Student’s Name: Yuxiong Xie

Instructor’s Name: Esther Namubiru

Class and Section: EAP 506 SECTION P03

Due Date: 11/24/2015

**Language Analysis**

Academic writing acts as the reflection of academic researches, and it has lots of clear rules on language and grammar choices. In particular, the uses of language present the best way for authors to describe ideas and for audiences to understand the main topic. Certainly, in the statistics field, there are also some linguistic features that can be distinguish from other areas. Thus, in order to help international graduate students to learn and grasp the formality conventions in statistical academic wring, this paper analyzed the article “Does data splitting improve prediction?” on language style, such as verbs choices, tense and noun phrases.

In the article, Faraway (2014) used a direct question as the title – “Does data splitting improve prediction?” – which can be the most effective way to explain his idea and draw audience’s attention to the article (Swales & Feak, 2012). In addition, there is also another question in this article. That is, “Will the gain from avoiding data reuse compensate for the loss in model selection and parameter estimation performance?” (Faraway, 2014, p.2). This question serves as a research issue which the author wants to answer, and it will inspire readers to consider and resonate with the author when they read this sentence. According to Swales and Feak (2012), direct questions can be used for putting forward argument and discuss problems; however, indirect questions seem to be more common. Faraway (2014) had a good distribution in this aspect. He only used direct questions to discuss the main topic; while applying indirect issue to introduce subtopics. For example, he described, “we could use scoring to judge how well the observed future responses match with the predictive distributions” (p.2). With the distribution, readers can easily distinguish and understand the main idea and sub-ideas that the author wants to show. What’s more, they can also have a clear framework when skimming this paper.

Moreover, academic writings have some other linguistic features. In the aspect of usage of noun phrase, academic writers add modifiers into noun phrases to make them complex; then sentences will explain ideas more accurately and efficiently (Swales and Feak, 2012). However, Faraway (2014) mostly introduced topics in the article by using simple noun phrases, such as “failures” and “predictions” (p.1); sometimes he also used complex noun phrases, like that “Using the same data to both select and estimate the parameters” and “Predictions based on statistical models” (p.1-2). The reason why the author wrote in this way is that statistical papers are too technical to having a good understanding with using complex sentences. Although complex statements make the papers more formal, the readers, relatively, will have more troubles in the articles’ comprehension. So that authors usually use simple noun phrase and sometimes use complex noun phrase with only one or two modifiers. For example, the noun phrase “Prediction based on statistical models” has only a modifier, which is added to make the meaning more clearly and detail – the statistics prediction is doing under models.

When authors are writing academic papers, they also prefer using single verbs to phrase verbs. Faraway (2014) almost used single verbs in his article. These following sentences are examples: “The second term represents the model selection…” and “The Box-Cox method selects the index …” (p.4,6). The single verbs make the sentences and article more formal and concise; and draw reader’s attention to understand the meaning easily (Walter, 2015). In addition, the author used present tense in every parts of this article. Faraway (2014) stated that, “Our simulations have also shown SD improving …” (p.11). In statistics, present tense can be used to “describe generally accepted facts” (“Common Error”, n.d.). Therefore, the author described methods and theories with present tense and did surveys based on that, he will express the points are realistic and the investigations are credible. In this article, he also often used lots of modal verbs, such as “can”, “should” and “may”, which explain the author’s emotion and aspiration. In particular, Faraway (2014) described his recommendation by using the verb “should”, which strongly suggests audiences to follow these advices and considerations.

As statistics papers, technical language is also one of the most important part. Because Statistical Science is little difficult for other field’s readers; and specialized vocabularies are strange for them. In this article, Faraway (2014) presented many technical vocabularies in the field, such as “Cross-validation”, “Model assessment”, “Model validation” and “Scoring”. Besides, he also took some acronyms which connect to statistics field. For instance, the author used “RMSE” as the abbreviation of “Root mean squared error”. Although Faraway (2014) did not define these technical words, he suggested the students in the same field need to spend more time to search and consider about the professional language. Thus, using specialized forms aims to make the specific readers to feel familiar with these languages clearly in this field.

What else, academic writing focuses on personal pronouns and contractions as well. Single authors prefer using the first-person pronouns to third-person pronouns, like “We” and “I” (Swales & Feak, 2012). Thus, through the whole article, Faraway (2014) used the first-person pronouns in every sections. For example, he introduced that, “We discuss the uses of data splitting and the methods…” in Introduction Section and “We make the following tentative recommendations…” in Conclusion Part (p.2,11). Using first-person pronouns, academic articles will be informal, but it is an easy way for author to write articles and for audiences to understand the topic, which just like “a face-to-face communication with readers”. Furthermore, Faraway (2014) did not use contractions and second-person pronouns, which seem to be informal and nonstandard; but he would choose the formal negatives to express his passive voice, just like that, “Simulations cannot simulate human judgment …” (p.11).

In conclusion, the analysis of language styles in this article states that most of language choices are formal when students write academic writing in statistical field, which includes complex noun phrases, single verbs, no contractions and technical language. However, there is also one informal selection that statistical papers prefer to using first-person pronouns. Therefore, this article presents these conventions about language formal choices in writing. In addition, language analysis also let graduate students understand that parts of language choices will have special types of rules in different professional areas, which will cause the result of identifying both formal and informal language in the same articles. Therefore, this article suggests students to get more information about the distinctive provisions when they are writing statistical articles.

**Reference**

Swales, J. M. & Feak, C. B. (2012). *Academic writing for graduate students, 3rd edition: essential skills and tasks*. Michigan ELT: University of Michigan Press.

Faraway, J. J. (2014). Does data splitting improve prediction?. *Statistics and Computing, 24,* 1-12. doi: 10.1007/s11222-014-9522-9

Walter, L. (2015). I won’t tolerate it! Replacing formal words with phrasal verbs. *New words.* Retrieved from <http://dictionaryblog.cambridge.org/2015/04/01/i-wont-tolerate-it-replacing-formal-words-with-phrasal-verbs/>

Ruf. Rice. Edu. (n.d.). *Common errors in student research papers.* Retrieved from <http://www.ruf.rice.edu/~bioslabs/tools/report/reporterror.html>