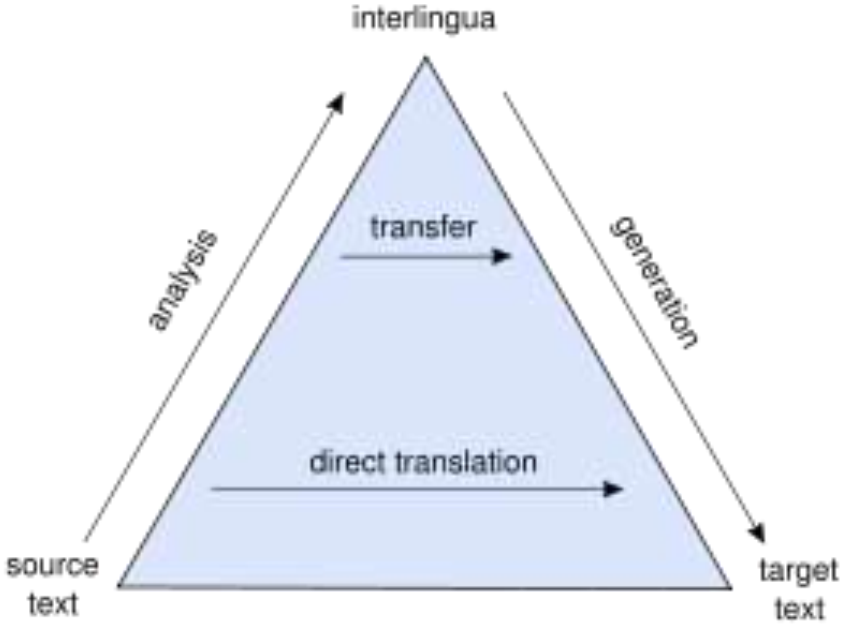
Putting Text-Level Linguistics into Statistical Machine Translation

Andrei Popescu-Belis
www.idiap.ch/~apbelis
Idiap Research Institute, Martigny, Switzerland

"Why Linguistics?" Conference – Tartu, 6 May 2015

Since its inception in the 1950s, and especially in the past 20 years, machine translation has made less and less use of linguistics



- State-of-the-art MT is slipping down the MT pyramid
- From rule-based, to example-based, to statistical systems
  - Within rule-based: from inter-lingua (representing meaning), to transfer (syntactic), to direct

The success of statistical MT

- "*Whenever I fire a linguist, our system performance improves*"
  - said Frederik Jelinek around 1980, marking the statistical turn in automatic speech recognition, followed later by machine translation

- What is statistical MT?
  - translation as a noisy channel (Weaver 1947, then Brown et al. 1993)

    1. Learn n-gram based translation and language models.
    2. Decode the source sentence: find the target sentence that maximizes the probabilities given by the translation model and the language model.

- State of the art performance
  - phrase-based or hierarchical SMT, or direct rule-based MT

Formal definition

- Goal: given *s*, find *t* which maximizes P(*t*|*s*)

- Rewritten using Bayes's theorem as:

$$\underset{t \in TL}{\operatorname{argmax}} P(t \mid s) = \underset{t \in TL}{\operatorname{argmax}} \left(P(s \mid t) \cdot P(t) \right)$$

translation model

language model (target)

Our problem

- SMT is efficient, has good coverage, is quite intelligible
  - suudlevad tudengid → kissing students [→ suudlemine õpilased]
- *But* it always translates sentence by sentence
  - it does not propagate information along a series of sentences

- Such information is crucial for correct *text* translation
  - referring information, lexical chains
    - noun phrases, terminology, pronouns
  - discourse relations, as signaled by connectives
  - verb tense, mode, aspect
  - style, register, politeness

- This information is not accurately captured by SMT

Plan of this talk

- Motivation and method
- Discourse-level linguistic features for SMT
  1. Disambiguation of English discourse connectives
  2. Translation of English verb tenses into French
  3. Towards coherent translation of referring expressions
     - pronoun post-editing from English to French
     - co-reference to compounds in German and Chinese
- Conclusion and perspectives

Setting and contributors

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

- Large collaboration started in 2010 supported by the Swiss National Science Foundation through two consecutive projects
  - COMTIS: Improving the coherence of MT by modeling inter-sentential relations
  - MODERN: Modeling discourse entities and relations for coherent MT

- Research groups and people
  - **Idiap NLP group** (lead): Thomas Meyer, Quang Luong, Najeh Hajlaoui, Xiao Pu, Jeevanthi Liyanapathirana, Lesly Miculicich, Catherine Gasnier
  - **University of Geneva, Department of Linguistics:** Jacques Moeschler, Sandrine Zufferey, Bruno Cartoni, Cristina Grisot, Sharid Loaiciga
  - **University of Geneva, CLCL group:** Paola Merlo, James Henderson, Andrea Gesmundo
  - **University of Zurich, Institute of Computational Linguistics**: Martin Volk, Mark Fishel, Laura Mascarell
  - **Utrecht Institute of Linguistics**: Ted Sanders, Jacqueline Evers-Vermeul, Martin Groen, Jet Hoek

1. MOTIVATION AND METHOD

Examples: problems with discourse connectives

- Source: Why has no air quality test been done on this particular building ***since*** we were elected?
- SMT: Pourquoi aucun test de qualité de l' air a été réalisé dans ce bâtiment ***car*** nous avons été élus ?
- Human: Comment se fait-il qu'aucun test de qualité de l'air n'ait été réalisé dans ce bâtiment ***depuis*** notre élection?

- Source: ***While*** no one wants to see public demonstration, I have to say I understand the anxiety …
- SMT: ***Bien que*** personne ne veut voir la démonstration publique, je dois dire que je comprends l'inquiétude …
- Human: ***Alors que*** personne ne veut voir de manifestations publiques, je dois dire que je comprends leur anxiété …

Example: problems with verb tenses

- Source: Grandmother ***drank*** three cups of coffee a day.
- SMT: Grand-mère ***a bu*** trois tasses de café par jour.
- Human: Grand-maman ***buvait*** trois tasses de café par jour.

- Source: ... that we ***support*** a system that ***is*** clearer than the current one ...
- SMT: ... que nous ***soutenir*** un système qui ***est*** plus claire que le système actuel ...
- Human: ... que nous ***soutenons*** un système qui ***soit*** plus clair que le système actuel ...

Example: problem with NP coherence

- Source: Am 3. Juni schleppten Joe, Mac und ich die erste Traglast zum Lager II, während ***die Träger*** die unteren Lager mit Vorräten versorgten. [..] Am nächsten Morgen kamen ***die Träger*** unbegleitet vom Lager II zu uns herauf, als wir noch in den Schlafsäcken lagen.

- SMT: Le 3 Juin Joe, Mac, et j'ai traîné la première charge au camp II, tandis que ***le support*** fourni avec le roulement inferieur fournitures. […] Le lendemain matin, ***le transporteur*** est arrive seul à partir de Camp II a nous, car nous étions encore dans leurs sacs de couchage.

- Human: Le 3, Joe, Mac et moi montâmes les premières charges au camp II, tandis que ***les porteurs*** faisaient la navette entre les camps inferieurs. […] Nous étions encore dans nos sacs de couchage, le lendemain matin, lorsque ***les porteurs*** arrivèrent du camp II.

Examples: problems with pronouns

- Source: The table is made of wood. ***It*** is magnificent.
- SMT: La table est faite de bois. ***Il*** est magnifique.
- Human: La table est en bois. ***Elle*** est magnifique.

- Source: The European commission must make good these omissions as soon as possible. ***It*** must also cooperate with the Member States …
- SMT: La commission européenne doit réparer ces omissions dès que possible. ***Il*** doit également coopérer avec les états membres …
- Human: … ***Elle*** …

Objective
(condensed view)

			1. Connective	2. Pronoun	3. Verb tense		
The matrix	*has been reduced*	*four times,*	*since*	*it*	*was*	*too large.*	
La matrice	*a été réduite*	*quatre fois,*	*depuis qu'*	*il*	*a été*	*trop grand.*	✘
			car	*elle*	*était*	*trop grande.*	✓

Current machine translation systems: red

Using longer-range dependencies: green

Why linguistics?

- Why is a text more than a series of unconnected sentences?
- Because text is coherent. Coherence is ensured by cohesion markers.

- What cohesion markers might help machine translation?
- The ones mentioned above: noun phrases, pronouns, tense, connectives. But also: lexical choice, style, register, etc.

- Why do current SMT systems mistranslate (some) cohesion markers?
- Because SMT translates sentence-by-sentence, while cohesion markers are longer range, often inter-sentential.

- How can we improve MT at the text level? (And possibly locally as well.)
- By giving it information about constraints across sentences.

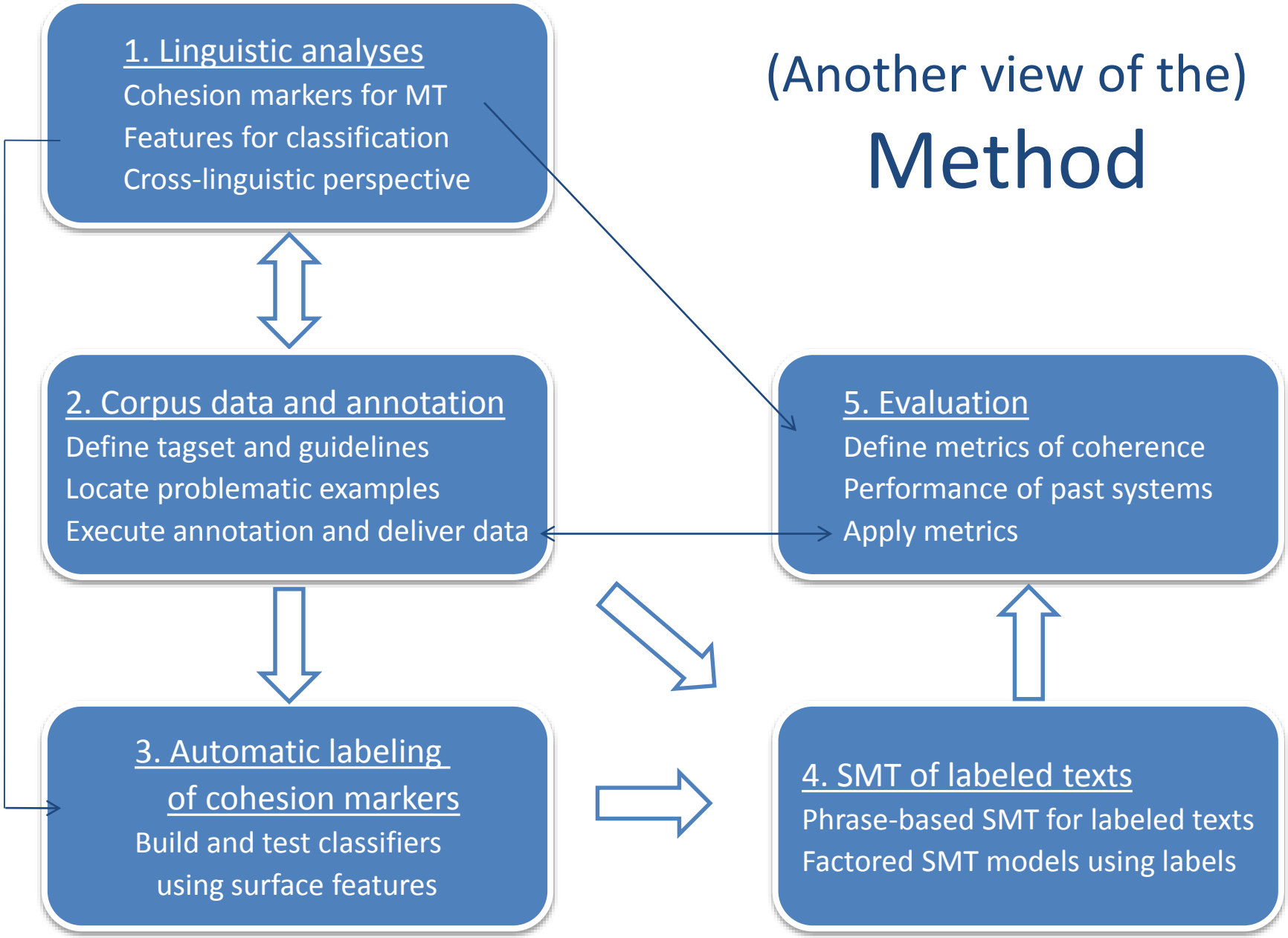
- How do we model such long-range constraints?
- Using insights from linguistics.

Method

1. Define and analyze the phenomena to target
   - design theoretical models, keeping in mind objective and tractability
   - propose features for automatic recognizers
2. Create data for training and evaluation
   - define labeling instructions
   - annotate data sets (which can also be used for corpus linguistics)
   - validate linguistic models through empirical studies
3. Automatic disambiguation (= labeling = classification = recognition)
   - design and implement automatic classifiers
   - e.g. using machine learning over annotated data, based on surface features
4. Combine the automatically-assigned labels with MT
   - adapt MT systems (SMT or RBMT) or design new text-level translation models and decoding algorithms
5. Evaluation
   - assess improvements for the targeted phenomena and overall quality

(Another view of the)

Method



Putting the method into application

- Phenomena discussed in this talk

  1. Discourse connectives 2. Verb tenses 3. Pronouns/NPs

- Languages
  - English, French, German, Italian, Arabic, Chinese

- Domains/corpora
  - parliamentary debates: Europarl (EU languages)
  - transcribed lectures: TED (ALL)
  - Alpine Club yearbooks: Text+Berg (FR, DE)
  - news: data from the Workshops on SMT (ALL)

2. DISAMBIGUATION OF ENGLISH DISCOURSE CONNECTIVES

What are discourse connectives?

- Small words, big effects
  - signal discourse relations between sentences or clauses
  - additional, temporal, causal, conditional, etc.
- Theoretical descriptions
  - Rhetorical Structure Theory (Mann and Thompson)
  - Discourse Representation Theory (Asher et al.)
  - Cognitive approach to Coherence Relations (Sanders et al.)
  - annotation-oriented: Penn Discourse Treebank (PDTB) (Prasad, Webber, Joshi et al.)

- Connectives are challenging for translation because they may convey different relations, which are translated differently
  - *while* contrastive or temporal: French *mais* or *pendant que*
  - *since* causal or temporal: French *puisque* or *depuis que*
- Wrong translations of connectives lead to:
  - low coherence or readability
  - distorted relationships between sentences
  - correct relations are sometimes impossible to recover

What are discourse connectives?

- Small words, big effects
  - signal discourse relations between sentences or clauses
  - additional, temporal, causal, conditional, etc.
- Theoretical descriptions
  - Rhetorical Structure Theory (Mann and Thompson)
  - Discourse Representation Theory (Asher et al.)
  - Cognitive approach to Coherence Relations (Sanders et al.)
  - annotation-oriented: Penn Discourse Treebank (PDTB) (Prasad, Webber, Joshi et al.)

- Connectives are challenging for translation because they may convey different relations, which are translated differently
  - *while* contrastive or temporal: French *mais* or *pendant que*
  - *since* causal or temporal: French *puisque* or *depuis que*

English: What stands between them and a verdict is this doctrine that has been criticized *since*_TEMPORAL it was first issued.
French reference: Seule cette doctrine critiquée *depuis*_TEMPORAL son introduction se trouve entre eux et un verdict.
French baseline MT: Ce qui se situe entre eux et un verdict est cette doctrine qui a été critiqué **parce qu'*_CAUSAL il a d'abord été publié.

Annotation of discourse connectives for translation (Cartoni, Meyer, Zufferey)

- PDTB: complex hierarchy of possible senses of connectives
  - difficult to annotate, not necessarily relevant to MT
- Annotation through translation spotting
  - annotators identify the human translation of each connective (in Europarl)
  - observed translations are clustered into *a posteriori* "senses" relevant to MT
  - fewer labels, cheaper to annotate (e.g. *while* has 21 PDTB labels vs. 5 here)

Connective	Training set			Testing set		
	EP	PDTB	Distribution of labels (%)	EP	PDTB	Distribution of labels (%)
although	168	312	Ct: 68.9; Cs: 31.1	15	16	Ct: 48.4; Cs: 51.6
however	348	450	Ct: 47.8; Cs: 52.2	70	35	Ct: 47.6; Cs: 52.4
meanwhile	102	177	Ct: 77.3; T: 22.7	28	14	Ct: 76.2; T: 23.8
since	339	174	Ca: 38.7; T: 59.6; T/Ca: 1.7	82	10	Ca: 30.4; T: 67.4; T/Ca: 2.2
(even) though	276	306	Ct: 33.3; Cs: 66.7	69	14	Ct: 33.7; Cs: 66.3
while	236	744	Ct: 14; Cs: 23; T: 15; T/Ct: 46.6; T/Ca: 1.4	58	37	Ct: 22.8; Cs: 33.7; T: 9.8; T/Ct:
yet	326	99	Ct: 23.2; Cs: 29.8; Adv: 47	77	2	Ct: 30.4; Cs: 19; Adv: 50.6
Total	1795	2262	–	399	128	–

Features for the automatic disambiguation of connectives

Hong Kong-NNP trade figures illustrate-PRESENT the toy makers' reliance on factories across the border-NN. -JOINT- In-IN 1989's first seven months, -JOINT- domestic exports fell-VBD-PAST-1 29%, to HK$3.87 billion-NN, -CONTRAST- while-IN re-exports-NN rose-VBD-PAST 56%, to HK$11.28 billion-NN.

- syntactic features
  - connective, punctuation, context words, context tree structures, auxiliary verbs
- WordNet antonymy features
  - similarity scores (word distance) and antonyms from the clauses
- TimeML features
- discourse relation features
  - discourse relations from a discourse parser
- polarity features
  - using a polarity lexicon, count positive and negative words, account for negation
- translational features
  - baseline translation (e.g. *tandis que*), sense from dictionary (*contrast*), position (*25)*

- Extracted from the current and the previous sentences

Automatic labeling of connectives
(Th. Meyer)

- For each (new, unseen) discourse connective
  - given the features extracted from the text
  - determine its most probable label (“sense”)

- Use of machine learning for classification
  - Maximum Entropy classifier (or decision trees)
  - trained on manually labeled data (PDTB and/or Europarl)
  - tested on unseen data

Example of scores on the PDTB

Connective	Number of occurrences and senses				F1 Score PT	F1 Score PT+
	Training set: total and per sense		Test set: total and per sense			
after	507	456 As, 51 As/Ca	25	22 As, 3 As/Ca	0.66	1.00
although	267	135 Cs, 118 Ct, 14 Cp	16	9 Ct, 7 Cs	0.60	0.66
however	176	121 Ct, 32 Cs, 23 Cp	14	13 Ct, 1 Cs	0.33	1.00
indeed	69	37 Cd, 24 R, 3 Ca, 3 E, 2 I	*2	2 R	*0.50	*0.50
meanwhile	117	66 Cj/S, 16 Cd, 16 S, 14 Ct/S, 5 Ct	10	5 S, 5 Ct/S	0.32	0.53
nevertheless	26	15 Ct, 11 Cs	6	4 Cs, 2 Ct	0.44	0.66
nonetheless	12	7 Cs, 3 Ct, 2 Cp	*1	1 Cs	*1.00	*1.00
rather	10	6 R, 2 Al, 1 Ca, 1 Ct	*1	1 Al	*0.00	*0.00
since	166	75 As, 83 Ca, 8 As/Ca	9	4 As, 3 Ca, 2 As/Ca	0.78	0.78
still	114	56 Cs, 51 Ct, 7 Cp	13	9 Ct, 4 Cs	0.60	0.66
then	145	136 As, 6 Cd, 3 As/Ca	6	5 As, 1 Cd	0.83	1.00
while	631	317 Ct, 140 S, 79 Cs, 41 Ct/S, 36 Cd, 18 Cp	37	19 Ct, 10 S, 4 Cs, 4 Ct/S	0.93	0.96
yet	80	46 Ct, 25 Cs, 9 Cp	*2	2 Ct	*0.5	*1.00
Total	**2,320**	–	**142**	–	**0.57**	**0.75**

Al: alternative, As: asynchronous, Ca: cause, Cd: condition, Cj: conjunction, Cp: comparison, Cs: concession, Ct: contrast, E: expansion, I: instantiation, R: restatement, S: synchrony

Performance of automatic connective labeling

Data	Method	although	however	meanwhile	since	(even) though	while	yet
Training (c.v.)	All_Features	0.69 ± 0.04	0.85 ± 0.05	0.86 ± 0.01	0.93 ± 0.05	0.77 ± 0.04	0.76 ± 0.04	0.88 ± 0.07
Test: Europarl and PDTB (WSJ s. 23)	Majority class	0.52	0.52	0.76	0.68	0.66	0.34	0.51
	All_Features	0.58	**0.73**	0.71	**0.90**	0.69	0.45	**0.78**
	Best	0.61	0.60	0.74	0.87	**0.71**	0.43	0.72
	All_Synt+Dep	**0.65**	0.67	**0.79**	0.89	0.7	**0.47**	0.72
Test: Europarl	All_Features	0.60	**0.69**	0.79	**0.90**	0.67	0.45	**0.78**
	Best	**0.80**	0.56	0.82	0.85	**0.72**	0.43	0.74
	All_Synt+Dep	0.73	0.66	**0.89**	0.88	0.71	**0.50**	0.73
Test: PDTB (WSJ s. 23)	All_Features	**0.56**	**0.83**	0.57	0.90	**0.79**	**0.46**	**1.0**
	Best	0.44	0.69	0.57	**1.0**	0.64	0.43	0.0
	All_Synt+Dep	**0.56**	0.69	0.57	**1.0**	0.64	0.43	0.50

- F1 scores: trained on 4057 occurrences, tested on 527
- Findings
  - scores compare well to human agreement levels (80-90%)
  - classifying each connective separately is better than jointly
  - using all features is the best option

How do we use labeled connectives in SMT?

Four possible methods have been tested

1. Replace in the system's phrase table all unambiguous occurrences of the connective with the correct one

2. Train the system on (a) manually or on (b) automatically labeled data, with labels concatenated to words (e.g., *while_Temporal*)

3. Use a connective-specific SMT system only when the connective labeler is confident enough (otherwise use a baseline one)

4. Use Factored Models as implemented in the Moses system
   - word-level linguistic labels are separate translation features
   - a model of labels is learned when training, then used when decoding

How do we measure the improvement of connective translation? (Meyer, Hajlaoui)

- Measuring translation quality
  - subjective measures: fluency, fidelity → too expensive for everyday use
  - objective, reference-based measures: BLEU (or METEOR, etc.)
    - comparison of a candidate text with one or more reference translations in terms of common n-grams (usually from 1 to 4)
  - connectives are not frequent → small effects on BLEU scores

- Count how many connectives are correctly translated:
  ACT metric [Accuracy of Connective Translation]
  - given a source sentence with a discourse connective *C*
  - use automatic alignment to find out:
    - how *C* is translated in the reference and in the candidate translations
  - compare the translations: identical? "synonymous"? incompatible? absent?

Improvement of connectives in SMT

1. Modified phrase table
   – tested on ~10,000 occurrences of 5 types: **34%** improved, **20%** degraded, **46%** unchanged

2. Concatenated labels
   (a) trained on manually labeled data: **26%** improved, **8%** degraded, **66%** unchanged
   (b) trained on automatically labeled data: **18%** improved, **14%** degraded, **68%** unchanged

3. Thresholding based on automatic labeler's confidence
   – with two connectives only: improvement of **0.2**-**0.4** BLEU points (quite significant)

4. **Factored models with Moses**

Languages	Test set	System	BLEU	Δ	p	ACT	Δ	p
EN/FR	nt2012	baseline	26.1			56.28		
		labeled connectives	25.8	-0.3	**	57.68	1.40	*
	nt2010	baseline	24.4			68.12		
		labeled connectives	24.3	-0.1	**	68.60	0.48	*
	nt2008+sy2009	baseline	28.9			61.36		
		labeled connectives	29.2	0.3	*	60.94	-0.42	*
EN/DE	nt2012	baseline	11.8			62.28		
		labeled connectives	11.8	0.0	n/s	65.08	2.80	**
	nt2010	baseline	15.0			62.42		
		labeled connectives	15.0	0.0	n/s	69.28	6.86	***
	nt2008+sy2009	baseline	13.0			71.06		
		labeled connectives	13.1	0.1	n/s	70.30	-0.76	n/s

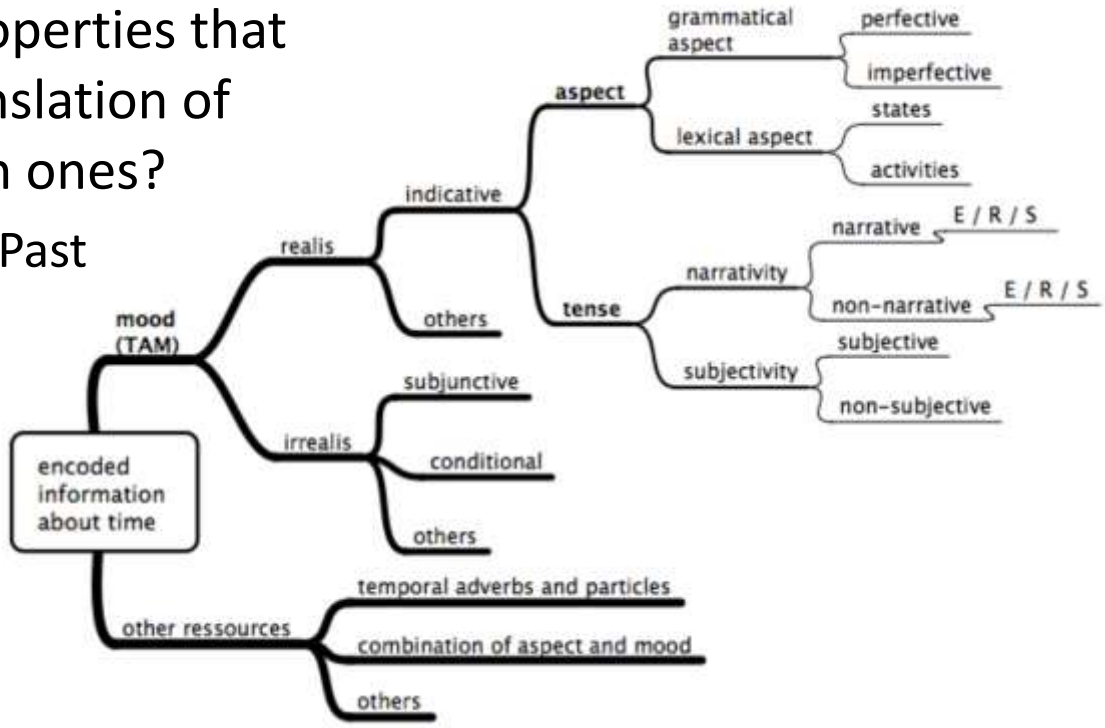
3. TRANSLATING VERB TENSES

Cross-lingual modeling of verb tenses
(Grisot and Moeschler)

- Two well-known models
  - event time, reference time, speech time (Reichenbach)
  - four classes of aspect (Vendler)
- What are the relevant properties that would enable correct translation of English tenses into French ones?
  - focus on English Simple Past
- Theoretical hypothesis:

  simple past narrative
  → *passé simple* or *passé composé*

  simple past non-narrative
  → *imparfait*



Empirical studies of tense translation

- Approaches: narrativity-based vs. general tense correlation

1. Annotation of narrativity (Grisot)
   - English/French parallel corpora (also Italian, Romanian)
   - 576 EN simple past verb phrases
   - inter-annotator agreement on 71% of instances, $\kappa = 0.44$
   - ► narrativity correctly predicts 80% of translated tenses

2. Annotation of translated tense for all English VPs (Loaiciga)
   - rules for precise alignment of VPs in Europarl
   - annotated ca. 320,000 VPs, with about 90% precision
   - ► confirmed divergencies between EN and FR tenses

Observed EN/FR tense divergencies for 322,086 verb phrases (Loaiciga)

French	English								
	Past continuous	Past perfect continuous	Past perfect	Present continuous	Present perfect continuous	Present perfect	Present	Simple past	Total
Imparfait	462 54%	7 27%	**365** **24%**	146 1%	18 2%	463 1%	1 510 1%	**8 060** **21%**	11 031 3%
Impératif				37 0%	1 0%	6 0%	203 0%	11 0%	258 0%
Passé composé	139 16%	2 8%	**214** **14%**	282 1%	325 33%	**26 521** **61%**	1253 1%	**19 402** **49%**	48 138 15%
Passé récent			1 0%	8 0%	3 0%	187 0%	2 0%	3 0%	204 0%
Passé simple	4 1%		6 0%	16 0%	2 0%	54 0%	42 0%	374 1%	498 0%
Plus-que-parfait	27 3%	8 31%	**782** **52%**	2 0%	4 0%	217 1%	22 0%	1 128 3%	2 190 1%
Présent	216 25%	9 35%	102 7%	18 077 96%	617 63%	**14 736** **34%**	211 334 97%	**9 779** **25%**	254 870 79%
Subjonctif	15 2%		28 2%	258 1%	6 1%	**1 053** **2%**	2 969 1%	568 1%	4 897 2%
Total	863 100%	26 100%	1 498 100%	18 826 100%	976 100%	43 237 100%	217 335 100%	39 325 100%	322 086 100%

Features for automatic prediction of narrativity or (directly) translated tense

If the situation were-VBD-PAST-SV-1-0 to change-VBP-INFINITIVE-0-0-0, it would-MD-CONDITIONAL-CSV-0-0 clearly also change-VBP-INFINITIVE-0-0-0 as far as-synch we are-VBP-PRESENT-CSV-0-0 concerned-VBN-0-0-0-0.

- all verbs in the current and previous sentences
- word positions
- verb POS and trees
- auxiliaries and tenses
- TimeML features
- temporal connectives (from a hand-crafted list)
- synchrony/asynchrony of the connectives
- semantic roles
- imparfait indicator: yes/no
- subjonctif indicator: yes/no

- Extracted from the current and the previous sentences

Automatic annotation: results

- Using a maximum entropy classifier

1. Automatic annotation of narrativity (+/-)
   - training on 458 instances, testing on 118
   - F1 score = 0.71, κ = 0.43
2. Prediction of translated tense
   - over 196’000 instances with 10-fold cross-validation
   - F1 score = 0.85 for c.-v.
   - F1 score = 0.83 on a held-out test set

Improvements of SMT using narrativity

- Scores from human evaluators
  - first, is the narrativity label correct?
  - then, are verb tenses and lexical choices improved?

Criterion	Rating	N.	%	Δ
Labeling	correct	147	71.0	
	incorrect	60	29.0	
Verb tense	+	35	17.0	+9.7
	=	157	75.8	
	−	15	7.2	
Lexical choice	+	19	9.2	+3.4
	=	176	85.0	
	−	12	5.8	

Improvements of SMT using predicted tense labels

- Oracle = the prefect prediction
- BLEU scores per target tense

	Baseline	Oracle	Predicted	# Sent.
Imparfait	24.10	25.32	24.57	122
Passé composé	29.80	30.82	30.08	359
Impératif	19.08	19.72	18.70	4
Passé simple	13.34	16.15	14.09	6
Plus-que-parfait	21.27	23.44	23.22	17
Présent	27.55	27.97	27.59	2,618
Subjonctif	26.81	27.72	26.07	78
Passé recent	24.54	30.50	30.08	3

- Manual evaluation of a sample

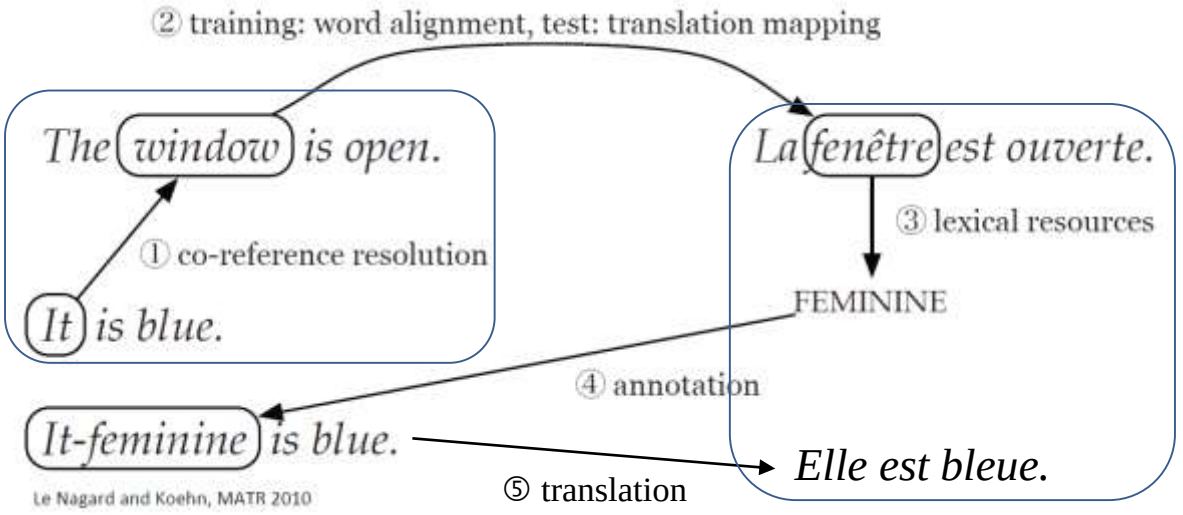
French tense	System	TAM		
		Incorrect	Correct	
			$\neq$ ref	= ref
Imparfait	Baseline	**82**	**15**	**41**
	Predicted	**42**	**23**	**73**
	Oracle	**13**	**4**	**121**

4. REFERENTIAL COHERENCE IN MT

Theoretical view

- If several expressions in a text refer to the same entity, then this information may help to translate them more correctly
  - noun phrases = coreferences | pronouns = anaphora

- Goal: determine referential links, then use them in an SMT system.
  - is this tractable?
  - not pure labeling



Application to pronouns

(J. Liyanapathirana, L. Miculicich, Q. Luong , C. Gasnier)

- Automatic anaphora resolution is far from perfect
  - tried a different approach: pronoun post-editing without anaphora resolution
- Training data (from Europarl)
  - surface features from EN source and FR target sentences (e.g. gender of surrounding nouns), source pronoun, candidate target pronoun
- System predicting if/how a candidate pronoun should be post-edited
  - initial results: 27% improved but 16% degraded
- Shared tasks at the *DiscoMT 2015 Workshop*
  - pronoun-focused translation | pronoun prediction

Coreference and compounds
(L. Mascarell, X. Pu, M. Fishel)

- Motivating example (German to French, Text+Berg corpus)
  - Source: … worauf das ***Bundesamt*** für Landestopographie in Aktion trat. Nur dieses ***Amt*** war in der Lage, [...]
  - SMT: Que ce ***poste*** était dans la situation, [...]
  - Human: Seul cet ***office*** était en mesure, […]

- Objective
  - use the coreference link between a compound (noted *XY*) and its subsequent mention by the head noun (noted *Y*) to improve the translation of *Y*

- Hypothesis
  - the translation of *Y* in *XY* is more accurate than the translation of *Y* alone

- Challenges
  - avoid non-compound and non-coreferent *XY*/*Y* pairs
  - correctly locate and replace the translations of *XY* and *Y*

Use of compound co-reference for SMT

1. CHINESE SOURCE SENTENCE	她以为自买了双两英寸的高跟鞋， 但实际上那是一双三英寸高的鞋。
2. SEGMENTATION, POS TAGGING, IDENTIFICATION OF COMPOUNDS AND THEIR CO-REFERENCE	她#PN 以为#VV 自#AD 买#VV 了#AS 双#CD 两#CD 英寸#NN 的#DEG **高跟鞋#NN** ， #PU 但#AD 实际上#AD 那#PN 是#VC 一#CD 双#M 三#CD 英寸#NN 高#VA 的#DEC **鞋#NN** 。#PU
3. BASELINE TRANSLATION INTO ENGLISH (STATISTICAL MT)	She thought since bought a pair of two inches high **heel**, but in fact it was a pair of three inches high **shoes**.
4. AUTOMATIC POST-EDITING OF THE BASELINE TRANSLATION USING COMPOUNDS	She thought since bought a pair of two inches high **heel**, but in fact it was a pair of three inches high **heel**.
5. COMPARISON WITH A HUMAN REFERENCE TRANSLATION	She thought she'd gotten a two-inch heel but she'd actually bought a three-inch **heel**. ✓

idiap

Improvement of SMT using compounds

- Test data for SMT: ZH/EN and DE/FR
  - training sets: about 200k sentences | tuning: about 2k sentences
  - testing: 800/500 sentences with ca. 250 *XY/Y* pairs
- BLEU scores
  - ZH/EN: very small improvement from 11.18 to 11.27
  - DE/FR: very small decrease from 27.65 to 27.48
- Comparison of the *Y* translations
  - our systems are closer to the reference than baseline
  - when our (best) system and the baseline both differ from reference
    - often due to the use of pronouns in translation
    - our system has an acceptable translation in 92% of the cases vs. 71% for the baseline

			CACHING		POST-EDITING	
			= ref	≠ ref	= ref	≠ ref
ZH/EN	BASELINE	= ref	59.3	*4.1*	42.3	*4.5*
		≠ ref	**13.8**	22.8	**20.3**	32.9
DE/FR	BASELINE	= ref	70.1	*10.3*	73.9	*5.0*
		≠ ref	**4.3**	15.2	**3.5**	17.5

5. CONCLUSIONS & PERSPECTIVES

Wrap up

- Long-range dependencies can be modeled thanks to linguistic theories, and their automatic annotation, although imperfect, can benefit SMT
- Genuine collaboration between: theoretical linguistics and pragmatics, corpus linguistics, natural language processing, and machine translation
- Some outputs
  - publications: available from COMTIS and MODERN websites
  - resources: annotations on Europarl of discourse connectives and verb phrases
  - software: automatic discourse connective labeler, ACT and APT metrics
- Discourse for Machine Translation workshops
  - at ACL 2013 and at EMNLP 2015 (submission deadline: June 28)
  - in 2015 with a shared task on pronoun translation

Challenges for the future

- For the above-mentioned cohesion markers
  - modeling, automatic classification, and SMT can all be improved
    - our consortium is now focusing on co-reference and nouns phrases
  - dilemma: invest research in the classifiers or in the MT?
- For other types of cohesion markers
  - all the work remains to be done
  - question: what are the most promising ones?
- New methods for integration with MT systems
  - text-level decoding for SMT | methods for other types of MT
- Ultimately, how do we integrate these complex, heterogeneous knowledge sources into efficient and robust MT systems?

Acknowledgments

FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

- Swiss National Science Foundation

  COMTIS: Improving the coherence of MT by modeling inter-sentential relations

  www.idiap.ch/project/comtis

  MODERN: Modeling discourse entities and relations for coherent MT

  www.idiap.ch/project/modern

- Collaborators on these projects
  - **Idiap NLP group**: Thomas Meyer, Quang Luong, Najeh Hajlaoui, Xiao Pu, Lesly Miculicich, Jeevanthi Liyanapathirana, Catherine Gasnier
  - **University of Geneva, Department of Linguistics:** Jacques Moeschler, Sandrine Zufferey, Bruno Cartoni, Cristina Grisot, Sharid Loaiciga
  - **University of Geneva, CLCL group:** Paola Merlo, James Henderson, Andrea Gesmundo
  - **University of Zurich, Institute of Computational Linguistics**: Martin Volk, Mark Fishel, Laura Mascarell
  - **Utrecht Institute of Linguistics**: Ted Sanders, Jacqueline Evers-Vermeul, Martin Groen, Jet Hoek

Bibliography

Meyer T., Hajlaoui N., & Popescu-Belis A. (2015) - Disambiguating Discourse Connectives for Statistical Machine Translation. To appear in *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*.

Grisot C. & Meyer T. (2014) - Cross-Linguistic Annotation of Narrativity for English/French Verb Tense Disambiguation. *Proceedings of LREC 2014 (9th International Conference on Language Resources and Evaluation)*, Reykjavik.

Loaiciga S., Meyer T. & Popescu-Belis A. (2014) - English-French Verb Phrase Alignment in Europarl for Tense Translation Modeling. *Proceedings of LREC 2014 (9th International Conference on Language Resources and Evaluation)*, Reykjavik.

Mascarell, Laura; Fishel, Mark; Korchagina, Natalia; & Volk, Martin (2014) - Enforcing Consistent Translation of German Compound Coreferences. *Proceedings of KONVENS 2014 (12th German Conference on Natural Language Processing)*, Hildesheim, Germany.

Cartoni B., Zufferey S., Meyer T. (2013) - "Annotating the meaning of discourse connectives by looking at their translation: The translation-spotting technique". *Dialogue & Discourse* : Beyond semantics: the challenges of annotating pragmatic and discourse phenomena. Vol. 4, No. 2, pp. 65-86.

Zufferey S. & Cartoni B. (2012) - English and French causal connectives in contrast. *Languages in Contrast*. 12(2): 232-250.

Meyer T., Grisot C. and Popescu-Belis A. (2013). Detecting Narrativity to Improve English to French Translation of Simple Past Verbs. In *Proceedings of the 1st DiscoMT Workshop at ACL 2013 (51th Annual Meeting of the Association for Computational Linguistics)*, Sofia, Bulgaria, pages 33-42.

Meyer T., Popescu-Belis A., Hajlaoui N., Gesmundo A. (2012). Machine Translation of Labeled Discourse Connectives. In *Proceedings of the Tenth Biennial Conference of the Association for Machine Translation in the Americas (AMTA)*, San Diego, CA.

Zufferey S., Degand L., Popescu-Belis A. & Sanders T. (2012) - Empirical validations of multilingual annotation schemes for discourse relations. *Proceedings of ISA-8 (8th Workshop on Interoperable Semantic Annotation)*, Pisa, p.77-84.

Meyer, T., Popescu-Belis, A. (2012). Using Sense-labeled Discourse Connectives for Statistical Machine Translation. *In Proceedings of the EACL 2012 Workshop on Hybrid Approaches to Machine Translation (HyTra)*, Avignon, France, pp. 129-138.

Popescu-Belis A., Meyer T., Liyanapathirana J., Cartoni B. & Zufferey S. (2012). Discourse-level Annotation over Europarl for Machine Translation: Connectives and Pronouns. *Proceedings of LREC 2012*, May 23-25 2012, Istanbul, Turkey.