

The SLE Example-Based Translation System

Pete Whitelock, Victor Poznanski
Sharp Laboratories of Europe Ltd.
Oxford
{pete, vp}@sharp.co.uk

Background

- 1970s – work by Sharp Corp. on E→J MT
Duet, Power E/J, これ一本 (PC package)
- 1997-2003 – SLE's Intelligent Dictionary
おまかせ 訳振り (lightweight English analysis)
- 2003-2005 – SLE work on bilingual example retrieval
and TM (~EBMT) functions for package
- 2005-2006 – SLE extend EBMT
 - Improve accuracy
 - include lightweight dependency analysis
 - Reduce resource utilisation for low power device
 - no language model
 - replace recursive use of TM with dictionary



Example

query

この階にコートがありますか

stored example

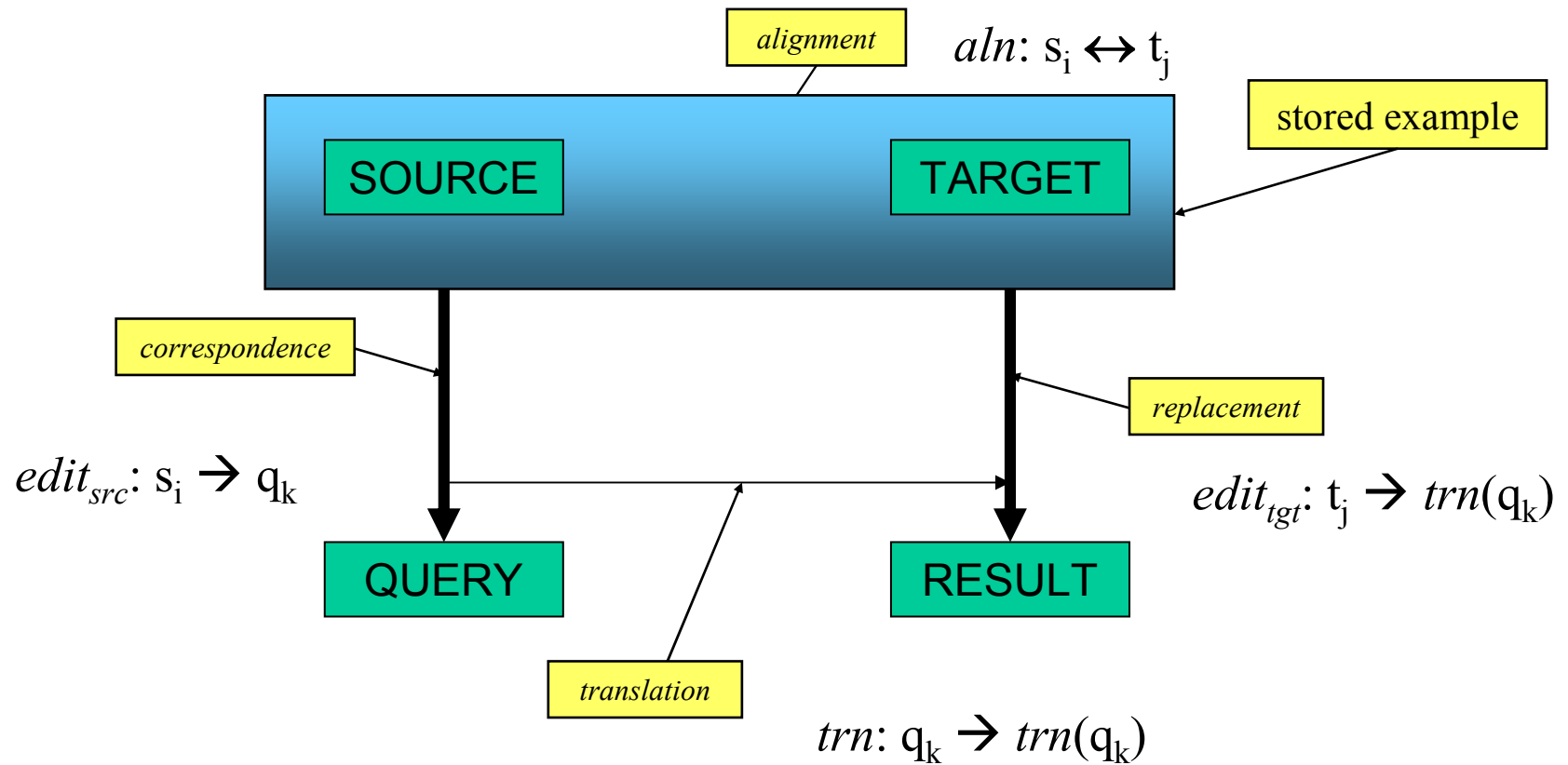
この階に子供服がありますか

Is this the floor for children's clothes?

result

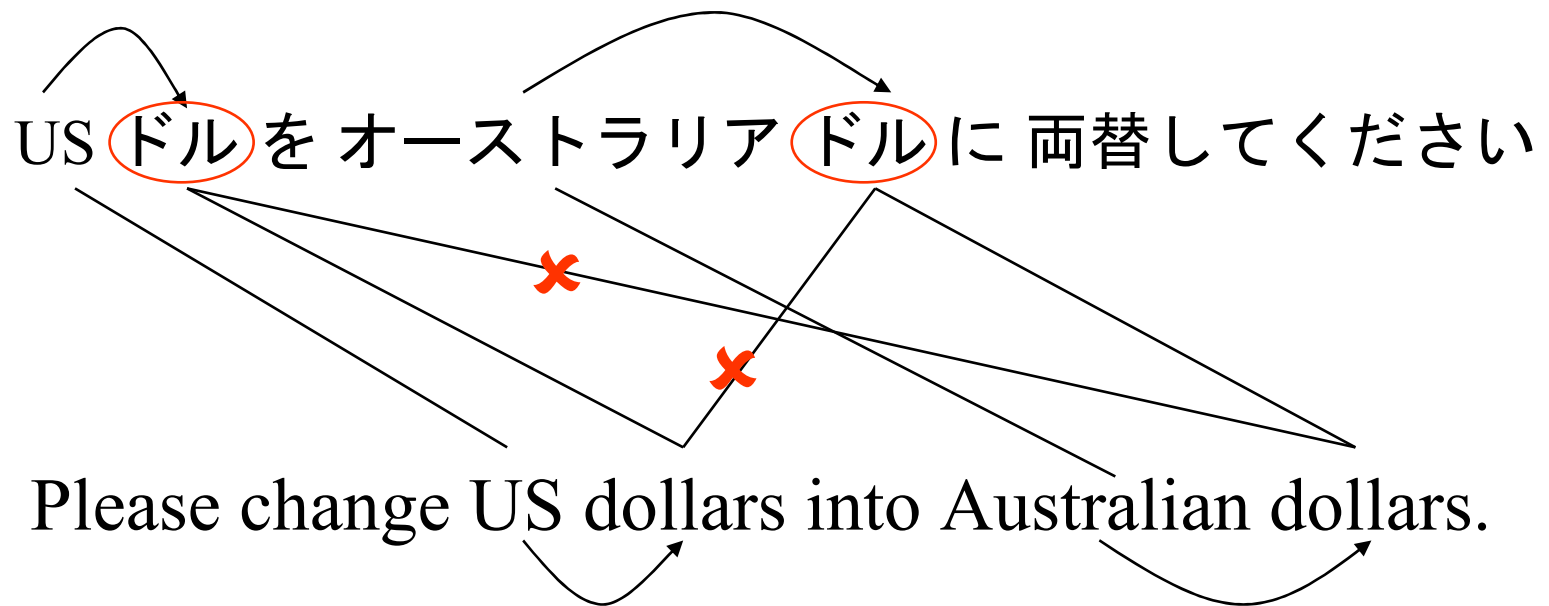
Is this the floor for coats?

Edit Model



Alignment

- Sentence aligned bitext =>
(partial) alignment of (sets of) words
- Bilingual dictionary/thesaurus + dependency analysis
 - Constraint propagation
 - Use dependency proximity to resolve ambiguity
 - Annotate examples with semantic (thesaurus) codes of bilingual entries used
- Special routines
 - Number parser for English and Japanese
 - Readings of kanji proper names as found by Japanese morphological analyser
 - Katakana to English edit rules +dp spell check



Correspondence

- Example base is indexed by words, lemmas, semantic codes
- Retrieve candidate examples which share index terms with query
 - don't prefer rare terms, so obtain templatic effect.
- Compute sequence of matched and unmatched stretches
 - determine score based on these
 - adjust score by semantic overlap in unmatched stretches
- Use example with best score as basis for translation

14206 CLOTHES

16224 ~~SPORTS PLACE~~

この階に コート がありますか

query

14206 CLOTHES

overlap = 5

この階に 子供服 がありますか

Is this the floor for children's clothes?



examples

16232 STORE

overlap = 3

この階に レストラン がありますか

Is there a restaurant on this floor?

Translation

- Dependency-analyse query
- Find set of bilingual entries matching query
- Find maximum consistent (tiling) set
 - prefer entries covering more items
 - prefer entries with semantic codes consistent with those determined in correspondance
 - prefer entry used most frequently in corpus
- Pass the set to replacement phase

Translation

PRIO 3.15 at_PREP ← +clocktime ⇔ NUM 時に
PRIO 2.05 he_PRON ⇔ 彼 → は
PRIO 1.1 get_V ← back_ADV ⇔ 戻る_VERB

彼 は | 5 時 | に | 戻り ます 。

彼 は | 月曜日 | に | 戻り ます 。

He'll be back on Monday.

~~He'll be back at 5 o'clock~~

Replacement

- Determine positions in basis to be substituted
 - by examination of alignment relation
- Apply TL edit
 - deletion
 - delete words at positions in basis
 - delete determiners and prepositions when their head is deleted
 - substitution
 - substitute head position in basis by translations of words in query
 - delete words at all other substitution positions in basis
 - insert
 - insertion is adnominal – add head noun to both sides of source edit
 - assimilate insertion to substitution
 - insertion is adverbial – add translation to start or end of sentence
 - handle cases of boundary friction
 - *a/an* alternation, removal of multiple prepositions
 - apply number agreement
 - generate inflected comparatives and superlatives
 - copy features from basis to output

Feature Copy

この 電車 は 定刻に 出発 の 予定 ですか。

この 便 は 定刻に 到着 の 予定 ですか。

Will this flight arrive on time?

Will this $t(\text{電車})$ $t(\text{出発})$ on time?

Will this train departure on time?

Will this train depart on time?

N→V ↓

Results

	BBS	SLE Example Base (~13,000)			SLE+IWSLT Example Base (~53,000)		
		EBMT Only	%age	+BBS	EBMT Only	%age	+BBS
devset2 (IWSLT2004)	.3524 [7.7607]	.4910 [7.6240]	70.5	.4063 [8.2176]	.5610 [8.927]	75.3	.4663 [8.8784]
devset3 (IWSLT2005)	.3137 [7.5425]	.4994 [7.8347]	66.0	.3930 [8.1415]	.5450 [8.1934]	72.7	.4411 [8.5965]
devset4 (IWSLT2006)	.1917 [5.5127]	.1537 [2.1997]	38.7	.1828 [5.5208]	.1313 [1.4768]	74.2	.1835 [5.6189]
test (asr 1best)						59.4	.1599 [5.3393]
test (correct)	.1797 [5.4599]					60.6	.1726 [5.6497]

