
Hybridisation through System Combination

Antonio Toral

Workshop on Hybrid MT for Irish
29th April 2016
Dublin City University

Outline

- What it* is
- How does it work
- MEMT
- In practice
- Final remarks

*system combination

What it is

Source “i am a teacher”

2 MT outputs

- Tá mé múinteoir
- Tá i i mo mhúinteoir

What it is

Source “i am a teacher”

2 MT outputs

- Tá mé múinteoir
- Tá i i mo mhúinteoir

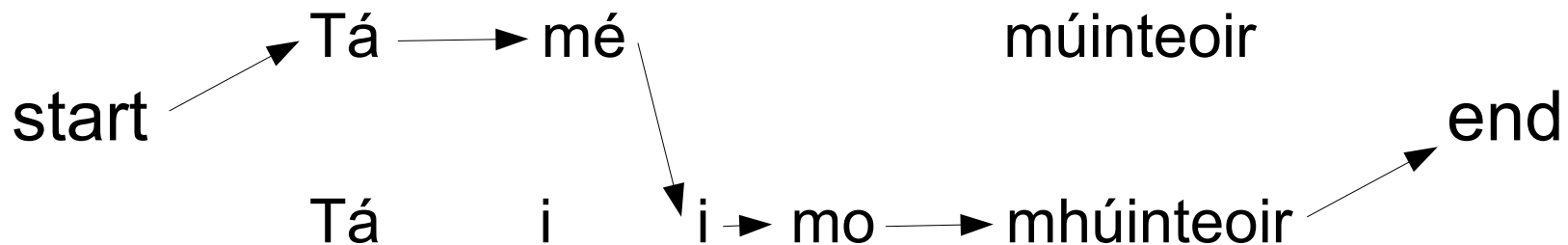
	Tá		mé		múinteoir					end
start										
	Tá		i		I		mo			mhúinteoir

What it is

Source “i am a teacher”

2 MT outputs

- Tá mé muinteoir
- Tá i i mo mhuinteoir



How does it work

1. Align MT outputs
2. Tune weights
3. Decode

How does it work. Alignment

Tá mé múinteoir
| | |
Tá i l mo mhúinteoir

How does it work. Tuning

- A set of weights
 - System confidence
 - Language model
- E.g. 2 systems, 1 weight per system, 1 LM
$$\lambda_1 \cdot S_1 + \lambda_2 \cdot S_2 + (1-\lambda_1-\lambda_2) \cdot LM$$

e.g. $\lambda_1 = 0.6$ and $\lambda_2 = 0.3$ then LM weight = 0.1

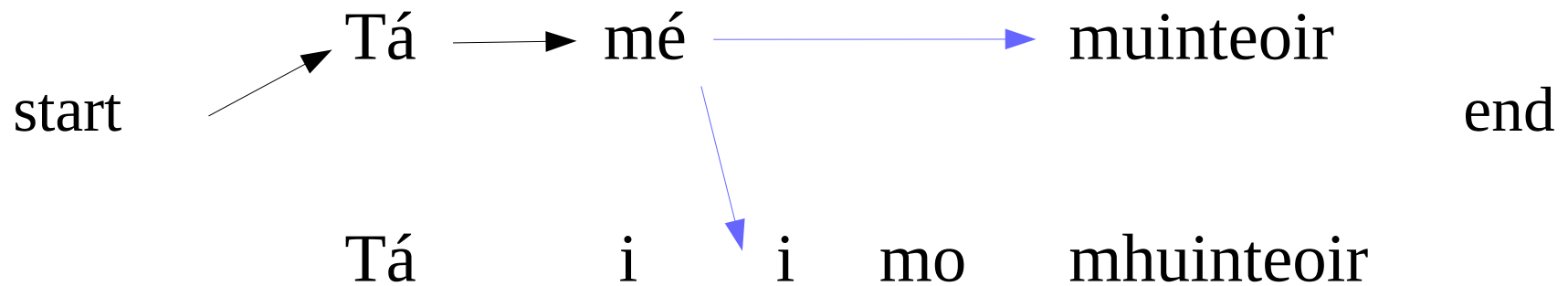
How does it work. Tuning

- Procedure similar to “common” MT tuning
 - Optimise weights on a dev set
 - Same tuning algorithms: MERT, MIRA, etc.
 - Instead of single MT output, we have a combined MT output

How does it work. Decoding

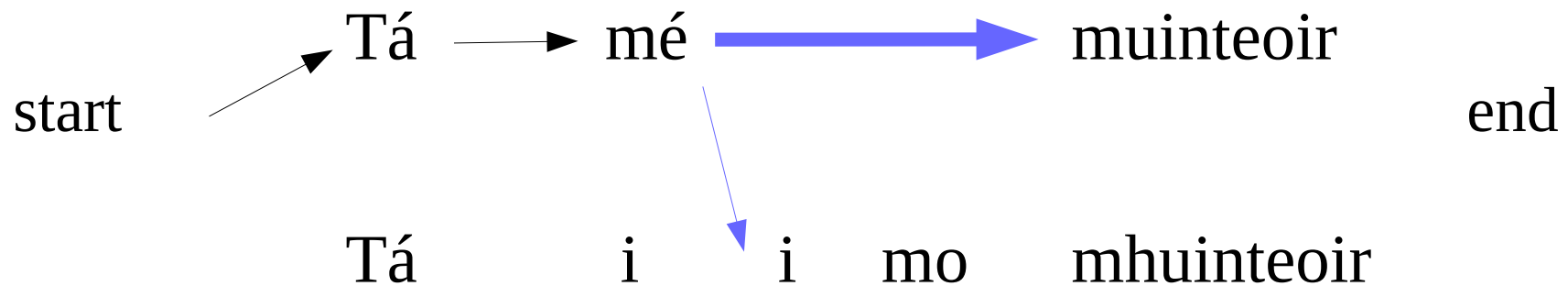
1. Decode test set with all the MT systems
2. Align the MT outputs
3. Obtain the best path by applying the weights learnt on the dev set

How does it work. Decoding



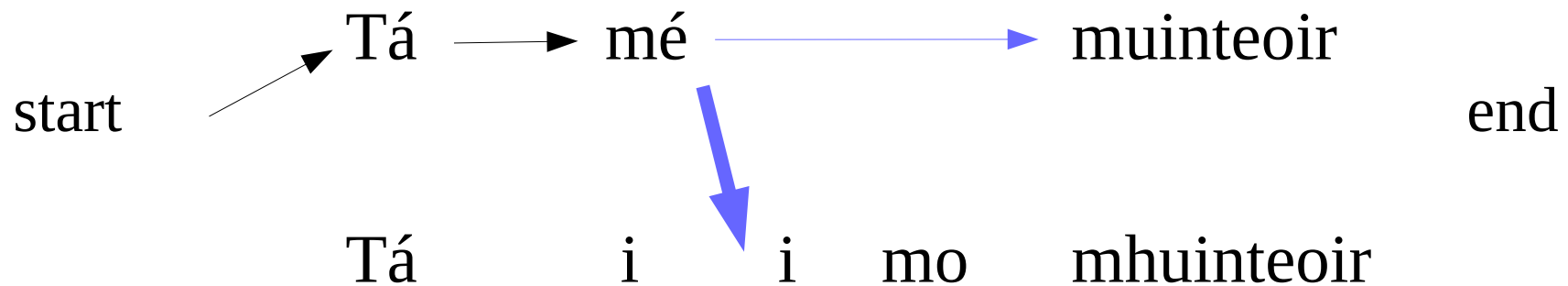
How does it work. Decoding

λ_1 = weight system 1
 λ_3 = weight LM
 $p_{LM} = p(\text{muinteoir}, \text{Tá mé})$



How does it work. Decoding

λ_2 = weight system 2
 λ_3 = weight LM
 $p_{LM} = p(i, \text{Tá mé})$



MEMT

- Open source tool for system combination
- Alignment with METEOR
- Tuning with ZMERT
- Features
 - 5 per system
 - Exact matches for 1-grams and 2-grams
 - Any match for 1-grams, 2-grams, and 3-grams (stems, synonyms, ...)
 - LM score
 - OOV count
- <https://github.com/kpu/MEMT>

In Practice

1

TweetMT 2015

In Practice. TweetMT 2015

- 3 language pairs
 - Closely related: Spanish – Catalan, Spanish – Portuguese
 - Unrelated: Spanish – Basque
- Systems
 - SMT
 - phrase based
 - As is
 - With data selection (ParFDA)
 - with morph segmentation (only for Basque)
 - hierarchical
 - RBMT (Apertium / Matxin)

In Practice. TweetMT 2015

	System	BLEU	TER
ES→CA	Moses (1)	82.21	0.1102
	cdec (2)	81.45	0.1128
	ParFDA (3)	82.37	0.1062
	Apertium (4)	78.17	0.1310
	1+2	81.71	0.1102
	1+4	82.37	0.1057
	1+2+4	81.93	0.1085
	1+3+4	82.45	0.1074

Results on dev

	System	BLEU	TER
ES→CA	DCU1 (1+4)	0.7669	0.1740
	DCU2 (1)	0.7899[†]	0.1626[†]
	DCU3 (1+2+4)	0.7630	0.1738
CA→ES	DCU1 (1+4)	0.7826	0.1506
	DCU2 (1+2+4)	0.7816	0.1500
	DCU3 (1+3+4)	0.7943[†]	0.1431[†]

Results on test

In Practice. TweetMT 2015

	System	BLEU	TER
ES→EU	Moses (1)	22.57	0.6116
	cdec (2)	23.7	0.5863
	ParFDA (3)	21.59	0.6181
	Matxin (4)	12.66	0.7436
	Morph (5)	5.20	0.8812
	1+2	23.18	0.5796
	1+4	18.36	0.6112
	1+2+4	23.58	0.5771
	1+2+4+5	24.07	0.5741
	1+2+3+4+5	24.42	0.5777
EU→ES	Moses (1)	24.21	0.6228
	cdec (2)	24.65	0.5911
	ParFDA (3)	22.25	0.6346
	Apertium (4)	18.36	0.6918
	Morph (5)	11.25	0.9655
	1+2	24.18	0.5883
	1+4	24.33	0.6076
	1+2+4	24.94	0.5831
1+2+4+5	25.21	0.5792	

	System	BLEU	TER
ES→EU	DCU1 (1+2+4)	0.2455	0.6533
	DCU2 (1+2+3+4+5)	0.2636[†]	0.6469[†]
	DCU3 (1+2+4+5)	0.2493	0.6553
EU→ES	DCU1 (2)	0.2687	0.6512
	DCU2 (1+2+4)	0.2698	0.6406
	DCU3 (1+2+4+5)	0.2728	0.6363

Results on test

In Practice. TweetMT 2015

	System	BLEU	TER
ES→PT	Moses (1)	29.21	0.6052
	cdec (2)	28.14	0.5962
	ParFDA (3)	27.74	0.6164
	Apertium (4)	24.96	0.6272
	1+2	28.76	0.5891
	1+4	26.58	0.6082
	1+2+4	27.00	0.5878
PT→ES	Moses (1)	30.47	0.5267
	cdec (2)	29.42	0.5254
	ParFDA (3)	29.63	0.5338
	Apertium (4)	27.52	0.5335
	1+2	29.9	0.5230
	1+4	30.01	0.5131
	1+2+4	29.89	0.5089

Results on dev

	System	BLEU	TER
ES→PT	DCU1 (1)	0.3595	0.5290
	DCU2 (1+2)	0.3711[†]	0.5157[†]
	DCU3 (1+2+4)	0.3687	0.5163
PT→ES	DCU1 (1)	0.4465	0.5767
	DCU2 (1+2)	0.4467	0.5627
	DCU3 (1+2+4)	0.4524[†]	0.5403[†]

Results on test

In Practice

2

WMT 2015

In Practice. WMT 2015

- English – Finnish
- Systems
 - SMT with different segmentations
 - SMT with factored models (only into Finnish)
 - Translation: word → word and POS
 - Hierarchical

In Practice. WMT 2015

Segmenter	text
None	kuntaliitoksen selvittämisessä
hfst-comp	‘kunta → ← liitoksen selvittämisessä
hfst-morph	kunta → ← liitokse → ← n selvittämisessä
Flatcat	kun → ← tali → ← itoksen selvittämisessä
Morfessor	kun → ← ta → ← liito → ← ksen selvittä → ← misessä
Gloss	municipality+annexation.Gen examination.Ine
Translation	examination regarding municipal annexation

In Practice. WMT 2015

Segmenter		text	
None		kuntaliitoksen selvittämisessä	2
hfst-comp		‘kunta→←liitoksen selvittämisessä	3
hfst-morph		kunta→←liitokse→←n selvittämisessä	5
Flatcat		kun→←tali→←itoksen selvittämisessä	5
Morfessor		kun→←ta→←liito→←ksen selvittämisessä	6
Gloss		municipality+annexation.Gen examination.Ine	
Translation		examination regarding municipal annexation	

In Practice. WMT 2015

System	Dev		Test	
	BLEU	TER	BLEU	TER
Phrase-Based	17.19	0.762	16.90	0.759
Hierarchical	16.98	0.768	15.93	0.773
HFST Comp	17.87	0.748	16.68	0.753
HFST Morph	18.64	0.735	17.22	0.752
Morfessor	16.83	0.769	15.96	0.756
Flatcat	16.78	0.766	17.33	0.741
Combination	19.66	0.719	18.77	0.726

Finnish → English

In Practice. WMT 2015

System	Dev		Test	
	BLEU	TER	BLEU	TER
Phrase-Based	13.51	0.827	12.33	0.843
Factored Model	13.08	0.827	11.89	0.847
Hierarchical	13.05	0.822	12.11	0.830
HFST Comp	13.57	0.814	12.66	0.828
HFST Morph	13.19	0.818	12.77	0.819
Morfessor	12.21	0.860	11.58	0.864
Flatcat	12.67	0.844	12.05	0.849
Combination	14.61	0.786	13.54	0.801

English → Finnish

In Practice

3

WMT 2016

In Practice. WMT 2016

- English – Finnish
- Systems
 - SMT with dynamic segmentation
 - Reranked with RNN LM
 - NMT

In Practice. WMT 2016

	BLEU	CHRF1
SMT	0.1562	51.0419
NMT	0.1830	52.4349
Combo (BLEU)	0.1638	51.7523

In Practice. WMT 2016

	BLEU	CHRF1
SMT	0.1562	51.0419
NMT	0.1830	52.4349
Combo (BLEU)	0.1638	51.7523
Combo (CHRF1)	0.1767	52.3657

In Practice. WMT 2016

- Given reference “i mo mhúinteoir” and 2 outputs:
 - MT1: “i mo múinteoir”
 - MT2: “i mo dalta”
- BLEU will score them the same
 - Word sequences matches: “i, mo, i mo”
- CHRF will score the 1st higher
 - Char sequences matches, e.g. 3-grams:
 - MT1: “imo, mom, úin, int, nte, teo, eoi, oir” omh, mhú, húi
 - MT2: “imo”

In Practice. WMT 2016

	BLEU	CHRF1
SMT	0.1562	51.0419
NMT	0.1830	52.4349
Combo (BLEU)	0.1638	51.7523
Combo (CHRF1)	0.1767	52.3657
+ reranking	0.1845	52.6478

Final remarks

- Is system combination useful?
 - Offline, batch translation ✓
 - online ☹️
 - Different and competitive MT systems ✓
 - Too similar ☹️
 - Important gap in performance ☹️

Final remarks

- BLEU doesn't like morph. rich languages
 - Range of metrics: BLEU, TER, METEOR, CHRF
- Reranking with neural LM
- RBMT, SMT...
 - And NMT

**Go raibh maith agaibh!
Ceisteanna?**

**Hybridisation through
System Combination**

Antonio Toral, DCU
