ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

The Assistant: A Solution for Expressive Language Disorder (ELD)

Shunshi Zhang

The Archer School for Girls, 11725 Sunset Blvd, Los Angeles, CA 90049, United States

Abstract. The Assistant is an app carefully designed to assist people with expressive language disorder, a lifelong condition that impacts a person's ability to express their own ideas when they speak. The purpose of this project is to assist the user themselves in practicing how to fluently express themselves, but also to "translate" their speech for those around them that may not understand what the user is saying. Automatic speech recognition (ASR) will play a crucial role in the development of this app, as it allows the system to convert any speech into printed text. The Assistant also has an educational aspect, teaching children and young adults that it is indeed possible to help those who struggle with the basics of speech and expression through the development of an app and progressive technology. In conclusion, The Assistant has a consistent and achievable goal and can be accessed by the population with expressive language disorder if it were launched.

Keywords: expressive language disorder, Automatic speech recognition, education data mining.

1. Introduction

5-10% of the US population has a communication disorder [1], or as many like to call them, a developmental language disorder (DLD; a communication disorder that persists into adulthood). Many people also have expressive language disorder (ELD), a disorder with similar traits that fall under the broader category of DLD, which is what this project is going to be based on. ELD is treatable, but not curable, and often goes untreated because many families don't see the point in investing money which cause unable to be fixed, or it's because they are lacking financially. As a result, many children diagnosed with ELD grow up with worsened effects of this disorder, such as extensive stuttering. According to The Understood Team, "People with expressive language disorder (ELD) often struggle to form sentences that make sense." [2] With automatic speech recognition (ASR), the app will take the user's input in the form of their own words, and provide "filler words" in between gaps in their speech, forming complete and comprehensible sentences, which is displayed in text.

Table1. Realistic input-output transactions

Verbal Input	ASR Speech-to-Text Translation	Adding "Filler" Words	Text Output*
"Regular oranges have…fat nor sodium, have vitamins A…C and much calcium."	Words detected: Regular oranges have…fat nor sodium, have vitamins A…C and much calcium.	Words added: 'Neither', 'but they do', 'and', 'they have'.	Regular oranges have neither fat nor sodium, but they do have vitamins A and C, and they have much calcium.

*Direct translation, not taking punctuation or fluency of sentence into consideration.

According to Speech Buddy, "Various studies put the number of children who struggle with literacy as part of their expressive language disorder at up to 75%." [3] Therefore, the app will have a read-aloud feature for individuals who struggle with identifying words and literacy in general. This paper will cover the 5 aspects of: Automatic speech recognition (ASR), models, educational

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023
data mining (EDM), evaluation, and embedded experiments. To define these terms, ASR is a type of voice technology that works to the best of its ability to turn speech into text. Major companies today like Microsoft, Google, and Apple use it to develop their products. Models in app will need both traditional machine learning model and time series model to demonstrate how The Assistant is relevant to today's world and the respective function parts of The Assistant. As will be explained in detail later on, the EDM portion of the paper will demonstrate the 2 types of data this app will collect, the data collected will help with user personalization and will take feedback from users to update and progress the app's functionality. Afterwards, there will be 2 criteria evaluation descriptions of the app will be based on, they are accuracy and instructional design. For the embedded experiments aspect of the paper, 2 specific experiments will be conducted to test the limitations of the built-in ASR, and the general functionality of the app. Based on the results, it will then be up to interpretation just how assistive The Assistant actually is, and how its functionality can be improved. Lastly, some questions to keep in mind while organizing the contents of this research paper are: How can The Assistant be made more accessible, easier to use? What will make people with expressive language disorder want to use this app? How will it improve the quality of their conversations with other people?

2. Models

Models are most often used to help illustrate ideas and to demonstrate purposes, which is essential when it comes to planning the uses of a yet-established assistive app such as The Assistant. As briefly mentioned in the introduction, the 2 types of models this app needs are a traditional machine learning model and a time series model. A traditional machine learning model can be used to perform two tasks, regression and classification. Regression can be used, in this case, to predict the number of "filler" words needed to form a complete output sentence based on the user's speech input. Regression often has a positive or negative numerical value as its output, but in this case it can only be a positive output since the amount of "filler" words cannot be negative.

Table 2. A concrete example of regression

User Input (x)	ASR Speech-to-Text Translation	Adding "Filler" Words	Text Output (y)	Number of "Filler" Words Needed
"Regular oranges have…fat nor sodium, have vitamins A…C and much calcium."	Words detected: Regular oranges have…fat nor sodium, have vitamins A…C and much calcium.	Words added: 'Neither', 'but they do', 'and', 'they have'.	Regular oranges have neither fat nor sodium, but they do have vitamins A and C, and they have much calcium	7

*Using the previous example in the Introduction.

Next, an example of a classification. A classification is exactly what it sounds like, it classifies any data provided and arranges them into a specific category. Based on the user's name, age, and sex (male or female), which are the data, the system is able to "prepare" itself for any slang or words that are outdated but still used in today's world. That way, words within a specific range, like abbreviations teens often use today ('ttyl', 'gtg', 'btw', 'idk', etc) will be expected. This is helpful information because part of The Assistant's job is to figure out the meaning of the user's speech, in order to add in "filler" words to form complete sentences as a text output, but if the system doesn't even know where to begin to look for filler words because the user's input is in "slang" or too "outdated", it won't be able to formulate an output.

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

Table3. A concrete example* of classification

User Age	User Sex (Male/Female)	User Name	User Input (x)	Type of Speech	"Filler" Words & Removed Words	Text Output (y)**
16	Male	Mark Roe	"Bro…idk man…not like, super…important. Um…it, uh…prom isn't until, like…January."	21st Century Slang Words translated: 'Idk': I don't know	Words added: 'It's' Words removed: 'Um', 'uh', 'it' Words replaced: 'I don't know' instead of 'idk'	Bro I don't know man it's not like, super important. Prom isn't until, like January.

*Hypothetical situation.

**Direct translation, not taking punctuation or fluency of sentence into consideration.

Lastly, a time series model is a timeline in a graph-like structure that stores data. These models are used to predict, or forecast events based on historical data. To predict whether if The Assistant will be widely used if it's released in 2023, a time series models is need to graph the population that has expressive language disorder (ELD), or some other type of developmental language disorder (DLD) to see if the line is increasing drastically or not, to be able to forecast whether more than 100k users (the goal) will actually use this app. If the population of diagnosed DLD patients has been around the same number in the past decade (2012-2022), then most likely it will also be around the same number in 2023, which means there really is no urgent need for The Assistant to be released in the coming year.

Disclaimer: DLD occurs most commonly in children in kindergarten, however, this condition persists into adulthood.

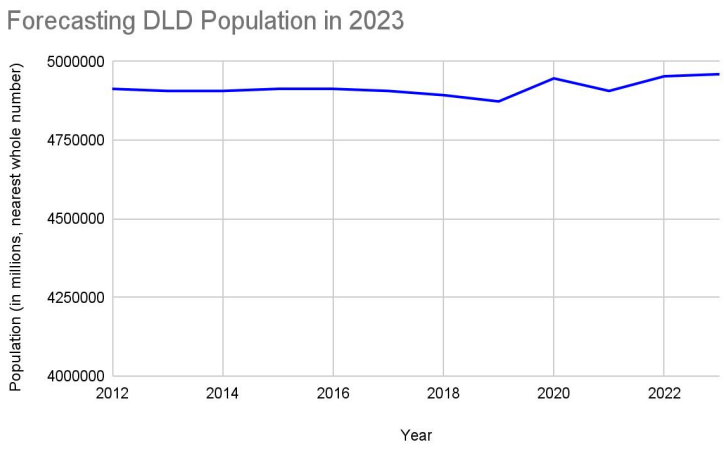

Figure 1. Forecasting DLD Population in 2023

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

Unfortunately, research did not help me find any data on the number of children and adults who have been diagnosed with either DLD or ELD in the past decade. However, I do know that 1 in 15 children has DLD, and ELD falls under this category, so I made a prediction of the population with DLD by multiplying the respective number of children (ages 0-17) in the United States by 1/15 in each of these years, to come up with the data on the graph. I do realize that this data is somewhat inaccurate but because the graph cannot be blank an estimation has been made. My prediction is that in 2023, there will be an all-time high number of children diagnosed with DLD (in the past decade) since the ChildStats website also provided the number of children in the United States in 2023. The number will be around 4,960,000 children (74,400,000*1/15) with developmental language disorder in 2023. This is crucial information because DLD is a common condition that often follows a child into adulthood, it is treatable but not curable. To conclude, the data shows a stable increase in population into the year 2023, so it would be safe to assume that the coming year is a good time to release The Assistant onto the market.

3. Educational Data Mining

Data in the form of numbers and percentages (numerical data) is two types of data The Assistant is able to collect. This type of data can be derived from a model. Taking the regression model, a traditional machine learning model, as an example (table below), it is evident that the number 7 determines how many "filler" words are needed to form a complete sentence output. The "filler" words are extremely important as they fill in the gaps in the user's expression and help the system produce a concrete, understandable text output that the user will be looking for in the Textbox, which makes the data useful.

Table 4 Take the traditional machine learning model, regression model, as an example

User Input (x)	ASR Speech-to-Text Translation	Adding "Filler" Words	Text Output (y)	Number of "Filler" Words Needed
"Regular oranges have…fat nor sodium, have vitamins A…C and much calcium."	Words detected: Regular oranges have…fat nor sodium, have vitamins A…C and much calcium.	Words added: 'Neither', 'but they do', 'and', 'they have'.	Regular oranges have neither fat nor sodium, but they do have vitamins A and C, and they have much calcium	7

Data can also come from user feedback.In this case, Data would not be a number, it would instead be in the form of words that can be taken into consideration and used to revise the features of The Assistant. For instance, if someone testing the app were to comment, "how can you make your app easier to navigate?", it would then be the right decision to be in the shoes of the user and see exactly what they mean. This is useful because in the process of building an app, it's always important to ask for user feedback so that the app can be improved in terms of accessibility and updated features. In this way, it is mutually beneficial since the user is the one actually needing to use The Assistant and the creator is the one following the users' suggestions so more people would want to download the app.

A possibility of what users might see (follow the steps):

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

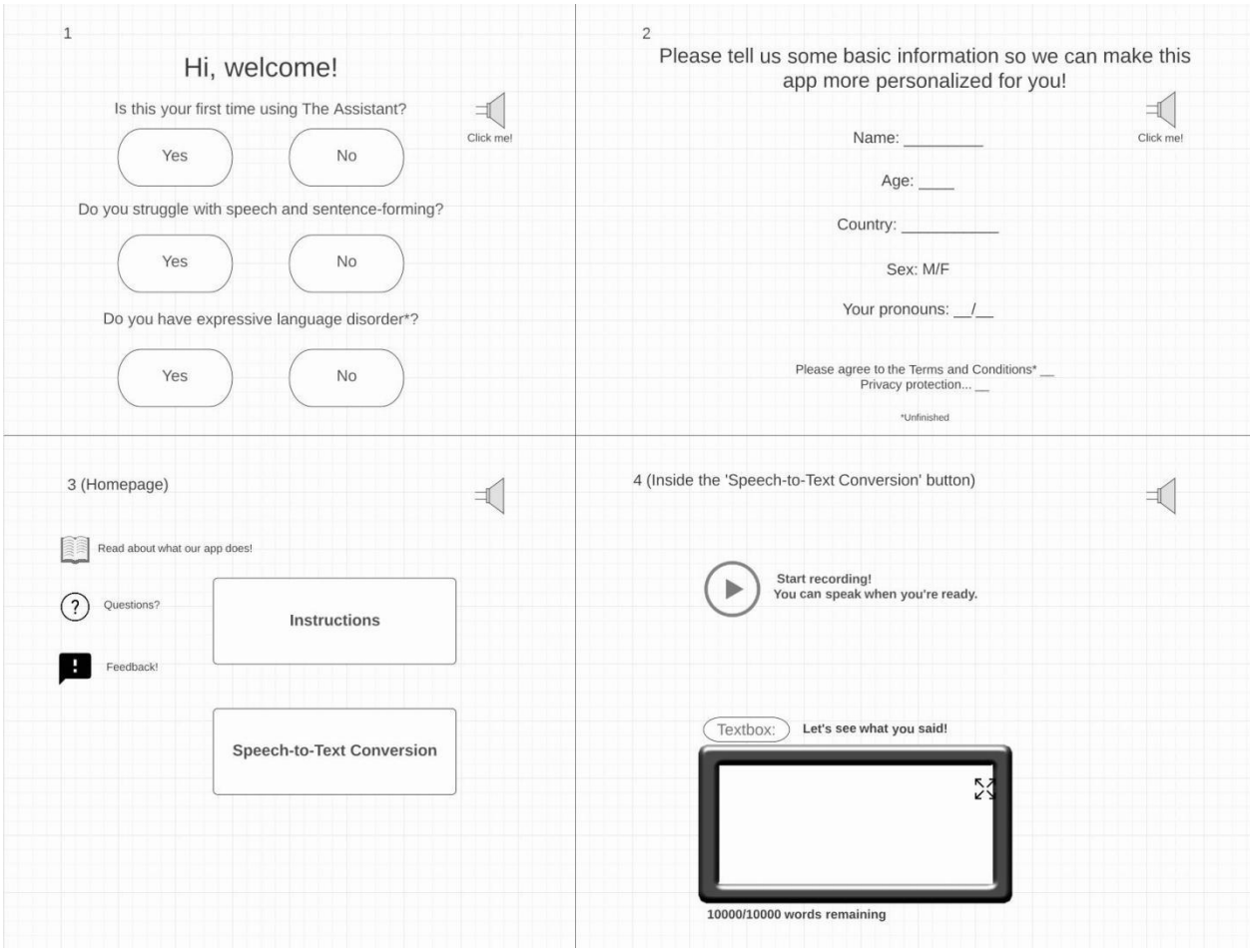

Table 2. A possibility of what users might see (follow the steps):

The various steps completed in order to use the app are shown and labeled. Referencing back to the user's feedback, it is text-heavy and colorless, making it somewhat unobvious which features are more important than the others. It would make sense to color coordinate the buttons and make the primary features the most visible and vibrant. The more important features are the speaker ('Click Me!') buttons, the smaller symbols for the users' use on the top left corner of the Homepage, the 'Speech-to-Text Conversion' button, the 'Start Recording', and lastly, the textbox (where the words are being printed). With color and highlights, the user will be able to navigate the app in less time and will immediately know which buttons are for instruction (purple & green), which are there to assist (yellow), and where they can find their printed text (red). After revision, this is what users will see:

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

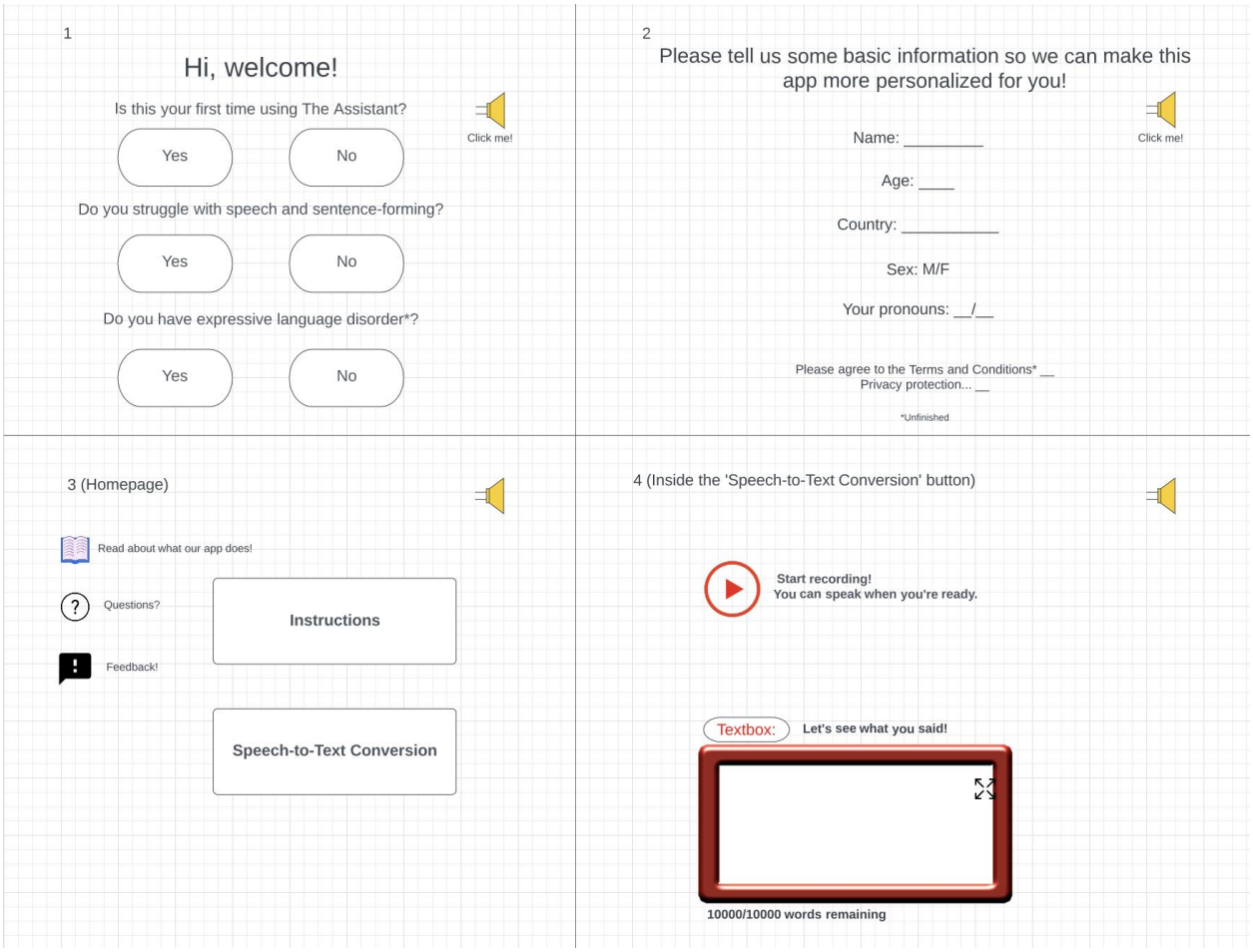

Table 3.The various steps completed using the application are displayed and labeled

4. Evaluation

Accuracy and instructional design are the 2 main criteria The Assistant will be assessed on. Accuracy is directly related to the input-output transaction, it's scored on how close in meaning the printed text is to what the user is trying to express (how satisfied they are with the sentences). Instructional design, in this context, is referring to the function The Assistant is programmed to do, how well it is designed, and which crowds of people it affects. A table has been created to demonstrate how "accuracy" can be calculated through a single example of the speech-to-text translation, with the addition of "filler" words:

Table 4. Show how to calculate "accuracy" with an example of speech-to-text translation

Original Verbal Input	ASR Speech-to-Text Translation	Adding "Filler" Words	Adjusted Text Output**	Percentage of Accuracy*
"My neighbor…cat, 14 almost, its color…orange."	Words detected: My neighbor…cat, 14 almost, its color…orange.	Words Added: 'Has a', 'it's', 'fur', 'is'	My neighbor has a cat, it's 14 almost, its fur color is orange.	Hypothetical user rates: 9/10 =90% satisfaction; mostly accurate

*A score out of 10, provided by users, based on their satisfaction with the printed sentence(s).
**Direct translation, not taking punctuation or fluency of sentence into consideration.

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

Referencing back to the table, the hypothetical user rated their level of satisfaction a 9/10, which is actually how "accurate" or similar the output is in meaning to what they had wanted to express. As a percentage, that would be 90%. The instructional design, however, judges the app on its main function, how creative the idea behind the design is, and the population it affects. The main function of The Assistant is a button on the homepage (reference my drawings), Speech to Text Conversions, that converts the user's spoken words into printed text in the Textbox. The idea behind the design was inspired by a friend of mine, who has a condition similar to expressive language disorder, where she couldn't speak properly and in complete sentences, only fragments of them. It became difficult as she got older to understand the meaning of her sentences, so I decided I would explore this idea further – an app that could translate any fragments of a sentence, no matter how incomprehensible, into something that resembles that of a normal sentence. Most likely, anyone else would evaluate the app in a similar way based on the 2 given criteria, since this is a non-profit assistive app meant to be downloaded on the users' mobile devices and used frequently. A social experiment will be conducted in the 'Embedded Experiments' section of the paper below relating to how randomly selected people of the population think of the basic functions of The Assistant. Lastly, the population this app affects are the people with expressive language disorder, or similar conditions that could potentially affect a person's ability to express themselves through speech. Family members of the users could also educate themselves on how to help more people with such language disorders. Thus, this is not only an assistive app, but also an educational one.

5. Embedded Experiments

The first of the 2 experiments will be to test the limitations of The Assistant's automatic speech recognition software as it is the sole function that carries the purpose of the app. Experiment 2 will test to see if The Assistant can work its way out of system errors (such as 'error: no translation') if definitions (in Python, a definition is something that is listed before any code is written, it defines a variable) were provided beforehand. The last experiment will be a bonus experiment (not embedded; a social experiment) and will also be asking the public's opinion on how useful they think the app will be if it ever gets released. The experiment regarding people's opinion will be a social experiment, all names anonymous and will reflect the general public's thoughts on The Assistant. A population of different, randomly selected people will be asked to answer a series of short questions after they have a basic understanding of the app and how it functions. The population's ages are between 11-48.

Experiment 1: Testing The Assistant's ASR Limitations (Short sample from: Living with Developmental Language Disorder, YouTube) [5]

Disclaimer: Expressive language disorder falls into a larger category known as developmental language disorder (DLD).

Table.5 Voice input text

Original Verbal Input	ASR Speech-to-Text Translation	Adding "Filler" Words	Adjusted Text Output*
"He also …say I speaks funny but uh…maybe that's just 'cuz hindsight's in 2020…I have, I'm just trying to form my words, umm…Like, I think it was definitely there, just no one	Words detected: He also …say I speaks funny but uh…maybe that's just 'cuz hindsight's in 2020…I have, I'm just trying to form my words, umm…Like, I think it was definitely	Switched/Removed words**: -Remove 'uh' -'Says' instead of 'say' -'speak' instead of 'speaks' -'2020's in hindsight' instead of	He also says I speak funny but maybe that's just 'cuz 2020's in hindsight…I have, I'm just trying to form my words. Like, I think it was definitely there, just no one around me

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

in…around me really paid attention." [1]	there, just no one in…around me really paid attention.	'hindsight's in 2020' -Remove 'umm' -Remove 'in'	really paid attention.

*Direct translation, not taking punctuation or fluency of sentence into consideration.

**New feature since it is necessary to switch and remove words in order to form sentences that actually would actually make sense when it's not just the sample sentence.

From this experiment, it can be concluded that people with ELD struggle not only with remembering words, but also the past, present, or future tense they are supposed to be in. Challenging the ASR system with a much more complex sample from a YouTube video proves The Assistant's capabilities when it comes to dissecting and reforming complicated errors in sentences the users may express. The results, judging by a much more coordinated output, can tell us that The Assistant functions properly and can handle complicated user inputs. Next, the second experiment will be listed.

Experiment 2: Can the System Work Its Way Out of an Error?

Table 6. Definition provided

Variables*:	Provided Definitions:
malfunction	malfunction = "System malfunction. Please contact customer services. They may be able to help you." print(malfunction)
undefined	undefined = "There are undefined terms. Please make sure there is no background noise. It is advised that you restart your recording." print(undefined)
exceeds_text_limit	exceeds_text_limit = "You have exceeded the maximum text limit of 10,000 words. The recording will stop now." print(exceeds_text_limit)

*All variables in Python are lowercase.

Table 7. Statement Filling

Original Verbal Input	ASR Speech-to-Text Translation	Adding "Filler" Words	Adjusted Text Output*	ERROR
"Our family is of Chinese 北京?@#?%...descent…uh my	Our family is of Chinese 北京?@#?%...descent…uh my	Switched/Removed words**: -Remove 'uh' -Replace 'says'	Our family is of Chinese 北京?@#?% descent, my	Error: There are undefined terms. Please make sure there is no

ISSN:2790-167X DOI: 10.56028/aehssr.4.1.367.2023

family always…they says to umm…follow family tradition of photo-taking…anytime when there's…important celebration."	family always…they says to umm…follow family tradition of photo-taking…anytime when there's…important celebration. 北京?@#?%: Undefined terms. Error.	with 'said' -Remove 'umm' -Remove 'they' -Remove 'when' Added words: -Add 'our' -Add 'an'	family always said to follow our family tradition of photo-taking anytime there's an important celebration.	background noise. It is advised that you restart your recording.

*Direct translation, not taking punctuation or fluency of sentence into consideration.

**New feature since it is necessary to switch and remove words in order to form sentences that actually would actually make sense when it's not just the sample sentence.

From this second experiment, it can be concluded that the app indeed is able to work its way out of system errors by first identifying the type of error that it is. Then, the system 'refers' to the provided definitions in the table and prints on the user's screen 'there are undefined terms. Please make sure there is no background noise. It is advised that you restart your recording' since the error identifies as 'undefined' under the list of variables.

6. Conclusion

The Assistant is an app designed to assist the population with expressive language disorder and most conditions under the larger umbrella, developmental language disorder. The 5 aspects covered in this paper summarize the basic functions of the app and the planning stages of it. Many realistic examples have been given to demonstrate the input-output conversion of the speech-to-text translation, therefore illustrating how automatic speech recognition is formally utilized in this app. Thank you.

References

[1] Quick Facts (Speech and Language Disorders). American Speech-Language-Hearing Association. https://www.asha.org/about/press-room/quick-facts/

[2] The Understood Team. What is expressive language disorder? Understood.

[3] https://www.understood.org/en/articles/what-is-expressive-language-disorder

[4] O., C. Build Reading Skills in Kids with Expressive Language Disorder. Speech Buddies.

[5] https://www.speechbuddy.com/blog/speech-therapy-techniques/build-reading-skills-in-kids-with-expressive-language-disorder/

[6] U.S. Census Bureau. (2021, July 1). POP1 Child population: Number of children (in millions) ages 0–17 in the United States by age, 1950–2021 and projected 2022–2050. Childstats. https://www.childstats.gov/americaschildren/tables/pop1.asp

[7] BoysTownHospital. (2019, January 30). Living with Developmental Language Disorder [Video]. YouTube. https://www.youtube.com/watch?v=KrOISXtCgVA