



# Announcements

---

- ▶ FP due next Wednesday. Check-ins returned
- ▶ A4, A5 back later this week
- ▶ eCIS evaluations

# Multilinguality



# Dealing with other languages

---

- ▶ Other languages present some challenges not seen in English at all!
- ▶ Some of our algorithms have been specified to English
  - ▶ Some structures like constituency parsing don't make sense for other languages
  - ▶ Neural methods are typically tuned to English-scale resources, may not be the best for other languages where less data is available
- ▶ Question:
  - 1) What other phenomena / challenges do we need to solve?
  - 2) How can we leverage existing resources to do better in other languages without just annotating massive data?



# This Lecture

---

- ▶ Morphological richness: effects and challenges
- ▶ Morphology tasks: analysis, inflection, word segmentation
- ▶ Cross-lingual tagging and parsing
- ▶ Cross-lingual word representations

# Morphology



# What is morphology?

---

- ▶ Study of how words form
- ▶ Derivational morphology: create a new *lexeme* from a base
  - estrangle (v) => estrangement (n)
  - become (v) => unbecoming (adj)
    - ▶ May not be totally regular: enflame => inflammable
- ▶ Inflectional morphology: word is inflected based on its context
  - I become / she becomes
    - ▶ Mostly applies to verbs and nouns



# Morphological Inflection

- In English: I arrive      you arrive      he/she/it arrives      [X] arrived  
we arrive      you arrive      they arrive

- In French:

|                 |                            | singular                 |                          |                          | plural                        |                             |                            |
|-----------------|----------------------------|--------------------------|--------------------------|--------------------------|-------------------------------|-----------------------------|----------------------------|
|                 |                            | first                    | second                   | third                    | first                         | second                      | third                      |
| indicative      |                            | je (j')                  | tu                       | il, elle                 | nous                          | vous                        | ils, elles                 |
| (simple tenses) | present                    | arrive<br>/a.ʁiv/        | arrives<br>/a.ʁiv/       | arrive<br>/a.ʁiv/        | arrivons<br>/a.ʁi.vɔ̃/        | arrivez<br>/a.ʁi.ve/        | arrivent<br>/a.ʁiv/        |
|                 | imperfect                  | arrivais<br>/a.ʁi.vɛ/    | arrivais<br>/a.ʁi.vɛ/    | arrivait<br>/a.ʁi.vɛ/    | arrivions<br>/a.ʁi.vjɔ̃/      | arriviez<br>/a.ʁi.vje/      | arrivaient<br>/a.ʁi.vɛ/    |
|                 | past historic <sup>2</sup> | arrivai<br>/a.ʁi.vɛ/     | arrivas<br>/a.ʁi.va/     | arriva<br>/a.ʁi.va/      | arrivâmes<br>/a.ʁi.vam/       | arrivâtes<br>/a.ʁi.vat/     | arrivèrent<br>/a.ʁi.vɛʁ/   |
|                 | future                     | arriverai<br>/a.ʁi.vʁɛ/  | arriveras<br>/a.ʁi.vʁa/  | arrivera<br>/a.ʁi.vʁa/   | arriverons<br>/a.ʁi.vʁɔ̃/     | arriverez<br>/a.ʁi.vʁe/     | arriveront<br>/a.ʁi.vʁɔ̃/  |
|                 | conditional                | arriverais<br>/a.ʁi.vʁɛ/ | arriverais<br>/a.ʁi.vʁɛ/ | arriverait<br>/a.ʁi.vʁɛ/ | arriverions<br>/a.ʁi.və.ʁjɔ̃/ | arriveriez<br>/a.ʁi.və.ʁje/ | arriveraient<br>/a.ʁi.vʁɛ/ |





# Morphological Inflection

## ► In Spanish:

|            |             | singular   |   |                       | plural               |                      |                        |
|------------|-------------|------------|---|-----------------------|----------------------|----------------------|------------------------|
|            |             | 1st person | 2nd person                                    | 3rd person            | 1st person           | 2nd person           | 3rd person             |
|            |             | yo         | tú<br>vos                                     | él/ella/ello<br>usted | nosotros<br>nosotras | vosotros<br>vosotras | ellos/ellas<br>ustedes |
| indicative | present     | llego      | llegas <sup>tú</sup><br>llegás <sup>vos</sup> | llega                 | llegamos             | llegáis              | llegan                 |
|            | imperfect   | llegaba    | llegabas                                      | llegaba               | llegábamos           | llegabais            | llegaban               |
|            | preterite   | llegué     | llegaste                                      | llegó                 | llegamos             | llegasteis           | llegaron               |
|            | future      | llegaré    | llegarás                                      | llegará               | llegaremos           | llegaréis            | llegarán               |
|            | conditional | llegaría   | llegarías                                     | llegaría              | llegaríamos          | llegaríais           | llegarían              |
|            |             |            |   |                       |                      |                      |                        |





# Noun Inflection

- ▶ Not just verbs either; gender, number, case complicate things

| Declension of Kind <span>[hide ▲]</span> |          |      |                             |        |         |
|--|----------|------|-----------------------------|--------|---------|
|  | singular |      |                             | plural |         |
|  | indef.   | def. | noun                        | def.   | noun    |
| nominative                               | ein      | das  | Kind                        | die    | Kinder  |
| genitive                                 | eines    | des  | Kindes,<br>Kinds            | der    | Kinder  |
| dative                                   | einem    | dem  | Kind,<br>Kinde <sup>1</sup> | den    | Kindern |
| accusative                               | ein      | das  | Kind                        | die    | Kinder  |

- ▶ Nominative: I/he/she, accusative: me/him/her, genitive: mine/his/hers
- ▶ Dative: merged with accusative in English, shows recipient of something  
I taught the children <=> Ich unterrichte die Kinder  
I give the children a book <=> Ich gebe den Kindern ein Buch



# Irregular Inflection

---

- ▶ Common words are often irregular
  - ▶ I am / you are / she is
  - ▶ Je suis / tu es / elle est
  - ▶ Soy / está / es
- ▶ Less common words typically fall into some regular *paradigm* — these are somewhat predictable



# Agglutinating Languages

- Finnish/Hungarian (Finno-Ugric), also Turkish: what a preposition would do in English is instead part of the verb

|                       |                       | active          | passive     |
|-----------------------|-----------------------|-----------------|-------------|
| 1st                   |                       | halata          |             |
| long 1st <sup>2</sup> |                       | halatakseen     |             |
| 2nd                   | inessive <sup>1</sup> | halatessa       | halattaessa |
|                       | instructive           | halaten         | —           |
| 3rd                   | inessive              | halaamassa      | —           |
|                       | elative               | halaamasta      | —           |
|                       | illative              | halaamaan       | —           |
|                       | adessive              | halaamalla      | —           |
|                       | abessive              | halaamatta      | —           |
|                       | instructive           | halaaman        | halattaman  |
| 4th                   | nominative            | halaaminen      |             |
|                       | partitive             | halaamista      |             |
| 5th <sup>2</sup>      |                       | halaamaisillaan |             |

| indicative mood       |                 | perfect             |             | pluperfect          |                        |
|-----------------------|-----------------|---------------------|-------------|---------------------|------------------------|
| present tense         |                 |                     |             |                     |                        |
| person                | positive        | negative            | person      | positive            | negative               |
| 1st sing.             | halaan          | en halua            | 1st sing.   | olen halannut       | en ole halannut        |
| 2nd sing.             | halaat          | et halua            | 2nd sing.   | olet halannut       | et ole halannut        |
| 3rd sing.             | halaa           | ei halua            | 3rd sing.   | on halannut         | ei ole halannut        |
| 1st plur.             | halaa           | emme halua          | 1st plur.   | olemme halanneet    | emme ole halanneet     |
| 2nd plur.             | halatte         | ette halua          | 2nd plur.   | olette halanneet    | ette ole halanneet     |
| 3rd plur.             | halavat         | evät halua          | 3rd plur.   | ovat halanneet      | evät ole halanneet     |
| passive               | halataan        | ei halata           | passive     | on halattu          | ei ole halattu         |
| past tense            |                 |                     |             |                     |                        |
| person                | positive        | negative            | person      | positive            | negative               |
| 1st sing.             | halasin         | en halannut         | 1st sing.   | olin halannut       | en ollut halannut      |
| 2nd sing.             | halasit         | et halannut         | 2nd sing.   | olit halannut       | et ollut halannut      |
| 3rd sing.             | halasi          | ei halannut         | 3rd sing.   | oli halannut        | ei ollut halannut      |
| 1st plur.             | halasimme       | emme halanneet      | 1st plur.   | olimme halanneet    | emme olleet halanneet  |
| 2nd plur.             | halasitte       | ette halanneet      | 2nd plur.   | olitte halanneet    | ette olleet halanneet  |
| 3rd plur.             | halasivat       | evät halanneet      | 3rd plur.   | olivat halanneet    | evät olleet halanneet  |
| passive               | halattiin       | ei halattu          | passive     | oli halattu         | ei ollut halattu       |
| conditional mood      |                 |                     |             |                     |                        |
| person                | positive        | negative            | person      | positive            | negative               |
| 1st sing.             | halaisin        | en halaisi          | 1st sing.   | olisin halannut     | en olisi halannut      |
| 2nd sing.             | halaisit        | et halaisi          | 2nd sing.   | olisit halannut     | et olisi halannut      |
| 3rd sing.             | halaisi         | ei halaisi          | 3rd sing.   | olisi halannut      | ei olisi halannut      |
| 1st plur.             | halaisimme      | emme halaisi        | 1st plur.   | olisimme halanneet  | emme olisi halanneet   |
| 2nd plur.             | halaisitte      | ette halaisi        | 2nd plur.   | olisitte halanneet  | ette olisi halanneet   |
| 3rd plur.             | halaisivat      | evät halaisi        | 3rd plur.   | olisivat halanneet  | evät olisi halanneet   |
| passive               | halattaisiin    | ei halattaisi       | passive     | olisi halattu       | ei olisi halattu       |
| imperative mood       |                 |                     |             |                     |                        |
| person                | positive        | negative            | person      | positive            | negative               |
| 1st sing.             | —               | —                   | 1st sing.   | —                   | —                      |
| 2nd sing.             | halaa           | älä halua           | 2nd sing.   | ole halannut        | älä ole halannut       |
| 3rd sing.             | älkään halako   | älkäälä halako      | 3rd sing.   | älkään olo halannut | älkäälä olo halannut   |
| 1st plur.             | halakaa         | älkää halako        | 1st plur.   | olkaamme halanneet  | älkäämme olo halanneet |
| 2nd plur.             | halatkaa        | älkää halako        | 2nd plur.   | olkaa halanneet     | älkää olo halanneet    |
| 3rd plur.             | halakoot        | älköt olo halako    | 3rd plur.   | olkoot halanneet    | älköt olo halanneet    |
| passive               | halattakoon     | älköt olo halattako | passive     | olkoon halattu      | älköt olo halattu      |
| potential mood        |                 |                     |             |                     |                        |
| person                | positive        | negative            | person      | positive            | negative               |
| 1st sing.             | halan           | en halanne          | 1st sing.   | lenen halannut      | en liene halannut      |
| 2nd sing.             | halat           | et halanne          | 2nd sing.   | lenet halannut      | et liene halannut      |
| 3rd sing.             | halanee         | ei halanne          | 3rd sing.   | lenee halannut      | ei liene halannut      |
| 1st plur.             | halanemme       | emme halanne        | 1st plur.   | lenemme halanneet   | emme liene halanneet   |
| 2nd plur.             | halanette       | ette halanne        | 2nd plur.   | lenette halanneet   | ette liene halanneet   |
| 3rd plur.             | halannevat      | evät halanne        | 3rd plur.   | lenevät halanneet   | evät liene halanneet   |
| passive               | halan           | ei halanne          | passive     | lenee halattu       | ei liene halattu       |
| nominal forms         |                 |                     |             |                     |                        |
| infinitives           | active          | passive             | participles | active              | passive                |
| 1st                   | halata          | halattava           | present     | halassa             | halattava              |
| long 1st <sup>2</sup> | halatakseen     | halattavaksi        | past        | halannut            | halattu                |
| 2nd                   | halatessa       | halattaessa         | past, 1, 3  | halaama             | halaamaton             |
| inessive <sup>1</sup> | halaten         | —                   | negative    | —                   | —                      |
| instructive           | halaten         | —                   |             |                     |                        |
| inessive              | halaamassa      | —                   |             |                     |                        |
| elative               | halaamasta      | —                   |             |                     |                        |
| illative              | halaamaan       | —                   |             |                     |                        |
| adessive              | halaamalla      | —                   |             |                     |                        |
| abessive              | halaamatta      | —                   |             |                     |                        |
| instructive           | halaaman        | halattaman          |             |                     |                        |
| nominative            | halaaminen      | —                   |             |                     |                        |
| partitive             | halaamista      | —                   |             |                     |                        |
| 5th <sup>2</sup>      | halaamaisillaan | —                   |             |                     |                        |

halata: “hug”

illative: “into”

adessive: “on”

- Many possible forms — and in newswire data, only a few are observed



# Morphologically-Rich Languages

---

- ▶ Many languages spoken all over the world have much richer morphology than English
- ▶ CoNLL 2006 / 2007: dependency parsing + morphological analyses for ~15 mostly Indo-European languages
- ▶ SPMRL shared tasks (2013-2014): Syntactic Parsing of Morphologically-Rich Languages
- ▶ Word piece / byte-pair encoding models for MT are pretty good at handling these if there's enough data





# Morphologically-Rich Languages



MORGAN & CLAYPOOL PUBLISHERS

## Linguistic Fundamentals for Natural Language Processing

*100 Essentials from  
Morphology and Syntax*

Emily M. Bender

**SYNTHESIS LECTURES ON  
HUMAN LANGUAGE TECHNOLOGIES**

Graeme Hirst, *Series Editor*

- ▶ Great resources for challenging your assumptions about language and for understanding multilingual models!

# Morphological Analysis/Inflection



# Morphological Analysis

---

- ▶ In English, lexical features on words and word vectors are pretty effective
- ▶ In other languages, **lots** more unseen words due to rich morphology!  
Affects parsing, translation, ...
- ▶ When we're building systems, we probably want to know base form + morphological features explicitly
- ▶ How to do this kind of *morphological analysis*?





# Morphological Analysis: Hungarian

But the government does not recommend reducing taxes.

Ám a kormány egyetlen adó csökkentését sem javasolja .

n=singular / case=nominative / proper=no  
deg=positive / n=singular / case=nominative  
n=singular / case=nominative / proper=no  
n=singular / case=accusative / proper=no / pperson=3rd / pnumber=singular  
mood=indicative / t=present / p=3rd / n=singular / def=yes



# Morphological Analysis

---

- ▶ Given a word in context, need to predict what its morphological features are
- ▶ Basic approach: combines two modules:
  - ▶ Lexicon: tells you what possibilities are for the word
  - ▶ Analyzer: statistical model that disambiguates
- ▶ Models are largely CRF-like: score morphological features in context
- ▶ Lots of work on Arabic inflection (high amounts of ambiguity)



# Morphological Inflection

- ▶ Inverse task of analysis: given base form + features, inflect the word
- ▶ Hard for unknown words — need models that generalize

*w i n d e n* →

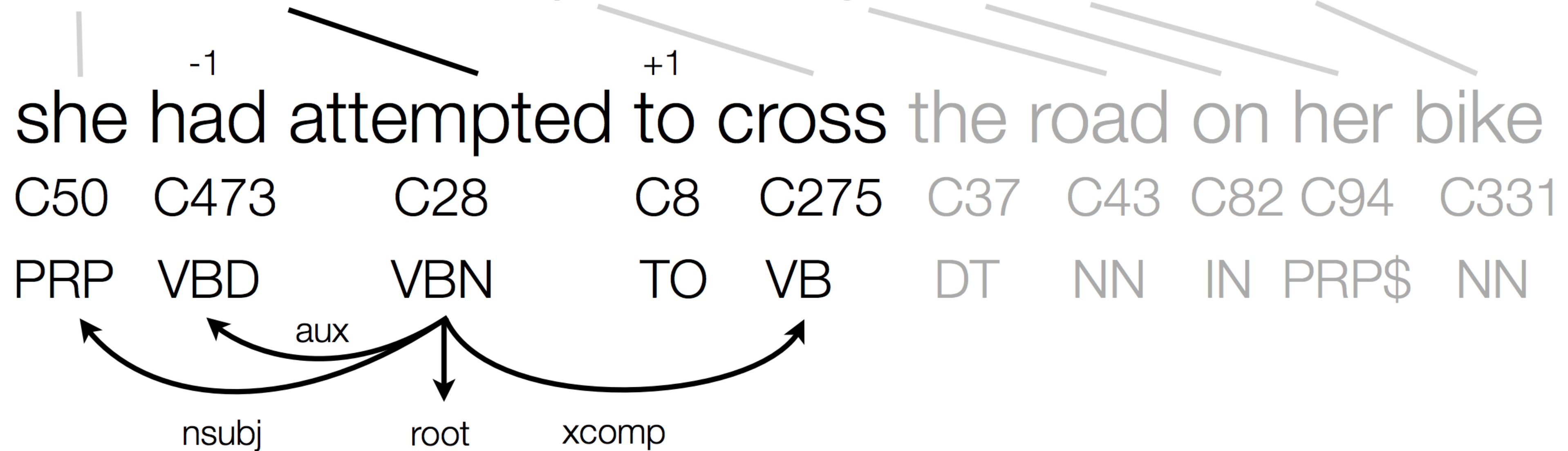
| conjugation of winden    |            |              |    |             |            | [hide ▲] |
|--------------------------|------------|--------------|----|-------------|------------|----------|
| infinitive               |            | winden       |    |             |            |          |
| present participle       |            | windend      |    |             |            |          |
| past participle          |            | gewunden     |    |             |            |          |
| auxiliary                |            | haben        |    |             |            |          |
|                          | indicative |              |    | subjunctive |            |          |
| present                  | ich winde  | wir winden   | i  | ich winde   | wir winden |          |
|                          | du windest | ihr windet   |    | du windest  | ihr windet |          |
|                          | er windet  | sie winden   |    | er winde    | sie winden |          |
| preterite                | ich wand   | wir wanden   | ii | ich wände   | wir wänden |          |
|                          | du wandest | ihr wandet   |    | du wändest  | ihr wändet |          |
|                          | er wand    | sie wanden   |    | er wände    | sie wänden |          |
| imperative               | winde (du) | windet (ihr) |    |             |            |          |
| composed forms of winden |            |              |    |             |            | [show ▼] |



# Morphological Inflection

σ:пытаться\_V + μ:mis-sfm-e

она **пыталась** пересечь пути на ее велосипеде



- ▶ Machine translation where phrase table is defined in terms of lemmas
- ▶ “Translate-and-inflect”: translate into uninflected words and predict inflection based on source side

Chahuneau et al. (2013)





# Other “Morphological” Analysis

- ▶ Word segmentation:  
some languages  
including Chinese are  
totally untokenized
- ▶ LSTMs over character  
embeddings / character  
bigram embeddings to  
predict word boundaries
- ▶ Having the right  
segmentation can help  
machine translation

冬天 (winter), 能 (can) 穿 (wear) 多少 (amount) 穿 (wear) 多少 (amount); 夏天 (summer), 能 (can) 穿 (wear) 多 (more) 少 (little) 穿 (wear) 多 (more) 少 (little)。

Without the word “夏天 (summer)” or “冬天 (winter)”, it is difficult to segment the phrase “能穿多少穿多少”.

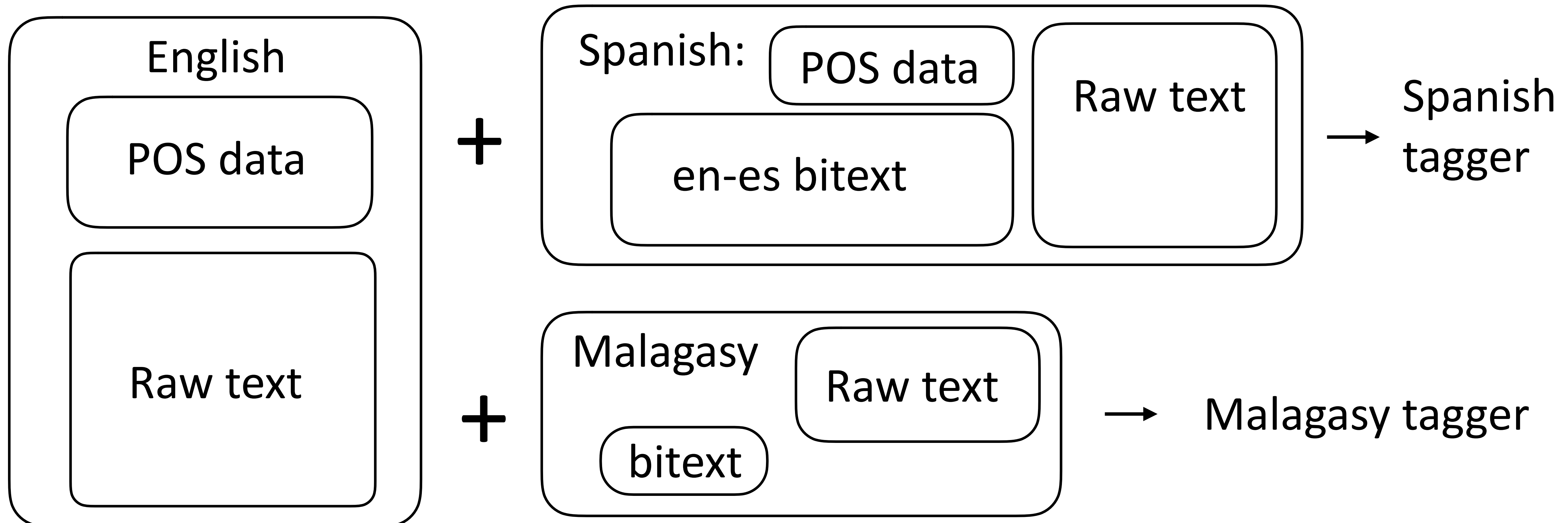
- separating nouns and pre-modifying adjectives:  
高血压 (*high blood pressure*)  
→ 高(*high*) 血压(*blood pressure*)
- separating compound nouns:  
民政部 (*Department of Internal Affairs*)  
→ 民政(*Internal Affairs*) 部(*Department*).

# Cross-Lingual Tagging and Parsing



# Cross-Lingual Tagging

- ▶ Labeling POS datasets is expensive
- ▶ Can we transfer annotation from *high-resource* languages (English, etc.) to *low-resource* languages?

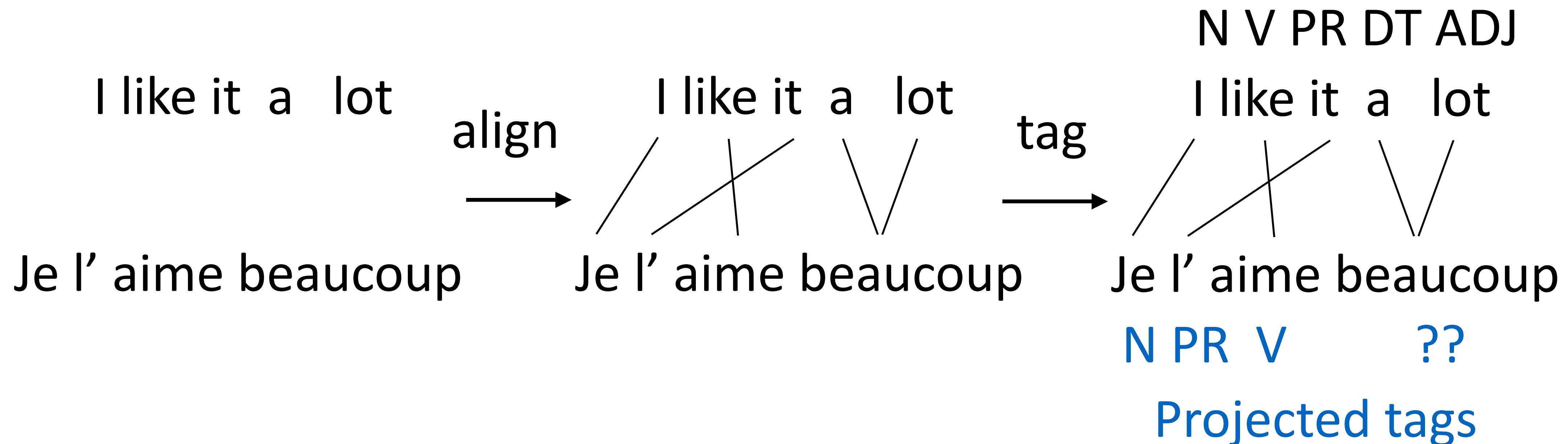






# Cross-Lingual Tagging

- Can we leverage word alignment here?

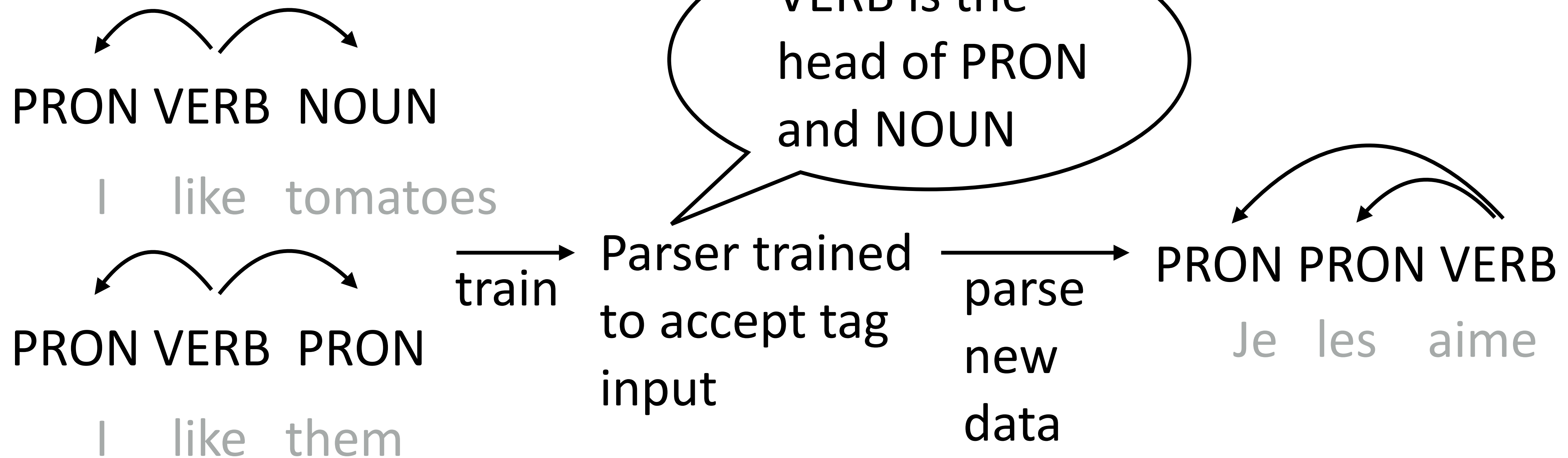


- Tag with English tagger, project across bitext, train French tagger?  
Works pretty well



# Cross-Lingual Parsing

- ▶ Now that we can POS tag other languages, can we parse them too?
- ▶ Direct transfer: train a parser over POS sequences in one language, then apply it to another language



McDonald et al. (2011)



# Cross-Lingual Parsing

|     | best-source |          | avg-source<br>gold-POS | gold-POS   |             | pred-POS   |             |
|-----|-------------|----------|------------------------|------------|-------------|------------|-------------|
|     | source      | gold-POS |                        | multi-dir. | multi-proj. | multi-dir. | multi-proj. |
| da  | it          | 48.6     | 46.3                   | 48.9       | 49.5        | 46.2       | 47.5        |
| de  | nl          | 55.8     | 48.9                   | 56.7       | 56.6        | 51.7       | 52.0        |
| el  | en          | 63.9     | 51.7                   | 60.1       | 65.1        | 58.5       | 63.0        |
| es  | it          | 68.4     | 53.2                   | 64.2       | 64.5        | 55.6       | 56.5        |
| it  | pt          | 69.1     | 58.5                   | 64.1       | 65.0        | 56.8       | 58.9        |
| nl  | el          | 62.1     | 49.9                   | 55.8       | 65.7        | 54.3       | 64.4        |
| pt  | it          | 74.8     | 61.6                   | 74.0       | 75.6        | 67.7       | 70.3        |
| sv  | pt          | 66.8     | 54.8                   | 65.3       | 68.0        | 58.3       | 62.1        |
| avg |             | 63.7     | 51.6                   | 61.1       | 63.8        | 56.1       | 59.3        |

► Multi-dir: transfer a parser trained on several source treebanks to the target language

► Multi-proj: more complex annotation projection approach

McDonald et al. (2011)

# Cross-Lingual Word Representations

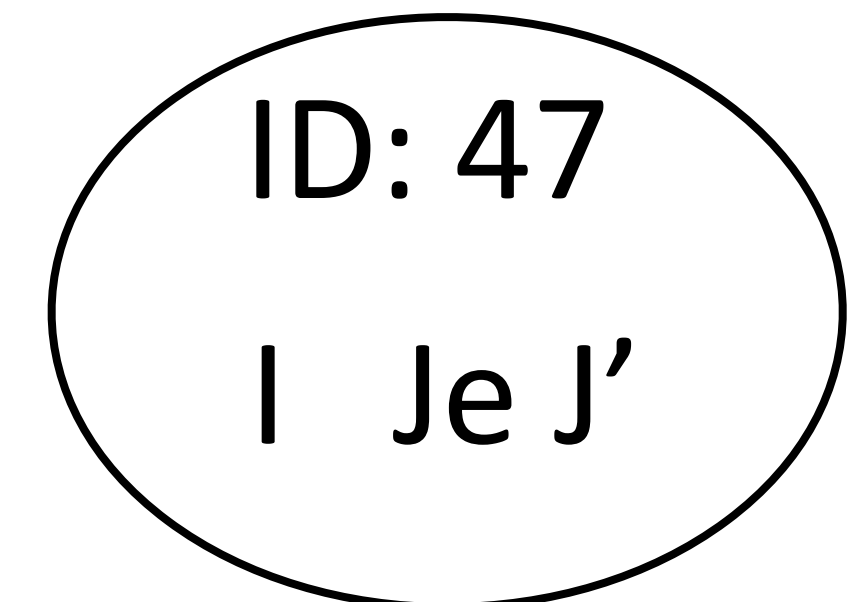
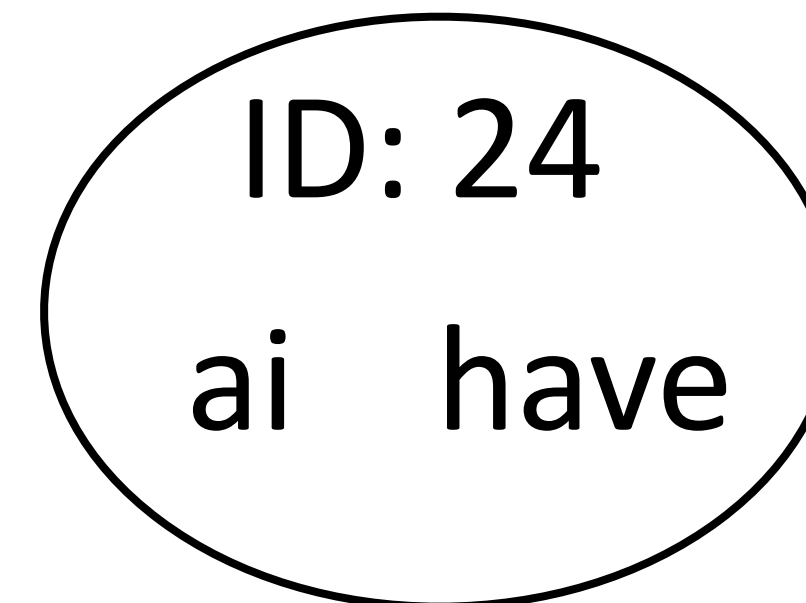


# Multilingual Embeddings

- ▶ Input: corpora in many languages. Output: embeddings where similar words *in different languages* have similar embeddings

I have an apple  
47 24 18 427

J' ai des oranges  
47 24 89 1981



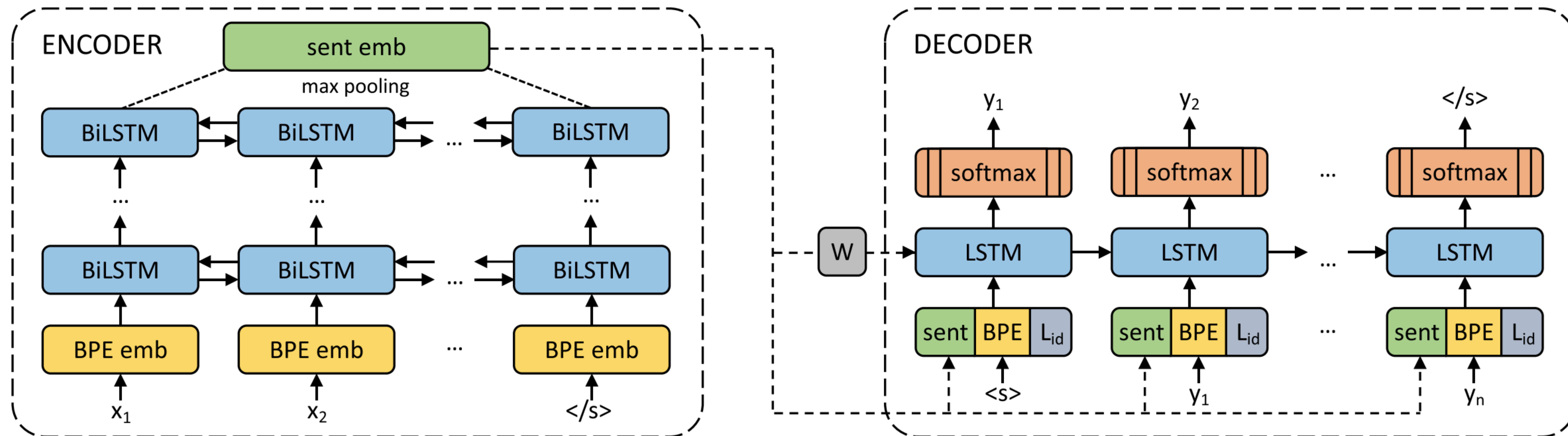
- ▶ multiCluster: use bilingual dictionaries to form clusters of words that are translations of one another, replace corpora with cluster IDs, train “monolingual” embeddings over all these corpora
- ▶ Works okay but not all that well

Ammar et al. (2019)





# Multilingual Sentence Embeddings



- ▶ Form BPE vocabulary over all corpora (50k merges); will include characters from every script
- ▶ Take a bunch of bitexts and train an MT model between a bunch of language pairs with shared parameters, use  $W$  as sentence embeddings

Artetxe et al. (2019)



# Multilingual Sentence Embeddings

|  |             | EN          | EN → XX     |             |             |             |             |             |             |             |             |             |             |             |             |             |
|--|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
|  |             |             | fr          | es          | de          | el          | bg          | ru          | tr          | ar          | vi          | th          | zh          | hi          | sw          | ur          |
| <b>Zero-Shot Transfer, one NLI system for all languages:</b> |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |             |
| Conneau et al.<br>(2018b)                                    | X-BiLSTM    | 73.7        | 67.7        | 68.7        | 67.7        | 68.9        | 67.9        | 65.4        | 64.2        | 64.8        | 66.4        | 64.1        | 65.8        | 64.1        | 55.7        | 58.4        |
|  | X-CBOW      | 64.5        | 60.3        | 60.7        | 61.0        | 60.5        | 60.4        | 57.8        | 58.7        | 57.5        | 58.8        | 56.9        | 58.8        | 56.3        | 50.4        | 52.2        |
| BERT uncased*  | Transformer | <u>81.4</u> | –           | <u>74.3</u> | 70.5        | –           | –           | –           | –           | 62.1        | –           | –           | 63.8        | –           | –           | 58.3        |
| Proposed method  | BiLSTM      | 73.9        | <b>71.9</b> | 72.9        | <u>72.6</u> | <b>72.8</b> | <b>74.2</b> | <b>72.1</b> | <b>69.7</b> | <b>71.4</b> | <b>72.0</b> | <b>69.2</b> | <u>71.4</u> | <b>65.5</b> | <b>62.2</b> | <u>61.0</u> |

- ▶ Train a system for NLI (entailment/neutral/contradiction of a sentence pair) on English and evaluate on other languages





# Multilingual BERT

- ▶ Take top 104 Wikipedias, train BERT on all of them simultaneously
- ▶ What does this look like?

Beethoven may have proposed unsuccessfully to Therese Malfatti, the supposed dedicatee of "Für Elise"; his status as a commoner may again have interfered with those plans.

当人们在马尔法蒂身后发现这部小曲的手稿时，便误认为上面写的是“Für Elise”（即《给爱丽丝》）[51]。

Кита́й (официально — Кита́йская Наро́дная Респу́блика, сокращённо — КНР; кит. трад. 中華人民共和國, упр. 中华人民

Devlin et al. (2019)



# Multilingual BERT: Results

| Fine-tuning \ Eval | EN           | DE           | NL           | ES           |
|--------------------|--------------|--------------|--------------|--------------|
| EN                 | <b>90.70</b> | 69.74        | 77.36        | 73.59        |
| DE                 | 73.83        | <b>82.00</b> | 76.25        | 70.03        |
| NL                 | 65.46        | 65.68        | <b>89.86</b> | 72.10        |
| ES                 | 65.38        | 59.40        | 64.39        | <b>87.18</b> |

Table 1: NER F1 results on the CoNLL data.

| Fine-tuning \ Eval | EN           | DE           | ES           | IT           |
|--------------------|--------------|--------------|--------------|--------------|
| EN                 | <b>96.82</b> | 89.40        | 85.91        | 91.60        |
| DE                 | 83.99        | <b>93.99</b> | 86.32        | 88.39        |
| ES                 | 81.64        | 88.87        | <b>96.71</b> | 93.71        |
| IT                 | 86.79        | 87.82        | 91.28        | <b>98.11</b> |

Table 2: POS accuracy on a subset of UD languages.

- ▶ Can transfer BERT directly across languages with some success
- ▶ ...but this evaluation is on languages that all share an alphabet



# Multilingual BERT: Results

|    | HI          | UR          |    | EN          | BG          | JA          |
|----|-------------|-------------|----|-------------|-------------|-------------|
| HI | <b>97.1</b> | 85.9        | EN | <b>96.8</b> | 87.1        | 49.4        |
| UR | 91.1        | <b>93.8</b> | BG | 82.2        | <b>98.9</b> | 51.6        |
|    |             |             | JA | 57.4        | 67.2        | <b>96.5</b> |

Table 4: POS accuracy on the UD test set for languages with different scripts. Row=fine-tuning, column=eval.

- ▶ Urdu (Arabic script) => Hindi (Devanagari). Transfers well despite different alphabets!
- ▶ Japanese => English: different script and very different syntax



# Multilingual BERT

- mBERT doesn't require word piece overlap between things to do well (but going from 0 overlap to some overlap helps a lot)

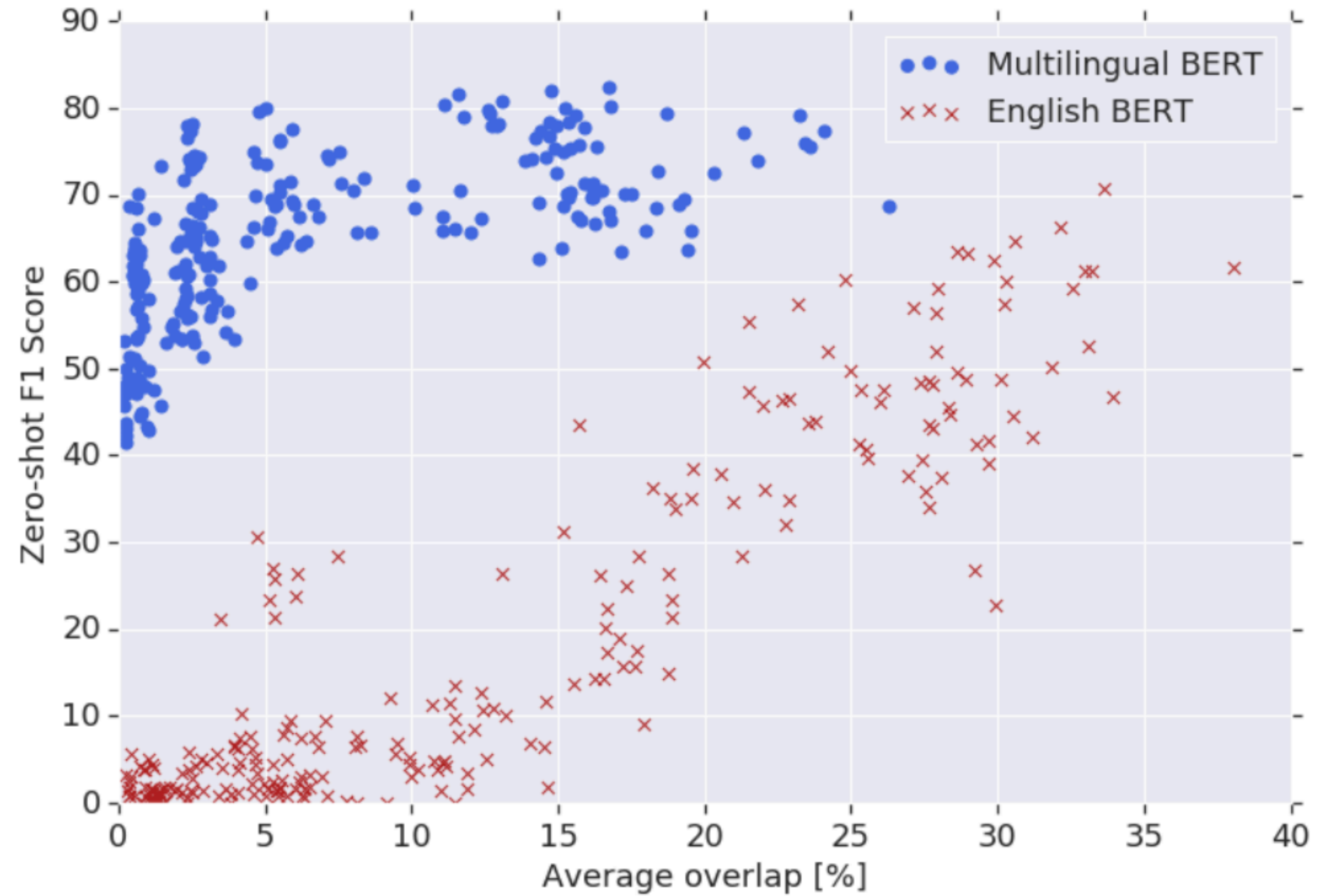


Figure 1: Zero-shot NER F1 score versus entity word piece overlap among 16 languages. While performance

Pires et al. (2019)





# Where are we now?

---

- ▶ Universal dependencies: treebanks (+ tags) for 70+ languages
- ▶ Many languages are still small, so projection techniques may still help
- ▶ More corpora in other languages, less and less reliance on structured tools like parsers, and pretraining on unlabeled data means that performance on other languages is better than ever
- ▶ BERT has pretrained multilingual models that seem to work pretty well



# Takeaways

---

- ▶ Many languages have richer morphology than English and pose distinct challenges
- ▶ Problems: how to analyze rich morphology, how to generate with it
- ▶ Can leverage resources for English using bitexts
- ▶ Next time: wrapup + discussion of ethics