# **English Teacher**

#### Enhancing Alana to help learn English

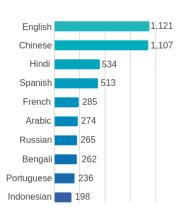
- ????? (You)
- Shubham Agarwal
- Verena Rieser

#### English - The "World language", "Lang of Internet", etc.



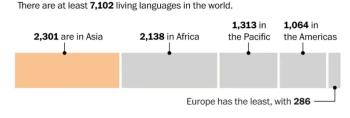
No. of native and non-native English speakers

Source: Ethnologue, 21st Edition

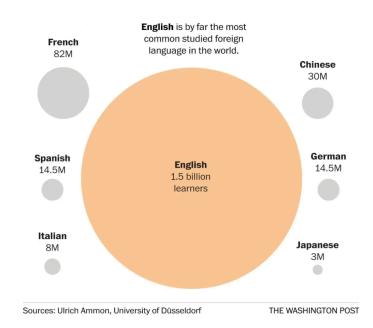


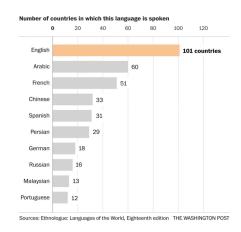
Top 10 languages by number of speakers (native + non-native)

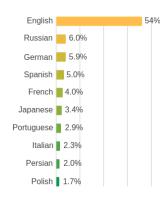
Source: Ethnologue, 21st edition



Sources: Ethnologue: Languages of the World, Eighteenth edition THE WASHINGTON POST







Top 10 languages on the internet

Source: W3Techs

#### Reference:

- 1. https://www.ethnologue.com/guides/ethnologue200
- 2. https://lemongrad.com/english-language-statistics/
- 3. https://en.wikipedia.org/wiki/English-speaking world
- 4. https://www.washingtonpost.com/news/worldviews/wp/2015/04/23/the-worlds-languages-in-7-maps-and-charts/

#### **Fun facts:**

- 1. 20% of the Earth's population speaks English
- 2. India has the second-largest English speaking population after the US

### Motivation

- Alana extension can be used by people who want to practice their English conversation.
- Should be able to practice chit-chat and also talk about topics such as music and movies.
- Eg. Take pointers from "Duolingo"/"Babble" apps and apply in a conversational setting (?)



- Possible target audience (Ideally all three):
  - Beginners
    - Children (<8yrs) with/- their mother tongue as English
    - Adults learning English for business
  - Intermediate
    - Teenagers preparing for exams or talking about superheroes / Disney characters
  - Expert
    - Student preparing for grad applications: TOEFL / IELTS



# Some categorized tasks

- Content / data sources
- Error diagnosis in NLU Modify NLU to understand utterances even if they are incorrect in some way - topics
- NLG
  - For repairing errors
  - In general responses
- DM for practicing English / selecting lessons (e.g. spot and remedy errors)
- Persona (and chit-chat) design and development
- Evaluation of language learning
- -> AI/DL techniques to automatically detect syntactically and semantically correct user exchanges
  - a. Clean data source -> Can use pre-trained networks (BERT)
  - b. Challenging problem:
  - Buffalo buffalo buffalo buffalo buffalo buffalo buffalo

Reference: https://en.wikipedia.org/wiki/List\_of\_linguistic\_example\_sentences

# WHY IS ENGLISH SO MUCH FUN?

ALL THE FAITH HE HAD HAD HAD HAD NO FFECT ON THE OUTCOME OF HIS LIFE.

BECAUSE THAT SENTENCE MAKES PERFECT SENSE.

more Fun on likealaugh.org

## Some References

- Curry et al, "Alana v2: Entertaining and Informative Open-domain Social Dialogue using Ontologies and Entity Linking", Alexa Prize Proceedings (2018)
- Litman et al. "Towards Using Conversations with Spoken Dialogue Systems in the Automated Assessment of Non-Native Speakers of English" In SIGDial (2016)
- Huang et al. "A chatbot for a dialogue-based second language learning system"
  EUROCALL (2017)
- Baur et al. "A Textbook-based Serious Game for Practising Spoken Language"
  2013.
- Bibauw et al. "Dialog Systems for language learning: state of the art and avenues for research on task-based agents" CALICO (2015)

### #CA2020



shubhamagarwal92



@shubhamag1992



sa201@hw.ac.uk

Feel free to drop a message!

Slides: https://github.com/shubhamagarwal92/CA\_2020/blob/master/project\_material/