Mark Streer - Final Research Project

The Brain in Translation:
Linguistic features of English research articles authored by Japanese neuroscientists

Navigation

1. Abstract

English is the lingua franca of international scientific communication. In Japan, scientists often draft research articles (RAs) in their native language, and have them translated into English by technical translators for publication. Despite broadly similar document structures, Japanese-language RAs differ from English RAs in their conventional diction and discourse, which can manifest as unusual or confusing constructions when transferred directly. This paper introduces a few source-language features typically present in English RAs authored by Japanese scientists, and substantiates their prevalence in a corpus analysis in a single domain. I constructed two mini-corpora composed of comparable RAs published in the same open-access journal (*PLOS One*)—10 by authors from Japanese institutions, and 10 by authors from Western institutions—in the field of functional magnetic resonance imaging (fMRI). The differences identified offer insights for Japanese researchers and translators alike about expected domain-specific language in neuroscience as well as original RAs generally.

2. Introduction

Original research articles (RAs) are reports of systematic research published in scientific journals. Scientists around the world use this genre to communicate the background, aims, and findings of their research with their peers. Their basic structure[1] is similarly universal, although section names and their order vary by subgenre and field. While RAs may be published in any language depending on expected audience, English is the modern *lingua franca* chosen by authors worldwide to reach the largest possible international readership; in the non-Anglophone world, this process is aided by technical translators.

Japan has a sizeable market for English technical translation thanks to the country's status as an export-driven economic, technological, and industrial power. Still, customers are often delivered translated documents loaded with 'translationese': unnatural syntax, word order, and high concentrations of

[1] I.e., "IMRaD": Introduction, Methods, Results, and Discussion

unnatural-sounding terminology resulting from direct transfer of words and grammar from the source language (Shuttleworth & Cowie 187). Regardless of mother tongue, verbatim transfer often occurs when an author/translator fails to recognize a corresponding language or genre norm in English, and instead cautiously chooses a word-level equivalent. Even when direct transfer more or less conveys the original meaning, there is greater risk of it being misinterpreted by not just L1 English readers, but also L2 English readers familiar with more-conventional expressions and phraseology.

This challenge is exacerbated by the vast divergence between Japanese and English. For example, Japanese rarely distinguishes between singular and plural, or definite and indefinite articles; it uses postpositions, not prepositions; its SOV word-order facilitates long relative clauses before nouns they modify; and grammaticalization of topic–subject as well as social hierarchies and roles (Refsing & Lundquist 18-20). Germanic in origin, English is an Indo-European language characterized by extensive adoption of scholarly and scientific terms from Greek, Latin, or descendant Romance languages; Japanese is an independent language family with extensive logographic and conceptual transfer from Chinese. For scientific concepts, Japanese imported new concepts using Chinese characters/compounds called *kanji/kango* through the modern era, and loanwords from European languages since the 1600s. More recently, English terms and especially acronyms are common sights in Japanese RAs, but typically confined to parentheses or explanatory footnotes adrift in seas of logographic prose.

Just like their English counterparts, Japanese RAs are sophisticated texts that require deep subject-matter expertise and familiarity with domain-specific conventions to comprehend, yet tend to be shorter in length and use smaller lexical and syntactical inventories. This may seem surprising in light of Japanese's higher tolerance for verbosity, which translators are encouraged to curtail (Wakabayashi 109-110). One plausible explanation is that information can generally be conveyed in fewer logographic than alphabetic characters: for example, Article 1 of the UN Declaration of Human Rights consists of 170 characters in English, but just 87 in Japanese and 43 in Chinese (Ingraham). There may also be human factors at play: for example, experts may negatively regard paraphrasing for the sake of readability, since their SME makes them highly sensitive to technical distinctions and nuance, incentivizing translators to err on the side of redundancy and literal transfer (Mossop et al. 72-73). Whatever the reason, characteristic lexical and syntactical aspects of Japanese-language RAs are expected to carry over to English RAs created by translation or written/revised by L1 Japanese authors themselves. This corresponds to a general tendency – widely recognized – for source-language features having straightforward equivalents in the target language to be over-represented in translated texts, termed *interference* (Eskola 96). Discourse differences are admittedly less pronounced for scientific texts than popular genres such as literature and short-form journalism (Wakabayashi 223), based on a decade of career experience revising scientific communications, I can attest to the prevalence of cryptic, confusing prose apparently caused by direct transfer from Japanese in technical documents drafted by both L1 English translators and L2 English scientists. Elucidating and categorizing the sources and manifestations of such translationese would serve as valuable reference for Japanese researchers and translators wanting to prioritize the substance and findings of their research over the style of its presentation.

This paper is my attempt to empirically identify linguistic features of English RAs written by Japanese authors originating from their native language by comparison of two mini-corpora. To minimize the effects of variation due to domain-specific language and journal-specific criteria I selected RAs from the same research subfield—neuroscience, specifically human functional magnetic resonance imaging (fMRI)[2]—and the same publication (*Public Library of Science ONE*). Below, after explaining the rationale and method of the corpora's construction, I will discuss selected example texts which exhibit Japanese linguistic conventions in lexical, syntactic, and pragmatic terms.

[2] This field was chosen for practical reasons: having earned a degree in neuroscience (B.S., 2009) and been involved in neuroimaging studies at three different research institutions in the past, I felt more qualified to judge the technical content and language of primary literature in this field than in any other.

3. Corpora construction and analysis

Two comparable mini-corpora were constructed: "JP-EN" consisting of 10 RAs authored exclusively by authors at Japanese institutions, and "EN-EN" consisting of 10 RAs originating from Western[3] institutions. All RAs were published in *PLOS One* over a ten-year period (2012-01-01 to 2021-12-31). Selecting this open-access journal helped to minimize selection bias due to licensing restrictions, and simplified corpus construction as the full texts of all papers are made available in XML format.[4]

First, the JP-EN corpus was prepared by searching the *PLOS One* online database using the search strings "fMRI study" and "Japan" over the date range above. These search results were filtered by Article Type (i.e., only "Research article"; no reviews, commentary, corrections, etc.) and Subject Area ("Functional magnetic resonance imaging", yielding 222 hits. These candidates were manually inspected to isolate RAs focused on higher-order cognitive functions (e.g., language, memory, decision-making), primarily in healthy adults; organic diseases typically discussed in terms of anatomical/histological characteristics (e.g., brain tumors) or physical symptoms (e.g., Parkinson's disease) were excluded. Finally, authorship information was inspected to confirm that each candidate had been authored exclusively by Japanese authors at Japanese institutions. Fourteen RAs met the screening criteria as outlined above.

Next, the EN-EN corpus was prepared. The *PLOS One* online database was searched again using the same date range and filters, but this time for only "fMRI study". Given the much larger number of search results (n=1,705), each article's URL and eight subject-area keywords were scraped using BeautifulSoup in Python, and compared for overlap with the subject areas of the fourteen RAs from Japan. For each RA from Japan, a 'partner' RA from a Western institution was located with high subject area overlap (3-5 out of 8 subject areas, including "fMRI"). The corpora were further reduced in size by removing JP-EN candidates lacking a strongly overlapping 'partner' paper in EN-EN. The titles, authors, URLs, and eight subject areas of the 20 RAs analyzed are provided in the Appendix.

Document texts were analyzed in Jupyter Notebook using scispaCy, a Python library of free NLP tools and language models specialized for processing biomedical/clinical texts (Neumann et al.), after slight preprocessing to align sentence boundaries and remove citations. At the word level, tokens were lemmatized and POS-tagged by scispaCy's native lemmatizer/morphological analyzer. Type–token ratio was calculated by dividing each document's unique token count by its total token count. To focus on general vocabulary rather than study- or author-specific terms, relative frequencies of words/tokens across the corpora were calculated only if they appeared in at least 4 RAs: ≥2 in each of JP-EN and EN-EN. Results are reported at the document level as raw counts and per-1,000-word frequencies, and as the mean of the counts/frequencies of the 10 documents in the respective corpus. RAs in the respective corpora were comparable in length (mean word count: 6852.3 v. 6453.2; p=0.562). Corpus means were compared by independent-samples two-tailed *t*-test to identify statistically anomalous differences in usage frequencies. Corpus analysis code is available to view on a public GitHub repository.[5]

---

[3] Note that this corpus is *not* designated "Anglophone" or "American/European". Originally, the plan was to include only papers from institutions in Anglophone countries (US, UK, Canada, Australia, NZ) to ensure (relatively) homogenous L1 English background across papers. However, few such papers were good matches for the JP-EN texts: to prioritize domain similarity, better matches from Western European institutions were included as needed. Each paper was manually inspected and confirmed to have genre-conventional domain-specific language: in any case, ESP education is strong in Western Europe (EF English First 2021), and the effects of linguistic transfer between Germanic languages can be assumed to be less pronounced than between Japanese and English (Refsing & Lundquist 18-19).

[4] https://plos.org/text-and-data-mining via "Download every PLOS Article" (downloaded 2022-04-25).

[5] https://github.com/streerm/Project-Portfolio under "The Brain in Translation" (last updated 2022-05-04).

Word frequency data

Table 1. Word usage frequency data are presented below (per 1,000 words). Green, yellow, and blue denote tokens discussed below as lexical, syntactic, and pragmatic features of documents in the corpora.

Rank	Token	Frequency (per 1,000 words) JP-EN	EN-EN	P (t test)	Token	Frequency (per 1,000 words) JP-EN	EN-EN	P (t test)
1	detailed	0.171	0.038	0.005	explore	0.029	0.232	0.011
2	obtain	0.600	0.195	0.009	active	0.079	0.425	0.029
3	common	0.586	0.073	0.012	remove	0.073	0.205	0.032
4	list	0.163	0.018	0.012	vary	0.066	0.265	0.045
5	thickness	0.184	0.035	0.013	establish	0.037	0.188	0.046
6	parametric	0.362	0.036	0.015	rating	0.190	0.844	0.047
7	respectively	0.832	0.133	0.020	control	1.417	4.373	0.048
8	adopt	0.152	0.010	0.020	analyze	0.117	0.283	0.048
9	activate	1.107	0.226	0.021	duration	0.114	0.420	0.049
10	study	6.116	4.144	0.025	precede	0.011	0.209	0.054
11	enhance	0.389	0.088	0.027	performance	0.131	2.071	0.055
12	sphere	0.179	0.053	0.032	versus	0.126	0.426	0.056
13	medial	0.699	0.278	0.035	relevant	0.139	0.375	0.058
14	brain	4.506	2.617	0.045	baseline	0.115	0.751	0.060
15	planar	0.151	0.091	0.047	affect	0.462	1.390	0.060
16	table	1.375	0.718	0.049	covariate	0.012	0.285	0.061
17	evaluate	0.350	0.068	0.049	presentation	0.331	0.588	0.063
18	Brodmann	0.322	0.046	0.050	enter	0.035	0.163	0.066
19	knowledge	0.315	0.132	0.050	spatially	0.060	0.226	0.071
20	contribute	0.326	0.089	0.051	literature	0.014	0.196	0.073
21	recently	0.159	0.033	0.054	perception	0.055	0.265	0.074
22	input	0.284	0.049	0.058	separate	0.132	0.407	0.074
23	uncorrected	0.534	0.162	0.071	increase	1.032	2.218	0.075
24	report	1.565	0.788	0.074	fact	0.108	0.301	0.077
25	confirm	0.444	0.239	0.075	lead	0.156	0.447	0.082
26	clarify	0.299	0.010	0.077	processing	0.905	1.648	0.083
27	indicate	1.734	1.045	0.082	cluster	0.618	2.348	0.083
28	corresponding	0.184	0.065	0.083	research	0.142	1.237	0.085
29	system	0.931	0.484	0.084	detail	0.144	0.376	0.089
30	think	0.362	0.128	0.084	hypothesis	0.250	0.550	0.095
31	different	1.293	0.753	0.086	able	0.081	0.276	0.096
32	gyrus	2.491	1.071	0.088	give	0.455	0.921	0.096
33	statistical	0.693	0.404	0.092	effect	1.814	3.915	0.096
34	property	0.490	0.060	0.095	sensitive	0.069	0.496	0.098
35	coefficient	0.342	0.028	0.096	overall	0.129	0.299	0.102
36	length	0.541	0.087	0.102	implicate	0.054	0.237	0.103
37	bilateral	1.597	0.783	0.102	choose	0.054	0.234	0.104
38	supramarginal	0.156	0.014	0.108	healthy	0.180	1.026	0.109
39	answer	0.753	0.102	0.109	target	0.330	3.733	0.113
40	picture	1.550	0.126	0.109	accuracy	0.351	1.581	0.121

4. Selected results & discussion

This section presents a few linguistic features of terminology and phraseology in Japanese RAs, along with direct/non-standard and indirect/standard equivalents in English RAs, and how they manifested in the corpus analysis. Salient results are presented with literature discussion.

4.1 Lexis: Type/token ratio

Translated texts' tendency to have lower lexical diversity than original texts in the same genre has been well documented in the literature (e.g. Al-Shabab 102-106; Laviosa 561-564; Lei & Yang 5-7). Here, while JP-EN papers did have marginally lower type-token ratios than EN-EN papers on average, this trend was not statistically significant (24.1% vs. 25.3%: $p = 0.390$). This surprising outcome ran counter to my own observations of semantic creep as English technical terms are translated into Japanese by authors and then back into English by translators, resembling the children's game 'Chinese whispers'.

Japanese metonymy often favors a concise *kanji* or *kango* being frequently repeated in technical texts as a kind of pronoun. For example, the term *hairetsu* (配列) is often rendered as "sequence", in the sense of an ordered sequence of nucleotides, a definition which applies to genetic material whether naturally occurring (e.g. *gene, primary transcript*) or bioengineered (*oligonucleotide*, *[PCR] probe*). Let's consider the case of a Japanese author citing a DNA molecule in a published English study, in which it was referred to as an "oligonucleotide". If local precision is essential, Japanese can access the specific technical terms as katakana loanwords (*oligonucleotide*→オリゴヌクレオチド), but their length and foreign etymology makes them highly marked and potentially distracting. In contrast, *hairetsu* consists of just two kanji, making it an inobtrusive and intuitive metonym. Bilingual dictionaries typically give "sequence" as the first entry for *hairetsu*; in the translation process, the translator might settle into simplistic one-to-one equivalence as "sequence" and miss opportunities to restore the more-precise diction of "oligonucleotide." Repeated across dozens of terms, this erosion of technical terminology results in lower lexical diversity in translated documents.

This study's finding of comparable lexical diversity across corpora reflects positively on the works included from Japanese scientists contained therein, dispelling concerns of damage to semantic precision, and supporting claims that educational attainment and years working in the field are better predictors of markers of sophisticated writing such as advanced phraseology and diverse vocabulary. Due to their inexperience in the RA genre and their field, students and junior scientists are especially prone to relying on formulaic sentence patterns and limited subsets of terminology. Based on their observation that English RAs written by native-Chinese 5th-year STEM PhD candidates were consistently less lexically diverse than RAs authored by native-English experts and even 'beginners' (4th-year undergraduates and masters students), Lei & Yang conclude that "academic expertise may play a more important role than nativeness in the writing of research articles" (1). Perhaps the authors represented in JP-EN were better positioned to write sophisticated English because of their advanced education and career stage: most have doctorates or medical degrees. Complementarily, since *PLOS One* is a highly competitive scientific journal with exacting editorial criteria, it could be that RAs that end up being published are invariably well-written and domain-adherent. Going forward, it may be more instructive to compare original and translated English RAs published in a less-competitive journal or less-technical domain.

4.2 Lexis: Signs of direct transfer

Source-language interference is evidenced by the fact that many of the tokens more prevalent in JP-EN than EN-EN are dictionary equivalents of conventional verbs widespread in Japanese RAs: *evaluate* (評価; *hyōka*), *report* (報告; *hōkoku*), *confirm* (確認; *kakunin*), *clarify* (明らかにする; *akiraka ni suru*), *obtain* (得る; *eru*), and *contribute* (貢献; *kōken*). The ubiquity of these English tokens – each the first entry under its Japanese equivalent in bilingual dictionaries – suggests Japanese authors/translators should be more attentive to varying their diction using synonyms or related words (e.g., *evaluate→assess, rate;*

clarify→identify, illuminate; contribute→help). This prescription notwithstanding, these verbs would be useful Japanese-specific features for machine-learning algorithms intended to automatically identify translationese or translated documents in English corpora (Volansky 105-106).

Two non-verb tokens deserve further commentary: *recently* and *study*. The former originates from *saikin* (最近), a common sentence-initial topic in the early paragraphs of Japanese RAs for framing the current work in the broader context of recent trends and research. In English, however, the present-perfect tense can adequately convey the same temporal frame without need for explicit mention of "recent(ly)" ("*We have ~~recently~~ demonstrated that*…") Secondly, *study*'s place in the list can be attributed to a Japanese preference for referring to previous works generally as *kenkyū* (研究: 'study') or variants (esp. 本研究では, 先行研究: 'in this study', 'preceding studies') rather than naming their authors. These features thus also constitute useful markers of translationese in English RAs translated from Japanese.

Conversely, what about the tokens more prevalent in EN-EN than JP-EN? There are fewer of them—just 9 tokens under $p<0.05$ compared with 17 in JP-EN—and they resist easy interpretation by POS or topic. One whose presence may be more than coincidental is *control*. Japanese does have an equivalent loanword, which native speakers use even in daily conversation (*kontorōru*; コントロール), but native *kango* are preferred in formal writing for reasons of register. Depending on context, *control*'s meaning – in the sense of influencing or directing behaviors or processes – can be conveyed using different Japanese terms: e.g., 品質管理: *quality control*, 妊娠調節: *birth control*. But at the word level, other English words are more accurate as direct equivalents (管理: *management*, 調節: *regulation*). *Control* may thus be under-represented in JP-EN because authors/translators preferentially selected *manage*, *regulate*, or other related Japanese terms whose closest meanings overlap incompletely with its own.

4.3 Syntax: Preference for passive voice

Passive voice's history in scientific writing pedagogy is one of controversy. Historically, it was advocated as a grammatical embodiment of social values in science, acting to emphasize processes and observations in the physical world as universal and observer-independent by distancing them from the discreteness and subjectivity of human experiments (Ding 137). More recently, the passive voice has fallen out of favor: besides its wordiness, the potential to conceal and obfuscate information – or even just appear to do so – makes a good case for strategic use of both active and passive voice in research reports (Ping Alvin 13-14). Unbound by proscriptions on the first person, English writers can clearly distinguish subjective elements such as motivations and inferences – e.g., *We (hypothesize | predict | can conclude) that*… – from ostensibly observer-independent findings and methods. On the other hand, Japanese is far more permissive of passive constructions, and technical prose is no exception. Given Japanese RAs' origin in Western scientific practices, their usage seems motivated by a similar desire to distance the observed from the observer. Still, direct transfer in every instance tends to produce English that sounds uncertain, imprecise, and even deliberately secretive or misleading (see *4.5*).

For now, let's consider the token *report*, which was significantly more common in JP-EN than EN-EN (1.565 v. 0.788 [per 1,000 words]; $p=0.074$). This verb corresponds to the Japanese *hōkoku suru* (報告する), which is typically used to relay the findings of previous research without an explicit agent as the passive construction *hōkoku sarete iru* (報告されている). The English equivalent *(is | was | has been) reported* is widespread in JP-EN, accounting for ~60% of matching tokens, compared to just ~30% in EN-EN. Some J-E translators are attentive to converting passive to active voice, but in the absence of a specific author, end up introducing the vague *studies* as the grammatical subject (→*studies have shown that…*): indeed, JP-EN accounted for ~90% of all instances of this usage across the corpora. Thus, even when passive voice is minimized in translated texts, the passive origins of many sentences are still detectable. One more point deserves mention: the grammatical subject of active-voice *report* in JP-EN RAs was almost always *study* or an author's name: in a semantic sense, equivalent to *wrote* or *published*. In EN-EN, in contrast *report*'s subject was more frequently patients or individuals, in the sense of *endorsing* or *complaining of [symptoms]*.

Table 2. Selected excerpts for *report* in active and passive voice	
EN-EN *report*	● All participants in the fMRI experiment reported normal hearing, normal or corrected to normal vision, and no histo… ● …participants reported significant reductions in feelings of awareness (t(10)= 3.65; p=0.004), control (t(10)= 3.19; p=0… ● …Osumi and colleagues have recently reported higher psychopa thic tendencies in healthy males being related to attenuate…
JP-EN *report* 1.565 (/1000) v. 0.788 EN-EN (p=0.07)	● For example, increases in white matter volume have been reported throughout childhood and adolescence, particularly… ● …clinical observations related to abnormal motions are frequently reported in people with ASC, raising the possibi… ● …left dominance has been reported in relation to the linguis tic nature of the task in previous studies on perspective p…

4.4 Pragmatics: Process orientation versus result orientation

Does this tendency toward passive voice bely something deeper? Japanese has been described as a situation-focused language, based on its preference for describing states using intransitive verbs without referencing the forces that caused them, while English is considered person-focused in contrast: for an everyday example, compare the Japanese *onaka ga suita* (lit. '[my] stomach emptied') with the English *I am hungry*. This parallels another dichotomy: while Japanese preferentially attends to the 'how' of events and processes, English emphasizes the 'what' of results and properties (Wakabayashi 92-94). Grammatically, one could even reformulate this difference between the two languages as 'verb-focus' versus 'noun- and adjective-focus.'

This feature seems exemplified in the corpora by the dyad of *activate/active*, which are significantly concentrated in JP-EN/EN-EN (respectively: 1.107 v. 0.226, p=0.021; 0.079 v. 0.425, p=0.029). The process/result in question is "activation", meaning enhanced blood flow or glucose uptake in a specific brain region on fMRI, purported to correspond to greater mental activity in the neurological or cognitive processes governed by that region. EN-EN RAs tend to describe this phenomenon using the adjective *active*, signifying either a property of a region under observation or else the result of a certain psychological task. In contrast, the regular use of both active and passive forms of *activate* (活性化・活動する) in JP-EN are artefacts of how the same phenomenon is instead encoded as a process in Japanese.

Table 3. Selected excerpts for English result orientation (*active*) versus Japanese process orientation (*activate*)	
EN-EN *active (adj.)* 0.425 (/1000) v. 0.079 JP-EN (p=0.03)	● Frontal…regions were shown to be active in both the block and event-related analyses, suggesting their importance… ● …a large cluster in the right and a smaller cluster in the left premotor cortex…were significantly more active in the… ● …all action observation ROIs were significantly more active for action conditions than for the rest conditions: left…
JP-EN *activate (v.)* 1.107 (/1000) v. 0.226 EN-EN (p=0.02)	● For example, the precuneus is activated during 'forgivabili ty' judgments in social scenarios and in the attribution… ● and that L. AG/SMG was also activated during the identifica tion of correct past-tense forms of verbs, probably as a… ● whereas implicit ES activated the lateral prefrontal cortex and medial/lateral parietal cortex, explicit ES activated…

Curiously, there is at least one example where Japanese appears to emphasize property over process: *detailed (adj.)* versus *(in) detail (adv.)*. *Detailed* corresponds to the adjective *shōsai na* (詳細な),

frequently present in collocations endemic to Japanese scientific prose (e.g. ～情報: *detailed information*, ～説明: *detailed explanation*). In EN-EN, in contrast, *detail* appeared most frequently as an adverbial phrase modifying a process (e.g., *explained in detail*). Global generalizations across languages, such as state/process orientation in English/Japanese, should thus be interpreted with caution: translators should be equally wary of exceptions to such rules that can still mark translated works as translationese.

4.5 Pragmatics: Passive-voice hedging (*is thought to*)

Let's return to the passive voice, now attending to its role in the discourse function of hedging. Passive forms are often assumed to merely express a passive meaning, but can also indicate non-intentional or involuntary actions that would be conventionally expressed with an agent in English (Wakabayashi 120). This usage category encompasses mental-state verbs expressing the writer's feelings or perceptions, as exemplified by *kangaerareru*, the passive/potential form of *kangaeru* (*to think*/*consider*) (Alfonso 952-953). Put simply, *kangaerareru* is a rhetorical device intended to couch such conclusions in a frame of objectivity ('This conclusion was reached naturally and unconsciously from the observations; in the face of the evidence, the same *would be thought* by anyone'). Likewise, the passive *to sareru* serves a similar hedging function, denoting a theory or consensus in a field while stopping short of declaring it established fact. Amateur J-E translators often relay such 'thinker-less thoughts' via passive constructions such as *is considered* or *is thought [that]*. The fact that a conclusion is putative, but perhaps not 100% established truth, can be expressed more conventionally in English using a modal like *may* or *could*, an adverb such as *probably* or *likely*, or even omitting the verb entirely (Wakabayashi 121).

Indeed, JP-EN has unusually many instances of passive *to be thought* serving this discourse function, which seems to be responsible for the lemma's higher prevalence in this corpus overall. In EN-EN, the word is rarely used to hedge; rather, it denotes *thinking* in a general sense, or else signifies the mental activity of study subjects. Such usages are also present in JP-EN; however, over half of the instances serve to hedge, compared with just a single instance in all of EN-EN.

Table 4. Selected excerpts exhibiting hedging (Japanese) and non-hedging (English) usages of *think*

EN-EN *think*	• Participants were presented with an example generation cue word for 12s and instructed to think of at least seven exam… • Our findings…suggest that thinking about harming them is more distressing, leads to more empathic responses, evokes … • Subjects were told to explicitly think about the rules and expected outcome signaled by each instruction cue.
JP-EN *think* 0.362 (/1000 v. 0.128 EN-EN (p=0.08)	• …regions are considered to underlie empathy…and uncertainty judgments, which are thought to be involved in social judgm… • The FP and CO networks are thought to be critical for adaptive control and stable set maintenance, respectively. • Furthermore, typing is thought to involve more parallel processing, such that prior to the execution of one letter …

4.6 Limitations

The results of this preliminary investigation should be interpreted cautiously for a couple of reasons. First, its focus was confined to relatively common tokens for methodological reasons: since each corpus contains only 10 documents, each a highly specialized work applying fMRI to visualize different cognitive processes, I decided to consider only tokens present in ≥20% of documents in both corpora. Some document-specific tokens that did not meet these criteria might nonetheless have divergent source/target distributions that would emerge in larger corpora. Second, the process by which each JP-EN paper was drafted and revised is unknown, and likely inconsistent across papers. The corpus is likely a mixture of texts revised by L1 English editors after being written by Japanese authors and other texts translated from Japanese by technical translators, potentially without revision. Regardless of the writer's

native language, knowledge of conventional language in English RAs would act to minimize source-language artefacts. Similar investigations confined exclusively to RAs translated from Japanese – or only RAs written by Japanese scientists in their second language – could help to determine which aspects of translationese originate in L2 Japanese reception versus L2 English production.

5. Conclusion

This study identified a number of lexical, syntactic, and pragmatic features characteristic to original RAs authored by Japanese neuroscientists, which differentiated them from comparable texts written by their Western peers. These differences can inform Japanese researchers and translators about how to improve their scientific communications by delivering domain-specific language expected by their audiences.

Science is implicitly seated in objectivity: its objects of inquiry are expected to be invariant across different observers. Neuroscience assumes discoveries about the grey-white mass of neural tissue inside the skull to be extrapolatable to all humans, regardless of whether they verbalize it as *brain* or 脳. But beyond the realm of the anatomical, are psychological insights about human cognition necessarily universal to the same degree? The Sapir–Whorf hypothesis holds that language shapes cognition: could differences in how "thought" is conceptualized by Japanese and English neuroscientists be identified based on differences in linguistic features across larger corpora? Going forward, I hope to explore this question in search of such insights at the intersection of linguistics and comparative psychology.

6. Works Cited

al-Shabab, Omar Sheikh. *Interpretation and the Language of Translation: Creativity and Conventions in Translation.* Janus Publishing Company, 1996.

Alfonso, Anthony. *Japanese Language Patterns: A Structural Approach.* Sophia University L. L. Center of Applied Linguistics, 1971.

Ding, Daniel D. "The Passive Voice and Social Values in Science." *Journal of Technical Writing and Communication*, vol. 32, no. 2, 2002, pp. 137-54.

Eskola, Sari. "Untypical Frequencies in Translated Language: A Corpus-Based Study on a Literary Corpus of Translated and Non-Translated Finnish." *Translation Universals: Do They Exist?*, edited by Pekka Kujamäki and Anna Mauranen, J. Benjamins Publishing Company, 2004, pp. 83-99.

First, EF Education. "EF English Proficiency Index - a Ranking of 112 Countries and Regions by English Skills." EF Education First, 2021, https://www.ef.com/wwen/epi/.

Ingraham, Christopher. "The Languages That Let You Say More with Less." *Washington Post*, 28 Sep. 2017, https://www.washingtonpost.com/news/wonk/wp/2017/09/28/the-languages-that-let-you-say-more-with-less/.

Laviosa, Sara. "Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose." *Meta: Translators' Journal*, vol. 43, no. 4, 1998, pp. 557-70.

Lei, Siyu and Ruiying Yang. "Lexical Richness in Research Articles: Corpus-Based Comparative Study among Advanced Chinese Learners of English, English Native Beginner Students and Experts." *Journal of English for Academic Purposes*, vol. 47, 2020, p. 100894.

Mossop, Brian et al. *Revising and Editing for Translators.* 4 ed., Routledge, 2019.

Neumann, Mark et al. "Scispacy: Fast and Robust Models for Biomedical Natural Language Processing." *Proceedings of the 18th BioNLP Workshop and Shared Task*, 2019, pp. 319-27.

Ping Alvin, Leong. "The Passive Voice in Scientific Writing. The Current Norm in Science Journals." *Journal of Science Communication*, vol. 13, no. 1, 2014, p. A03.

Refsing, Kirsten and Lita Lundquist. *Translating Japanese Texts.* Museum Tusculanum, 2009.

Shuttleworth, Mark and Moira Cowie. *Dictionary of Translation Studies.* St. Jerome Publishing, 1997.

Volansky, Verod et al. "On the Features of Translationese." *Digital Scholarship in the Humanities*, vol. 30, no. 1, 2015, pp. 98-118.

Wakabayashi, Judy. *Japanese-English Translation: An Advanced Guide.* Routledge, 2020.

7. Appendix

Supplementary Table 1. Research articles included in corpora

ID	First author	Date	Title (URL hyperlink)	Subject Area 1	Subject Area 2	Subject Area 3	Subject Area 4	Subject Area 5	Subject Area 6	Subject Area 7	Subject Area 8
JP	Nakata	2019	Negative BOLD responses during hand and foot movements: An fMRI study	fMRI	Parietal lobe	Neural networks	Hands	Frontal lobe	Right hemisphere	Body limbs	Temporal lobe
EN (NL)	Jansma*	2013	fMRI Guided rTMS Evidence for Reduced Left Prefrontal Involvement after Task Practice	fMRI	Parietal lobe	Neural networks	Hands	Prefrontal cortex	TMS	Learning	Reaction time
JP	Nakai	2017	Sense of Accomplishment Is Modulated by a Proper Level of Instruction and Represented in the Brain Reward System	fMRI	Emotions	Sensory perception	Neuro-imaging	Human learning	Problem solving	Prefrontal cortex	Motivation
EN (UK)	Deeley	2013	Using Hypnotic Suggestion to Model Loss of Control and Awareness of Movements: An Exploratory fMRI Study	fMRI	Emotions	Sensory perception	Neuro-imaging	Body limbs	Eyes	Cerebellum	Hands
JP	Higashiyama	2015	The Neural Basis of Typewriting: A Functional MRI Study	fMRI	Working memory	Cerebellum	Neuro-imaging	Hands	Thalamus	Cognition	Lesions
EN (AU)	Neale	2015	Functional Activation during the Rapid Visual Information Processing Task in a Middle Aged Cohort: An fMRI Study	fMRI	Working memory	Cerebellum	Neuro-imaging	Parietal lobe	Vigilance	Reaction time	Attention
JP	Murakami	2015	Neural Networks for Mindfulness and Emotion Suppression	fMRI	Cognition	Emotions	Prefrontal cortex	Psychological stress	Artificial intelligence	Neural networks	Amygdala
EN (AT)	Majdandžić*	2012	The Human Factor: Behavioral and Neural Correlates of Humanized Perception in Moral Decision Making	fMRI	Cognition	Emotions	Prefrontal cortex	Decision making	MRI	Priming	Prosocial behavior
JP	Itahashi	2014	Altered Network Topologies and Hub Organization in Adults with Autism: A Resting-State fMRI Study	fMRI	Neural networks	Adolescents	Clustering coefficients	Adults	Specimen disruption	Autism	Cognitive impairment
EN (DE)	Lidzba*	2013	Complex Visual Search in Children and Adolescents: Effects of Age and Performance on fMRI Activation	fMRI	Neural networks	Adolescents	Parietal lobe	Attention	Occipital lobe	Vision	Frontal lobe
JP	Yomogida	2014	The Neural Basis of Event Simulation: An fMRI Study	fMRI	Cognition	Prefrontal cortex	Social cognition	Cingulate cortex	Theory of mind	Echo planar imaging	Perception
EN (CA)	Dixon	2012	The Decision to Engage Cognitive Control Is Driven by Expected Reward-Value: Neural and Behavioral Evidence	fMRI	Cognition	Prefrontal cortex	Sensory cortex	Motivation	Habits	Decision making	Learning

JP	Minamoto	2014	Extrapunitive and Intropunitive Individuals Activate Different Parts of the Prefrontal Cortex under an Ego-Blocking Frustration	fMRI	Aggression	Prefrontal cortex	Behavior	Emotions	Parietal lobe	Cognition	Hypo-thalamus
EN (DE)	Pawliczek*	2013	Anger under Control: Neural Correlates of Frustration as a Function of Trait Aggression	fMRI	Aggression	Prefrontal cortex	Behavior	Emotions	Amygdala	Impulsivity	Frontal lobe
JP	Ohta	2013	Syntactic Computation in the Human Brain: The Degree of Merger as a Key Factor	fMRI	Language	Memory	Linguistic morphology	Vowels	DTI	Syntax	Grammar
EN (US)	Himmelstein	2018	Linguistic analysis of the autobiographical memories of individuals with major depressive disorder	fMRI	Language	Memory	Depression	Amygdala	Mental health	Memory recall	Emotions
JP	Watanabe	2012	Diminished Medial Prefrontal Activity behind Autistic Social Judgments of Incongruent Information	fMRI	Emotions	Facial expressions	Behavior	Amygdala	ASD	Autism	Social communication
EN (US)	Kim	2016	Identifying Core Affect in Individuals from fMRI Responses to Dynamic Naturalistic Audiovisual Stimuli	fMRI	Emotions	Facial expressions	Behavior	CNS	Sensory perception	Neuro-imaging	Vision
JP	Tamura	2012	Neural Network Development in Late Adolescents during Observation of Risk-Taking Action	fMRI	Emotions	Neuro-imaging	Personality	Social psychology	Adolescents	Depression	Anxiety
EN (US)	Sobhani	2012	Interpersonal Liking Modulates Motor-Related Neural Regions	fMRI	Emotions	Neuro-imaging	Personality	Social communication	Macaque**	Monkeys**	Sensory perception

Abbr.: ASD: autism spectrum disorder, AT: Austria, AU: Australia, CA: Canada, CNS: central nervous system, DE: Germany, DTI: diffusion tensor imaging, MRI (fMRI): (functional) magnetic resonance imaging, NL: Netherlands, TMS: transcranial magnetic stimulation, UK: United Kingdom, US: United States.

* Denotes RAs whose author affiliations include only institutions in non-Anglophone Western countries.

** Keywords refer to primate research only mentioned in study introduction; not relevant to study scope or methods.