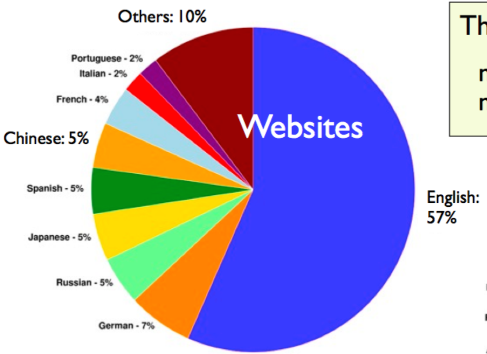
Treebank Translation for Cross-Lingual Parser Induction

Jörg Tiedemann[1] **Željko Agić**[2] Joakim Nivre[1]

[1]Department of Linguistics and Philology, Uppsala University
[2]Department of Linguistics, University of Potsdam

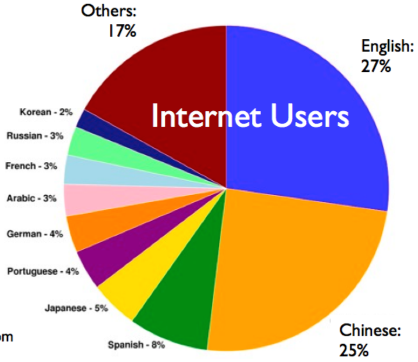
CoNLL 2014, 2014-06-27

Motivation



The World is not English only

many languages on the Web;
most are under-resourced



> 2 billion Internet users
> 12 billion indexed web pages

Sources: W3Techs.com, Internet World Stats, WorldWideWebSize.com

Motivation

There are languages out there that require processing, but lack the required resources (Bender, 2011; Bender, 2013).

- most of World languages under-resourced (META-NET LWPs, 2012)
- uniform language processing
  - lack of resources
  - *balkanization* – the one-scheme-per-language rule

- we focus on dependency parsing
- Is there a dependency treebank for... Croatian? Slovene?

Approaches

- annotation projection
- model transfer

- unsupervised
  - not addressed here
  - performance generally below previous two

Annotation projection

- take a parallel corpus
- word-align it
- parse it for syntactic dependencies
- project the annotation via alignment

- some variations
  - one side of parallel corpus is a treebank (rare)
  - word alignments are manual (rare)
  - usually relies on automatic word alignment and dependency parsing

    (Yarowsky et al., 2001; Hwa et al., 2005)

✓ language-specific features

✗ noise from parsing, alignment, projection

Model transfer

- train model on source language treebank
- rely on common features
- apply model on target language

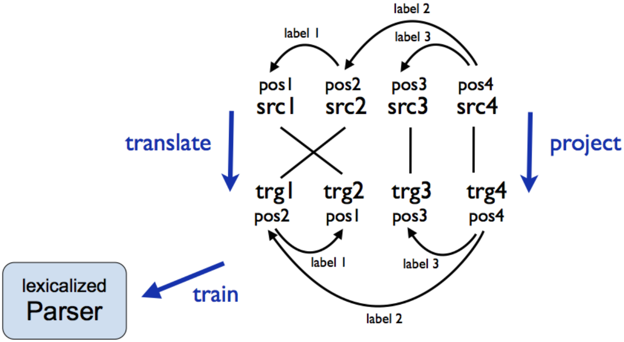
- approaches
  - delexicalization (Zeman & Resnik, 2008; McDonald et al., 2013)
  - data point selection (Søgaard, 2011)
  - multi-source transfer (McDonald et al., 2011)
  - cross-lingual word clusters (Täckström et al., 2012)

✓ no resources required for target, no alignment and projection noise

✗ poor feature model

Treebank translation

- train a source-target SMT system
- translate source treebank into target language
- project annotations
- train dependency parser on synthetic treebank
- do parsing



Treebank translation

- differs from annotation projection
  - ✓ no source parsing noise
  - ✓ word alignment not separated, better for synthetic data
- and from model transfer
  - ✓ lexicalization
  - ✓ allows full feature set in target language
  - ✓ no assumptions on language universals
- potential issues
  - ✗ annotation projection noise still remains
  - ✗ quality of SMT

Setup

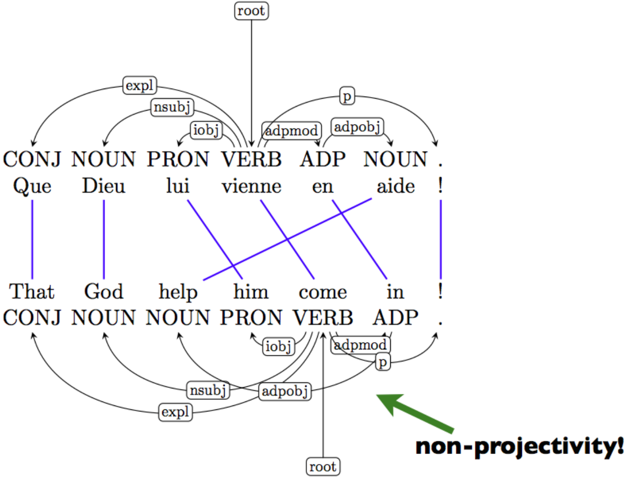
- treebanks
  - Google Universal Treebanks 1.0 (McDonald et al., 2013)
  - Universal POS (Petrov et al., 2012)
  - (adapted) Stanford Dependencies
  - excluded Korean as outlier: 5 languages
  - reliable cross-lingual dependency parsing assessment
  - existing train-dev-test split
- parsing
  - MaltParser (Nivre et al., 2007)
  - MaltOptimizer chooses optimal configuration (Ballesteros & Nivre, 2012)
- translation
  - Moses (Koehn et al., 2007), Europarl (Koehn, 2005)

Translation

- three scenarios
  - dictionary lookup
    - replace each word by default translation
    - no reordering
  - word-to-word
    - single-word translation table
    - distance-based reordering
    - 5-gram language model
  - phrase-based
    - standard phrase-based SMT model
- effects on non-projectivity
- projection requirements

Projection

- trivial for dictionary lookup
- same for word-to-word translation, non-projectivity occurs



Projection

- projection for phrase-based models
- multi-word alignments (m:n)
- labels must be projected as well
- one solution: dummy nodes (Hwa et al., 2005)

- our approach
  - use SMT phrase membership and phrase alignment information
  - use tree attachment heuristics

Projection

Input: source tree S, target sentence T, word alignment A, phrase segmentation P
Output: syntactic heads head[], word attributes attr[]

```
treeSize = max_distance_to_root(S) ;
attr = [] ;
head = [] ;
for t ∈ T do
    if is_unaligned_trg(t,A) then
        for t' ∈ in_trg_phrase(t,P) do
            [s_x,..,s_y] = aligned_to(t') ;
            ŝ = find_highest([s_x,..,s_y],S) ;
            t̂ = find_aligned(ŝ,S,T,A) ;
            attr[t] = DUMMY ;
            head[t] = t̂ ;
        end
    else
        [s_x,..,s_y] = aligned_to(t) ;
        s = find_highest([s_x,..,s_y],S) ;
        attr[t] = attr(s) ;
        ŝ = head_of(s,S) ;
        t̂ = find_aligned(ŝ,S,T,A) ;
        if t̂ == t then
            [s_x,..,s_y] = in_src_phrase(s,P) ;
            s* = find_highest([s_x,..,s_y],S) ;
            ŝ = head_of(s*,S) ;
            t̂ = find_aligned(ŝ,S,T,A) ;
            head[t] = t̂ ;
        end
    end
end
```

use phrase segmentation

attach to highest node

function: find_aligned:

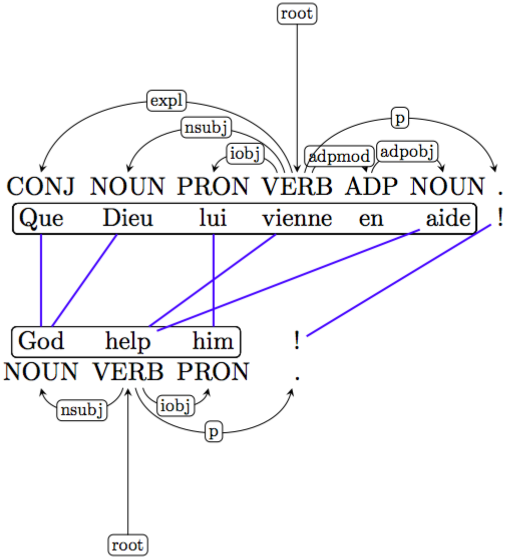
Input: node s, source tree S with root ROOT, target sentence T, word alignment A
Output: node t*

```
if s == ROOT then
    return ROOT ;
end
while is_unaligned_src(s,A) do
    s = head_of(s,S) ;
    if s == ROOT then
        return ROOT ;
    end
end
p = 0 ;
t* = undef ;
for t' ∈ aligned(s,A) do
    if position(t',T) > p then
        t* = t' ;
        p = position(t',T) ;
    end
end
return t* ;
```

walk up the tree if unaligned

heuristics for multiple targets: take right-most

Projection


root
expl
nsubj
iobj
p
adpmod
adpobj
CONJ NOUN PRON VERB ADP NOUN .
Que Dieu lui vienne en aide !
God help him !
NOUN VERB PRON .
nsubj
iobj
p
root

Results

Baseline

Monolingual					
	de	en	es	fr	sv
	72.13	87.50	78.54	77.51	81.28
Delexicalized					
	de	en	es	fr	sv
de	62.71	43.20	46.09	46.09	50.64
en	46.62	77.66	55.65	56.46	**57.68**
es	44.03	46.73	68.21	**57.91**	53.82
fr	43.91	46.75	**59.65**	67.51	52.01
sv	**50.69**	**49.13**	53.62	51.97	70.22
McDonald et al. (2013)					
	de	en	es	fr	sv
de	64.84	47.09	48.14	49.59	53.57
en	48.11	78.54	56.86	58.20	**57.04**
es	45.52	47.87	70.29	**63.65**	53.09
fr	45.96	47.41	**62.56**	73.37	52.25
sv	**52.19**	**49.71**	54.72	54.96	70.90

Results

Delexicalized models

	de	en	es	fr	sv
Word-to-word					
	de	en	es	fr	sv
de	–	48.12 (4.92)	50.84 (4.75)	52.92 (6.83)	55.52 (4.88)
en	49.53 (2.91)	–	57.41 (1.76)	**58.53** (2.07)	**57.82** (0.14)
es	45.48 (1.45)	48.46 (1.73)	–	58.29 (0.38)	55.25 (1.43)
fr	46.59 (2.68)	47.88 (1.13)	**59.72** (0.07)	–	52.31 (0.30)
sv	**52.16** (1.47)	**49.14** (0.01)	56.50 (2.88)	56.71 (4.74)	–
Phrase-based					
	de	en	es	fr	sv
de	–	45.43 (2.23)	47.26 (1.17)	49.14 (3.05)	53.37 (2.73)
en	49.16 (2.54)	–	57.12 (1.47)	**58.23** (1.77)	**58.23** (0.55)
es	46.75 (2.72)	46.82 (0.09)	–	58.22 (0.31)	54.14 (0.32)
fr	48.02 (4.11)	**49.06** (2.31)	**60.23** (0.58)	–	55.24 (3.23)
sv	**50.96** (0.27)	$46.12^{-3.01}$	55.95 (2.33)	54.71 (2.74)	–

Results

Lexicalized models

Lookup

	de	en	es	fr	sv
de	–	48.63 (5.43)	52.66 (6.57)	52.06 (5.97)	58.78 (8.14)
en	48.59 (1.97)	–	57.79 (2.14)	57.80 (1.34)	**62.21** (4.53)
es	47.36 (3.33)	49.13 (2.40)	–	**62.24** (4.33)	57.50 (3.68)
fr	47.57 (3.66)	**54.06** (7.31)	**66.31** (6.66)	–	57.73 (5.72)
sv	**51.88** (1.19)	48.84 (0.29)	54.74 (1.12)	52.95 (0.98)	–

Word-to-word

	de	en	es	fr	sv
de	–	51.86 (3.74)	55.90 (5.06)	57.77 (4.85)	61.65 (6.13)
en	**53.80** (4.27)	–	60.76 (3.35)	63.32 (4.79)	**62.93** (5.11)
es	49.94 (4.46)	49.93 (1.47)	–	**65.60** (7.31)	59.22 (3.97)
fr	52.07 (5.48)	**54.44** (6.56)	**65.63** (5.91)	–	57.67 (5.36)
sv	53.18 (1.02)	50.91 (1.77)	60.82 (4.32)	59.14 (2.43)	–

Phrase-based

	de	en	es	fr	sv
de	–	50.89 (5.46)	52.54 (5.28)	54.99 (5.85)	59.46 (6.09)
en	**53.71** (4.55)	–	60.70 (3.58)	62.89 (4.66)	**64.01** (5.78)
es	49.59 (2.84)	48.35 (1.53)	–	**64.88** (6.66)	58.99 (4.85)
fr	51.83 (3.81)	**53.81** (4.75)	**65.55** (5.32)	–	59.01 (3.77)
sv	53.22 (2.26)	49.06 (2.94)	58.41 (2.46)	58.04 (3.33)	–

Conclusions

- substantial improvements
  - delexicalized up to +6.38 LAS
  - lexicalized up to +7.31 LAS
- phrase-based projection fails to deliver
  - quality of SMT
  - unreliable POS mappings, link ambiguity
  - no tree constraints
- overall results very positive
  - lexical features
  - reordering
  - per-language parser optimization
- future work
  - better translation
  - better projection (Tiedemann, 2014)
  - multi-synthetic-source transfer using n-best lists
  - closely related languages (Agić et al., 2012)

Thank you for your attention. ☺

Non-projectivity

Original					
	de	en	es	fr	sv
	14.0	0.00	7.90	13.3	4.20
Word-to-word					
	de	en	es	fr	sv
de	–	49.1	62.6	52.8	60.4
en	43.3	–	27.6	34.8	0.00
es	54.9	25.1	–	12.3	18.3
fr	68.2	39.6	32.8	–	57.8
sv	34.1	5.20	21.6	33.7	–
Phrase-based					
	de	en	es	fr	sv
de	–	51.5	57.3	58.8	46.8
en	49.3	–	50.3	61.7	14.6
es	65.9	66.7	–	62.8	49.0
fr	58.0	53.7	44.7	–	38.2
sv	43.9	43.6	49.6	57.1	–

Link ambiguity