Inferring syntactic rules for word alignment through Inductive Logic Programming

Sylwia Ozdowska, Vincent Claveau

CLLE-ERSS - Univ. of Toulouse
Toulouse, France

IRISA-CNRS
Rennes, France

May 19, 2010

Word alignment

Definition and use

- link occurrences of words (or phrases) that are in a translation relationship in parallel corpora
- usefulness of word alignment (Véronis 00)
  - acquisition of bilingual lexical resources, machine translation, cross-lingual information retrieval...

Existing techniques

- most approaches:
  - statistical alignment models (Brown *et al.* 93)
  - lexicon-based alignment models (Gale & Church 91)
- growing interest for syntax-informed models (Wu 00 ; Yamada & Knight 01 ; Gildea 03 ; Lin & Cherry 03)

Syntax and alignment

Debili & Zribi's hypothesis (96)

- if two words are translations of each other in aligned sentences, then their respective governors and dependents may be translations of each other

ALIBI (Ozdowska, 06)

- rule-based system for English/French
- principle: from two aligned anchor words (AW), the alignment link is projected to syntactically connected words

Syntax and alignment

Debili & Zribi's hypothesis (96)

- if two words are translations of each other in aligned sentences, then their respective governors and dependents may be translations of each other

ALIBI (Ozdowska, 06)

- rule-based system for English/French
- principle: from two aligned anchor words (AW), the alignment link is projected to syntactically connected words

The Community banned imports of ivory

La Communauté a interdit l'importation d'ivoire

Syntax and alignment

Debili & Zribi's hypothesis (96)

- if two words are translations of each other in aligned sentences, then their respective governors and dependents may be translations of each other

ALIBI (Ozdowska, 06)

- rule-based system for English/French
- principle: from two aligned anchor words (AW), the alignment link is projected to syntactically connected words

suj

The *Community* *banned* *imports of ivory*

|

La *Communauté* *a interdit* *l'importation d'ivoire*

suj

Syntax and alignment

Debili & Zribi's hypothesis (96)

- if two words are translations of each other in aligned sentences, then their respective governors and dependents may be translations of each other

ALIBI (Ozdowska, 06)

- rule-based system for English/French
- principle: from two aligned anchor words (AW), the alignment link is projected to syntactically connected words

suj

The *Community* ***banned*** *imports of ivory*

La *Communauté* ***a interdit*** *l'importation d'ivoire*

suj

Syntax and alignment

Syntactic propagation rules

- key component of the alignment system
- isomorphism (identical syntactic path): V-subj-N / V-subj-N

subj

The ***Community*** ***banned*** *imports of ivory*

La ***Communauté a interdit*** *l'importation d'ivoire*

subj

Syntax and alignment

Syntactic propagation rules

- key component of the alignment system
- isomorphism (identical syntactic path): V-subj-N / V-subj-N
- non-isomorphism (compatible pattern): V-obj-N / V-pp+pcomp-N

obj

. . . ***affects*** *cell* ***stability***

. . . ***intervient*** *sur la* ***stabilité*** *des cellules*

pp pcomp

Syntax and alignement

Syntactic propagation rules

- key component of the alignment system
- isomorphism (identical syntactic path): V-subj-N / V-subj-N
- non-isomorphism (compatible pattern): V-obj-N / V-pp+pcomp-N

Manual-encoding of the rules

- yields good results...
- ... yet defining these propagation rules is an issue
  - necessitate an expert in both languages
  - tedious task to be carried out for any new pair of languages, of parsers...

$\Rightarrow$ **machine learning of the propagation rules**

Machine learning of alignment rules

Supervised approach

- examples are pairs of words, linked by a syntactic path in both languages

Inductive Logic Programming (ILP)

- highly expressive, symbolic ML technique (Muggleton 95)
  - examples and output in first order logic (Prolog)
- natural way to encode relations and external knowledge
  - eg. translation and syntactic relations with simple predicates: x is the subject of y = subj(x, y)
- outputs human readable rules, making a linguistic analysis possible

Inductive Logic Programming

Theoretical framework of ILP

- infer a set of rules H (Horn clauses)...
- ...from examples E^+ (and possibly counter-examples E^-)
- ...and a Background Knowledge B
- ...such as $B \wedge H \wedge E^- \not\models \square$ and $B \wedge H \models E^+$

In our case

- H: syntactic propagation rules
- E^+: pairs of AW (no counter-examples)
- B: dependency relations and AW

Machine learning of alignment rules

In practice

- training data
  - aligned sentence

private	*sector*	*companies* /	*les*	*entreprises*	*du*	*secteur*	*privé*
e1	e2	e3	f1	f2	f3	f4	f5

  - dependency relations and AW in *B*

adj(e2,e1).	det(f2,f1).	pcomp(f3,f4).	aw(e2,f4).
nn(e3,e2).	pp(f2,f3).	adj(f4,f5).	aw(e3,f2).

- several rules generated for each example, organized in a lattice
  - for ex., align(E,F) :- nn(E,E2), pp(F,F3), pcomp(F3,F4), aw(E2,F4).

. . . *E2* *. . .* *E* *. . .*

nn

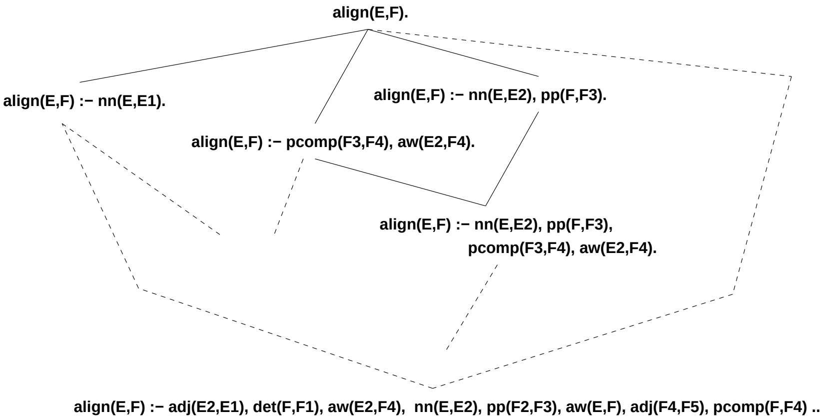
. . . *F* *. . .* *F3* *. . .* *F4* *. . .*

pp pcomp

Machine learning of alignment rules

Search lattice built on one example

- each rule of the lattice is scored wrt the other examples
- the best one is kept in *H*



The whole picture

Alignment algorithm

1. generate the examples: anchoring
   - cognates: string similarity (Fluhr *et al.* 00)
   - lexicon: simple cooccurrence model (Gale & Church 92)
2. parse the bitext
   - Syntex FR and Syntex EN (Bourigault 07)
3. infer propagation rules with ILP
   - ALEPH implementation (Srinivasan 01)
4. apply the rules to any bitext (after parsing and anchoring)
5. consider found alignments as anchors and goto 4

Experiments

Questions about ILP

- performance for the alignment task?
- interpretability of the inferred rules?

Questions about training data

- influence of the type of the training corpus?
- influence of the size of the training corpus?

Performance evaluation

Evaluation framework

- training dataset
  - HANSARD corpus (RALI, Univ. of Montreal)
  - Canadian parliamentary debates
  - 1000 sentences used for the training
- test set: HLT'03 dataset
  - 447 sentences from the Hansards ($\neq$ training corpus)
  - sure alignments S (inter-annotator agreement on S) and probable alignments P (multi-word expressions, free translations...)
- evaluation in precision (P), recall (R) and f-measure (F)

Performance evaluation

Results on S alignments from HLT'03 data set

System	ALIBI	ILP	Ralign	XRCE	BiBr	ProAlign
P	0.89	0.82	0.72	0.55	0.63	0.72
R	0.67	0.74	0.81	0.93	0.74	0.91
F	0.76	0.78	0.76	0.69	0.68	0.80

- Performance comparable with existing alignment systems (Mihalcea & Pedersen 03)
  - higher P
  - lower R

Performance evaluation

Cause of errors

Misalignments

- mostly caused by parsing errors
  - adjective *federal* was wrongly attached to *carpenters* leading to the misalignment *carpenter* / *gouvernement* in *federal government carpenters get $ 6.42* / *Les menuisiers du gouvernement fédéral touchent $ 6.42*.
- caused by overgeneralization
  - *gouvernement* and *legislation* are misaligned in the sentence pair: *good legislation has been brought in by Liberal governments* / *les gouvernements libéraux ont apporté de bonnes mesures législatives*.

Non detected alignments

- lack of anchor pairs and of dependency relations

Type of training corpus 1/2

Corpora

- HANSARD
- INRA
  - Institut National de la Recherche Agronomique
  - research and popular science articles on agronomy
  - $\sim$ 300 000 word tokens
- JOC
  - ARCADE Project (Véronis & Langlais 00)
  - various questions and answers dealt with at the European Commission
  - $\sim$ 400 000 word tokens
- 1 000 sentences for each corpus used for training (separately)

Type of training corpus 2/2

Performance on HLT'03 test set

Training corpus	HANSARD	JOC	INRA
P	82.08%	80.65%	**83.16%**
R	74.09%	**74.10%**	66.90%
F	**77.88%**	77.20%	74.15%

- Little differences with respect to the type of training corpus (except R on INRA)
- F-measure slightly improves if training and test are done on the same type of corpus

Inferred rules

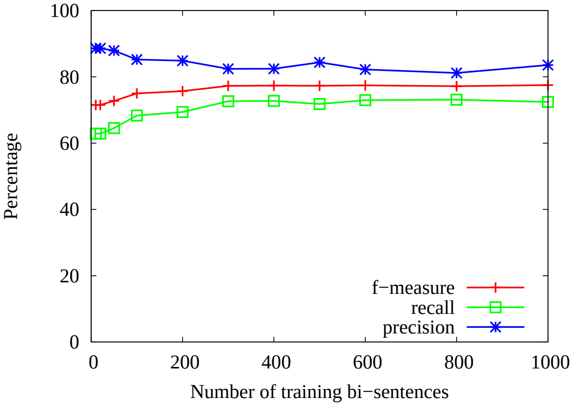
Genericity

- $\sim$ 60 rules learned from each corpus of 1000 sentences
  - 38 rules shared across the three corpora
  - 13 to 21 corpus-specific rules

Comparison with human-generated rules

- all identical rules encoded in ALIBI were inferred
- most of compatible rules encoded in ALIBI were inferred
- new rules not encoded in ALIBI were found

Size of the training corpus



- 300 to 1000 sentences: little variation in P and R
- $<$ 300: P increases and R decreases
- 10 sentences: 70% f-measure

Concluding remarks

About our approach

- fully automatic approach
  - supervised ML approach bootstrapped by the generation of anchors
- yields good performance
- inferred rules give insights on case of isomorphisms and non-isomorphisms between the two languages

No free lunch

- approach chiefly based on syntax
  - makes the most of knowledge embedded in parsers, thus requires few training data
  - dependent on the existence and quality of the parsers

Pespectives

Improvements

- enrich Background knowledge
  - add information like PoS, lemmas...
  - use the score from statistical alignment approaches
- find strategies to deal with partial syntactic analysis
- extension to dependency tree alignment

Application

- portability to different parsers
- portability to different language pairs

Inferring syntactic rules for word alignment through Inductive Logic Programming

Sylwia Ozdowska, Vincent Claveau

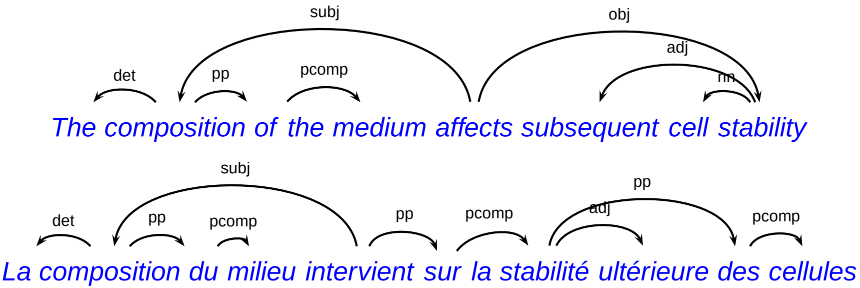
CLLE-ERSS - Univ. of Toulouse
Toulouse, France

IRISA-CNRS
Rennes, France

May 19, 2010

Parsers

- SYNTEX fr and SYNTEX en (Bourigault 07)
  - input : POS tagged sentences (TreeTagger (Schmidt 94))
  - output : dependency relations for each sentence



- Both parsers designed according to the same architecture
- Performance: SYNTEX fr > SYNTEX en

Corpora

- INRA
  - Institut National de la Recherche Agronomique
  - research and popular science articles on agronomy
  - $\sim$ 300 000 word tokens
- JOC
  - ARCADE Project (Véronis & Langlais 00)
  - various questions and answers dealt with at the European Commission
  - $\sim$ 400 000 word tokens
- HANSARD
  - RALI (University of Montreal)
  - Canadian parliamentary debates
  - $\sim$ 250 000 word tokens

Evaluation of overall performance

Method

- Evaluation of precision (P), recall (R) and f-measure (F)
- Cross-corpus evaluation
- Human annotation task
  - 120 test sentences for each corpus
  - annotation guidelines (Melamed 98, Véronis 98)
  - 3 human judges
- Human annotation output for each corpus
  - 60 sentences annotated by 2 persons
  - 60 sentences annotated by 1 person

Human annotation task

Inter-annotator agreement estimation

	J1J2	J1J3	J2J3
INRA	0.90	0.89	0.88
JOC	0.87	0.86	0.85
HANSARD	0.76	0.82	0.72

- Overall agreement between pairs of annotators
- Lower agreement on HANSARD than on INRA and JOC

Human annotation task

Different annotation schemes

- Segmentation level

J1 *The* ***[allis shad]*** $_1$ ***[is considered to be]*** $_2$ *a vulnerable species*
La ***[grande alose]*** $_1$ ***[est considérée comme]*** $_2$ *une espèce vulnérable*

J2 *The* ***allis*** $_1$ ***shad*** $_2$ ***[is considered]*** $_3$ ***[to be]*** $_4$ *a vulnerable species*
La ***grande*** $_1$ ***alose*** $_2$ ***[est considérée]*** $_3$ ***comme*** $_4$ *une espèce vulnérable*

Human annotation task

Different annotation schemes

- Segmentation level

  J1 *The **[allis shad]$_1$** **[is considered to be]$_2$** a vulnerable species*
  *La **[grande alose]$_1$** **[est considérée comme]$_2$** une espèce vulnérable*

  J2 *The **allis$_1$** **shad$_2$** **[is considered]$_3$** **[to be]$_4$** a vulnerable species*
  *La **grande$_1$** **alose$_2$** **[est considérée]$_3$** **comme$_4$** une espèce vulnérable*

- NULL alignments

  J1 *. . . that there is any change **[in the balance of ways and means]$_1$***
  *. . . avoir apporté le moindre changement **[au niveau de l'ensemble]$_1$***

  J2 *. . . that there is any change **[in the balance of ways and means]$_0$***
  *. . . avoir apporté le moindre changement **[au niveau de l'ensemble]$_0$***

Human annotation task

Types of correspondences

	1-1	NULL	chunks
INRA	64%	15%	21%
JOC	51%	22%	27%
HANSARD	43%	21%	36%

- 1-1 alignments: INRA $>$ JOC $>$ HANSARD
- chunk alignments: INRA $<$ JOC $<$ HANSARD

Evaluation of overall performance

Results

	CLA	ALIBI	GIZA++	ALIBI
		INRA		
P	0.96	0.90 (−0.06)	0.95	0.91 (−0.04)
R	0.45	0.62 (+0.17)	0.66	0.75 (+0.09)
F	0.61	0.73 (+0.12)	0.78	0.82 (+0.04)
		JOC		
P	0.96	0.87 (−0.09)	0.93	0.87 (−0.06)
R	0.43	0.57 (+0.14)	0.58	0.67 (+0.09)
F	0.60	0.69 (+0.09)	0.71	0.75 (+0.04)
		HANSARD		
P	0.95	0.85 (−0.10)	0.89	0.82 (−0.07)
R	0.28	0.40 (+0.12)	0.43	0.53 (+0.10)
F	0.43	0.55 (+0.12)	0.58	0.64 (+0.06)

Evaluation of overall performance

Results

	CLA	ALIBI	GIZA++	ALIBI
	INRA			
P	0.96	0.90 (−0.06)	0.95	0.91 (−0.04)
R	0.45	0.62 (+0.17)	0.66	0.75 (+0.09)
F	0.61	0.73 (+0.12)	0.78	0.82 (+0.04)
	JOC			
P	0.96	0.87 (−0.09)	0.93	0.87 (−0.06)
R	0.43	0.57 (+0.14)	0.58	0.67 (+0.09)
F	0.60	0.69 (+0.09)	0.71	0.75 (+0.04)
	HANSARD			
P	0.95	0.85 (−0.10)	0.89	0.82 (−0.07)
R	0.28	0.40 (+0.12)	0.43	0.53 (+0.10)
F	0.43	0.55 (+0.12)	0.58	0.64 (+0.06)

Evaluation of overall performance

Results

	CLA	ALIBI	GIZA++	ALIBI
	INRA			
P	0.96	0.90 (−0.06)	0.95	0.91 (−0.04)
R	0.45	0.62 (+0.17)	0.66	0.75 (+0.09)
F	0.61	0.73 (+0.12)	0.78	0.82 (+0.04)
	JOC			
P	0.96	0.87 (−0.09)	0.93	0.87 (−0.06)
R	0.43	0.57 (+0.14)	0.58	0.67 (+0.09)
F	0.60	0.69 (+0.09)	0.71	0.75 (+0.04)
	HANSARD			
P	0.95	0.85 (−0.10)	0.89	0.82 (−0.07)
R	0.28	0.40 (+0.12)	0.43	0.53 (+0.10)
F	0.43	0.55 (+0.12)	0.58	0.64 (+0.06)

Evaluation of overall performance

Results

	CLA	ALIBI	GIZA++	ALIBI
		INRA		
P	0.96	0.90 (−0.06)	0.95	0.91 (−0.04)
R	0.45	0.62 (+0.17)	0.66	0.75 (+0.09)
F	0.61	0.73 (+0.12)	0.78	0.82 (+0.04)
		JOC		
P	0.96	0.87 (−0.09)	0.93	0.87 (−0.06)
R	0.43	0.57 (+0.14)	0.58	0.67 (+0.09)
F	0.60	0.69 (+0.09)	0.71	0.75 (+0.04)
		HANSARD		
P	0.95	0.85 (−0.10)	0.89	0.82 (−0.07)
R	0.28	0.40 (+0.12)	0.43	0.53 (+0.10)
F	0.43	0.55 (+0.12)	0.58	0.64 (+0.06)

Evaluation of overall performance

Results

	CLA	ALIBI	GIZA++	ALIBI
	INRA			
P	0.96	0.90 (−0.06)	0.95	0.91 (−0.04)
R	0.45	0.62 (+0.17)	0.66	0.75 (+0.09)
F	0.61	0.73 (+0.12)	0.78	0.82 (+0.04)
	JOC			
P	0.96	0.87 (−0.09)	0.93	0.87 (−0.06)
R	0.43	0.57 (+0.14)	0.58	0.67 (+0.09)
F	0.60	0.69 (+0.09)	0.71	0.75 (+0.04)
	HANSARD			
P	0.95	0.85 (−0.10)	0.89	0.82 (−0.07)
R	0.28	0.40 (+0.12)	0.43	0.53 (+0.10)
F	0.43	0.55 (+0.12)	0.58	0.64 (+0.06)

Evaluation of overall performances

Results

	CLA	ALIBI	GIZA++	ALIBI
		INRA		
P	0.96	0.90 (−0.06)	0.95	0.91 (−0.04)
R	0.45	0.62 (+0.17)	0.66	0.75 (+0.09)
F	0.61	0.73 (+0.12)	0.78	0.82 (+0.04)
		JOC		
P	0.96	0.87 (−0.09)	0.93	0.87 (−0.06)
R	0.43	0.57 (+0.14)	0.58	0.67 (+0.09)
F	0.60	0.69 (+0.09)	0.71	0.75 (+0.04)
		HANSARD		
P	0.95	0.85 (−0.10)	0.89	0.82 (−0.07)
R	0.28	0.40 (+0.12)	0.43	0.53 (+0.10)
F	0.43	0.55 (+0.12)	0.58	0.64 (+0.06)

Evaluation of overall performances

Results

	CLA	ALIBI	GIZA++	ALIBI
		INRA		
P	0.96	0.90 (−0.06)	0.95	0.91 (−0.04)
R	0.45	0.62 (+0.17)	0.66	0.75 (+0.09)
F	0.61	0.73 (+0.12)	0.78	**0.82** (+0.04)
		JOC		
P	0.96	0.87 (−0.09)	0.93	0.87 (−0.06)
R	0.43	0.57 (+0.14)	0.58	0.67 (+0.09)
F	0.60	0.69 (+0.09)	0.71	**0.75** (+0.04)
		HANSARD		
P	0.95	0.85 (−0.10)	0.89	0.82 (−0.07)
R	0.28	0.40 (+0.12)	0.43	0.53 (+0.10)
F	0.43	0.55 (+0.12)	0.58	**0.64** (+0.06)

Evaluation of overall performances

To sum up

- F-measure increased over baselines
- Best strategy: ALIBI bootstrapped with GIZA++
- Much higher performances on INRA than on JOC and HANSARD
  - INRA: $F = 0.82$
  - JOC: $F = 0.75$
  - HANSARD: $F = 0.64$
- HANSARD
  - inter-annotator agreement $--$
  - chunk alignments $++$
  - performances $--$