ] | v.  | 'duck, dive, lower oneself'   |
| Dec 21 | wisyata-reke | [wɪsɑtɑrɛkɛ]  | n.  | 'solstice'                    |
| Dec 22 | ihitso       | [iħɪtso]      | n.  | 'celebration, observance'     |
| Dec 23 | iyanto       | [ijɑndo]      | n.  | 'complaint, grievance'        |
| Dec 24 | tiqua        | [tɪkʷɑ]       | n.  | 'evening'                     |
| Dec 25 | ootsihe      | [oːt͡sihɛ]    | n.  | 'gladness, joy, pleasure'     |
| Dec 26 | konta        | [kɔnda]       | n.  | 'box, case, container'        |

# Results

Dec 17–Dec 31

| Date   | Word           | Pronunciation  | POS | Meaning                         |
|--------|----------------|----------------|-----|---------------------------------|
| Dec 27 | tsweni-i-terit | [tʰwɛniˈtɛrit] | n.  | 'footprint'                     |
| Dec 28 | keeriyerot     | [kɛ:riːjɛrot]  | v.  | 'snuggle, cuddle'               |
| Dec 29 | sarat          | [sarət]        | v.  | 'solve, settle, find an answer' |
| Dec 30 | sike           | [sɪkɛ]         | n.  | 'beard, whiskers (on a human)'  |
| Dec 31 | iketsi         | [ɪkɛtsɪ]       | n.  | 'ending, conclusion, finale'    |



# Results

## Additional Commentary

- Original goal was just to fill in some lexical gaps
- But I found it tricky to see Skerre's gaps (in a timely fashion on a given day)
- So I invented “what I knew”: the word for a given day had something to do with that date
  - For instance: the word for December 23 was **iyanto** ‘grievance’ in honor of Festivus, the faux holiday from *Seinfeld*, where it is customary to air grievances

# Unexpected results

## A Way to Engage Others With Skerre

- An excellent way to engage my friends – whether conlanger, linguist, or other – with Skerre
  - As linguistic items go, single words are fairly relatable, even to those who have no background in linguistics
  - Many friends of my friends were aware that I conlang, but I think they appreciated a chance to “see the language in action” – Lexember was great for this
  - Likes and comments indicated that many of my friends found Lexember interesting to follow

# Unexpected results

## Suggestions for Improvement

- Interestingly, I even got suggestions for improvement from my friends
  - The inclusion of a pronunciation in IPA was at the behest of one of my IPA-knowledgeable friends, midway through the month
  - A few comments (from conlangers/linguists) helped me better work out which sense(s) I wanted the word to have

## Research Enters the Picture

- Quickly, each day's Lexember word invention became a lengthy and research-intensive endeavor
  - Why?
- The short answer: realism
  - My insatiable desire to model my conlangs after patterns in actual natlangs
  - Always asking myself: How do natlangs do it?

# Word Creation Options

## General Possibilities

- In this domain, the question becomes: How do natlangs create new words?
- Three (broad) possibilities
  - 1 Coin – re-work preexisting resources (usually morphological, maybe sometimes phonological) to make a new word
  - 2 Borrow – use someone else's word
  - 3 Extend – add a meaning to an existing word

# Word Creation Options

As Applied to Skerre

- Extend seemed a little ridiculous to apply to Lexember
  - With a relatively small number of words in Skerre already, using Extend would create a pidgin-like lexicon
  - A pidgin-like lexicon is not the end product I want
- Borrow was unappealing due to my purist feelings towards Skerre
  - My mantra: Skerre should be full of Skerre words
- Coin – best option for me (maybe not an accident since I love morphology)

## Coining Choices

- Choosing coining means still more decisions – what kind of coining to do?
- Coining style tends to correlate with overall morphological type
  - Isolating  $\Rightarrow$  predominantly compounding
  - Synthetic  $\Rightarrow$  moderate amount of affixal morphology
  - Polysynthetic  $\Rightarrow$  large amounts of compounding and affixal morphology in some combination
- Independently, Skerre is synthetic, so the moderate path it was.

# The Nature of the Lexicon

- Coining results in a collection of morphologically complex words
- Pretty clear that all languages have some mix of morphologically simplex words and complex words
- Question: Which type of word tends to encode which meaning(s)?
- A related question: How can we tell which meaning will be encoded in which way?



# Which meaning(s) will be encoded in which way?

## Conlanger's Intuition

- One tried and true 'solution': conlanger's intuition
- This has been the traditional way I invented Skerre vocabulary
- But has the potential to be troublesome:
  - Long-time Skerre word for 'decorate': *hasera*
  - Always possible to have an opaque stem for a word – English does this a lot
  - But 'decorate' seems to be related to 'decor' → the English word does seem to be morphologically complex (at least historically)
  - Is having a opaque stem for this concept very natlang-like?

# Which meaning(s) will be encoded in which way?

Derive and Derive Some More

- Another possible ‘solution’: select a small number of basic ‘concepts’ and derive everything else
- This solution seems common to auxlangs due to a belief that this is easy to learn
- Esperanto exhibits this trend in an interesting way:

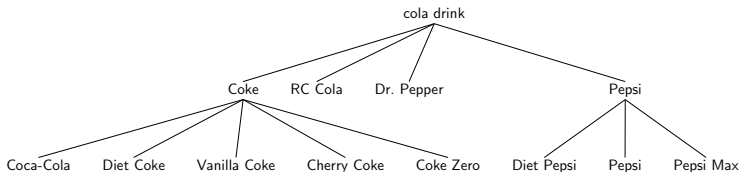
|   | Base   |         | Derived   |                          |
|---|--|---------|-----------|--------------------------|
| • | alta   | ‘high’  | malalta   | ‘low’ (‘non-high’)       |
|   | lumo   | ‘light’ | mallumo   | ‘darkness’ (‘non-light’) |
|   | trinki   | ‘drink’ | maltrinki | ‘urinate’ (‘disdrink’)   |
| • | How realistic is this: do any natlangs have ‘urinate’ as a derived word with the base for ‘drink’? |         |           |                          |

## How to Better Answer the ‘Encoding’ Question

- Using ‘pure’ intuition or deciding to go the ‘let’s derive everything’ route seems to be prone to yielding unnatural results.
- How might we more rigorously figure out which words are more likely to simplex or complex (in natlangs, with the goal of importing it into a conlang)?
- Work on lexical categorization – or even more general work on the nature of categories – might be able to help

## Some help from Lexical Categorization

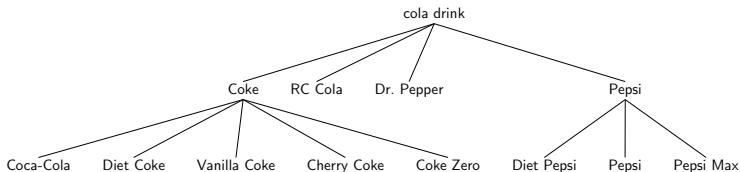
- Consider:



- This is a lexical relation network (aka taxonomy)

## Some help from Lexical Categorization

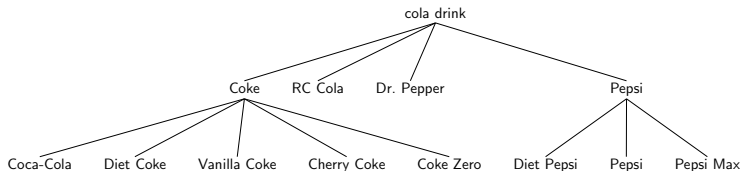
- Consider:



- This is a lexical relation network (aka taxonomy)
  - Organizes words/expressions by the 'kind of' relation (x is a kind of y)
  - Lower nodes are subtypes of the corresponding higher node

# Basic Level Categories

## The Idea



### • Observe:

- Middle line: Have the most basic “names” and seemingly are conceptually simplest (psychologists have verified that it is) → Basic Level Categories
- Lines above and below the middle line: More complex names (all compounds) and more conceptual complex → not basic

# Basic Level Categories

## The Application

- The Basic/non-basic divide looks to correlate with the morphologically simple/complex divide
- Suggests a general mode for creation:
  - Basic Level Category concepts → morphologically simple
  - Non-Basic Level Category concepts → morphologically complex
- Task for the conlanger: figure out whether a particular meaning is Basic Level or not

## Issues (that strike me)

- But using Basic Level Categories might not be *the* solution, as there are some issues with it:
  - ① Are Basic Level Categories always equivalent across different conceptual domains?
  - ② Do all languages organize their conceptual domains in the same fashion?
  - ③ How to deal with other relationships beyond the ‘kind of’ relationship?
- These questions are not easy to answer (in general) and trying to answer them may get in the way of actual conlanging
- What might be a way to get at some of the issues without completely solving the problem(s)?



## An Idea

- A thought: use Wiktionary to help
- Wiktionary (<http://www.wiktionary.org/>): a collection of free online user-created multilingual dictionaries
  - Different wiktionaries are written in different description languages
  - However: each wiktionary is not confined to only the description language's words
  - Thus, the English Wiktionary contains information about French, Latin, Russian, Chinese, Navajo, etc.
  - Fun for any language lover

## Some General Benefits of Utilizing Wiktionary

- Gets at a question that is important to me:  
How do actual natlangs do it?
- Could indirectly answer the question *Do all languages organize their conceptual domains in the same fashion?*
- Could allow me to indirectly answer or to bypass the other issues noted about using Basic Level Categories

## The Good Parts

The two key portions of Wiktionary for me:

- 1 Etymological information about a given word
- 2 Translations into (a myriad of) different languages for a given sense of a word

## The Good Parts

The two key portions of Wiktionary for me:

- 1 Etymological information about a given word
  - Useful because it often gives morphological information about the word
- 2 Translations into (a myriad of) different languages for a given sense of a word

## The Good Parts

The two key portions of Wiktionary for me:

- 1 Etymological information about a given word
  - Useful because it often gives morphological information about the word
- 2 Translations into (a myriad of) different languages for a given sense of a word
  - Useful because it gives an accessible means to look at the variation across a large number of languages

## The Ends of the Means

- I used the etymology and translation information to answer questions like
  - 1 Should this word be morphologically complex or simple?
  - 2 If complex, what are the possible bases for this word?
  - 3 If complex, what kind of affixes do languages choose to relate the base and the derived meaning?

# Drawbacks

- Wiktionary not without its own drawbacks
  - 1 Translations are heavily biased towards major languages of the world: Patterns that are specific to languages spoken by small populations are not well-represented
  - 2 Some entries lack etymological information → etymological dead-ends are common
  - 3 Information from the various languages varies significantly in quality and often requires some expertise in the language in question

## An example

- Let's consider an example: *celebration*
- <http://en.wiktionary.org/wiki/celebration>



## Lessons from the hunt

Stepping back for details of the *celebration* search

- Did not come up with as clear a picture of the possible bases as we might have liked
- It did seem that there was a propensity for treating *celebration* as a ‘action nominalization’ – some sort of deverbal noun that refers to the name of the event denoted by the original verb

# Application of Lessons to Skerre

## Nominalizations in Skerre

- 'Action nominalization' find is significant, given the morphology options in Skerre
- Nominalizations form a significant chunk of the available derivational options in Skerre
  - Nominalizations create new nouns from verbs
  - Semantically, either refer to the situation (denoted by the verb) itself or to a participant of the situation

# Application of Lessons to Skerre

## The 'Action' Nominalization

- The 'Action' Nominalization
  - *I-* (with allomorphy): 'the name of the event denoted by the verb'
  - Some Examples:

| Derived |              | Base |              |
|---------|--------------|------|--------------|
| iwesi   | 'sleep' (n.) | wesi | 'sleep' (v.) |
| ihera   | 'movement'   | hera | 'move        |
| isawo   | 'obsession'  | sawo | 'obsess'     |

# Application of Lessons to Skerre

## The Instrument/Product Nominalization

- The Instrument/Product Nominalization
  - *T-* (with allomorphy): ‘an instrument characteristically involved with the verb or product resulting from the event of the verb’
  - Some Examples:

| Derived |                    | Base  |               |
|---------|--------------------|-------|---------------|
| triis   | ‘tool’             | riis  | ‘use, employ’ |
| titeyan | ‘remedy, antidote’ | teyan | ‘heal’        |
| tsosi   | ‘food’             | hosi  | ‘eat’         |

## Application of Lessons to Skerre

### Which Affix to Use?

- In terms of the rough semantics given for each construction, *celebration* is, in principle, compatible with either nominalization pattern
  - Could be viewed as a name for the event of *celebrating*
  - Could be viewed as the product of the verb *celebrate*
- But the cross-linguistic data (to the extent we saw it) seems to suggest that the concept involved with *celebration* is more commonly encoded as an ‘action nominalization’– a name for the event of *celebrating*
- So the *i-* nominalization construction seemed to be the way to go

## What about the base?

- Recall: the cross-linguistic check via Wiktionary provided very little inspiration for possible bases
- But, as I was inventing this word, I noticed I already had a base *hitso* that meant ‘observe’ or ‘celebrate’
- Rather than re-invent the wheel, I went ahead and just used this word as the base
- Maybe I was too quick to go with a pre-existing base, but there were other things to do on Dec 22
- So the word for ‘observance’ or ‘celebration’ in Skerre became *ihitso*

## Closing Comments

### Lexember: A Great Experience

Lexember is fun!

- I heartily encourage people to participate in the future
- Don't be dissuaded if you think the methods in this talk are the only way to approach this activity, yet they seem quite difficult
  - There are probably lots of different way of making Lexember fun
- Who knows: this may be a useful way to engage more people in your language

## Closing Comments

### The Structure of the Lexicon

- The lexicon may be difficult, but it's not mean
  - The lexicon perceived as difficult by language learners and theoretical linguists alike
  - But I am convinced that there are patterns to be found in the lexicon → especially with regard to how certain meanings are expressed
- Figuring out the structure of the lexicon need not be something you have to intuit alone
  - Existing natlangs might help pave the way to better understanding



## Closing Comments

### Wiktionary

- Wiktionary can be a useful tool for exploring the lexical differences between languages
  - Better than a paper dictionary because cross-linguistic comparison right at your fingertips
- Hope that the exploration of *celebration* gave you a feel for how I used Wiktionary
  - Feel free alter my methods as you see fit

# The end

Thanks for listening and happy conlanging!