

**NLP**

# Introduction to NLP

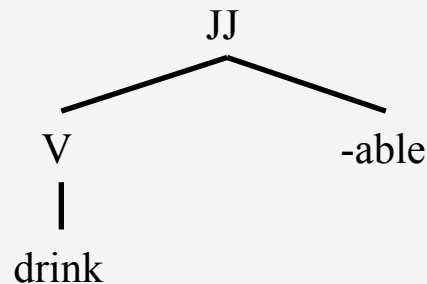
## *Morphology and the Lexicon*

## Mental Lexicon

- What is the meaning of cat? Its pronunciation? Part of speech?
- What is the meaning of wug?
- What is the meaning of cluvious?
- Compare traftful and traftless?
- Morphology of these words
- Intuition and productivity
- “Runs”
- Allomorphs – “cats/oxen”, “played/swung”
- Affixes

# Derivational Morphology

- Er (many examples)
- What do these morphemes mean?
  - Ness, able, ing, re, un, er (adj)
  - JJ  $\rightarrow$  V + “-able”
- Recursion:
  - unconcernednesses
- Ambiguous – undoable
- Not ambiguous – unbelievable – why?



## Answer to the Quiz

- Undoable
  - unable to be done
  - able to be undone
- Unbelievable
  - unable to be believed
  - ? able to be unbelieved

# Morphological Examples

- Reduplication
  - amigo = friend, amimígo = friends (in Pangasinan) [Rubino 2001]
  - savali = he travels, savavali = they travel (in Samoan)
- Circumfixes
  - spielen – gespielt (in German)
- Pig Latin
  - appyhay
- Verlan
  - “céfran”, “ripou” (from “l’envers”, “Français”, “pourri”)
- Massa-*freakin*’-chusetts
  - where can you insert “freakin’” in “education”?

## Answer to the Quiz

- The “freakin’” infix is inserted
- ... to the left of the syllable that bears the main stress
  - edu-*freakin'*-cation
  - \* educa-*freakin'*-tion
  - \* e-*freakin'*-ducation
- though there can be exceptions

## Morphemes

- Stems, affixes
- Concatenative morphology
- Templatic morphology (e.g., Semitic languages):
  - lmd (learn), lamad (he studied), limed (he taught), lumad (he was taught)



# Inflectional Morphology

- Tense, number, person, mood, aspect
- Five verb forms in English
- 40+ forms in French
- Six cases in Russian:  
<http://www.departments.bucknell.edu/russian/language/case.html>
- Up to 40,000 forms in Turkish (you cause X to cause Y to ... do Z)

# Morphological Analysis

- sleeps = sleep + V + 3P + SG
- done = do + V + PP

# Turkish Vowel Harmony

	Front		Back	
	Unrounded	Rounded	Unrounded	Rounded
High	i	ü	ɪ	u
Low	e	ö	a	o

- Back vowels
  - in the room → oda**da**
  - at the door → kapı**da**
- Front vowels
  - at home → ev**de**
  - at the lake → göl**de**
  - on the bridge → köprü**de**

# NACLO Problem

- Turkish
  - [www.naclo.cs.cmu.edu/problems2010/F.pdf](http://www.naclo.cs.cmu.edu/problems2010/F.pdf)

# NACLO Solution

- Turkish
  - [www.naclo.cs.cmu.edu/problems2010/FS.pdf](http://www.naclo.cs.cmu.edu/problems2010/FS.pdf)

# Agglutinative Languages

- How does English become Turkish?

if      we      will      be      able      to      make      ...      become      strong

if      we      will      be      able      to      make      ...      become      strong

...      strong      become      to      make      be      able      will      if      we

...      sağlam      +laş      +tır      +abil      +ecek      +se      +k



... sağlamlaştıracaksak

アメフト	amefuto	Ame(rican) Foot(ball)
アイスクリーム	aisu kurīmu	ice cream
アイドル	aidoru	idol
アパート	apāto	apartment
バイク	baiku	bike
バリアフリー	bariafurī	barrier free
コンピューター	konpyūtā	computer
デスク	desuku	desk (at a news agency)
ドラマ	dorama	drama (on TV)
エレベーター	erebētā	elevator
エスカレーター	esukarētā	escalator
フライドポテト	furaidopoteto	fried potato (French fries)
グラス	gurasu	glass (for drinking)
ハッピーエンド	happīendo	happy end(ing)
ホットケーキ	hottokēki	hotcake (pancake)
カシューナッツ	kashū nattsu	cashew nut
コーヒー	kōhī	coffee
クラブ	kurabu	club
キーボード	kībōdo	keyboard
キャンペーン	kyanpēn	campaign
キャップ	kyappu	cap
パソコン	pāsokon	perso(nal) com(puter)
パーソナルコンピューター	pāsonaru konpyūtā	personal computer
レジュメ	rejume	resume
レストラン	resutoran	restaurant
リモコン	rimokon	remo(te) con(trol)
サラダ	sarada	salad
タバコ	tabako	tobacco
テレビゲーム	terebigēmu	television game
ゼミナール	zemināru	seminar

# Introduction to NLP

*Other Levels of Linguistic Analysis*



# Semantics

- Semantics
  - Lexical semantics and compositional semantics
- Lexical Semantics
  - Hypernyms, hyponyms, antonyms, meronyms and holonyms (part-whole relationship, tire is a meronym of car), synonyms, homonyms
  - Senses of words, polysemous words
  - Collocations
    - *white hair, white wine*
  - Idioms
    - *to kick the bucket*
- *Compositional Semantics*
  - How to understand the meaning of a sentence based on the meaning of its components.

# Pragmatics

- The study of how knowledge about the world and language conventions interact with literal meaning.
- Speech acts
- Resolution of anaphoric relations
- Modeling of speech acts in dialogue

## Other Areas

- Sociolinguistics
  - interactions of social organization and language.
- Historical linguistics
  - change over time.
- Linguistic typology
- Language acquisition
  - L1 and L2
- Psycholinguistics
  - real-time production and perception of language

**NLP**