From yesterday
Senses vs Homophones

• Homophones
  • Two different words that sound the same.
    • *ruler* [king/queen] vs *ruler* [measuring thing]
    • *bank* [building for money] vs *bank* [side of river]

• Senses
  • *Same* word used in different ways
    • *media* ['the media': newspapers, etc] vs *media* [e.g. CD, DVD]
Headwords

• How different do words have to be before they count as separate headwords?
• Discuss: how many headwords?
  • Bank
  • Count
  • On
  • Field
• No single answer: relies on speaker and linguist’s intuitions
Headwords

• Morphologically related forms?
  • Often (e.g. in corporate dictionaries):
    • Inflectional morphology (e.g. singular vs plural) not separate headwords [and not listed] unless
      • Very different semantics [brother ~ brethren]
      • Irregular forms [child ~ children; bring ~ brought]
    • Derivational morphology (e.g. augmentatives, diminutives, etc) not listed unless
      • Not productive
      • Accompanied by meaning change
• Phrasal compounds
  • Usually listed (come on, come off, come over, etc)
Examples

• Headwords vs subentries

• Headwords vs ‘return all items’:
  
  [http://chamacoco.swarthmore.edu/?fields=all&q=dog](http://chamacoco.swarthmore.edu/?fields=all&q=dog)
Illustrating entries

• Adding pictures
• Where from? (copyright issues, cf. wikipedia creative commons license)
• How to pick which entries to illustrate?
  • Cultural items?
  • Flora/fauna?
  • Anything that you have pictures for? (sourcing illustrations can be a good way to get others involved in the dictionary)
Dictionary Scope?

• How big a dictionary do you plan?
  • Everything available
  • What we can do before the money runs out
  • First draft in 6 months with what we have by then, second draft in 12 months
  • Launch web site at 500 entries, then continue adding.
  • Start with plants and animals book, then a series of leaflets on different semantic domains, then combine and expand into dictionary
  • Compile all words and glosses, then add definitions, examples, etc as possible
Dictionary scope

• How much grammatical information to include in an entry?
  • Everything available? (nice to be comprehensive, but might overwhelm learners)
  • Include paradigms? (takes up space, not needed for fluent speakers, but helpful for learners)
What other sorts of information to include?

- Dialectal representation? One dialect or several?

  **Separating dialectal information:**
  - [http://www.pledari.ch/mypledari/index2.php](http://www.pledari.ch/mypledari/index2.php) (Surmiran)

  **Including words from all dialects as headwords:**
Words to include/leave out?

• Words to include/leave out?
  • Include everything?
  • Swear words
  • Taboo words
  • words used only by some sections of the community [cf intellectual property rights discussed by Marsha and Alice on Monday]
  • Loanwords (when does a loan become native?)
  • bound forms (word pieces)
  • productively formed words [e.g. compounds]
  • Idioms
• Personal names? Place names? (in Appendix instead?)
Today:

• Getting dictionary data

• Using existing materials

• (Getting your own materials)
Methods overview

• Get some material and make a user profile
• Make a preliminary wordlist
• **Plan your entries**
• **Analyze your data** (including translation)
• Synthesize (edit, merge)
• Disseminate (print out, publish, put on web, etc).- later
USING EXISTING MATERIALS
• Method #1 - Produce a list from your own knowledge, or pull some stuff off the web
• Method #2 - Use a corpus to check an existing wordlist
• Method #3 - Create a wordlist using a corpus

(what you had:
• Corpus searching.)
Demo of #2

• TextStat demo
GETTING YOUR OWN DATA
Using the DDP/translation equivalents

- [http://www.sil.org/computing/ddp/DDP_downloads.htm](http://www.sil.org/computing/ddp/DDP_downloads.htm)
- Set of structured questions in many semantic domains
- Designed to be ‘culture-neutral’ (this is good and bad…)
- Designed to allow effective brainstorming of vocabulary and example sentences.
Using templates

• Templates are checklists of information for a semantic field.

• E.g. for a verb:
  • Is it transitive or intransitive?
  • Conjugation class, etc
  • What case marking do the verb’s arguments take?
  • Can it take a subordinate clause?
  • What derivational morphology can the verb take? (passive, applicative, etc?) un-, re-
  • collocations? (words that tend to go together)
  • Related words
Things to watch out for

- examples vs definitions
- ranges of term (e.g. English ‘hand’ \neq Bardi ‘nimarl’)
- specialized meanings, specialized terminology
- polysemy vs homophony (1 word, two meanings vs 2 words, two meanings)
WORKFLOW
‘Workflow’

• Sequence for working with data so that
  • You work consistently
  • Work gets done in a logical order
  • Data don’t get overlooked
Workflow diagram
General workflow for getting word data

- Every time you come across a new word, ask about it, add it to the dictionary, and include an example from context.
- Periodically, go through your data for questions, missing information, etc.

Pros:
- increases the dictionary size rapidly (at least initially)
- allows for common items to emerge early.

Cons:
- unsystematic
- easy to get only partial information
- can interrupt the flow of other work.
Eliciting words for dictionaries

- Work by semantic field
- Work in small groups, if possible.
- Record everything.
TASKS
Task 1: planning entries

- How many entries would you create out of this Baonan dataset?
Let’s focus on these:

- ñebeb
- ñebron
- ñebróna

• Is this three entries, two, or one?
One possible solution:

• ...we might create two entries:


  Ĥeëbê'h a <adj> liar.  Cf. Amdo Tibetan Ĥeëfca.
Task 2

- Take your list of ten words and do a corpus search for a couple; use those searches to plan fuller entries.
- (Optional: use a couple of the words to think about a template for a semantic field.)
Additional…
Ways of getting data (not exhaustive!)

• pointing at things
• translation equivalents
• brainstorming words by semantic field
• vernacular definitions
• using books of pictures
• translation equivalents vs discussing meanings with speakers.
  • can be difficulties in translation in both directions
  • might want to make this clear in entries
  • example: Bardi ‘grandmother’ terms;
    • mother’s mother vs father’s mother
    • but people are most likely to look up ‘grandmother’ or ‘granny’
Task:

• In 5 minutes, come up with as many words for things in the sky as you can (in any language you like).
• In the second 5-10 minutes, discuss your list with a partner, compare, see if you can add to it.