The Construction of a Customized Medical Corpus for Assisting Chinese Clinicians in English Research Article Writing

Xiaowen Wang[1], Yong Gao[2], and Tianyong Hao[3](✉)

[1] School of English and Education, Guangdong University of Foreign Studies, Guangzhou, China
[2] Reproductive Medicine Centre, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China
[3] School of Informatics, Guangdong University of Foreign Studies, Guangzhou, China

55736436@qq.com, gaoyong9971@163.com, haoty@126.com

Abstract. A great number of clinicians in mainland China are under increasing pressure to publish their research results on international journals, and they urgently need support for writing research articles in English to compensate their limited English level. Though corpus has been proved to be a useful resource to assist second language learning and writing, research on corpus-assisted medical English writing is very sparse. This paper is concerned with the construction and application of a customized medical corpus for Chinese clinicians to aid their research article writing in English. With the support of a research project, this is the first customized medical corpus built under the joint collaboration between computer-linguistic researchers and clinicians in mainland China to directly serve the actual needs of clinicians. In particular, we report a case of how urologists apply the corpus – CCUT (Customized Corpus for Urology Team) in article writing under the situated assistance of linguistic researchers. The corpus has been found useful in assisting them in choosing the word of appropriate semantic relations, finding grammatical patterns different from general English in specialized medical context, learning how to use unfamiliar medical terms and revising "Chinglish" (unidiomatic) expressions.

Keywords: Customized medical corpus ·Chinese clinicians ·research article writing ·CCUT

1 Introduction

Corpus, as defined by Sinclair (1994:2), is "a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language". In recent years, corpus-based researches have been increasingly applied to second language writing from pedagogical perspectives, and concordancing is for many reasons widely regarded as a useful tool in the writing

adfa, p. 1, 2011.

class (Yoon, 2011:131). Scholars such as Yoon (2011:130-139) have proved that specialized corpora compiled for specific genres or disciplines enable learners to discover vocabulary, word combinations and grammatical patterns. Corpora can also be a good reference resource with which learners check if a specific element of the writing is correct. However, previous studies mainly focus on practices in classroom settings, and few studies have explored how professional practitioners actually exploit the corpora in specific professional genre areas, especially medical field where linguistic support is badly needed.

In fact, clinicians in mainland China are under increasing pressure to publish their research results internationally on the most prestigious journals possible (Qiu, 2010). Assessments for job tenure and promotion require high publication outputs, as do competitive applications for research grant funding. It is almost axiomatic that this now means writing the manuscripts in English (Ammon, 2001; Belcher, 2007), and in the style of English that meets the requirements of the journals concerned (Burrough-Boenisch, 2003; Langdon-Neuner, 2007; Cargill et al, 2012). This situation suggests that a rapidly increasing number of Chinese clinicians need help to enhance their ability to write research articles in English, and a customized medical corpus could be an extremely useful resource to help them. However, so far there has not been any reported medical English corpus available for Chinese clinicians to use for this purpose, no matter in China or abroad.

Computer-linguistic researchers and clinicians in mainland China are jointly working on a project to build the first customized medical corpus directly serving the actual needs of clinicians for English research article writing in specifically targeted medical domain. As the pilot study of this project, we constructed a Customized Corpus for the Urology Team (CCUT), the target users of which are from a research group in one of the top 3A hospitals in China undertaking some natural science projects at national and provincial levels. Most of the members have had very limited practice in scientific writing in English, and some of them turn to local English experts for help, but those experts (no matter native or non-native English speakers) also have difficulties in dealing effectively with the specific language features and discourses of the medicine content. With the help of CCUT, however, linguistic researchers can help the clinicians carry out corpus-assisted English research article writing, and observe their behaviors in using the corpus. One of the team members, Dr. A, has used the CCUT under situated assistance from the linguistic researchers while writing a research article manuscript, which has latter been published on a SCI indexed journal. His behavior of using CCUT will be reported as a case below.

2 Literature review

A considerable number of studies have been conducted on corpus use in second language teaching, especially writing instruction. Among them, most are about writing in general English (Todd, 2001; Cresswell, 2007; Gaskell & Cobb, 2004; Yoon & Hirvela, 2004; Yoon, 2008; Kennedy & Miceli, 2010; Flowerdew, 2010), and only a few are related to writing in an ESP (English for specific purposes) field, such as computer, business, forestry, and law. In the field of forestry, Friginal (2013) investigates the use of corpora to develop the research report writing skills of college-level students enrolled in a professional forestry program. In the field of business English, Walker (2011) examines how a corpus-based study of the factors which influences collocation can help in the teaching. In the field of legal education, Hafner and Candlin (2007) explore the relationship between student use of online corpus tools and academic and professional discourse practices in a professional legal training course at The City University of Hong Kong. In the field of computer science, Chang & Kuo (2011) take a corpus-based, genre-analytic approach to teaching materials development with a corpus of 60 research articles. Gavioli (2005) shows how the analysis of smaller specialized corpora can be used to heighten awareness of key lexical, grammatical or textual issues amongst learners of ESP. Although researchers focus on different aspects of corpora that could help students' writing improvement, it is commonly agreed that most students in such corpus-based teaching class find the corpus approach beneficial to students' writing practices. However, those studies are pedagogically oriented, only focusing on teaching of students in classroom settings with the help of one or more corpora. Although teaching of ESP students is supposed to be career-oriented, how professional practitioners actually make use of corpora to improve their writing in workplace are hardly touched in the literature we found. What is more, research on corpus-assisted medical English writing is even sparse.

Meanwhile, there is also a lack of medical English corpora available for Chinese clinicians. To our knowledge, there are only very few medical English corpora which can be openly used for language learning. GENIA, a corpus of articles extracted from MEDLINE database, is widely used in biomedical language processing (Kim, 2003). However, focusing on biological reactions concerning transcription factors in human blood cells, it only selected articles with the MeSH terms (Kim, 2003), so it is less applicable for clinicians in other medical fields. Other bio-medical English corpora are relatively small in size and mostly only open to limited users (Gurulingappa, 2012). In China, PCMW (Chen & Ge, 2011) is a valuable large scale English-Chinese parallel corpus of medical works, covering about 15 medical domains, such as paediatrics, gynecology, surgery, etc. Mainly targeted on the resource construction of computer (-aided) translation, it is not openly applicable to medical professionals at the moment. Other medical English corpora reported in China are mainly used for linguistic research in stylistics (Ping 2010) or lexicology (Wang, 2008; Wang, 2010). As discussed above, so far there has not been any report about direct application of medical English corpora by clinicians in their L2 learning. In order to bridge computational-linguistic research with the clinicians' actual needs, we establish a customized medical English corpus in the present study.

3 The construction of the customized medical corpus

Inspired by Zheng (2012)'s Eco-dialogical model of interaction, we design a model for the construction of customized medical corpus. As shown in Fig. 1, clinicians provide raw data, based on which the computer-linguistic researchers design and construct the corpus customized to the clinicians' needs in their specific medical domain. Then the computer-linguistic researchers assist them in analyzing data while they write their articles, and provide step-by-step training for corpus use. Finally, the clinicians provide feedback to the constructors so that they can further adapt and improve the corpus accordingly. The clinicians and computer-linguistic researchers work together to achieve meaning perception and realize values in actual writing action in a dynamic and cyclic way.

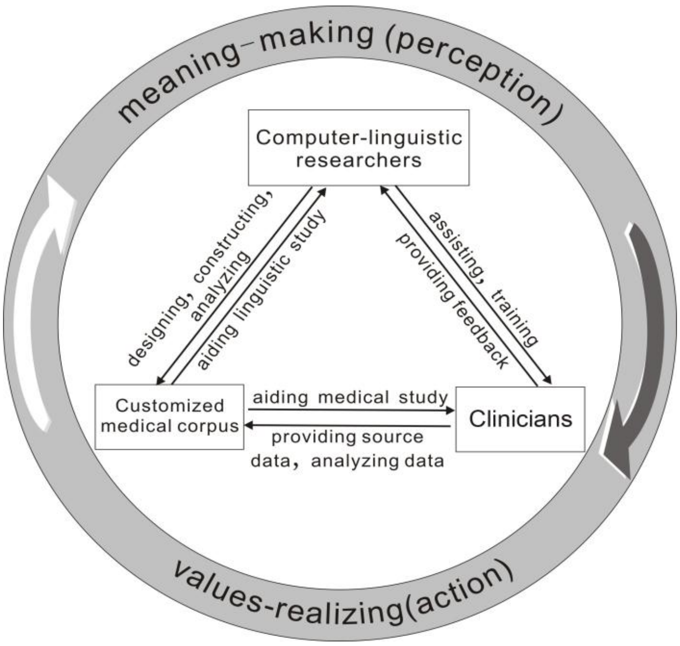


Fig. 1. Model for the construction of customized medical corpus

Specifically, the construction procedure can be divided into the following stages:

1) Needs analysis: The computer-linguistic researchers and clinicians worked together to analyze the needs of target users through discussions and surveys so as to provide suggestions for corpus design.

2) Data collection: The source texts were directly collected from the medical team members, which include 240 medical research articles they downloaded from the PubMed database and shared within the team as core reference readings in recent years. As the team members have the same research direction—application of stem-cell technology in the field of urology, so the source texts mainly fall in the fields of stem-cell and urology.

3) Data cleaning and processing: Linguistic researchers converted pdf files the clinicians provided to txt format, and proofread all texts for two times. Illustrations irrelevant with the language information were deleted. Errors and unrecognizable codes were corrected or substituted after collating with the original pdf text and

consulting the urologists when necessary. A corpus of 1453138 word tokens in total was built (shown in Fig. 2).

4) Corpus sharing on the cloud: The corpus was uploaded to the cloud platform for members in the collaborative project to share.

5) Data application: In the application, the corpus analysis tool we chose for clinicians to use is AntConc (Anthony), a free software relatively easy to operate. Basic functions of AntConc, such as word search, KWIC display, collocates, and clusters, are introduced to the urology team members by the linguistic researchers while assisting their English writing.

6) Feedback collection for corpus improvement: Clinicians upload their feedback to the cloud, and the computer-linguistic researchers will summarize the feedback to further adapt the corpus for their needs in the next step.

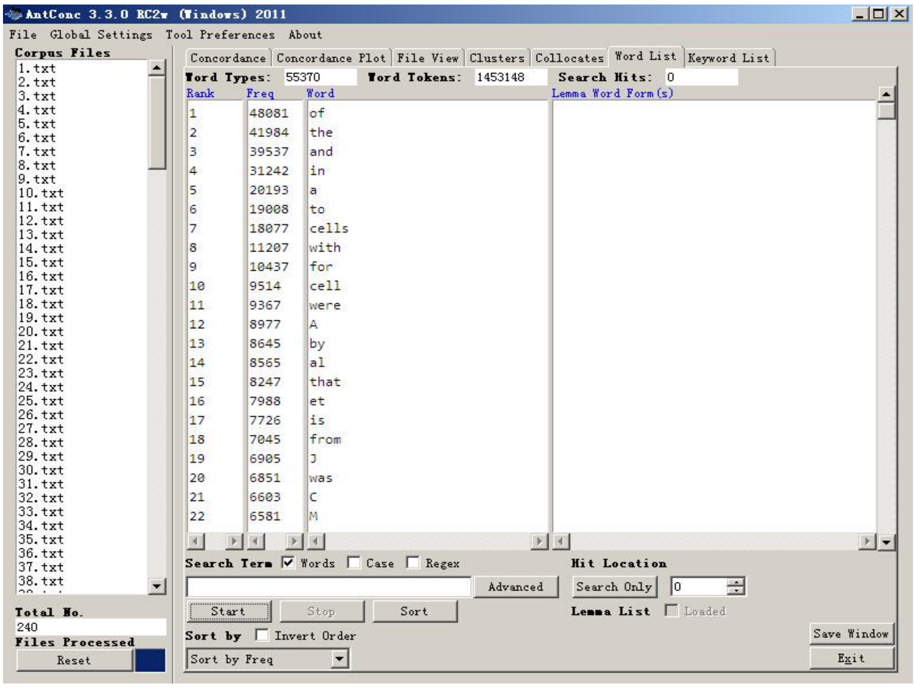


Fig. 2. The wordlist of CCUT (a screenshot in AntConc)

The corpus we constructed could be used to implement a number of functions, including generating concordance lists (key word in context), visual concordance plots, wordlists, and key wordlists, extracting collocation and colligation, extracting terminology, computing collocate salience, creating wordsketches (summary of the word’s grammatical and collocational behavior) and distributional thesaurus (showing similar words in terms of grammatical and collocational behavior), and designing “minitext” (extracts of concordance list for pedagogical use). The basic function - key word in context (KWIC) is similar to the search function of some online databases such as Google Scholar, but the KWIC function can allow more complicated and flexible search to discover collocational and grammatical rules and patterns by inputting regular expressions. Moreover, the concordance results of KWIC in CCUT is much more targeted and thus more useful for clinicians since its source texts are directly related to the clinicians' own medical research fields. An example of another function - wordlist is shown in Fig. 2, in which the wordlist of CCUT is generated to show the frequency of words in the corpus, based on which linguistic researchers could further develop a syllabus of graded professional words for future English training in the targeted medical domain.

4 Corpus use: a case study

Below we report a randomly selected user (Dr. A)'s experience of applying CCUT while writing medical research article manuscript under the assistance of the linguistic researcher (the first author). Dr. A is a 35-year-old urologist. Under great pressure of publication, he is always very interested in using CCUT to help his English writing. We started working with him since 2012 when he was a novice clinician. Over these years, by guiding him to perform corpus analysis, the linguistic researcher could at the same time observe his corpus use behavior directly. Based on our observation, his use of corpus mainly falls into the following four purposes：

1) Choosing a word of appropriate semantic relations in medical context

Clinicians may feel hard to use even a common English word in the specialized medical context. For example, while writing a research article manuscript, Dr. A turned to the linguistic researcher about how to express "获得" (get) in English when describing the process of getting a certain cell.

The linguistic researcher then guided him to analyze the target word in terms of semantic relations proposed by Sinclair (1991, 1996). As shown by Sinclair (1996), corpus work accounts for at least four types of meaningful relations that words entertain with other words around them. In corpus linguistics, these are called: collocation, colligation, semantic preference and semantic prosody. Collocation is defined as "the occurrence of two or more words within a short space of each other" (Sinclair 1991:170). Colligation is, instead, the relationship between a word and a grammatical class of words. Semantic preference is the relationship between a word and a semantic class of words. Semantic prosody does not only have to do with the relationship between words, but it also involves the way words affect each other with their meanings. "Prosody" is applied particularly to the way in which words or expressions create an aura of meaning capable of affecting words around them (Gavioli, 2005: 45-46).

A comparison of the selected concordance lines of "aquire" and "obtain" is shown in Table 1. The search for "acquire" in CCUT with AntConc provided 44 occurrences, showing that this is a relatively frequently used word in a medical context. However, at the collocates on the right of the node, most of the collocates were not biomedical entities, but mainly some abstract nature, characteristics or capacity of certain biomedical entities, such as "properties", "characteristics", "ability", and "expression". The search for "obtain" in CCUT provided 57 occurences. We found the semantic relations of this word have special patterns in the medical context. For collocation, it is mostly collocated with biomedical entities, such as cell, tissue, fraction, material, and gene expression profiles. For colligation, "to obtain +NP" or "be+ adj. +obtain + NP" is the most salient pattern. For semantic preference, word combinations like "attempts to, able/unable to, could not, it took many years to, hard/difficult to" are associated with "obtain", which all seemed to show certain difficulty in the obtaining action. Therefore, when it comes to the semantic prosody, "obtain" has the connotational associations of "successfulness after hard efforts". All of the four scales of semantic relations of "obtain" matched Dr. A's situation, i.e., to get some cells after great efforts of scientific research, so he confirmed that "obtain" is the best word to choose, rather than "acquire". At the same time, he also made clear how to use "acquire" in other situations.

Table 1. Comparison of the concordance lines of "acquire" and "obtain"

Words	Concordance lines in CCUT (extract)		
acquire	regulate VE-cadherin and CD105 expression and	**acquire**	the capacity to generate multilineage
	genic Leydig cells which, in addition, rapidly	**acquire**	neuronal and glial properties. These
	fter their displacement from vessel walls, first	**acquire**	steroi-dogenic properties expression
	Marion RM, Strati K, Li H, et al. Telomeres	**acquire**	embryonic stem cell characteristics
	experience hypoxia, they dedifferentiate and	**acquire**	stem cell 445 neural crest-like features
	HIF-2a protein expression, differentiate, and	**acquire**	expression of SNS markers. In ypoxic
	o not only lose their WT function but also to	**acquire**	new properties, including the ability t
	ved ECs (Xiao et al., 2006). ESC-derived ECs	**acquire**	cobblestone morphology (Cho et al.,
	ntrolling the risk for metastasis. RCCs usually	**acquire**	metastatic potential when their size
obtain	repair the body (1), but **it took many years to**	obtain	hard evidence in support of this theory.
	ing nature of PDA, **it is nearly impossible to**	obtain	pure tumor tissues without a contaminati
	RRK2 - PD iPSC. Unfortunately, we **could not**	obtain	a clear signal for SNCA in immunoblots
	AC)-based methods **offers alternative ways to**	obtain	genetically corrected iPS cells [17-19].
	Urce of cells when primary cells **are difficult to**	obtain	in sufficient numbers for in vitro studies or
	rdiac tissues from DCM patients **are difficult to**	obtain	and do not survive in long-term culture.
	ack of cell surface markers hindered **attempts to**	obtain	purified SLC fractions [24]. Once isolate
	is less well understood. A recent study **sought to**	obtain	a specic marker for peritubular myoid c
	repair. One of the **major problems has been to**	obtain	MSC populations free of hematopoietic
	ells. However, in these cases, we **were unable to**	obtain	any hES-like colonies at all. Because

2) Finding medical grammatical patterns different from general English

While writing the manuscript, Dr. A came up with another question: Should he use "mouse cell", or "mouse's cell"? By intuition he thinks "mouse's cell" is grammatically correct, but he feels like seldom seeing "mouse's cell" in medical literature reading.

To answer this question, we searched "mouse 's" in BNC, and found occurrences such as "mouse's body", "mouse's ear", "mouse's tail", "mouse's dulled wintering heart". A look into the collocates of "mouse" in BNC indicated that "'s" is its most frequently used collocate.

Surprisingly, a search for “mouse’s” in CCUT showed no concordance hit at all, but a further search for “mouse * cell*” instead provided 101 occurrences (shown in Fig. 3). To make this clear, we worked out in AntConc the 2-word clusters of “mouse” sorted by frequency with “mouse” on the left in CCUT, the tops 3 expressions on the list are “mouse embryonic”, “mouse model”, “mouse testis”, which in general English may be expressed as “mouse’s embryonic…”, “mouse’s model”, “mouse’s testis”.

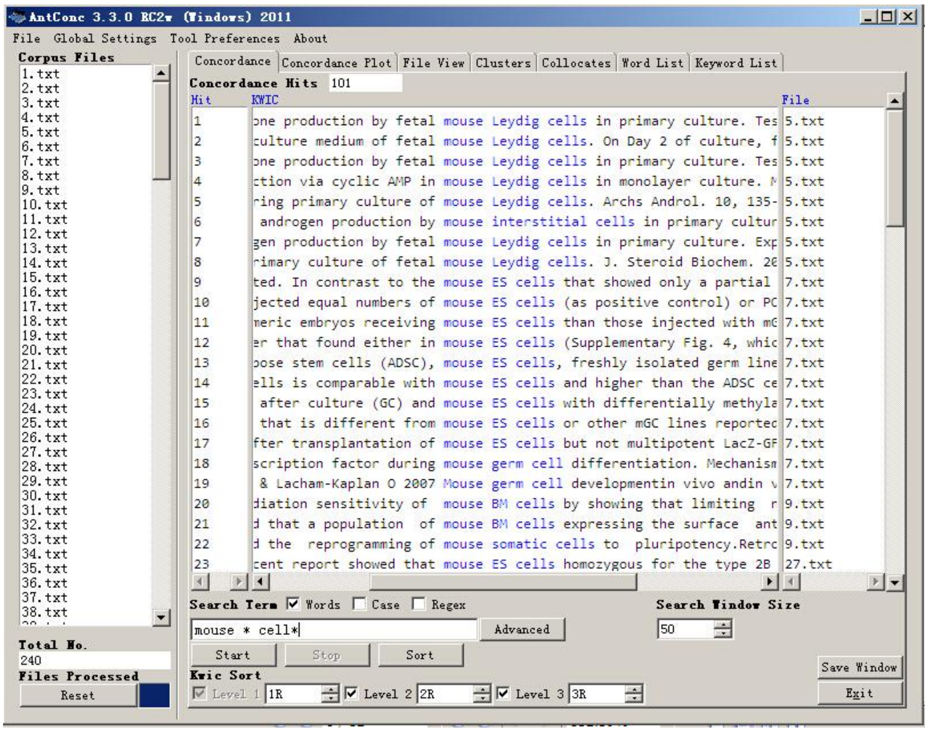


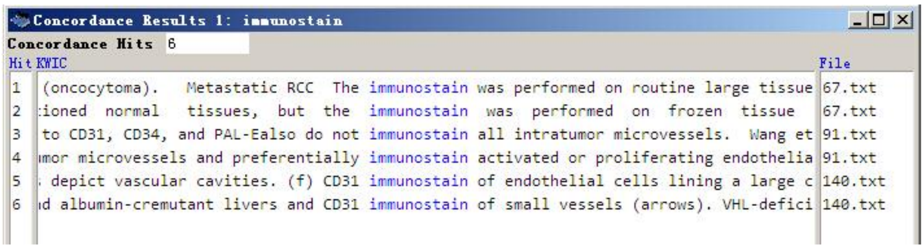
Fig. 3. The screenshot for the concordance results of “mouse * cell*” in CCUT

Being aware of the difference in grammatical patterns - “mouse +NP” in medical English vs “ mouse(’s) + NP” in general English, Dr. A understood that he should follow the conventions in the medical context and wrote in his manuscript “mouse LCs” instead of “mouse’s LCs” and “mouse testes” instead of “mouse’s testes”.

3) Learning how to use unfamiliar medical terms

For unfamiliar medical terms, the clinicians have to firstly consult medical English dictionaries to find the potential English equivalent words, but the example sentences might be insufficient. Though CCUT is not useful in the first step of searching word-to-word equivalents, it provides more information than the dictionaries about how the word is used in context. For example, Dr. A searched an unfamiliar term - “免疫荧光染色” (immunostain) in the new Comprehensive Medical English

Dictionary (KingYee Ltd., 2016), a dictionary regarded as the most comprehensive and useful medical English dictionary in his circle, but he only found its equivalent English expression "immunostain" with no example sentence at all. So we searched "immunostain" in CCUT. The results are shown below.



Concordance Results 1: immunostain

Concordance Hits 6

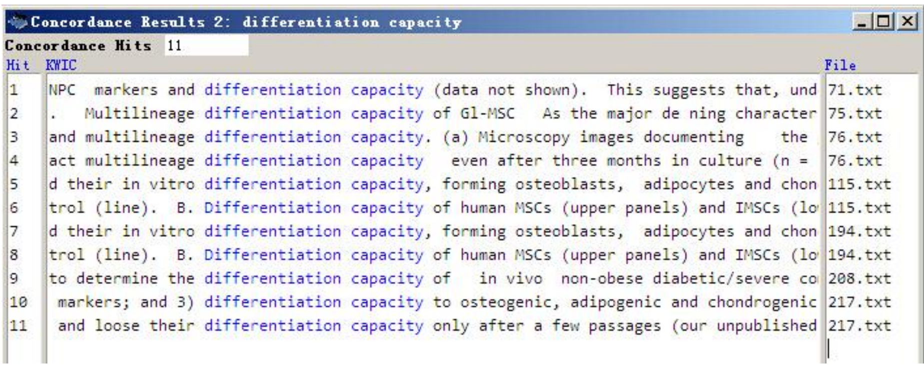
Hit	KWIC	File
1	(oncocytoma). Metastatic RCC The immunostain was performed on routine large tissue	67.txt
2	:ioned normal tissues, but the immunostain was performed on frozen tissue	67.txt
3	to CD31, CD34, and PAL-Ealso do not immunostain all intratumor microvessels. Wang et	91.txt
4	ımor microvessels and preferentially immunostain activated or proliferating endothelia	91.txt
5	: depict vascular cavities. (f) CD31 immunostain of endothelial cells lining a large c	140.txt
6	ıd albumin-cremutant livers and CD31 immunostain of small vessels (arrows). VHL-defici	140.txt

Fig. 4. The concordance results of "immunostain" in CCUT

Though there were only 6 occurrences, we noticed the recurrent pattern of "the immunostain was performed on…", so Dr. A learned to use this pattern to make his sentence in the manuscript. By analyzing concordances, corpus users can grasp the meanings and functions of the structures that were presented to them much better than when they were presented in the traditional fashion (Gavioli, 2005: 28).

4) Revising "Chinglish" (unidiomatic) expressions

The word combination Dr. A used for a subtitle in his manuscript was originally "In vitro Nes-GFP+ cells differentiation", but he realized it might be "Chinglish" (unidiomatic), and asked whether he needed to revise it in a more idiomatic way. To check whether this expression is unidiomatic, a search of "differentiation capacity" was generated, with 11 occurrences found (Shown in Fig. 5).



Concordance Results 2: differentiation capacity

Concordance Hits 11

Hit	KWIC	File
1	NPC markers and differentiation capacity (data not shown). This suggests that, und	71.txt
2	. Multilineage differentiation capacity of Gl-MSC As the major de ning character	75.txt
3	and multilineage differentiation capacity. (a) Microscopy images documenting the	76.txt
4	act multilineage differentiation capacity even after three months in culture (n =	76.txt
5	d their in vitro differentiation capacity, forming osteoblasts, adipocytes and chon	115.txt
6	trol (line). B. Differentiation capacity of human MSCs (upper panels) and IMSCs (lo	115.txt
7	d their in vitro differentiation capacity, forming osteoblasts, adipocytes and chon	194.txt
8	trol (line). B. Differentiation capacity of human MSCs (upper panels) and IMSCs (lo	194.txt
9	to determine the differentiation capacity of in vivo non-obese diabetic/severe co	208.txt
10	markers; and 3) differentiation capacity to osteogenic, adipogenic and chondrogenic	217.txt
11	and loose their differentiation capacity only after a few passages (our unpublished	217.txt

Fig. 5. The concordance results of "differentiation capacity" in CCUT

In all the 11 instances, the word on the left of "differentiation capacity" were all attributive adjuncts, such as "in vitro", "multilineage", and no expression of "[cell name] + differentiation capacity" was found. Therefore, the word "Nes-GFP^{+} cells" might not be suitable to be put before "differentiation capacity". From Line 5 and 7, we found that "in vitro" could be put before "differentiation capacity" to be an attributive adjunct, so we decided to use the word combination of "in vitro differentiation capacity".

As for where to put "Nes-GFP$^+$ cells", in the concordance lines we found "differentiation capacity +of +[entity name]" in Line 2, 6, 8, 9 (in case the entity differentiates) and "differentiation capacity + to + [entity name]" in Line 10 (in case the entity is the result of differentiation). So Dr. A revised the original subtitle "in vitro Nes-GFP$^+$ cells differentiation capacity" to "in vitro differentiation capacity of Nes-GFP$^+$ cells".

As the Chinese language is characteristic of parataxis (words are connected by implicit coherence), and the English language is characterized by hypotaxis (words are connected by explicit cohesive devices), the missing of the connective "of" in the original subtitle might be caused by the negative transfer from Chinese. Just as this case shows, clinicians can use the CCUT corpus to explore "idiomatic" areas of language and even repair the negative transfer from their mother language.

5 Discussion

The corpus under our construction is different from other kinds of non-customized corpora traditionally built by the corpus-linguists in that: 1) It is jointly developed by medical and computer-linguistic researchers; 2) Though it could also be used for linguistic analysis, it is user-oriented in that its primary function is to serve the clinicians; 3) Its source texts are provided by the clinicians so that it is highly related to the specific research domain of the users; 4) Computer-linguistic researchers not only help the clinicians build the corpus, but also provide situated assistance, corpus use training, and guidance for data analysis step by step to make sure the users can really make effective use of the corpus; 5) Computer-linguistic researchers collect user feedbacks from the clinicians and further develop the corpus according to target users' needs; 6) While the computer-linguistic researchers play the roles of designers, constructors, data analysts, assistants and trainers, the clinicians play the roles of source providers, main corpus users, data analysts, feedback providers, and trainees. The relationships among the computer-linguistic researchers, clinicians and corpus are dialogical and dynamic.

With the joint efforts of the computer-linguistic researchers and the clinicians, the customized corpus has been proved to provide targeted language learners with invaluable information, especially the recurrent, conventional lexical and syntactic patterns in specialized context, the usage of specialized words that could not be found sufficiently described in dictionaries, and the ways to test one's intuitions so as to repair certain negative transfer from Chinese to English. In the situated guidance, it is essentially important for linguistic researchers to guide the clinicians to gradually increase their sensitivity and awareness of the conventions in the medical context through the clues provided by the corpus. If such kind of training keeps running, the clinicians could enhance their language intuitions, and get more and more familiar with the discourse of their specialized community. However, one thing worth noting is how to transfer “from maximum guidance to maximum independence” (Gavioli, 2005: 127) so that the clinicians can be independent analyzers finally. After the clinicians get familiar enough with the corpus in the situated guidance, training workshops will be organized to summarize the ways of corpus use, and multi-media demo videos, guide books will be offered to help them.

There are some limitations for the current preliminary study. The ways of application are still limited, and empirical evaluations need to be collected after larger-scale application. As the pilot study for our project on the customized medical English corpus, CCUT is limited to the domain of stem cell and urology. As the research goes, our project will extend the customized medical corpus to include more medical domains.

6 Conclusions

This paper discusses the construction of customized medical corpus CCUT and shows how urologists from the medical research team applied CCUT to aid their research article manuscript writing in English. With the situated guidance of the linguistic researchers, CCUT has been effectively used to help clinicians choose the word of appropriate semantic relations, find grammatical patterns different from general English in specialized medical context, make use of unfamiliar medical terms and revise unidiomatic expressions. Our case study presented that it is not only possible but also worthwhile to introduce clinicians to corpus linguistics through a dialogical, cyclic and goal-oriented collaboration between the computer-linguistic researchers and clinicians on customized corpus.

Acknowledgements. This work was supported by the Science and Technology Project of Guangdong Province, China (2016A040403113) and Innovative School Project in Higher Education of Guangdong, China (GWTP-LH-2014-02).

References

1. Ammon, U.: The dominance of English as a language of science. Effects on other languages and language communities. Mouton de Gruyter, New York (2001)
2. Belcher, D.D.: Seeking acceptance in an English-only research world. Journal of Second Language Writing 16(1): 1-22 (2007)
3. Burrough-Boenisch, J.: Shapers of published NNS research articles. Journal of Second Language Writing 12: 223-243 (2003)
4. Cargill, M., et al.: Educating Chinese scientists to write for international journals: Addressing the divide between science and technology education and English language teaching. English for Specific Purposes 31: 60-69 (2012)
5. Chang, C., Kuo, C.: A corpus-based approach to online materials development for writing research articles. English for Specific Purposes 30: 222-234 (2011)
6. Chen, X.X., Ge, S.L.: The Construction of English-Chinese Parallel Corpus of Medical Works Based on Self-Coded Python Programs. Procedia Engineering 24: 598-603 (2011)

7. Cresswell, A.: Getting to ‘know’ connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Quereda, & S. Juan (Eds.), Corpora in the foreign language classroom. Amsterdam, Netherlands: Rodopi (2007)
8. Flowerdew, L.: Using corpora for writing instruction. In A. O’Keeffe, & M. McCarthy (Eds.), The Routledge handbook of corpus linguistics. pp. 444-457 (2010)
9. Friginal, E.: Developing research report writing skills using corpora. English for Specific Purposes 32: 208–220 (2013)
10. Gaskell, D., Cobb, T.: Can learners use concordance feedback for writing errors? System 3: 2301–319 (2004)
11. Gavioli, L.: Exploring Corpora for ESP Learning. John Benjamins B.V. (2005)
12. Gurulingappa, H.: Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. Journal of Biomedical Informatics 45(5): 885-892 (2012)
13. Hafner, C.A., Candlin, C.N.: Corpus tools as an affordance to learning in professional legal education’.Journal of English for Academic Purposes 6: 303-318 (2007)
14. Kennedy, C., Miceli, T.: Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. Language Learning and Technology, 14(1): 28–44 (2010)
15. Kim, J-D., Ohta, T., Tateisi, Y., Tsujii, J.: Genia corpus-semantically annotated corpus for bio-textmining. Bioinformatics, 19(suppl 1): 180-182 (2003)
16. Langdon-Neuner, E: Let them write English. Revista do Colégio Brasileiro de Cirurgiões 34(4): 272-276 (2007)
17. Lee, D., Swales, J.: A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. English for Specific Purposes 25: 56-75 (2006)
18. Ping, W.: Corpus-based contrastive study on stylistics of Foreign and Chinese medical research article abstracts in English. Journal of Xianning University 30(3): 90-91 (2010)
19. Qiu, J.: Publish or perish in China. Nature 463: 142-143 (2010)
20. Sinclair, J. : Corpus, Concordance, Collocation. OUP, Oxford (1991)
21. Sinclair, J.: Corpus Typology: A Framework for Classification. In: Melchers G., Warren B. (eds), Studies in Anglistics. Almquist and Wiksell International, Stockhom. 17-34 (1995)
22. Sinclair, J.: The search for units of meaning. Textus 9: 75-106 (1996)
23. Walker, C.: How a corpus-based study of the factors which influence collocation can help in the teaching of business English. English for Specific Purposes 30: 101-112 (2011)
24. Wang, L.: Application of corpus and concordance tools to medical English vocabulary teaching. China Medical Education Technology 22(5): 427-430 (2008)
25. Wang, S.: Corpus-based Medical English Vocabulary Teaching. J. Gansu College of TCM 2010(6): 59-61 (2010)
26. Todd, R.W.: Induction from self-selected concordances and self-correction. System 29, 91–102 (2001)
27. Yoon, H., Hirvela, A.: ESL student attitudes towards corpus use in L2 writing. Journal of Second Language Writing 13: 257-283 (2004)
28. Yoon, H.: More than a linguistic reference: The influence of corpus technology on L2 academic writing. Language Learning & Technology 12: 31-49 (2008)
29. Yoon, C.: Concordancing in L2 writing class: An overview of research and issues. Journal of English for Academic Purposes 10: 130-139 (2011)
30. Zheng, D.: Caring in the dynamics of design and languaging: exploring second language learning in 3D virtual spaces. Language Sciences 34: 543-558 (2012)