Computational Linguistics and Chinese Language Processing
Vol. 27, No. 1, June 2022, pp. 75-110

An N-gram Approach to Identifying the Chinese Linguistic Signals for the Problem-Solution Pattern in Annotated Online Health News

Chen-Yu Chester Hsieh* and Yu-Yun Chang+

Abstract

This article will report the results of an exploratory project that combined the annotation of the Problem-Solution (PS) textual pattern in online health news and the quantitative and qualitative methods of corpus linguistics to investigate the linguistic features of particular rhetorical moves. A total of 120 journalistic texts written in Chinese were collected from a Taiwan-based journalistic website that focused on providing news related to health and medicine and were annotated with the four components of the PS pattern. To identify signals in the genre for the elements of the PS move structure, an n-gram approach was then implemented to extract frequent lexicogrammatical sequences from the corpus in general and from the Problem and Response moves in particular. The results showed that the linguistic features found in the retrieved sequences tended to fall within a range of categories, such as abstract nouns, medical terms, and modal verbs, which not only served as functions relevant to the rhetorical move in which they were used but also reflected characteristics specific to the health news genre and the Chinese language. The findings and annotated data generated from the current project will thus provide a solid foundation for future research and applications.

Keywords: **N**-gram, Problem-Solution Pattern, Health News, Annotation, Journalistic Discourse

* Research Center for Translation, Compilation and Language Education, National Academy for Educational Research, New Taipei City, Taiwan
E-mail: chesterhugues@gmail.com
+ Graduate Institute of Linguistics, National Chengchi University, Taipei City, Taiwan
E-mail: yuyun@nccu.edu.tw
The author for correspondence is Yu-Yun Chang.

1. Introduction

As problem solving constitutes an essential activity and ability in various fields (Jonassen, 2000; Charles, 2011; Handford & Matous, 2015; Heffernan & Teufel, 2018), the Problem-Solution (PS) pattern is arguably one of the most common rhetorical structures across cultures and text types (e.g., Hoey, 2001; Flowerdew, 2003; Charles, 2011; Ratanakul, 2017). According to Hoey (1983, 2001), the PS pattern (also known as the SPRE [Situation-Problem-Response-Evaluation] or SPSE [Situation-Problem-Solution-Evaluation] pattern) is typically composed of four moves: Situation (i.e., background information regarding the topic or problem), Problem (i.e., issues that need to be addressed or fixed), Response/Solution (i.e., proposals or methods for dealing with the problem), and Evaluation (i.e., assessments of the mentioned solutions/responses).

Due to its prominence and popularity in English texts, the PS rhetorical pattern has garnered the attention of several text/corpus linguists (e.g., Hoey, 1983, 2001; Scott, 2000; Flowerdew, 2003, 2008; Charles, 2011). Previous findings have revealed that the PS pattern appears in a wide range of genres, such as advertisements (Hoey, 2001), academic reports (Flowerdew, 2003), research articles (Charles, 2011; Heffernan & Teufel, 2018), journalistic texts (Ali, 2013; Ratanakul, 2018), and even TED talks, (Ratanakul, 2017). The moves of the pattern may be ordered in different sequences, with some of the components repeated or excluded (Hoey, 2001; Belmonte, 2009; Ratanakul, 2018), and they are often signaled by particular identifiable linguistic features, such as abstract nouns (Flowerdew, 2003, 2008) and adverbials (Charles, 2011), and are supported by various communicative functions (Belmonte, 2009; Ali, 2013).

Despite the number of studies on this topic, a few research gaps still remain. First, most, if not all, of the past studies drew on data written or delivered in English. However, no evidence indicates that this discourse pattern is restricted to the English language. Recent research has suggested that this rhetorical pattern is prevalent in English texts produced both inside and outside the Inner Circle (Kachru, 1992), where English is learned and used as the first or native language (e.g., Ali, 2013; Handford & Matous, 2015; Ratanakul, 2018). Scholars such as Belmonte (2009) and Ali (2013) have also advocated the compilation of multilingual corpora or a cross-linguistic comparison of the PS pattern. Applying this textual structure in the analysis of articles composed in Chinese would allow for a closer look at how this textual pattern is signaled and realized in typologically and culturally distinctive languages.

Another major gap in the literature pertains to the genres that have been investigated. Although journalistic writing is one of the most often studied text types with respect to the SPRE pattern, only business news texts and opinion articles have been examined in more detail (e.g., Belmonte, 2009; Ali, 2013). Medical and health news, which presents information pertinent to medical conditions and their treatments (李松濤 & 鄔啟柔, 2017; 李松濤 & 許

文怡, 2020), has not yet been scrutinized under the framework of the PS pattern. As has been noted by scholars, journalistic texts regarding medical and health issues have substantial influence over readers' well-being and lifestyle decision-making (Fu *et al*., 2020; De Coninck *et al*., 2020), and may even shape their news and scientific literacies (李松濤 & 鄔啟柔, 2017; 李松濤 & 許文怡, 2020); as such, this particular genre is worthy of further investigation.

Finally, most of the previous studies on the PS textual structure adopted a keyword analysis approach (Scott, 2000; Flowerdew, 2003, 2008) or simply focused on a given part of speech or communicative act (Charles, 2011; Ali, 2013). Although these methods may be suitable for English, a language that has clearer word boundaries, less disputable syntactic categories, and more literature related to the topic in question, they may not be as applicable to analyzing texts written in languages that do not have the above characteristics, such as Chinese (Fu *et al*., 2020; Tian *et al*., 2020). As a result, an even more data-driven approach, such as n-gram analysis, may be preferable for an exploratory investigation of the linguistic signals for an understudied rhetorical pattern in articles composed in Chinese.

In light of the above gaps, the current paper aims to report the results of a preliminary project that examined the PS pattern in online health-related journalistic articles from one of the most popular and influential health news websites based in Taiwan by integrating the annotation of the PS discourse structure and the methods of corpus linguistics. In addition to not having been investigated under the SPRE framework, the genre of online health news was chosen because it has become one of the major sources of medical information for the general public, with real-life consequences (Fox & Duggan, 2013), and because gaining more knowledge about how this genre is composed and organized will allow for opportunities to improve media and science literacies (李松濤 & 鄔啟柔, 2017; 李松濤 & 許文怡, 2020) and to develop algorithms that can help detect and retrieve information about particular medical conditions and recommended solutions, facilitating the process of solving health-related problems (cf. Ghenai & Mejova, 2018; Waszak *et al*., 2018; Kumari *et al*., 2021; Tsirintani, 2021). More specifically, the current study set out to answer the following research questions:

RQ 1: What are the distribution patterns of the SPRE moves in online health news articles written in Chinese?

RQ 2: What are the most common Chinese n-grams (trigrams in this study) in the online health news articles and what features do they display?

RQ 3: How do the most frequent trigrams reflect the characteristics of this genre?

RQ 4: What are the most frequent and most distinctive trigrams in the Problem and Response moves of the online health news articles? What are some of their generalizable features?

RQ 5: How do these trigrams help achieve the rhetorical functions of the Problem and Response

moves in the online health news articles?

By answering the above questions, this study demonstrated that the n-gram approach was a useful framework for retrieving the linguistic signals of specific move components, the results of which will add to prior research on move structures, journalistic genres, and the annotation of Chinese texts. Moreover, by using a small set of data, the findings of the current study, which was exploratory in nature, will serve as the foundation for future research using a larger amount and greater diversity of data.

2. Related Works

As one of the earlier scholars that paid close attention to the PS pattern in English, Hoey (2001) put forth a number of insightful observations about this discourse structure. First, the PS pattern is created by the writer of a text to respond to a series of questions although the order of the moves is not fixed. Second, the linguistic signals in the PS pattern are typically lexical items, including words that are explicitly related to the moves, such as "solution" and "problem," and evaluative devices, such as "unfortunately" and "have no money." Third, the Situation move is often retrospectively identified and an intervening stage, such as planning or recommendation, can be found between the Problem move and the Solution move. Lastly, the PS pattern can be recycled when more than one solution is presented. Therefore, in lieu of the Situation-Problem-Response-Evaluation pattern, the move structure can be as complex as Situation → Problem → Response 1 → positive Evaluation → negative Evaluation → Response 2 → positive Evaluation (Hoey, 2001: 134).

2.1 The PS Pattern in Academic Writing

Based on Hoey's (2001) insights, a number of researchers have investigated the realization of the PS pattern in various contexts. One of the most investigated genres in the literature is academic writing (Flowerdew, 2003; Charles, 2011; Heffernan & Teufel, 2018). For example, conducting a keyword analysis to compare the PS pattern in the technical academic reports composed by novice and professional writers, Flowerdew (2003; 2008) observed that the Problem move was closely linked to causal relationships, such as Reason-Result and Means-Purpose, and was often marked by lexical devices, including connectives such as "therefore" and "as a result," causative verbs such as "lead to" and "avoid," and abstract nouns such as "problems" and "impacts." Moreover, the nominalization forms of verbs such as "implementation" and "recommendation" together with verbs in the present perfect such as "recommended" were found to be indicative of the Solution element in reports composed by professional writers (Flowerdew, 2008).

With a focus on the adverbials of "result" and "contrast" in theses written by native-speaker writers in politics and in materials science, Charles (2011) analyzed the functions and co-occurrence patterns of the most frequent adverbials of "result" (i.e., "thus") and "contrast" (i.e., "however") in the PS pattern and found that despite the differences in number between disciplines, in general, "however" was used to mark the Problem element and "thus" was used to signal the Evaluation of the Response move when the former preceded the latter. On the other hand, when the order was reversed, while "however" was still the signal for the Problem move, the preceding "thus" marked the Situation move instead.

More recently, Heffernan and Teufel (2018), drawing on a machine learning approach and a corpus of problem and solution statements, specified 15 features to help develop classifiers that could identify Problem and Solution elements in scientific texts, including n-grams, polarity, and syntax. Heffernan and Teufel (2018) asserted that their model accurately differentiated problems/solutions from non-problems/solutions and that syntactic information, documenting, and word embedding were the three best features that allowed them to achieve the target task.

2.2 The PS Pattern in Journalistic Texts

Another major line of research looked into how the PS pattern was realized in various types of journalistic genres, including feature articles (Scott, 2000), opinion pieces (Belmonte, 2009; Ratanakul, 2018), and business news (Ali, 2013). Scott (2000), one of the earliest to examine the use of the PS pattern in journalist texts, found that, surprisingly, the nouns "problem" and "solution" were not especially frequent and thus not "key" enough in the Guardian's feature articles. Even when the two nouns were characterized as keywords, their signaling function tended to be local rather than global, in contrast to what had been commonly presumed.

Belmonte (2009), investigating the rhetorical organization of editorials and op-eds in USA Today, identified a number of characteristics specific to those genres with regard to the PS textual pattern. First, the textual pattern in the two genres tended to revolve around the Problem and Evaluation elements, while the Solution move was not as prominent. This indicated that it was a genre convention for the editorial writer to leave the solution to the problem unspecified. Second, different components of the PS structure displayed the tendency to occur in different parts of the articles, each with a particular rhetorical function. For example, when the Evaluation element was presented at the end of a text, it was usually more negative, creating a feeling of discomfort or dissatisfaction. Third, despite the fact that editorials and op-eds are both, in general, evaluative texts supported by other rhetorical roles, such as elaboration and justification, and by other communicative acts, such as statements and assertions, the PS pattern in editorials tended to show a more impersonal style, while that in op-eds was more often presented in a more involved style, with more frequent use of shared knowledge assertions and recommendations.

Similarly, Ratanakul (2018) carried out a move analysis of opinion columns in two newspapers online, the New York Times and China Daily, based on the SPRE framework. Partially echoing Belmonte's (2009) findings, Ratanakul (2018) also noted that the Problem move constituted the most central component of the PS pattern in the articles, although the Response element was found to be the second most frequent move in the data, in contrast to what Belmonte (2009) reported. In addition, Ratanakul (2018) pinpointed several features of each move in the PS pattern, both obligatory and optional, including causes of the problem (Problem) and a call for action (Solution), which indicated that each component of the target discourse structure could be further divided and analyzed.

On the other hand, Ali (2013) examined the PS textual pattern and its communicative functions, such as informative, evaluative, explicative, and predictive, found in journalistic articles in business magazines published in Malaysia (i.e., Malaysian Business [MB]) and the United Kingdom (i.e., Management Today [MT]), respectively. In agreement with previous studies such as Belmonte (2009), Ali's (2013) findings suggested that elements in the PS pattern, such as Problem and Solution, contained multiple speech and communicative acts and that the frequency and distribution of each act in the moves differed across sources of the texts. Nevertheless, no one-to-one connections were found between the communicative acts and the Problem and Solution components in the business news articles. For example, informative, evaluative, and predictive acts were identified in both the Problem and Solution moves.

All in all, the works reviewed above point to the fact that while the PS pattern is a prevalent rhetorical structure across genres, factors such as topics, text types, and language variations influence the realization of the PS pattern. An analysis of this pattern in languages other than English, as several researchers have suggested, is thus warranted. The present study aimed to contribute to this line of research.

2.3 Functional Classification of N-grams

In the realm of applied linguistics, n-grams are studied under different labels, such as "lexical bundles" (Biber *et al*., 2004) and "multiword expressions/sequences" (e.g., Staples *et al*., 2013). One of the most widely adopted functional taxonomies was proposed by Biber *et al*. (2004), in which lexical bundles serve as the functions of "stance," "referring," and "discourse organizing." Stance bundles allow users to mark an epistemic or attitudinal modality (e.g., 'I don't know if' and 'if you want to'), while referring bundles are used to identify an entity, convey imprecision, specify an attribute, or refer to a particular place, time, or part of the text (e.g., 'that's one of the', 'something like that', 'there's a lot of', and 'at the same time'). Finally, discourse organizing bundles introduce a focus or manage topics (e.g., 'take a look at' and 'on the other hand').

While Biber *et al*.'s (2004) framework is the most general, commonly used functional

taxonomy, a number of more genre-specific classifications have also been proposed. For example, Biber and Gray (2013) modified the taxonomy put forth by Biber *et al*. (2004) for the analysis of responses in the TOEFL iBT speaking and writing tests. To accommodate the purpose of the writing tasks, they divided the stance bundles into two subgroups, "personal/epistemic" and "attitudinal/evaluative," and the discourse organizing bundles into three subtypes, "information source," "information organizer" (marking more specific information in the writing), and "discourse organizer" (serving more general discourse functions in the discourse), expanding the taxonomy into a five-way categorization (notably, this classification does not include the referring function). Biber and Gray (2013) found that the distribution pattern varied in accordance with the task type and the writer's performance level.

To investigate research articles, master's theses, and dissertations from different disciplines, Hyland (2008) developed a functional taxonomy that grouped n-gram expressions according to whether they were "research-oriented," "text-oriented," or "participant-oriented." In that taxonomy, research-oriented bundles refer to expressions that facilitate writers' presentation of experiences and activities in the real world, such as location and procedure, whereas text-oriented bundles pertain to the organization of texts, such as transition signals. Finally, participant-oriented bundles are sequences of words that involve the writer and/or the reader of the text and may serve functions related to stance and engagement (cf. Hyland, 2008).

On the other hand, some researchers categorized n-gram expressions based on their function in rhetorical moves. For example, analyzing the sentence-initial multiword expressions in Arts and Humanities PhD dissertation abstracts, Li *et al*.'s (2020) tailor-made functional taxonomy consists of "background bundles," "purpose bundles," "method bundles," "findings bundles," "implications bundles," and "structure bundles." According to Li *et al*. (2020), each of these types helped achieve the rhetorical function of a specific move in the abstracts, such as stating the research purpose or outlining the structure.

3. Methodology

3.1 Data Collection and Annotation

We targeted the health topic for the analysis of the moves in the PS discourse structure since it has been widely abused and propagated through content farm articles on the Internet. Several studies have worked on analyzing and detecting the spread of health misinformation (Ghenai & Mejova, 2018; Waszak *et al*., 2018; Kumari *et al*., 2021; Tsirintani, 2021). An increasing number of people prefer to seek health and medical solutions by themselves first rather than directly consult medical professionals. The Pew Research center conducted a survey in 2013 and found that 59% of adults sought health information online and 3% were harmed by misinformation (Fox & Duggan, 2013).

To build a corpus annotated with the moves in the PS pattern, articles published on the Heho 健康 website were retrieved and collected.[1] Heho is a popular health- related website in Taiwan that provides credible health-related information contributed and certified by healthcare professionals. Although there are other platforms that publish and disseminate health information to the public, the content is often not provided by reliable sources and may have been released by content farms. Since many people prefer to seek solutions for health issues online beforehand, we focused on investigating the PS discourse structure for content that involved questions frequently asked by people and solutions given by medical experts. To the best of our knowledge, no research has explored the online health information issue under the scope of the PS discourse structure. Thus, 120 articles (139,131 tokens) published by 10 journalists in the series 醫生說 'The Doctor Says' under the 請問專家 'Asking Experts' topic were collected, with the publication dates ranging from March 11 to July 12, 2021. The retrieved articles in the series drew on various expert sources and involved specific health-related problems and solutions. Another characteristic of the articles was that the main problem highlighted by the authors was addressed directly in the articles' titles.

To annotate the articles, we adopted the definitions of the PS moves specified in Ratanakul (2018: 235) to identify each element of the PS pattern, as listed below:

- Situation: background information on situations; facts about people, issues, events, places involved in the issue of discussion
- Problem: aspect of a situation requiring a response, need, dilemma, puzzle, or obstacle under discussion; weaknesses inherent to the current situation
- Response: solution(s) to the problem; discussion of a way(s) to deal [with] or to solve the problem
- Evaluation: assessment of the effectiveness of the proposed solution(s); if there is more than one solution, which solution is the best?

To increase the efficiency of the annotation task, we utilized one of the prevalently adopted annotation tools in corpus linguistics, GATE.[2] Although the definition of the annotation scheme was provided, most previous studies that annotated the PS pattern simply made up or took some examples from the texts for analysis, without annotating the entire article, because the annotation scheme as well as the PS pattern were proposed at the discourse level instead of the sentence level, which made it difficult to set clear boundaries for annotation. In this study, we followed the definition of the annotation scheme and annotated the PS pattern throughout all the articles. As each of the moves could occur more than once in an article, several moves

[1] The Heho 健康 website is available at: https://heho.com.tw/

[2] The GATE annotation tool is available at: https://gate.ac.uk/download/

in the same article were annotated with the same move label. Thus, each article was labeled with several moves of Problem, Response, Situation, and Evaluation. In the dataset, since the titles of the articles revealed the main problems to be discussed, each annotator was asked to first identify the problem(s) in an article based on its title. The title of one of the retrieved articles is shown in (1). As suggested by the title, the main problem discussed in this article was 剖腹產後會有沾黏風險 'there is a risk of adhesion after a C- section'. Based on the identified problem, the PS moves in the article were then annotated accordingly.

(1) 剖腹產後會有沾黏風險！產科權威提醒「抗沾黏」3 大步驟
'There is a risk of adhesion after a C-section! The obstetrician authority suggests three steps for anti-adhesion'

To ensure that the annotators followed and understood the annotation scheme consistently, a training session with labeled articles was held before the annotation task was performed. In order to maintain annotation quality during the task, a checkpoint meeting was arranged. In the meeting, all the annotation issues raised by the annotators were discussed. The annotators were asked to modify their annotations based on the discussions. Since it would take more time for the annotators to label the PS pattern throughout the entire article, each article was annotated by one annotator. Four annotators were recruited for this annotation task. All 120 articles collected were equally distributed to the four annotators, so each annotator was assigned 30 different articles.

3.2 Data Analysis

In this study, the trigrams were character-based instead of word-based under the following considerations. Character-based n-grams have been widely applied in different tasks, such as text classification (Cavnar & Trenkle, 1994), spam filtering (Kanaris *et al*., 2007), authorship attribution (Escalante *et al*., 2011), and plagiarism detection (Kuta & Kitowski, 2014). Some studies have demonstrated that character-based n-grams are more effective in generating word embeddings for unknown words (Wieting *et al*., 2016; Bojanowski *et al*., 2017) and are more informative in performing topic categorization and document summary (Giannakopoulos & Karkaletsis, 2009). Since Chinese is not a space delimited language, the issue of defining a Chinese word has long been discussed and is still disputable (Ng & Low, 2004; Tian *et al*., 2020). Although several Chinese segmenters have been released, segmentation results are still hard to evaluate and segmentation errors have led to coarse-grained data analysis. Moreover, based on the limited number of annotated articles, processing word-level n-grams in a dataset has also led to fewer n-gram types, resulting in less statistical effectiveness in analyzing PS

patterns.[3] Therefore, we considered each Chinese character the basic unit of analysis in this study and targeted trigrams for further exploration. We also applied the Natural Language Toolkit in Python to retrieve trigrams from the collected dataset.[4]

N-gram patterns with a larger n may occur only once in a dataset, resulting in relative frequencies that are close to zero. It is difficult to observe the PS pattern based on raw frequencies and relative frequencies, since both values cannot be placed on a normalized scale for exploration. Different scaling approaches (e.g., z-score and Min-Max) may be applied to help normalize the dataset before investigation. The normalization techniques of z-score and Min-Max scaling are different. Z-score scaling takes the entire distribution of a dataset into consideration by calculating the mean and standard deviation, and the range of the z-score is affected by the dataset distribution accordingly, either positive or negative. Min-Max scaling has the benefit of scaling all the values into positives and transforms the values into a range between 0 and 1. Since it is easier to inspect a dataset with a fixed range of normalized values, we chose to apply Min-Max scaling in this study.

After the target trigrams were retrieved and selected, the trigram expressions most prevalent in the two core moves of the PS pattern (i.e., Problem and Response) were identified and treated as signals for the respective moves. To analyze how these signals helped achieve the communicative goals of the moves, the trigrams were manually checked and categorized structurally according to the part of speech and semantic features of the main component and functionally based on Biber *et al.*'s (2004) tripartite taxonomy. The functional categories consisted of (i) stance (including epistemic and attitudinal stances); (ii) discourse organizing (including marking the information sources, managing the topic and focus, and signaling other

[3] We performed word-level trigrams with the jieba segmenter in our preliminary study, as suggested by one of the reviewers. After segmentation, the highest frequency of a word-level trigram in the corpus was 22, which was significantly lower than the highest frequency of a character-level trigram (i.e., 113). The top four word-level trigrams were 異位性皮膚炎 'atopic dermatitis', 機器人手臂 'robotic arms', 李孝貞醫師說 'the doctor Xiao-Zhen Li says', and 的情況下 'situation of'. Compared with the character-level trigram results listed in Table 2, the word-level trigrams were mostly technical or topic-specific terms, rather than carrying the linguistic cues of the PS pattern. Thus, character-level trigrams were taken into consideration in exploring the PS pattern in this study.

[4] In the present study, trigram sequences in particular were examined for further analyses for a number of reasons. First, a number of bigram sequences did not convey a complete meaning (e.g., 。」and ，如) and thus did not produce enough PS patterns for analysis. Second, the number of sequences was too high for more qualitative analysis. Third, sequences that included more than three characters were either specific technical terms instead of patterns, such as 攝護腺癌 'prostate cancer' (4-gram), 代謝症候群 'metabolic syndrome' (5-gram), and 異位性皮膚炎 'atopic dermatitis' (6-gram), or too few in frequency for generalization. As a result, trigram sequences were the best unit of analysis for the current research.

discourse relations); and (iii) referring (including identifying the focus of the move and referring to circumstantial elements such as time and space).

4. Results and Discussion

4.1 Descriptive Statistics of the Annotated Dataset

Among the 120 articles collected, 113 articles were judged as containing the PS pattern (henceforth target articles), and the remaining seven articles were ignored and left unannotated since they were not the focus of this research. The structure of each article contained an average of 13 paragraphs (SD=4.45) and 1,159 tokens (SD=370), and 1,212 items were identified and annotated based on the four moves.

Table 1. Number of items in each of the four PS moves in the target articles

Moves	Situation	Problem	Response	Evaluation
Number of Items	156	357	672	27

Table 1 above presents the number of moves observed in the target articles. As can be seen, among the components of the PS pattern, the Response move and Problem move were substantially more addressed and employed than the other two moves (e.g., Situation and Evaluation) (cf. Scott, 2000; Hoey, 2001). This tendency may have been partly due to the characteristics of online health news (i.e., readers want to see the relevance of the problem and identify the recommended solution(s) within a short time, and they do not prefer to read too much technical information [cf. 葉蓉慧 & 黃暐超, 2017]).

To further investigate the linguistic features of the PS pattern in a corpus-driven and quantitative way, we extracted trigrams from the corpus. Since the raw frequency of n-grams can be biased by the corpus size, all the trigrams were further normalized by Min-Max scaling. Table 2 below presents the top 10 frequently occurring trigrams with their corresponding statistical scores:

Table 2. Top 10 frequently occurring trigrams

Trigrams	Frequency	Relative Frequency	Min-Max
，如果	113	0.0009	1.0000
說：「	85	0.0007	0.7500
，所以)	80	0.0006	0.7054
，因為	75	0.0006	0.6607
科醫師	74	0.0006	0.6518
，可以	74	0.0006	0.6518
症狀，	71	0.0006	0.6250
的時候	70	0.0006	0.6161
攝護腺	69	0.0005	0.6071
的問題	69	0.0005	0.6071

As shown in Table 2, the relative frequencies of the trigrams were rather small as the largest value was only 0.0009, and they were highly sensitive to the corpus distribution and thus were not suitable to be used for observing patterns. After transforming the values via Min-Max scaling, the values were larger and thus easier to utilize for exploring the patterns.

Figure 1 shows an overview of the top 20 trigrams transformed via the Min-Max scaling approach:

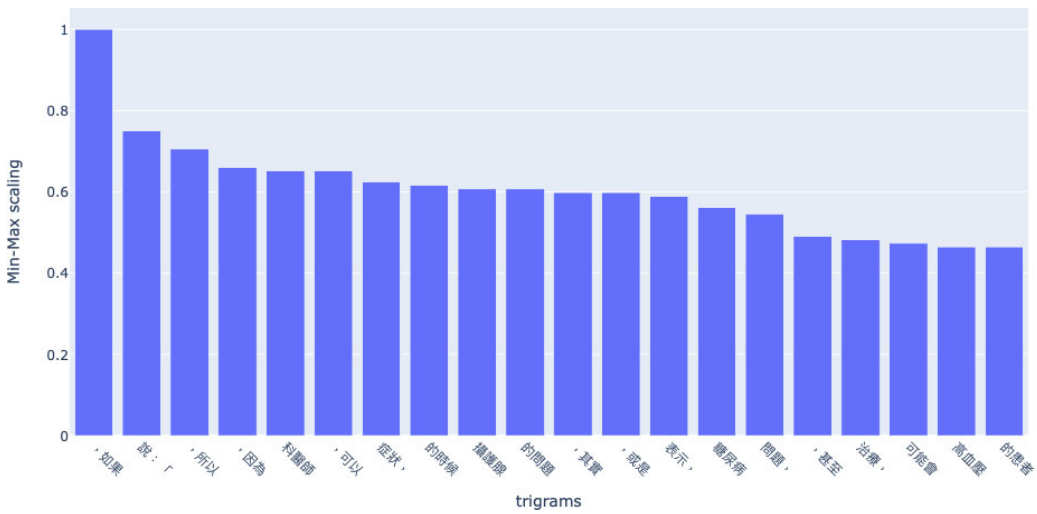

Figure 1. Min-Max scaling of the top 20 trigrams

The threshold of the Min-Max scale value for the trigram expressions that would be further analyzed was set at 0.3 to avoid hapax legomenon. Since the dataset was small, a Min-Max

scaling threshold of 0.2 may have resulted in instances of hapax legomenon. The threshold of 0.3 allowed a total of 58 trigram sequence types to be selected. Items containing terms that were too technical or topic specific, such as 攝護腺 'prostate', 糖尿病 'diabetes', and 高血壓 'hypertension', were eliminated from the list. Lastly, trigrams with a DP value higher than 0.85 based on the dispersion measure developed by Gries (2008) were excluded to avoid trigrams that occurred skewedly in a limited set of texts.[5] As a result, 48 types of trigram items were included for later analysis, and the full list is displayed in Table 3. It should be noted that, as can be seen in the following table, trigrams that included punctuation marks were retained for further analysis. Although in English-based n-gram studies punctuation marks are usually not counted as an element of an n-gram because they are found to convey discourse relationships (Yue, 2006) and occupy the same amount of space as a character does in written traditional Chinese, they were not eliminated from the data. These trigrams were then manually examined for structural (part-of-speech) and functional analysis.

Table 3. Target trigrams for further analysis

，可以	，可能	，因此	，因為	，如果	，其實
，或是	，所以	，甚至	，通常	，像是	來說，
治療，	表示，	症狀，	問題，	疾病，	問題。
解釋：	說：「	釋：「	常見的	這樣的	的治療
的狀況	的風險	的原因	的時候	的疾病	的症狀
的方式	的問題	的患者	的情況	的傷口	治醫師
是因為	可能會	有可能	就可以	越來越	會出現
會造成	臨床上	醫師說	科主治	科醫師	主治醫

4.2 General Observations of the Frequent Trigram Expressions

Far from being highly diversified, the most frequent trigram sequences were restricted to merely a number of categories, including connectives, adverbials, abstract nouns, health- and medicine-related terms, quotative devices, and a few others. This suggested that the genre of online health news as well as the PS pattern in that genre tended to feature the use of specific lexicogrammatical signals. For example, echoing Flowerdew's (2003, 2008) observations, linguistic devices that marked the PS pattern were predominantly indicative of the Reason-Result relationship, including 因為 'because', 所以 'so', 因此 'therefore', 造成 'to

[5] A DP value of a trigram is a number between 0 and 1. The higher the value, the more skewed the distribution of the trigram in the corpus is. See Gries (2008) for more detailed information about this dispersion measure.

cause', and 的原因 'the reason of'. Similar to the findings of previous studies (e.g., Hoey, 2001; Flowerdew, 2003, 2008), abstract nouns that were semantically connected to components of the PS pattern, explicitly or implicitly, such as 問題 'problem', 風險 'risk', 狀況 'condition', 方式 'method; approach', 原因 'reason', and 情況 'situation', were also found in the most frequent sequences. More notably, most abstract nouns were preceded by the grammatical morpheme 的, which links abstract nouns to another lexical or clausal unit. The preceding morpheme 的 can also serve as (part of) a shell noun construction (Schmid, 2000) that links abstract nouns to a co-referring phrase or clause (cf. Hsieh, 2020).

Results not reported in previous studies were also retrieved from the health news corpus. First, the conditional marker 如果 'if' constituted the most frequent trigram sequence in the dataset. As pointed out by Lin (2019), conditionals play a crucial role in both spoken and written medical communication, serving multiple functions such as presenting suggestions, explaining successive treatment, and calling for attention. Another group of lexical items that was also less often mentioned in the literature was modal verbs, including 可以 'can' and 可能 'may', which supported communicative acts such as making predictions and recommendations in this genre (Hsieh, 2005). Finally, in correspondence with Flowerdew's (2008) findings, verbs that indicated causal relationships such as 造成 'to cause' were also found in the most recurrent trigram sequences, although not as prevalent as Flowerdew (2008) suggested.

Finally, features that were heavily linked to a particular genre or topic were spotted in the list of trigram sequences as well. For example, several nouns in the realm of physiology and medicine were involved in many of the sequences, including 醫師 'doctor', 症狀 'symptom', and 傷口 'wound'. On the other hand, insofar as the data were journalistic texts in essence, reporting verbs such as 說 'say', 表示 'indicate', and 解釋 'explain' (Liu & Chiang, 2008) and typographical markers for quotations such as colons and quotation marks, which introduced quotes from the sources, were also found to be highly frequent (cf. Waugh, 1995), a tendency that has, again, rarely been reported in prior research. As will be shown in the following, many of these topic- and genre-oriented features were also signals for a particular component in the PS textual pattern.

4.3 Trigram Sequences in Rhetorical Moves

This section will focus on the trigram sequences that most frequently occurred in the Problem and Response moves, respectively, because these two moves not only theoretically constituted the most essential components of the PS pattern but also empirically accounted for the most move instances and contained the most high-frequency trigram expressions in the corpus. As will be presented later, although the Problem and Response moves shared a few types of trigram sequences with regard to their parts of speech, the sequences most commonly found in each of the two moves in fact differed from each other semantically and functionally, reflecting the

purposes that each move was intended to achieve.

4.3.1 Signals for the Problem Move

To begin with, a majority of the frequent trigram expressions in the corpus introduced above most frequently occurred in, and thus could be considered signals for, the Problem move. This move contained the widest range of trigram signals, including abstract nouns, medical terms, adverbials, connectives, reporting verbs, causal verbs, modal verbs, and others, as displayed in the following Table 4:

Table 4. Structural categories of the trigram signals for the Problem move

Abstract Nouns	**Medical Terms**	**Reporting Verbs**	**Modal Verbs**
的時候 'time of' 的問題 'problem of' 的原因 'reason of' 的狀況 'situation of' 問題。 'problem.' 的風險 'risk of' 問題，'problem,'	科醫師 'doctor of the Division of' 主治醫 'attending physician' 治醫師 'attending physician' 科主治 'attending physician at the Division of' 症狀， 'symptom,' 疾病， 'disease,' 的疾病 'disease of' 臨床上 'clinically'	說：「 'say' 醫師說 'doctor says' 表示， 'indicate,'	可能會 'probably will' ，可能 ', probably'
Other Verbs	**Connectives**	**Adverbials**	**Others**
會造成 'will cause' 會出現 'will appear'	，所以 ', so' 是因為 'be because'	，其實 ', actually' ，甚至 ', even ，通常 ', usually'	常見的 'common' ，像是 ', such as' 來說， 'as for,' 這樣的 'such; this type of'

With regard to abstract nouns, in agreement with Hoey (2001) and Flowerdew (2003, 2008), the Problem move tended to be marked by nouns that denoted or implied a problem, such as 問題 'problem' and 風險 'risk', and nouns that indicated a Reason-Result relationship, such as 原因 'reason', as illustrated in (2) and (3), respectively. As mentioned earlier, these abstract nouns were mostly preceded by the functional morpheme 的 in the trigram sequences, which linked the abstract nouns to their co-referring components. In addition, nouns that referred to a circumstance, such as 狀況 'situation' and 時候 'time', were also found to be frequent trigram signals for the Problem move, as in examples (4) and (5), respectively.

However, it is interesting to note that while 狀況 'situation' denoted the meaning of the situation or circumstance, it was preceded by the description of the problem, as illustrated in (4) below. In other words, in contrast to 時候 'time' in (5), which was used to establish the context for the following discourse, 狀況 'situation' in fact more often functioned as a shell noun (Schmid, 2000) or an evaluation carrier (Mahlberg, 2005) that marked a problematic situation.

(2) Example of 的問題 'problem of'

眼科醫師最擔心除了近視外，更擔憂高度近視、兩眼視差大的問題。

'In addition to myopia, ophthalmologists are more concerned about the problem of high myopia and a significant difference in vision between eyes.'

(3) Example of 的原因 'reason of'

而造成這種可怕情況的原因，就是「快樂缺氧」，又稱為「隱形缺氧」。

'What causes such a horrifying situation to happen is "happy hypoxia," also known as "silent hypoxia".'

(4) Example of 的狀況 'situation of'

雖然這些人可能還沒有出現器官實質上的損害，但其實已經出現功能上變差的狀況。

'Although the organs of these people have not displayed physical damage, they actually have shown functional decline.'

(5) Example of 的時候 'time of'

不過有些人刷牙的時候都沒有流血，但就是牙齦腫腫的，壓到也會有點痛，搞不清楚自己現在到底是發生了什麼問題。

'However, some people do not bleed when brushing their teeth. It's just that their gums are swollen and would hurt when pressed. They aren't really clear what has happened to them exactly.'

As for the medical terms that served as signals for the Problem component, many of them were related to physicians, such as 科醫師 'doctor from the division of' and 主治醫(師) 'attending physician', as shown in (6) and (7), respectively. This may have been partly because the journalists often quoted the opinions of practicing physicians as trustworthy sources in the Problem move, as illustrated in the following extracts. These doctor-related trigram expressions thus appeared to be not only signals for the Problem element but also indicators of the

journalistic genre and the medical topic.

(6) Example of 科醫師 'doctor from the division of'

精神科醫師指出：「一般人看了天災人禍感到焦慮、不安、擔憂，也會出現『急性壓力障礙』」。

'A psychiatrist points out that when witnessing disasters and accidents, people mostly would feel anxious, disturbed and worried and would develop acute stress disorder.'

(7) Example of 主治醫(師) 'attending physician'

長安神經內科醫療中心主治醫師陳惠萱表示，很多人常常會有「慢性疲勞」的問題。

'Huixuan Chen, attending physician at the Neurology Therapy Center of Everan Hospital, notes that many people would have the problem of chronic fatigue.'

Another group of medical terms that was often found in this move included nouns that indicated a problem, such as 症狀 'symptom' and 的疾病 'disease of', as shown in (8) and (9), respectively, below. These terms were often utilized to further explain the problem in question, as illustrated in the following examples. This group of frequent trigram expressions demonstrated that in addition to more general abstract nouns, such as problem and risk, nouns that were more topic specific yet less explicitly related to the concept of a problem also served similar functions. This highlighted the importance of examining and analyzing a rhetorical structure like the PS pattern in the context of a specific genre.

(8) Example of 症狀， 'symptom,'

「腳痛」是很多人常常會有的症狀，可能是因為肌肉拉傷、抽筋，或是筋膜炎等原因導致。

'"Foot pain" is a common symptom that many people may experience, which may happen due to muscle strains, cramps, fasciitis, or other causes.'

(9) Example of 的疾病 'disease of'

失眠這個問題，李信謙解釋：「它是一種『慢性狀況、急性惡化』的疾病。

'Regarding the problem of insomnia, Xinqian Li explains that it is a chronic disease that may deteriorate acutely.'

In addition to nominal sequences, many of the verb-based trigram expressions were also more likely to be found in the Problem move. One type of verb that frequently occurred in this move was reporting verbs, such as 說 'say' and 表示 'indicate'. Similar to and along with physician-related terms, reporting verbs were often used by the journalists of health news to cite an authoritative source in the Problem element, as exemplified in (10) below. One of the trigram signals contained both the noun 醫師 'physician' and the reporting verb 說 'say', as exemplified in (11) below. Again, this pattern showcased the characteristics of the health news genre, in addition to being a signal for the Problem move.

(10) Example of 表示， 'indicate,'

中國醫藥大學新竹附設醫院婦產科醫師許馨予表示，許多人對於哺乳議題有迷思。

'Xinyu Hsu, obstetrician at China Medical University Hsinchu Hospital, notes that many people have myths about breastfeeding.'

(11) Example of 醫師說 'doctor says'

蔡醫師說，多年前因對此疾病認知不足，泛視神經脊髓炎常被診斷成多發性硬化症。

'Dr. Tsai says that many years ago, due to the lack of knowledge about this disease, neuromyelitis opticaspectrum disorder was often diagnosed as multiple sclerosis.'

Other types of verbs that tended to occur in the Problem move included modals such as 可能 'may' and 會 'will', causal verbs such as 造成 'cause', and change-of-state verbs such as 出現 'appear', as illustrated in (12), (13), and (14), respectively, below. In the data, all of these verbs were mostly utilized to introduce the problem into the discourse, as illustrated in the following examples. These frequent verbs and the trigram sequences in which they occurred displayed a strong negative semantic prosody (Sinclair, 1991), while the verbs themselves did not appear to be semantically negative (cf. Stubbs, 1995; Tao, 2003). Also noteworthy is the prevalence of the epistemic modal 可能 'may', which helped express uncertainty in the medical discourse (cf. Lin, 2019) and, as the data showed, in the Problem move of health news.

(12) Example of 可能會 'possibly will'

但隨著年紀增長，或是不當使用膝蓋，膝關節可能會提早退化。

'But as one ages or uses their knees incorrectly, their knee joints may possibly start to deteriorate at an earlier age.'

(13) Example of 會造成 'will cause'

再加上精緻飲食、高糖高油的飲食、高血壓高血脂等慢性疾病的影響，不只讓大腦變得遲鈍，甚至可能會造成「失智症」。

'Along with the impact of refined foods, high-sugar and high-fat diets and chronic diseases such as hypertension and high cholesterol, it may not only cloud the brain but also lead to "dementia".'

(14) Example of 會出現 'will appear'

皮膚科醫師提醒，使用 3C 產品除了要擔心用眼過度、近視外，也可能會出現皮膚症狀！

'A dermatologist warns that in addition to overworked eyes and myopia, using electronic devices may also lead to the emergence of skin conditions!'

In agreement with the findings of previous studies, connectives that indicated the Reason-Result relationship, such as 所以 'so' and 因為 'because', as in (15) and (16), respectively, and adverbials of contrast, such as 其實 'actually' and 甚至 'even', were found to be signals for the Problem element (Flowerdew, 2008; Charles, 2011). The two adverbials exemplified in (17) and (18), respectively, allowed the writer to direct the reader's attention to the "real problem" or the potential harm that may be caused by the problem in question. In contrast, less often reported in the literature and yet found in the frequent trigram signals for the Problem move was the adverbial 通常 'usually'. As illustrated in (19), the trigram sequence served as a hedge for the following clause. Similar to the modal verb 可能 'may', which was discussed earlier, the high frequency of the adverbial 通常 'usually' may have also reflected the uncertainty inherent in medical science and discourse (Lin, 2019).

(15) Example of ，所以 ', so'

而有些人在一開始只是「偶發頭痛」，常容易被忽略，所以會有 2% 左右的人演變成慢性頭痛。

'And at the beginning, some people only have "occasional headaches" and thus overlook them, so approximately 2% of the people would develop chronic headache.'

(16) Example of 是因為 'be because'

《歐洲心臟病學雜誌》指出有四分之一的心臟病發作是因為高血壓所引起。

'European Journal of Cardiology points out that a quarter of heart attacks were induced

due to hypertension.'

(17) Example of adverbial ，其實 ', actually'
林明秀也提醒，其實會產生疤痕，並不是傷口大小來決定的，而是傷口的深度及嚴重度。
'Mingxiu Lin also emphasizes that the formation of scars in fact depends not on the size of the wound but on its depth and severity.'

(18) Example of ，甚至 ', even'
然而，中軸性脊椎關節炎與運動過度的狀況不同，更無法透過休息改善，甚至越久不動會越嚴重。
'However, axial spondyloarthritis is different from overexercise. It cannot be improved by rest and may even be worsened by being sedentary.'

(19) Example of ，通常 ', usually'
另外，通常有狐臭的人，還容易伴隨有多汗症。
'In addition, usually, people with armpit odor tend to have hyperhidrosis as well.'

Finally, the Problem move was also signaled by less common types of trigram expressions, such as the noun modifiers 常見的 'commonly seen' and 這樣的 'such; of this kind', and topic introducers, such as 像是 'such as' and 來說 'as for; with respect to'. Despite the variety in meaning and parts of speech, these trigram signals served as discourse management functions, such as topicalizing, as in the examples of 常見的 'commonly seen' in (20) and 來說 'as for' in (21), introducing examples, as in the example of 像是 'such as' in (22), and referring, as in the example of 這樣的 'such; of this kind' in (23):

(20) Example of 常見的 'commonly seen'
濕疹最常見的是搔癢與起疹子。
'The most common (symptoms) of eczema are itches and rashes.'

(21) Example of 來說， 'as for,'

但對於而耳鼻喉科的醫師來說，雖然喉嚨卡卡是一個小症狀，卻是很多疾病的共同警訊。

'But for ENT doctors, although a minor symptom, a sore throat is the common warning sign of several diseases.'

(22) Example of ，像是 ', such as'

靜脈曲張常發生在久坐久站的職業類別，像是廚師、櫃姐、老師、保全等常受此困擾。

'Individuals with jobs that require them to sit or stand for a long time often develop varicose veins. Cooks, salesclerks, teachers, and security guards, for example, are often bothered by this condition.'

(23) Example of 這樣的 'such; of this kind'

神經內科醫師解釋，這樣的情況容易發生在中年或已停經的女性。

'A neurologist explains that such conditions are often found in middle-aged or menopausal women.'

With respect to the functional distribution of trigrams in the Problem move, as can be seen in Table 5 below, most of the trigrams were used to serve the discourse organizing and referring functions, and only a couple of modal verb (e.g., 可能會 'probably will') and adverb (e.g., ，其實 ', actually') trigrams were used to serve stance functions. This may have been partly due to the written nature of the online health news articles (cf. Biber *et al*., 2004), which is in correspondence with the findings of previous studies on lexical bundles in Chinese journalistic texts (Hsu, 2021). Another distinguishable feature that was observed in this list was that several trigrams in the discourse organizing category were employed to mark the information source, a genre-specific characteristic discussed earlier (cf. Biber & Gray, 2013). A number of trigrams were employed to introduce a topic for later discussion, such as 的原因 'reason of', 會造成 'will cause', and 來說， 'as for,'. Moreover, trigrams containing nouns that were semantically associated with the Problem move, such as 問題 'problem' and 風險 'risk', that identified the focus of the article or move mostly fell under the referring category. Trigrams such as 臨床上 'clinically' and 這樣的 'such; this type of' were used to specify the attribute of a topic, while trigrams such as ，通常 ', usually' and 的時候 'time of' referred to a time-related concept (e.g., frequency) or point in the texts.

Table 5. Functional categories of the trigram signals for the Problem move

Types	Examples
I. STANCE	
A. Epistemic stance	可能會 'probably will' ，可能 ', probably'
B. Attitudinal stance	，其實 ', actually' ，甚至 ', even'
II. DISCOURSE ORGANIZING	
A. Information source	說：「 'say' 醫師說 'doctor says' 表示， 'indicate,' 科醫師 'doctor of the Division of' 主治醫 'attending physician' 治醫師 'attending physician' 科主治 'attending physician at the Division of'
B. Topic management	的原因 'reason of' 會造成 'will cause' 會出現 'will appear' ，像是 ', such as' 來說， 'as for,'
C. Discourse connection	，所以 ', so' 是因為 'be because'
III. REFERRING	
A. Identification/focus	的問題 'problem of' 問題。 'problem.' 的風險 'risk of' 問題， 'problem,' 的狀況， 'situation of,' 症狀， 'symptom,' 疾病， 'disease,' 的疾病 'disease of'
B. Specification of attributes	臨床上 'clinically' 這樣的 'such; this type of'
C. Time reference	，通常 ', usually' 的時候 'time of'

4.3.2 Signals for the Response Move

Despite the fact that more instances of the Response move were found in the corpus, fewer types of trigram signals were retrieved compared with those for the Problem move. The more frequent trigram signals for the Response move included those that involved abstract nouns (e.g., 情況 'situation' and 方式 'method'), medical terms (e.g., 傷口 'wound' and 治療 'cure; treatment'), modal verbs (e.g., 可以 'can' and 有可能 'having the possibility; probably'), connectives (e.g., 如果 'if', 因為 'because', and 或是 'or'), and others (e.g., 越來越 'more and more'), as displayed in Table 6:

Table 6. Structural categories of the trigram signals for the Response move

Abstract Nouns	**Medical Terms**	**Modal Verbs**
的情況 'situation of' 的方式 'method of'	的傷口 'wound of' 的治療 'treatment of'	，可以 ', can' 就可以 'just can' 有可能 'having the possibility; probably'
Connectives	**Others**	
，如果 ', if' ，因為 ', because' ，或是 ', or'	越來越 'more and more'	

Unlike the Problem move, in which a Chinese counterpart for the abstract noun problem was found to be one of the prevalent signals, the Response move was not signaled by trigram expressions that involved nouns that denoted a solution in Chinese. This was partly due to the fact that the nominal counterpart of 解法 or 解方 'solution', for example, is not as commonly used in Chinese. Instead, nouns that were less explicitly linked to the function of the move, such as 方式 'method; approach', were involved in the sequences, as in (24). On the other hand, circumstantial abstract nouns, such as 情況 'situation', were also found to be frequent in this move. However, as illustrated in (25) below, despite being a near-synonym of 狀況 'situation' discussed earlier, 情況 'situation' was not used to present a problem; instead, it was deployed to mark the context in which the recommended solution should be implemented or would be appropriate.

(24) Example of 的方式 'method of'

如果傷口略大，可以透過輕壓的方式幫助止血。

'If the wound is fairly big, you can try (the method of) pressing lightly on the wound to stop bleeding.'

(25) Example of 的情況 'situation of'

而若遇到疫情較嚴重的情況下，醫師仍建議患者應規律就醫回診、用藥。

'Even if the situation of the pandemic becomes more serious, physicians still recommend that patients should visit the doctor and take medication regularly.'

With respect to medical terms, in stark contrast to the Problem move, no human-related nouns were found to serve as signals for the Response move. This may have been in part because the Response move was typically subsequent to the Problem move, in which the main source of information was mentioned the first time and thus with the full title listed, and as a result, the source was usually referred to solely by name in the Response move, such as the more frequent health- and medicine-related 'cure; treatment' in (26), and nouns that referred to things that could or should be treated in particular ways, such as 傷口 'wound' as exemplified in (27) below:

(26) Example of 的治療 'treatment of'

除了靠藥物的治療外，建議也需規律運動。

'In addition to medical treatment, regular exercise is also recommended.'

(27) Example of 的傷口 'wound of'

這種類型傷口可能需要縫合，尤其是傷及真皮層的傷口更要小心護理，應該就醫評估，是否要進一步治療。

'This type of wound may require sutures. Wounds that damage the dermis in particular should be treated with extra care. One should consult a doctor to assess if further treatment is needed.'

Similar to the Problem move, the Response move was marked by trigram sequences that involved modal verbs. However, rather than being signaled by epistemic modals such as 可能 'may' and 會 'will', the Response move more often featured the modal verb 可以 'can', as illustrated in (28) and (29) below. As pointed out by researchers such as Wang (2018), in addition to its "ability" reading, 可以 'can' can also be used to present a suggestion or recommendation, as illustrated in (28) and (29), which explains why this modal verb was a frequent signal for the Response move whose communicative function was to recommend a way to solve the problem in question.

(28) Example of ，可以 ‘, can’

所以，如果想選用抗沾黏產品，可以注意幾個要點。

‘Therefore, you can pay attention to a few key points when choosing anti-adhesion products.’

(29) Example of 就可以 ‘then can’

這時候就可以透過現在很流行的筋膜槍，來讓我們身上的肌肉放鬆。

‘Then we can use the now very popular fascia gun to relax our muscles.’

It should be noted that although 可能 ‘possibility’ was also found in one of the trigram signals for the Response move, it was different from the use of 可能 ‘may’ in the Problem move. First, it was used as a noun-like item in the Response move because it was preceded by the verb 有 ‘to have’. Second, in lieu of identifying the key component of the move (i.e., the solution), the trigram sequence 有可能 ‘having the possibility; probably’ presented the contextual information that supported the construction of the Response move, as exemplified in (30) below:

(30) Example of 有可能 ‘having the possibility; probably’

孩子大一點，半夜醒來，有可能是因為做夢或白天刺激過多造成，並不一定是沒吃飽。

‘As the child grows older, if they wake up at midnight, it may be probably due to dreaming or overstimulation during the daytime, not necessarily due to not having enough food.’

As for connectives that signaled the Response move, the markers of reason, such as 因為 ‘because’, were found to have achieved this function. However, more interestingly, connectives that were less often reported in the literature, such as the conditional marker 如果 ‘if’ and the disjunctive marker 或是 ‘or’, were frequent in the Response element in the corpus. The high frequency of these two connectives may have been due to the fact that the Response part of a medical news article involved the act of giving advice or making suggestions. As shown in (31) and (32) below, these two markers, respectively, often introduced the solution or response to the problem in question. This pattern, albeit distinct from Flowerdew’s (2008) findings, corresponded with the observations in previous research, that conditionals and expressions indicating alternativeness were often employed to support the advice-giving act in the discourse (Hsieh, 2019; Lin, 2019).

(31) Example of ，如果 ‘, if’

但在夏天或是劇烈運動後，如果臉部感到很油膩則可以再多洗一次。

‘But in summer or after strenuous exercise, if your face still feels greasy, then you can wash it one more time.’

(32) Example of ，或是 ‘, or’

民眾日常飲食建議多攝取富含維生素 B 群、C、D 的蔬果，如：蘋果、酪梨、奇異果等，或是 含有豐富礦物質的深綠色蔬菜。

‘People are recommended to consume more vegetables and fruits that are rich in Vitamins B, C, and D, such as apples, avocados, and kiwis, or leafy greens, which contain abundant minerals.’

Finally, the comparative modifier 越來越 ‘more and more’, similar to the abstract noun 情況 ‘situation’ and the conditional marker 如果 ‘if’ discussed earlier, also functioned to present the condition or context for the solution presented, as illustrated in (33) below:

(33) Example of 越來越 ‘more and more’

如果針眼越來越大顆，還是直接去看眼科醫生最為有用。

‘If the sty gets bigger and bigger, it is still better to just go to an ophthalmologist.’

Compared with the Problem move, the Response move only showed a slight preference for the referring function, with much less diverse discourse organizing trigrams, as illustrated in Table 7 below. Stance functions in this move were served largely by epistemic modal verb trigrams, such as 就可以 ‘just can’ and 有可能 ‘having the possibility; probably’, whereas discourse organizing functions were achieved by trigrams containing grammatical connectives, such as 如果 ‘, if’, 因為 ‘, because’, and 或是 ‘, or’. Similar to the pattern discussed in the previous section, noun-based trigrams that more explicitly indicated the Response move, such as 的方式 ‘method of’, 的傷口 ‘wound of’, and 的治療 ‘treatment of’, tended to be used to serve the referential function, suggesting the importance of referring trigrams in signaling the moves in the PS discourse structure in this genre. Finally, as mentioned earlier, while 狀況 ‘situation’ and 情況 ‘situation’ appeared to be near-synonymous, they in fact played distinctive roles in this genre of text. In the Problem move, trigrams that contained 狀況 ‘situation’ functioned to identify the focus of the move (i.e., the problem). In contrast, trigrams that involved 情況 ‘situation’ only referred to contextual information, such as the

time and condition, in the Response move. This showed the importance of a more qualitative, function-oriented analysis of the results.

Table 7. Functional categories of the trigram signals for the Response move

Types	Examples
I. STANCE	
A. Epistemic stance	，可以 ‘, can’ 就可以 ‘just can’ 有可能 ‘having the possibility; probably’
II. DISCOURSE ORGANIZING	
A. Discourse connection	，如果 ‘, if’ ，因為 ‘, because’ ，或是 ‘, or’
III. REFERRING	
A. Identification/focus	的方式 ‘method of’ 的傷口 ‘wound of’ 的治療 ‘treatment of’
B. Specification of attributes	越來越 ‘more and more’
C. Time reference	的情況 ‘situation of’

5. Conclusion

As one of the first projects that examined the Problem-Solution pattern in Chinese discourse, the current study constructed a small corpus of online health news articles annotated with elements of the PS textual pattern. An n-gram approach was then adopted in an attempt to search for the linguistic signals in this genre for each move in the PS pattern, and the findings were fruitful. The results showed that high-frequency trigram expressions retrieved from the corpus displayed some of the same characteristics reported in prior studies, such as indicating the Reason-Result relationship, marking contrast, and signaling the key components of the PS pattern, including Problem and Solution. On the other hand, features that have been mentioned in the literature, such as quotative devices and medically relevant terms, were also frequently used in the recurrent trigram expressions, and these features helped achieve functions that were highly relevant to the genre of online health news.

A closer look at the Problem and Response components in the online health news articles revealed that each of the key moves in the PS pattern preferred particular trigram sequences. For example, the Problem move in the dataset was often signaled by abstract nouns explicitly

or implicitly linked to the notion of problems and by devices for citing an authoritative source, such as a reporting verb and the noun 主治醫(師) 'attending physician' in Chinese. This move also tended to be marked by sequences that contained modals that conveyed prediction and uncertainty or verbs that indicated a (usually negative) change of state with negative semantic prosody. Regarding the functions that the trigrams served, the signals in the Problem move tended to fall under the discourse organizing and referring categories. A number of the expressions were used to cite the information source or refer to move-related topics, such as problems, risks, and diseases.

On the other hand, trigram signals for the Response move often involved linguistic features that enabled the writer to present suggestions or advice, such as dynamic verbs, conditionals, and disjunctive markers. Moreover, as the counterpart of the noun "solution" in Chinese is not frequently used, the abstract nouns found in the trigram signals were less explicitly related to "solution," which displayed a language-specific characteristic. In contrast to the Problem move, the Response move displayed only a mild preference for referring trigrams. The stance and discourse organizing trigrams in the Response move also served less diverse functions. However, similar to the Problem move, the referring trigrams in the Response move played an important role in signaling the rhetorical structure.

Given the findings summarized above, the present study has a number of implications. First, while most of the previous studies on the PS pattern adopted the methods of keyword analysis to investigate the linguistic cues of the move structure, the current research demonstrated that the n-gram or multiword approach is a potential alternative framework for identifying lexicogrammatical signals for particular rhetorical functions and components, as this approach enabled us to examine linguistic resources beyond the constraints of traditional, prescriptive definitions of linguistic units (Cortes, 2013; Tian *et al*., 2020).

Moreover, the findings presented in this article also exemplified the interaction between a rhetorical structure, such as the PS pattern, and a journalistic genre, in the case of this study, online health news articles written in Mandarin Chinese. As observed in the trigram signals retrieved from the corpus, while a popular rhetorical structure, especially in writing, the PS pattern featured linguistic markers that were characteristic of the genre for which the pattern was specifically utilized. The results also suggested that a few of the move or component signals were formulated to achieve communicative functions that were prominent or specific to the genre, such as citing a source or giving advice. These communicative acts in turn helped the writer to achieve the rhetorical function(s) of each move, such as identifying the problem and presenting the solution (cf. Belmonte, 2009; Ali, 2013). This points to the importance of putting the PS pattern in the context of a particular genre and of including more types of text in research on the PS structure.

On the basis of the current research, a number of directions for future research were

identified. First, a larger number of texts of the same (or different) genre(s) from different sources should be collected to see whether and to what extent the conclusions drawn in this article still apply and whether trigram expressions are the best unit of analysis. Second, in addition to the analysis of the relationship between frequent trigrams and the components of the Problem-Solution pattern, future studies should look into the distribution of communicative acts in each move of this genre and the mapping between communicative acts and n-gram sequences, as a few previous studies have examined (Belmonte, 2009; Ali, 2013). Lastly and probably most excitingly, the discourse data annotated for and the concluding findings in the current research project will serve as the foundation for training models for the automatic detection and annotation of the PS pattern and for developing other related applications, such as chatbots for medical purposes and algorithms that translate or even produce journalistic texts on health-related topics.

References

Ali, A. M. (2013). Combining problem-solution categories and communicative acts: An analysis of Malaysian and British business journalistic texts. *World Applied Sciences Journal*, *21*,174-185. https://doi.org/10.5829/idosi.wasj.2013.21.sltl.2152

Belmonte, I. A. (2009). Toward a genre-based characterization of the problem–solution textual pattern in English newspaper editorials and op-eds. *Text & Talk*, *29*(4), 393-414. https://doi.org/10.1515/TEXT.2009.021

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, *25*(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Biber, D., & Gray, B. (2013). *Discourse characteristics of writing and speaking task types on the TOEFL IBT® test: A lexico-grammatical analysis* (RR-13-04, TOEFLiBT-19). ETS Research Report Series.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135-146. https://doi.org/10.1162/tacl_a_00051

Cavnar, W. B., & Trenkle, J. M.. (1994). N-gram-based text categorization. In *Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval*, Vol. 161175.

Charles, M. (2011). Adverbials of result: Phraseology and functions in the problem–solution pattern. *Journal of English for Academic Purposes*, *10*(1), 47-60. https://doi.org/10.1016/j.jeap.2011.01.002

Cortes, V. (2013). The purpose of this study is to: Connecting lexical bundles and moves in research article introductions. *Journal of English for Academic Purposes*, *12*(1), 33-43. https://doi.org/10.1016/j.jeap.2012.11.002

De Coninck, D., d'Haenens, L., & Matthijs, K. (2020). Forgotten key players in public health: News media as agents of information and persuasion during the COVID-19 pandemic. *Public Health*, 183,65-66. https://doi.org/10.1016/j.puhe.2020.05.011

Escalante, H. J., Solorio, T., & Montes-y-Gómez, M. (2011). Local histograms of character n-grams for authorship attribution. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, 288-298.

Flowerdew, L. (2003). A combined corpus and systemic-functional analysis of the problem-solution pattern in a student and professional corpus of technical writing. *TESOL Quarterly*, 37(3), 489-511. https://doi.org/10.2307/3588401

Flowerdew, L. (2008). *Corpus-based analyses of the problem-solution pattern: A phraseological approach*. John Benjamins.

Fox, S., & Duggan, M. (2013). *Health online 2013*. PEW RESEARCH CENTER

Fu, J., Liu, P., Zhang, Q., & Huang, X.-J. (2020). Is Chinese word segmentation a solved task? Rethinking neural Chinese word segmentation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 5676-5686. https://doi.org/10.18653/v1/2020.emnlp-main.457

Ghenai, A., & Mejova, Y. (2018). Fake cures: User-centric modeling of health misinformation in social media. *Proceedings of the ACM on Human-Computer Interaction*, *2*(CSCW), 1-20. https://doi.org/10.1145/3274327

Giannakopoulos, G., & Karkaletsis, V. (2009). N-gram graphs: Representing documents and document sets in summary system evaluation. In *Proceedings of Text Analysis Conference TAC 2009*.

Gries, S. T. (2008). Dispersions and adjusted frequencies in corpora. *International Journa l of Corpus Linguistics*, *13*(4), 403-437. https://doi.org/10.1075/ijcl.13.4.02gri

Handford, M., & Matous, P. (2015). Problem-solving discourse on an international construction site: Patterns and practices. *English for Specific Purposes*, *38*, 85-98. https://doi.org/10.1016/j.esp.2014.12.002

Heffernan, K., & Teufel, S. (2018). Identifying problems and solutions in scientific text. *Scientometrics*, *116*(2), 1367-1382. https://doi.org/10.1007/s11192-018-2718-6

Hoey, M. (1983). *On the surface of discourse*. Allen & Unwin.

Hoey, M. (2001). *Textual interaction: An introduction to written discourse analysis*. Psychology Press.

Hsieh, C.-L. (2005). Modal verbs and modal adverbs in Chinese: An investigation into the semantic source. *UST Working Papers in Linguistics*, *1*, 31-58.

Hsieh, C.-Y. C. (2019). *Language, intersubjectivity, and institutional interaction: Advice-giving directives in Taiwan EFL writing tutorials* (Doctoral dissertation). National Taiwan University.

Hsieh, C.-Y. C. (2020). Meaning in repair: The abstract noun yisi 'meaning/intention' in the management of intersubjectivity in Mandarin conversation. *Taiwan Journal of Linguistics*, *18*(2), 39-88. https://doi.org/10.6519/TJL.202007_18(2).0002

Hsu, C.-C. (2021). The structure and function of lexical bundles in Chinese conversation and news. *Taiwan Journal of Chinese as a Second Language*, *22*, 69-96. https://doi.org/10.29748/TJCSL.202106_(22).0003

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, *27*(1), 4-21. https://doi.org/10.1016/j.esp.2007.06.001

Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, *48*(4), 63-85. https://doi.org/10.1007/BF02300500

Kachru, B. B. (1992). *The other tongue: English across cultures*. University of Illinois Press.

Kanaris, I., Kanaris, K., Houvardas, I., & Stamatatos, E. (2007). Words versus character n-grams for anti-spam filtering. *International Journal on Artificial Intelligence Tools*, *16*(06), 1047-1067. https://doi.org/10.1142/S0218213007003692

Kumari, S., Reddy, H. K., Kulkarni, C. S., & Gowthami, V. (2021). Debunking health fake news with domain specific pre-trained model. *Global Transitions Proceedings*, *2*(2), 267-272. https://doi.org/10.1016/j.gltp.2021.08.038

Kuta, M., & Kitowski, J. (2014). Optimisation of character n-gram profiles method for intrinsic plagiarism detection. In *Proceedings of the 13th International Conference on Artificial Intelligence and Soft Computing (ICAISC 2014)*, 500-511. https://doi.org/10.1007/978-3-319-07176-3_44

Li, L., Franken, M., & Wu, S. (2020). Bundle-driven move analysis: Sentence initial lexical bundles in PhD abstracts. *English for Specific Purposes*, *60*, 85-97. https://doi.org/10.1016/j.esp.2020.04.006

Lin, W.-H. (2019). Expressing uncertainty with conditionals in medical discourse: A comparison across genres. In: Tao, H., Chen, HJ. (eds) *Chinese for Specific and Professional Purposes. Chinese Language Learning Sciences*. Springer. https://doi.org/10.1007/978-981-13-9505-5_10 213–243.

Liu, M.-C., & Chiang, T.-Y. (2008). The construction of Mandarin VerbNet: A frame-based study of statement verbs. *Language and Linguistics*, *9*(2), 239-270.

Mahlberg, M. (2005). *English general nouns: A corpus theoretical approach*. John Benjamins.

Ng, H. T., & Low, J. K. (2004). Chinese part-of-speech tagging: One-at-a-time or all-at-once? Word-based or character-based? In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 277-284.

Ratanakul, S. (2017). A study of problem-solution discourse: Examining TED talks through the lens of move analysis. *LEARN Journal: Language Education Acquisition Research Network*, *10*(2), 25-46.

Ratanakul, S. (2018). A move analysis of problem-solution discourse: A pedagogical guide for opinion and academic writing. *Arab World English Journal*, *9*(3), 233-247. https://doi.org/10.24093/awej/vol9no3.16

Schmid, H.-J. (2000). *English abstract nouns as conceptual shells: From corpus to cognition*. De Gruyter Mouton.

Scott, M. (2000). Mapping key words to problem and solution. In M. Scott & G. Thompson (Eds.), *Patterns of Text: In honour of Michael Hoey* (pp. 109-127). John Benjamins.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Staples, S., Egbert, J., Biber, D., & McClair, A. (2013). Formulaic sequences and eap writing development: Lexical bundles in the TOFEL iBT writing section. *Journal of English for Academic Purposes*, *12*(3), 214-225. https://doi.org/10.1016/j.jeap.2013.05.002

Stubbs, M. (1995). Collocations and semantic profiles: On the cause of the trouble with quantitative studies. *Functions of Language*, *2*(1), 23-55. https://doi.org/10.1075/fol.2.1.03stu

Tao, H. (2003). Toward an emergent view of lexical semantics. *Language and Linguistics*, *4*(4), 837-856.

Tian, Y., Song, Y., Ao, X., Xia, F., Quan, X., Zhang, T., & Wang, Y. (2020). Joint Chinese word segmentation and part-of-speech tagging via two-way attentions of auto-analyzed knowledge. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 8286-8296. https://doi.org/10.18653/v1/2020.acl-main.735

Tsirintani, M. (2021). Fake news and disinformation in health care-challenges and technology tools. In *Public health and informatics (Proceedings of MIE 2021)*, 318-321. IOS Press.

Wang, P. Y. (2018). *A cognitive-pragmatic study on modal verbs of possibility in Chinese* (Doctoral dissertation). The Pennsylvania State University.

Waszak, P. M., Kasprzycka-Waszak, W., & Kubanek, A. (2018). The spread of medical fake news in social media—The pilot quantitative study. *Health Policy and Technology*, *7*(2), 115-118. https://doi.org/10.1016/j.hlpt.2018.03.002

Waugh, L. R. (1995). Reported speech in journalistic discourse: The relation of function and text. *Text & Talk*, *15*(1),129-73. https://doi.org/10.1515/text.1.1995.15.1.129

Wieting, J., Bansal, M., Gimpel, K., & Livescu, K. (2016). Charagram: Embedding words and sentences via character n-grams. In arXiv preprint arXiv:1607.02789

Yue, M. (2006). Discursive usage of six Chinese punctuation marks. In *Proceedings of the COLING/ACL 2006 Student Research Workshop*, 43-48.

李松濤、許文怡 (2020)。科學傳播歷程中程序性知識特徵的框架探究-以飲食保健類科學研究新聞為例。**科學教育學刊**，**28**(2)，143－168。https://doi.org/10.6173/CJSE.202006_28(2).0003。[Lee, S. T. & Hsu, W.-Y. (2020). A Framing Exploration of Procedural Knowledge Characteristics in Science Communication Process: Examples of Research Based Science News Regarding Healthy Diet. *Chinese Journal of Science Education*, *28*(2), 143-168.]

李松濤、鄔啟柔 (2017)。科學新聞傳播內容與模式之探究以飲食,疾病與醫藥類新聞為例。**科學傳播論文集 8**，17－30。 https://doi.org/10.6930/scicomm.201705.0002。[Lee, S. T. & Wu, C. R. (2017). An exploration on the content and model of science news communication: A case study of news stories related to food, disease, and medicine. *Collected Papers on Science Communication 8*, 17-30.]

葉蓉慧、黃暐超 (2017)。社群網路時代民眾對新聞回應:以臉書民眾對台灣流感議題回應為例。**科學傳播論文集 8**，103－113。https://doi.org/10.6930/scicomm.201705.0006。[Yeh, J. H. B. & Huang, W. C. (2017). Readers' responses to news in the social media era: A case study of Facebook users' responses to flu-related issues in Taiwan. *Collected Papers on Science Communication 8*, 103-113.]

Appendix A. Statistics of the 58 trigrams with a Min-Max scaling value larger than 0.3

Trigrams	Frequency	Relative Frequency	Min-Max Scaling
(，如果)	113	0.000895724	1.0000
(說：「)	85	0.000673774	0.7500
(，所以)	80	0.000634141	0.7054
(，因為)	75	0.000594507	0.6607
(科醫師)	74	0.00058658	0.6518
(可以)	74	0.00058658	0.6518
(症狀，)	71	0.0005628	0.6250
(的時候)	70	0.000554873	0.6161
(攝護腺)	69	0.000546946	0.6071
(的問題)	69	0.000546946	0.6071
(，其實)	68	0.000539019	0.5982
(，或是)	68	0.000539019	0.5982
(表示，)	67	0.000531093	0.5893
(糖尿病)	64	0.000507312	0.5625
(問題，)	62	0.000491459	0.5446
(，甚至)	56	0.000443898	0.4911
(治療，)	55	0.000435972	0.4821
(可能會)	54	0.000428045	0.4732
(高血壓)	53	0.000420118	0.4643
(的患者)	53	0.000420118	0.4643
(主治醫)	52	0.000412191	0.4554
(治醫師)	52	0.000412191	0.4554
(就可以)	52	0.000412191	0.4554
(這樣的)	50	0.000396338	0.4375
(的症狀)	49	0.000388411	0.4286

(護腺癌)	48	0.000380484	0.4196
(是因為)	48	0.000380484	0.4196
(，可能)	47	0.000372558	0.4107
(有可能)	47	0.000372558	0.4107
(的原因)	46	0.000364631	0.4018
(，因此)	46	0.000364631	0.4018
(疾病，)	44	0.000348777	0.3839
(膝關節)	44	0.000348777	0.3839
(的狀況)	43	0.000340851	0.3750
(性皮膚)	43	0.000340851	0.3750
(的情況)	43	0.000340851	0.3750
(，通常)	43	0.000340851	0.3750
(皮膚炎)	42	0.000332924	0.3661
(的傷口)	41	0.000324997	0.3571
(疤產品)	40	0.00031707	0.3482
(醫師說)	40	0.00031707	0.3482
(問題。)	40	0.00031707	0.3482
(的治療)	40	0.00031707	0.3482
(會造成)	40	0.00031707	0.3482
(疤痕，)	40	0.00031707	0.3482
(釋：「)	39	0.000309144	0.3393
(解釋：)	39	0.000309144	0.3393
(越來越)	39	0.000309144	0.3393
(常見的)	37	0.00029329	0.3214
(的疾病)	37	0.00029329	0.3214
(的方式)	37	0.00029329	0.3214
(會出現)	36	0.000285363	0.3125
(疤痕的)	36	0.000285363	0.3125
(科主治)	36	0.000285363	0.3125

(臨床上)	36	0.000285363	0.3125
(，像是)	36	0.000285363	0.3125
(來說，)	35	0.000277436	0.3036
(的風險)	35	0.000277436	0.3036