# A Corpus-Based Contrastive Critical Genre Analysis of Universities' Introductory Discourses

YU Weiwei

Zhejiang University, Hangzhou, China

Educational discourses, or to be more specific, universities' introductory discourses have not been studied from the contrastive perspective using critical genre analysis theories. This study uses Fairclough's framework of discourse analysis to take a corpus-based contrastive study between the Chinese and global universities' introductory discourses from three aspects. Firstly, the textual dimension will be explored using the corpus-based approach. Secondly, it studies from the discursive dimension from intertextuality and interdiscursivity. Finally, the social cultural perspective study will be given about the corresponding social cultural interpretation on the differences between Chinese and global universities' introduction discourses.

*Keywords:* university introductory discourses, discourse analysis, corpus-based, intertextuality, social cultural practice

## Introduction

Educational discourses, as a special institutional genre, have been mentioned in Fairclough's (1992) *Discourse and Social Change*. From the perspective of intertextuality of educational discourses with advertising discourses, Fairclough commented that "education is one of a number of domains whose orders of discourse are being colonized by the advertising genre" (Fairclough, 1989, pp. 208-211). According to his analysis, on the one hand, universities try to sell them to themselves, but they also have to impose regulations of control to them (p. 214). Therefore, the educational discourse is a mixture or appropriation of persuasion and information. So, universities also carry the color of promotion. Fairclough establishes a framework by analyzing three aspects: the textual, discursive, and social cultural practice. In the contemporary world, Bhatia establishes a three-space model of discourse analysis, integrating a number of different perspectives on genre analysis, which mainly shifts genre analysis direction from text-internal to text-external factors, such as social-pragmatic, institutional, and ethnographic, etc. (2004, p. 171). This shift broadens the conventional dimension on lexical-grammatical dimension and bridges professional genre with professional practice. So, this study tries to integrate the two models together and studies universities' introductory discourses from text, and discursive and social cultural factors. Although Fairclough has made an exploratory study on educational discourses, no previous studies have covered a contrastive study between inter-cultural ones using a corpus-based approach to deal with corpora of a large scale, thus this study is justified in this sense.

In Fairclough's framework, he always follows the overall progression from: (i) analysis of discourse practices (at a "macro" level), focusing upon the intertextuality and interdiscursivity of discourse samples; to (ii)

---

YU Weiwei, Ph.D. candidate, School of International Studies, Zhejiang University.

analysis of texts (plus "micro" aspects of discourse practice); and to (iii) analysis of the social practice of which the discourse is a part (1992, p. 231). Fairclough offers a guideline of studying discourse practice from interdiscursivity or constitutive intertextuality, intertextual chains, conditions of discourse practice or social contexts and manifest intertextuality. Fairclough says that "the constitutive intertextuality of a text is constituted through a novel configuration of existing types, and specifically the emergence of hybrid information-and-publicity (or 'telling-and-selling') discourse" or to make it simple, the mixing of different types of discourses and "in manifest intertextuality, other texts are explicitly present in the text under analysis; they are 'manifestly' marked or cued by features on the surface of the text, such as quotation marks" (1992, p. 104). Intertextual chains are concerned with the distribution of different types of discourses, which are related to each other in that they can be transformed in regular and predictable ways (Fairclough, 1992, p. 130). He studies text from interactional control, cohesion, politeness, ethos, grammar, transitivity, theme, modality, word meaning, wording, and metaphor. Social practice is studied from the perspective of social matrix of discourse or its social relations and structures in terms of conventions, orders of discourse (which are characterized by vocabularies, turn-taking systems, etc.) and ideological and political effects of discourse. This paper will focus on the manifest intertextuality and interdiscursivity parts.

## Background

According to information from the Ministry of Education of China website, the list of Chinese higher education institutions is divided into five types, 2,101 state and 386 non-state regular institutions of higher education, 354 state and two non-state institutions of adult higher education, and 309 state approved independent colleges. Here, the top 20 Chinese universities are all national key universities under the direct jurisdiction of Ministry of Education. And among the top 20 global universities, nearly all of them are private universities except nine universities. As Fairclough says, no particular significance should be attached to the choice of the university or the materials as similar tendencies are evident in other entries and other educational introductions (1992, p. 210). Here, the author only chose the most prestigious, top universities on league tables both in China and in the world.

## Materials and Methods

The study of educational institutional discourses is based on general introductory discoursesor overviews of top 20 Chinese universities, published by China Association of Alumnae and top 20 global universities, published by Times Higher Education and International Development Program (IDP) Educational Group, all of which are acquired from the official websites. The general descriptive statistics can be seen from Table 1. Since Chinese university introductory discourses are mostly translated from Chinese version and with no further detailed division as available links, such as "History" "Facts" and "Statistics", while global universities as studied in this paper are almost all with many subdivision labels and picture links to choose from, and the author only chooses the immediate introduction and not the whole text, so there is a big difference in the total words and other descriptive statistics between Chinese and global universities.

Having been annotated by JClaws and analyzed afterwards by Wordsmith 5 on the wordlist, keywords, collocation, concordance, clusters, the materials were studied on the lexical-grammatical features including tense, voice, and modality, which are examined to explore the features on textual dimension. When choosing reference corpora, considering that the aim of this study is not to study universities' introductory discourses as a

genre compared to general English but the contrastive study between Chinese and global universities, the author chooses global universities' introductory discourses as the reference corpus when calculating keywords of Chinese ones and vice versa. In this way, the author can get a contrastive result on Chinese and global universities' introductory discourses. Among them, the author will focus on the past, simple present, present perfect, and the passive voice and the modal words. Moreover, collocation and keyword clusters will be examined further. On the social cultural level, the author shall discuss the potential contextual or professional practice explanations.

Table 1

*Descriptive Statistics of Corpora*

|  | Types (distinct words) | Token (running words) in text | Sentences | Paragraphs | Standardized type/Token ratio (TTR) | Mean word length |
|---|---|---|---|---|---|---|
| Chinese | 2,714 | 18,063 | 17,792 | 4,952 | 38.87 | 5.5 |
| Global | 1,426 | 5,062 | 4,656 | 20 | / | 5.42 |

## Results

Firstly, as seen from Table 2, after annotation and calculation, the following results are acquired on wordlist, keyword, and lexical-grammatical features. Of the top 10 words from both Chinese and global universities, they are more or less similar except that global universities have more figures while Chinese ones more definitive "The", Chinese corpus has more "University". The words "Has" and "National" are two words that do not stand in global university corpus' top 10 words, while "Students" is the one that does not stand in Chinese university introduction Corpus's top 10.

Table 2

*Wordlist and Relative Frequency*

| Ranking | Wordlist (China) | Relative frequency (%) | Wordlist (Global) | Relative frequency |
|---|---|---|---|---|
| 1 | The | 5.8 | # | 5.97 |
| 2 | # | 5.19 | The | 5.24 |
| 3 | And | 4.98 | And | 4.76 |
| 4 | Of | 4.71 | Of | 3.83 |
| 5 | In | 2.34 | In | 2.17 |
| 6 | University | 2.33 | To | 1.70 |
| 7 | A | 1.40 | A | 1.58 |
| 8 | To | 1.27 | University | 1.32 |
| 9 | Has | 1.15 | Students | 1.28 |
| 10 | National | 1.06 | Is | 0.95 |

As for the keywords, when choosing one as the other's reference corpus, the findings revealed a lot of interesting results. As suggested by Table 3, it seems that for the positive keywords, there are "China", "National", "Chinese", "Key", "Has", "RUC", and "Ministry"; and for the negative keywords, there are "Faculty", "Students", "AMP", "About", and "Our". When taking a look at those keywords of the global universities, the positive keywords are "Our", "About", "AMP", "Harvard", "Oxford", "Stanford", "Community", "U", "MIT", "We", "Hopkins", "Johns", "Chicago", "Facts", "Students", "Penn", "Faculty", "Princeton", and "UCL"; and the negative keywords are "Has" and "National". How should this be understood?

Firstly, let us look at the Chinese corpus; it makes sense for the seven positive keywords for all of the 20 Chinese universities are state universities and there is a sense of authority when mentioning that the university is a "national key university under the jurisdiction of Ministry of Education"; that is the best-selling point for the university to attract a lot of eyes, in the largest sense not the "Faculty" or "Students". The Chinese universities will not put too much emphasis on "Our", because it is not that objective and after all; the university is a public one, not theirs really. On the counterpart of the global university corpus, among the positive keywords, apart from the proper names, such as "AMP", "Harvard", "Oxford", "Stanford", "MIT", "Hopkins", "Johns", "Chicago", "Penn", "Princeton", and "UCL", there were words, like "Our", "About", "Community", "U", "We", "Facts", "Students", and "Faculty"; and the negative keyword are "Has" and "National". Here, we can see as students are their main "customers"; they regard themselves as "Communities" and "Students" and "Faculty" as their residents, which form the main body of the global universities. This just corresponds with the use of "Our" and "We".

Table 3

*Keyword and RC*

| Ranking | Keyword (China) | RC (%) | Keyword (Global) | Relative frequency |
|---------|-----------------|--------|------------------|--------------------|
| 1 | China | 50.52 | Our | 60.27 |
| 2 | National | 37.76 | About | 58.65 |
| 3 | Chinese | 37.63 | AMP | 49.28 |
| 4 | Key | 36.11 | Harvard | 45.61 |
| 5 | Has | 33.85 | Oxford | 42.57 |
| 6 | RUC | 25.23 | Stanford | 39.52 |
| 7 | Ministry | 24.74 | Community | 39.52 |
| 8 | Faculty | -24.40 | U | 39.52 |
| 9 | Students | -27.42 | MIT | 36.48 |
| 10 | AMP | -49.28 | We | 34.28 |
| 11 | About | -58.65 | Hopkins | 32.81 |
| 12 | Our | -60.27 | Johns | 32.81 |
| | | | Chicago | 30.40 |
| | | | Facts | 29.92 |
| | | | Students | 27.42 |
| | | | Penn | 27.36 |
| | | | Faculty | 24.40 |
| | | | Princeton | 24.32 |
| | | | UCL | 24.32 |
| | | | Has | -33.58 |
| | | | National | -37.76 |

When we take a look at the past tense features of the Chinese corpus, as can be seen from Table 4, it indicates that in colligate dimension, apart from the functional words on the top of the list, there were collocates, like "Established" "Founded" "National" "Visited" "Won" "Renamed" and "The first", most of which were related to the history and the past achievement and important events, including famous figures' visiting of the universities.

Table 4

*Colligates of the Past Tense in the Chinese Corpus*

| Ranking | Past tense (China) | Frequency | Clusters | Frequency |
|---------|--------------------|-----------|---------|-----------|
| 1 | was | 60 | University was | 18 |
| 2 | The | 49 | It was | 14 |
| 3 | at | 48 | The university | 14 |
| 4 | university | 26 | Was founded | 9 |
| 5 | of | 25 | Was established | 7 |
| 6 | and | 25 | The national | 7 |
| 7 | in | 20 | Was the | 7 |
| 8 | were | 18 | The first | 7 |
| 9 | It | 14 | Visited RUC | 6 |
| 10 | A | 13 | By the | 6 |
| 11 | to | 13 | Was renamed | 5 |
| 12 | Its | 10 | Won the | 5 |
| 13 | founded | 9 | | |
| 14 | Established | 8 | | |
| 15 | National | 7 | | |
| 16 | Visited | 7 | | |
| 17 | Won | 7 | | |
| 18 | An | 7 | | |
| 19 | Research | 7 | | |
| 20 | Renamed | 6 | | |

As to the global corpus, the most frequent colligates with past tense cases are "Was", "The", and "And" with frequency of 12, seven, and five respectively. The corresponding clusters are "the + noun" with a frequency of six as can be seen from Table 5.

Table 5

*Past Tense of Global Corpus*

| Past tense (Global) | Frequency | Clusters | Frequency |
|---------------------|-----------|---------|-----------|
| Was | 12 | the | 6 |
| The | 7 | | |
| And | 5 | | |

Table 6 shows the present perfect division in Chinese corpus. As can be seen from Table 6, most of the 116 collocates here are concerned with "Has established" (6) and "Has become" (5). Besides, there are others, like "Has received, Created, Educated, Cultivated, Participated, Reached, Made, Undertaken, Developed, Won, Ranked, Gained, Produced", etc. these words are all concerned with their past achievement, the rewards they have gained, and kinds of facts. When taking a look at the present perfect of the global corpus, it appeared that there were altogether 13 cases with collocates, such as has "_become" "_produced" "_received" "_pioneered" "_made" "_served" "_studied" "_succeeded" "_come", and no significant collocates and clusters.

Table 6

*Present Perfect of Chinese Corpus*

| Ranking | Present perfect (China) | Frequency | Clusters | Frequency |
|---|---|---|---|---|
| 1 | Has | 28 | University has | 42 |
| 2 | been | 28 | The university | 19 |
| 3 | the | 13 | university | 19 |
| 4 | university | 12 | Has been | 18 |
| 5 | Has (established) | 6 | IT has | 13 |
| 6 | Has (established) | 6 | University has been | 12 |
| 7 | has (become) | 5 | AND has | 9 |
| 8 | (Has) become | 5 | RUC has | 9 |
| 9 | | | HUST has | 7 |
| 10 | | | More than | 5 |
| 11 | | | Shandong University | 5 |
| 12 | | | NJU has | 5 |

To take a look at the passive voice as suggested from Table 7, results of the Chinese corpus are as follows. Similar to the past tense, most of the passive voice cases were concerned with past history, such as was "_located" "_founded" "_established" "_awarded" and "_renamed". So, these are all facts belonging to the past. These facts were used to achieve authoritative and objective effects. On the passive voice part of global corpus, it seems also insignificant, while there were 34 cases, among which were cases, like be "_felt" "_recognized" "_published" "_known" "_answered" "_founded" "_awarded" "_honored" "_developed" "_known" "_organized", etc.

Table 7

*Passive Voice of Chinese Corpus*

| Ranking | Passive voice (China) | Frequency | Clusters | Frequency |
|---|---|---|---|---|
| 1 | located | 10 | The | 34 |
| 2 | founded | 9 | Have been | 22 |
| 3 | Renamed | 8 | University was | 17 |
| 4 | Established | 8 | Has been | 13 |
| 5 | Awarded | 7 | The university | 12 |
| 6 | Was established | 7 | Is located | 7 |
| 7 | AT located | 7 | Founded in | 7 |
| 8 | University renamed | 7 | University is | 7 |
| 9 | been | 6 | which | 6 |
| 10 | was | 5 | It was | 6 |
| 11 | THE located | 5 | Located in | 5 |
| 12 | IN located | 5 | As the | 5 |
| 13 | | | Campus is | 5 |

As shown in Table 8, the modal words in both Chinese and global corpora are a little bit different in types and frequency. In both corpora, the most frequent modal word is "Can", but it is more frequent in the global corpus. In the Chinese corpus, there are "May" "Would" and "Could" besides "Can" and "Should". When taking a look at the source text, the author found that most of the modal words are used when describing the equipment and facilities of the universities, so this is ability. While for the global universities, four out of the six instances of "Can" are used when concerning the students or their families, such as "Can afford", "Can

read", "Can enter", and "Can study". In this sense, it carries more regulative information or permission rather than ability and achieves intertextuality of "Telling" and "Selling". This is different from "Can" in Chinese corpus.

As to the quotation, there exists a big difference between the Chinese and global corpora. As is shown from Table 8, there were 51 cases of quotation in the Chinese corpus and only four cases in global one. Among the 51 Chinese cases, one was from the speech of former president Jiang Zemin, and the rest 50 were all labels of programs or projects, such as "project 211" and "project 985"; famous figures, campaigns, honors, or changes in names concerned with the university's development, such as "four tutors", "May 20th Movement", "the most beautiful university", and "Tsinghua School"; and educational philosophy of the university, such as "seeking truth". In this way, through using the third party discourses, it adds objectivity and intertextuality, while the global quotation is more concerned with the original words by someone or the honor.

Table 8

*Modal Words of Chinese Corpus*

| Ranking | Modal words (China) | Frequency | Quotation (51) | Frequency | Modal words (Global) | Frequency | Quotation (4) | Frequency |
|---|---|---|---|---|---|---|---|---|
| 1 | can | 3 | Nanjing Uni. | 19 | can | 6 | Stanford Uni. Of | 2 |
| 2 | may | 2 | People's Uni. | 8 | should | 1 | California Berkeley | 2 |
| 3 | would | 2 | Wuhan Uni. | 8 | | | | |
| 4 | could | 1 | Fudan Uni. | 6 | | | | |
| 5 | should | 1 | Tsinghua Uni. | 6 | | | | |
| 6 | | | Zhejiang Uni. | 4 | | | | |

## Discussions and Analyses

In this part, the author shall discuss the above results from discursive dimension on interdiscursivity and manifest intertextuality. According to Bakhtin (1986, p. 94),

> intertextuality is basically the property texts have of being full of snatches of other texts, which may be explicitly demarcated or merged in, and which the text may assimilate, contradict, ironically echo, and so forth. In terms of production, an intertextual perspective stresses the historicity of texts: how they always constitute additions to existing "chains of speech communication". (as cited in Fairclough 1992, p. 84)

Therefore, simply speaking, intertextuality is the mixing of two or more discourses either in an explicit way (manifest intertextuality) or implicit way (interdiscursivity or constitutive intertextuality). Here, in the university introduction discourses, there are a lot of intertextualities, which will be studied as follows. Kristeva observed that intertextuality implies "the insertion of history (society) into a text and of this text into history" (1986, p. 39, as cited in Fairclough, 1992, p. 102). So, let us firstly look at how universities integrate the past history into its current introduction discourses.

First of all, the author shall discuss the constitutive intertextuality or interdiscursivity between historical and introductory discourses. As analyzed above, the uses of past tenses, present perfect and passive voices are mostly concerned with the past history and past achievements of the universities, which really have no direct relationship with the universities today. Yet through the mixing of historical and introductory discourses of universities, by connecting the traditional events or fame as facts with the present, a university creates a feeling

that "we are the same good as we were in the past because we had the tradition of seeking the truth". In this way, the universities carry the same fame or good name with the past. Just as Bhatia says,

> the main intention of placing the two genres within the boundaries of the same article is that such textual proximity is likely to lend the two genres the same factual reliability and hence credibility' that is often presupposed from the use of facts as history (2008, p. 168).

But the fact is that staff, faculty, or organizations do not exist any more and the universities have even changed names for many times; how can they be the same good as they were hundreds of years ago. Moreover, as people's values and outlook have changed a lot, how can one expect the previous universities are judged good in today's people's eyes. Uses of the past, the present perfect, and passive voice are similar in global corpus, but they are not so significant as those in the Chinese corpus. Since the present perfect is a negative keyword in the global corpus.

Furthermore, let us take a look at the quotation part or manifest intertextuality. As studied above, we get the data that there are 51 and four quotations in the Chinese and global corpora respectively. These quotations as indicated by quotation marks are what Authier-Revuz (1982) and Maingueneau (1987) called manifest intertextuality. We are informed that the quotations in the Chinese corpus are mainly concerned with all kinds of projects or programs, labels, honors, and direct quotation of speech. The common feature is that they are all from the third party, and all of them are good comments or authoritative labels, such as "Project 211 or 985". Here, the represented discourses or those that are in quotation marks are translated to the representing discourses of university introductions, the same is that their authority is also translated to the representing discourses to support their position, and thus establishes their credibility from outside voices. In the global corpus, the quotation marks are not that much as in the Chinese ones; on the one hand, it is due to the size, but more important it is owing to the social cultural practice, so in the following part, the author shall discuss the social cultural practice behind this phenomenon.

Why are these different results acquired? On the one hand, it is because of the professional practice. As can be seen from the official websites, nearly all of the "about us" parts of global websites are divided into many sections with available links, such as "History", "Facts", "Statistics", "Students" "Faculty", "Maps", "Tour around the campus", etc. This is also in accordance with the commercialism I the sense that nowadays companies offer more choices and more kinds of products for customers. While Chinese introductory discourses are translated from Chinese versions directly and do not take the intended audience's needs into consideration, thus lacking the international concern for the foreign audiences. On the other hand, it is a result of cultural practice. It is a custom of China that the good will survive in long struggle, so Chinese deeply rely on history and the past to judge whether a person is good or not, a little like word-of-mouth communication. Moreover, Chinese worship the authoritative power, so when the words are from some authorities, they will carry great force and people have a strong tendency to believe. In a word, there are social cultural explanations for this phenomenon.

## Conclusions

This paper discusses the universities introductory discourses from three dimensions, namely the lexico-grammatical dimension using the corpus-based approach, the discursive dimension through intertextuality and interdiscursivity, and the social cultural practice explanations of this phenomenon. Therefore,

conclusions are reached that there are differences in the grammatical features, such as tenses, passive voices, and modal words uses. Secondly, the intertextuality in Chinese universities' discourses is more significant in that they mix introductory discourses with historical ones and integrate third-party quotations into their introductory discourses. Finally, this linguistic phenomenon results from the social cultural reasons, such as the professional practice as well as the cultural custom. Due to the limits of abilities, there are limitations of this study, such as the balance of size of corpora and the using of frequency not relative frequency of quotation cases in Chinese and global corpora.

So, further studies can examine by using balanced corpora and discuss in detail from the cognitive, social, critical, and ethnographic perspectives.

## References

Authier-Revuz, J. (1982). *Heterogeneite monuee et heterogeneite constitutive: Elements pour une approche de I'autre dans le discours* (*Heterogeneity demonstrations and heterogeneity constituent: elements of an approach in the speech*). Germany: Universitätsbibliothek Johann Christian Senckenberg.

Bakhtin, M. (1986). *Speech genres and other late essays*. (M. Holquist Ed., C. Emerson and M. Holquist, Trans.). Austin: University of Texas Press.

Bhatia, V. K. (2004). *Worlds of written discourse: A genre-based view*. London: Continuum International.

Bhatia, V. K. (2008). Genre-analysis, ESP and professional practice. *English for Specific Purposes, 27*, 171, 168.

Fairclough, N. (1989). *Language and power*. London: Longman.

Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.

Foucault, M. (1981). *The archaeology of knowledge*. New York: Pantheon Books.

Kristeva, J. (1986). Word, dialogue and novel. In T. Moi (Ed.), *The Kristeva reader* (pp. 34-61). Oxford: Basil Blackwell.

Maingueneau, D. (1987). *Nouuelles tendances en analyse du discours* (*New tendencies in discourse analysis*). Paris: Hachette.