Multi Language Support for Virtual Assistants

Sierra Kaplan-Nelson, Max Farr

Mentor: Mehrad Moradshahi

Broad Topic (everything we do now in many other languages)

أعطني معلومات
عن الانتخابات

- Speech recognition, speech -> text
- Machine translation
- Data collection
- **Question answering**
- **Semantic parsing**
- Guided learning
- Chatbots
- Etc., etc., ...

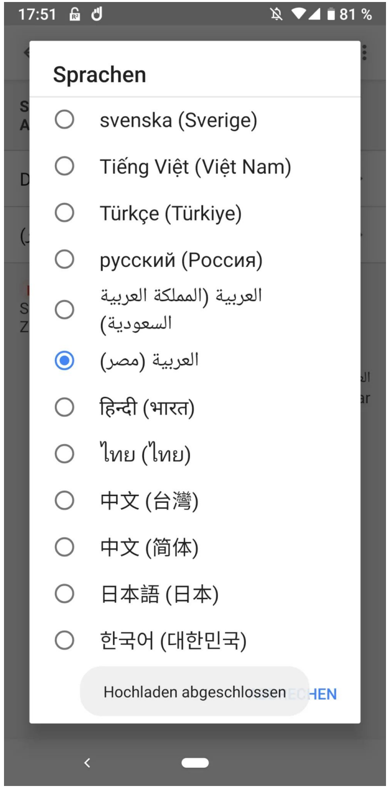
Overview of Machine Language Translation

أعطني معلومات
عن الانتخابات

- Previously all done via rules-based methods
- For awhile hybrid machine translation was the norm, where sentences were pre-processed using a rules engine before fed through an ML model
- Now almost all done by deep neural networks
- VAs in some ways are using hybrid machine translation since they can use templates

State of the Art VAs in Other Languages

- Google VA has most languages
  - Issues detecting accents
  - Started to employ AI on sound wave visualizations to improve language detection and spelling correction techniques to reduce errors by 29%
  - Supporting new language also involves localization that can take a month
- Question answering in other languages is active research topic, currently performs much worse than English
- VAs that perform specific tasks, like helping children learn, are almost exclusively in English



Arabic VA for Autistic Children (2019)

- Teaches both social behavior and academic skills, mostly using hardcoded flow diagrams and quizzes

[Autistic Innovative Assistant (AIA): an Android application for Arabic autism children]() (Sweidan, Salameh, Zakarneh & Darabkh)

Multi Language Question Answering

Supervised Learning to Improve Arabic Question Similarity Detection

- Arabic is poorly-informatized (not many knowledge graphs etc.)
- Uses rules to separate questions by broad type
- Created dataset of pairs questions from ejaaba.com (answer.com in Arabic) and hand labeled them as similar “Yes” or “No”
- Used paraphrasing to generate more “Yes” pairs
- Hybrid learning approach combining string and semantic similarity

ID	**Scope**	**Question words**	**Paraphrased words**
TimexF	Time - Factoid	متى, ايان “When”	في اي وقت “in what time” في اي سنة “in what year” ما هو تاريخ “what is the date”
LocF	Location - Factoid	أين Where	ما موقع “What is the location” في اي مدينة “in what city” “in which country” في اي دولة
NVF	Numeric value - Factoid	كم How many How Much	ما طول “what is the length” ما هي المسافة “what is the distance” ما عرض “what is the width”
NEF	Named Entity - Factoid	لمن Whose	الى من “for whom” من هو “Who is”

Novel Approach towards Arabic Question Similarity Detection (Daoud)

Multilingual Extractive Reading Comprehension (2018)

- Most high quality large datasets are annotated in English

- Seeks to increase RC in other languages without costly process of creating new large training datasets

- Translates question AND document context from language L into English with attentive NMT model and get answer in English

Multilingual Extractive Reading Comprehension by Runtime Machine Translation (Asai, Eriguchi, Hashimoto, and Tsuruoka)

Multilingual Extractive Reading Comprehension

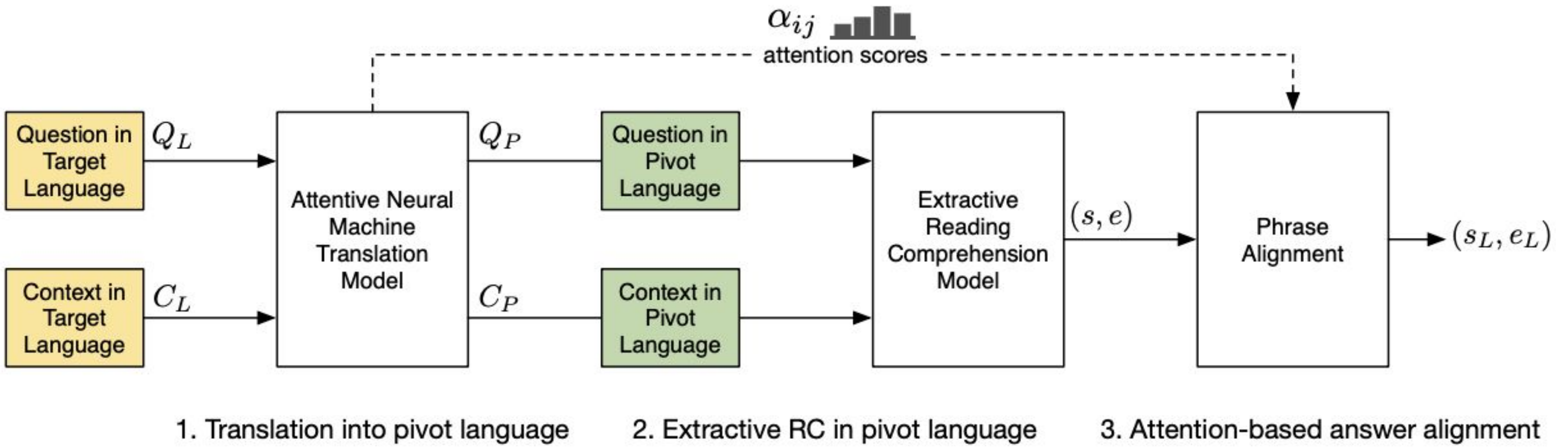


Figure 1: Overview of our method. $\boldsymbol{\alpha}_{ij}$ are the attention weights (attention distribution) in the NMT model. (s, e) and $(s_\mathrm{L}, e_\mathrm{L})$ are the answer spans in the pivot language (e.g. English) and target language L, respectively.

Multilingual Extractive Reading Comprehension by Runtime Machine Translation (Asai, Eriguchi, Hashimoto, and Tsuruoka)

Multilingual Extractive Reading Comprehension

- Recover answer in context in L using soft alignments from NMT
  - Alignment in this context is the start and end of the span in the text containing answer
- Found that how well questions are translated **significantly** affects performance
  - Using paraphrased questions decreased accuracy
  - Oversampling high quality translations in training improves performance
- Found that this method improved performance over just back translating English results with Google translate

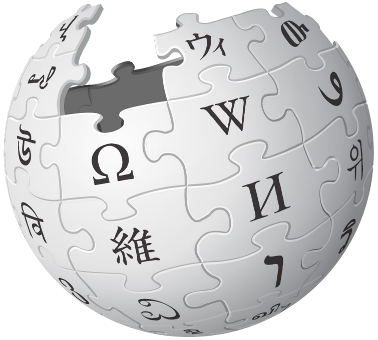
Method	Japanese		French	
	F1	EM	F1	EM
Our method	**52.19**	**37.00**	**61.88**	**40.67**
Back-translation by using Google Translate	42.60	24.77	44.02	23.54

Table 3: RC results of our method and the baseline on Japanese and French SQuAD. The BiDAF model trained on the original English SQuAD dataset achieves an F1 score of 77.1 and an EM score of 67.2.

Multilingual Extractive Reading Comprehension by Runtime Machine Translation (Asai, Eriguchi, Hashimoto, and Tsuruoka)

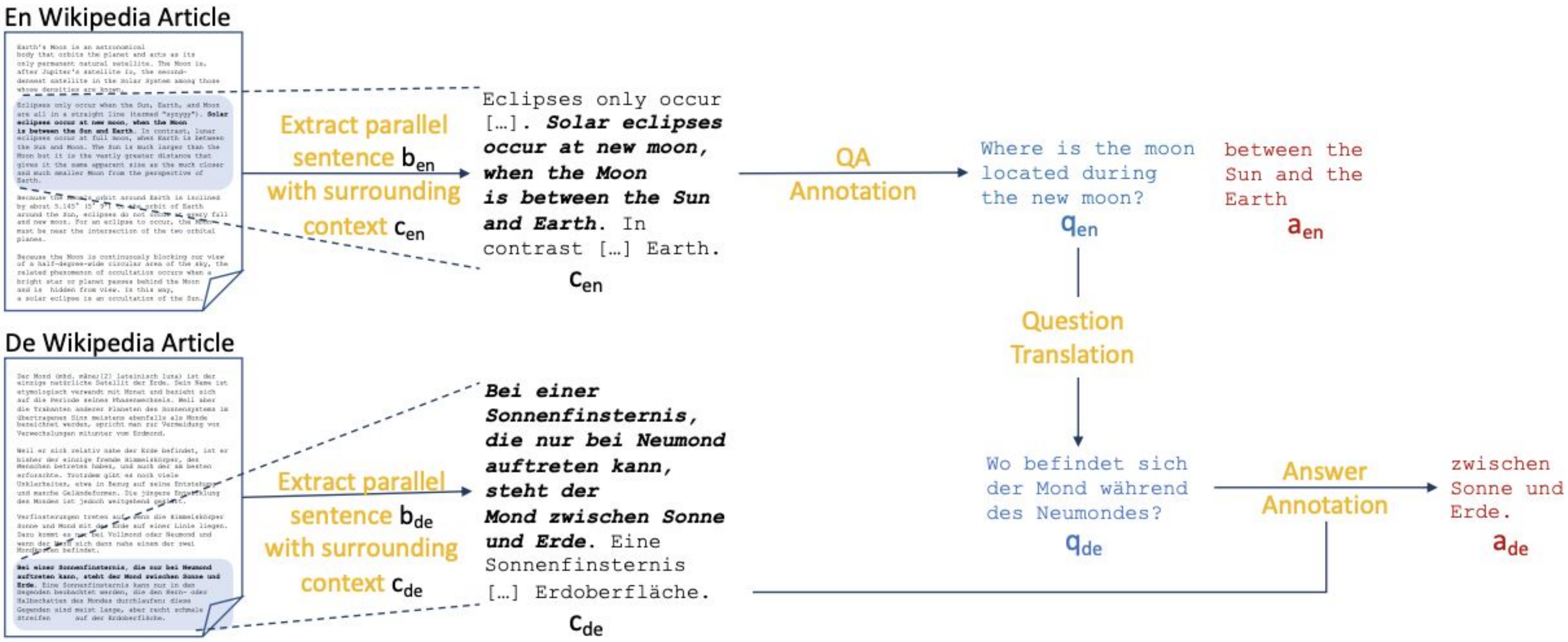
MLQA: Evaluating Cross-lingual Extractive Question Answering (2020)

- Benchmark datasets to compare with SQUAD to help speed up QA improvements in other languages
- Contains QA instances in 7 languages: English, Arabic, German, Spanish, Hindi, Vietnamese and Simplified Chinese
- MLQA has over 12K instances in English and 5K in each other language, with each instance parallel between 4 languages on average.
- Pulled text from Wikipedia articles that exist in many languages, then employed crowdsourced annotators



Multilingual Extractive Reading Comprehension by Runtime Machine Translation (Asai, Eriguchi, Hashimoto, and Tsuruoka)

MLQA: Evaluating Cross-lingual Extractive Question Answering (2020)



Multilingual Extractive Reading Comprehension by Runtime Machine Translation (Asai, Eriguchi, Hashimoto, and Tsuruoka)

Quiz 1

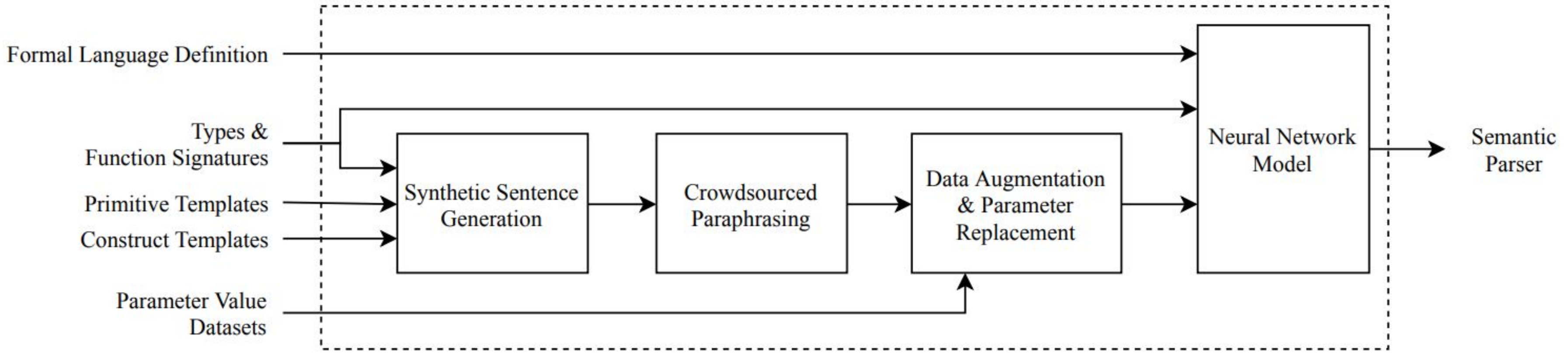
In what respect do you think multilingual semantic parsing differs from multilingual question answering?

Multi Language Semantic Parsing

Templated-based data generation

Genie methodology:

- Developers write templates to synthesize data
- Generate more natural data using crowdsourced paraphrases and data augmentation
- Combine paraphrases with the synthesized data, to train a semantic parser



Finding Data in Other Languages

Structured:

- Any websites using Schema.org metadata can be scraped to find relevant properties in each domain

General:

- Wikipedia and other open websites allow scraping but some knowledge is required to properly extract the values

Prior work

Datasets:

- ATIS: Airline Travel Information System
- GeoQuery: The functional query language used in the Geoquery domain
- Overnight: In seven domains covering various linguistic phenomena
- NLMaps: A Natural Language Interface to Query OpenStreetMap

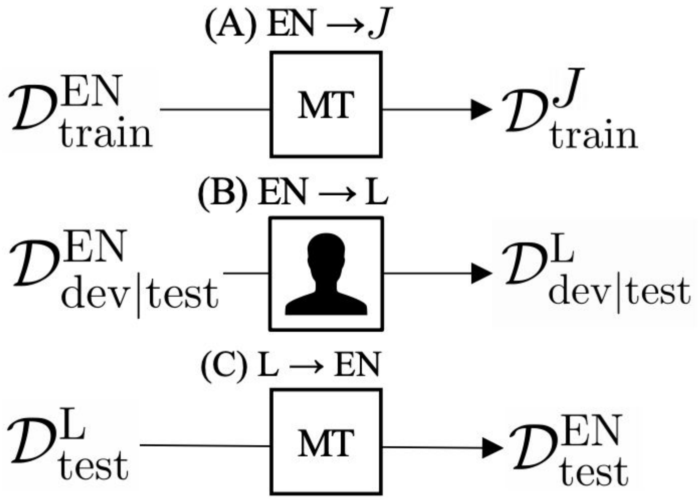
Methods:

- Polyglot decoder for source-code generation from API documentation
- Ensemble monolingual hybrid tree parsers to generate a single parse tree
- Find multilingual representations based on dependencies or embeddings of logical forms
- Bootstrapping from English to another language without parallel data

Bootstrapping a Crosslingual Semantic Parser

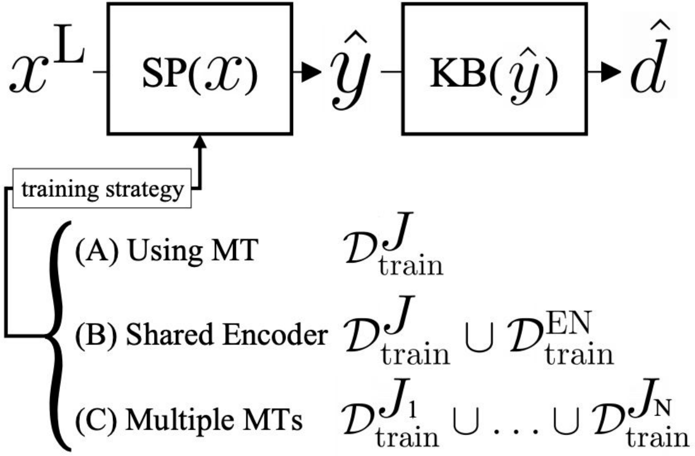
Bootstrapping a Crosslingual Semantic Parser

- Train data is translated using multiple public machine translation APIs
- Dev and test are human translated



Bootstrapping a Crosslingual Semantic Parser

- Train with three different train sets



	DE (MT)	ZH (MT)
Backtranslation to EN	53.9	57.8
+BERT-base	56.4	58.9
SEQ2SEQ	61.0	55.2
+BERT-(de/zh)	64.8	57.3
Shared Encoder	64.1	58.7
+BERT-ML	66.4	59.9
MT-Paraphrase	62.2	64.5
+BERT-ML	**67.8**	65.0
+Shared Encoder	66.6	**68.1**
MT-Ensemble	63.9	57.5
+BERT-ML	64.3	65.5
+Shared Encoder	65.7	67.8

Paraphrasing in Other Languages

- English dataset is synthesized and does not perfectly match with how humans write queries.
- Paraphrasing is used to generate more natural examples to cover a bigger space of all possible utterances
- Translation models can act as paraphrases although we won't have much control over the generated response.
- More sophisticated paraphrasing for other languages has become possible with the recent introduction of mBART (already has 5 citations!) and MarianMT models.

Marian: Fast Neural Machine Translation in C++
Multilingual Denoising Pre-training for Neural Machine Translation

Quiz 2

Why is it better to train a single encoder on multiple languages compared to training one encoder for each language?

Preliminary Error Analysis on Spanish

Error Analysis of Current Results - Spanish

Translating synthesized English sentences to Spanish can result in nonsense

¿cuál es el número de teléfono de la oficina más banh mi nha trang subs

English: What is the office phone number more banh mi nha trang subs

¿el blended bistro & boba en local pond tiene una opinión todavía ?

English: Does the blended bistro & boba at local pond still have an opinion?

lo que hace el restaurante nimi v. reseña de ?

English: what does the restaurant nimi v. review of?

Error Analysis of Current Results - Spanish

Often filters on location instead of cuisine type

Example Question:

buscar un restaurante dim sum .

Correct Response:

now => (@org.schema.Restaurant.Restaurant) filter param:servesCuisine =~ " dim sum " => notify

Gives response:

now => (@org.schema.Restaurant.Restaurant) filter param:geo == location: " dim sum " => notify

Error Analysis of Current Results - Spanish

Has difficulty with cuisines made up of two words (Asian fusion), thinks one of them is a description or restaurant name. This could be a problem with other params that can be 1 - many words long.

Example Question:

¿hay restaurantes fusión asiática cercanos con opiniones 10 estrellas ?

Gives Response:

now => (@org.schema.Restaurant.Restaurant) filter @org.schema.Restaurant.Review { and param:description =~ " fusión " and param:reviewRating.ratingValue == 10 and param:servesCuisine =~ " asiática " => notify

Error Analysis of Current Results - Spanish

Sometimes generates random syntax:

¿cuáles son los últimos comentarios y puntuaciones de este restaurante ?

English: What are some of the most recent reviews of this restaurant?

Gives:

now => [param:aggregateRating.ratingValue , param:reviewRating.ratingValue] of ((@org.schema.Restaurant.Restaurant) filter param:geo == location:current_location) => notify

what does this even mean?

Room for Improvement

- Templates to make sure that common grammar patterns create correct parameters (cuisine vs. location)
- AND hook up model with database to understand if a word is cuisine or something else
- Better ML to create paraphrased sentences in other languages to avoid nonsense

Quiz 3

Why is translation-based data synthesis method a practical alternative to template-based sentence generation?