

The Past, the Present, the Future of Spoken Language Translation Evaluation

Marcello Federico
FBK, Trento, Italy

InterACT25, 14-15 July 2016, Baden Baden, Germany

What is IWSLT?

- **International Workshop Spoken Language Translation**

- Annual event since 2004 (12 editions)
- Evaluation Campaign + Technical Workshop

- **Organisers**

- Chairs: KIT (Germany), FBK (Italy), CMU (USA)
- International steering committee

- **Important activities**

- WIT3.fbk.eu website with parallel data of TED talks
- Evaluation methods

Core Tasks

- **Automatic Speech Recognition (ASR)**

- English *speech* → English *transcript*

- **Machine Translation (MT)**

- English text → French text

- **Spoken Language Translation (SLT)**

- English *speech* → French text

- English *transcript* → French text

Son of the STARs



1992 - 2006

from demonstrations to
open evaluations



1993-2000

demonstration systems +
internal evaluations



2003-2007

integration + open evaluations

C-STAR > IWSLT (Kyoto 2005)

Is this a C-STAR
or a IWSLT
dinner?



Son of the STARs



1992 - 2006

from demonstrations to
open evaluations



1993-2000

demonstration systems +
internal evaluations



2003-2007

integration + open evaluations

TC-STAR – (Seville, 2003)

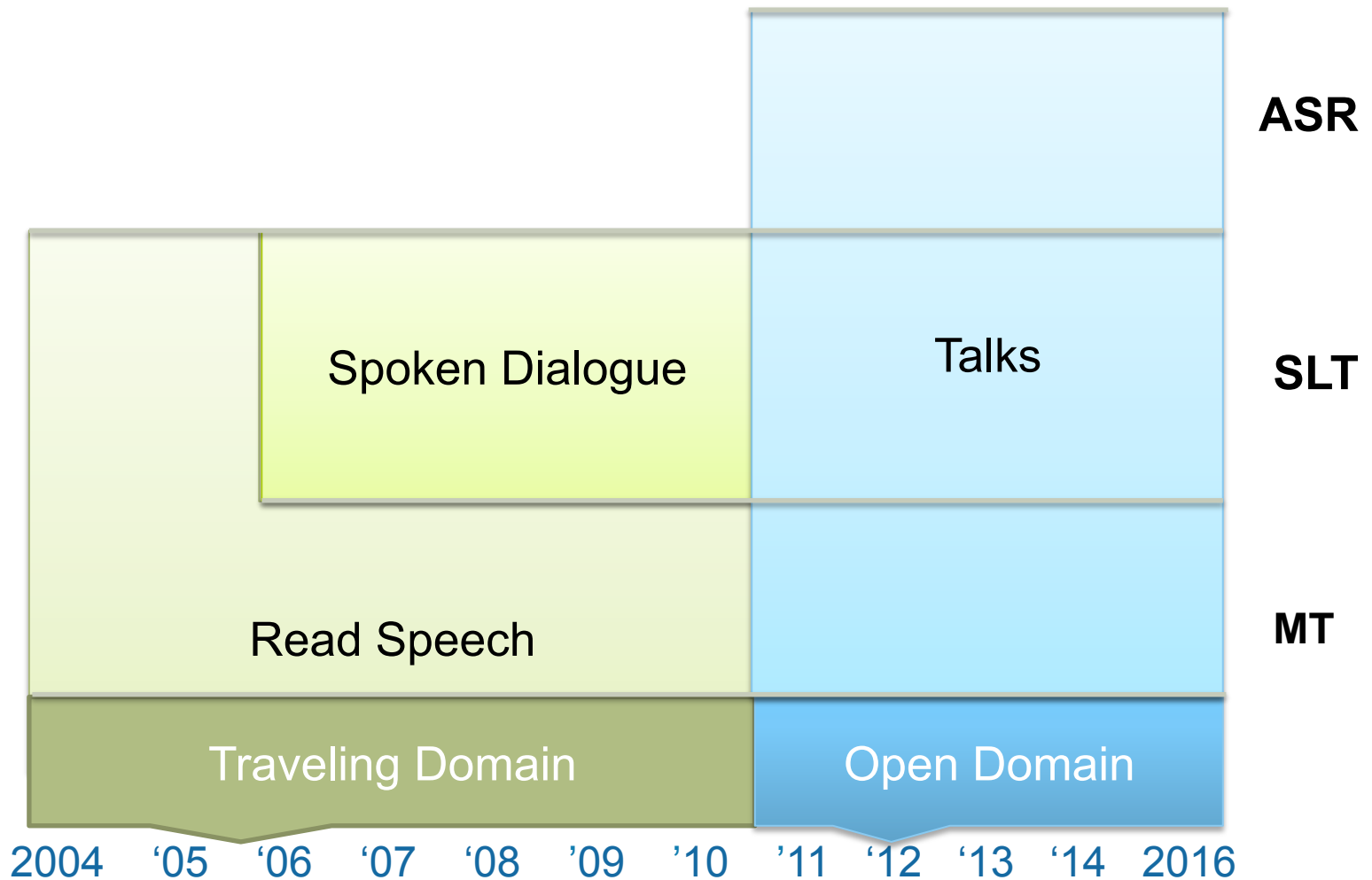
We should
evaluate all
our systems

...and make
them
accessible
via UIMA

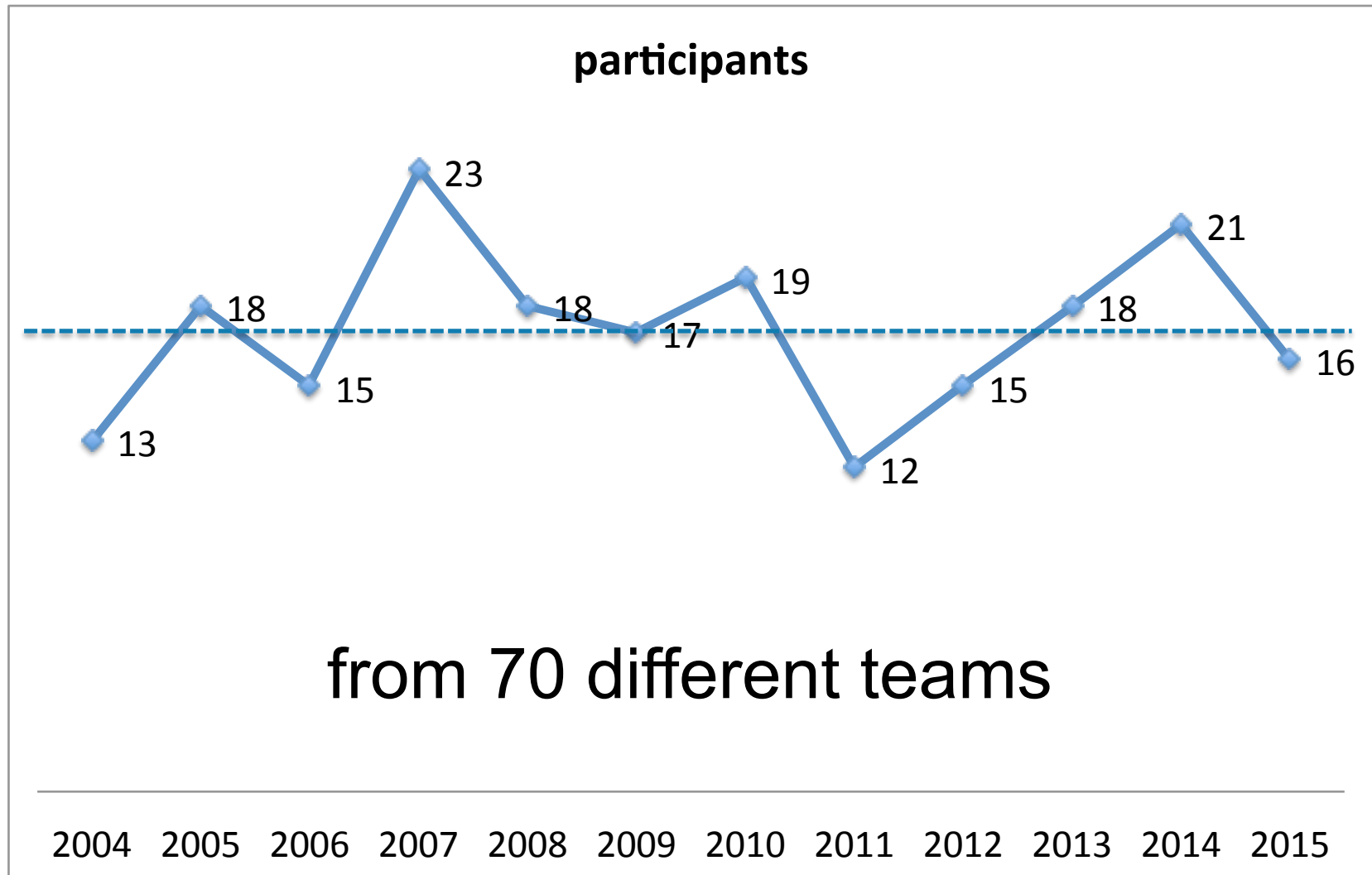
Couldn't we
just simply
evaluate?



IWSLT Evaluation Scenarios



IWSLT participants



TED Talks

TED Ideas worth spreading

Themes: TED Conferences, TED Community, About TED
Speakers: TEDx Events **NEW**, TED Blog
Talks: TED Prize
Translations **NEW**: TED Fellows

Search

Riveting talks by remarkable people, free to the world
Available in العربية, Deutsch, हिन्दी, ไทย, Русский, and more More about the [TED Open Translation Project](#).

Resize by:

- ☒ Newest releases
- ☐ Date filmed
- ☐ Most languages
- ☐ Most emailed this week
- ☐ Most comments this week
- ☐ Rated jaw-dropping
- ☐ ... persuasive
- ☐ ... courageous
- ☐ ... ingenious
- ☐ ... fascinating
- ☐ ... inspiring
- ☐ ... beautiful
- ☐ ... funny
- ☐ ... informative

Show talks related to:

- ☒ Technology
- ☐ Entertainment
- ☐ Design
- ☐ Business
- ☐ Science
- ☐ Global issues
- ☐ All

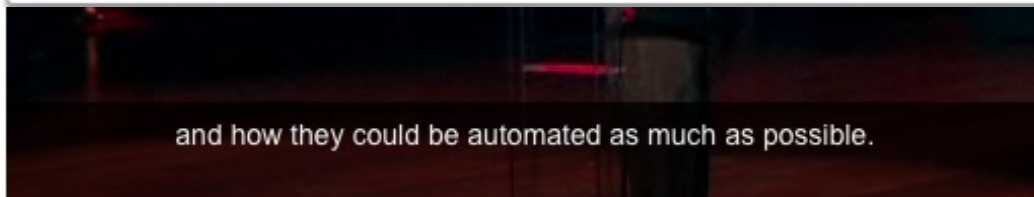
[View all tags >](#)

Featured Talks:

- Dan Phillips: Creative houses from reclaimed stuff**
- Tom Chatfield: 7 ways games reward the**
- Jason Fried: Why work doesn't happen at work**
- Miwa Matreyek's glorious visions**
- Shimon Steinberg: Natural pest control ...**
- R.A. Mashelkar: Breakthrough designs**
- Heribert Watzke: The brain in your gut**
- Ze Frank's web playroom**
- Natalie Jeremijenko: The art of the eco-**
- Conrad Wolfram: Teaching kids real math with computers**
- Peter Haas: Haiti's disaster of**
- Barbara Block: Tagging tuna in the**
- David Bismark: E-voting without fraud**
- Eben Bayer: Are mushrooms the new**
- Sebastian Seung: I am my connectome**

- TED LLC is non-profit
- Two annual events
- Short talks
- Variety of topics
- Website with:
 - Videos
 - Transcripts
 - Translations
- CC License

Human task: subtitling and translating



- ✓ segment audio
- ✓ transcribe and annotate
- ✓ split into captions
- ✓ translate captions

TED Talks Translations

	Nov '10	Nov '11	Nov '12	Nov '13	Nov '14	Nov '15
Talks (EN)	800	1,080	1,395	1,650	1,875	2,095
Languages	80	83	93	103	105	109
Translators	4,000	6,823	8,382	11,010	18,699	15,487
Translations	12,500	24,287 +94%	32,707 +34%	49,607 +52%	65,290 +32%	83,265 +28%

Challenges

- **Language modelling**

- Limited in-domain training data
- Variability of topics and styles

- **Acoustic modelling**

- Speaker: accent, fluency, speaking rate, style, , ...
- Noise: some mumble, applause, laughs, music, ...

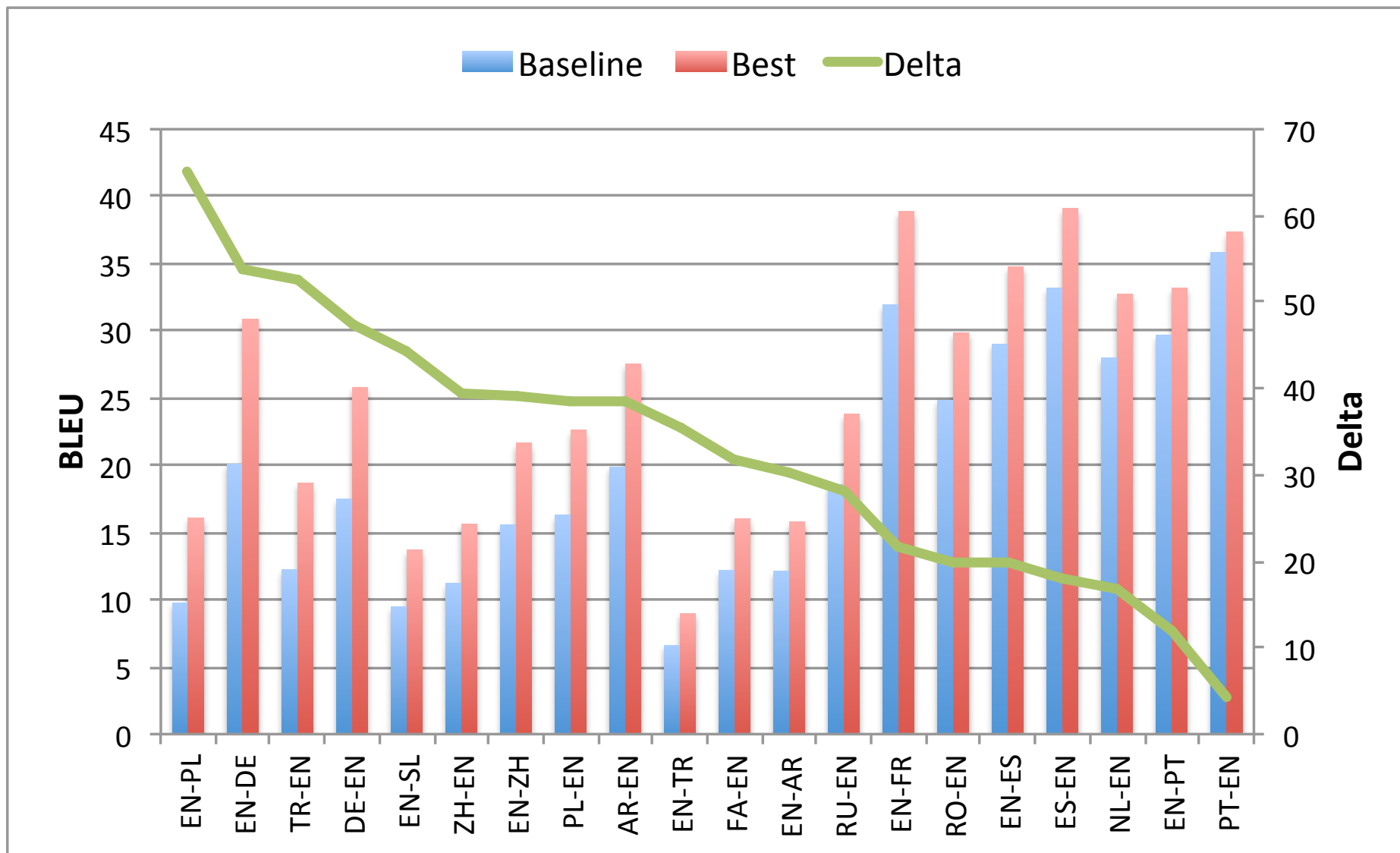
- **Translation modelling**

- Distant and under-resourced languages
- Morphologically rich languages

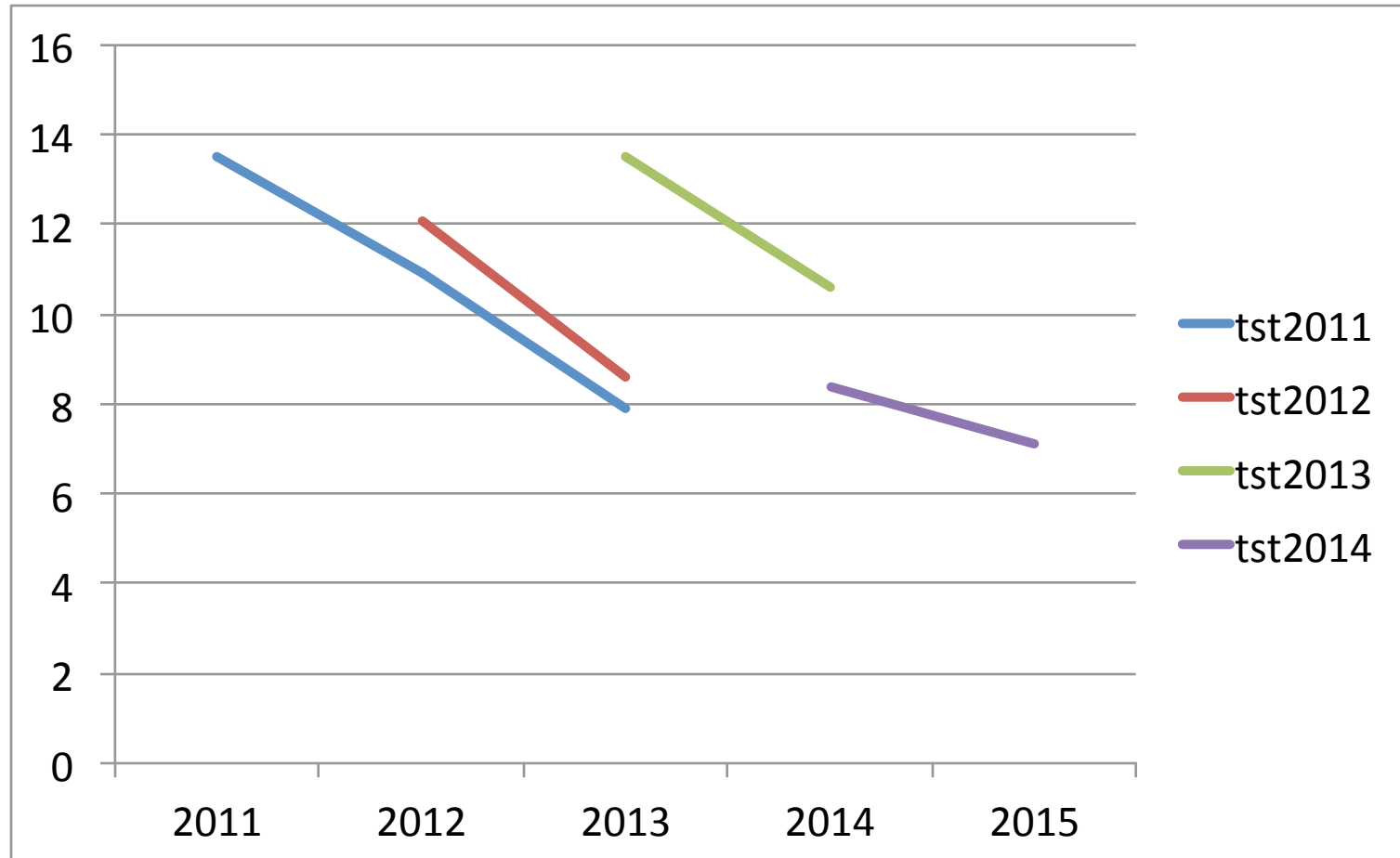
- **Speech Translation**

- Detection and removal of non-speech events
- **Subtitling and translating**

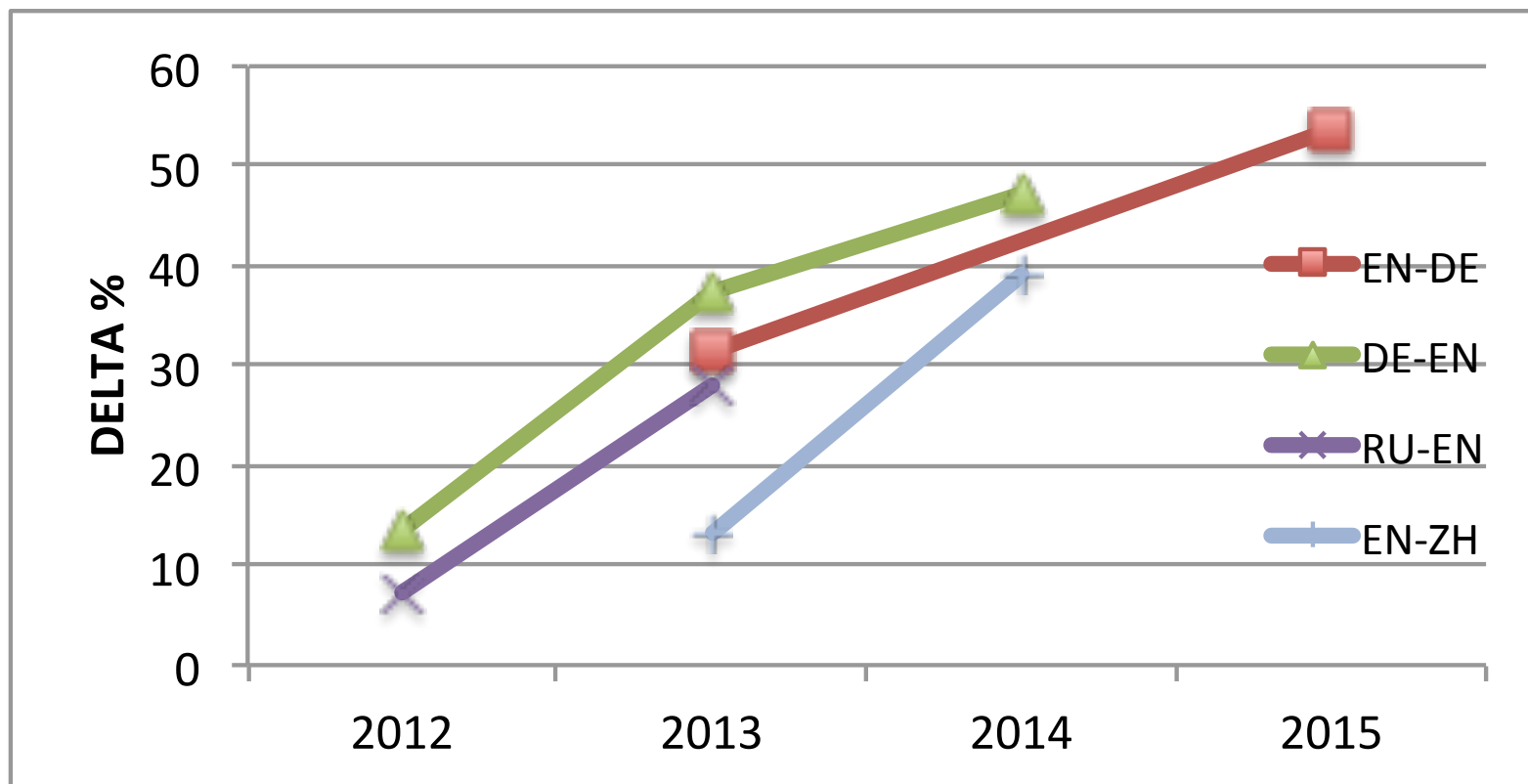
Best Measured MT Performance



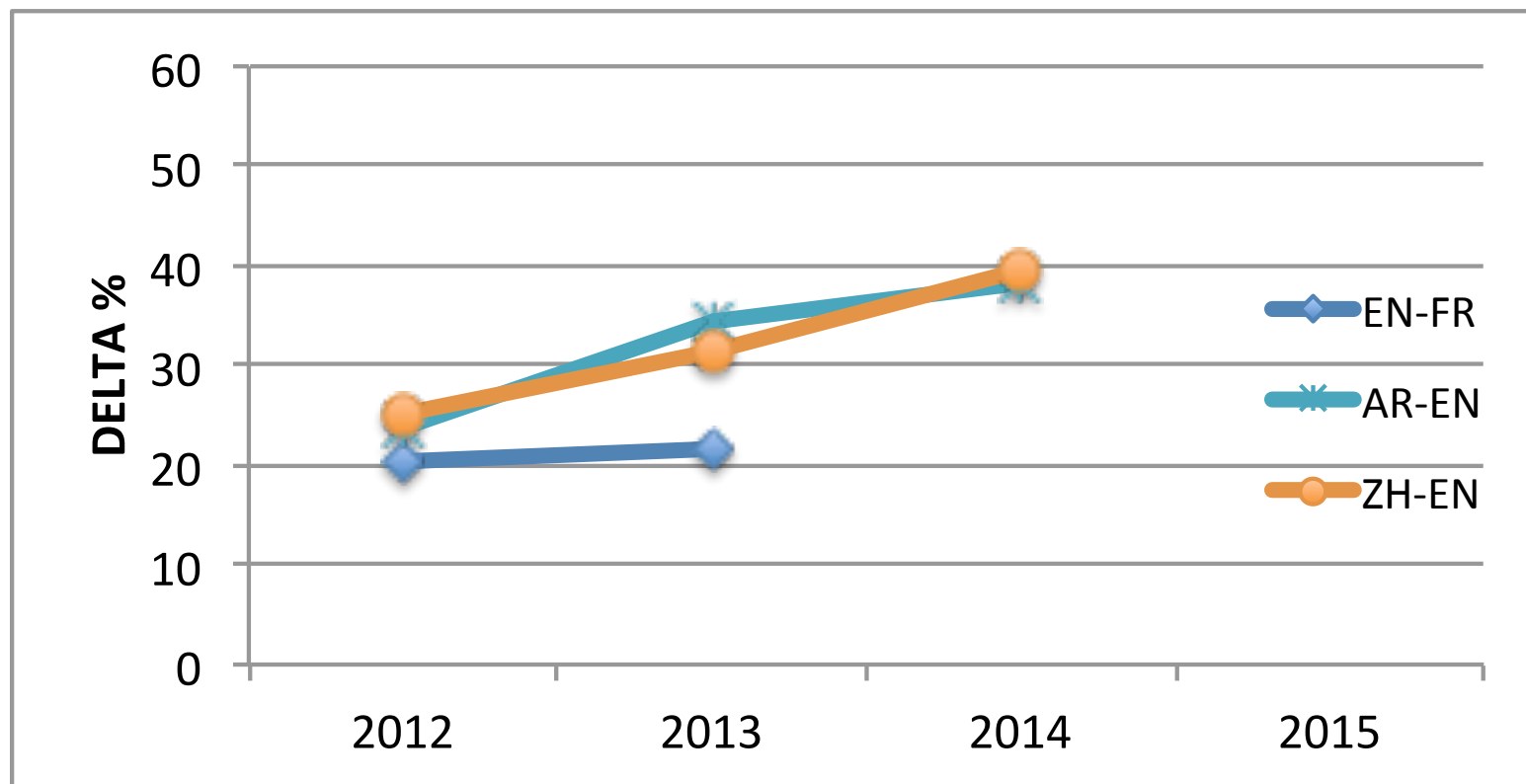
Progress in ASR En (best systems WER%)



Best MT Performance over Time



Best MT Performance over Time



Human Evaluation

- **Pairwise ranking** [2010-2012]

- *native volunteers judge quality of MT outputs*
- pros: relatively cheap
- cons: no re-use of data

- **Post-editing effort** [2013-2015]

- *professional translators post-edit MT outputs*
- cons: more expensive (~7 cents/word)
- pros: multiple references as by product
- pros: permits profiling of MT systems

Skype Translator Task



Challenges

- **Language modelling**

- Limited in-domain training data
- Variability of topics and styles

- **Acoustic modelling**

- Speaker: accent, fluency, speaking rate, style, , ...
- Noise: spontaneous speech, disfluencies

- **Translation modelling**

- Morphologically rich languages

- **Speech Translation**

- From spontaneous speech to polished text
- **Simultaneous translation**



IWSLT 2016
Seattle, US
December 08-09, 2016
www.iwslt.org

International Workshop on Spoken Language Translation

Credits

➤ **Language resources**

- TED LLC, USA (Talk data)
- Workshop Machine Translation (Giga and news data)
- DFKI, Germany (United Nations data)
- Microsoft Research (Skype Translator Task)
- Qatar Computing Research Institute (QED Lectures)

➤ **Funding**

- H2020 CSA CRACKER
- H2020 QT21

Questions?