Hundred Sentences and XFST Assignments

2-19-14
Your assignment

• Make an elicitation checklist of at least 100 sentences covering these things:
  – Transitive and intransitive verbs.
  – Semantic verb classes: stativity, dynamicity, and telicity.
  – Tense/mood/aspect
  – Special sentence types:
    • Existential
    • Copula
    • Possession
    • Questions
  – Speech acts:
    • Statement
    • Command/prohibition
    • Question
  – Negation
Your assignment continued

• Noun types
  – Pronoun
  – Common
  – Proper
  – Count/Mass
  – Concrete/Abstract

• Noun phrases
  – Definiteness
  – Possession
  – Proximity
  – Diminutive/Augmentative
  – Quantity
    • Cardinal numbers
    • Quantifiers: all, each, both, some
  – Ordinal (first, second)
  – Partitive
    • The top of the tree
    • The bottom of the tree
Your assignment continued

• Adjectives
  – Comparative
  – Superlative
  – Intensive (very)
  – Unintensive (a little)

• Adpositions and/or case marking

• Typical modifiers of nouns and verbs.
  – Location, time, manner, etc.
Example: Belele (Bwele)  
(Inspiration from Warlpiri and Iñupiaq)

Bele-le  
word-generic
“Wordkind”

Fwe-le bar -la -to  
Bird-gener sing-hab-pres
“Birds sing”

Fwe-n bar -la -to  
Bird-def sing-hab-pres
“The bird sings”

Fwe-n bar-na  
Bird-def sing-past
“The bird sang”

Fwe-n bar- go- na  
Bird-def sing-incep-past
“The bird started to sing”

Intransitive sentences.  SV word order.

Verb morphology: root-aspect-tense

Aspects:
habitual  -la
Punctual  Ø
Inceptive  -go

Tenses:
Present  -to
Past  -na
Example continued

• Compound noun:
  First element not marked for number and definiteness
  Nar -bi ndo –ki -n
  happy -one day –pl –def
  “the happy days”

  There are adjective stems, but they have to be turned into nouns in order to use them. -bi turns an adjective into a noun.

• Two noun phrases (either order, or can be discontinuous):
  Nar -bi -ki -n ndo –ki -n
  happy -one –pl-def day –pl –def
  “the happy ones the days”

  ndo –ki -n nar -bi -ki -n
day –pl –def happy -one –pl-def
  “the days the happy ones”

  Nouns and their modifiers can be in either order or can be separated if the modifier is marked for number and definiteness.
Example continued

• Diminutive postbase
  Konik-nu -n
  Boy -dim –def
  “The small boy”

• Compound noun
  Nu -bi konik-an
  Small-one boy -def
  “The small boy”

• Two noun phrases
  Konik-an nu -bi -n
  Boy -def small-one-def
  “The boy the small one”

  nu -bi -n konik-an
  Small-one-def boy -def
  “The small one the boy”
References

• Mary B. Rossman and Mary W. Mills, *Graded Sentences for Analysis*, Noble and Noble, 1922.

• The Comrie-Smith checklist:
  http://www.eva.mpg.de/lingua/tools-at-lingboard/questionnaire/linguaQ.php


• FrameNet:  https://framenet.icsi.berkeley.edu/fndrupal/
  – Thanks, Chuck Fillmore, and rest in peace! (1929-2014)
Suggestions

• You don’t have to do all of the following.
• You can add your own ideas too.
Intransitive verbs

• Activities: play, study, work, write, fight, read, eat, dance, speak, swim, graze
• Activities with sound: Bark, hum, sing, laugh, cry, shout, purr, roar,
• Activity with body movement but not change of location: Nod, wave, smile, bow,
  – English: nod a greeting; smile a broad smile
• Aspectual: Start, stop, begin, continue, keep
• Change of location with mostly neutral manner: Come, go, arrive, depart, fall, turn (away, around), travel, wander, gather, arise
• Manner of motion: Walk, run, strut, roll, jump, march, limp, hop, scamper, plunge, flit,
  – English: walk/roll/limp/hop across the room vs
  – vs cross the room walking
  – vs go walking across the room
• Change of state: melt, freeze, grow, bloom, improve, fade,
  – English: grow tall, freeze solid vs
  – freeze and become solid
• Look, listen,
• pass (time passes), happen, occur
• Shine, blow, dawn, flow, clear away, sparkle, blaze, rustle, glow, glitter
• Strike (clock), toll, flutter,
• Locate: Wait, stand, rest, sit, lie, tower, hang, live, extend,
• Glide, float, waft, (smoke) curl up, wave, bend, sway,
• Know, understand,
• Grieve, rejoice,
Telicity

• I walked home (Telic)
  – If I start walking home and stop, I haven’t walked home.
  – I walked home in an hour.

• I walked (Atelic)
  – If I start walking and stop, I have walked.
  – I walked for an hour.
Things to think about:

• Aspectuals (start, finish, etc.): verbs or not?
  – Slavic languages have inceptive/inchoative aspect.
  – What does it mean for an event to start?
    • Start dancing.
  – For a noun to start?
    • The movie started.

• Differential encoding (case or agreement) based on animacy, definiteness, TAM, speech act, stativity, evidentiality.
  – Different case marker for agent/patient or volition/non-volition
  – Yidiny (Australia): She cooked fish vs she happened to cook fish.
  – Hindi: To you befalls to go. (You have to go.)

• Different uses of TAM depending on stativity, telicity, evidentiality.
  – The clock stands on the shelf (present time)
  – I am talking (present time)
  – I talk (habitual)
  – Spanish: likelihood of morphological future, present, or periphrastic future with “ir” depends on verb class.

• Lexicalization of manner and motion:
  – float (manner+motion) into (direction) the cave,
  – enter (direction+motion) the cave floating (manner)
  – go (motion) floating (manner) into (direction) the cave.
Copula sentences

• Identity:
  – Clark Kent is Superman.
  – Lori and Alan are the instructors.

• Role:
  – She is a doctor.

• Description/Property:
  – She is tall.
Existential sentences

• There are rabbits on the lawn.
• There were students arrested.
• There were students yawning in class.
• There are no unicorns.
Possession Sentences

• In some human languages existential and possession sentences are the same.
• In some human languages location (The book is on the table) and possession (The book is at me) are the same.
  – Existential:
    • There is a book to/at me
  – Have:
    • I have a book.
  – Be:
    • To/at me is a book.
• Alienable and inalienable possession may be different:
  – kinship and body parts vs ownership
  – other part-whole relations (the top of the tree)
Negation

• With each tense, aspect, mood, speech act.
  – They did not yawn.
  – They are not yawning.
  – They will not have been yawning.
  – Nobody yawned.
  – Didn’t you yawn?
  – Don’t yawn!
Negation Sentences from ParGram 2014
Ignore extra numbers: PowerPoint outsmarted me

Negation
1. The house is not big.
2. The cat is not in the house.
3. The houses are not big.
4. The cats are not in the house.
5. The cat is not in the house either.
6. Neither is the cat in the house.
7. The girl did not buy the apple.
8. The girl did not buy apples.
9. The girl did not buy any apples.
10. The girl did not buy anything.
11. The girl bought no apples.
12. The girl bought nothing.
13. The girl never bought anything.
14. The girl did not buy some books.
15. The boy does not think that the girl is tall.
16. The boy said that the girl is not tall.
17. The boy cannot watch the film.
18. The boy cannot not watch the film.
19. The woman asked the girl not to open the door for strangers.
20. The woman saw the boy not eating the fish.
21. Not being hungry, the man did not cook fish.
22. It is not the dog that the girl saw.
23. It is not the dog that the girl did not see.
24. Do not eat the fish.
25. Do not eat anything.

ParGram is a group of people who write high coverage precision grammars with the LFG parser (XLE) in many languages. They have been meeting once or twice per year for 20 years.
Possession sentences from ParGram 2014

26. The dog has a bone.
27. The farmer owns a cow.
28. The cat has dark fur.
29. That tractor has a bright red color.
30. The tree has a big branch.
31. Dogs have four legs.
32. The farmer has a fear of spiders.
33. The girl has two brothers.
34. The girl doesn't have a father.
35. NP: the dog's bone
36. NP: the farmer's cow
37. NP: the cat's dark fur
38. NP: the bright red color of the tractor
39. NP: the tree's branch
40. NP: the branch of the tree
41. NP: the farmer's fear of spiders
42. NP: the fear of spiders of the farmer
43. NP: the girl's two brothers
44. NP: the two brothers of the girl
45. NP: the farmer's cows' milk
46. NP: the milk of the cows of the farmer
Transitive verbs

- Cause motion: push, pull, move,
- Cause motion with manner: throw, roll, bounce
- Cause motion with path: insert, remove,
- Cause motion with manner: wipe, blow, sweep
- Move in place: shake, wave,
- Cause change of state: melt, freeze, break, bake
- Contact: touch, kick, hit, tap,
- Create: create, make, build, form,
- Create with method: bake, draw, sew
- Attach: sew, glue,
- Fill: load, fill,
  - English: load the truck with hay/load hay onto the truck
- Cover: paint, cover, spray, smear
  - English: spray the wall with paint/spray paint on the wall
- Agent, Patient, Recipient: give, hand, award, present
  - English: hand something to someone/hand someone something
- Convey information with or without manner: tell, whisper, say, scream,
- Cause emotion: frighten, scare, startle, surprise, anger, please
- Experience emotion: fear, like, love, hate
- Activities: write, read, sing, say
- Other: share, shelter, risk
Transitive sentences

- In human languages, the most typical sentences have a definite, animate agent acting on an indefinite inanimate patient, and having an effect on it. Telic and past tense (completed) are probably more typical than incomplete. (Transitivity in Grammar and Discourse. Paul J. Hopper; Sandra A. Thompson. Language, Vol. 56, No. 2. (Jun., 1980), pp. 251-299.)
  - Fifth most highly cited paper in Language, the journal of the Linguistic Society of America
  - Prototypical sentence: The girl caught a fish.
What happens in human languages when transitive sentences aren’t typical?

• Grammar happens!

• Differential object marking:
  – Spanish uses “a” before animate direct objects.
  – Hebrew uses “et” before definite direct objects.
  – Hindi “ko” is obligatory for animate direct objects.
  – Turkish and Farsi, accusative marker is used for specific direct objects.
  – Swahili and other Bantu languages: there is an agreement marker on the verb for definite direct objects.
  – Chinese uses “ba” construction for telic sentences with definite direct objects.
What happens in human languages when transitive sentences aren’t typical?

• Chinese often uses existential sentences for inanimate subjects of transitive verbs:
  – “Have student read book” (A student read a book) is more common than “Student read book” (A student read a book).

• Inverse morphemes:
  – Algonquian, Athabascan, Mapudunun (Chile, Argentina)
  – When the patient outranks the agent in definiteness, animacy, or person, an extra morpheme appears on the verb and the patient acts like the subject. Kind of like obligatory passivization, but they might also have a separate passive construction.
Morphology in intransitive and transitive sentences

- **Subject:** gender, number, animacy, agentivity, definiteness, proximity
- **Object:** gender, number, animacy, affectedness, definiteness, proximity
- **Verb:** TAM, stativity, telicity, evidentiality, semantic class
- **Speech act:** question, statement, command/request, prohibition, promise, offer, etc.
- **Negation**
Adpositional phrases

• Do you have adpositions (prepositions or postpositions)?
  – If yes, do you have case marking?
  – If yes, what case do you use with different adpositions?
• If you don’t have adpositions, you probably have a lot of case markers (like Finnish).
• Or maybe you’ve reverted to an old stage of Indo-European with preverbs.
• Or maybe you have applicative morphemes like Bantu languages.
Modifiers of sentences

• Certainty: surely, probably, maybe, without doubt
• Manner: hesitantly, hastily, haltingly, patiently
• Time when: tomorrow, today
• Time duration: in an hour, for an hour
• Location:
• Still/yet
Nouns

• Pronouns
  – Some human languages have different case marking systems for pronouns and common nouns.
    • English has case only on pronouns.
    • Some languages have nominative-accusative pronouns and ergative-absolutive common nouns.

• Common nouns

• Proper nouns

• Any sub-classes of nouns?
  – Gender or other noun classes
  – Count and mass nouns
Noun phrases

• Definiteness
• Possession
• Proximity
• Diminutive/Augmentative
• Quantity
  – Cardinal numbers
  – Quantifiers: all, each, both, some
• Ordinal (first, second)
• Partitive
  – The top of the tree
  – The bottom of the tree
Adjectives

• Comparative
  – Happier

• Superlative
  – Happiest

• Very
  – Very happy

• A little
  – A little happy
  – Slightly happy
Why is it so hard to make a small set of sentences that covers everything?

• Things interact:
  – Tense and question formation
    • The student yawned.
    • Did the student yawn.
  – Definiteness and possession
    • My book (definite)
    • A book of mine (indefinite)
Does this seem like a lot?

• We are not even touching on
  – Coordinate structures
  – Embedded clauses
    • Relative clauses
    • Complement clauses
    • Subordinate clauses
  – Passive voice
  – Conditional (if-then)
  – Reflexive pronouns
    • (around 200 things to check in Comrie-Smith Checklist)
  – Scope of negation phenomena
  – Information structure (old and new information)
  – Location:
    • Comrie-Smith checklist has an extensive list of movement/location distinctions
      (movement to a place under something, movement under something and keep going, movement to a place inside something, movement through the inside and keep going, etc.)
    • Melissa Bowerman had even more.
  – Etc.
Morphology Assignment
Underlying forms and surface forms

• Phonology:
  – Underlying: ɪNkʌmplɪt (incomplete)
  – Surface: ɪŋkʌmplɪt

• Orthography
  – Underlying: happy+ness
  – Surface: happiness
A terrible twist of terminology in XFST

• The underlying form is called the upper form.
• The surface form is called the lower form.
• Think of it this way:
  – in phonology class, if you are going to derive a word using a sequence of rules, you write the underlying form at the top of the page/board and write the derivation under it.
Underlying and surface forms

• Underlying forms can be glosses of morphemes.

• Underlying: happy[AdjCmpr]
  • Surface: happier

• Underlying: happy[NDer][Npl]
  • Surface: happinesses
Where do those things in square brackets come from?

• You make them up.
• I decided to call –er [AdjCmpr] for “adjective, comparative”
• You could call it something else.
You have to make up labels for your morphemes
What will you do for this assignment?

• You will use a program called XFST (Xerox Finite State Transducer).
• If you give it an underlying form, it will give you the surface form.
• If you give it a surface form, it will give you the underlying form.

• XFST software: 
• Tutorial slides by Ken Beesley and Lauri Karttunen: 
Simplest way

Multichar_Symbols  [NSg]  [NPl]  [Adj]  [AdjCmpr]  [AdjSupr]  [NDer]

LEXICON Root
0:0   NRoot ;
0:0   AdjRoot ;

LEXICON AdjRoot
happy[Adj]:happy    # ;
happy:happi         AdjSuffs ;

LEXICON AdjSuffs
[AdjCmpr]:er        # ;
[AdjSupr]:est       # ;
[NDer]:ness         NSuffs-es ;

LEXICON NRoot
book:book           NSuffs-s ;
pencil:pencil       NSuffs-s ;
child[NSg]:child    # ;
child[NPl]:children # ;
rash:rash           NSuffs-es ;

LEXICON NSuffs-s
[NSg]:0    # ;
[NPl]:s     # ;

LEXICON NSuffs-es
[NSg]:0    # ;
[NPl]:es   # ;
Type "help" to list all commands available or "help help" for further help.

xfst[0]: read lexc lex-without-rules.t
lex-without-rules.txt  lex-without-rules.txt~
xfst[0]: read lexc lex-without-rules.txt
Opening input file 'lex-without-rules.txt'
February 13, 2014 04:56:33 GMT
Reading UTF-8 text from 'lex-without-rules.txt'
Root...2, AdjRoot...2, AdjSuffs...3, NRoot...5, NSuffixs-s...2, NSuffixs-es...2
Building lexicon...Minimizing...Done!
3.4 Kb. 35 states, 44 arcs, 13 paths.
Closing 'lex-without-rules.txt'
xfst[1]: apply up children
child[NPL]
xfst[1]: apply up books
book[NPL]
xfst[1]:

xfst[1]: apply down child[NPL]
children
xfst[1]:

xfst[1]:
Another way

LEXICON NRroot
book;book  NSuffs;
pencil;pencil  NSuffs;
child[NSg]:child  #;
child[NPl]:children #;
rash;rash  NSuffs;

LEXICON NSuffs
[NSg]:0  # ;
[NPl]:s  # ;

s -> es || sh _ .# ;
Break the job down into small pieces

- Test Suite Sentences (More detail later)
- Design your inflectional paradigms.
  - Give a square bracket name to each morpheme.
- Design your derivational morphemes.
  - Give a square bracket name to each morpheme.
- Install XFST and learn to run it.
- Write a file called a lexc file.
- Optional: write a rule file.
- Test:
  - If you input a word that is well-formed in your language, the output is the underlying form.
  - If the word is not ambiguous, you should get one underlying form.
  - If the word is ambiguous (e.g., [unlock]able, un[lockable]), you should get one underlying form for each meaning.
  - If you input an ill-formed word (e.g., happyness), there should be no underlying form and the FST will output three question marks (???).