

Investigating the sustainability of linguistics as a discipline in Mainland China based on the national research funding: A bibliometric analysis (2000-2020)

Lei Gu¹

¹Shanghai Jiao Tong University

October 05, 2022

Abstract

National research funding regulates government research expenditures and thus greatly affects the sustainability of the academic disciplines. This study examined the sustainability of linguistics based on national research funding over the past 20 years (2000-2020). In particular, we used a bibliometric approach to examine the evolutionary trends in the number of topics, important factors, and major research themes in the field of linguistics based on the National Social Science Foundation (NSSF). The main findings are three-fold and summarized as follows. First, a significant increase in awarded grants in linguistics was found over the examined periods. Second, influential factors such as regions, institutions, and academic ranks significantly affected the likelihood of receiving research grants. Third, eight major research themes were identified, along with their evolving trends. These identified research themes were similar to those in previous studies and focused on basic research. The results further indicate: (1) linguistics as an academic discipline in China has undergone sustainable growth over the examined periods; (2) close attention should be given to critical factors that affect the chance of receiving research grants; (3) research grants reflect government's strategic plans to sustain the growth of academic disciplines; (4) research grants can work as a helpful tool for mining research information. Practically, while the present study has taken China as the representative case, the methods in use nevertheless have high ecological validity, thus being applicable to future studies concerning mining research themes from voluminous, diachronic text data. Besides, the current results can also serve as practical references for prospective researchers and applicants in their decision-making process.

**Investigating the sustainability of linguistics as a discipline in Mainland China
based on the national research funding: A bibliometric analysis (2000-2020)**

Lei Gu

koolui@sjtu.edu.cn

School of Foreign Languages, Shanghai Jiao Tong University

Abstract: National research funding regulates government research expenditures and thus greatly affects the sustainability of the academic disciplines. This study examined the sustainability of linguistics based on national research funding over the past 20 years (2000-2020). In particular, we used a bibliometric approach to examine the evolutionary trends in the number of topics, important factors, and major research themes in the field of linguistics based on the National Social Science Foundation (NSSFC). The main findings are three-fold and summarized as follows. First, a significant increase in awarded grants in linguistics was found over the examined periods. Second, influential factors such as regions, institutions, and academic ranks significantly affected the likelihood of receiving research grants. Third, eight major research themes were identified, along with their evolving trends. These identified research themes were similar to those in previous studies and focused on basic research. The results further indicate: (1) linguistics as an academic discipline in China has undergone sustainable growth over the examined periods; (2) close attention should be given to critical factors that affect the chance of receiving research grants; (3) research grants reflect government's strategic plans to sustain the growth of academic disciplines; (4) research grants can work as a helpful tool for mining research information. Practically, while the present study has taken China as the representative case, the methods in use nevertheless have high ecological validity, thus being applicable to future studies concerning mining research themes from voluminous, diachronic text data. Besides, the current results can also serve as practical references for prospective researchers and applicants in their decision-making process.

Keywords: bibliometric analysis; linguistics; research grants; research trends; NSSFC

1. Introduction

Bibliometric analysis (a quantitative analysis that detects systematic patterns from electronic references) assists in discovering research themes and related evolving trends, which in return keeps researchers up to date with fronts of scientific research (Fang et al., 2018; Lei et al., 2020; Lei & Liu, 2019). Empirically, researchers have adopted the bibliometric method and identified research themes for different academic disciplines, including linguistics (e.g., Lei & Liu, 2019), accounting (e.g., Fang et al., 2018), and library science (e.g., Garcia et al., 2019). Previous studies have primarily employed electronic references for extracting bibliographic information, while comparatively fewer studies have taken advantage of the research grants for information inquiry (Jacob & Lefgren, 2011). In fact, research grants are equally informative as electronic references concerning information acquisition (Jiang et al., 2020; Wang et al., 2019).

Moreover, research grants also function as an essential administrative regulation whereby the government distributes financial resources between both collegial and non-collegial research institutions, ultimately steering the national research towards planned directions (Benner & Sandström, 2000; Payne & Siow, 2003). Previous studies have found positive causal effects between financial support and sustainable research productivity (Auranen & Nieminen, 2010; Hicks et al., 2019; Rosenbloom et al., 2015), indicating that financial incentives work to ensure academic sustainability. However, in the face of the economic downturn in the twenty-first century, especially after the outbreak of the COVID-19 pandemic, many countries are facing budget constraints and have to reconsider their budget planning & spending for nonprofit sections, including academia. For example, governments of European countries have continued to cut down their financial support for research and development over the last two decades (OECD, 2020). In particular, governments reduce financial subsidies for colleges but allocate increasing financial support to non-collegial institutions for better market applicability (Payne & Siow, 2003). Moreover,

developed countries, including the United States, continue to spend disproportionately low funding on non-profitable research disciplines, including humanities and arts (National Science Foundation, 2020), consequently hindering the sustainable development of academic research.

By contrast, the Chinese government maintains balanced support for both natural science and social science. National Social Science Fund of China (NSSFC), administered by the State Council since 1986, serves as the most influential and wide-ranging state-level research grants in China to support academic research in humanities, arts, and social sciences. NSSFC accords with governments' policy regarding socio-economic planning and contributes to the sustainable growth of academic disciplines in humanities, arts, and social sciences (NSSFC, 2021). For example, Jiang et al. (2020) examined whether NSSFC oversaw sustainable growth in translation studies in China. They found that NSSFC facilitated translation studies as an academic discipline, as reflected by the sustainable growth in the number of grants, research focuses, and frequently examined issues of research programs. In essence, NSSFC oversees the sustainable development and research innovation in humanities, arts, and social sciences in China.

However, winning NSSFC grants is still a challenging task. Beside meticulous preparation, external factors might also affect the chance of receiving the grants. First of all, researchers from elite universities than non-elite institutions are more likely to be awarded with grants. China classifies its higher education institutions broadly into three categories: 'Project 985' (39 universities), 'Project 211' (about 100 universities), and regular (university not listed in both Projects) (Chan & Mok, 2001; Mok, 2000). It is worth noting that elite universities are often given a preferential allocation of resources (Zong & Zhang, 2019). Besides, the academic ranks of researchers make a difference. Senior researchers often obtain a higher likelihood of winning the grants than junior colleagues, as the former group accumulates more research experiences and academic outputs, consequently making them stand out in a crowd of applicants. Last, geographical regions play a role. Specifically, wealthy regions are more likely than poor regions to gather competent researchers who have higher odds of winning the research grants.

The present study adopted a quantitative bibliometric method to explore the sustainable

growth in linguistics as an academic discipline in Mainland China. Previous studies have primarily adopted qualitative approaches to discover research themes in linguistic research. For example, Sahragard and Meihami (2016) reviewed 428 research articles drawn from three top-ranking applied linguistic journals to analyze the diachronic trends of thematic structures for a range of 40 years (1975-2015). In a similar vein, Wang and Zhou (2018) conducted a full-scale qualitative review regarding systematic-function linguistics in China from 2010 to 2016. They manually identified eight representative research themes (theoretical explorations, syntax, discourse analysis, studies of Chinese, translation studies, language comparison, appraisal theory, grammatical metaphor, and others). In the meantime, the literature also records studies that employed bibliometric techniques to extract research themes and analyze their corresponding trends. For example, Zhang (2020) examined the field of second language acquisition (1997-2018) with the keywords and co-citations/citations of articles from 15 top SLA journals. With comparable methods, Huan and Guan (2020) conducted a bibliometric analysis of discourse analysis between 1979 to 2018, and identified the essential research themes including ‘corpus linguistics’, ‘digital conversation analysis’, ‘discursive news values approach’, ‘membership categorization analysis’, ‘multimodal analysis’, and ‘social media. In addition, past studies have primarily used the time-series methods to depict the evolving trends of discovered themes. For example, Lin and Lei (2020) adopted the first-order autoregressive model to predict the publication trend in linguistic research in the examined decades (2000-2019).

To summarize, previous studies have made contributions to the bibliometric analysis of linguistics but still have some limitations in their scopes and methods. To begin with, qualitative methods that primarily feature human justification and judgment are infeasible in the face of massive unstructured text data, as human-dependent methods are both labor-intensive and error-prone. Besides, previous studies have primarily focused on electronic references for bibliometric analysis, while few sketch the scientific landscape using research grants which provide equally valuable bibliographic information (Jiang et al., 2020). In this case, more efforts are required to examine the relations between research funds and the sustainability of academic disciplines when influential factors are simultaneously considered. Hence, the present study examines the sustainable growth and fronts of linguistics as an

academic discipline in China by applying a bibliometric method to national research grants. In particular, the present study sets out to address three research questions.

1. What are the evolving trends for the number of grants in the area of linguistics?
2. What are the systematic variations regarding the likelihood of receiving grants by regions, institutions and academic ranks?
3. What are the major research themes and progressive trends in the area of linguistics?

2. Methods

2.1 Data collection

NSSFC releases separate lists of awarded research grants annually, including the ‘major subsidized projects’, ‘basic subsidized projects’, and ‘specialized projects for young scholars’ (it should be noted that there are other types of projects released by the NSSFC, but the projects selected are the most representative of the provinces.). As the present study aims to examine the sustainable development in linguistics from national research funding, we thus decided to focus on the three lists of subsidized projects in the area of linguistics. In practice, we first ran a self-written Python script to acquire the complete grant information of the three subsidized projects in linguistics from the NSSFC official database (<http://fz.people.com.cn/skygb/sk/index.php/index/index/>) between 2000 and 2020 (the database had its last update in 2020). Apart from the titles of projects and awarding years, the data acquisition also collected other necessary information of the selected projects, including information about the applicants (e.g., institutions, regions, and academic ranks). Then, the collected data were stored in an Excel 2007 spreadsheet for follow-up data processing and analysis.

2.2 Processing and analysis

To capture the changes in the number of grants over time, we charted the growth of the total number of projects over the past twenty years (2000-2020). In addition, we tabled and visualized the changes of grant number by grants types (‘major subsidized projects’, ‘basic subsidized projects’, ‘specialized projects for young scholars’), regions (east, middle, west,

northeast), by institutions ('Project 985', 'Project 211', regular), and by academic ranks (full professor, associate professor, lecturer, teaching assistant). The regions were categorized as per China's National Bureau of Statistics (National Bureau of Statistics, 2020), which divides China's 31 provinces into four geographic regions: east ($N = 10$), middle ($N = 6$), northeast ($N = 3$), west ($N = 12$). For institutions, it is worth noting that '985 Project' lists a total of 39 universities that concurrently appear in 'Project 211'. To avoid double-counting, we removed the same universities from 'Project 211'. In addition, it merits attention that lecturer in China is the entry-level rank for tenured faculty members, equivalent to the rank of assistant professor at universities in the United States. Moreover, considering raw frequencies do not adjust for unequal design (e.g., regions with more universities tend to receive more grants), we applied weights to the raw data. In practice, we calculated the normalized frequency by dividing the raw frequencies against the number of candidates for each factor. For example, if the eastern region consists of ten provinces with 134 'major subsidized projects' in total, the normalized frequency equals 13.4 (134 divided by 10).

Moreover, as we assumed systematic variations in the likelihood of winning grants, we performed multiple multinomial logistic regressions on the associations between the grants and the influential factors. Multinomial logistic regression was appropriate for the present study in that the predictors (region, institution categories, academic ranks) were categorical, while the outcome variable (levels of grants) was of non-ordinal categories (Field, 2009). Besides, we took the year of awards as a covariate to better control the time effect. The alpha value was set at 0.05 for all statistical tests to determine whether the results reached statistical significance. We reported the odds ratio to measure the magnitude of the associations between the predictors and the outcome variable.

In order to identify research themes, the present study employed the non-negative matrix factorization (NMF) (Fu et al., 2021; Ray et al., 2019), a matrix factorization-based topic modeling approach, to obtain latent thematic structures from the titles of the projects. Following previous quantitative studies that analyzed the research trends in academic disciplines (e.g., Fang et al., 2018; Wang et al., 2019), we likewise ran a self-written Python script to remove non-Chinese symbols from the titles of projects, such that the cleaned text was appropriate for subsequent tokenization (tokenization is the splitting of text into

individual tokens, each representing a term). Later, we tokenized the cleaned text with the Jieba library in Python. There were a total of two rounds of tokenization. In the first round, we tokenized the cleaned text as per a customized user dictionary that comprised keywords of 17,921 published articles concerning linguistics research (these articles were likewise published between 2000 and 2021 and were all listed by Chinese Social Science Citation Index/CSSCI). As the user dictionary might not contain all technical terms regarding project titles in linguistics, the first round was then supplemented with a second round of tokenization that added the unidentified terms to the customized dictionary. Last, we removed the stopwords that did not contain concrete meanings (e.g., auxiliary words).

After text cleaning, we applied the term frequency-inverse document frequency (TF-IDF) vectorizer to the cleaned titles and then conducted the matrix factorization decomposition by using the scikit-learn library in Python. We adopted the deterministic non-negative double singular value decomposition initialization to handle the instability from random initialization (Boutsidis & Gallopoulos, 2008). To assess the correct number of themes, we created word vectors using a word2vec model (Mikolov et al., 2013) with the Gensim library in Python. The k (number of themes) was kept between a minimum of five ($k_{\min} = 4$) and a maximum of fifteen ($k_{\max} = 15$) for a more reasonable interpretation. The selection process kept iterating and finally found the final k with the highest mean coherence score. The research themes were then described by the top 10 descriptors found for each theme (this action was also for more reasonable explanation).

Once the topic model successfully identified the research themes, a normalized frequency was computed to depict the changes of the identified research themes by dividing the raw frequency of a given theme in a given year against the number of all themes in that year. According to Lei et al. (2020), the normalized frequency ensures comparability across years since the raw frequency of a given theme tends to fluctuate according to the total number of themes in a given year. For modeling the changes of thematic structures, we did not use autoregressive models as previous studies did (Fang et al., 2018; Lei et al., 2020) because autoregression handles autocorrelated data that features self-repetition in fixed time intervals (e.g., stock prices). For the present study, the research themes were based on high-level summaries that were ‘hidden’ to researchers, making autoregressive models

methodologically inappropriate. Instead, we depicted variation tendencies of research themes with the local regression with a smoothing function (LOWESS) that fitted localized subsets of data (Isnanto, 2011).

$$\text{normalized frequency} = \frac{\text{raw frequency of a theme in a given year}}{\text{total number of all themes in a given year}}$$

3. Results and discussion

3.1 Evolving trends for the number of grants

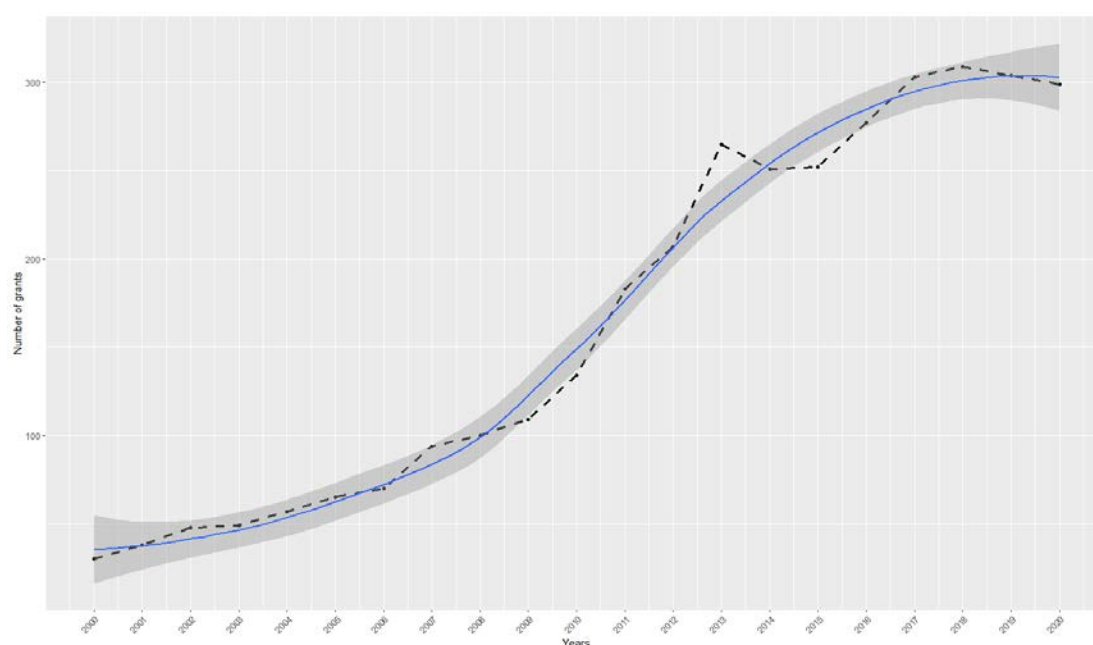
The data collection process selected a total of 3,444 awarded grants in linguistics over the last two decades. **Table 1** presents the evolving trend of the total number of grants over the examined periods. The growth of research grants is also illustrated in **Figure 1**. Following Lin and Lei (2020), a linear regression model was fit, and the results showed that the examined decades significantly predicted the total number of grants awarded per year ($F(1,19) = 341, p < .001$, adjusted $R^2 = 0.94$). In particular, the positive coefficient ($\beta = 16.53, SE = 0.90, t = 18.47, p < 0.001$) showed that as years raised by one unit, the total number of grants increased by 16 units. Such a stable upward trend in the total number of grants (despite a minor dip in recent years) further indicates that linguistics has been receiving closer attention over time. The results also reveal that national research funding contributes to the sustainable growth of linguistics as an academic discipline in China, as an increasing number of research programs have been receiving financial supports from the government.

Table 1. Total number of awarded grants per year

Year	Number of Grants
2020	299
2019	304
2018	309
2017	303
2016	277
2015	252
2014	251
2013	265
2012	207

2011	183
2010	134
2009	109
2008	100
2007	94
2006	70
2005	65
2004	57
2003	49
2002	48
2001	38
Total	3444

Figure 1. *Evolving trends for the total number of awarded grants over time*



Besides the trends of the total number, we present in **Table 2** and **Figure 2** the annual sum of grants by different categories (level of grants, regions, institutions, and academic ranks). Similar to the total number of awarded grants, we conducted fourteen linear regressions on the relationships between year of awarding and the number of awarded grants by particular categories. The statistical results are summarized in **Table 3**. On the whole, there were upward trends for all categories (despite some slight twists), and all trends reached statistical significance. These results further reveal: (1) the government's continued investments in linguistics; (2) country-level strategic plans to encourage basic research despite limited market applicability; (3) sustainable growth in linguistics as an academic discipline.

Table 2. Yearly number of grants by different categories

Year	Grant			Region				Institution			Rank			
	Major	Basic	Young	East	Middle	West	NE	P-985	P-211	Regular	FP	AP	LT	TA
2020	28	219	52	172	48	60	19	76	61	162	122	101	68	1
2019	20	233	51	165	51	67	21	62	71	171	119	117	66	2
2018	26	229	54	168	61	58	22	70	61	178	116	121	71	1
2017	22	218	63	156	55	64	28	65	79	159	112	112	78	1
2016	20	193	64	149	55	50	23	62	52	163	90	106	78	1
2015	12	189	51	141	52	41	18	62	58	132	99	103	49	0
2014	20	172	59	144	49	37	21	62	45	144	107	88	54	1
2013	17	160	88	152	53	43	17	59	54	152	109	98	51	0
2012	9	133	65	107	39	46	15	40	49	118	102	67	36	0
2011	9	116	58	104	32	32	15	44	51	88	92	63	28	0
2010	3	93	38	82	22	19	11	41	32	61	79	33	21	0
2009	0	74	35	69	17	20	3	25	29	55	52	41	16	0
2008	2	70	28	62	19	13	6	30	24	46	58	30	11	0
2007	5	65	24	51	22	12	9	32	26	36	54	32	7	0
2006	2	52	16	48	11	9	2	28	21	21	42	19	9	0
2005	2	51	12	34	19	7	5	18	18	29	35	26	4	0
2004	2	43	12	33	10	11	3	20	14	23	33	16	7	0
2003	1	40	8	32	7	7	3	18	14	17	31	14	4	0
2002	1	39	8	24	12	10	2	17	11	20	28	18	1	0
2001	3	29	6	29	1	6	2	17	11	10	25	9	3	0
2000	0	23	7	21	4	2	3	13	8	9	14	11	5	0
Total	204	2441	799	1943	639	614	248	861	789	1794	1519	1225	667	7

Note. NE = northeast; P-985 = ‘Project 985’; P-211 = ‘Project 211’; FP = full professor; AP = associate professor; LT = lecturer; TA = teaching assistant

Figure 2. Evolution in the number of awarded grants by different categories over time

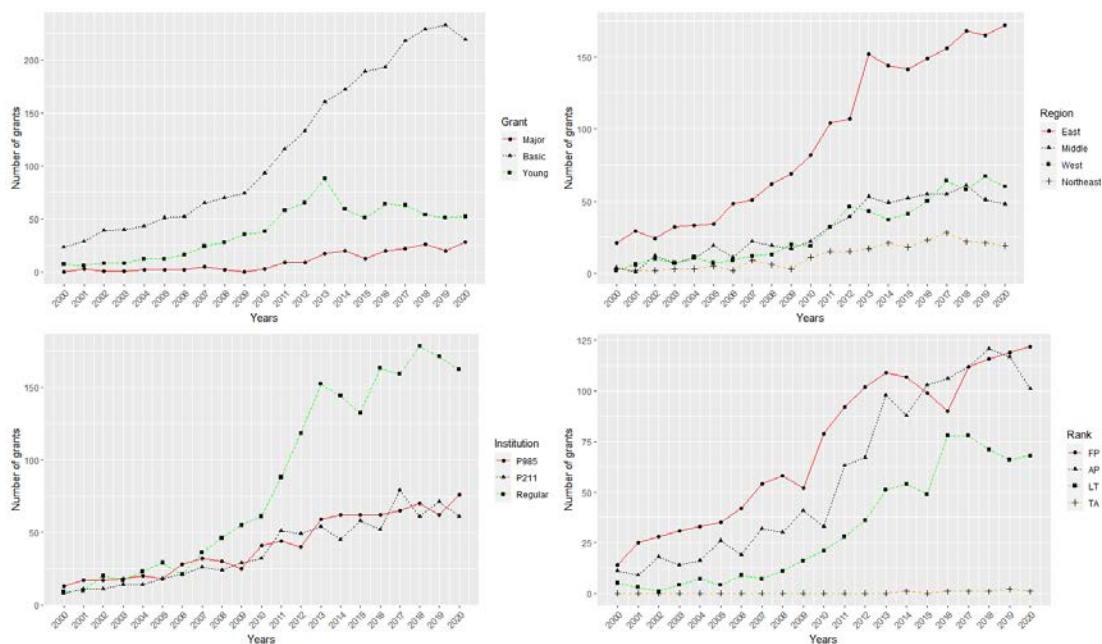


Table 3 Results of linear regressions on the number of grants by categories over time

		β	SE	t	p	F	p	Adj R^2
Grants	Major	1.39	0.15	9.19	< 0.001	84.38	< 0.001	0.81
	Basic	11.80	0.61	19.28	< 0.001	371.61	< 0.001	0.95
	Young	3.35	0.49	6.87	< 0.001	47.20	< 0.001	0.70
Regions	East	8.83	0.46	19.04	< 0.001	362.56	< 0.001	0.95
	Middle	3.06	0.25	12.25	< 0.001	150.01	< 0.001	0.88
	West	3.38	0.24	13.78	< 0.001	189.98	< 0.001	0.90
	NE	1.27	0.13	10.04	< 0.001	100.87	< 0.001	0.83
Institutions	P-985	3.27	0.20	16.40	< 0.001	268.90	< 0.001	0.93
	P-211	3.38	0.24	13.97	< 0.001	195.13	< 0.001	0.91
	Regular	9.88	0.66	15.01	< 0.001	225.32	< 0.001	0.92
Ranks	FP	5.73	0.36	16.08	< 0.001	258.45	< 0.001	0.93
	AP	6.33	0.46	13.79	< 0.001	190.13	< 0.001	0.90
	LT	4.29	0.36	11.82	< 0.001	139.74	< 0.001	0.87
	TA	0.07	0.01	4.79	< 0.001	22.96	< 0.001	0.52

Note. NE = northeast; P-985 = ‘Project 985’; P-211 = ‘Project 211’; FP = full professor; AP = associate professor; LT = lecturer; TA = teaching assistant

3.2 The systematic variations in the types of grants

On account of external effects on the sustainable growth of research grants, we examined to what extent major factors such as regions (east, middle, northeast, west), institutions (‘Project 985’, ‘Project 211’, regular), and academic ranks (full professor, associate professor, lecturer, teaching assistant) affected the normalized number of grants awarded and the likelihood of receiving different types of grants (‘major subsidized projects’, ‘basic subsidized projects’, ‘specialized projects for young scholars’).

Table 4 presents the normalized number of grants received by different categories. Four general patterns were detected for the distribution of awarded grants. First, economically developed regions received more grants than economically less developed regions. According to **Figure 3**, the eastern region came first in the number of grants received, followed by the middle, northeastern, and western regions. Second, elite universities received more grants than less elite institutions. For instance, universities listed in the ‘985 Project’ received more grants than universities listed otherwise. Third, applicants with higher academic ranks were awarded more grants than lower-ranked researchers. For example, full professors were awarded the most grants, whereas teaching assistants received the fewest grants. However, it

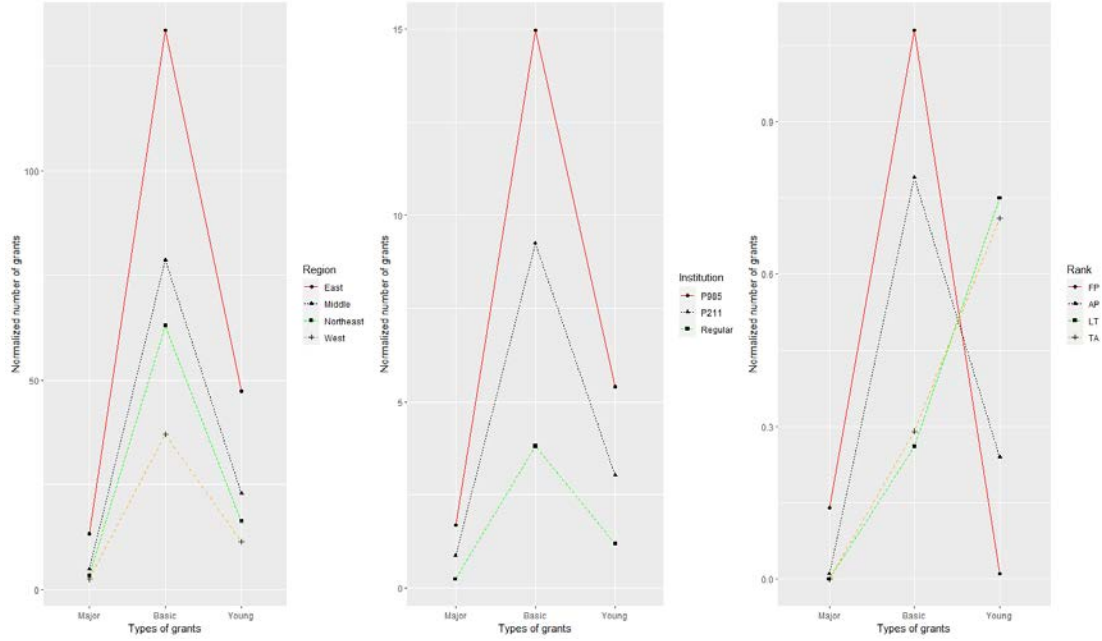
is interesting to note that junior researchers such as lecturers and teaching assistants reaped more ‘specialized projects for young scholars’. This result might owe to the NSSFC’s regulations that applicants for the specialized projects must be under 35 years old (NSSFC, 2021), which senior researchers are often not qualified for. Fourth, the ‘basic subsidized projects’ outnumbered other types of grants in most situations (except for young scholars applying for the specialized projects), indicating that basic projects function as the mainstay for the national research funding.

Table 4. *Normalized numbers for grants by different categories*

		Types of grants		
	Predictor	Major	Basic	Young
Regions	East	13.40	133.50	47.40
	Middle	4.83	78.67	23.00
	Northeast	3.33	63.00	16.33
	West	2.58	37.08	11.50
Institutions	P-985	1.69	14.97	5.41
	P-211	0.87	9.25	3.03
	Regular	0.25	3.81	1.19
Ranks	FP	0.14	1.08	0.01
	AP	0.01	0.79	0.24
	LT	0.00	0.26	0.75
	TA	0.00	0.29	0.71

Note. P-985 = ‘Project 985’; P-211 = ‘Project 211’; FP = full professor; AP = associate professor; LT = lecturer; TA = teaching assistant

Figure 3. *Normalized numbers of grants by different categories*



Multiple logistic regression analyses were performed on the associations between types of grants and different factors to examine systematic variations in the likelihood of receiving research grants. **Table 5** presents the results of the logistic regression analyses run on the associations between types of grants and regions. The odds statistics showed that in comparison to more affluent regions that received ‘major subsidized projects’, less developed regions had a higher probability of receiving ‘basic subsidized projects’ and ‘specialized projects for young scholars’ (odds ratios were larger than one). In addition, western regions were less likely than northeastern regions to receive ‘basic subsidized projects’ (odds ratios were less than one). However, when baseline for comparison was the ‘basic subsidized projects’, western regions were 1.19 times more likely to receive ‘specialized projects for young scholars’ than northeastern regions. These results indicate that regions with less economic growth have a higher probability of receiving basic projects and specialized projects than more competitive projects (e.g., major subsidized projects).

Table 5. Results of logistic regression analyses (regions) on the likelihood of different grant types

Grant types	Predictor	β	SE	Z	p	OR
Basic vs. Major	Middle vs. East	0.46	0.05	8.54	< 0.001	1.59
	Northeast vs. Middle	0.24	0.04	6.33	< 0.001	1.27

Young vs. Major	West vs. Northeast	-0.26	0.04	-6.84	< 0.001	0.77
	Middle vs. East	0.28	0.05	5.62	< 0.001	1.33
	Northeast vs. Middle	0.12	0.03	3.42	< 0.001	1.13
Young vs. Basic	West vs. Northeast	-0.09	0.04	-2.42	0.016	0.92
	Middle vs. East	-0.18	0.10	-1.69	0.09	0.84
	Northeast vs. Middle	-0.11	0.07	-1.59	0.112	0.89
	West vs. Northeast	0.18	0.07	2.40	0.016	1.19

Note. OR = odds ratio; Model χ^2 (8) = 38.60, $p < 0.001$

Table 6 displays the results of the logistic regression analyses run on the associations between types of grants and institutions. The results displayed no systematic grants variations between the two elite university groups (coefficients were all insignificant). By contrast, regular schools were more likely to receive less competitive research grants than elite schools (odds ratios were larger than one). However, the differences between regular and elite schools were somehow negligible regarding the basic and specialized grants. These results suggest that elite schools are more likely to receive grants (Zong & Zhang, 2019), but such a privilege is limited to more competitive projects (e.g., ‘major subsidized projects’).

Table 6. Results of logistic regression analyses (institutions) on the likelihood for different grant types

Grant type	Predictor	<i>B</i>	<i>SE</i>	<i>Z</i>	<i>p</i>	OR
Basic vs. Major	P-211 vs. P-985	0.05	0.06	0.82	0.41	1.05
	Regular vs. P-211	0.40	0.05	7.59	< 0.001	1.50
Young vs. Major	P-211 vs. P-985	-0.06	0.06	-1.01	0.31	0.95
	Regular vs. P-211	0.38	0.05	7.79	< 0.001	1.46
Young vs. Basic	P-211 vs. P-985	-0.10	0.12	-0.86	0.39	0.90
	Regular vs. P-211	-0.02	0.10	-0.17	0.86	0.98

Note. P-211 = ‘Project 211’; P-985 = ‘Project 985’; OR = odds ratio; Model χ^2 (6) = 20.98, $p = 0.002$

Table 7 presents the results of the logistic regression analyses run on the associations between types of grants and academic ranks. The log ratios of grants were compared between researchers of different academic ranks. The results showed that all odds ratios were larger than one, indicating that when the baseline for comparison are more competitive grants (e.g., ‘major subsidized projects’), lower ranked researchers have a higher probability of receiving basic grants and specialized grants than higher ranked researchers.

Table 7. Results of logistic regression analyses (academic ranks) on the likelihood for different grant types

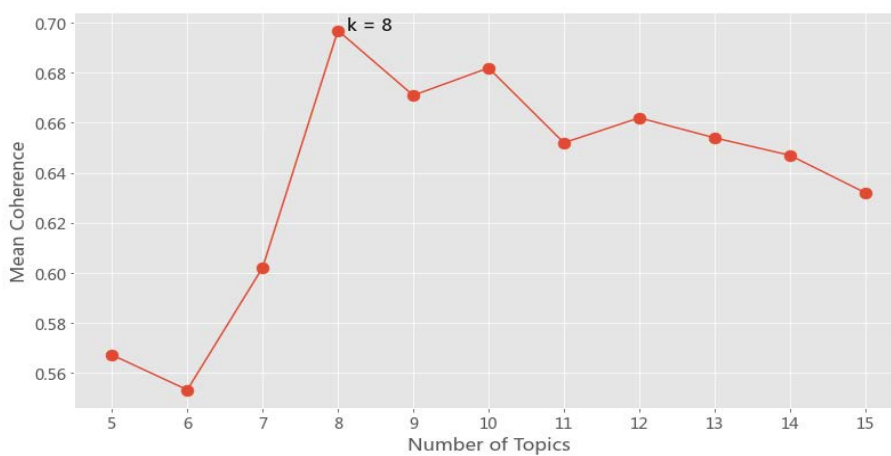
Grant type	Predictor	<i>B</i>	<i>SE</i>	<i>Z</i>	<i>p</i>	OR
Basic vs. Major	AP vs. FP	2.34	0.03	78.52	< 0.001	10.36
	LT vs. AP	1.07	0.06	18.83	< 0.001	2.91
	TA vs. LT	4.51	0.01	8507.90	< 0.001	90.90
Young vs. Major	AP vs. FP	6.23	0.03	216.04	< 0.001	510.16
	LT vs. AP	3.65	0.06	64.43	< 0.001	38.49
	TA vs. LT	4.70	0.01	8868.30	< 0.001	110.60
Young vs. Basic	AP vs. FP	3.89	0.06	66.84	< 0.001	48.94
	LT vs. AP	2.60	0.11	22.85	< 0.001	13.40
	TA vs. LT	0.17	0.00	148.81	< 0.001	1.19

Note. FP = full professor; AP = associate professor; LT = lecturer; TA = teaching assistant; OR = odds ratio; Model χ^2 (8) = 1784.27, $p < 0.001$

3.3 Major research themes and corresponding trends

Our word2vec model calculated the coherence scores for each number in k . As shown in **Figure 4**, the most appropriate number of topics was $k = 8$, as when $k = 8$, the highest coherence score was reached, and a further rise in k did not result in improved coherence scores. In this case, there should be eight topics for linguistic research over the examined period (2000-2020) based on the 3,444 titles of projects from the approved grants of National research funding.

Figure 4 Coherence scores and number of topics



The identified research topics were summarized in **Table 8**, together with the top 10 associated descriptors and total counts. Based on the raw counts of topics, the most frequent

topic was Topic 1 (N = 820) while the least frequent topic was Topic 2 (N = 212). The eight research topics covered a wide range of research focuses, including Topic 1 (Chinese language, typology, lexicalization, and historical evolution of language), Topic 2 (contrastive study between English and Chinese), Topic 3 (English-Chinese translation and English language usage in China), Topic 4 (comparative study of grammar in linguistic varieties), Topic 5 (syntax-semantic analysis of modern Chinese), Topic 6 (language typology, language contact, and language evolution), Topic 7 (database building and language research), Topic 8 (corpus-based language research).

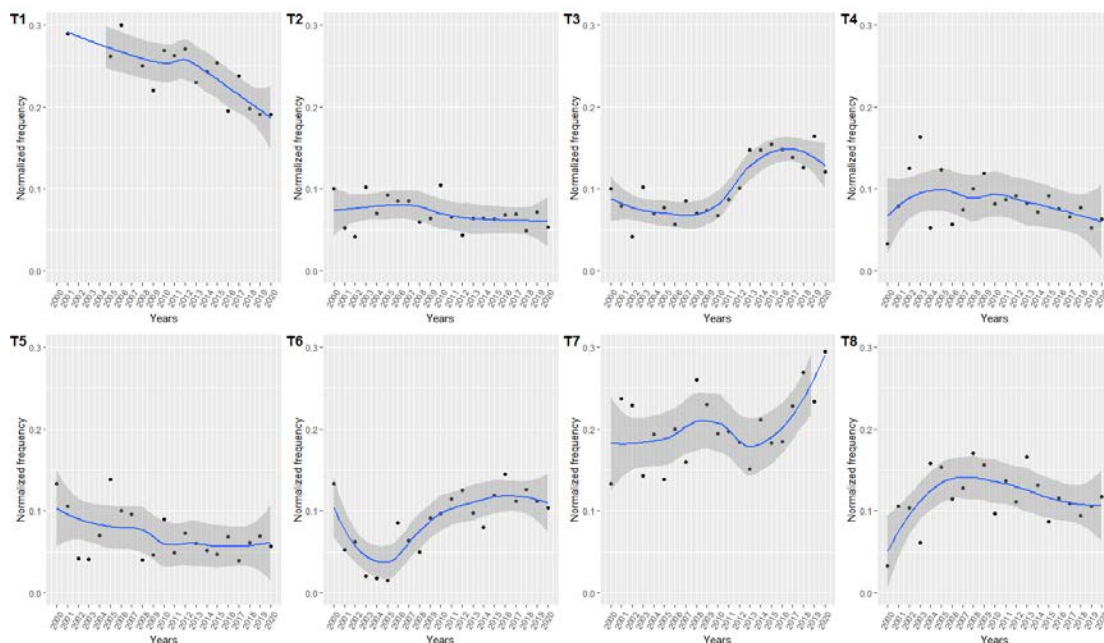
Table 8 *Research topics identified*

Topics	Top 5 terms describing the topics	Count
Topic 1	Chinese, lexicalization, typology, construction, historical evolution	820
Topic 2	Contrastive study, Chinese-English, function, cognition, semantics	212
Topic 3	China, English translation, English language, domestic, language use	432
Topic 4	Comparative study, linguistic varieties, grammar, investigation, features	284
Topic 5	Modern Chinese, syntax-semantics, structures, constructions, events	221
Topic 6	Language typology, contact, variation, evolution, grammar study	355
Topic 7	Database, literature, vocabulary research, text analysis, language	730
Topic 8	Corpus-based, cognitive research, English, discourse, novels	390

To depict the thematic changes of the eight research topics over time, we employed the normalized number of grants (as shown in **Figure 4**). The results showed one topic with evident rising trends (Topic 7), three topics with clear declining trends (Topic 1, Topic 3, Topic 4), and four topics with constant trends (Topic 2, Topic 5, Topic 6, Topic 8). The results indicate that despite sustainable growth in the total number of research grants (cf **Figure 1** and **Figure 2** in subsection 3.1), research themes found in grants nonetheless have their own developmental tracks, which calls for a close inspection into their idiosyncratic variations. Besides, most identified research themes focus on basic research in linguistics instead of practical application, which reflects the government's strategic plans to maintain the sustainable growth of fundamental research that nevertheless promises less market applicability (National Science Foundation, 2020). Finally, these results also offer valuable references for prospective researchers and applicants when making decisions for research and

applications, which contributes to the sustainable growth of linguistics as an academic discipline (Jiang et al., 2020; Wang et al., 2019).

Figure 4 *Topics trends in the examined decades*



In addition, compared to previous studies that examined linguistic research topics with electronic references, the present study showed that research grants could offer equally useful bibliographic information. For example, Topic 1 (comparative study of grammar in linguistic varieties and between Chinese & English languages) was close to previous study that reviewed systematic-function linguistics in China (e.g., Wang & Zhou, 2018). In a similar vein, Topic 8 (corpus-based language research) echoed the research themes ‘corpus linguistics’, ‘multimodal analysis’ identified by previous discourse analyses (e.g., Huan & Guan, 2020). In sum, these results indicate that linguistic research in China maintains sustainable growth as an academic discipline over time while keeping pace with international trends at the same time.

4. Conclusion

The present study intended to address to what extent national research funding contributed to the sustainable development of academic disciplines. Specifically, it explored

the sustainability of linguistics as an academic discipline based on the National Social Science Funding of China. Methodologically, we collected a total of 3,444 national research grants and measured their evolution over the last two decades (2000-2020). Besides, we also inspected the systematic variations of research grants by inspecting how major factors affected the likelihood of receiving grants. In addition, we obtained the underlying research themes by mining the titles of the items and describing their changes over time. Consequently, we found that linguistic research in China generally maintained sustainable growth over the past twenty years, as reflected by the rising trends in the number of awarded research grants. We also found that key factors such as regions, institutions, and academic ranks did impact the likelihood of receiving grants, with developed regions, elite universities, and senior researchers obtaining a higher chance of being awarded. Last, we identified eight research themes with different evolving patterns and noticed that most topics focused on basic research in linguistics. These identified research topics resembled thematically those from previous studies. In all, these results indicate: (1) sustainable growth in linguistics as an academic discipline in China; (2) further efforts are needed for balancing research funding across different regions, institutions, and researchers; (3) research grants can reflect governments' strategic plans to maintain sustainable growth of academic disciplines; (4) research grants can advance our knowledge in academic disciplines while providing adequate, accurate, and up-to-date information for bibliometric analysis and decision-making.

References

- Auranen, O., & Nieminen, M. (2010). University research funding and publication performance—An international comparison. *Research Policy*, 39(6), 822-834. <https://doi.org/10.1016/j.respol.2010.03.003>
- Benner, M., & Sandström, U. (2000). Institutionalizing the triple helix: Research funding and norms in the academic system. *Research Policy*, 29(2), 291-301. [https://doi.org/10.1016/S0048-7333\(99\)00067-0](https://doi.org/10.1016/S0048-7333(99)00067-0)
- Boutsidis, C., & Gallopoulos, E. (2008). SVD based initialization: A head start for nonnegative matrix factorization. *Pattern Recognition*, 41(4), 1350-1362. <https://doi.org/10.1016/j.patcog.2007.09.010>
- Chan, D., & Mok, K.-H. (2001). Educational reforms and coping strategies under the tidal wave of marketisation: A comparative study of Hong Kong and the mainland. *Comparative Education*, 37(1), 21-41. <https://doi.org/10.1080/03050060020020417>
- Fang, D., Yang, H., Gao, B., & Li, X. (2018). Discovering research topics from library electronic references using latent Dirichlet allocation. *Library Hi Tech*, 36(3), 400-410.

<https://doi.org/10.1108/LHT-06-2017-0132>

- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Sage.
- Fu, Q., Zhuang, Y., Gu, J., Zhu, Y., & Guo, X. (2021). Agreeing to disagree: Choosing among eight topic-modeling methods. *Big Data Research*, 23, 100173. <https://doi.org/10.1016/j.bdr.2020.100173>
- Garcia, P., Lueck, J., & Yakel, E. (2019). The pedagogical promise of primary sources: Research trends, persistent gaps, and new directions. *The Journal of Academic Librarianship*, 45(2), 94-101. <https://doi.org/10.1016/j.acalib.2019.01.004>
- Hicks, D. J., Coil, D. A., Stahmer, C. G., & Eisen, J. A. (2019). Network analysis to evaluate the impact of research funding on research community consolidation. *PLoS ONE*, 14(6), e0218273. <https://doi.org/10.1371/journal.pone.0218273>
- Huan, C., & Guan, X. (2020). Sketching landscapes in discourse analysis (1978–2018): A bibliometric study. *Discourse Studies*, 22(6), 697-719. <https://doi.org/10.1177/1461445620928814>
- Isnanto, R. (2011). Comparison on several smoothing methods in nonparametric regression. *Jurnal Sistem Komputer*, 1(1), 41-48.
- Jacob, B. A., & Lefgren, L. (2011). The impact of research grant funding on scientific productivity. *Journal of Public Economics*, 95(9-10), 1168-1177. <https://doi.org/10.1016/j.jpubeco.2011.05.005>
- Jiang, Z., Wu, Y., & Tsung, L. (2020). National Research Funding for Sustainable Growth in Translation Studies as an Academic Discipline in China. *Sustainability*, 12(18), 7241. <https://doi.org/10.3390/su12187241>
- Lei, L., Deng, Y., & Liu, D. (2020). Examining research topics with a dependency-based noun phrase extraction method: A case in accounting. *Library Hi Tech*. <https://doi.org/10.1108/LHT-12-2019-0247>
- Lei, L., & Liu, D. (2019). Research trends in applied linguistics from 2005 to 2016: A bibliometric analysis and its implications. *Applied Linguistics*, 40(3), 540-561. <https://doi.org/10.1093/applin/amy003>
- Lin, Z., & Lei, L. (2020). The Research Trends of Multilingualism in Applied Linguistics and Education (2000–2019): A Bibliometric Analysis. *Sustainability*, 12(15), 6058. <https://doi.org/10.3390/su12156058>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv:1301.3781*.
- Mok, K. H. (2000). Impact of globalization: A study of quality assurance systems of higher education in Hong Kong and Singapore. *Comparative Education Review*, 44(2), 148-174. <https://doi.org/10.1086/447601>
- National Bureau of Statistics. (2020). 2020 年 1—6 月份全国房地产开发投资和销售情况. Retrieved on 27 July 2021 from http://www.stats.gov.cn/xxgk/sjfb/zxfb2020/202007/t20200716_1776285.html
- National Science Foundation. (2020). Fiscal year 2021 budget requests to congress. Retrieved on 27 July 2021 from <https://www.nsf.gov/pubs/2020/nsf20005/nsf20005.pdf>
- NSSFC. (2021). Application instructions on NSSFC Funding: 2010 - 2021. Retrieved on 27 July 2021 from <http://www.nopss.gov.cn/GB/219469/431027/index.html>
- OECD. (2020). Main science and technology indicators. Retrieved on 27 July 2021 from <https://www.oecd.org/sti/msti.htm>
- Payne, A. A., & Siow, A. (2003). Does federal research funding increase university research output? *Advances in Economic Analysis & Policy*, 3(1), 1018-1018. <https://doi.org/10.2202/1538-0637.1018>
- Ray, S. K., Ahmad, A., & Kumar, C. A. (2019). Review and implementation of topic modeling in Hindi. *Applied Artificial Intelligence*, 33(11), 979-1007. <https://doi.org/10.1080/08839514.2019.1661576>
- Rosenbloom, J. L., Ginther, D. K., Juhl, T., & Heppert, J. A. (2015). The effects of research & development funding on scientific productivity: Academic chemistry, 1990-2009. *PLoS ONE*, 10(9), e0138176. <https://doi.org/10.1371/journal.pone.0138176>
- Sahragard, R., & Meihami, H. (2016). A diachronic study on the information provided by the research titles of

- applied linguistics journals. *Scientometrics*, 108(3), 1315-1331.
<https://doi.org/10.1007/s11192-016-2049-4>
- Wang, J., Fan, Y., Feng, L., Ye, Z., & Zhang, H. (2019). Research hotspot prediction and regular evolutionary pattern identification based on NSFC grants using NMF and semantic retrieval. *IEEE Access*, 7, 123776-123787. <https://doi.org/10.1109/ACCESS.2019.2938393>
- Wang, Y., & Zhou, Y. (2018). Systemic-functional linguistics in China (2010–2016). *WORD*, 64(1), 9-37. <https://doi.org/10.1080/00437956.2018.1427085>
- Zhang, X. J. S. i. S. L. A. (2020). A bibliometric analysis of second language acquisition between 1997 and 2018. *Studies in Second Language Acquisition*, 42(1), 199-222. <https://doi.org/10.1017/S0272263119000573>
- Zong, X., & Zhang, W. (2019). Establishing world-class universities in China: Deploying a quasi-experimental design to evaluate the net effects of Project 985. *Studies in Higher Education*, 44(3), 417-431. <https://doi.org/10.1080/03075079.2017.1368475>