

Received : Sep 26, 2021  
Revised : Oct 19, 2021  
Accepted: Nov 25, 2021  
Published: Nov 30, 2021

## The Incipits of Semi-Popular Articles about Language Learning

Yolanda Puteri Madani  
*English Language Education Study Program*  
*Universitas Negeri Jakarta*  
*Jakarta, Indonesia*  
*yolandaputerim99@gmail.com*

I Wayan Darya  
*English Language Education Study Program*  
*Universitas Negeri Jakarta*  
*Jakarta, Indonesia*  
*iwayandarya@unj.ac.id*

### Abstract

This study aims to identify the type of organization incipit used and analyze the language features of organization incipit in semi-popular science articles. This research was conducted by using qualitative methods and employing a corpus-driven analysis approach. The approaches are adopted from Hyland's framework of proximity, specifically the organization facet of incipits position. The result revealed that incipits in the semi-popular article can be broadly categorized into 5 categories and the most used category is 'research finding', followed by 'establishment of a common ground', then 'contextualization of the article', after that is 'research problem' category, and the least identified category is 'benefit of the finding', it means there are 5 ways of strategy to attract readers. This study revealed the category can be identified by looking at the language features each incipit has, whether it is through the word frequency, concordance, and content organization. The study is important to help aspiring writers learn to convey the result of their research in a semi-popular way, which is the one that is professional but can be understood by an amateur specialist, especially about language learning.

**Keywords:** *Semi-popular Science Articles, Science Popularization, Proximity, Corpus-driven analysis approach*

### INTRODUCTION

Scientific disciplines are often seen as inaccessible, difficult, and not connected to everyday life (Pelger, 2015), which is not what it is supposed to be. Scientific research is expected to go beyond the borders of science and to the lay public, because as Albert Einstein astutely stated, "It is just as important to make knowledge live and to keep it alive as to solve specific problems" (Einstein, [1932] 1954: 70). To connect the science community and their scientific research with the lay public, science popularization is needed.

Popularization, as defined by Calsamiglia and van Dijk (2004), is "a vast class of various types of communicative events or genres that involve the transformation of specialized knowledge into 'everyday' or 'lay knowledge'". Thus, science popularization is when scientific information is formulated in a way that can make lay public construct their understanding of the information and integrate it with their initial knowledge by defining specific terminology, giving examples, or using metaphors, to promote the scientific information to the non-expert audience and shape a public discussion (Calsamiglia and van Dijk, 2004; Perrault, 2013).

In 2010, Hyland introduced the term proximity that is used to identify the writers' strategy to communicate and make science more accessible to the lay public by taking the content in a science research article and adjusting its specialized terminology to a more widely used word. There are 5 facets of proximity that Hyland introduced, they are organization, argument, credibility, stance, and engagement. Those 5 facets are important for science popularization, but the organization facet is probably the most predominant because, in popular science genres, the main claim is positioned at the incipits part of the article. This facet's role is to observe how the writers adapt the pattern and positions of information within the popular article to match the target readers.

There are two distinct genres in the science popularization article continuum, they are popular science articles and semi-popular science articles. Popular science articles are written by journalists and produced for the lay public who read scientific news not to gain professional knowledge but simply to keep updated on the development (Hyland, 2010). Meanwhile, semi-popular articles are written by researchers or journalists who are trained to communicate scientific information (Ciapuscio, 2003) with the target audience of readers with prior knowledge of the discipline who needs to gain more understanding of the discipline (Muñoz, 2015), such as undergraduate students. As Muñoz (2015) stated, a semi-popular science article is a less specialized and less conceptually complex research article that is more accessible for students to fill in the knowledge gap between general English and specialized text used in research articles.

The previous study that is similar to this is a study conducted by Scotto di Carlo in 2014, which analyzed TED talks based on Hyland's concept of proximity (2010) and stated TED talk as another popular genre. The findings of the research show that, in terms of organization, most speakers would start the speech by recounting a personal experience (43%). The result is not surprising because the TED speech aims to create direct contact between scientists and the audience, which is the lay public.

That aim relates to this study which goal is to identify the type of organization incipit used and analyze the language features of organization incipit in semi-popular science articles to know how the organization proximity is used in the incipits part of the semi-popular article with its aim to attract readers, which is based on Hyland (2010) who states that the main claim of the popular article is foregrounded at the incipits and to connect the science community and their scientific research with the lay public. The study is important to help aspiring writers learn to convey the result of their research in a semi-popular way, which is the one that is professional but can be understood by an amateur specialist, especially about language learning.

### RESEARCH METHOD

This research uses qualitative analysis methods and employs a corpus-driven analysis approach. Qualitative analysis is used to identify the type of incipits, meanwhile, a corpus-driven analysis approach is employed to identify the structure and the language features of the incipits.

The data of this study are words, phrases, and sentences of the incipits part of the article. The data source is 81 semi-popular science articles about language learning from various universities or other science affiliation websites which were gathered through Science Daily websites as the tool. The 81 articles were categorized into three sub-corpora based on the topic and focus each character brings with the assumption that the article would use a different kind of rhetorical features for each sub-corpus.

| Sub-corpora                                 | Description   | Total article |
|---|---|---------------|
| The nature of language learning articles.   | Discuss the natural causes, phenomena, or processes about the acquisition of language skills. | 51            |
| The benefits of language learning articles. | Discuss the benefit, profits, and rewards that came from the research.                        | 11            |
| The methods of language                     | Discuss the process, practice, manner, or   | 19            |

|                   |                                   |  |
|-------------------|-----------------------------------|--|
| learning articles | technique of acquiring a language |  |
|-------------------|-----------------------------------|--|

Table 1. The sub-corpora description.

The data collection procedures of this study start with collecting the incipits part of 81 semi-popular articles about language learning. They are written by various scientists and specialized journalists with diverse institutional backgrounds from Europe, America, and Asia. Having diverse writers from various backgrounds is important to minimize a biased result, and to bring a generalized study result.

There are a lot of popular science article websites, but this study chooses Science Daily because it relates to the aims of this study. This website provides language learning articles more than other websites, written by researchers or journalists who are trained to communicate scientific information and has a target audience of amateur experts of the discipline, like university students, which is the same target audience as the semi-popular article. By using Science Daily websites, the articles from various university websites are collected easily.

The incipit part of the article is categorized to find out the trends and patterns that semi-popular science articles about language learning use to start their article. This analysis is helped by a second-rater with the same educational background as the first rater and is known to be a capable student, that's why she is believed to be qualified as the second-rater. The categorization was conducted in the following ways:

1. Collect all the data consisting of the incipits part of 81 semi-popular articles about language learning.
2. Divide the data into 3 topics, which are 'the nature of language learning', 'the benefits of language learning', and 'the methods of language learning'
3. The two raters independently conduct analysis and coding of a sample of 15 articles, 5 articles from each topic.
4. Both raters adjust their findings of the coding rubric or the categorization of each incipit until they reach an agreement. If the percentage of their agreement is low, they need to analyze the data together. If the agreement is more than 70%, they can continue.
5. After they reach an agreement, both raters independently analyze the rest of the articles, which are other 66 articles from all three topics.
6. Once again if there are any significant discrepancies between the rater's coding or categorization, it will be discussed until they reach an agreement.

To analyze the language features of the incipits, corpus-driven is applied to collect corpus data that were derived systematically from the recurrent patterns and the frequency distributions that emerge from the language in context, so it is suitable with the research aim that is to find out the recurrent patterns of sentences, and the distribution of words frequency that appears from the language in semi-popular article and each of its 3 sub-corpora. This part is conducted in the following ways:

1. Copy only the incipits part of the 81 articles into a document.
2. Separate them based on topics, category, and category per topic, then save it in a .txt format.
3. Open the AntConc software, insert a file through the Open File(s) menu.
4. Choose the desired file in a .txt format.
5. Analyze the frequency using the word-list tool by choosing the word-list menu and then clicking Start.
6. Check the word's concordance by clicking the word.

Through the concordance result from the software, we can conclude which word/word cluster is used and how it is presented in the incipits to construct a certain topic or certain category.

## RESULTS AND DISCUSSION

The positioning of information and the language use applied in the incipits part of the article needs to be constructed in a strategic way that can attract the readers. Partly refers to the incipit category in Scotto di Carlo (2014), the incipits category identified in this research are grouped into 5 categories, they are research findings, the establishment of common ground, contextualization of the article, research problem, and the benefit of the findings.

| NO | ICIPITS CATEGORY                   | DESCRIPTION                                      |
|----|------------------------------------|--|
| 1  | Research Findings                  | The finding of the research in the article.      |
| 2  | The Establishment of Common Ground | Establishing common knowledge about the topic.   |
| 3  | Contextualization of the Article   | A rough contextualization of the entire article. |
| 4  | Research Problem                   | The issue/theory the research aims to clear.     |
| 5  | Benefit of the Finding             | Promoting the benefit of the research finding.   |

Table 2. Incipits used in the semi-popular science articles on language learning from Science Daily.

With the help from the corpus analysis tool software, AntConc, the results of the word frequency and concordance analysis of the incipits are generated so it can be contextually analyzed to identify its use and importance in representing the message of incipits in the semi-popular article about language learning.

### A. Results

The incipits categorization results of the semi-popular article about language learning revealed that the highest-ranking category used in all 81 articles is 'research finding', as displayed in detail in Table 3.

| NO    | ICIPITS CATEGORY                   | TOTAL | % OF TOTAL |
|-------|------------------------------------|-------|------------|
| 1     | Research Findings                  | 43    | 53,09%     |
| 2     | The Establishment of Common Ground | 17    | 20,99%     |
| 3     | Contextualization of the Article   | 13    | 16,05%     |
| 4     | Research Problem                   | 6     | 7,41%      |
| 5     | Benefit of the Finding             | 2     | 2,47%      |
| TOTAL |                                    | 81    |            |

Table 3. The number and percentage of each incipit category used in semi-popular science of language learning articles.

With the help from AntConc software as the tool, some aspects of the category can be identified by looking at the language features each incipit has, whether it is through the word frequency or its concordance. In detail for each category, the language features that helped to identify the category are:

1. The language features of the research finding category consist of the words '**new**' (to emphasize the novelty of the finding), '**study**' (to emphasize the research), and '**according**' (to give credit), the last word usually followed by the affiliation or the researcher's name that did the research.
2. The language features of the establishment of a common ground category consist of the content organization, which starts by discussing the general fact of the research subject or by rhetorical questions.
3. The language features of contextualization of the article category consist of its content organization that starts with discussing the general fact of the research subject and the mentions of affiliation or the researcher's name that did the research. In some cases, the two-sentence uses the word '**now**' as a conjunction.
4. The language features for the research problem category consist of interrogative words, such as the word '**how**', and some open the incipits with the conjunction '**although**'.

5. The language features of the benefit of the findings category consist of its content that mentions the research finding and adding a positive or persuasive sentence to highlight the usefulness of the finding.

Besides the language features that help to identify every category, there are also language features that help to identify the semi-popular articles about language learning itself. They are **‘language’** which is identified 126 times across topics; the word **‘new’** which is identified 55 times across topics; the word **“learn/learning”** which is identified 34 times across the topics, and the word **“study”** identified 29 times across the topics.

The percentage of category incipits resulting in the whole 81 articles are different in the 3 sub-corpora of the nature of language learning sub-corpus, the benefits language learning sub-corpus, and the methods of language learning sub-corpus.

1. The nature of language learning.

The number and percentage of the categories in this sub-corpus show us there are similarities in the ranking with the general result from all 81 articles as displayed in Table 4. This specific sub-corpus can also be identified using the word **‘brain’**, which is the highest frequency identified word by 20 times.

| NO    | INCIPITS CATEGORY                  | TOTAL | % OF TOTAL |
|-------|------------------------------------|-------|------------|
| 1     | Research Findings                  | 28    | 54.90%     |
| 2     | The Establishment of Common Ground | 12    | 23.53%     |
| 3     | Contextualization of the Article   | 7     | 13.73%     |
| 4     | Research Problem                   | 3     | 5.88%      |
| 5     | Benefit of the Finding             | 1     | 1.96%      |
| TOTAL |                                    | 51    |            |

Table 4. The number and percentage of each incipit category used in ‘the nature of language learning’ topic.

There are differences and similarities between the incipits of each category in this corpus and it can be identified by looking at the identified language features each of them has. In detail, the language features of the categories in this topic are:

- 1) Research finding.

The word **‘new’** in this category is identified to emphasize the novelty of finding; the word **‘study’** is used to emphasize the research itself, and the word **‘according’** is used to give credit to the researchers, and as an addition, this category would mention affiliation or researchers’ name. Look at the sample below:

- A new study carried out at the University of Haifa sheds light on how first and second languages are represented in the brain of a bilingual person. A unique single case study that was tested by Dr. Raphiq Ibrahim of the Department of Learning Disabilities and published in the Behavioral and Brain Functions journal, showed that first and second languages are represented in different places in the brain. (University of Haifa; July 9, 2009).

- 2) The establishment of common ground.

The organization of the content consisted of a brief introduction to the article through the discussion of the general fact of the topic. Here is the sample:

- For most native English-speakers, learning the Mandarin Chinese language from scratch is no easy task. (University of Delaware; May 8, 2019)

3) Contextualization of the article.

The organization of its content discusses general facts of the research, and then followed by the discussion of research finding that mentions affiliation or researchers, and in 5 out of 7 incipits. They also use the word ‘**Now**’ as a conjunction. See the following sample.

- Researchers have long debated whether or not language and music depend on common processes in the mind. Now, researchers at Georgetown University Medical Center have found evidence that the processing of music and language do indeed depend on some of the same brain systems. (Georgetown University Medical Center; September 28, 2007)

4) Research problem.

The language feature in this category mostly uses interrogative words such as ‘**how**’, but there is one starting with an ‘**although**’ conjunction that gives a skeptical effect to the statement after it.

- Although researchers have long agreed that girls have superior language abilities than boys until now no one has clearly provided a biological basis that may account for their differences. (Northwestern University; March 5, 2008)

5) Benefit of the finding.

The mentions of research findings followed by subjective and positive words are ‘**good news**’.

- Recent research into how we learn is set to help people in their efforts to read a second or foreign language (SFL) more effectively. This will be good news for those struggling to develop linguistic skills in preparation for a move abroad or to help in understanding foreign language forms, reports, contracts, and instructions. (Economic and Social Research Council (ESRC); March 6, 2013)

2. The benefit of language learning

The number and percentage of the Research Finding category make the category rank first even in this sub-corpus, and it shows that even though the rank is similar to the general result, the rest of the category ranking is very different, as displayed in Table 5. This specific sub-corpus can also be identified using the word ‘**speak**’, which is the highest frequency with 7 times occurrences.

| NO    | INCIPITS CATEGORY                  | TOTAL | % OF TOTAL |
|-------|------------------------------------|-------|------------|
| 1     | Research Findings                  | 5     | 45.45%     |
| 2     | The Establishment of Common Ground | 1     | 9.09%      |
| 3     | Contextualization of the Article   | 4     | 36.36%     |
| 4     | Research Problem                   | 0     | 00.00%     |
| 5     | Benefit of the Finding             | 1     | 9.09%      |
| TOTAL |                                    | 11    |            |

Table 5. The number and percentage of each incipit category used in ‘the benefit of language learning’ topic.

The incipits of each category in this corpus showed similarities and differences with the general result, and it can be identified by looking at the identified language features each of them has. In detail, the language features of the categories in this topic are:

## 1) Research finding.

The language features of this category in this topic can be identified by how it mentions affiliation, researchers' names, or background.

- Canadian scientists have found astonishing evidence that the lifelong use of two languages can help delay the onset of dementia symptoms by four years compared to people who are monolingual. (Baycrest Centre for Geriatric Care; January 12, 2007)

## 2) The establishment of common ground.

The organization of the content consisted of the existence of a rhetorical question to set a prior knowledge of the article.

- If you could save the lives of five people by pushing another bystander in front of a train to his death, would you do it? And should it make any difference if that choice is presented in a language you speak, but isn't your native tongue? (University of Chicago; August 17, 2017)

## 3) Contextualization of the article.

The organization of its content discusses general facts of the research and then followed by mentioning affiliation or researchers.

- Today, in Sweden a large proportion of school pupils have a foreign background and switch between their home language and Swedish on a daily basis. However, bilingualism in itself does not bring any advantages when it comes to learning a foreign language. A new thesis in German from the University of Gothenburg shows that bilingual upper secondary school pupils perform somewhat below average in language tests, compared with those who only speak one language. (University of Gothenburg; May 9, 2011)

## 5) Benefit of the finding.

The use of persuasive language about the importance of the topic.

- Want to learn to code? Put down the math book. Practice those communication skills instead. (University of Washington; March 2, 2020)

## 3. The methods of language learning

The number and percentage result of each incipits category in the methods of language learning sub-corpus shows that even though the first and second rank category is similar to the general one, the rest of the ranks are different, as displayed in Table 6. This specific sub-corpus can also be identified using the word '**university**', which is the highest frequency identified word by 14 times.

| NO    | INCIPITS CATEGORY                  | TOTAL | % OF TOTAL |
|-------|------------------------------------|-------|------------|
| 1     | Research Findings                  | 10    | 52.63%     |
| 2     | The Establishment of Common Ground | 4     | 21.05%     |
| 3     | Contextualization of the Article   | 2     | 10.53%     |
| 4     | Research Problem                   | 3     | 15.79%     |
| 5     | Benefit of the Finding             | 0     | 0.00%      |
| TOTAL |                                    | 19    |            |

Table 6. The number and percentage of each incipit category used in 'the methods of language learning' topic.

There are differences and similarities between the incipits of each category in this corpus and it can be identified by looking at the identified language features each of them has. In detail, the language features of the categories in this topic are:

1) Research finding.

The word **'new'** in this category is identified to highlight the finding novelty as it followed by words like **'study'**, **'research'**, and **'thesis'**, and the word **'according'** to give credit to the researchers, and as an addition, this category would mention affiliation or researchers' name.

- New research by neuroscientists at the University of Pittsburgh and University of California San Francisco (UCSF) revealed that a simple, earbud-like device developed at UCSF that imperceptibly stimulates a key nerve leading to the brain could significantly improve the wearer's ability to learn the sounds of a new language. This device may have wide-ranging applications for boosting other kinds of learning as well. (University of Pittsburgh; August 6, 2020)

2) The establishment of common ground.

The organization of the content consisted of a brief introduction to the article topic through general facts.

- Babies learn language best by interacting with people rather than passively through a video or audio recording. But it's been unclear what aspects of social interactions make them so important for learning. (University of Washington; July 27, 2015)

3) Contextualization of the article.

The organization of its content discusses general facts of the research and then followed by a discussion of the research finding, which in addition would mention affiliation or researchers' name.

- Your literacy skills in your first language heavily influence the learning of a foreign language. Thus, anyone who reads and writes German well is likely to transfer this advantage to English -- regardless of the age of onset of foreign language learning. Foreign language lessons at an early age, however, pay off less than was previously assumed. In fact, they can even have a negative impact on the first language in the short run, as a linguist from the University of Zurich reveals in her long-term study involving 200 Zurich high-school children. (University of Zurich; December 10, 2014)

4) Research problem.

The language feature here uses interrogative words such as **'how'** or mentioning the aims of the research.

- Experts say English slang and regional dialect should not be banned from classrooms but when you're getting to grips with a second language how helpful is it to learn non-standard. (Lancaster University; February 17, 2021)



*B. Discussion*

Identifying the type of organization incipit used in semi-popular language learning articles helps the aspiring writers to know what kind of incipits or intro that the writer can use to convey their message. For example, the most identified incipits category is the research finding, which means the article that uses this category starts theirs by highlighting the importance, and the significance of the new finding. That result shows that highlighting the finding is the most possible option they could pick. But the result of the rank of other categories also gives the writers an option of what they can highlight in the incipits, for example, they could choose the establishment of a common ground category if they want to set a common knowledge between the reader and the article content. The third one is the contextualization of the article category that equips the reader with a thorough prior knowledge for the research and states the findings of the research in the incipits, to give the readers a whole graph of the article and make them curious about the detail. The research problem category highlights the aims of the research and wants to give the readers a sense of direction the research wants to discuss. The least identified incipits category is the benefit of the finding, which is used to promote the benefit and utility of the findings that can be applied by the reader.

Analyzing the language features of organization incipit in semi-popular language learning articles could help the writer to construct and write the article. If the writers choose the research finding category, they already know they must include words like **‘new’**, and then followed by words like **‘study’**, and to give a credit to the researchers they need to include a word like **‘according’** before mentioning the affiliation or researchers’ name. With the establishment of a common ground category, the writer will highlight the content of the incipits which mentions the context of the article without discussing the research directly. For the contextualization of the article category, the writer would discuss the context thoroughly without discussing the main topic of the research, they also would discuss the finding of the research and mention affiliation or researchers’ name. If they choose the research problem category, they need to use interrogative words or mention the research aims. Lastly, for the benefit of the finding category, the writers would use a positive connotation to describe the finding.

The idea of using each categories’ language features to identify the category, and the name of some categories is adapted from Scotto di Carlo’s(2014) study. The categories of the starts of TED talks are ‘sharing a personal story/experience’ ‘contextualization of the talk’, ‘establishing a common ground with the audience’, ‘historical contextualization of the topic’, ‘humorous opening’ and ‘meta-reference relating to the ted context’. Meanwhile, the categories of incipits used in semi-popular science articles about language learning are ‘research finding’, ‘establishment of a common ground’, ‘contextualization of the article’, ‘research problem’, and ‘benefit of the finding’. Looking at the categories, there are two categories with the same name in either study, they are the ‘contextualization of the talk’, ‘establishing a common ground with the audience’ and ‘establishment of a common ground’, ‘contextualization of the article’, but the rest are different. The difference in other categories’ names is expected as each study has a different aim. The category in Scotto di Carlo’s study shows how TED speakers start their speech, meanwhile this research shows how the writers of semi-popular science articles attract the readers to read the article, so the categories are not the same.

**CONCLUSION AND RECOMMENDATION**

It can be concluded that the semi-popular genre uses 5 categories of incipit as various ways of strategy to attract readers. The category varies from aims and can be identified from its content and/or language features. The most used category is ‘research finding’, followed by ‘establishment of a common ground’, then ‘contextualization of the article’, after that is ‘research problem’ category, and the least identified category is ‘benefit of the finding’. This study identifies each category’s language features, such as word frequency, concordance, and content organization, to differentiate one category from the others. For example, the ‘research finding’ category uses words like ‘new’, ‘study’, and ‘according’ that are usually followed by the mentions of affiliation or researchers’ names. The ‘establishment of common ground’ category language features is in their content organization as they usually start their article by equipping the readers with a

general fact of the research topic as prior knowledge. The next one is the ‘contextualization of article’ category, and it can be identified through its content organization that starts by discussing a general fact of the research topic, then followed by mentioning the affiliation or researchers’ name. The fourth most used category is the ‘research problem’ category, the language features consist of the use of interrogative words like ‘how’ and using rhetorical questions. The last category is the ‘benefit of the finding’, this category uses persuasive and positive words to compliment the finding to promote them.

After implementing this study, the researcher suggests aspiring semi-popular article writers consider using the identified category and express the scientific information using the language features to increase the lay public’s interest in reading scientific information.

## REFERENCES

- Calsamiglia, H., and van Dijk, T. A. (2004). Popularization discourse and knowledge about the genome. *Discourse & Society*, 15(4), 369–389. <https://doi.org/10.1177/0957926504043705>
- Ciapuscio, G. (2003). Formulation and reformulation procedures in verbal interactions between experts and (semi-) laypersons. *Discourse Studies*, 5(2), 207- 233.
- Einstein, A. (1932). *Die Naturwissenschaften* 20: 913 (cited in (1954) *Ideas and Opinions*, trans. S. Bargmann, p. 70. New York: Crown Publishers, Inc.).
- Hyland, K. (2010). Constructing proximity: Relating to readers in popular and professional science. *Journal of English for Academic Purposes*, 9(2), 116–127. <https://doi.org/10.1016/j.jeap.2010.02.003>
- Muñoz, V. L. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes*, 39, 26-44. <https://doi.org/10.1016/j.esp.2015.04.001>
- Pelger, S., Nilsson, P. (2016). Popular Science Writing to Support Students’ Learning of Science and Scientific Literacy. *Research in Science Education*, 46, 439–456. <https://doi.org/10.1007/s11165-015-9465-y>
- Perrault, S. T. (2013). *Communicating Popular Science: From Deficit to Democracy*. London: Palgrave Macmillan.
- Scotto di Carlo, G. (2014). The role of proximity in online popularizations: The case of TED talks. *Discourse Studies*, 16(5), 591–606. <https://doi.org/10.1177/1461445614538565>