

Machine Translation System Combination with Flexible Word Ordering



Kenneth Heafield, Greg Hanneman, and Alon Lavie

Language Technologies Institute

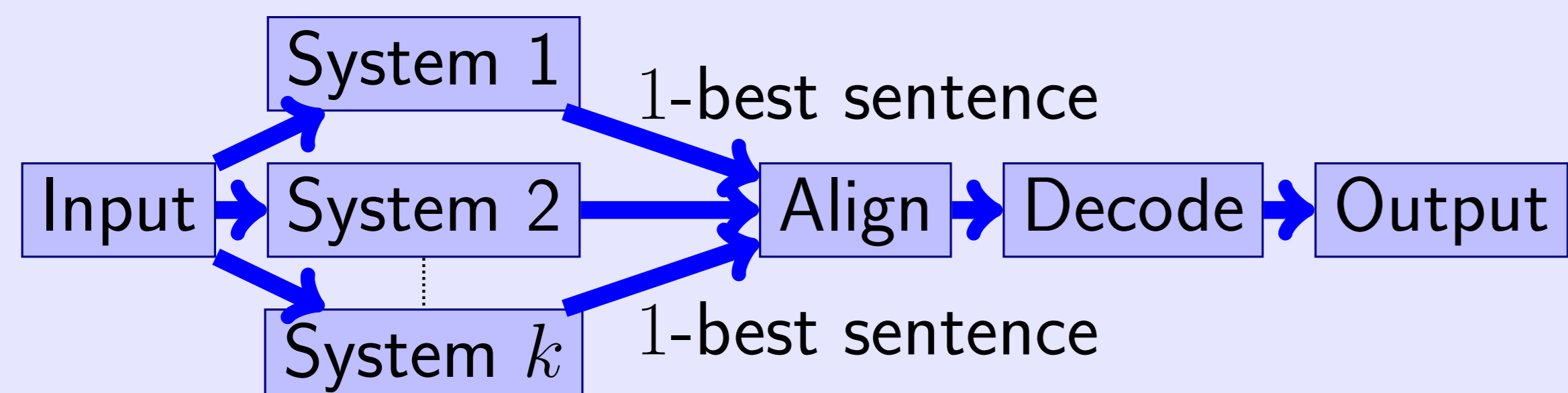
Carnegie Mellon University

{kheafiel,ghannema,alavie}@cs.cmu.edu

Carnegie Mellon

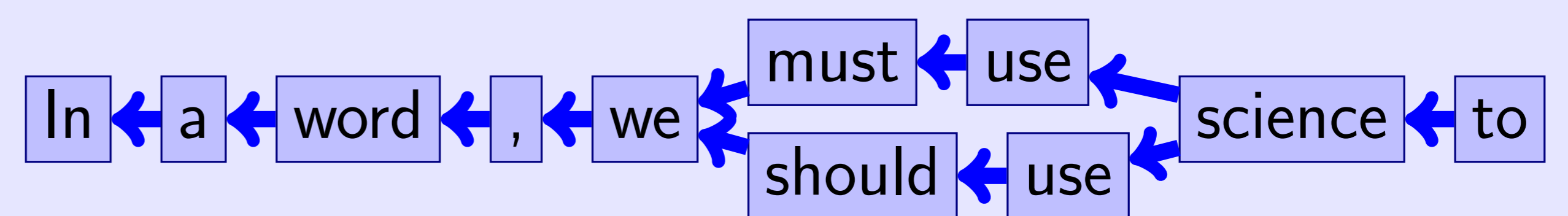
Overview

Gather 1-best translation system hypotheses
Align hypotheses using stemming and synonymy
Decode by following and switching between hypotheses
Output synthetic combination



Decoder Hypothesis

Features: Scores from our four features
Used: Set of used words
Phrase: Current phrase, if any
Word: Most recently decoded word
History: Set of pointers to preceding hypotheses



Hypotheses are recombined when used, phrase, and n -gram agree.

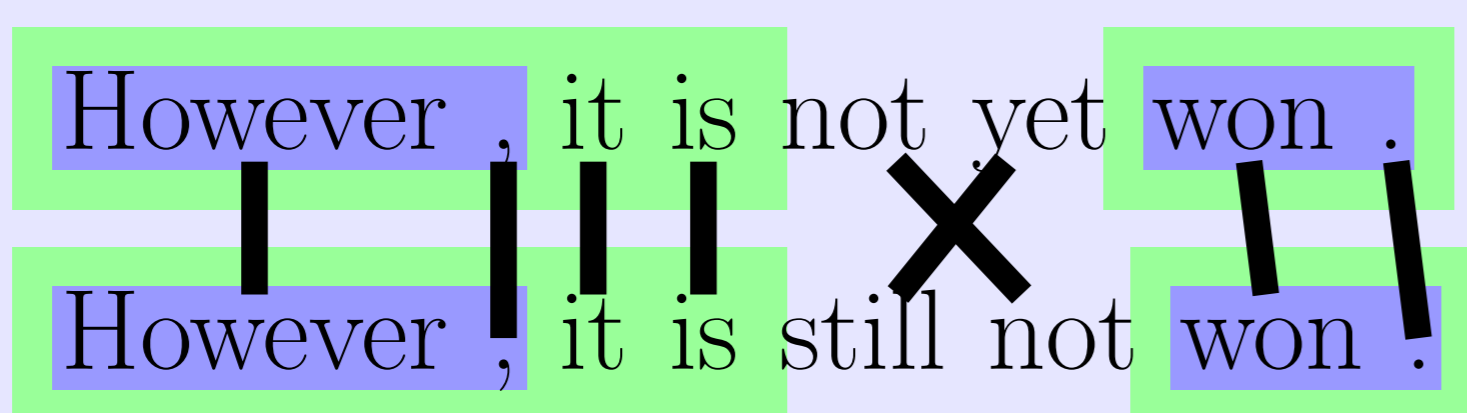
Alignment

Match surface, stems, and WordNet synsets
Minimize crossing alignments
Speculate using part of speech when neighbors align



Phrases

Detect phrases using maximal consecutive alignments
Tie punctuation to the preceding word
Constrain decoding to complete phrases if possible

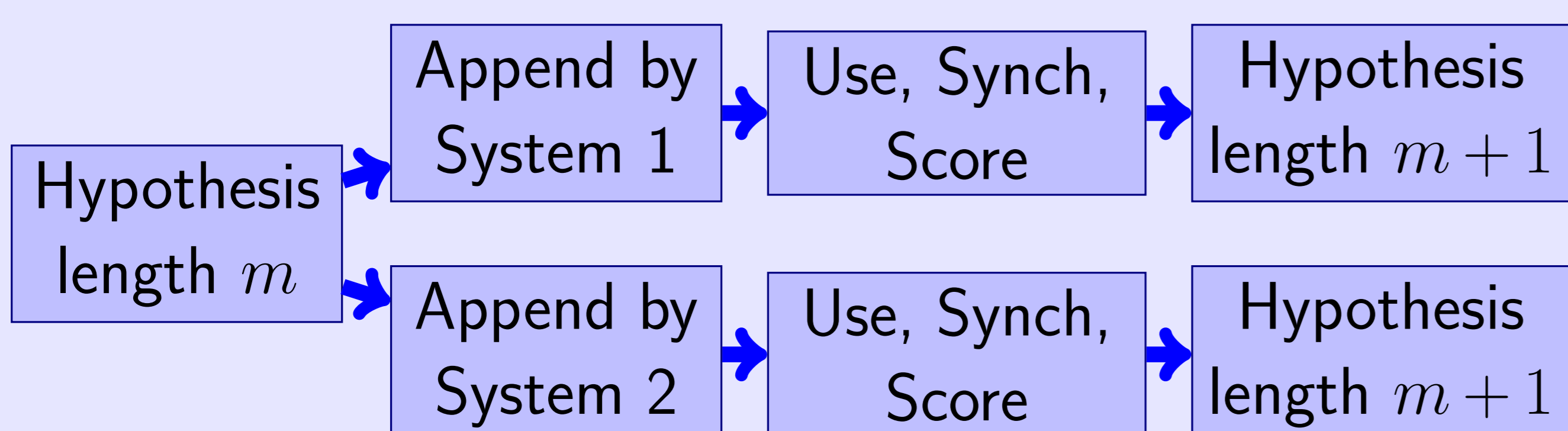


Decoding

Start at the beginning of sentence
Append the first unused word from a system
Use the appended word and those aligned with it
Synchronize to remove lingering unused words

Score using our four features

Loop until all hypotheses reach end of sentence



Synchronization Example

1. The decoder can pick the first unused word from either system.

Most people always takes over a cell phone .
 The majority of the people is always a mobile .

2. Suppose the decoder picks "Most", marking it used.

Most people always takes over a cell phone .

3. Looking at alignments, system 2 is behind by 4 words.

Most people always takes over a cell phone .
 The majority of the people is always a mobile .
 4 3 2 1 0

4. Words are marked used to synchronize within tolerance.

The majority of the people is always a mobile .

Scoring

Using MERT, we compute an overall score based on four features:

Alignment: Sum of confidence in systems aligned to word

Language: Score from a language model

n -Gram: e^n when the language model found an n -gram

Overlap: Size of overlap with previous n -gram

Results

	In	Entry	BLEU	MET	TER
cz	combo		.2171	.4993	61.20
	google		.2118	.4923	59.89
de	combo		.2226	.4989	59.56
	google		.2131	.4983	61.37
es	combo		.2824	.5490	54.06
	google		.2868	.5505	53.86
fr	combo		.3000	.5589	53.73
	google		.3113	.5656	51.74
hu	combo		.1377	.3831	72.23
	morpho		.0989	.3778	81.65
xx	combo		.3080	.5637	52.31
	google-fr		.3113	.5656	51.74

Example

Source: Trotz allem gibt es genügend Gründe dafür, warum man sich einen eigenständigen Player zulegen sollte.

Translate:

Google: Despite everything, there are plenty of reasons why a player should be independent.

UKA: Even so, there are plenty of reasons why you should be a stand-alone player.

Stuttgart: Despite all of this there are enough reasons why we should establish an independent player.

UMD: Despite everything, there is plenty of reasons why we should create an independent player.

Combine

Combination: Even so, there are plenty of reasons why a player should be independent.

Compare

Reference: In spite of this, there are many reasons to get a separate MP3 player.