

INFORMATION PACKAGING IN ENGLISH MEDICAL PAPERS PUBLISHED IN A CHINESE JOURNAL

Gao Wenyan

Zhuhai Campus , Beijing Institute of Technology

Abstract

This study attempts to investigate possible differences between English medical papers published in an international journal and a Chinese journal . Features explored include lexical density and nominalizations , both of which may affect the way information is packaged in academic English .

Key words

information packaging ; lexical density ; nominalization

1 . Information packaging in academic papers

Information packaging is an important research area in EAP studies . Among other things , the Systemic-Functional approach to language attempts to analyze lexical density and nominalization in written English as opposed to spoken English .

1 .1 Lexical density in academic writing

Among the many linguistic features which characterize English academic writing , the use of extensive nominal groups in which a sequence of modifiers is piled to the left of a head noun has been well noted in the literature (Galve 1998) . Such long nominal constructions produce a high lexical density , which Halliday & Martin (1993) define as " a measure of the density of information in any passage of text , according to how tightly the lexical items (content words) have been packed into the grammatical structure " . Halliday (1985) analyzed the functions of written English , and indicated that written English is characterized by a lexically higher density . Eggins (1994) went further by showing that spoken English is 33 % lexical , while 42 % of the words in written English are lexical . Guillen Galve (1998) also believes that the complexity of written language is a lexical complexity rather than a syntactic one because written language attains a high density of lexical items , which have a higher information content . The following examples (from Halliday 1985) show the difference in the density of lexical items between written and spoken English .

- 1) The use of this method of control unquestionably leads to safer and faster train running in the most adverse conditions . (written English)
- 2) You can control the trains this way and if you do that you can be quite sure that they'll be able to run more safely and more quickly than they would otherwise no matter how bad the weather gets . (spoken English)

As illustrated in examples above , in 1) (written English) , 55 % (11 words out of 20) of the words are content words , while in 2) (spoken English) , only 30 % (11 words out of 40) of the words are content words . The density of information in written English is often considerably higher than that in spoken English .

In medical writing, a typical written register, the density of information may be even higher, though nouns tend to predominate over adjectives in prenominal modifiers (Galve 1998).

- 3) Continuous intra arterial pressure monitoring allowed rapid correction of any hypotension occurring during institution of anaesthesia, or resulting from intraoperative blood loss.
- 4) limb flexion reflex withdrawal time

With a higher proportion of words in the text carrying content information, the amount of information that gets packed into such texts, according to Halliday (1985), is worth noticing, and this tendency to package as much matter as possible into the text is commonplace in medical English. Researchers in the scholarly field seem to have reached the consensus that high lexical density, which is often measured by dividing the number of content words (nouns, regular/main verbs, and most adjectives and adverbs) by the total number of words in the text (Eggins 1994), is characteristic of scientific English. The high content of such lexical items in the written text renders the text lexically complex but syntactically simple.

1.2 Nominalization in academic writing

Another phenomenon widely found in scientific English (medical English in particular) is nominalization, the turning of a verb or an adjective, with or without morphological transformation, into a noun, as in the following examples:

- 5) The police investigated the murder. the police's investigation of the murder
- 6) The difference is obvious. the obvious difference

Such substitution of one grammatical class, or one grammatical structure, by another, is referred to as grammatical metaphors in Systemic-Functional Grammar (Halliday & Martin 1993; Thompson 1996). Chafe & Danielewicz (1987) observed frequencies of nominalizations in different registers, and found that nominalizations occur most frequently in academic papers, 92 occurrences per 1000 words (9.2%) as compared to 27 occurrences per 1000 words (2.7%) in conversations, 56 occurrences per 1000 words (5.6%) in lectures, and 55 occurrences per 1000 words (5.5%) in letters.

Nominalizations serve a number of purposes in the written language (from Wang 2007). They may turn actions or processes (verbs) into concepts (nouns), with the result that the tone of writing sounds more abstract and more formal, as in 7b) below:

- 7) a) Water evaporates when it is heated.
- b) Evaporation occurs more quickly when water is heated.

Nominalizations may help pack more abstract ideas in a single sentence, as in 8b) (from Wang 2007)

- 8) a) Every day shops lose thousands of dollars worth of valuable items. And this affects us all because prices increase and we have to pay extra.
- b) The daily loss of thousands of dollars worth of valuable stock ultimately affects us all through an increase in prices.

With nominalizations, more information can be compressed into each nominal group, so the academic writer may concisely refer to recurring abstract ideas.

Bloor & Bloor (2000) summarize the SFL view of grammatical metaphor and conclude that there are three types of grammatical metaphors: experiential grammatical metaphor in which a process is turned into a noun (nominalization), logical grammatical metaphor in which ideas are organized in an incongruent form at the level of discourse, and interpersonal grammatical metaphor in which modal verbs are expressed by adverbs, adjectives, or nouns. Grammatical metaphor, according to Halliday & Matthiessen (2004), is a language resource to condense information by expressing concepts in an incongruent form valued in scientific and academic registers as a way of expressing objectification and abstraction. Nominalization is often thought of as "the single most powerful resource for creating grammatical metaphor", which probably "evolved first in scientific and technical registers" and then gradually spread to other areas of adult discourse and became a mark of prestige and power (Halliday &

Matthiessen 2004).

While a large number of discussions about the interrelationship between grammatical metaphor , lexical density and nominalization in written English can be found in the literature ,hardly any research has been done to reveal whether Chinese academic writers can make use of these means to achieve their desired goals when they write for English medical journals . By comparing medical research papers published in an internationally renowned English journal with medical research papers in an English medical journal published in China ,this study attempts to analyze whether Chinese medical writers of English can effectively employ necessary means to package an adequate amount of information in a concise manner as required by English for Academic Purposes (EAP).

2 . Methodology

2 .1 Research questions

This study attempts to answer two questions related to information packaging in Chinese academic writers English medical papers .

- 1) How is the lexical density of Chinese academic writers English medical papers as compared to that in medical papers published in internationally renowned journals ?
- 2) Can Chinese medical writers of English use an adequate proportion of nominalizations to achieve their information packaging goals ?

In answering the first question ,following Eggins (1994) ,lexical density will be calculated as

Lexical density = $C / T \times 100 \%$,where

C = total number of content words (nouns ,regular /main verbs ,and most adjectives and adverbs) ;
and

T = total number of words (tokens)

and the lexical density of Chinese academic writers English medical papers will be compared to that of the medical papers published in an internationally renowned English journal .

In answering the second question ,the number of nominalizations from verbs (nouns suffixed with -ment , -ion , -ure , -ing , a /ence and their plural forms)and adjectives (nouns suffixed with -ty , -ness , -ure and their plural forms)in both datasets will be calculated and compared .

2 .2 Data collection

Two datasets are analyzed in this study . The first set (hereafter called NNS for short) is comprised of the five medical research papers in Journal of Medical Colleges of PLA (2003 ,18(2)) and the five papers in Journal of Medical Colleges of PLA (2003 ,18(6)) , totalling to 10 papers in all . The second set (hereafter called NS for short) is comprised of the five medical research papers in Annals of Surgery (Vol 218 ,No .1) ,published in 1993 by J . B . Lippincott Company . Annals of Surgery is an international journal renowned in the medical field . Papers in both datasets were scanned with a scanner and processed with an OCR computer application . The output of the OCR application was then manually proofread twice to eliminate chances of scanning errors .

Basic statistics about the two datasets are given below .

Table 1 . Information about the datasets

	Number of papers	Number of words (tokens)
NNS	10	17 322
NS	5	15 368

It can be seen that the two datasets are not of the same size . Therefore ,before frequencies were compared in the analysis ,normalization was conducted . The normalized frequency in this study is given on the percentage basis .

2.3 Text processing

As the output of the OCR software was a number of files in the Microsoft Word format, which makes text analysis difficult, the 15 files were converted into the pure text format (*.txt) for further processing and analysis.

In order to retrieve the frequencies of content words and nominalizations in the texts, the texts were tagged with a POS (part of speech) tagger.

2.4 Data analysis

For the first research question, the lexical density for both datasets was calculated and converted to statistics on a percentage basis for better comparison. Meanwhile, raw frequencies of content words in both datasets were also analyzed with the Chi square test to see whether the difference between the two frequencies is statistically significant. During this process, the total size of the datasets (in terms of the total number of words) was also taken into consideration.

For the second research question, raw frequencies of nominalizations in both datasets were analyzed with the Chi square test. The total size of the datasets was also taken into consideration to guarantee comparability.

3. Results and discussion

3.1 Lexical density in NS and NNS medical research papers

After the number of content words in both datasets were retrieved and calculated, a significance test (Chi square) was conducted, producing the following statistics:

Table 2. Frequencies of content words and lexical density

NS		NNS		Chi square
Freq.	Lexical density (%)	Freq.	Lexical density (%)	
9037	58.804	10397	60.022	0.848

As seen in Table 2, the lexical density of medical research papers published in the international journal is 58.804%, while the lexical density of medical research papers in the English journal published in China is 60.022%. This is consistent with Guillen Galve (1998), who reports that the lexical density for written English is over 40%, and that in scientific writing, lexical density can go to a percentage of 55-75. Also, the results show that the lexical density of medical research papers in the English journal published in China is not significantly lower than that of the papers published in international journals. In addition, the results show that lexical density alone does not indicate any information packaging problem in Chinese medical writer's texts, though this does not necessarily mean that their texts are indeed free of any information packaging problems.

3.2 Nominalizations in NS and NNS medical research papers

Nouns nominalized both from verbs and from adjectives were retrieved from the datasets. The frequencies of these nominalizations in the two datasets were then tested for difference (with the Chi-square test), and results of the analysis are presented in the following table:

Table 3. Nominalizations in NS and NNS medical research papers

NS		NNS		Chi square
Freq.	Normalized Freq. (%)	Freq.	Normalized Freq. (%)	
1280	8.329	1075	6.260	47.179 **

** Difference is significant on the 0.01 level.

As shown in Table 3, there are 1280 nominalizations in the NS journal articles, while only 1075

nominalizations are found in the NNSjournal articles . While the total size of the NNS dataset is a bit larger than the size of the NSdataset , fewer nominalizations are used in the NNSdataset . The Chi square test indicates a significant difference in the frequency of nominalizations between the two datasets . Such findings seem to indicate that Chinese medical researchers are less aware that the use of nominalizations may help construct a more academic and more formal written style suitable for EAP . Besides , many of the ideas in the English papers written by Chinese researchers may have not been presented in an abstract and objectivized way as required by English for Academic Purposes .

However , a closer look at the statistics in Table 2 and Table 3 may reveal something more interesting . While nominalizations may help produce texts of a higher lexical density , so that more information could be packaged within concise texts , Chinese medical researchers underuse of nominalizations did not lead to a decrease in the lexical density . In other words , compared with writers of medical texts published in international journals , Chinese medical research writers underuse nominalizations , but they do not underuse content words . This leads us to believe that they may overuse other lexical items which are content words . A second investigation into the data produced the figures in the following table :

Table 4 . Overuse and underuse of lexical items

	NS		NNS		Chi square
	Freq .	Normalized Freq . (%)	Freq .	Normalized Freq . (%)	
Nominalizations	1280	8 .329	1075	6 .260	47 .179 **
Other nouns	4108	26 .731	5341	30 .834	30 .563 **
Lexical verbs	1150	7 .483	1520	8 .775	15 .103 **
Adjectives	2072	13 .483	1931	11 .148	30 .640 **
Adverbs	477	3 .104	530	3 .06	0 .0377

** Difference is significant at the 0 .01 level .

As shown in Table 4 , the frequencies of other nouns (non nominalizations) , lexical verbs and adjectives in the NNSdataset are all significantly different from those in the NSdataset . In other words , Chinese writers of medical research papers underuse nominalizations (chi square =47 .179 , $p < 0 .01$) , but at the same time they overuse other nouns (chi square =30 .563 , $p < 0 .01$) and lexical verbs (chi square =15 .103 , $p < 0 .01$) . Besides , they overuse adjectives (chi square =11 .148 , $p < 0 .01$) . As most of the lexical categories of content words are either overused or underused , further investigations may be necessary to see whether the major or most frequent grammatical structures in Chinese medical researchers English papers are different from those in medical papers published in international journals .

4 . Conclusion

This study explores the lexicogrammatical differences between English medical research papers published in a Chinese journal and those published in an international journal . Results of the study indicate that the way information is packaged in these two types of medical papers may be different . Although the lexical density in the two datasets is not significantly different , the frequency of nominalizations in the two datasets is significantly different , along with the differences in the frequencies of non nominalizations , lexical verbs and adjectives . The differences in these major means of information packaging seem to indicate a great difference in how information is packaged in the papers in the two journals . More such explorations , which incorporate more data , may be necessary to further explore the lexicogrammatical differences between the papers .

References

Bloor , T . & M . Bloor . 2000 . The Functional Analysis of English : A Hallidayan Approach . Beijing : Foreign Language Teaching and Research Press .

- Chafe ,W . &J . Danielewicz . 1987 . Properties of spoken and written language . In R . Horowitz &S . J . Samueles (eds .) . Comprehending Oral and Written Language . San Diego ,CA :Academic Press Inc . pp . 83 -113 .
- Eggins ,S . 1994 . An Introduction to Systemic Functional Linguistics . London :Pinter .
- Galve ,J . G . 1998 . The textual interplay of grammatical metaphor on the nominalizations occurring in written medical English . Journal of Pragmatics 30 :363 -385 .
- Halliday ,M . A . K . 1985 . An Introduction to Functional Grammar . London :Edward Arnold .
- Halliday ,M . A . K . &J . R . Martin . 1993 . Writing Science :Literacy and Discursive Power . London :Falmer .
- Halliday ,M . A . K . &C . Matthiessen . 2004 . An Introduction to Functional Grammar (3rd ed .) . London :Arnold .
- Thompson ,G . 1996 . Introducing Functional Grammar . London :Arnold .
- Wang ,X . F . 2007 . Grammatical concepts and their application in foreign language teaching . In P . L . Jeffery (ed .) . AARE Conference Paper Abstracts 2006 . Available at <http://www.aare.edu.au/b6pap/abs06.htm> (accessed 08/06/15) .

(...continued from p .54)

- 史旭升、吕艳萍 2004 ,图式理论在英语听力教学中的应用,《外语电化教学》第4期。
- 束定芳、庄智象 1996 《现代外语教学——理论、实践与方法》。上海外语教育出版社。
- 田星 2003 ,元认知与外语听力教学,《四川外语学院学报》第6期。
- 袁凤识、肖德法 2006 元认知策略在TEM4中的运用及其与成绩的关系研究,《外语与外语教学》第3期。
- 周天豪 2004 基于元认知理论的传递式听力教学策略《外语电化教学》第4期。
- 周相利 2002 ,图式理论在英语听力教学中的应用,《外语与外语教学》第10期。

(...continued from p .65)

- Talbert ,J . E . 1993 . Constructing a school wide professional community :the negotiated order of a performing arts school . In J . W . Little & M . W . McLaughlin (eds .) Teachers Work :Individuals , Colleagues ,and Contexts . New York :Teachers College Press .
- Wenger ,E . 1998 . Communities of Practice :Learning , Meaning , and Identity . Cambridge , U .K . ; New York ,N .Y . :Cambridge University Press .
- Wenger E . , R . A . McDermott & W . M . Snyder . 2002 . A Guide to Managing Knowledge :Cultivating Communities of Practice . Boston ,Massachusetts :Harvard Business School Press .