
Advanced Natural Language Processing

Lecture 2

Morphology

Frank Keller (slides by Philipp Koehn)

23 September 2011



How Many Different Words?

10,000 sentences from the Europarl corpus

Language	Different words
English	16k
French	22k
Dutch	24k
Italian	25k
Portuguese	26k
Spanish	26k
Danish	29k
Swedish	30k
German	32k
Greek	33k
Finnish	55k

Why the difference? Morphology.

Morphemes: Stems and Affixes

- Two types of morphemes
 - stems: *small*, *cat*, *walk*
 - affixes: *+ed*, *un+*
- Four types of affixes
 - suffix
 - prefix
 - infix
 - circumfix

Stem vs. Root vs. Lemma

- Stem
 - the part of the word that is common to all its inflected variants
 - stem of **produce** and **production** is **produc**
- Root
 - primary lexical unit of a word, cannot be reduced into smaller constituents
 - the root of **interrupt** is **rupt** (although that has no meaning)
- Lemma
 - the canonical form, dictionary form, or citation form of a set of words
 - **fly**, **flies**, **flew** and **flying** are forms of same lexeme, with **fly** as lemma

Suffix

- Plural of nouns

cat+s

- Comparative and superlative of adjectives

small+er

- Formation of adverbs

great+ly

- Verb tenses

walk+ed

- All inflectional morphology in English uses suffixes

Prefix

- In English: meaning changing particles

- Adjectives

un+friendly
dis+interested

- Verbs

re+consider

- German verb prefix *zer* implies destruction

Infix

- In English: inserting profanity for emphasis

abso+bloody+lutely
unbe+fucking+lievable

- Why not:

ab+bloody+solutely

Circumfix

- No example in English
- German past participle of verbs:

ge+sag+t (German)

Not that easy...

- Affixes are not always simply attached
- Some consonants of the lemma may be changed or removed
 - walk+ed
 - frame+d
 - emit+ted
 - eas(-y)+ier
- Typically due to phonetic reasons

Irregular Forms

- Some words have irregular forms:
 - is, was, been
 - eat, ate, eaten
 - go, went, gone
- Frequent words more likely to have irregular forms
- Morphology reduces the need to create completely new words; irregular forms run counter to that.

Inflectional Morphology

- In English
 - nouns are inflected for number (plural: +s) and for possessive case (+’s)
 - verbs are inflected for tense (+ed, +ing) and a special 3rd person singular present form (+s)
 - adjective are inflected in comparative (+er) and superlative (+est)
 - determiners are not inflected
- In German
 - nouns are inflected for number and case
 - verbs are inflected for tense, person, and number
 - adjectives are inflected for number, case, gender, and definiteness
 - determiners are inflected for number, case and gender

Forms of German the

Case	Singular			Plural		
	masc	fem	neut	masc	fem	neut
nominative (subject)	der	die	das	die	die	die
genitive (possessive)	des	der	des	der	der	der
dative (indirect object)	dem	der	dem	den	den	den
accusative (direct object)	den	die	das	die	die	die

Not only many different forms, but each form is highly ambiguous:
syncretism

Why Morphology?

- Alternatives
 - Some languages have no verb tenses
 - use explicit time references (*yesterday*)
 - Case inflection determines roles of noun phrase
 - use fixed word order instead
 - Case-marked noun phrases often play the same role as prepositional phrases
- There is value in redundancy and subtly added information...

Inflectional vs. Derivational Morphology

- Derivational morphology
 - change part of speech or meaning of a word
 - not driven by syntactic relations outside the word
- Inflectional morphology
 - does not change basic meaning or part of speech
 - expresses grammatical features or indicates relations between different words
 - applies to all words of the same part of speech
- Inflectional affixes are attached before derivational affixes:

govern+ment+s

Derivational Morphology

- Changing the part of speech, e.g. noun to verb

word → wordify

- What does that mean?

Derivational Morphology

- Changing the part of speech, e.g. noun to verb

word → wordify

- What does that mean?
- Consulting Google:
 - 8,840 hits
 - e.g., wordify mugs, tshirts and magnets

Derivational Morphology

- Changing the verb back to a noun

wordify → wordification (2,350 hits on Google)

- A person who engages in wordification

wordification → wordificator (8 hits on Google)

- A person who wordifies

wordify → wordifier (2,820 hits on Google)

- What is the difference between a wordifier and a wordificator?

Derivational Morphology

- Turning [wordification](#) into a ideology:

[wordification](#) → [wordificationism](#)

- 1 hit on Google

I think you're confusing the term "Democracy" with "Capitalism"; I think you mean "Has Capitalism failed"?

No. It hasn't.

I agree, Hambone; I'm just trying to correct the [wordificationism](#).

Where in the world did you get the word "[wordificationism](#)"? Not in the Merriam-Webster dictionary, not in the Thesaurus...

Derivational Morphology

- A adherent of wordificationism

wordificationism → wordificationist

- 0 hits on Google
- We created a new word!

Compounds

- Creating new words by merging multiple words
- Examples in English:

home work → homework
web site → website

- Spelling can vary, but arguably the same compound:

web site or web-site or website

Acronyms

- Guardian:
David Cameron plans to save millions by cutting quangos
- What is a quango?

Acronyms

- Guardian:
David Cameron plans to save millions by cutting quangos
- What is a quango?
- An Australian animal?

Acronyms

- Guardian:

David Cameron plans to save millions by cutting quangos

- What is a quango?

- An Australian animal?

- No:

quasi non-governmental organization

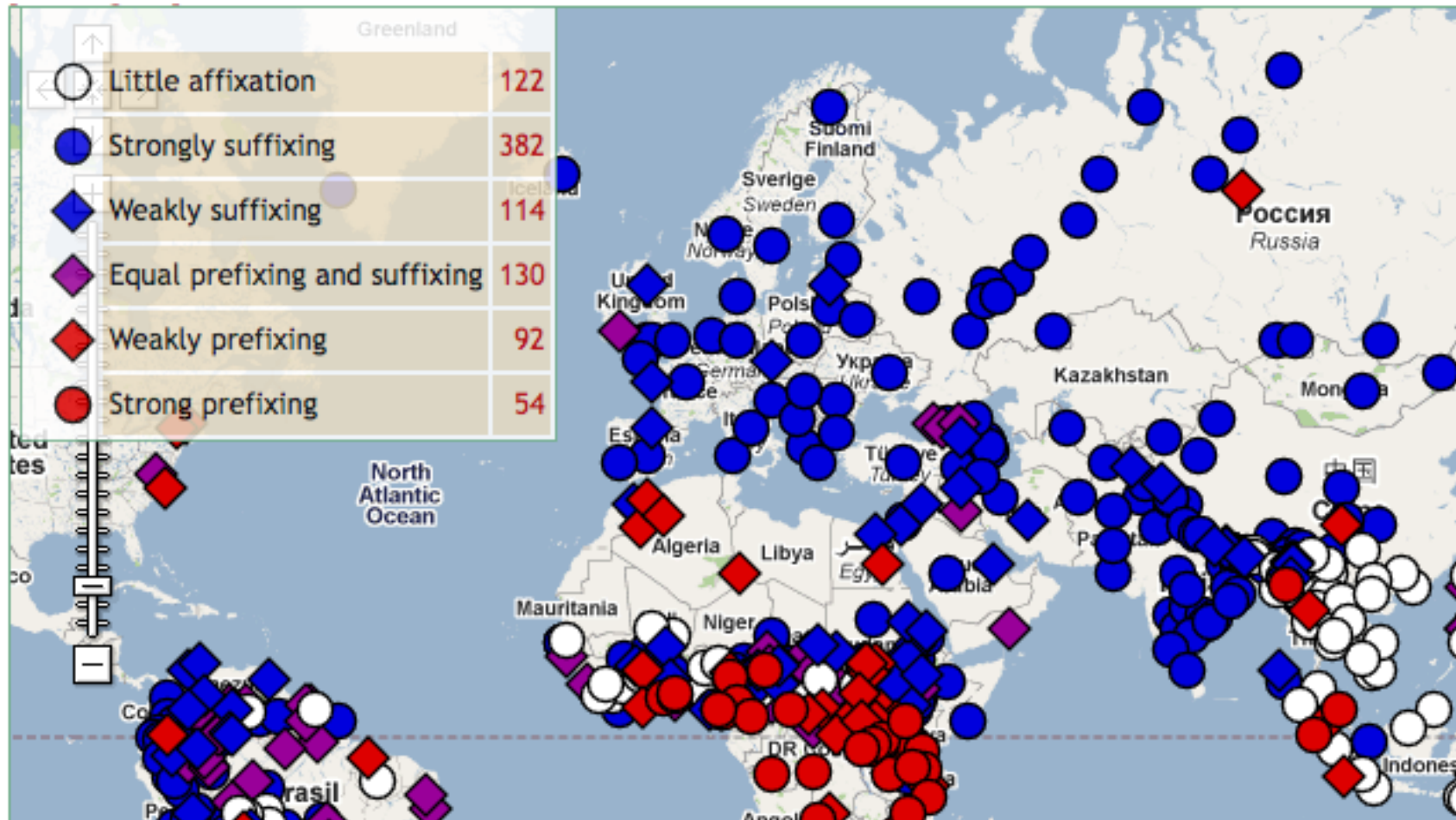
Acronyms

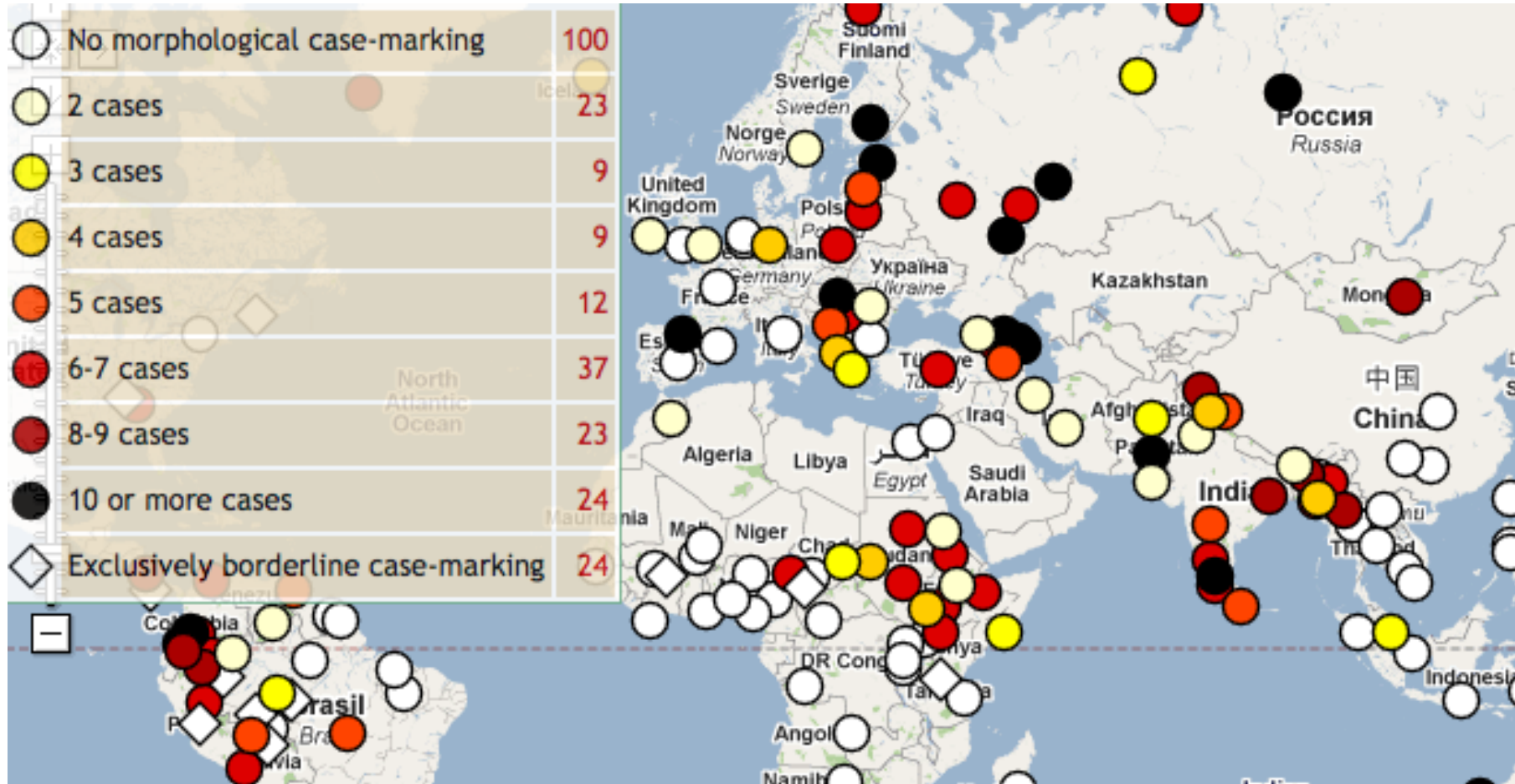
- Another example: Wikileaks / Guardian, document 2007-081-100110-0444:

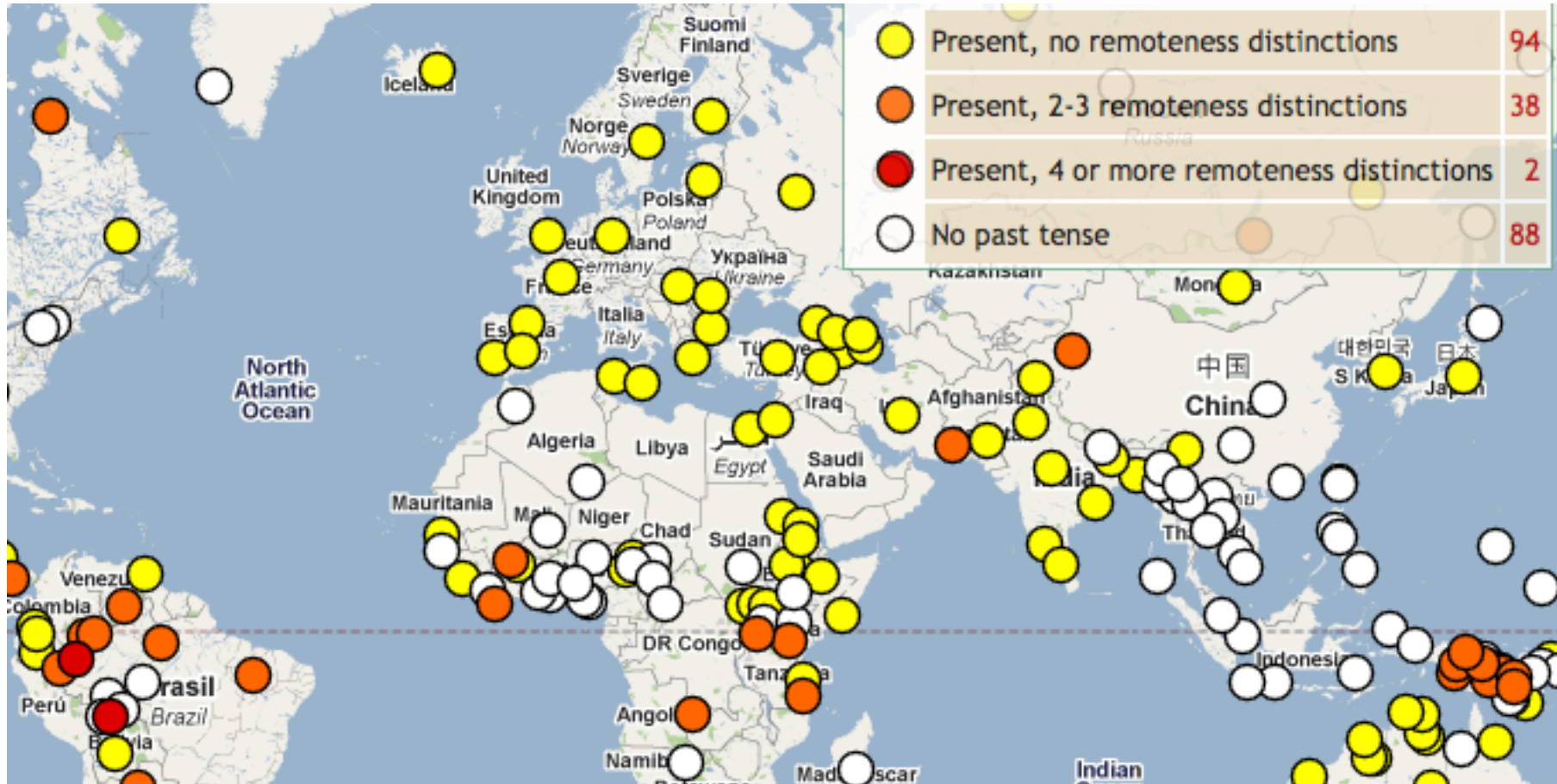
OGA operating in TF Catamount sector moved into Malekshay for operation. LN Shum Khan ran at the sight of the approaching CFA's. CF utilized the escalation of force doctrine and shouted to stop, fired warning shots and then fired to wound. The LN was hit in the ankle and treated by Element medics on scene. It was determined through discussions with local Elders that the man was a deaf mute that was nervous of the CF operation. Solatia was made in the form of supplies and the Element mission progressed

Different Languages

- Languages differ a lot in morphology
- Examples from The World Atlas of Language Structures Online (wals.info)
 - prefixes vs. suffixes
 - cases (zero to more than ten)
 - past tense remoteness distinctions







Next Lecture: Computation Models of Morphology