



Title	SNSデータを用いた中国語圏の生物多様性への関心の調査と比較
Author(s)	
Citation	令和3（2021）年度学部学生による自主研究奨励事業研究成果報告書. 2022
Version Type	VoR
URL	https://hdl.handle.net/11094/85628
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka

令和3年度大阪大学未来基金「学部学生による自主研究奨励事業」研究成果報告書

ふりがな 氏名	しゅう げつゆ ZHOU YUERU	学部 学科	工学部 環境・エネルギー工学科	学年	2年
ふりがな 共同 研究者氏名	さとう たいが 佐藤 大雅	学部 学科	工学部 環境・エネルギー工学科	学年	2年
	よしか まさたか 吉賀 正峻		歯学部 歯学科		2年
	かわかみ さとる 川上 理		理学部 化学科	学年	2年
	む えそら MU JIANGTIAN		基礎工学部 情報工学科		1年
アドバイザー教員 氏名	まつい たかのり 松井 孝典	所属	工学研究科		
研究課題名	SNS データを用いた中国語圏の生物多様性への関心の調査と比較				
研究成果の概要	研究目的、研究計画、研究方法、研究経過、研究成果等について記述すること。必要に応じて用紙を追加してもよい。(先行する研究を引用する場合は、「阪大生のためのアカデミックライティング入門」に従い、盗作剽窃にならないように引用部分を明示し文末に参考文献リストをつけること。)				

Purpose

Biodiversity includes genetic diversity, species diversity, and ecosystem diversity, and it is from this diversity that we enjoy the blessings of nature and live a quality life. Nature's blessings from biodiversity have supported food production and pharmaceutical development, purified water and soil, and satisfied our aesthetic needs. However, over the past 50 years, biodiversity and nature have been deteriorating at an unprecedented rate due to direct and indirect factors. The direct factors contributing to the deterioration of nature are (in descending order of impact): changes in land and sea use, direct use of organisms, climate change, pollution, and invasion of alien species. Indirect (underlying) factors that affect these five direct factors include production and consumption patterns resulting from human social values and behaviors, demography and population transition, trade, technological innovation, and local to global governance. To achieve the SDG indicators for 2030 and the Aichi Targets for 2050, transformative change in economic, social, political, technological and underlying human values is indispensable. In order to create measures to bring about a change in human values and to promote a change in lifestyle, it is necessary to conduct a survey on people's awareness of biodiversity. A method of analyzing big data in social networking services and examining users' thoughts can be considered for awareness surveys.¹⁾

According to a paper by Ishida et al.²⁾ of our department, English-speaking and Japanese-speaking people have different interests in biodiversity. English-speaking people are more interested in biodiversity and species loss and tend to use governance to conserve ecosystems. In contrast, the Japanese-speaking population is more interested in current events and news about biodiversity and has a more bystander-like attitude. The goal of the SDGs for 2030 is to create sustainable societies around the world, but this study focused only on the English and Japanese-speaking regions of Twitter, and not on the other linguistic regions of the United Nations. The other language spheres of the UN were not addressed. However, changing lifestyles and mainstreaming biodiversity requires policy and education measures and the understanding of people with different cultural backgrounds from all over the world. In order to promote mainstreaming biodiversity, we must first gather the thoughts of people from all over the world and pick up the thoughts of different cultures. China, as a major UN country, support the population that will exceed 1.4 billion in 2019. There are various ideas born from its huge population.

We conducted this research because we thought that picking up and analyzing the ideas about biodiversity from many people in China would be the base for biodiversity mainstreaming policies.

Method

This study used text mining to investigate the awareness of topics related to biodiversity and ecosystem services in Simplified Chinese on Weibo and Traditional Chinese on Twitter (hereafter referred to as "Chinese-speaking countries"). First, we examine which categories each of the Chinese-speaking keywords collected by the data falls into. Next, we will perform morphological analysis on those words and extract and visualize about 300 words in order of frequency. Furthermore, based on the six categories of the "conceptual framework" proposed in the IPBES "Toward the Integration of Science and Policy in the Field of Biodiversity," we will investigate which categories of words people tend to tweet³⁾ IPBES conceptual framework. Finally, the survey results will be used to compare the awareness of biodiversity in Simplified Chinese, Traditional Chinese, and English-speaking countries, and the results of the comparison will be visualized. We will also use Networkx to examine word co-occurrence relationships and their frequencies, create network diagrams, visualize communities, and investigate topics in specific language spheres.

Plan

May~July

All researchers will learn the basic operations of Python and the principles of morphological analysis based on Python. We will also test the feasibility of this research. First, we will obtain API tokens for weibo and Twitter, and run a trial run with the data mining code published on Github to verify access permissions. In addition, we will try to perform morphological analysis using the Chinese morphological analysis code "结巴中文 (<https://github.com/fxsjy/jieba>)" on Github.

July~October

1. It connects to Weibo's API and goes back over the past two years to collect data on sentences containing "生物多样性 OR 生态多样性" in Simplified Chinese. It also connects to Twitter's API and collects data of Tweets containing "生物多样性" in Traditional Chinese.
2. After removing unnecessary information such as URLs and RTs (转发), the noun keywords contained in the text are extracted from the morphological analysis program 结巴.
3. Analyze the words extracted in step 3 and identify the words with high frequency of occurrence.
4. Rank important words based on the number of times they appear and the attention span of the extracted words.
5. Then, an international comparative analysis will be conducted by visualizing the English-speaking, Simplified Chinese and Traditional Chinese regions in WordCloud.
6. We also identify the location of interest in biodiversity by classifying the extracted nouns in the categories of the conceptual framework presented by IPBES.

November - December

1. We will examine the data we have collected so far and discuss how Chinese-speaking people think about biodiversity. At the same time, we will compare the views of people in Traditional Chinese, Simplified Chinese and English-speaking countries and draw out the characteristics of each.
2. Examine how the obtained data can be applied.
3. He and his co-researchers will present the results of their research at domestic and international conferences on sustainability and attempt to engage in dialogue and exchange opinions with the international community through Matsui.

Progress

The progress of the research is generally in line with the plan, but there are some items that differ from the plan.

May~July:

Due to the policy of the Chinese authorities, we were unable to obtain access to the Weibo API, so we only collected tweets from Chinese-speaking countries on Twitter.

July~October:

In addition to the Wordcloud module as a visualization tool, the Networkx module was used for network analysis.

November - December:

On December 17, I participated in an international conference called Eco-health and Ecosystem service in Asia, where I presented my research results and answered questions. We will continue further research with Ishida.

Result

1. WordCloud Visualization according to Frequencies



Figure 1 Words in the top 90% of occurrences in Chinese mochi



Figure 2: English translation of the word cloud in Figure 1.



Figure 3 Words in the top 80-90% of occurrences in Chinese mochi



Figure 4: English translation of the word cloud in Figure 3.



Fig. 5 Words in the top 80% of occurrences in Simplified Chinese



Figure 6: English translation of the word cloud in Figure 5.



Figure 7: Words in the top 70-80% of occurrences in Simplified Chinese



Figure 8: English translation of the word cloud in Figure 7.



Figure 9: Words muttered in Japanese

We found that many of the tweets in Traditional Chinese were directly related to biodiversity and its loss, such as [environment, Greenhouse, nature, destroy]. This is thought to be linked to ecotourism. In addition, many words that indirectly affect biodiversity such as [politics, place, media, government] were also muttered. It is thought that these words are associated with products such as chests of drawers that have eco-friendliness as a selling point. In addition to words related to the ecosystem itself and its loss, such as [Nature, Destroy, Ocean], words related to anthropogenic assets, such as [society, economy, smart chain, people], [cell phone, information construct In addition, words related to the digital ecosystem and bitcoin, such as [cell phone, information construct, block chain, smart chain], were often muttered. As for the words that were muttered in Japanese, there were many topical words such as [conservation conference, environment, ecology, COP, international, treaty]. This is thought to be influenced by the policies promoted by the Ministry of the Environment and the CBD.

2. Result of Network Analysis

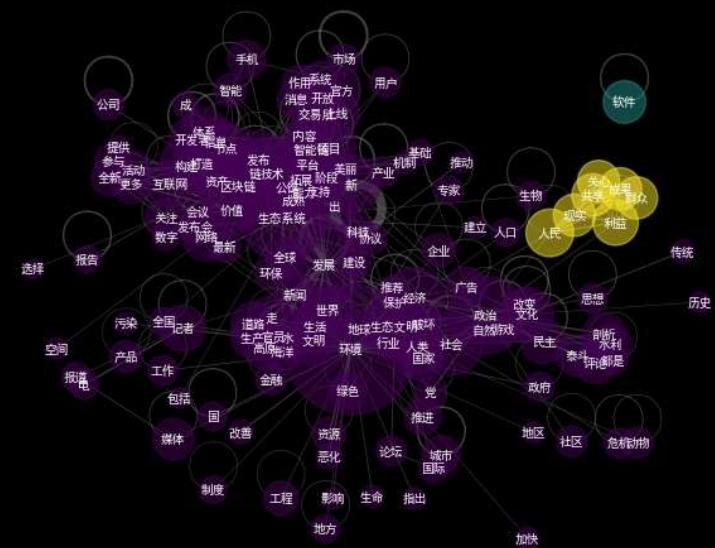


Figure 10: Community diagram generated from 170 Simplified Chinese words.

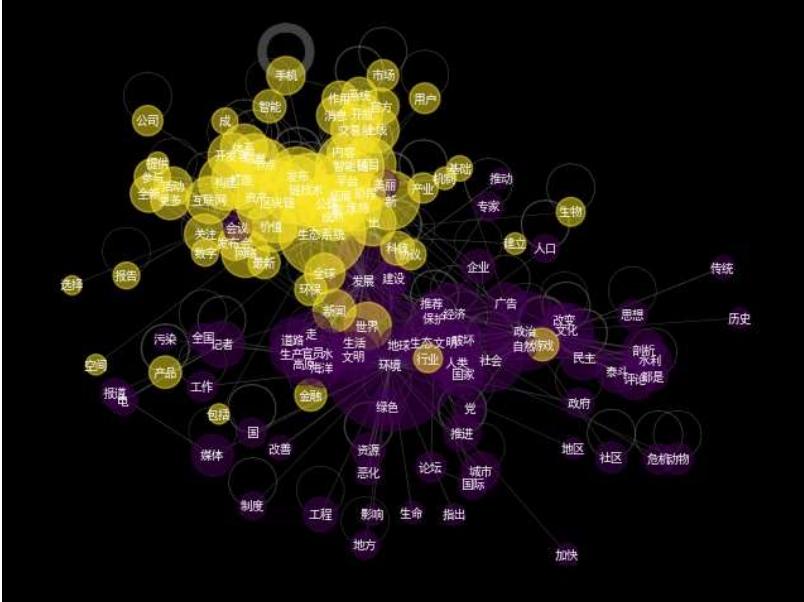


Figure 11: Community diagram generated from 138 Simplified Chinese words.

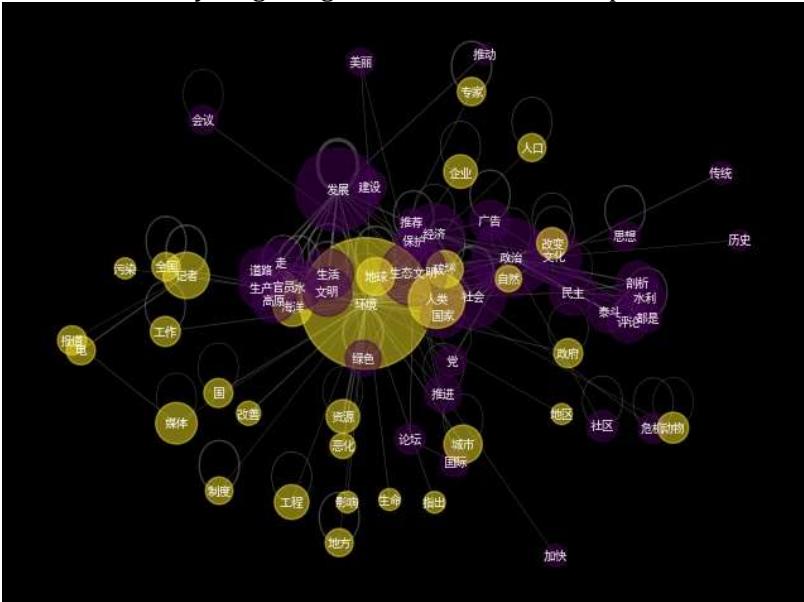


Figure 12: Community diagram generated from 70 Simplified Chinese words.

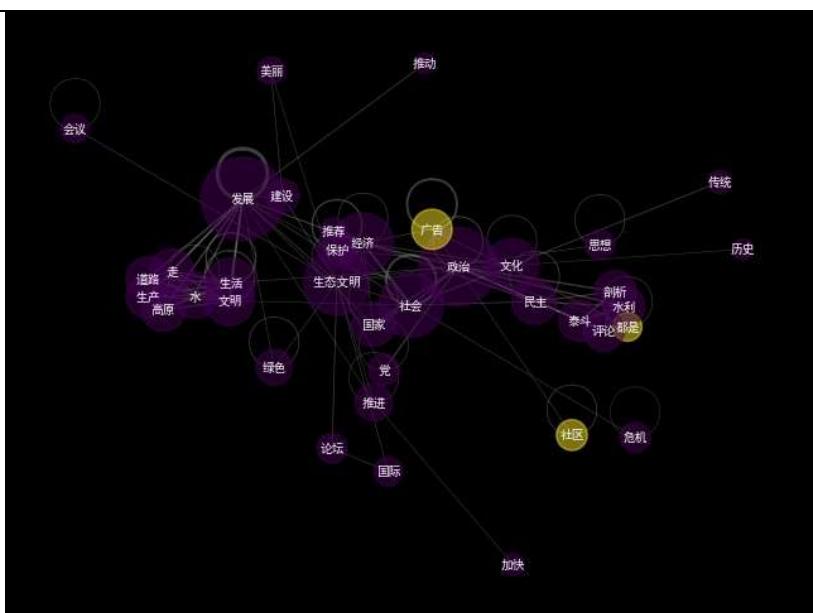


Figure 13: Community diagram generated from 38 Simplified Chinese words.

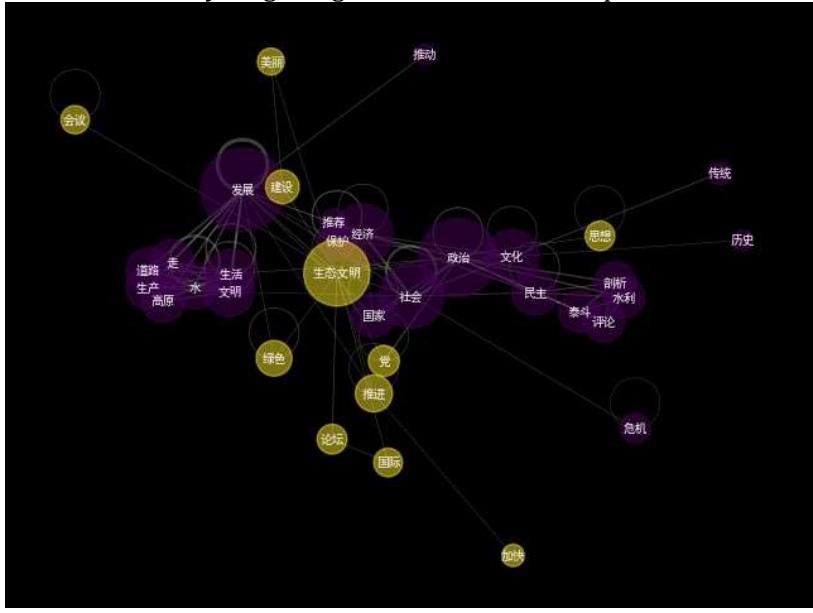


Figure 14: Community diagram generated from 35 Simplified Chinese words.

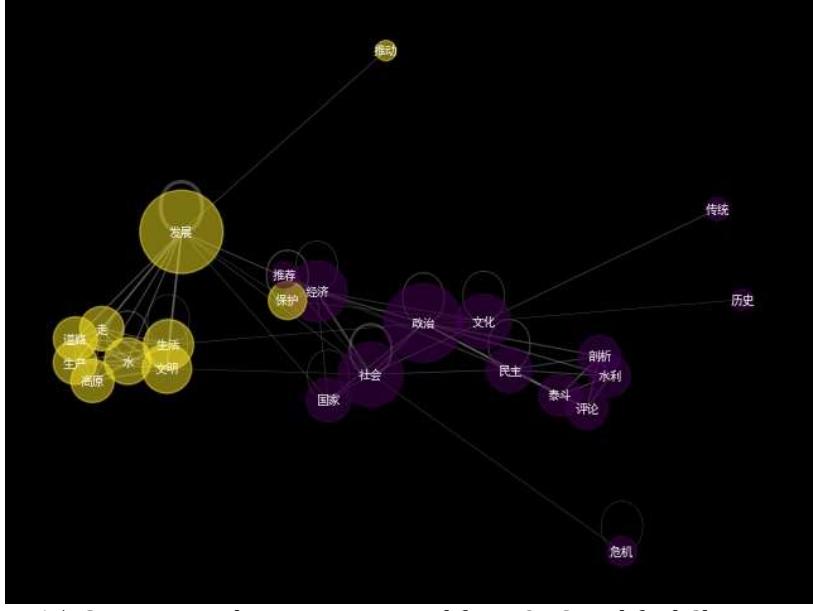


Figure 15: Community diagram generated from 24 Simplified Chinese words.

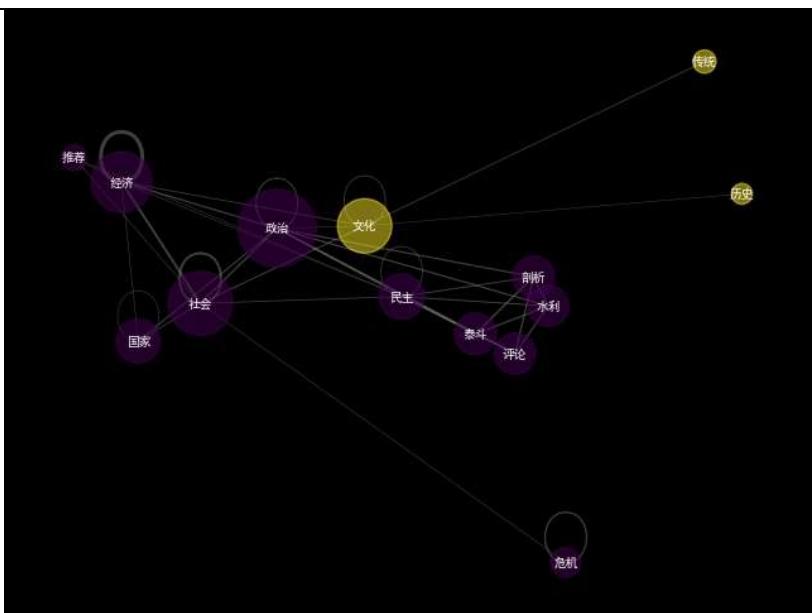


Figure 16: Community diagram generated from 14 Simplified Chinese words.

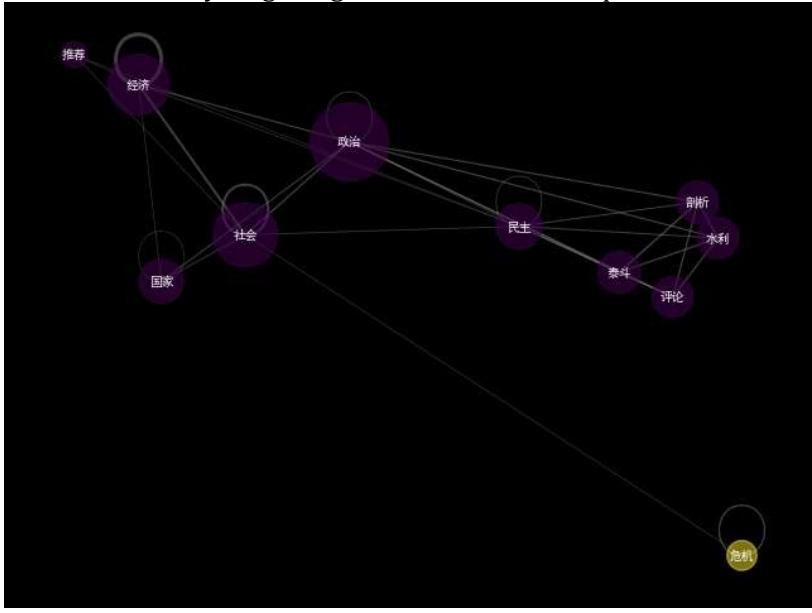


Figure 17: Community diagram generated from 11 Simplified Chinese words.

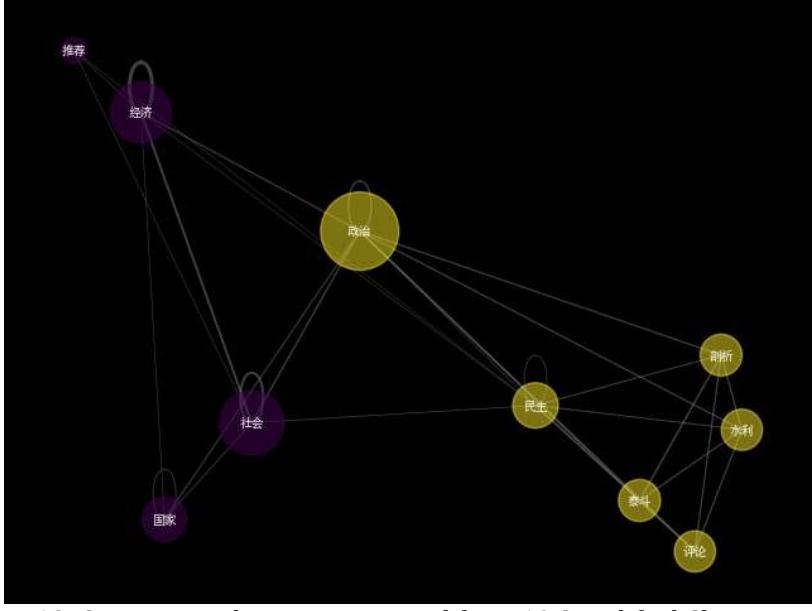


Figure 18: Community diagram generated from 10 Simplified Chinese words.

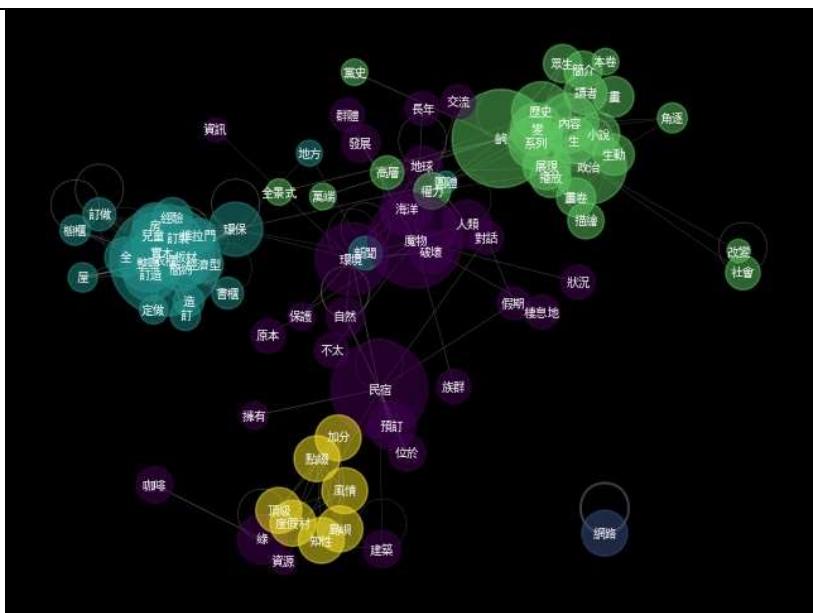


Figure 19: Community diagram generated from 147 Traditional Chinese words.

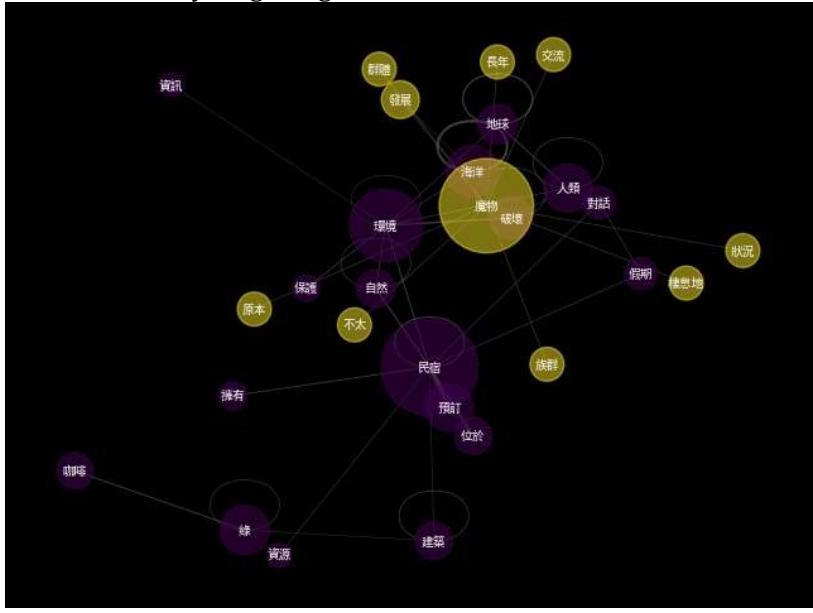


Figure 20: Community diagram generated from 28 Traditional Chinese words.

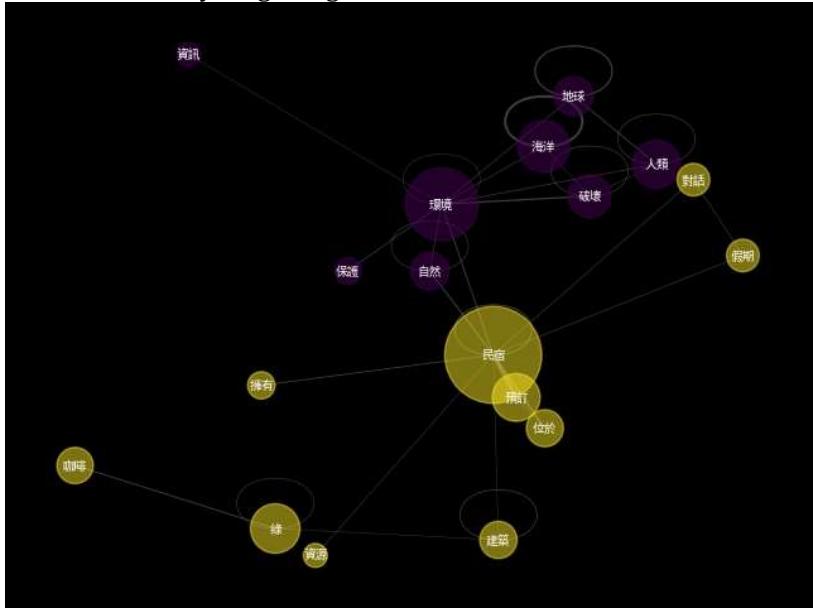


Figure 21: Community diagram generated from 18 Traditional Chinese words.

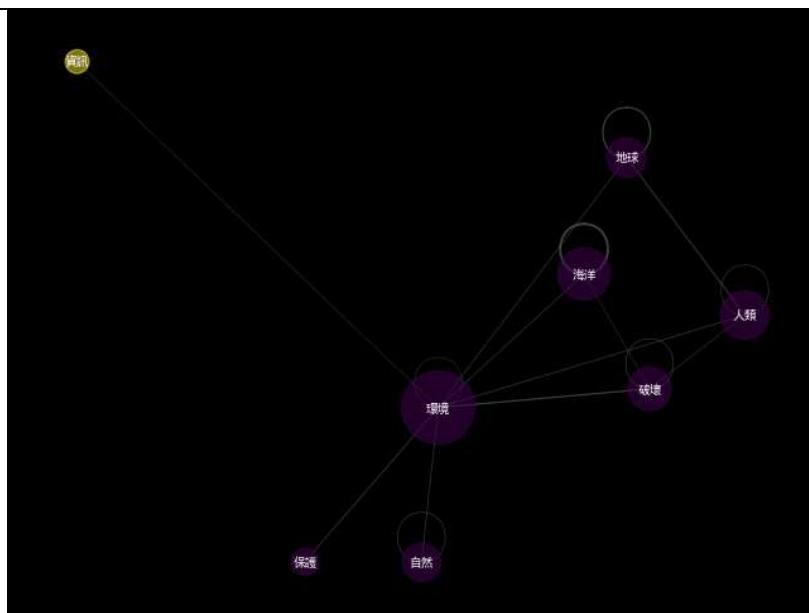


Figure 22: Community diagram generated from eight Traditional Chinese words.

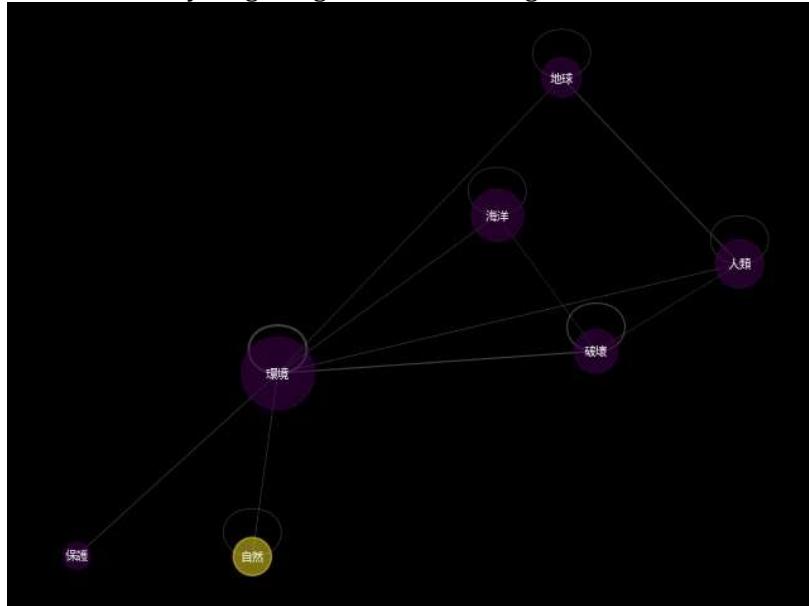


Figure 23: Community diagram generated from seven Traditional Chinese words.

There are seven unique communities in the nine Simplified Chinese community diagrams from Figure 10 to Figure 18. The community ['the masses', 'people', 'shared', 'Outcome', 'Benefit', 'care', 'Reality'] refers to the policy promotion that the CCP uses to praise itself. ['come out', 'supply', 'Block chain', 'platform', 'Smart chain', 'mechanism', 'value', 'number', 'chain', 'market', 'report', 'beauty', 'user', 'intelligent', 'technology', 'content', 'construct', 'space', 'release', 'more', 'include', 'commoder', 'the internet', 'company', 'Environmentally friendly', 'Base', 'Activity', 'focus on', 'up to date', 'industry', 'brand new', 'come', 'choose', 'Conference', 'Developer', 'Build', 'worldwide', 'open', 'node', 'effect', 'support', 'expand', 'Exchange', 'online', 'the Internet', 'Mature', 'protocol', 'cell phone', 'availability', 'world', 'biology', 'ecosystem', 'system', 'participate', 'product', 'establish', 'project', 'assets', 'finance', 'official', 'The communities ['new', 'news', 'game', 'stage', 'technology', 'information'] refer to the digital ecosystem. ['Poke', 'Earth', 'population', 'destroy', 'environment', 'Work', 'country', 'animal', 'improve', 'Ocean', 'Change', 'Officials', 'city', 'resource', 'media', 'nature', 'reporter', 'deterioration', 'Report', 'life', 'National', 'place', 'Electricity', 'government', 'Pollution', 'The community ['area', 'Influence', 'enterprise', 'expert', 'Humanity'] refers to the current state of the natural environment in China. The community ['Ecological Civilization', 'internationality', 'Thought', 'Advancement', 'Meeting', 'accelerate', 'forum', 'green', 'party'] means ecological civilization. civilization. The community ['life', 'go', 'the way', 'production', 'promote', 'civilization', 'water', 'protect', 'plateau', 'develop'] means sustainable growth. The community ['comment', 'analyze', 'dean', 'water conservancy', 'politics', 'democracy'] stands for the management of water resources.

There are six unique communities in the five Traditional Chinese community diagrams from Figure 19 to Figure 23. {'bookcase', 'child', 'environment friendly', 'order', 'cupboard', 'Order', 'Simple', 'Solid wood', 'Plate', 'group', 'Complete', 'make', 'House', 'news', 'Economic', 'Experience', 'Sliding doors', 'wardrobe', 'overall', 'Room', 'place', 'Customize'} is a community of eco-friendly products. The communities {'Sliding doors', 'wardrobe', 'overall', 'Room', 'place', 'Customize'} are meant to advertise products with eco-friendly selling points. The communities {'Intelligence', 'Embellish', 'Resort', 'Add points', 'islands', 'Style', 'Top-level'} and {'have', 'resource', 'lie in', 'green', 'coffee', 'Guesthouse', 'architecture', 'Holiday', 'Booking', 'dialogue'} communities mean ecotourism. The communities {'situation', 'communicate', 'Bad', 'Ethnic group', 'Monster', 'develop', 'original', 'Many years', 'Habitat'} are about the setting of a fictional world where demon-like creatures appear. This community is about the setting of a fictional world where demonic creatures appear. The community {'protect', 'Humanity', 'Ocean', 'destroy', 'environment', 'Earth'} is about protecting the ocean.

Discussion

We examined the commonality and differences in Simplified Chinese, Traditional Chinese, and Japanese at the tokens level and the communities level, respectively. At the tokens level, we extracted unique words and commonly muttered words based on the top 1% of Japanese, Simplified Chinese, and Traditional Chinese tokens, respectively. Next, we drew a Venn diagram of the unique and commonly muttered words based on their frequencies, creating a wordcloud diagram.

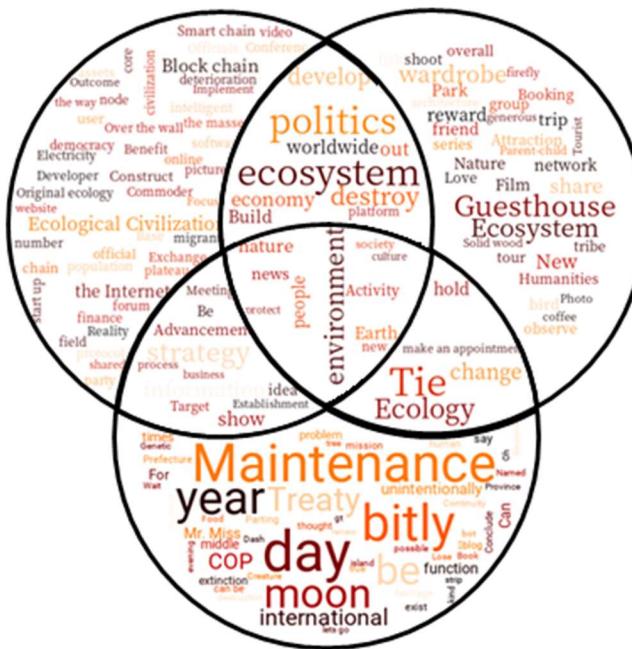


Figure 24: Distinctive words muttered in a language area

In order to examine the commonalities and differences in detail, the words in the WordCloud and the diagrams of the Networkx analysis were classified based on the IPBES Framework.

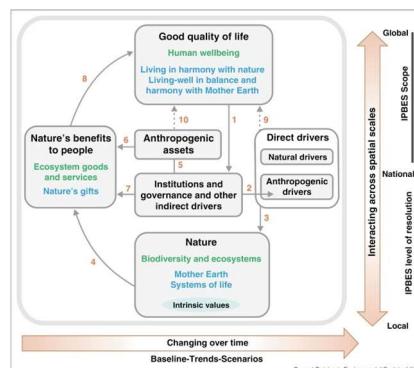


図 25 IPBES Conceptual Framework

There are 6 different elements in IPBES Conceptual Framework. The first one is 'Nature'. 'Nature' in the context of IPBES refers to the natural world with an emphasis on the diversity of living organisms and their interactions among themselves and with their environment. The second one is 'Anthropogenic assets'. 'Anthropogenic assets' refers to built infrastructure, health facilities, knowledge (including ILK and technical or scientific knowledge, as well as formal and nonformal education), technology (both physical objects and procedures), and financial assets, among others. The third one is 'Nature benefit to people'. 'Nature's benefits to people' refers to all the benefits that humanity — individuals, communities, societies, nations or humanity as a whole — in rural and urban settings — obtains from nature. Ecosystem goods and services — including provisioning, regulating and cultural services [8] — all fall in this category. The fourth one is 'Institutions and governance systems and other indirect drivers'. 'Institutions and governance systems and other indirect drivers' are the ways in which people and societies organize themselves and their interactions with nature at different scales. The fifth one is 'Direct drivers'. 'Direct drivers', both natural and anthropogenic, affect nature directly. 'Natural direct drivers' are those that are not the result of human activities and whose occurrence is beyond human control. The sixth one is 'Good quality of life'. 'Good quality of life' is the achievement of a fulfilled human life. Although what it entails varies considerably within and among different societies and cultures, everybody wants to be free from poverty and disease, have a long and fulfilling life, and access to freedoms and rights.

1. Commonality Analysis using IPBES Conceptual Framework



Fig. 26 Commonly visualized word groups in Simplified Chinese, Traditional Chinese, and Japanese

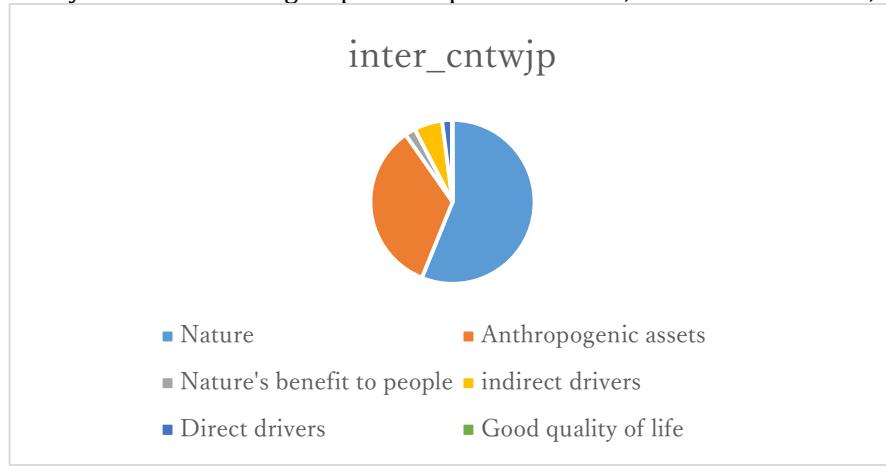


Figure 27. Commonly visualized word categories.

The fan chart Figure 27 shows that the words that are included in the three communities. These words mainly relate to "Nature, Anthropogenic assets". This is because that In COP 10, the concept of nature and its vital contribution to people are promoted. As a result, users in twitter understood and began to use these words.

2. difference analysis according to unique words



Figure 28. Word clouds visualized only in Simplified Chinese.

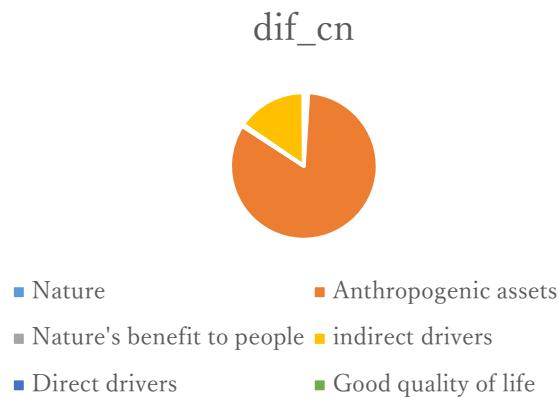


Figure 29. Word group categories visualized only in Simplified Chinese.

Figure 29 is the visualization of words in simplified Chinese community mainly relate to "anthropogenic assets". This is because that users in China prefer to consider technology and social problems relate to themselves, rather than to consider the nature.

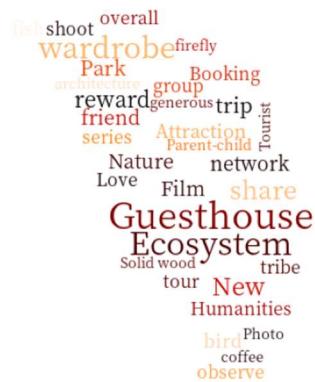


Figure 30: Word clouds visualized only in Traditional Chinese.

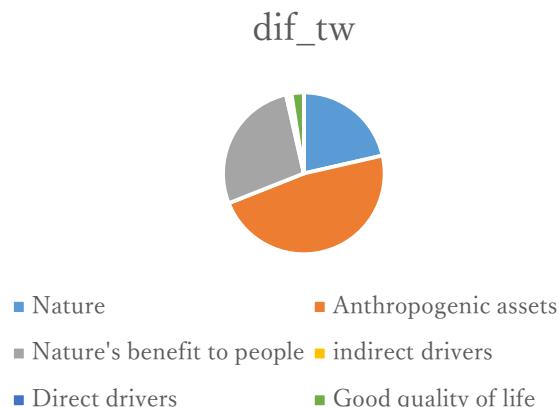


Figure 31. Word categories visualized in Traditional Chinese only.

Figure 31 shows the visualization of words in traditional community. These words are mainly related to Anthropogenic assets and Direct drivers.



Figure 32: Word clouds visualized only in the Japanese-speaking world

dif_jp



- Nature
- Anthropogenic assets
- Nature's benefit to people
- Indirect drivers
- Direct drivers
- Good quality of life
- UN Days & conference

Figure 33. Word categories visualized only in the Japanese-speaking world.

Figure 33 is the visualization of words in Japanese community are roughly related to UN days and Nature itself. These is because that during COP 10, users post many tweets related to nature. Also, these tweets include Japanese government promotion of their policies.

3. Commonality and Difference in topics

	Simplified Chinese	Traditional Chinese
Nature	Degrading nature, crisis	Ecosystem Conservation
Anthropogenic assets	Digital ecosystem, Marketing,	Internet, twitter bot
Benefits to People	Cultural Service	Ecotourism, nature-friendly products
Indirect Drivers	Good ecosystem governance, Ecological Civilization, economy	
Direct Drivers	Sustainable Development, water management	Environment Destruction of earth and ocean
Good quality of life		
Others		

Table 1: Categorization of the topics extracted from tweets

Table 1 shows the categorization of the topics that extracted from networkx analysis. In simplified Chinese community, the topics are degrading nature, crisis, digital ecosystem, cultural service, good ecosystem governance, ecological civilization, economy, sustainable development, and water management. In traditional Chinese community, the topics are ecosystem conservation, internet, twitter bot, ecotourism, nature-friendly products and environment destruction of earth and ocean.

The categories of topics in simplified and traditional Chinese community are almost same but with difference relates to indirect drivers. Topics, such as degrading nature, crisis, digital ecosystem, cultural service, economy, sustainable development, water management, ecosystem conservation, internet, twitter bot, ecotourism, nature-friendly products ,and environment destruction of earth and ocean, in simplified and traditional Chinese belong to same categories. However, the specific topics are different. In addition, topics which could be categorized as good quality of life did not appear. This indicate that users in simplified and traditional Chinese may not realize or seldom receive happiness from nature. Moreover, In the row of indirect drivers, topics in traditional Chinese community did not appear in networkx analysis. 2 reasons why the difference was found can be suggested. The first reason is that compare to traditional Chinese, China authority and Chinese media would like to promote politics slogan and news to praise the governance of Chinese authority. Receiving these messages, users may repeat or discuss the content of this messages including ecosystem governance and ecological civilization. The second reason is the existence of bots that repeating the slogan and news from Chinese authority and media. As a result, tweets from bots were collected and topics relate to indirect drivers were collected.

Conclusion

In this study, text mining was applied to tweets and analyze semantics toward biodiversity using wordcloud and networkX library. As a result, we found that the semantics in different languages communities show significant difference. In simplified Chinese community, the topics relate to biodiversity is 'ecosystem civilization' from 'indirect drivers' and 'digital ecosystem' from 'anthropogenic assets'. In traditional Chinese community, the topics are ecotourism and commercial promotion of eco-friendly products from 'benefit to people'. In Japanese community, the topics relate to biodiversity is the policy promotion of Japanese government from 'indirect drivers'. The reason why the unique topics in each community appears is different culture context and policies.

In the future, we would like to collect tweets data using more keywords and call for fellow researchers to commit the analysis of semantics. In this study, we used 'biodiversity' and 'eco' to collect tweets. However, in different culture, people will use frank words in their culture context, such as 'nature', 'environment' and so on. Thus, it is essential to cooperate with fellow researchers from different culture backgrounds and discuss the keywords for every languages communities. Moreover, translating tweets into English will lose their unique culture-based meaning. Therefore, we need to analyze tweets in their own culture context with fellow researchers.

References

- 1) IPBES (2019): Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. S. Díaz, J. Settele, E. S. Brondízio E.S., H. T. Ngo, M. Guèze, J. Agard, A. Arneth, P. Balvanera, K. A. Brauman, S. H. M. Butchart, K. M. A. Chan, L. A. Garibaldi, K. Ichii, J. Liu, S. M. Subramanian, G. F. Midgley, P. Miloslavich, Z. Molnár, D. Obura, A. Pfaff, S. Polasky, A. Purvis, J. Razzaque, B. Reyers, R. Roy Chowdhury, Y. J. Shin, I. J. Visseren-Hamakers, K. J. Willis, and C. N. Zayas (eds.). IPBES secretariat, Bonn, Germany. 56 pages.
- 2) Twitter データを用いた生物多様性に対する関心の探索的研究, Shu Ishida, 2021
- 3) S. Díaz et al. The IPBES Conceptual Framework — connecting nature and people, Current Opinion in Environmental Sustainability, Volume 14, 2015, Pages 1-16, ISSN 1877-3435