International Journal of English Language & Translation Studies
ISSN: 2308-5460

Creating a Translation Glossary Using Free Software: A Study of Its Feasibility with Japanese Source Text

[PP: 151-160]

Stephen Crabbe
School of Languages and Area Studies, University of Portsmouth
United Kingdom

David Heath
College of Intercultural Studies, Kanto Gakuin University
Japan

ABSTRACT

In this paper, we (a) explain how translators can benefit from creating their own glossaries; and (b) evaluate how easily a translation glossary can be created from Japanese source text using free software applications. As our study shows, a major hurdle arises from the fact that Japanese text does not include spaces; it must be segmented, i.e., broken into "usable chunks" (Fahey, 2016), before a concordancer (in our case, AntConc 3.2.4) can be used to analyze it for glossary creation. We segmented our Japanese text using an application (ChaSen 2.1) designed for this purpose. This application's output was problematic, forcing us to devise workarounds that became labour-intensive and time-consuming. Our completed glossary (shown in Appendix 1) is fit for purpose, but the complications in the process of creating it call into question the feasibility of using free software to make translation glossaries from text written in Japanese.

Keywords: *Translation glossary creation, Japanese text, Concordancers, Text segmentation, AntConc 3.2.4, ChaSen 2.1*

ARTICLE INFO	The paper received on	Reviewed on	Accepted after revisions on
	05/08/2017	**05/09/2017**	**16/10/2017**

Suggested citation:
Crabbe, S. & Heath, D. (2017). Creating a Translation Glossary Using Free Software: A Study of Its Feasibility with Japanese Source Text. *International Journal of English Language & Translation Studies.* 5*(3). 151-160.*

1. Introduction

In this paper, we draw on our experience as professional Japanese-to-English translators and translation scholars to (a) explain how translators can benefit from creating their own glossaries; and (b) evaluate how easily a translation glossary can be created from Japanese source text using the free software applications AntConc 3.2.4 (Anthony, 2014) (a concordancer) and ChaSen 2.1 (Matsuda, 2000) (a segmenter for Japanese text). We take a concordancer-based approach to glossary term selection (as opposed to using automatic term selection tools) as, inter alia, it is fundamentally "simple" (Muegge, 2013) and gives a degree of control that can be valuable in addressing the challenges "of "noise" (i.e., invalid term candidates) and "silence" (i.e., missing legitimate term candidates)" (Muegge, 2013).

For translators (especially those working with texts on technical or otherwise specialized subjects), a key to translation quality is "lexical congruency" (Stitt, 2017), i.e., using target-language terminology consistently. Simply stated, it is important (and arguably essential) to always use the same term as a label for the same thing or concept (Stitt, 2017). One method that translators use to maintain "lexical congruency" (Stitt, 2017) is to develop glossaries. So what *is* a glossary?

> "A glossary is essentially a list of terms in one or more languages. [...] the most basic glossary will simply contain lists of terms and their equivalents in one or more foreign languages. [...] At the other end of the glossary spectrum, you will find richly detailed glossaries containing definitions, examples of usage, synonyms, related terms, usage notes, etc. These are the glossaries which every translation student [...] dreams of having because they can use them to understand terms, to identify equivalents, to learn how to use terms [...]". (Bowker & Pearson, 2002, pp. 137-138)

A glossary has some similarities to a dictionary. However, dictionaries are often less useful than glossaries for translation that involves language for special purposes (LSP). One shortcoming of dictionaries "is their inherent incompleteness. The world around us and the language used to describe it are evolving

all the time, which means that printed dictionaries go out of date very quickly" (Bowker & Pearson, 2002, p. 15).

Another shortcoming of dictionaries is their size. Bowker and Pearson (2002, p. 15) make clear that "Although it is possible to compile large, multi-volume dictionaries that attempt to cover a specialized subject, not many people will be able to afford such dictionaries and [...] would not want to carry them around". Because of size limitations, "lexicographers who create [...] dictionaries have to choose which information to include and which to leave out. Unfortunately, their choices do not correspond with the needs of LSP users" (Bowker & Pearson, 2002, p. 15).

Dictionaries are also criticized for not giving enough "contextual or usage information. LSP learners must pay attention to how terms are used, which means that in addition to information about what a term means, they also need information about how to use that term in a sentence" (Bowker & Pearson, 2002, p. 16). Further, "most dictionaries [...] cannot easily provide information about how frequently a given term is used" (Bowker & Pearson, 2002, p. 16) even though such information can facilitate informed decisions about the appropriateness of lexical choices (Bowker & Pearson, 2002, p. 16).

A self-created glossary based on a corpus ("a body of text" (Bowker & Pearson, 2002, p. 9)) of the translator's own choice or design can be free of the aforementioned shortcomings of dictionaries. But how can translators create their own glossaries using freely available software? And how easy is this process when the source text is written in Japanese?

2. Literature Review

Lexicography (the activity of editing and/or compiling dictionaries) was originally a slow and painstaking process. The effort to define a word and sort its uses involved working with "slips of paper (called *citations*), each consisting of a quoted passage containing the word under discussion" (Landau, 2001, p. 44). Compilation of the first edition of the *Oxford English Dictionary* "took 70 long years of terrible labour" (Winchester, 2004, p. XXV). And despite the effort involved, citation-based dictionaries were fundamentally flawed. Content selection depended heavily on lexicographers' intuition and was thus subject to their "prejudices and preferences" (Krishnamurthy, 2002, p. 23). Further, they were inherently incomplete. Even the *Oxford English Dictionary* "managed only a piecemeal coverage" (Krishnamurthy, 2002, p. 23). Today, printed dictionaries still suffer from "inherent incompleteness" (Bowker & Pearson, 2002, p. 15), and from inclusion of "linguistic deadwood" (Bowker & Pearson, 2002, p. 15).

Lexicography underwent a dramatic change from the mid-1980s to the mid-1990s owing to vast increases in the power of file servers and to vast increases in the power of hard drives in desktop computers (Landau, 2001, p. 2). Perhaps most importantly, computers enabled lexicographers to collate "huge electronic collections of naturally occurring language (called *corpora*, singular *corpus*, meaning "body" in Latin)" (Landau, 2001, p. 2) and use them "to study and analyze language use in ways that were not possible before" (Landau, 2001, p. 2). Computer-held corpora can be massive. For instance, the Collins Corpus contains more than 4.5 billion words ("The Collins Corpus", 2016).

A large computer-held corpus "can be far more comprehensive and balanced than any individual's language experience" (Krishnamurthy, 2002, p. 23). Perhaps its chief merit is that it can give objective evidence of real-world language usage in terms of "how words are used, what they mean, which words are used together, and how often words are used" ("The Collins Corpus", 2016).

Computer-held corpora can be of great benefit to translators. They can be of particular benefit to technical translators, who need to learn and replicate the real-world usage of LSP, i.e., "the language that is used to discuss specialized fields of knowledge" (Bowker & Pearson, 2002, p. 25). As Bowker and Pearson (2002, p. 19) point out:

> "Since corpora are comprised of texts that have been written by subject field experts, LSP learners have before them a body of evidence pertaining to the function and usage of words and expressions in the LSP of the field. Moreover, with the help of corpus analysis tools, you can sort these contexts so that meaningful patterns are revealed. In addition, a corpus can give an LSP learner a good idea of how a term or expression *cannot* be used."

A computer-held LSP corpus and a concordancer—a computer program that allows the user to see each occurrence of a chosen word in its immediate context as a key-word-in-context (KWIC) concordance

and to perform statistical analysis on the corpus—can enable a translator to create an LSP glossary as an aid for producing target-language text that conforms to the real-world usage of target-language LSP terms.

By using the concordancer to (a) list the words in the corpus in order of frequency and/or alphabetically and (b) produce, sort, and compare KWIC concordances, the translator can identify term candidates for the glossary, ascertain which term candidates are actual terms (words and/or compounds "that are used in a specialized domain and have a clearly defined meaning" (Bowker & Pearson, 2002, p. 145)) worthy of inclusion in the glossary, and collate examples of real-world usage of the terms. Using the same tools, the translator can also "gain conceptual information, such as knowledge about the characteristics of the concepts behind the terms and the relationships concepts have with each other" (Bowker & Pearson, 2002, p. 39). The translator can use such conceptual information to produce source- and/or target-language definitions of the terms, optionally combining said information with his/her own knowledge and/or with definitions in other sources, e.g., conventional LSP dictionaries.

With some corpus-processing programs, the process of identifying term candidates can be semi-automated by means of a function that identifies "words which occur with an unusually high frequency in a text or corpus when that text or corpus is compared with another corpus" (Bowker & Pearson, 2002, pp. 114-115) and ranks words "according to 'keyness' rather than according to frequency" (Bowker & Pearson, 2002, p. 115) such that "the 'key' words float to the top" (Bowker & Pearson, 2002, p. 115). (We did not use such a function in our study as we created our glossary using a single corpus.)

Conventional monolingual LSP dictionaries "tend to concentrate on providing information about the meaning rather than the usage of terms. Consequently, they will not usually provide grammatical information or examples of usage" (Bowker & Pearson, 2002, p. 139). And in conventional bi-/multi-lingual LSP dictionaries, "definitions are rarely provided and the emphasis is mainly on providing equivalents and examples of usage" (Bowker & Pearson, 2002, p. 140). An LSP glossary produced using a computer-held LSP corpus and a concordancer can be free of all of these shortcomings and can thus be of significantly greater utility. The benefits of glossary compilation are highlighted by translation providers such as Integro Languages (2017) and Lionbridge (2016). Moreover, corpus building and glossary compilation are, as highlighted by the European Graduate Placement Scheme's occupational standards for European postgraduate translation students on work placement, key practical skills for providers of translation services (European Graduate Placement Scheme, n.d.).

3. Methodology

3.1 Corpus Design

The corpus we selected for our study is the source text of one of our own past Japanese-to-English translation projects: a product guidebook produced in 2009 by a Japanese automaker to give overseas distributors an overview of a car (an updated version of an existing model) that the automaker was preparing to launch. (For confidentiality reasons, we are excluding identifying information about the automaker from this paper.) The product guidebook's recent publication date suggests that the corpus adequately reflects "the current state of the language and subject field" (Bowker & Pearson, 2002, p. 54).

The corpus was written by a subject expert (a native-Japanese-speaking automotive copywriter) with editorial oversight from subject experts (automaker headquarters staff responsible for providing overseas distributors with product information and marketing materials). The authorship and editorial oversight suggest that the corpus contains "more authentic examples of LSP use" (Bowker & Pearson, 2002, p. 54) than it would have contained if it had been written by people who are not proven experts. We infer from translating similar Japanese texts that the users of the target text are also subject experts.

The corpus contains about 18,000 characters. Based on the Japanese-to-English translators' rule of thumb that 400 Japanese characters (the number that fit on a traditional Japanese manuscript sheet) of source text correspond to about 200 words of English target text, the corpus corresponds to about 9,000 English words. Bowker and Pearson (2002, p. 48) say that corpora ranging from about 10,000 words to several hundreds of thousands of words have proved useful in terms of enabling LSP claims to be made based on statistical frequency. By this measure, the size of our corpus appears to be minimally acceptable.

The product guidebook contains chapters on the car's design (i.e., styling); driving dynamics (engines, transmissions, and technologies related to steering, handling, and ride quality); craftsmanship (measures taken by the automaker to create a refined look and feel); and safety. It complies with Bowker and Pearson's (2002, p. 49) recommendation to use full texts (rather than extracts) in order to avoid accidentally eliminating useful content. However, the breadth of its coverage (the whole car) suggested from the outset that the number of times a given term appears—and the number of contexts in which it appears—could be small.

Partly in light of experience of translating texts similar to our corpus and partly in light of secondary literature (e.g., Takeuchi, Kageura, Koyama, Daille, & Romary, 2003), we assumed from the outset that much (perhaps most) of the lexical content relevant to glossary production consisted of nouns and/or noun-based expressions. Also, our corpus reflects the strong tendency of Japanese to omit subjects and leave the reader to infer them from context. For instance, a passage about the car's styling contains the following sentence: 「エクステリアでは、新しいファミリーフェイスを採用しました。」 [lit. On the exterior, [we] adopted [a] new family face.], where the omitted subject can be inferred as the automaker.

3.2 User Assumptions and Glossary Design

Our assumptions about the likely user of our glossary influenced our criteria for term selection and our design of glossary entries.

We have been translating technical texts for decades. We know from this experience that a translator can become overwhelmed with work under intense deadline pressure and need other translators' help. Also, our experience suggests that native-English-speaking Japanese-to-English translators with specialized automotive knowledge are few and far between. Consequently, the intended user of our glossary for the purposes of this study is a native-English-speaking Japanese-to-English freelance translator who is technically inclined and has an interest in cars but is not thoroughly familiar with key terms and concepts in distributor-oriented texts written in Japanese by Japanese automakers. (We excluded native Japanese speakers from our user hypothesis for two reasons: (1) Our experience suggests that their output is more prone to being unduly affected by what Baker (1992, p. 54) calls the "engrossing effect of the source text patterning". (2) The Japan Translation Federation states in its guide for translation buyers that 「外国語の文書を母国語に翻訳するのがプロの原則です」 [lit. It is a fundamental principle that professional translators work into their native languages.] (Japan Translation Federation, 2012, p. 15).) While bearing in mind the relevance of the frequency list produced by our concordancer, we therefore strove to

- exclude from the glossary any term for which we felt that a literal translation would, even if the translator did not have a complete grasp of the concept behind it, be likely to be correct;
- include any term for which we felt that a literal translation would not be correct owing, for example, to idiosyncratic usage of the term by the automaker or by the wider Japanese motor industry; and
- exclude what Bowker and Pearson (2002, p. 103) say is often called "subtechnical vocabulary, i.e., vocabulary that is used in specialized domains but not exclusively in any one domain".

We know from our professional experience that it is possible to know the meaning of a Japanese term that contains kanji (the Chinese-rooted logograms used in Japanese writing) without being able to remember its pronunciation (or without even knowing its pronunciation in the first place). Knowing the correct pronunciation can be vital for project-related meetings and telephone calls. For any term that includes kanji (with or without an auxiliary verb in hiragana (one of the two Japanese syllabaries used in conjunction with kanji)), we therefore added the pronunciation of the whole term in hiragana in brackets. We assumed that the glossary user would not need a romanized representation of any Japanese term.

Each entry in our glossary begins with the Japanese term in question (shown without a romanized representation) and continues with the term's word class (e.g., noun), our suggested English term, the domain in which the terminology is used, the source of our information (in most cases our own research and/or knowledge, signified by our combined initials, SCDH), and an example of a context in which the

Japanese term occurs within our corpus. Some entries also include a note on, inter alia, idiosyncratic use of the Japanese term by the automaker. This design for glossary entries enables us to give the user comprehensive information that s/he can use for translation without needing to refer to more sources. A sample glossary entry is shown below.

緊急制動 [きんきゅうせいどう]
Grammar noun *English* emergency braking *Domain* automobiles *Definition* Using a vehicle's brakes to bring the vehicle to a stop as quickly as possible (typically in order to avoid an accident). *Source* SCDH (July 2017) *Context*「４輪の ABS センサーから緊急制動を検出し、緊急制動信号を発信する。」

3.3 Software Selection

Our professional experience suggests that relatively few freelance Japanese-to-English translators are keen to spend money on software when functionally comparable freeware is available. Our experience also suggests that relatively few freelance Japanese-to-English translators can use programming languages (e.g., Python) or a command-line interface and that most freelance Japanese-to-English translators use a Windows or Macintosh operating system. Further, our experience suggests that confidentiality requirements imposed by commercial translation clients preclude any uploading of source text to third-party online services. We therefore decided that any software application we used for glossary creation should be Windows- and/or Macintosh-compatible freeware with a simple double-click installer and an intuitive graphical user interface.

One essential software application was a concordancer. Methods for using a concordancer in glossary creation are, we feel, adequately explained in secondary literature, e.g., Bowker and Pearson (2002). A number of concordancers are available for widely used operating systems. We selected the free concordancer AntConc 3.2.4 (Anthony, 2014). The version we selected is not the latest, which is AntConc 3.4.4 (Anthony, 2016). We used this earlier version as we were already familiar with it and were satisfied with its functionality for the purposes of our study.

Another essential application was a segmenter for Japanese text. We selected the free segmenter ChaSen 2.1 (Matsuda, 2000). The age of the application and an apparent lack of updates from its developer initially gave us pause. However, we were reassured by evidence that it has continued to be used in Japanese linguistic research, e.g., Breen (2010, pp. 13-22). Also, AntConc's developer, Laurence Anthony, had stated in personal communication with one of the authors that ChaSen was the most common application of its kind in Japan. Late in our study, we became aware that Anthony had released a segmenter, SegmentAnt (Anthony, 2017), that also appeared to meet our criteria. We intend to utilize this free software application in a future study.

4. Analysis and Discussion

Japanese text typically does not include spaces to show where one word ends and the next begins. This characteristic of Japanese text was not a problem for concordancing, but it forced us to extensively process the corpus before we could use our concordancer, AntConc 3.2.4 (Anthony, 2014), to create frequency and alphabetical lists.

The initial challenge in this study was to parse the corpus. AntConc 3.2.4 (Anthony, 2014) does not have the ability to parse texts. Notwithstanding the existence of ChaSen 2.1 (Matsuda, 2000), we initially experimented with manual segmenting, i.e., parsing the corpus by manually inserting spaces. Since we had assumed from the outset that much (perhaps most) of the lexical content relevant to glossary production consisted of nouns and/or noun-based expressions, our manual parsing involved, inter alia, splitting nouns away from modifiers that cause them to function verbally or adjectivally. Our rationale for splitting nouns away from modifiers was that we would at least be able to use the concordancer to identify every instance of noun-based compounds. Manually parsing the corpus was tedious and time-consuming; it involved about 10,000 depressions of the space bar and arrow keys on the computer keyboard and took about 10 hours. Unfortunately, the results proved unusable as, despite our best intentions, we had not been consistent in our splitting of nouns away from modifiers. At this point, we decided to parse our corpus with ChaSen 2.1 (Matsuda, 2000).

ChaSen 2.1 (Matsuda, 2000) did not yield immediately usable results as it parsed many multi-character terms incorrectly. (For example, it split「フェイスリフト」[lit. facelift] into its two constituent nouns and showed them as separate terms.) We had to clean up the results by, inter alia, manually removing hundreds of line breaks—a process that took several hours.

Even more manual processing then proved necessary as the frequency and alphabetical lists shown by AntConc 3.2.4 (Anthony, 2014) at this stage contained a great deal of "noise" (Bowker & Pearson, 2002, p. 169) in the form of numerals, English words, and noun modifiers. (A sample of the frequency list at this stage is shown in Appendix 2.) Some of the noun modifiers were written in hiragana. We considered keeping them in the corpus and using the concordancer to create a stop list for them, but we realized that such a stop list was not viable as it would have also caught genuine term candidates that were written in hiragana. Manually removing the "noise" (Bowker & Pearson, 2002, p. 169) took several hours. The manual cleanup necessitated further extra work, but we were at least confident that the results would be internally more consistent than the results of our earlier, abortive manual parsing. The resulting corpus content is predominantly nominal. Since we had assumed from the outset that many or all of our term candidates would be nominal, we were not unduly concerned about the loss of non-nominal content.

We were now able to use AntConc 3.2.4 (Anthony, 2014) to produce a usable frequency list (see the sample in Appendix 3) and a usable alphabetical list (see the sample in Appendix 4). The frequency list was of essential utility. However, the alphabetical list suggested that the frequency list was not a sufficient basis for deciding which terms to include in the glossary. Notably, the alphabetical list revealed that certain terms appeared in the corpus both in isolation and as parts of larger compounds. Whereas the frequency list showed the term 「減衰」 [lit. damping] in 904th place with a single appearance, for example, the alphabetical list revealed that the term also appeared in compounds such as 「減衰力」 [lit. damping force] and 「振動減衰性」 [lit. vibration-damping performance]. By additionally using AntConc 3.2.4 (Anthony, 2014) to produce KWIC concordances, left-sorted concordances, and right-sorted concordances for term candidates, we were able to discover the full range of compounds containing term candidates. The noun modifiers appearing before and/or after term candidates appeared to be "subtechnical vocabulary" (Bowker & Pearson, 2002, p. 103). We assumed that literal translation of such noun modifiers would yield correct translations provided the terms they modified were correctly translated. We therefore excluded such noun modifiers from the glossary.

In light of our user assumptions, we feel that our glossary (shown in Appendix 1) is fit for purpose. It is certainly free of the main shortcomings of dictionaries (outlined earlier in this paper). One potential enhancement to our glossary relates to formatting. We created the glossary as text blocks (one block per entry) to give ourselves maximal freedom to lengthen, shorten, and otherwise manipulate the entries as we refined them. Had we instead created the glossary in a Microsoft Excel spreadsheet, it would potentially have been more readily convertible into a termbase for computer-assisted translation software.

5. Summing Up

The advantages of a corpus-based glossary over a conventional dictionary are underscored by Firth's observation (1957, p. 179, cited by Storjohann, 2010, p. 6) that we "shall know the meaning of a word by the company it keeps". That said, our experience in this study of taking a corpus-based approach to the creation of a translation glossary suggests that such an undertaking is challenging when the corpus language is Japanese. The main challenge appears to be rooted in the fact that Japanese typically does not use spaces to mark boundaries between words. The need to parse the corpus using ChaSen 2.1 (Matsuda, 2000) and then spend many hours manually cleaning up the results before we could analyze them with AntConc 3.2.4 (Anthony, 2014) made glossary production extremely time-consuming and made us suspect that Japanese is unsuited to such an undertaking. Our suspicion is underscored by the existence of a University of Tokyo website ("Senmon yōgo kīwādo jidō chūshutsu sābisu gensen web", n.d.) that gives access to a system that automatically extracts domain-specific terms from inputted Japanese texts, thereby apparently precluding the need to parse Japanese texts with software such as ChaSen 2.1 (Matsuda, 2000), clean them up manually, and analyze them with a concordancer.

However, we remain convinced of the fundamental value of translation glossaries. We see no reason to doubt that Japanese-to-English translators (especially those working with texts on technical or otherwise specialized subjects) can benefit long-term from taking the time to create them. For a follow-up study, therefore, we plan to investigate whether other techniques

and/or other free software applications e.g., SegmentAnt (Anthony, 2017), would enable translation glossaries to be created from Japanese source text more quickly and easily.

About the Authors

Dr. Stephen Crabbe (PhD) is a Senior Lecturer in Applied Linguistics and Translation (Japanese to English) at the University of Portsmouth in the UK. Prior to coming to Portsmouth, he worked in Japan as a translator and interpreter. His research interests include written and visual technical/professional communication, English language learning and teaching in Japan and Japan studies, and these research interests are reflected in his teaching, publications and presentations.

David Heath is an Associate Professor responsible for translation studies at Kanto Gakuin University in Japan. He is also the managing director of a translation-focused Japanese media company that serves the TV and automotive industries. He holds a distinction-ranked MA in Translation Studies from the University of Portsmouth. He is a Chartered Linguist and a Fellow of the Chartered Institute of Linguists.

References

Anthony, L. (2014). AntConc (Version 3.2.4) [Computer software]. Tokyo, Japan: Waseda University. Retrieved March 13, 2017, from http://www.laurenceanthony.net/

Anthony, L. (2016). AntConc (Version 3.4.4) [Computer software]. Tokyo, Japan: Waseda University. Retrieved March 13, 2017, from http://www.laurenceanthony.net/

Anthony, L. (2017). SegmentAnt (Version 1.1.2) [Computer software]. Tokyo, Japan: Waseda University. Retrieved March 13, 2017, from http://www.laurenceanthony.net/

Baker, M. (1992). *In other words*. Abingdon, UK: Routledge.

Bowker, L., & Pearson, J. (2002). *Working with specialized language: a practical guide to using corpora*. London, UK: Routledge.

Breen, J. (2010). Identification of neologisms in Japanese by corpus analysis. In S. Granger, & M. Paquot (Eds.), *Proceedings of eLex 2009: eLexicography in the 21st century: new challenges, new applications* (pp. 13-22). Louvain, Belgium: Presses Universitaires de Louvain.

European Graduate Placement Scheme. (n.d.). *Occupational standards for European postgraduate translation students on work placement.* Retrieved October 13, 2017, from http://www.e-gps.org/wp-content/uploads/2014/05/Occupational ESF.pdf

Fahey, R. (2016). *Japanese text analysis in Python.* Retrieved August 19, 2017, from http://www.robfahey.co.uk/blog/japanese-text-analysis-in-python/

Goodsell, D.L. (1995). Damping. In *Dictionary of Automotive Engineering*. (p. 57). Warrendale, PA: Society of Automotive Engineers.

Integro Languages. (2017). *4 reasons why glossary creation before translation is so important.* Retrieved October 13, 2017, from http://www.integrolanguages.com/4-reasons-why-glossary-creation-before-translation-is-so-important/

Japan Translation Federation. (2012). *翻訳で失敗しないために翻訳発注の手引き* [lit. For not getting it wrong with translation: a guide to ordering translation]. Retrieved August 1, 2017, from http://www.jtf.jp/pdf/translation_order.pdf

Krishnamurthy, R. (2002, July). The corpus revolution in EFL dictionaries. *Kernerman Dictionary News*. 23-27.

Landau, S. (2001). *Dictionaries: the art and craft of lexicography*. (2nd ed.) Cambridge, UK: Cambridge University Press.

Lionbridge. (2016). *How to create a translation style guide and terminology glossary.* Retrieved October 13, 2017, from http://content.lionbridge.com/how-to-create-a-translation-style-guide-and-terminology-glossary/

Matsuda, H. (2000). ChaSen (2.1) [Computer software]. Nara, Japan: Nara Institute of Science and Technology. Retrieved March 13, 2017, from https://ja.osdn.net/projects/chasen-legacy/releases/27515

Muegge, U. (2013). *10 things you should know about automatic terminology extraction.* Retrieved August 1, 2017, from http://linguagreca.com/blog/2013/09/automatic-terminology-extraction/

Senmon yōgo kīwādo jidō chūshutsu sābisu gensen web [Terminology keyword automatic extraction service gensen web]. (n.d.). Retrieved July 20, 2017, from http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html

Stitt, R. (2016). *The essentials of consistent terminology in academic and professional translation.* Retrieved July 19, 2017, from https://www.ulatus.com/translation-blog/the-essentials-of-consistent-terminology-in-academic-and-professional-translation/

Storjohann, P. (2010). Lexico-semantic relations in theory and practice. In. P. Storjohann (Ed.), *Lexical-semantic*

relations: theoretical and practical perspectives. (pp. 5-13). Amsterdam, The Netherlands: John Benjamins Publishing.

Takeuchi, K., Kageura, K., Koyama, T., Daille, B., & Romary, L. (2003). Pattern based term extraction using ACABIT system. *Language Processing, 10*(4). Retrieved July 18, 2017, from https://arxiv.org/ftp/arxiv/papers/0907/0907.2452.pdf

The Collins Corpus. (2016). Retrieved August 15, 2017, from https://collins.co.uk/page/The+Collins+Corpus?

Winchester, S. (2004). *The meaning of everything: the story of the Oxford English Dictionary*. Oxford, UK: Oxford University Press.

Appendix 1: Glossary

Notes:

1. For confidentiality reasons, this rendering of our glossary shows the name of the automaker as "ABC", the name of the car model as "XYZ", and the names of proprietary body colours as "Colour 1" and "Colour 2".

2. SCDH stands for Stephen Crabbe and David Heath.

か)

緊急制動 [きんきゅうせいどう] ***Grammar*** *noun English* **emergency braking** *Domain automobiles Definition Using a vehicle's brakes to bring the vehicle to a stop as quickly as possible (typically in order to avoid an accident).* ***Source*** *SCDH (July 2017)* ***Context***「4 輪の ABS センサーから緊急制動を検出し、緊急制動信号を発信する。」

減衰 [げんすい] ***Grammar*** *noun English* **damping** *Domain automobiles Definition Dissipation of energy in a vibrating system, usually by mechanical friction or fluid flow through an orifice.* ***Source*** *Dictionary of Automotive Engineering (1995)* ***Context***「サスクロスの振動減衰性をアップ。」 ***Note***「減衰」*is used not only in isolation but also in compounds such as*「振動減衰性」[しんどうげんすいせい] *(typically rendered as "vibration-damping performance") and*「減衰力」[げんすいりょく] *(typically rendered as "damping force").*

さ)

サスクロス ***Grammar*** *noun English* **suspension crossmember** *Domain automobiles Definition A beam that forms a solid link between suspension components on a left-hand wheel and suspension components on the opposite, right-hand wheel.* ***Source*** *SCDH (July 2017)* ***Context***「外力が加わった際のサスクロス位置決め剛性を高めるとともに、サスクロスの内力を増加させ、振動減衰性を向上させています。」 ***Note*** *ABC typically writes "crossmember" as one word in product-information publications for distributors. It is possible that the term is written as two words, i.e., "cross member", in other ABC publications and in publications by other automakers.*

しっかり感 [しっかりかん] ***Grammar*** *noun English* **stability** *Domain automobiles Definition The feeling of steadiness given by a suspension system that adequately isolates the body from external forces.* ***Source*** *SCDH (July 2017)* ***Context***「操舵時のリアのしっかり感を向上させた。」

集中ディスプレイ [しゅうちゅうでぃすぷれい] ***Grammar*** *noun English* **centre display** *Domain automobiles Definition A display that is positioned approximately in the centre of a vehicle's instrument panel (typically separate from the speedometer and any other meter) and shows various types of information (e.g., the current time, the temperature setting of the air conditioner, and the settings of the audio system).* ***Source*** *SCDH (July 2017)* ***Context***「集中ディスプレイでは、時計の時刻調整を簡単に出来るよう、新たに時計調整スイッチを採用しました。」

浄化性能 [じょうかせいのう] ***Grammar*** *noun English* **emission-reduction performance** *Domain automobiles Definition The effectiveness with which a vehicle's exhaust system minimizes emissions of harmful substances.* ***Source*** *SCDH (July 2017)* ***Context***「冷間始動時の早期活性・浄化性能を向上させ、排出ガスのクリーン化を追求しています。」

ステアリングスイッチ ***Grammar*** *noun English* **steering-wheel switch; switch on the steering wheel** *Domain automobiles Definition Any of the switches incorporated into a steering wheel to enable the driver to control vehicle systems (e.g., the audio system) without letting go of the steering wheel.* ***Source*** *SCDH (July 2017)* ***Context***「ステアリングスイッチに、ABC 車として初めてハンズフリートーク操作専用ボタンを採用。」 ***Note*** *Some ABC publications use*「ステアリングスイッチ」*regardless of whether the switch is a rocker switch or a push-button. If the switch is a push-button, "steering-wheel button" or "button on the steering wheel" is a more appropriate rendering.*

設定する [せっていする] ***Grammar*** *verb English* **to make available** *Domain automobiles Definition To make a vehicle feature, e.g., a colour or technology, available with a particular model.* ***Source*** *SCDH (July 2017)* ***Context***「エクステリアカラーは全 10 色を設定しています。」***Note*** *Some ABC publications include this usage of*「設定する」*in addition to the more conventional usage, which typically refers to establishing a setting, e.g., setting a temperature with an air conditioner.*

操縦安定性 [そうじゅうあんていせい] ***Grammar*** *noun English* **handling stability** *Domain automobiles Definition A measure (usually expressed in terms of a cline from worse to better rather than numerically) of how faithfully a vehicle responds to the driver's steering inputs and how stable the vehicle remains when subjected to forces from outside.* ***Source*** *SCDH (July 2017)* ***Context***「フェイスリフト XYZ でも、新しいファミリーフェイスを始めとする新デザインと、操縦安定性に寄与

する CD 値のさらなる低減を両立させるべく、空力性能の開発を行いました。」 ***Note***「操縦安定性」*is sometimes shortened to*「操安性」[そうあんせい]. *The term "handling stability" is the established rendering for ABC product-information publications aimed at distributors. A rendering that better reflects the etymology of the Japanese term and appears to have greater currency is "handling and stability". It may be advisable to ask the source-text author whether s/he has a preference.*

た)

チューニングする ***Grammar*** *verb English* **to tune** *Domain automobiles Definition To adjust the design and/or operating variables of an engine or other vehicle system (e.g., the steering system) to achieve optimal performance. Source SCDH (July 2017)* ***Context***「パワーステアリングのアシスト特性をチューニングし、制御マップを変更。」 ***Note*** *Where the source text does not explicitly state the purpose of the tuning, "optimize" or "enhance" may be a more suitable rendering.*

トップレベル ***Grammar*** *adjective English* **among the best; some of the best** *Domain automobiles Definition An arguably disingenuous description used by ABC for a vehicle attribute (e.g., fuel economy or engine power) that is better than the corresponding attributes of most competing vehicles but is not the best.* ***Source*** *SCDH (July 2017)* ***Context***「超小型タービンの採用により、クラストップレベルのレスポンスを実現。」 ***Note*** *ABC uses the term*「トップレベル」*not only by itself but also in compounds such as*「クラストップレベル」*and*「世界トップレベル」.

トレーリングブッシュ ***Grammar*** *noun English* **trailing-arm bush** *Domain automobiles Definition A bush (a cylindrical sleeve forming a bearing surface for a shaft or pin) in one of the trailing arms of a vehicle's rear suspension.* ***Source*** *Dictionary of Automotive Engineering (1995) and SCDH (July 2017)* ***Context***「タイヤの動きを後上方に逃がしてショックを緩和するため、リアのトレーリングブッシュのすぐりを前傾。」

は)

ピアノブラック ***Grammar*** *noun English* **piano black** *Domain automobiles Definition A smooth, glossy, black finish that looks and feels like the finish on the black keys of a piano.* ***Source*** *SCDH (July 2017)* ***Context***「センターパネル周辺では、現行 XYZ で採用していたシルバーペイントの加飾を廃止し、新たに艶感や厚み、滑らかさのあるピアノブラックの加飾を採用。」

フェイスリフト ***Grammar*** *noun English* **facelift** *Domain automobiles Definition A change (or collection of changes) to a vehicle model mid-way through the model's production run. A facelift is less extensive than a full redesign. It typically consists of aesthetic updates but may also include updates to technologies such as the engine. It enables an automaker to freshen an aging model and thereby maintain customer interest in it until the next full redesign.* ***Source*** *SCDH (July 2017)* ***Context***「今回のフェイスリフトからの新採用色としては、Colour 1、Colour 2 を用意しています。」 ***Note*** *ABC often uses*「フェイスリフト」*as part of a compound noun, e.g.,*「フェイスリフト XYZ」. *In this case, the established English rendering is the adjective "refined", e.g., "the refined XYZ".*

踏み換える[ふみかえる] ***Grammar*** *verb English See Definition. Domain automobiles Definition To release the brake pedal and press the accelerator pedal or vice versa.* ***Source*** *SCDH (July 2017)* ***Context***「ドライバーがブレーキを踏み換え、発進に必要なトルクが発生するまでの間、停車状態を維持。」

フラット感[ふらっとかん] ***Grammar*** *noun English* **smoothness** *Domain automobiles Definition A feeling of levelness given by a vehicle's suspension system.* ***Source*** *SCDH (July 2017)* ***Context***「ストラットマウント特性変更:バネ定数ダウンにより、上下方向の振動を適度に逃がしてフラット感を向上。」 ***Note***「フラット感」*tends to be used to describe smoothness in terms of a ride whereby the body does not tip, roll, or bounce to any extent that could be felt by occupants.*「マイルド感」*is also rendered as "smoothness" but tends to be used to describe smoothness in terms of an absence of vibration and harshness in the ride.*

ブルブル感[ぶるぶるかん] ***Grammar*** *noun English* **shake; judder** *Domain automobiles Definition An unpleasant, juddering sensation resulting from failure of a vehicle's suspension system to adequately damp vibration and/or from flexing of an insufficiently stiff body.* ***Source*** *SCDH (July 2017)* ***Context***「キャビンの小刻みな動きとして感じやすい 4Hz～9Hz の上下振動エネルギーと、ブルブル感や減衰の悪さを感じやすい 10～14Hz の前後振動エネルギーを低減」 ***Note*** *If the source text explicitly states that the*「ブルブル感」*results from flexing of an insufficiently stiff body when the vehicle goes over bumps, the appropriate rendering is "scuttle shake".*

プレミアム感[ぷれみあむかん] ***Grammar*** *noun English* **premium identity** *Domain automobiles Definition A sense of superior quality conveyed by a vehicle or by some feature(s) of a vehicle. Source SCDH (July 2017)* ***Context***「スポーティでありながらアグレッシブになりすぎず、プレミアム感が漂うフロントビューを創り上げました。」 ***Note*** *If*「プレミアム感」*clearly applies to the appearance and/or tactile quality of a physical object, "premium look", "premium feel", or "premium look and feel" may be a more appropriate rendering.*

ま)

マイルド感[マイルドかん] ***Grammar*** *noun English* **smoothness** *Domain automobiles Definition An absence of vibration and harshness in the ride given by a vehicle. Source SCDH (July 2017) Context*「ブッシュのストッパーを廃止して荷重がかかった際のたわみ特性をリニアにし、マイルド感を向上」 ***Note***「フラット感」*is also rendered as "smoothness" but tends to be used to describe smoothness in terms of a ride whereby the*

body does not tip, roll, or bounce to any extent that could be felt by occupants.

ら)

リアコンビランプ ***Grammar*** *noun English* **rear combination lamp** *Domain automobiles Definition A rear lamp unit containing a number of lamps with separate functions, e.g., making the vehicle visible from behind in darkness, showing when the vehicle is turning (or about to turn) a corner, and showing when the driver is pressing the brake pedal.* ***Source*** *SCDH (July 2017)* ***Context*** 「リアビューでは、リアコンビランプのデザインを新しくしています。」

Appendix 2: Sample of frequency list before removal of noise

16	59	より
17	58	する
18	53	現行
19	50	まし
20	46	せ
21	45	が
22	44	な
23	41	感
24	40	性
25	40	採用
26	39	向上
27	36	変更
28	33	R
29	33	図
30	32	操作
31	30	から
32	29	grade

Note:

The first column shows where each term ranks in order of frequency of occurrence in the source text. The second column shows the number of occurrences.

Appendix 3: Sample of usable frequency list

27	10	ステアリングスイッチ
28	10	バネ定数
29	10	従来
30	10	新
31	10	発進
32	10	色
33	9	インチ
34	9	エンジン
35	9	ディーゼルエンジン
36	9	加速度
37	9	走行
38	9	車
39	8	インテリア
40	8	チューニング

Note:

The first column shows where each term ranks in order of frequency of occurrence in the source text. The second column shows the number of occurrences.

Appendix 4: Sample of usable alphabetical list

73	1	アクリルカバー
74	1	アグレッシブ
75	2	アシスト特性
76	2	アスファルト
77	1	アスレティックさ
78	1	アッパーボディ
79	7	アップ
80	7	アルミホイール
81	1	アンダーボディ
82	2	イメージ
83	9	インチ
84	8	インテリア
85	5	インテリアカラー
86	1	インテリアデザイン

Note:

The first column shows where each term ranks in alphabetical order. The second column shows the number of occurrences.