Recent Trends in Languages Studied in Psycholinguistics: A 2018-2022 Survey

Masatoshi Koizumi
(Tohoku University)

【**Abstract**】Based on the investigation of 2,004 papers in psycholinguistics published in major international academic journals from January 1, 2018, to December 31, 2022, it was found that research was conducted on 64 languages belonging to 32 language families. The language that appeared most frequently as the subject of study was English, accounting for 42.1% of the total. Chinese ranked second (14.6%), followed by German in third place (9.4%). Japanese ranked eighth (1.9%). Out of the 64 languages, 27 belonged to the Indo-European language family. Comparing these results to a similar survey conducted approximately a decade ago by Anand et al. (2011), the number of languages studied increased from 57 to 64, indicating an increase in diversity. Additionally, Chinese made a significant leap from seventh place to second place, while Japanese dropped from second place to eighth place.*

Keywords: comprehension, production, first language acquisition, minority languages, Japanese

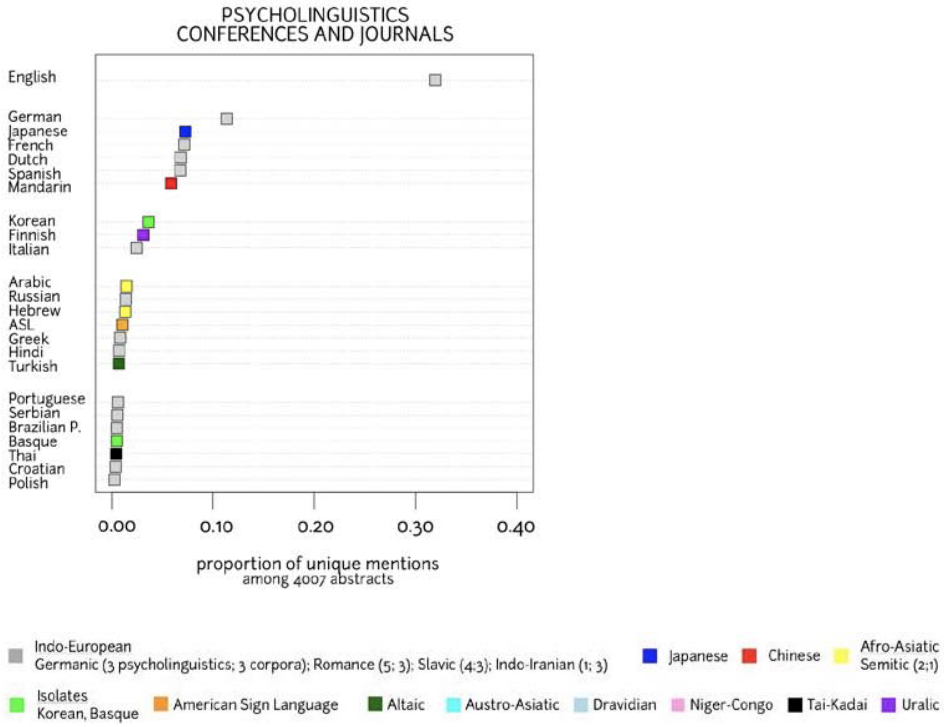
1. Introduction

Currently, there are over 7,000 languages in use around the world (Eberhard et al., 2023). However, psycholinguistic research exhibits a pronounced inclination towards “major languages” primarily spoken in economically prosperous regions, thus displaying a significant bias. According to Anand et al. (2011), English constitutes one-third of the total research output in major psycholinguistic studies (comprising over 4,000 papers and conference abstracts), with only 10 languages (English, German, Japanese, French, Dutch, Spanish, Mandarin, Korean, French, Italian) accounting for 85% of the research (Figure 1). When accounting for all languages that have been studied at least once, the total number amounts to only 57. The majority of these languages belong to the Indo-European language family. Since the publication of Anand et al. (2011), there has been increasing recognition of the importance of diversifying the languages studied in psycholinguistics (Norcliffe et al., 2015; Koizumi, 2023). Therefore, this study

* This paper is an extensively revised version of an academic trends research report submitted to the Research Center for Science Systems, Japan Society for the Promotion of Science in March 2023. The research received partial support from their grant and JSPS KAKENHI (JP19H05589). I express my gratitude to Kaito Yamamoto for assistance with data gathering and Sachiko Kiyama for conducting statistical analyses.

Figure 1

Distribution of Target Languages in Psycholinguistic Research (Adopted from Anand et al., 2011)



investigates the extent to which languages studied in psycholinguistics have diversified over the past decade.

2. Method

This study investigated the diversification of languages studied in experimental psycholinguistic research over the past decade. Specifically, we selected 36 journals in the Linguistics and Language category from Scimago Journal & Country Rank (https://www.scimagojr.com/), as listed in (1). The selection was based on their high SCImago Journal Rank indicator or H-index and their substantial publication rate of research papers on psycholinguistics. By reading all the abstracts (and the main texts if necessary) from these 36 journals, we extracted 2,004 research articles on language production, language comprehension, or first language acquisition published between January 1, 2018, and December 31, 2022. We examined the names and language families of the languages studied in each article using Ethnologue Classification (https://www.ethnologue.com/).

(1) Journals included in the study

1. Brain and Cognition
2. Brain and Language
3. Cognition
4. Cognitive Brain Research
5. Cognitive Science
6. Cognitive Psychology
7. First Language
8. Frontiers in Human Neuroscience
9. Frontiers in Psychology
10. Glossa
11. Human Brain Mapping
12. Journal of Child Language
13. Journal of Cognitive Neuroscience
14. Journal of Cognitive Psychology
15. Journal of Cognitive Science
16. Journal of Experimental Psychology: Learning, Memory, and Cognition
17. Journal of Memory and Language
18. Journal of Neurolinguistics
19. Journal of Phonetics
20. Journal of Pragmatics
21. Journal of Psycholinguistic Research
22. Journal of Speech, Language, and Hearing Research
23. Language
24. Language Acquisition
25. Language and Cognition
26. Language and Speech
27. Language, Cognition and Neuroscience
28. Lingua
29. Linguistic Inquiry
30. Natural Language and Linguistic Theory
31. NeuroImage
32. PLOS ONE
33. Proceedings of the National Academy of Sciences
34. Psychological Science

35. Quarterly Journal of Experimental Psychology

36. Syntax

3. Results

The results revealed 64 languages from 32 families (Figure 2; Table 1). English was the most frequently studied, accounting for 42.1% of the total studies. Chinese ranked second (14.6%), followed by German (9.4%). Japanese ranked eighth (1.9%). Research in the top ten languages accounted for 87.7% of all studies. Of the 64 languages studied, 27 belonged to the Indo-European family. Thirty languages were examined in only one study, whereas eight were studied in only two studies.

We conducted a decision tree analysis using language families as factors to examine the frequency of research on each language family over the past five years. The results showed a significant division among Indo-European (IE), Japanese-Ryukyuan (JR), and Sino-Tibetan (ST) language families and all other language families (Figure 3). The IE, JR, and ST language families had an average of 57.625 experimental studies per language published over the past five years. In contrast, other language families had only 5.000 per language, indicating a strong bias toward these three language families in experimental research. The number of minority languages studied in psycholinguistics is extremely limited, with very low research frequency.

Figure 2

Distribution of Target Languages in Psycholinguistic Research (2018-2022)

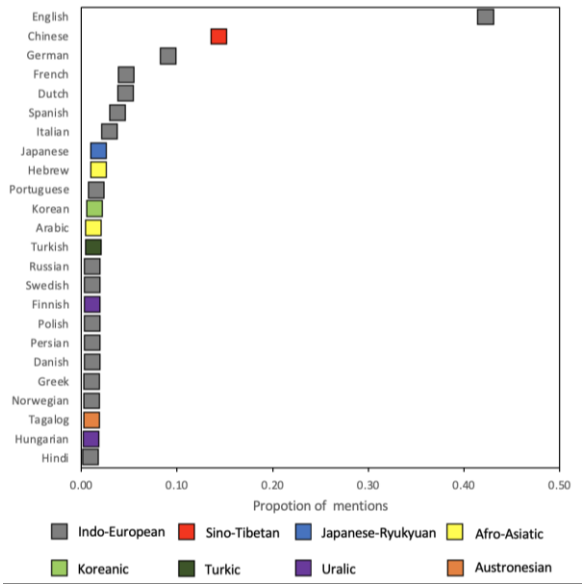
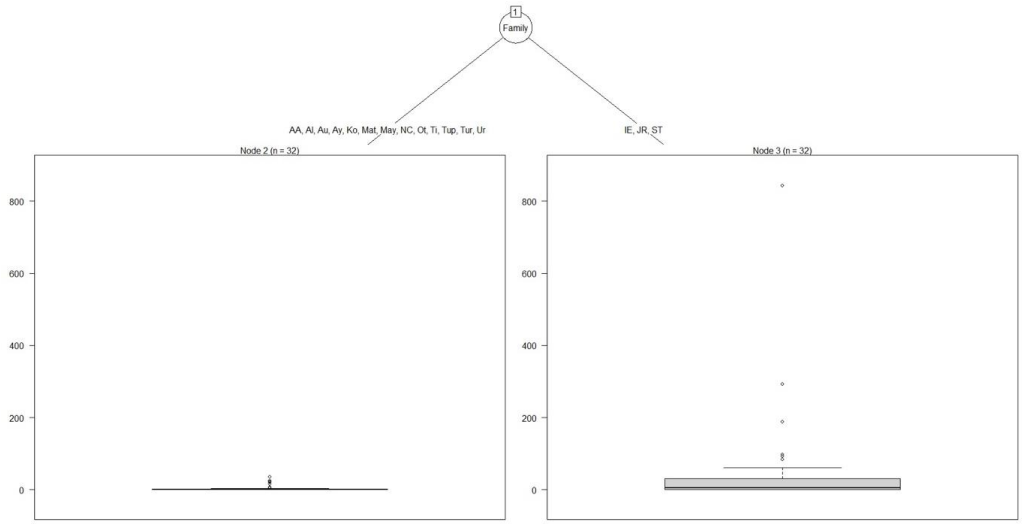


Table 1

Target Languages in Psycholinguistic Research (2018-2022)

Language	Family of language	Number	Percentage
English	Indo-European	843	42.07%
Chinese	Sino-Tibetan	293	14.62%
German	Indo-European	188	9.38%
French	Indo-European	97	4.84%
Dutch	Indo-European	93	4.64%
Spanish	Indo-European	84	4.19%
Italian	Indo-European	61	3.04%
Japanese	Japanese-Ryukyuan	37	1.85%
Hebrew	Afro-Asiatic	35	1.75%
Portuguese	Indo-European	26	1.30%
Korean	Koreanic	25	1.25%
Arabic	Afro-Asiatic	22	1.10%
Turkish	Turkic	21	1.05%
Russian	Indo-European	20	1.00%
Swedish	Indo-European	19	0.95%
Finnish	Uralic	16	0.80%
Polish	Indo-European	15	0.75%
Persian	Indo-European	14	0.70%
Danish	Indo-European	12	0.60%
Greek	Indo-European	10	0.50%
Norwegian	Indo-European	7	0.35%
Tagalog	Austronesian	6	0.30%
Hungarian	Uralic	5	0.25%
Czech	Indo-European	3	0.15%
Estonian	Uralic	3	0.15%
Hindi	Indo-European	3	0.15%
Akan	Niger-Congo	2	0.10%
Basque	Indo-European	2	0.10%
Catalan	Indo-European	2	0.10%
Georgian	Indo-European	2	0.10%
Icelandic	Indo-European	2	0.10%
Indonesian	Austronesian	2	0.10%
Northern East Cree	Algic	2	0.10%
Slovenian	Indo-European	2	0.10%
Armenian	Indo-European	1	0.05%
Aymara	Aymaran	1	0.05%
Bantu	Niger-Congo	1	0.05%
Bengali	Indo-European	1	0.05%
Castilian	Indo-European	1	0.05%
Chamorro	Austronesian	1	0.05%
Croatian	Indo-European	1	0.05%
Drehu	Austronesian	1	0.05%
Farsi	Indo-European	1	0.05%
Flemish	Indo-European	1	0.05%
Irish	Indo-European	1	0.05%
K'iche'	Mayan	1	0.05%
Lebanese	Afro-Asiatic	1	0.05%
Lithuanian	Indo-European	1	0.05%
Malay	Austronesian	1	0.05%
Maltese	Afro-Asiatic	1	0.05%
Murrinhpatha	Austronesian	1	0.05%
Niuean	Austronesian	1	0.05%
Norhtern Pale	Otomanguean	1	0.05%
Nungon	Austronesian	1	0.05%
Saudi Arabic	Afro-Asiatic	1	0.05%
Serbian	Indo-European	1	0.05%
Ticuna	Ticuna	1	0.05%
Tzotzil	Mayan	1	0.05%
Vietnamese	Austronesian	1	0.05%
Wichi	Matacoan	1	0.05%
Yoloxochitl Mixtec	Otomanguean	1	0.05%
Yudja(Tupi)	Tupian	1	0.05%
Kaqchikel	Mayan	1	0.05%
Zulu	Niger-Congo	1	0.05%
Total		**2004**	**100.00%**

Figure 3
Decision Tree Analysis of Research Frequency by Language Families



Note. AA (Afro-Asiatic), Al (Algic), Au (Austronesian), Ay (Aymaran), Ko (Koreanic), IE (Indo-European), JR (Japanese-Ryukyuan), Mat (Matacoan), May (Mayan), NC (Niger-Congo), Ot (Otomanguean), ST (Sino-Tibetan), Ti (Ticuna), Tup (Tupian), Tur (Turkic), Ur (Uralic).

4. Comparison with Anand, Chung, and Wagers (2011)

While Anand et al. (2011) conducted an extensive investigation encompassing over 4,000 abstracts of scholarly articles and conference presentations, this study focused exclusively on a sample of 2,004 academic papers. As such, it is worth noting that a direct comparison between the two studies may not be entirely warranted. Nonetheless, we conducted a comparative analysis of the findings for reference.

The number of languages studied has increased from 57 to 64, suggesting an increase in diversity. However, the percentage of studies focusing on English rose significantly from approximately one-third to 42.1%, indicating a growing concentration of English-focused studies. One possible reason for the higher proportion of English-focused research in this study is the impact of the COVID-19 pandemic. However, even when considering the two years before the pandemic (2018 and 2019), the proportion of English-focused research was still high at 43.1%. This suggests that the difference between the results of the two studies may not be solely attributed to the impact of the pandemic. Another possible factor is that Anand et al. (2011) included both journal articles and conference presentations, whereas this study focused only on journal articles. Studies targeting languages other than English, especially those on minority

languages, might be more frequently presented at conferences but published in specialized journals with a narrower focus. This may lead to their exclusion from this study. For example, three psycholinguistic studies on Seediq (Austronesian, Taiwan) were published in the *Journal of East Asian Linguistics* during the five years covered by this study (Ono et al., 2020; Sato et al., 2020; Yano et al., 2019) but are not included in Table 1. The *Journal of East Asian Linguistics* was not included in our list of target journals as it specializes in formal approaches to East Asian languages and publishes relatively few psycholinguistic studies. As earlier versions of the three papers under discussion were presented at international conferences, they would have been counted if the current study had included conference presentations. It is necessary to investigate the languages targeted in conference presentations from 2018 to 2022 to accurately determine whether the proportion of English-focused research has genuinely increased in psycholinguistics over the past decade.

Another notable difference between Anand et al. (2011) and the present study is the significant rise of Chinese from 7th (approximately 6%) to 2nd (14.6%) place, while Japanese fell from 2nd (approximately 8%) to 8th (1.9%) place. This trend is similar to the changes in the rankings of China and Japan in the top 10% of highly cited papers across all academic fields, as reported in Ministry of Education, Culture, Sports, Science and Technology (2023). This indicates a decline in Japan's international standing in academia, which is a matter of serious concern.

References

Anand, P., Chung, S., & Wagers, M. (2011). Widening the net: Challenges for gathering linguistic data in the digital age. Unpublished manuscript. University of California, Santa Cruz.

Eberhard, David M., Gary F. Simons, & Charles D. Fennig (Eds.). (2023). Ethnologue: Languages of the World. Twenty-sixth edition. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com. Accessed July 10, 2023.

Ministry of Education, Culture, Sports, Science and Technology (2023). *FY2023 White Paper on Education, Culture, Sports, Science and Technology*. (https://www.mext.go.jp/b_menu/hakusho/html/hpaa202301/1421221_00014.html, accessed July 21, 2023)

Norcliffe, E., Harris, A. C., & Jaeger, T. F. (2015). Cross-linguistic psycholinguistics and its critical role in theory development: Early beginnings and recent advances. *Language, Cognition and Neuroscience*, *30*(9), 1009-1032. https://doi.org/10.1080/23273798.2015.1080373.

Koizumi, M. (2023). Constituent Order in Language and Thought: A Case Study in Field-Based Psycholinguistics. Cambridge University Press.

Ono, H., Kim, J., Sato, M., Tang, A. A. Y., & Koizumi, M. (2020). Syntax and processing in Seediq: A behavioral study. *Journal of East Asian Linguistics*, 29(2), 237–258. https://doi.org/10.1007/s10831-020-09207-7.

Sato, M., Niikuni, K., Schafer, A. J., & Koizumi, M. (2020) Agentive versus non-agentive motions immediately influence event apprehension and description: an eye-tracking study in a VOS language. *Journal of East Asian Linguistics* 29(2) 211-236. https://doi.org/10.1007/s10831-020-09205-9.

Yano, M., Niikuni, K., Ono, H., Sato, M., Tang, A. A. Y., & Koizumi, M. (2019). Syntax and processing in Seediq: An event-related potential study. *Journal of East Asian Linguistics*, *28*(4), 395-419. https://doi.org/10.1007/s10831-019-09200-9.

心理言語学における研究対象言語の最近の動向

小泉政利

【要旨】2018 年 1 月 1 日から 2022 年 12 月 31 日までの 5 年間に主要な国際学術誌に掲載された心理言語学の論文 2004 本を調査した結果、32 語族の 64 言語が研究対象になっていた。取り上げられた回数（論文数）が一番多かったのは英語で、全体の 42.1%を占めた。2 位は中国語（14.6%）、3 位はドイツ語（9.4%）で、日本語は 8 位（1.9%）であった。64 言語中 27 言語が印欧語族に属していた。Anand et al. (2011) で報告されている約 10 年前の類似の調査結果と比較すると、研究対象が 57 言語から 64 言語に微増しており、多様性が増してきているようにみえる。また、中国語が 7 位（約 6%）から 2 位（14.6%）に躍進する一方、日本語が 2 位（約 8%）から 8 位（1.9%）に後退した。これは、全学術分野の Top10%補正論文数における中国および日本のランクの変動と酷似している。