**Inverse Relation Between Inflectional Productivity and Word Frequency**

*Seara Chen*

*McGill University*

[*siyu.chen2@mail.mcgill.ca*](mailto:siyu.chen2@mail.mcgill.ca)

*ID: 260670714*

**GUIDELINE**

* 10 pages (to 20) Graded by Tim yay!!!
* presentation (15-20 minutes)

**Component**

* Background
  + Motivation (state the motivation in between like in presentation)
* Related Works
  + Differences in productivity
* Hypothesis
  + Goal accomplished for establishing the pipeline
* ACTUAL CONTENT
  + Data
    - Why is this particular data selected? -- why particularly portugese
    - How data is cleaned
  + Stage one segmentation (with a chart, and graph)
  + Stage two clustering

Explain the pipeline we did here --- with diagram

* + Analysis
* Discussion and Future work
  + Research progress
* Conclusion
* Bibliography
* Appendix
  + Code to github

**Overview (DONE)**

In this project aims for an attempt to quantitatively demonstrate the inverse relationship between word token frequency and inflectional productivity – a common understanding in linguistic, although never formally proven nor demonstrated. Through demonstrating this relationship on Portuguese, the project also constructs a pipeline of that could be used to analyse other languages, so that the relationship can be generalized to other Latin-alphabet based languages in future work.

**Background**

In natural languages, verbs are marked with inflections, which often carries information about the person, plurality, tense and aspect. Person, meaning the pronouns associate with the verb; the plurality indicating the number of people/actors being involved; tense, referring to when the action took place; and finally aspect, describes how a verb extends over time. This well of information can all be found through how a verb is marked with inflections or conjugated, thus making inflection an important characteristic to a verb.

However, even within the same language, verbs are conjugated differently. Based on the different ways in which verbs are marked with inflections, we can divide verbs into different *inflectional classes*. For example, in Spanish, the regular verbs, meaning verbs are that ends with “ar”, “er” or “ir, are conjugated differently in present tense (details see **Table 1**).

|  |  |  |  |
| --- | --- | --- | --- |
|  | -ar verbs | -er verbs | -ir verbs |
| yo | -o | -o | -o |
| tú | -as | -es | -es |
| usted él ella | -a | -e | -e |
| nosotros/as | -amos | -emos | -imos |
| vosotros/as | -áis | -éis | -ís |
| ustedes ellos ellas | -an | -en | -en |

**Table 1** Spanish regular verbs in present tense

Now on the other hand

* Define productivity
  + Define productivity at inflectional level
  + Notion of frequency/ regular