Karlsruher Institut für Technologie

معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation عضو في مؤسسة قطر

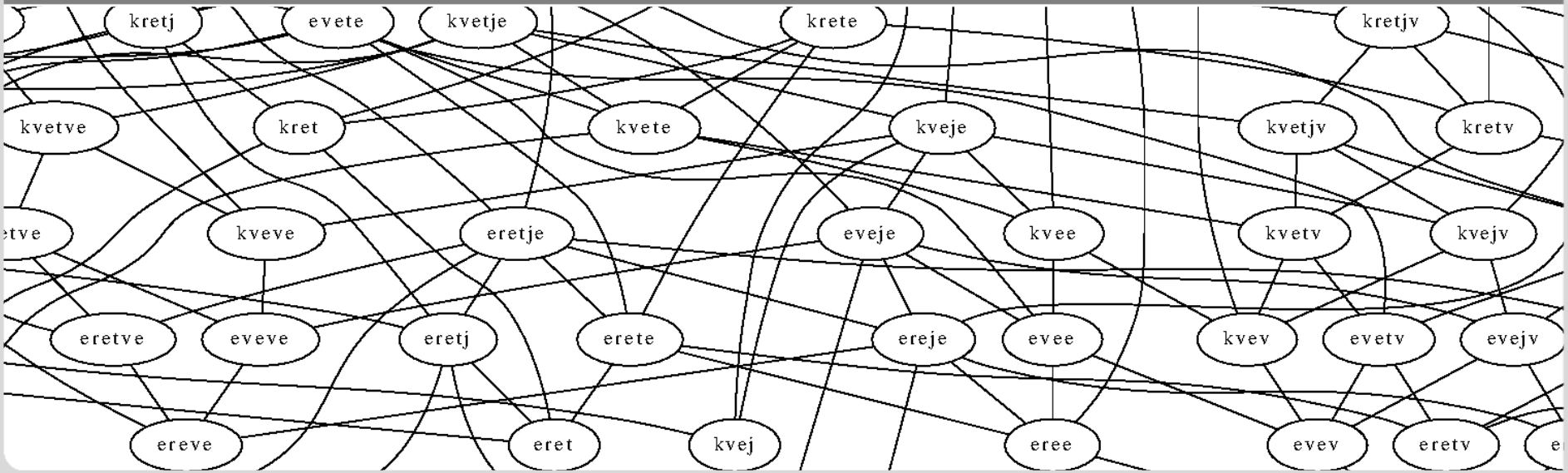
Cognitive Systems Lab

Cross-Lingual Lexical Language Discovery from Audio Data Using Multiple Translations

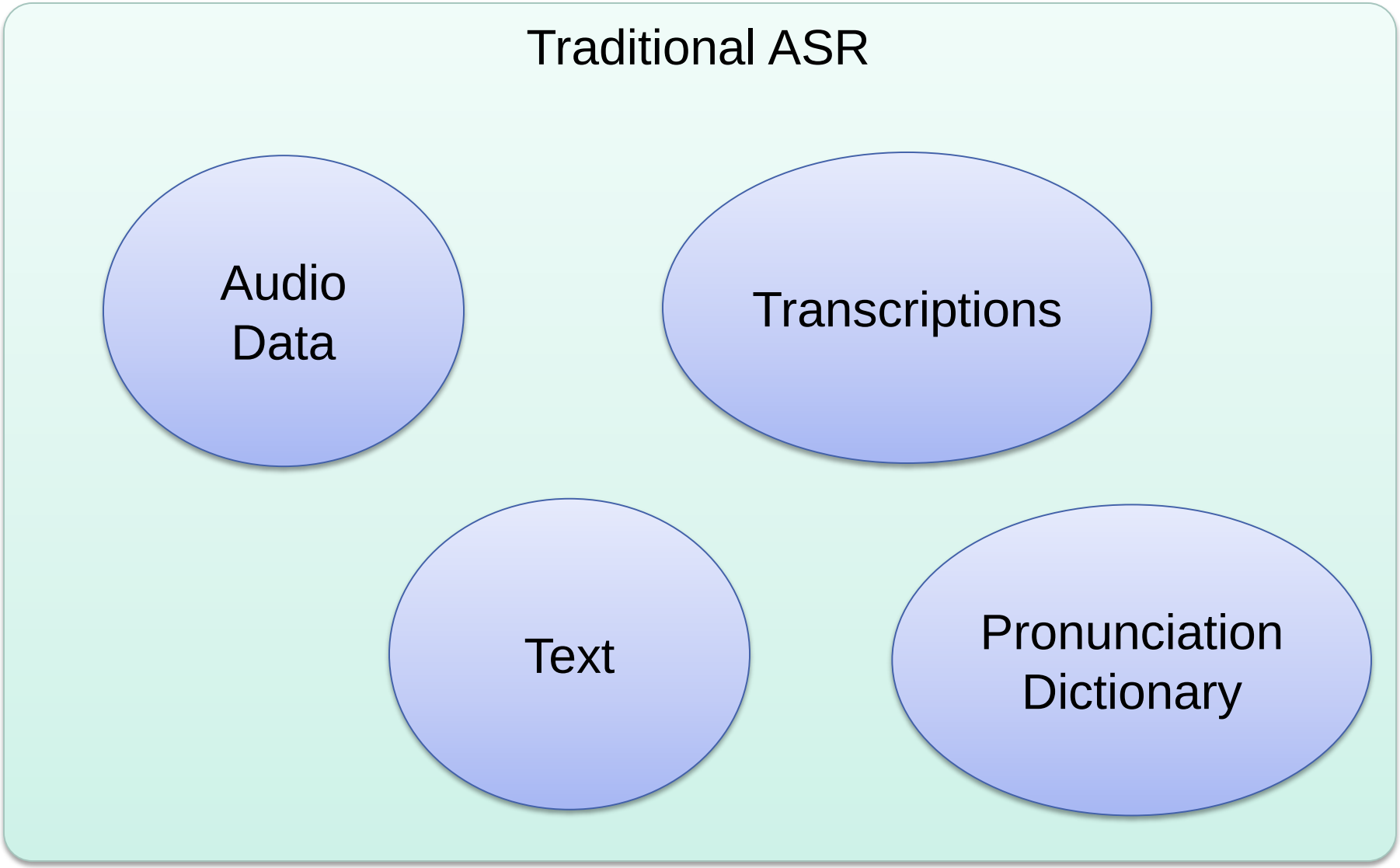
Felix Stahlberg, Tim Schlippe, Stephan Vogel, Tanja Schultz

24 April 2015

40th International Conference on Acoustics Speech and Signal Processing (ICASSP 2015)



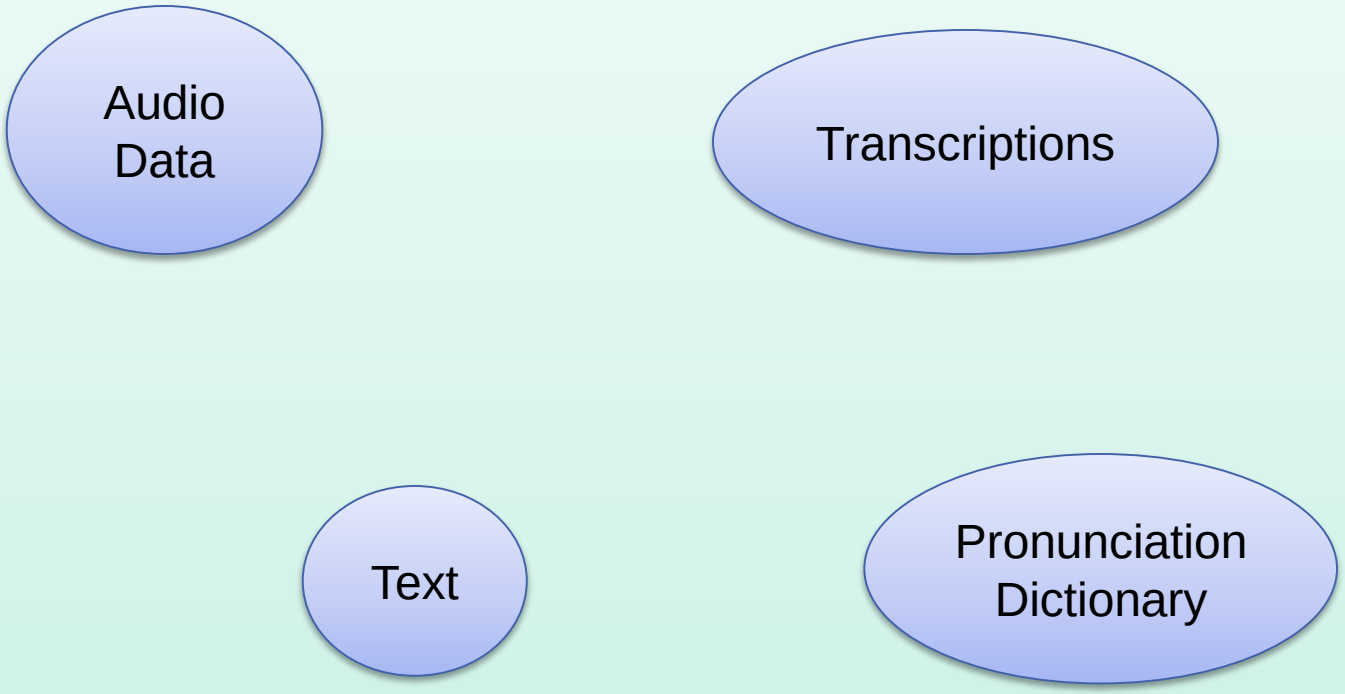
Resources for Automatic Speech Recognition



Resources for Automatic Speech Recognition

Traditional ASR

Under-Resourced ASR (Besacier et al., 2014)

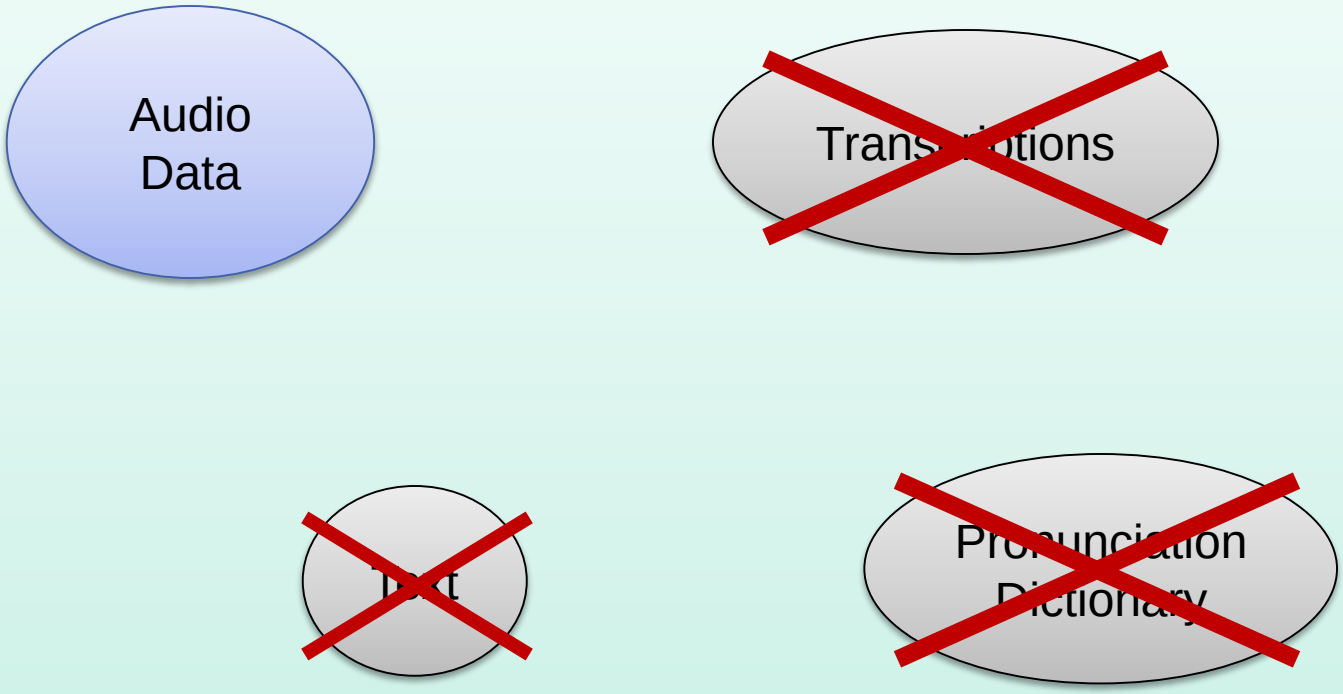


Resources for Automatic Speech Recognition

Traditional ASR

Under-Resourced ASR (Besacier et al., 2014)

Zero-Resource ASR (Jansen et al., 2012)

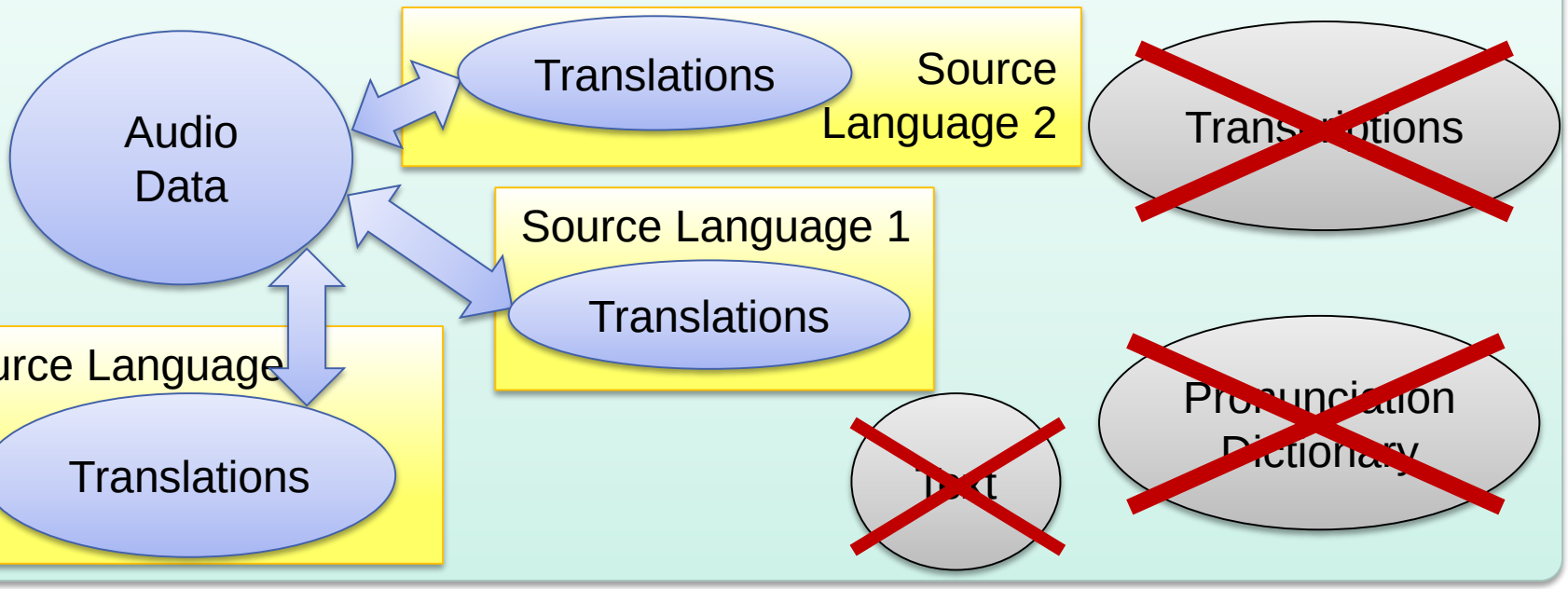


Resources for Automatic Speech Recognition

Traditional ASR

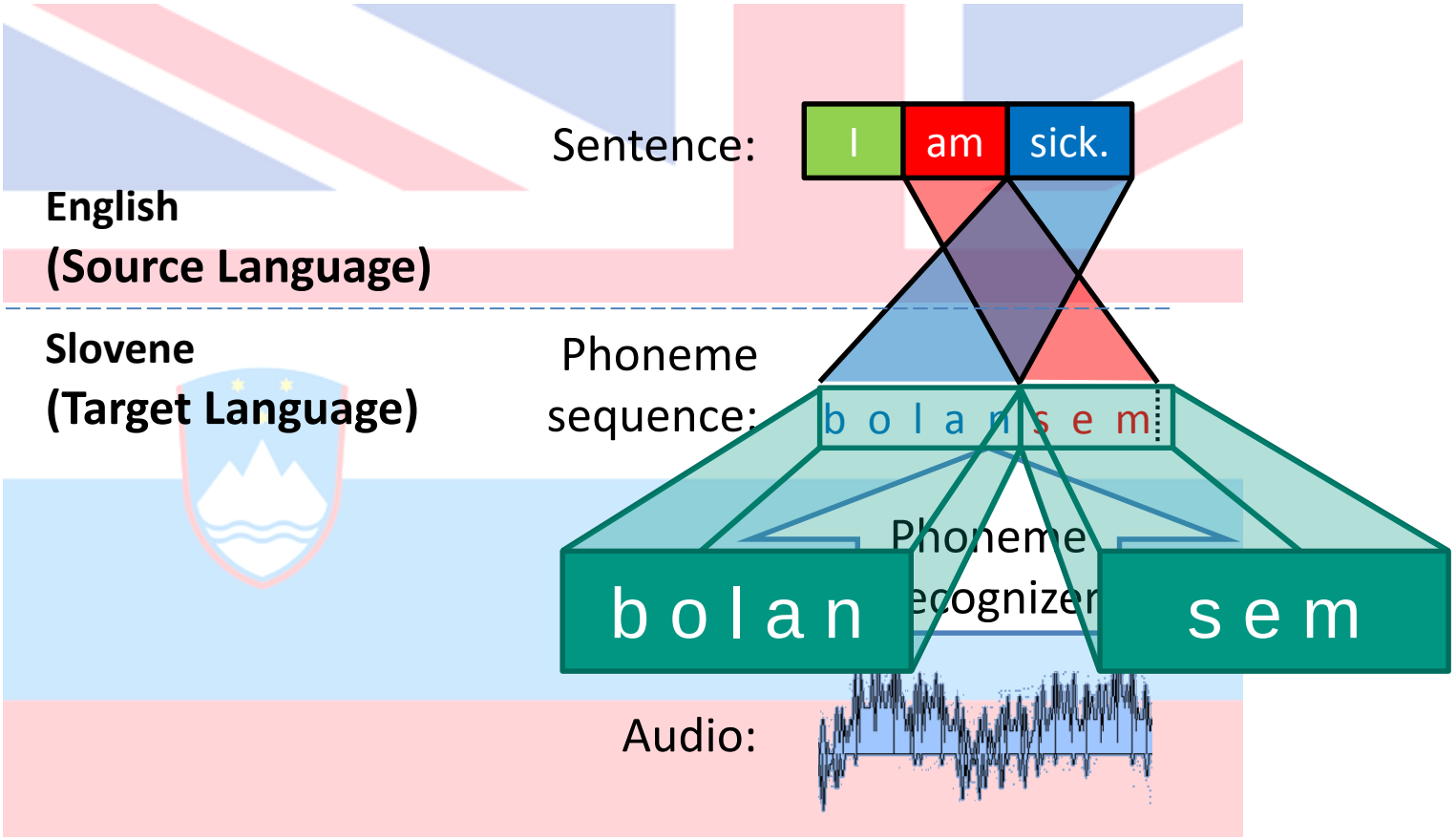
Under-Resourced ASR (Besacier et al., 2014)

Cross-Lingual Language Discovery



Zero-Resource ASR (Jansen et al., 2012)

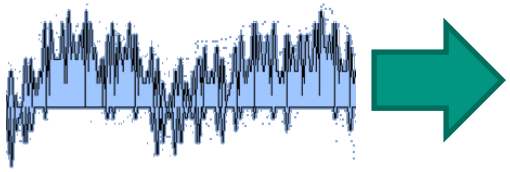
Cross-Lingual Word-to-Phoneme Alignment



PISA Alignment Tool: http://pisa.googlecode.com/

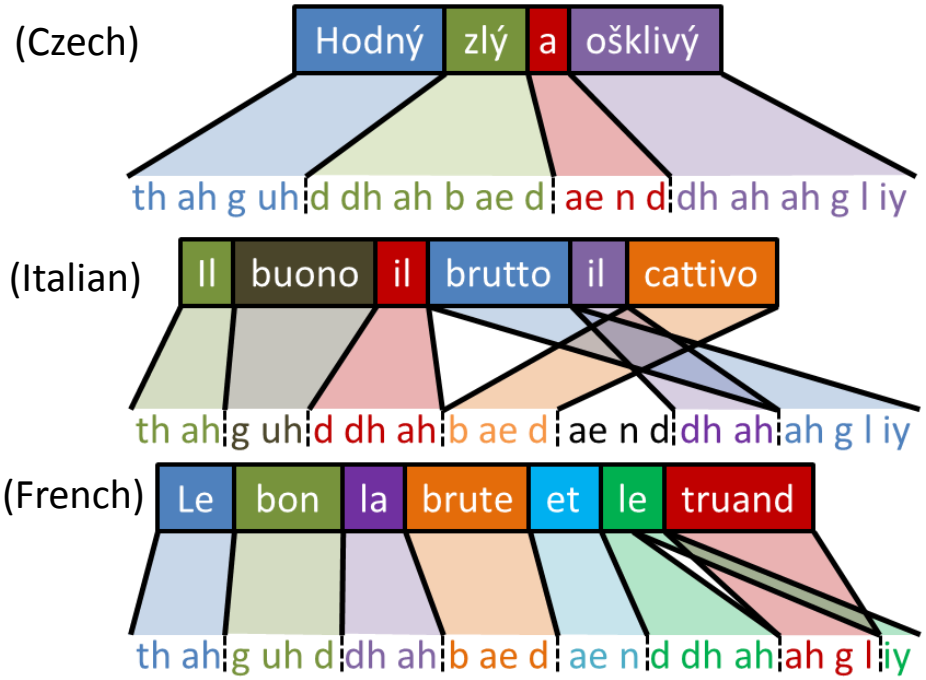
Multiple Translations (Step 1)

Step 0 (Phoneme Recognition)



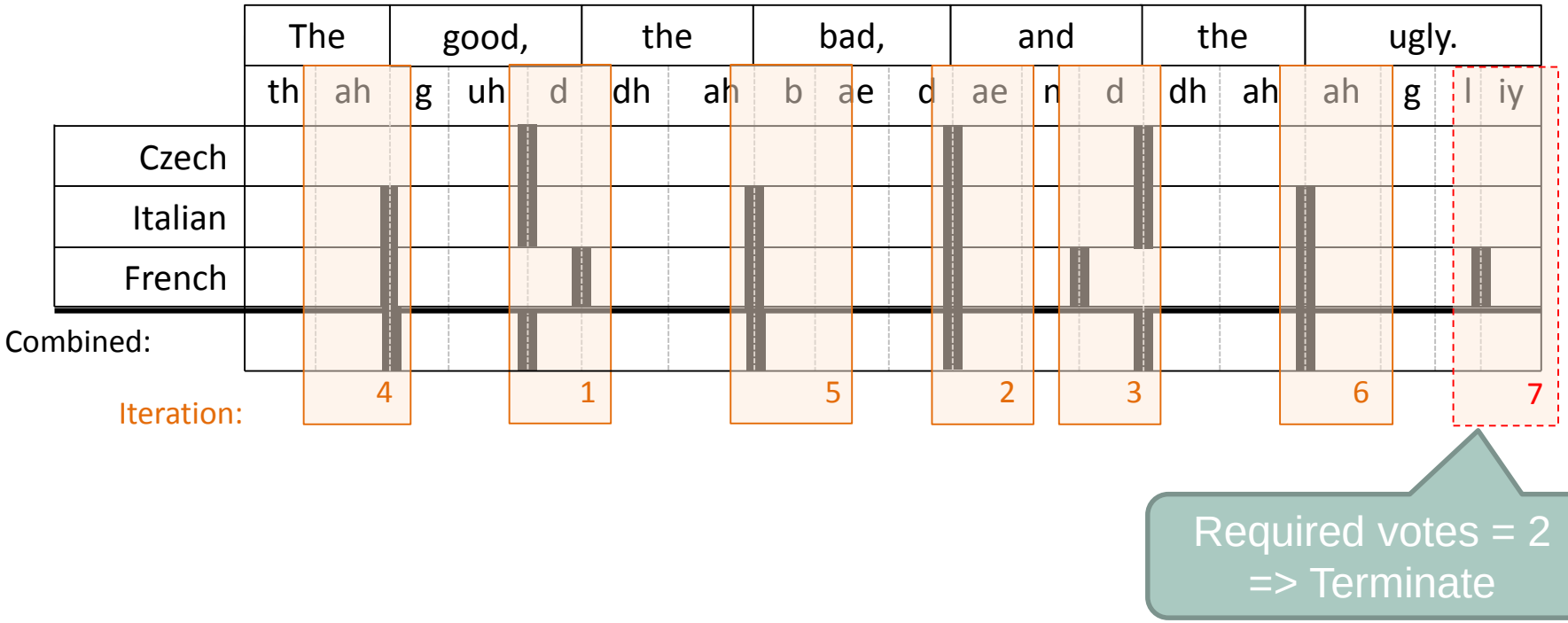
th ah g uh d dh da b ae d ae n d dh ah ah g l iy

Step 1 (Word-to-phoneme Alignment)



Multiple Translations (Step 2)

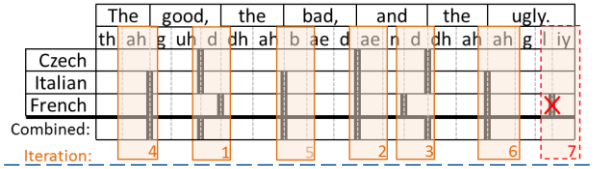
Step 2 (Segmentation Combination – voting with position tolerance)



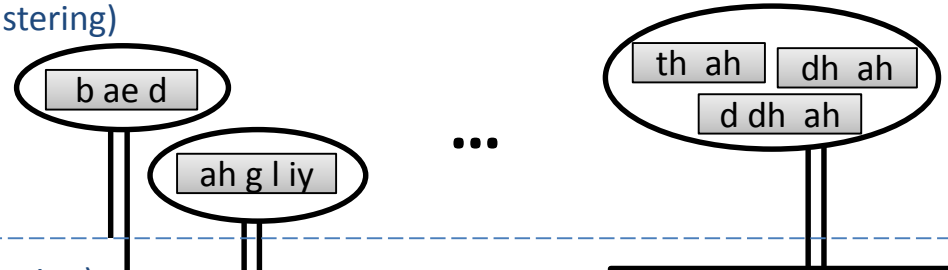
Multiple Translations (Overview)

Step 1 (Word-to-phoneme Alignment)

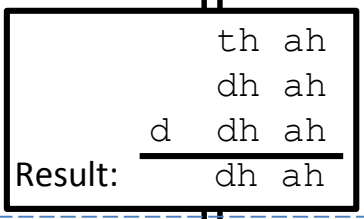
Step 2 (Segmentation Combination)



Step 3 (Phoneme Sequence Clustering)



Step 4 (Phoneme Level Combination)



Step 5 (Dictionary Generation)

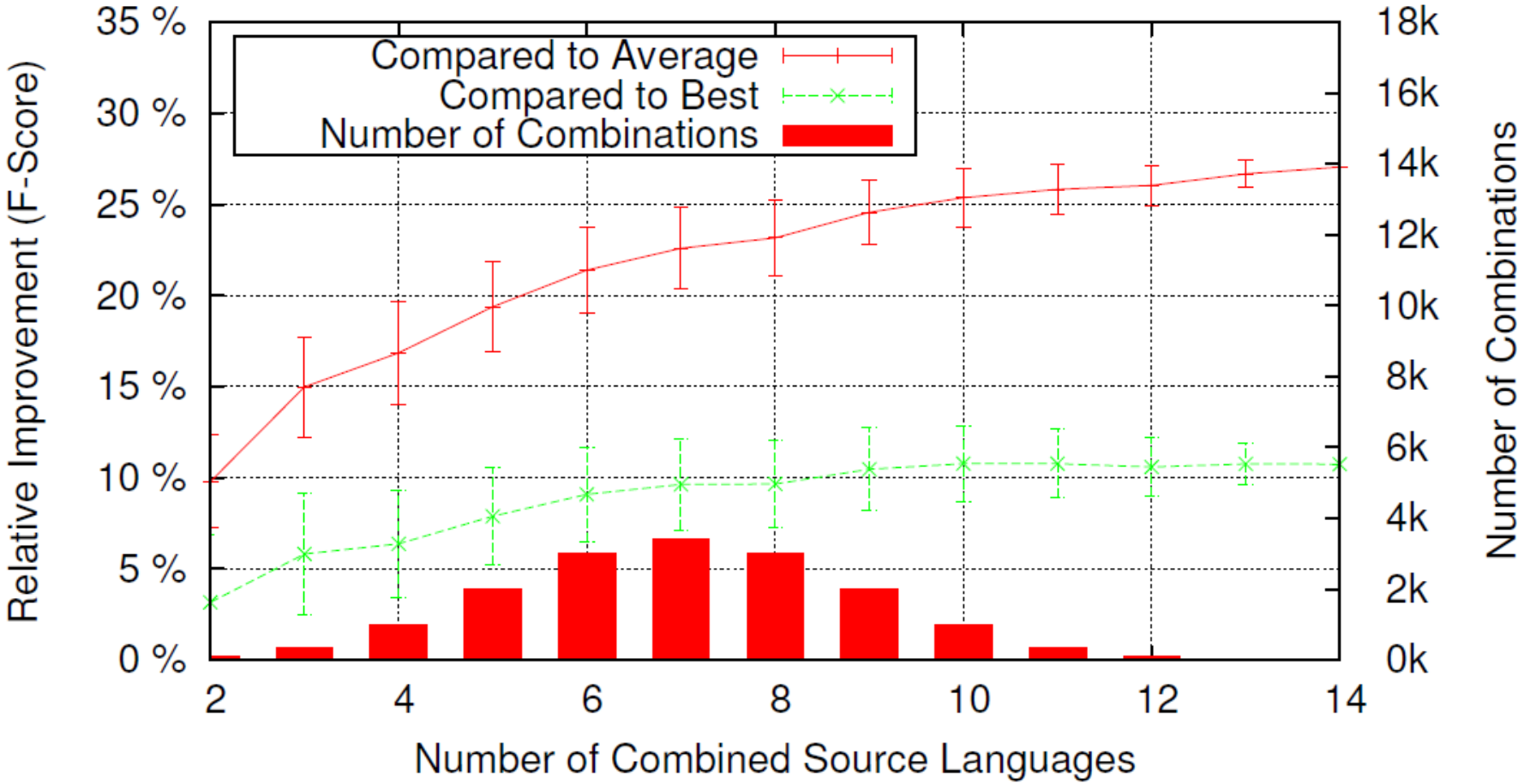
Label	Pronunciation
1	dh ah
2	g uh d
3	b ae d
4	ae n d
5	ah g l iy

Experimental Setup

- 23k parallel verses from the Christian Bible
- Target Language: English
- Phoneme Recognition:
  - Oracle (0%PER)
  - English phoneme recognizer (13.1%PER)
- Source Languages:

ID	Source Language	Full Source Language Bible Version Name	F-score (in %)
es3	Spanish	La Biblia de las Américas	77.5
es2	Spanish	Reina-Valera 1960	74.2
pt2	Portuguese	João Ferreira de Almeida Atualizada	73.2
fr2	French	Louis Segond	72.9
de1	German	Schlachter 2000	72.1
de2	German	Luther Bibel	72.0
it	Italian	Nuova Riveduta 2006	71.8
fr1	French	Segond 21	67.6
da	Danish	Dette er Biblen på dansk	67.4
pt1	Portuguese	Nova Versão Internacional	66.7
es1	Spanish	Nueva Versión Internacional	63.5
bg	Bulgarian	Bulgarian Bible	64.1
se	Swedish	Levande Bibeln	51.7
cs	Czech	Bible 21	51.6

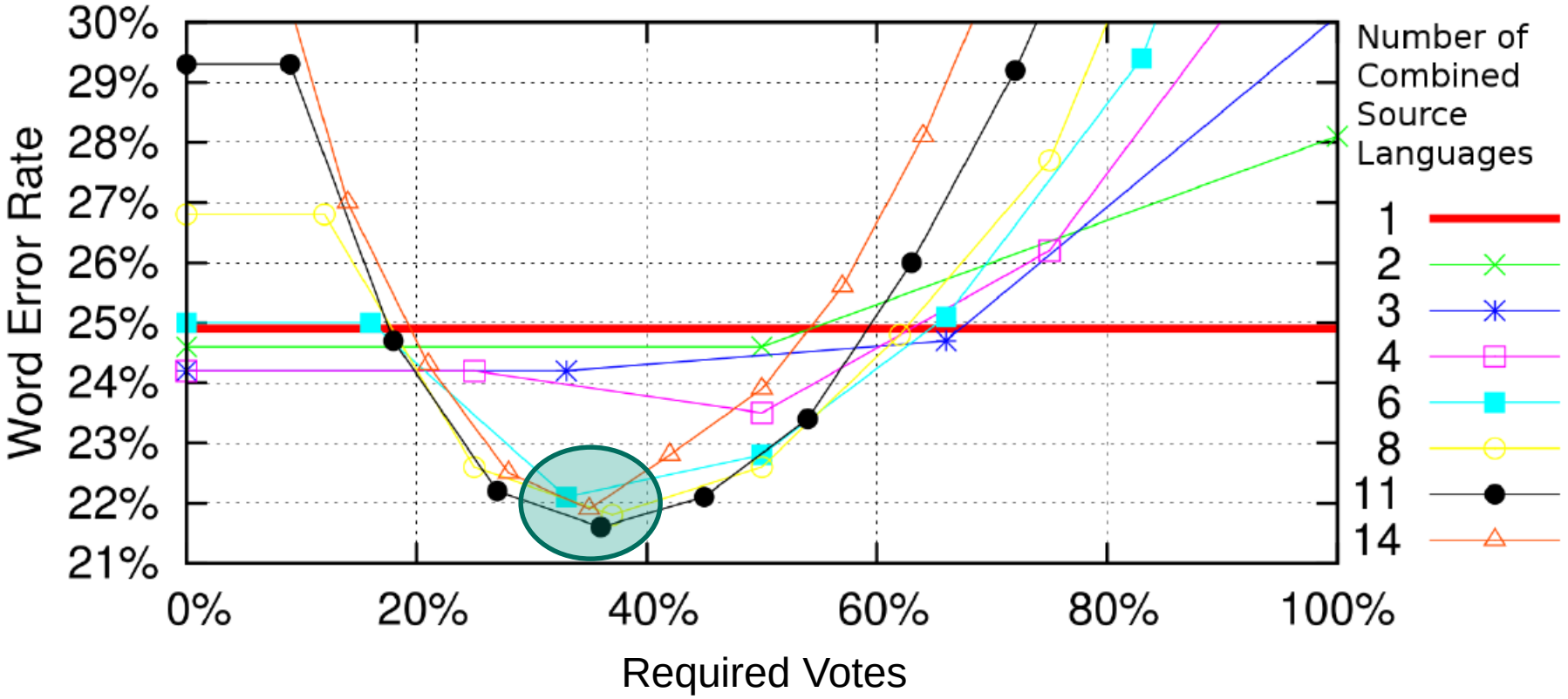
Segmentation Improvements over Number of Source Languages



- Required votes = 25%

Phoneme Recognizer: 0% PER

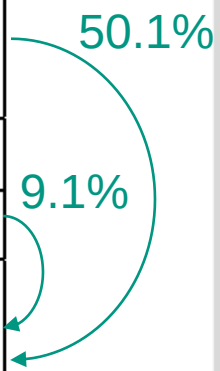
ASR Word Error Rate over Required Votes



Phoneme Recognizer: 0% PER

Overview of Improvements

Required Resource	Method	WER (in %)
target language phoneme sequence	Adaptor Grammars (*colloc2* grammar)	59.9
+ 1 source translation translation (*es3*)	GIZA++	37.3
	PISA	32.9
+ 8 additional translations	segmentation combination ($\epsilon = 33\%$)	29.9



Adaptor Grammars: (Johnson et al, 2006)

Phoneme Recognizer: 13.1% PER

Conclusion

- Cross-lingual lexical language discovery for non-written ASR in a speech-to-speech translation scenario
- Multiple translations help to improve ASR performance for non-written languages
  - 9.1% relative improvement in Word Error Rate compared to only one source language
  - 50.1% relative improvement in Word Error Rate compared to monolingual word segmentation

References

- Besacier, et al. "Automatic speech recognition for under-resourced languages: A survey." *Speech Communication* 56 (2014): 85-100.
- Jansen, et al. "A summary of the 2012 JHU CLSP workshop on zero resource speech technologies and models of early language acquisition."*ICASSP*. 2013.
- Stüker, Sebastian, and Alex Waibel. "Towards human translations guided language discovery for ASR systems." *SLTU*. 2008.
- Stahlberg, et al., “Word Segmentation Through Cross-Lingual Word- to-Phoneme Alignment,” in *SLT*, 2012
- Stahlberg, et al. "Word segmentation and pronunciation extraction from phoneme sequences through cross-lingual word-to-phoneme alignment."*Computer Speech & Language* (2014).
- Johnson, Mark, Thomas L. Griffiths, and Sharon Goldwater. "Adaptor grammars: A framework for specifying compositional nonparametric Bayesian models." *Advances in neural information processing systems*. 2006.

THANK YOU