

Surprise Language Evaluation: Rapid-Response Cross-Language IR

Maryland: Douglas W. Oard, Marine Carpuat, Petra Galuscakova,
Joseph Barrow, Suraj Nair, Xing Niu, Han-Chin Shing,
Weijia Xu, Elena Zotkina

Columbia: Kathleen McKeown, Smaranda Muresan, Efsun Selin Kayi,
Ramy Eskander, Chris Kedzie, Yan Virin

Yale: Dragomir Radev, Rui Zhang

Cambridge: Mark Gales, Anton Ragni

Edinburgh: Kenneth Heafield



Looking Backward

- 1966: ALPAC report
 - Refocus investment on enabling technologies
- 1988: IBM's Candide MT system
 - Data-driven approach
- 2003: DARPA TIDES surprise languages
 - Cebuano and Hindi
- 2017: IARPA MATERIAL program

Surprise Language Exercises

TIDES (2003)

English users / Docs in X
Time constrained
Research-oriented

MATERIAL (2019)

English users / Docs in X
Time constrained
Research-oriented

Zero-resource start
Digital text
Collaborative

Language pack start
Digital text and speech
Competitive

TIDES Schedule (2003)

Cebuano

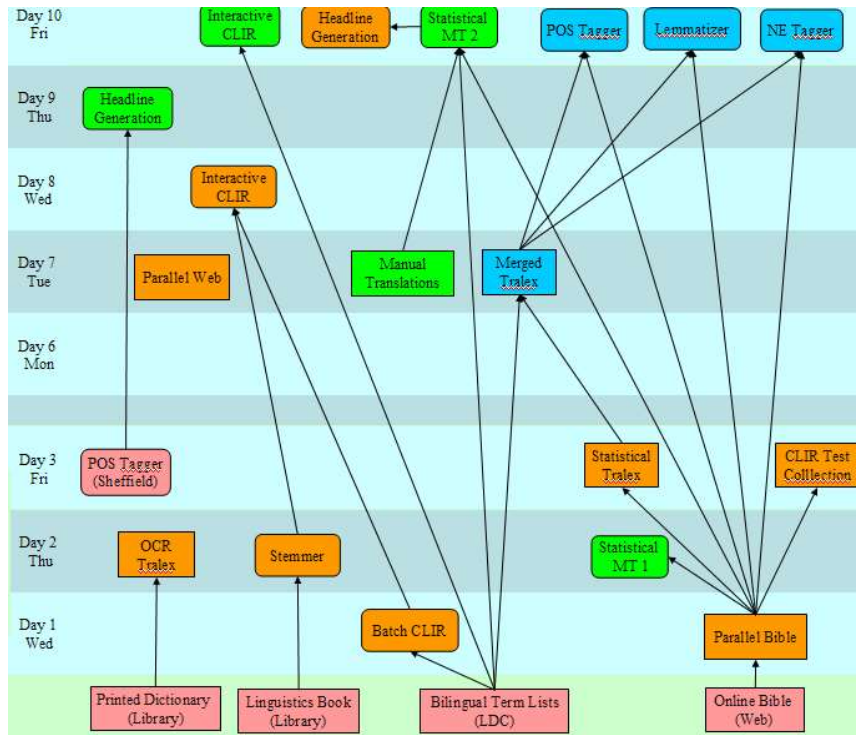
- Announce: March 5
- Test Data:
- Stop Work: March 14

Hindi

Jun 1
Jun 27
Jun 30

Cebuano Resources

- Bible: 913K words
- Examples of usage: 214K words (OCR)
- Communist Party Newsletter: 138K words
- Term list: 20K entries
- Web pages: 58K words
- Manual news translation
 - Discriminative training: 6K words
 - MT Evaluation: 13K words



Example Cebuano Translation

question transparent is our government ?

of salem arellano , mindanao scoop , **17 november 2002**

of so that day the **seminar that was held in america**

that from the four big official of the seven the place in

mindanao run until is in davao . the purpose of the

seminar , added of members orlando maglinao , is the

resistance to cause the **corruption in the government**

is be , ue , ue of our country .

Cebuano CLIR at Maryland

- Starting Point: iCLEF 2002 German system
 - Interface: “synonyms”/examples (parallel)/MT
 - Back end: InQuery/Pirkola’s method
- 3-day porting effort
 - Cebuano indexing (no stemming)
 - One-best gloss translation (bilingual term list)
- Informal Evaluation
 - 2 Cebuano native speakers (at ISI)

Hindi Results

- **Several components**
 - POS tags, morphology, time expression, parsing
- **5 evaluated tasks**
 - CLIR (English queries)
 - Topic tracking (English examples)
 - Machine translation into English
 - English “Headline” generation
 - Entity tagging
- **5 demos**
 - Interactive CLIR (2 systems)
 - Cross-language QA
 - Machine translation
 - Cross-document entity tracking

Hindi Resources

- Much more content available than for Cebuano
 - Total: 4.2 million words
- Large and diverse
 - Web, news, dictionaries, handbooks, hand translated, ...
- Huge effort: data conversion/cleaning/debugging
 - Many non-standard encodings
 - Often: no converters available or available converters do not work properly



Translation Elicitation Server

- Johns Hopkins University (David Yarowsky)

- People voluntarily translated large numbers of Hindi news sentences for nightly prizes at a novel Johns Hopkins University website
- Performance is measured by Bleu score on 20% randomly interspersed test sentences
- Allows immediate way to rank and reward quality translations and exclude junk
- **Result: 300,000 words of perfectly sentence-aligned bitext (exactly on genre) for 1-2 cents/word within ~5 days**
 - Much cheaper than 25 cents/word for translation services or 5 cents/word for a prior MT-group's recruitment of local students

Sample Interface:


अशोक लाटा, राजपाल मनमानी ने वार्ड नार७ बतरा थिल्डिआ से सिद्धी क्वॅलोनी, वार्ड नार८, १३ एवा॥१४ में सघन जनसंपर्क किया।

user (English) translations typed here...

डॉ. शर्मा ने इसके बाद बायई क्षेत्र में भी जनसंपर्क किया।

and here

Other options:

(1) View this page in alternate encoding  User choice of 2-3 encoding alternatives

© Standard UTF-8 Encoding (already installed in Windows) -- भारत

- Observed exponential growth in usage (before prizes ended)
 - **viral advertising** via family, friends, newgroups, ...
- **\$0 in recruitment, advertising, and administrative costs**
 - Nightly incentive rewards given automatically via amazon.com gift certificates to email addresses (any \$ amount, no fee)
 - no need for hiring overhead. Rewards only given for proven high quality work already performed (prizes not salary).
 - **immediate positive feedback encourages continued use**
- Direct immediate access to worldwide labor market fluent in source language

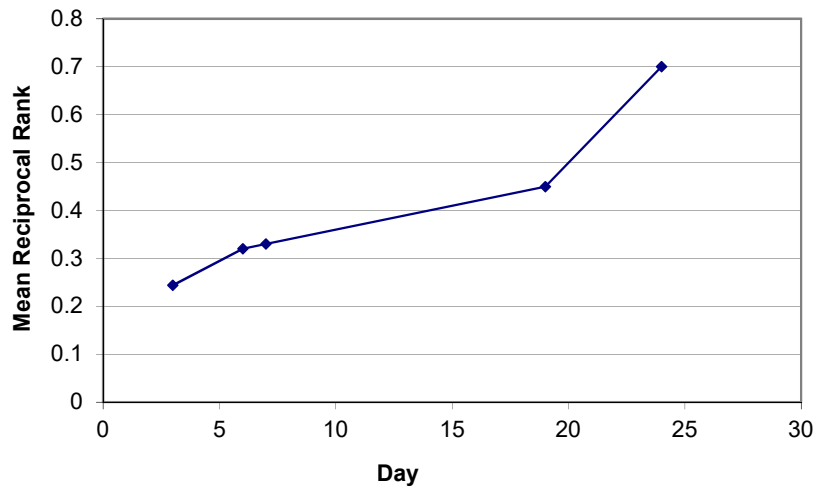
Example Hindi Translation

- **Indonesian City of Bali in October last year** in the bomb blast in **the case of imam accused** India of the sea on Monday began to be averted. **The attack on getting and its plan** to make the charges and decide **if it were found guilty, he death sentence** of May. Indonesia of the police said that the imam sea bomb blasts in his hand claim to be accepted. A **night Club** and time in the **bomb blast in more than 200 people were killed and several injured** were in which **most foreign nationals**. ...

Hindi CLIR

- N-grams (trigrams best for UTF-8)
- Relative Average Term Frequency (Kwok)
- Scanned bilingual dictionary (Oxford)
- More topics for test collection (29)
- Weighted structured queries (IBM lexicon)
- Alternative stemmers (U Mass, Berkeley)
- Blind relevance feedback
- Transliteration
- Noun phrase translation
- MIRACLE integration (ISI MT, BBN headlines)

Hindi CLIR Formative Evaluation



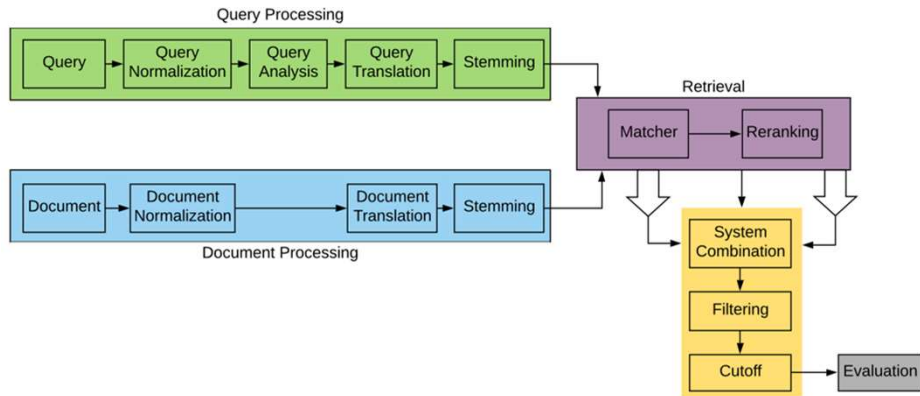
19 known item queries

Some Challenges in 2003

- Formative evaluation
- Synchronize variable-rate efforts
 - More like soccer than football
- Integration
- Capturing lessons learned



MATERIAL in 2019: CLIR Pipeline



Lithuanian ASR (Cambridge)



Day	Description	WER (%)		
		CTS	News	Topic
1	BABEL OP2 build	48.2	—	—
2	Baseline GMM-HMM	55.2	—	—
3	Baseline NN-HMM	41.1	62.9	53.1
4	Web language model	39.1	38.1	33.2
5	Speed perturbation	37.9	37.5	32.2
.
N	More text and audio	35.4	22.0	21.1

- Systems distributed to the team within 5 days marked in blue

Lithuanian MT

corpus	sentences
MATERIAL-BUILD (train)	42,000
ParaCrawl-BiCleaner-v4 (train)	815,713
MATERIAL-BUILD (dev)	1,654



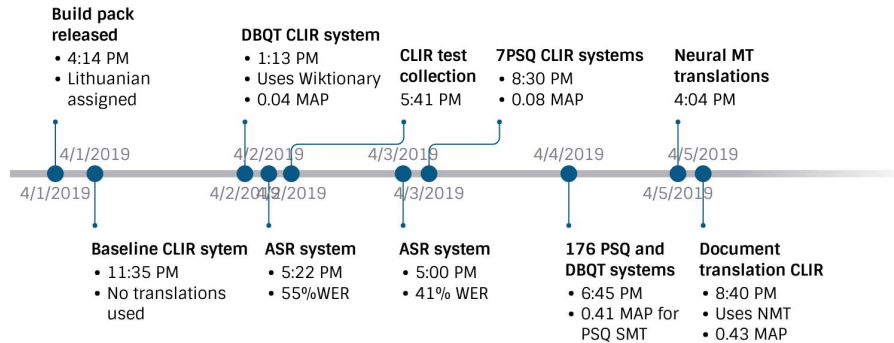
Edinburgh

newstest2019	BLEU-4	1-gram prec.
en-lt	21.52	51.4
lt-en	28.07	60.5

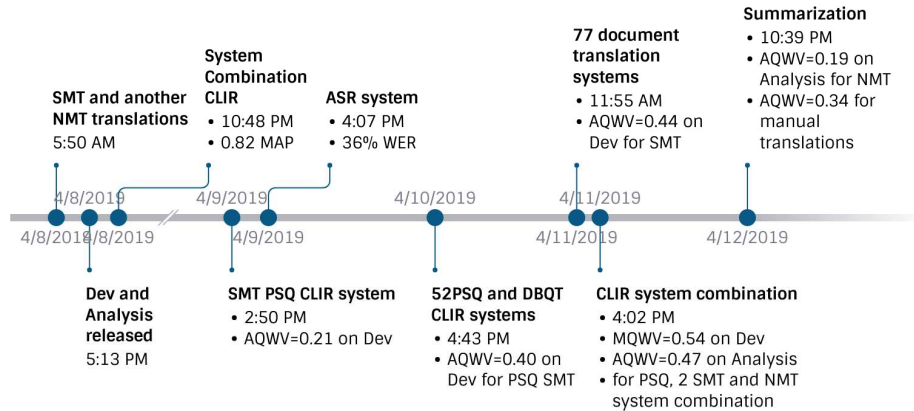
Maryland

newstest2019	System	BLEU-4	1-gram prec.
en-lt	SMT	13.00	44.6
	NMT	4.69	22.0
lt-en	SMT	20.73	56.2
	NMT	16.25	50.3

Lithuanian 5 Day Exercise: Week 1



Lithuanian ¹⁰~~8~~ Day Exercise: Week 2





Example Summary: “food shortage”

CLOSE MATCH (food shortage):

...can cause stress, some food products, air change, lack of food, misunderstandings, as well as many other factors.

Klasterinis headache pain It is quite rare, strong headache, which is more widespread between men than women. Klasterinis headache may arise one time during the day..



Example Summary: “food shortage”

Machine
Translation
Summary

CLOSE MATCH (food shortage):

...can cause stress, some food products, air change, lack of food, misunderstandings, as well as many other factors. Klasterinis headache pain It is quite rare, strong headache, which is more widespread between men than women. Klasterinis headache may arise one time during the day..

Human
Translation
Summary

CLOSE MATCH (food shortage):

...increases; it's often hereditary (to family members in one or several generations); a migraine can be caused by stress, some food products, changes in weather, lack of food, insomnia, also numerous other factors. cluster headache it's a quite rare, severe headache that's more prevalent among...



Human Evaluation on Query 1 - Analysis

	AQWV
CLIR (machine translation)	0.47
+ E2E (manual translation)	0.34
+ E2E (machine translation)	0.19

Some Lessons Learned

- Build on:
 - Existing infrastructure
 - Existing team
- Language packs enable rapid progress
 - Reuse it when the core technology improves
- Provide IR eval data on day 1

Lithuanian Surprise Language Hall of Fame

