Ng Ren Yong & Ernest Liu

A0216658W, A0216284E

Statement of Contribution

Ren Yong: Background Research, Data Analysis

Ernest: Data Analysis, Web Scraping, Data Visualization

Abstract

In pursuit of economic growth, Singapore has experienced an influx of non-English trained foreign workers. To improve English communication among this demographic, many private language agencies offer a standard English curriculum. In this paper, we critique the generalized curriculum utilized by these entities. Using the network of word definitions, we further propose that a vocation-specific curriculum would be more effective at improving workplace communication in the shortest amount of time.

Designing an efficient English curriculum

Network of Word Definitions

Introduction (237):

English is one of Singapore’s four official languages, along with Malay, Chinese, and Tamil. As the lingua franca of Singapore, Singaporeans have benefitted economically from opportunities abroad arising from sharing a common language with the rest of the world. To further remain economically competitive despite a declining population, Singapore is home to over 1.1 million low to medium-skill foreign workers. Construction workers and domestic helpers comprise most of these immigrants, accounting for over three-quarters of the construction industry and 30% of the entire service industry in 2020. These workers hail from culturally distinct countries such as China, Indonesia, India, Bangladesh, Myanmar, and the Philippines, as part of bilateral agreements between Singapore and these countries. As Singapore continues to experience an influx of foreign workers speaking a variety of mother tongues, the importance of English as the unifying language of communication becomes increasingly crucial. Without a satisfactory command of English, these foreign workers are relegated to working only with those who speak the same tongue. This leads to workforce rigidity where employment opportunities are limited by the language of the environment.

This paper investigates the current methods used to improve and assess the standard of English among foreign workers in Singapore. Using various parameters to analyse the network of word definitions, we hypothesize that a vocation-specific English course would be superior to existing generalized courses for the purpose of improving workplace communication within a short period of time.

Current Measures (304):

For work permit holders, which includes low-skilled workers such as construction workers and domestic helpers, English proficiency is not a strict requirement. However, starting from mid-2010, some work permit holders must pass the Service Literacy Test (SLT) to qualify as a “higher-skilled worker”. This includes people working in the retail, hotel, and food and beverage sectors. The test ensures basic English proficiency among service workers which enables them to perform their jobs more efficiently, improving service standards. Work permit holders who can speak basic English can also communicate with local workers and integrate better into their workplaces as well as the society.

The Service literacy test is a decent first step in boosting the English literacy of the foreign worker industry. Evidently, the government does see an incentive in providing English education to this demographic. However, considering that no progress has been made in the past 12 years, it is likely that the initiative was not successful and further plans have been abandoned. A possible reason is the lack of accreditation of English language training providers, leading to a lack of quality control.

However, we propose that a more plausible reason is that a generalized curriculum is not the ideal method of instruction. A generalized curriculum aims to create a stable foundation for more specialized words to be taught in the future. In such a curriculum, frequency lists are often the basis of curriculum design, with words that appear more frequently being taught first. However, this conflicts with the goal of improving communication in a workplace environment in the shortest amount of time. This is because the frequency of vocabulary used in a specific domain will differ vastly from normal conversational English. In other words, a generalized curriculum will exclude certain specialized words, while including some general words that are unnecessary for workplace communication.

Vocation-Specific curriculum (204):

Designing a curriculum is a complicated process that involves the careful selection of relevant vocabulary and grammar. Zoe Enser, a specialist advisor for English, opines that it is crucial to consider the purpose of the curriculum and concentrate on the content that truly matters. Today, vocation-tailored English education is still a novel concept but has been gaining immense interest. In Sri Lanka, the vocation specific *English for Specific Purposes* (ESP), when blended with the traditional *English for General Purposes* (EGP), was more successful in preparing students for the workforce than EGP alone.

Hence, we propose a vocation-specific curriculum that only teaches words that are relevant to the occupation, starting with basic words that define them. These words are selected by analysing the network of word definitions to find the most important words for each vocation.

For English, it is estimated that about 1000 vocabulary words are necessary for basic conversational skills, but more than 3000 words are required to understand specialized terminologies. While a short English course will not be able to impart all these words to the learner, it is important to build up a foundation of simple vocation-specific words to enable them to pick up words more easily when on the job.

Methodology (313):

In order to construct the network to determine which words should be deemed as “important”, we have scraped the definition of words off the internet. These definitions are taken from vocabulary.com.

We first start off with a few root words which have been decided using a language-learning mobile application called Duolingo. In this application, users are taught words from various categories such as “in the kitchen”, “sports”, “food” and much more. Thus, we decided that this would be a good starting point to obtain words from the categories of each occupation. For example, maids will have to learn English in several contexts such as “cooking”, “housekeeping” and “shopping”. Construction workers will have to learn English in the context of “safety”, “materials”, and “construction techniques”. We sampled 2 words from each category’s vocabulary list, and these served as the root nodes for building our network. The words that were chosen are:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Occupation** | **Construction worker** | | | **Domestic helper** | | |
| **Category** | **Safety** | **Materials** | **Construction Techniques** | **Cooking** | **Housekeeping** | **Shopping** |
| Word 1 | Hazard | Scaffold | Weld | Fork | Cleaning | Grocery |
| Word 2 | Crane | Cement | Construct | Spatula | Detergent | Discount |

After deciding on the root nodes, we scraped off their definitions from vocabulary.com. Learners of the word will have to understand every word in the definition. Hence, we scraped off the definitions of all the words contained in this word’s definition. We repeat this process recursively for 3 minutes, or till the algorithm terminates. However, since words such as “the” and “and” are expected to show up in most of these definitions, we decided to ignore all conjunctions, articles, prepositions, and pronouns which show up in the top 100 most used English words. This gives us a dataset of about 1300 words per root node, which calculates to about 7800 words per occupation.

After this dataset was obtained, we built the network on Gephi. If a word A contains another word B in its definition, a directed edge is drawn from the node representing word A, to the one representing node B. With these few steps, we obtain the following graphs plotted below.

Network (Construction worker):

Chart

Description automatically generated with low confidenceChart

Description automatically generated with medium confidence

Network (Domestic Helper):

Chart

Description automatically generated with low confidenceBackground pattern

Description automatically generated

Analysis (130):

For this analysis, we define “simple words” as those in the 100 most frequently used words, while “complex words” are those outside of this 100.

In the above diagrams, the nodes are coloured based on the word’s “In-degree”. The “In-Degree” in this network refers to the number of times the word appears in the definition of another word. Also relevant, but not visually represented above, is the “Out-Degree” which represents the number of complex words used in the definition of the selected word.

The size of the node corresponds with the Betweenness centrality. This number is a representation of the amount of influence a node has over the flow of information in a graph. A word with higher betweenness centrality would be crucial in bridging words from two unrelated categories.

Domestic Helper (Top 20 words sorted by In-degree and betweenness centrality):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Word | In-degree | Out-Degree | Word | Betweenness Centrality |
| space | 9 | 2 | nature | 86135.03 |
| parts | 9 | 2 | planet | 82831.38 |
| whole | 8 | 3 | authority | 77860.19 |
| others | 8 | 1 | world | 74615.86 |
| purpose | 8 | 1 | force | 62936.17 |
| thing | 7 | 1 | close | 62924.42 |
| nature | 6 | 4 | unit | 59624.19 |
| are | 6 | 3 | inferred | 56253.12 |
| knowledge | 6 | 3 | general | 55066.24 |
| group | 6 | 1 | whole | 54494.77 |
| more | 5 | 5 | conclude | 54207.88 |
| force | 5 | 3 | physics | 52287.59 |
| lacking | 5 | 2 | are | 52056.81 |
| amount | 5 | 2 | order | 50074.4 |
| located | 5 | 2 | character | 47909.74 |
| effort | 5 | 1 | approved | 45990.88 |
| scope | 5 | 1 | special | 45802.98 |
| distinguished | 5 | 1 | adapted | 45716.7 |
| general | 4 | 4 | abstraction | 45534.73 |
| structure | 4 | 3 | mind | 45284.41 |

Construction Worker (Top 20 words sorted by In-degree and betweenness centrality):

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Word | In-degree | Out-Degree | Word | Betweenness Centrality |
| senses | 6 | 2 | basic | 48032.5 |
| mind | 5 | 2 | essential | 48003.5 |
| amount | 5 | 2 | applying | 41653.33 |
| process | 4 | 3 | necessary | 40940.5 |
| position | 4 | 1 | general | 35338.33 |
| objects | 3 | 3 | provide | 34876.5 |
| extent | 3 | 2 | process | 33579.83 |
| change | 3 | 2 | cause | 33403.83 |
| class | 3 | 2 | put | 33374.83 |
| ability | 3 | 2 | unique | 31014.33 |
| more | 3 | 2 | particular | 30985.33 |
| perceived | 3 | 2 | belonging | 29538.33 |
| region | 3 | 2 | reduced | 27598.33 |
| essential | 3 | 1 | made | 27420.33 |
| particular | 3 | 1 | abstraction | 27325.17 |
| relation | 3 | 1 | relation | 27296.17 |
| limited | 3 | 1 | kind | 24776.33 |
| ones | 3 | 1 | local | 23544 |
| nature | 3 | 1 | parts | 23515 |
| attribute | 3 | 1 | showing | 23445.17 |

Analysis (Degree Centrality) (295):

One parameter to assess the importance of a word is its degree centrality, but more specifically its In-degree centrality. The “In-Degree” is a measure of its utility in defining the meaning of other complex words (words not in the top 100 list). This parameter helps us to structure the curriculum by ordering the sequence which we teach words to foreign workers. Higher “In-Degree” suggests the word should be taught first because many complex words require understanding of this word in question. On the other hand, words with higher “Out-Degree” should be taught later because it suggests the word in question requires understanding of many other complex words.

Ideally, we should start with words with high “In-Degree” and low “Out-Degree”. For domestic helpers, this would include “purpose”, “others”, and “space”. These words are not only useful in defining many other complex words but can also be easily defined by very few complex words.

*Purpose: what* ***something*** *is used for*

*other: not the* ***same*** *one*

*space: an* ***empty******area***

On the other hand, construction workers should start with words like “senses”, “mind”, and “amount”. Learning these words first will make the greatest number of occupationally relevant complex words accessible to the learner.

If curriculum designers only consider the “In-Degree” of a word, the words “nature” and “knowledge” may be classified as important words to be taught first. At first glance, these words should be taught early to domestic helpers, as evidenced by its high In-Degree. However, due to its high Out-Degree, these words are difficult to explain without prerequisite understanding of the complex words used to define them. This might lead to more confusion as educators must ensure that the complex terms used to define them must also be taught.

*Nature:* ***Phenomena*** *of the* ***natural physical world***

*Knowledge: the* ***psychological result*** *of* ***learning***

Therefore, the choice to include certain vocabulary for short-term vocational English language curriculum design is ultimately a balance between In-Degree and Out-Degree.

Analysis (Betweenness Centrality) (91):

Another key parameter in curriculum design is the betweenness centrality of words. Betweenness centrality helps us to identify which words can be used to bridge two different categories. For construction workers, the words with the highest betweenness centrality are “basic”, “essential” and “applying”. One commonality shared by these words with high betweenness centrality is that they are used to define words that are related to the original three chosen categories (Safety, Materials, Construction Techniques). These words serve as connecting words and can be used to introduce words from a new category.

Limitations (314):

Our model has several limitations, with one glaring problem being the oversimplification of effective curriculum design. In our model, we identify the best set of vocabulary to be included for a specific vocation without consideration for how grammar and sentence structure could be taught. In typical curriculum design, vocabulary and grammar work in tandem to provide learners a clearer understanding of the English language. Mindlessly including vocabulary suggested by our network without proper grammar pedagogy would instead hinder the learning process.

For this project, our algorithm was also limited by time and computing power. As a result, we were only able to use 6 root words for each of the two occupations. The sample size is too small and thus the result would likely be tough to replicate. For future studies, we propose that at least 8 root words over at least 4 categories should be used to obtain more accurate results.

Lastly, the “In” and “Out” degree centrality of a word may not accurately reflect the true “difficulty” of a word. Under our model, words not within the top 100 frequently used words are considered “complex”. Using this model, a word that requires many “complex” words to define (high Out-Degree) is a difficult word. However, due to the broadness of our definition of a “complex” word, the true difficulty of a word is often under or overstated.

*Nature:* ***Phenomena*** *of the* ***natural physical world***

*General:* ***Affecting*** *or* ***Concerning*** *all or* ***most******people*** *or things*

Although the definitions of the words “Nature” and “General” both contain four “complex” words, it is likely that a layman would agree that based on the above definition, “General” is an easier word to learn. This is because although all 8 bolded words are considered “complex”, those in the definition of “general” are more frequently used. This problem can be mitigated by tightening the definition of a “complex” word, such as only including words not in the top 300 frequently used words instead of top 100.

Conclusion (111):

In conclusion, we believe that there is an incentive in providing foreign workers in Singapore a English education to improve workplace and societal communication. However, the generalized English curriculum used in current education centres for these purposes can be improved. We propose that curriculum designers focus mainly on vocabulary that are used on the job, as well as the words that define them. Using our methodology, the words to include in the curriculum can be assessed by three criteria: 1) High “In-Degree” centrality 2) Low “Out-Degree” centrality 3) high betweenness centrality. While our methodology makes several critical assumptions, further studies implementing our proposed solutions should improve the accuracy of the result.

1999 Words excluding tables, headers, and examples

**Bibliography:**

Sources on foreign worker numbers:

https://data.gov.sg/dataset/foreign-workforce-numbers-annual?resource\_id=7e472f55-d032-4d26-ba82-73f8b5e527d9

https://data.gov.sg/dataset/foreign-workforce-numbers-annual?resource\_id=e5274807-bcc7-4d37-b39f-ab06fec24647

Sources on English curriculum:

https://www.fluentu.com/blog/how-many-words-do-i-need-to-know/

https://www.teachwire.net/news/4-steps-to-a-great-english-curriculum

https://www.cpsctech.org/2020/09/english-need-closer-examination-of.html

Sources on current measures:

https://www.mom.gov.sg/newsroom/press-releases/2010/service-literacy-test-to-up-service-standards-in-retail-hotel-and-fb-industries

https://www.mom.gov.sg/newsroom/press-releases/2010/service-literacy-test-to-start-on-3-may-2010