

Gabriel Marzinotto

1-Orange Labs, Lannion, France

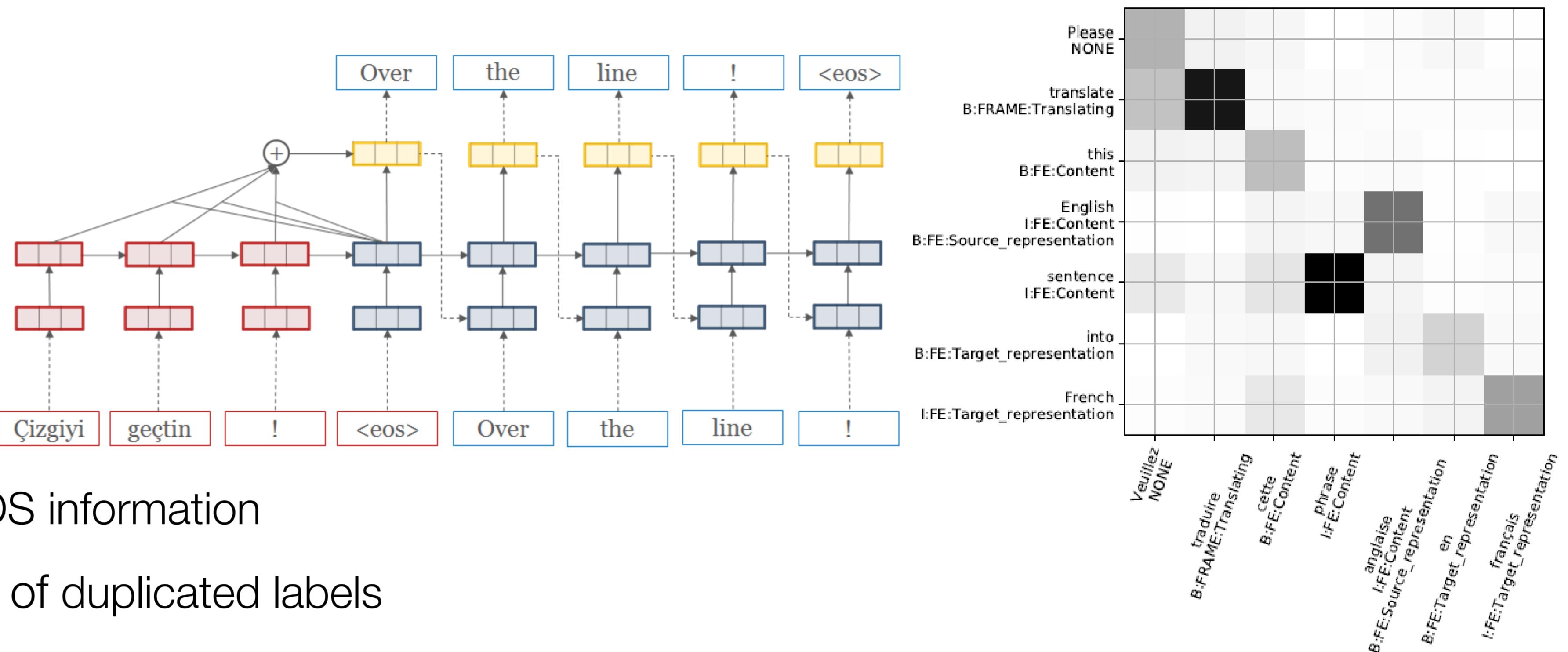
gabriel.marzinotto@orange.com

Objectives

- ✓ Project FrameNet annotations into other languages using attention-based neural machine translation (NMT) models.
- ✓ Evaluate the approach using a multilingual BERT-based FrameNet parser

Translation and Label Projection Model

- NMT with encoder-decoder attention
- Use attention to align source and target languages
- Simple rules to project labels
 - Project labels only for verbs, nouns and adjectives
 - Complete Frame Elements (w/determinants and prepositions)
- Score label projections using the attention matrix and POS information
- Evaluate the projection : 19% labels (deletions) and 10% of duplicated labels



Human Translation vs Machine Translation

EN: *We have a huge vested interest in it, partly because it's education that's meant to take us into this future that we can't grasp.*

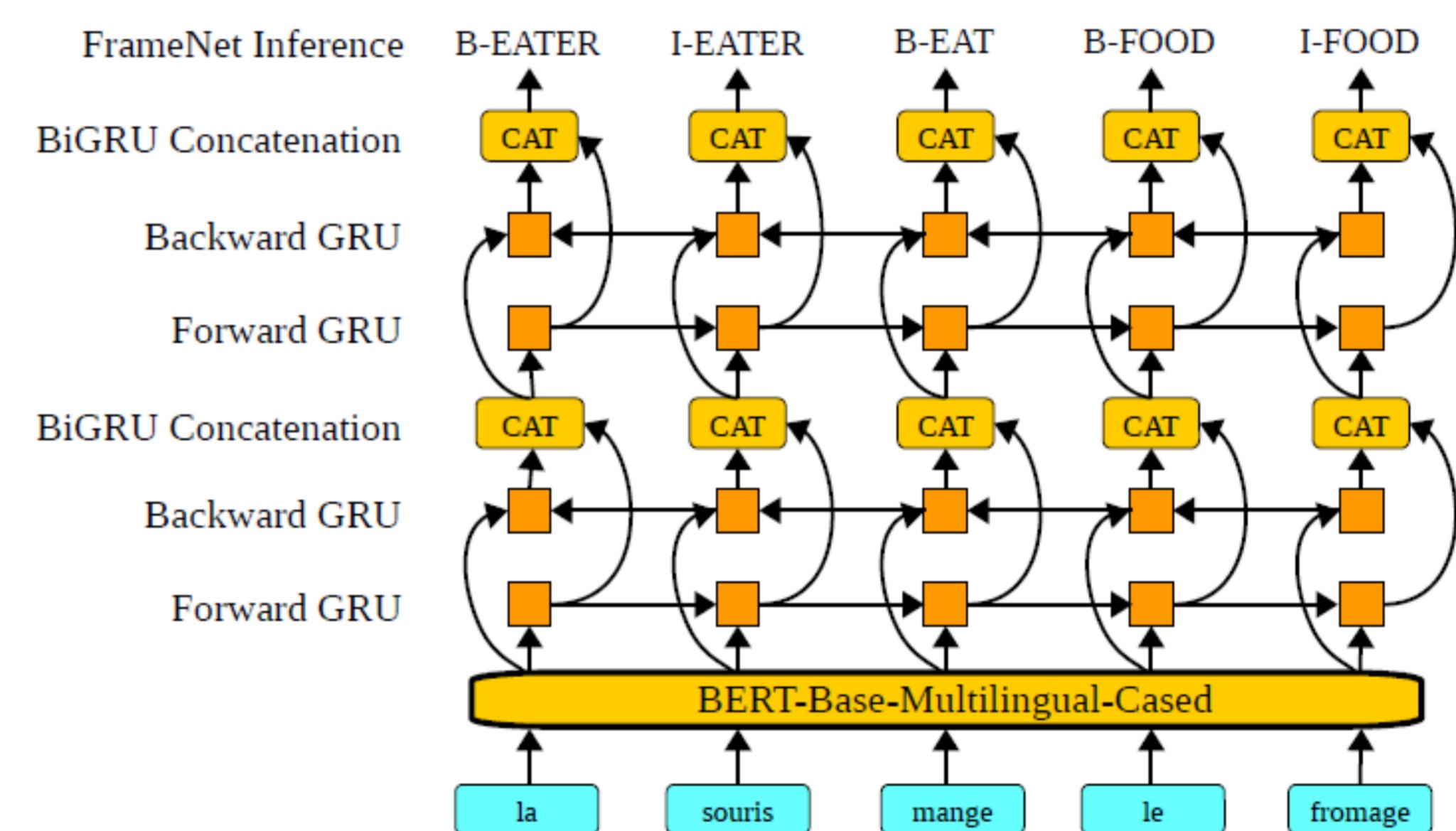
PT-Human: *Nos interessamos tanto por ela em parte porque é da educação o papel de nos conduzir a esse futuro misterioso.*

PT-Machine: *Temos um grande interesse, em parte porque é a educação que nos leva a esse futuro que não podemos compreender.*

Frame	EN	PT-HT	PT-MT
Size	huge.a	—	grande.a
Stimulus_focus	interest.n	interessar.v	interesse.n
Degree	—	tanto.adv	—
Causation	partly.adv	em parte.adv	em parte.adv
Education	because.c	porque.c	porque.c
Purpose	education.n	educação.n	educação.n
Performers.roles	mean.v	—	—
Bringing	take.v	papel.n	—
Goal	into.prep	conduzir.v	levar.v
Temporal_colloc	future.n	futuro.n	futuro.n
Certainty	—	misterioso.a	—
Capability	can.v	—	poder.v
Grasp	grasp.v	—	compreender.v

Semantic Parsing Models

- Semantic Parsing as a ‘Sequence Tagging Problem’
- 4 Layer bidirectional GRU on top of ML-BERT
- Parsing one trigger at a time
- Viterbi decoding and FrameNet Coherence filter to validate outputs



FrameNet Corpora et évaluation intrinsèque

✓ English FrameNet has much more frames we would like to exploit in French

✓ We translate SemEval-07-EN to obtain SemEval07-FR

	Lang	Data Source	Nb. Diff. Frames	Nb. Diff. LU	Nb. Diff. FE	Nb. Diff. Words	Nb. Sentences w/LU	Nb. Annot. LU	Avg. Nb. Annot. per LU
SemEval-07	EN	Journals	720	3,197	754	14,150	4,020	24,770	7.7
ASFALDA	FR	Journals	121	782	140	33,955	13,154	16,167	20.7
CALOR	FR	Encyclopedias	53	145	148	72,127	22,603	31,440	215.4

Automatic Parsing Experiments

- ✓ Training parsers using translated corpora improves performances on the target language.
- ✓ Frame Identification benefits more than argument identification
- ✓ Ultimately combining the original corpus and its translation yields further improvement
- ✓ However, there are important differences between translated annotations and gold annotations

	Test : ASFALDA		Test : CALOR	
	Frame Id.	Arg Id. pred frames	Frame Id.	Arg Id. pred frames
TRAINING CORPUS	ACC	F1-W	ACC	F1-W
CALOR	-	-	98.8	67.8
ASFALDA	73.3	52.6	-	-
SemEval-07-EN	42.2	19.7	77.5	30.5
SemEval-07-FR	73.3	19.1	78.1	31.0
SemEval-07-EN+FR	74.8	20.9	79.4	32.4

Performance of a biGRU+BERT on the Gold French FrameNet corpora using different training data

Conclusion

- ✓ Simple method to project FrameNet annotations into other languages using attention-based neural machine translation (NMT) models..
- ✓ A multilingual BERT-based FrameNet parser gives a strong baseline to evaluate how relevant the translated corpus is.
- ✓ Our method shows modest gains on English to French setting and shows that there is still room for improvement in the alignments.