

Developing a Corpus of YouTube How-to Videos: Investigating Spoken Procedural Monologues by Native-Level Speakers of English

YouTube ハウツー動画コーパスの開発：英語母語話者による
手続き的モノログ発話の研究

Mark HAMMOND*

マーク・ハモンド

Abstract

This study investigates procedural monologues (PM) produced by native-level speakers of English. The term PM, as used here, represents texts produced when a single speaker gives how-to instructions while demonstrating a task involving hands-on manipulation of physical objects. Taking a register analysis perspective, the study aims to identify salient communicative functions of PM produced by native-level speakers of English. Towards this goal, the paper describes the development of a specialized corpus of YouTube *how-to* videos (e.g., how to use tools, machinery, culinary equipment, and other hands-on activities) and preliminary analysis of two features found in Biber and Egbert's 2018 multi-dimensional analysis of written how-to/instructional texts found on the searchable World Wide Web. The paper also includes a discussion of a planned future study of original PM data produced by Japanese university undergraduates studying English as a foreign language to be used for comparative analysis.

1. Introduction

Spoken *how-to* instructions are featured in a variety of contexts in our professional and daily lives. This type of discourse may involve simple directions on how to connect a printer to a computer or purchase train tickets from a vending machine, but may also call for more complex instructions, such as those delivered by a master craftsperson on how to use a pottery wheel or by a factory supervisor on the use of specialized machinery and tools. Although there may be some interactive exchanges between speakers and listeners, generally how-to instructions are produced in a monologue delivered face-to-face or by recorded video, such as those found on the YouTube platform. These spoken texts, referred to henceforth as *procedural monologues* (PM), are synchronously produced with a demonstration of the task at hand, involve the manipulation of physical objects, and follow sequential steps to achieve a specific pre-determined goal.

Although a large body of literature exists that is concerned with technical writing of procedural texts and directions, there is little attention given to similar spoken discourse. Addressing this gap, this paper outlines the development and preliminary findings of an originally compiled specialized corpus of PM by native-level speakers of English. As a researcher focused on supporting language learning

* Kanazawa University, Institute of Liberal Arts and Science, Faculty of Foreign Language Studies

needs of practitioners of Japanese craft (伝統工芸), PM are a central focus of my current project (see Hammond, 2019–2022) to investigate the genre of demonstrative artist talks that focus on creative technique of pottery and ceramic art. Adding to my previous analysis of these talks (Hammond, 2020), the study presented here takes a more general view of PM by exploring a corpus that encompasses a wide range of how-to tasks to better understand the fundamental linguistic characteristics of spoken instructions. Such an understanding may support pedagogical approaches relevant not only to my own interests in demonstrations of traditional craft but also in contexts of English for specific purposes (ESP) that include this type of procedural discourse.

Taking a register analysis perspective to better understand the linguistic and functional characteristics of PM, this preliminary study is centered on two linguistic features (personal pronouns and modal verbs) that frequently occur in written how-to/instructional discourse, as shown in recent research by Biber and Egbert (2018). Focused on these features, the purpose of this paper is two-fold. The first is to determine what linguistic constructions associated with personal pronouns and modal verbs are frequently found in PM produced by native-level speakers of English. The second is to interpret how such constructions are used in terms of communicative functions relevant to the situational contexts of PM.

The paper first overviews relevant literature, with emphasis on research (Biber & Egbert, 2018) associated with the reference corpus used in the study. Subsequently, the methods employed in compiling and analyzing the corpus are described, followed by a report of key findings and contextual examples of salient patterns of language use. The paper then addresses implications of results and describes plans for future research, which involve a learner corpus of PM produced by Japanese undergraduates and the development of learning materials focused on procedural discourse embedded in technical demonstrations of the ceramic arts.

2. Literature Review

To a certain extent, elements of spoken procedural discourse have previously been investigated in a range of contexts and fields of research. These include the rhetorical structure of procedures (Eiriksdottir & Catrambone, 2011; Farkas, 1999) and levels of specificity of procedural steps as a tool to evaluate aphasia caused by brain injuries in the field of neurology (Ulatowska et al., 1990; Weinrich et al., 2002). Training in vocational trades such as plumbing, carpentry, and automotive repair have also been the focus of several research projects that touch on aspects of spoken procedural texts. These include studies from New Zealand based researchers (Coxhead et al., 2020; Coxhead & Demecheleer, 2018; Parkinson et al., 2017) who have included spoken instructional exchanges in corpora to explore lexis and discourse features of vocational language, and in corpus-based investigations of analogical discourse in spoken texts collected at vocational schools in Switzerland (Filliettaz et al., 2010). Other studies have centered on specific varieties of how-to videos: modality assessment in computer

software tutorial videos (Morain & Swarts, 2012), and interdiscursive performance in make-up and beauty tutorial videos (Bhatia, 2018).

Nevertheless, there appears to be a lack of studies that specifically focus on PM by taking a register analysis perspective to explore corpora compiled of discourse by a single speaker, such as what could be expected in a YouTube how-to video. However, some relevancy to how-to type discourse can be found in Biber and Egbert's (2018) extensive project focused on corpus-based analysis of register variation found on the searchable World Wide Web. Employing a multi-dimensional analysis approach (see Biber, 1992), the researchers identified 25 sub-registers in the Corpus of Online Registers of English (CORE), which was compiled for the project by categorizing 48,571 web based documents and contains approximately 54 million words. One of these categories is the 1.4 million-word *How-to/Instructional* sub-register, which is concerned with step-by-step instruction to perform a certain task. The study showed that such texts have linguistic characteristics notably unique compared to the other sub-registers. These features include 2nd person pronouns, conditional subordinators, possibility modals, present tense verbs, and *to*-clauses. Although the How-to/Instructional sub-register is based on written text from English-speaking countries, it is not possible to confirm the corpus as native level. However, the general communicative purpose is the same, making it well suited as a reference to investigate similar spoken texts. The pervasive linguistic features in this sub-register of the CORE, particularly pronouns and modals verbs, provide a useful point of departure for preliminary analysis of a specialized corpus of spoken PM compiled for this study, which is describe in the following sections.

3. Method

The study centers on an originally compiled corpus of PM produced by native-level English speakers, tentatively named the Hands-on Procedural Instructions Corpus (HandPIC). In the following section, the development of the corpus is described including criteria for text selection and categories of tasks. Additionally, the study's method of analysis is explained to show how the HandPIC was compared to the How-to/Instructional sub-register of the CORE (Biber & Egbert, 2018), henceforth referred to as HI-CORE.

3.1 Corpus description

The HandPIC is compiled of transcribed spoken texts from 100 videos publicly posted on the YouTube platform. After preliminarily screening parts of approximately 500 videos, selection for inclusion in the corpus was based on the following criteria: a single speaker with native-level proficiency (as determined by the researcher), three to five minutes in length, primarily live-recorded (i.e., without scripted voice-overs or heavy reliance on text titles or post-production elements) and having a clearly stated step-by-step hands-on task. There are 77 videos by male speakers and 23 by female speakers, generally reflecting the observed gender ratio of the how-to video genre on YouTube.

Although the study is not concerned with any aspects of regional dialects, for consistency, only speakers with North American accents were selected, based on my intuition as a native speaker of American English.

Transcription of the spoken text of videos was facilitated by use of the auto-generated text feature embedded in YouTube. All auto-generated text was reviewed for accuracy, corrected, and punctuated to represent a spoken utterance. In total, the 100 texts comprised 48,321 words. In a few cases, text unrelated to the task at hand was deleted from the transcript, such as lengthy self-promotion or appeals for the listener to subscribe to the speaker's YouTube channel.

Regarding the content of videos, 12 thematic categories of tasks, as listed in Table 1, were used as a guideline to avoid over saturation of any one context.

Table 1

Thematic categories and examples of HandPIC texts

Category (number of texts)	Examples (<i>How to...</i>)
Hand/Power Tools (15)	<i>Start a chain saw</i>
Sport Technique (10)	<i>Throw a football</i>
Food Preparation (10)	<i>Cut pineapple</i>
Computer Software (10)	<i>Copy/paste on an iPad</i>
Computer Hardware (5)	<i>Remove a hard drive</i>
Scientific Equipment (5)	<i>Prepare a petri dish</i>
Emergencies (5)	<i>Use a fire extinguisher</i>
Repairing Items (5)	<i>Mend an extension cord</i>
Assembling Items (5)	<i>Assemble a saxophone</i>
Hair/Skin Care (5)	<i>Trim a beard</i>
Nursing Practice (5)	<i>Tape an ankle</i>
Miscellaneous (20)	<i>Load a film camera</i>

Acknowledging that these categories do not necessarily represent the full spectrum of the how-to genre posted on YouTube, the themes cover a wider variety of content observed during the preliminary screening before selection.

3.2 Analysis

Analytical methods primarily centered on comparing the HandPIC to the HI-CORE, which was used not only as a reference corpus to compare frequency, but also as a guide to what linguistic features would be investigated. Two of the most distinctive features from the HI-CORE, personal pronouns, and modal verbs, were selected for the scope of this preliminary study based on the distinct differences in frequency and communicative purpose that were first observed during the transcription process.

The Sketch Engine corpora management platform (Kilgarriff et al., 2014) was used for all quantitative analysis. Although the full version of the CORE (see Davies, 2016), including the HI-CORE sub-register, is publicly available for searches via an online portal, it was more practical and expedient to use the same analytical tools in Sketch Engine for comparative aspects of this study. Via a request to Jesse Egbert (personal communication, August 13th, 2021), I obtained all 1,392 text files (1.4 million words) that were originally used to compile the HI-CORE. With both sets of data loaded into Sketch Engine, frequency of patterns related to pronouns and modal verbs were extracted and normalized at the rate of occurrence per 10,000 words. In addition, relevant examples of text used to make qualitative interpretation of communicative functions were extracted by the Sketch Engine tools related to parts of speech tagging, corpus query language (CQL) searches, and collocation analysis.

4. Results

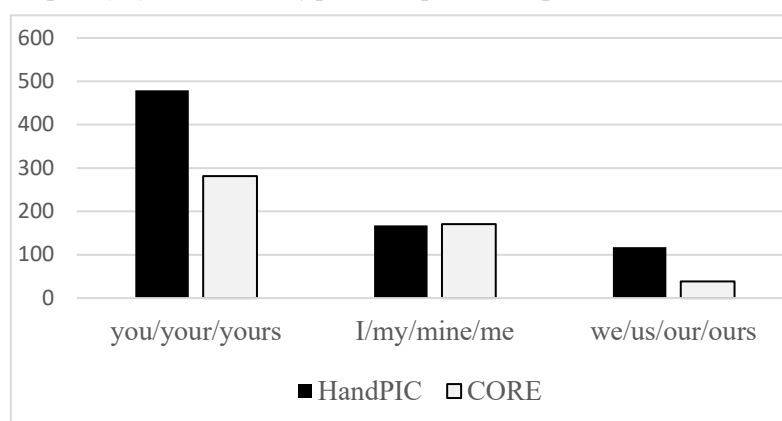
Addressing the questions at the core of the study, results presented here show the quantitative difference in frequency of the patterns associated with pronouns and modal verbs in both corpora as well as an interpretation of communicative function with examples of texts from the HandPIC for reference. All results of frequency are shown at a normalized rate of occurrence per 10,000-words.

4.1 Personal Pronouns

As shown in Figure 1, there is a significant difference in the frequency of personal pronouns, which in total were used much more frequently in the HandPIC (764.68) than in the HI-CORE (490.27).

Figure 1

Frequency of occurrence of personal pronouns (per 10,000 words)



Notably, 2nd person pronouns (479.51) were most common in the HandPIC, compared to the HI-CORE (281.25). While 1st person singular pronouns were close to identical (167.78 in the HandPIC and 170.62 in the HI-CORE), there was a substantial difference in frequency of 1st person plural

pronouns, which occurred more than three times more frequently in the HandPIC (117.39) than in the HI-CORE (38.40).

These differences in frequency of personal pronouns may be interpreted in several ways. Firstly, pronouns are used more indiscriminately in the HandPIC to suit the personal perspective of the speaker. Unlike written texts that may rely more on imperative forms, it is common for speakers in the HandPIC to use a variety of personal pronouns when describing a procedural action. For example, written how-to instructions may tend to use *Cut the string in half*, but in PM it is also feasible to use *You cut the string in half*; *I cut the string in half*; or *We cut the string in half*. Functionally, all these examples seem to serve the same purpose of directing a procedural step, but the variety of pronouns used in the HandPIC may allow the speaker to take a stance that is more inclusive at a personal level than what may be expected from written texts.

A second possible communicative function associated with pronouns is that they are often the subject of utterances that signal a forthcoming action that the speaker feels is worth paying attention to. This takes the form of PRONOUN + am/are + going to, as illustrated in these examples:

And finally, we are just going to put the screws in.

And now I am going to barely open the lid.

Next you are going to pull on the film advance lever.

This construction occurred at a very high frequency (47.80 in the HandPIC, compared to only 1.64 in the HI-CORE) and was often used with some temporal reference such as *first of all*, *next*, *now*, *before*, *after*, or *finally*.

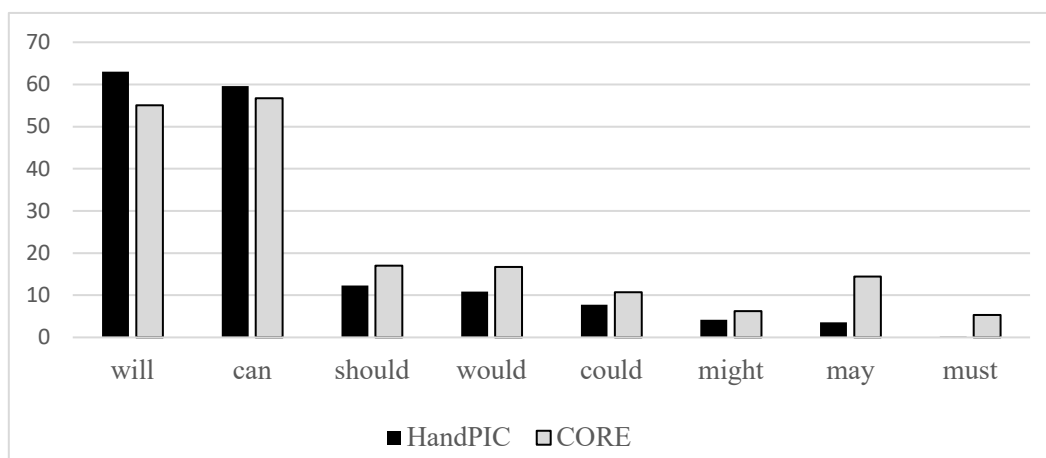
A third functional use of pronouns is associated with the introduction of objects used in the demonstration of a task, which took the form of PRONOUN + *have* + NOUN, as in this example: *We have the strap and then we have this plastic piece*. This pattern was found at a rate of 10.37 in the HandPIC but was extremely infrequent in the HI-CORE at 0.02. Often at the start of a PM, speakers used this construction to clarify and describe physical objects needed to complete the how-to task.

4.2 Modal verbs

Turning to modal verbs, the HandPIC has fewer occurrences of modals than the HI-CORE, as seen in Figure 2. While *might*, *may*, and *must* were too infrequent to make any meaningful interpretation of communicative function, some patterns of language use were observed after investigating collocations of the five most frequently occurring modals: *will*, *can*, *should*, *would*, and *could*.

Figure 2

Frequency of occurrence of modal verbs (per 10,000 words)



One frequently occurring pattern involves the confirmation of expected results of a procedural step, which often include locational reference. This took the form of two constructions. The first is MODAL + *see* (e.g., *At this point you can see we have a really hot fire going on right in the middle*), which occurred significantly higher (9.03) in the HandPIC compared to the HI-CORE (0.17). The second takes the form of MODAL + *be* + LOCATION (e.g., *The fitted end should be to your right*), which occurred at the rate of 3.79 in the HandPIC as opposed to only 0.17 in the HI-CORE.

An additional function of modals in the HandPIC occurred when they were collocated with the verb *use* in utterances that suggested alternative objects, materials, technique, or tools, as seen in the two following examples: *This project mat works great but you can also use cardboard*; *If you want to be an absolute fanatic, you can use a torque wrench*. This occurred at a rate of 4.15 in the HandPIC and 2.06 in the HI-CORE. Such utterances may convey that the listener has some flexibility to adjust or modify the speaker’s instructions to best suit their own approach or available resources.

Finally, the modal *can* was used with *go ahead and* + VERB (1.98 in the HandPIC, 0.01 in the HI-CORE) to signal a progression to the next procedural step, as seen in the example: *At this point we can go ahead and start to open up the back panels on your laptop computer*. This pattern was also used with demonstrative pronouns (*this/that*) such as *So we can go ahead and do that for our entire pineapple*, to indicate that the next action is a repeat of a previously explained step.

5. Discussion

It is important to acknowledge that the study is a preliminary investigation of a developing corpus and is limited in scope. Nevertheless, results show significant difference between spoken discourse found in the how-to genre of YouTube videos and that of written procedural documents available on the World Wide Web. The following discussion addresses implications of these differences and describes plans to use findings in future research.

One unexpected finding of the study was how indiscriminately pronouns seemed to be used by speakers in the HandPIC texts. There was not always a clearly apparent default choice of pronoun, as seen in these examples:

So you just keep wrapping it around and hold the tip in your hand.

What we need to do first of all is we need to wash our hands.

So then I click my home button.

In the HI-CORE, Biber and Egbert (2018) point out that the primary focus is on the reader of the text, who is referred to directly as *you*. Accordingly, 2nd person pronouns are the most significantly co-occurring linguistic feature in the register of written how-to/instructional texts. This is also the case in the HandPIC, but a notable difference is that the 1st personal plural (*we*) occurs more than three times as frequently as in the HI-CORE. This may imply that the speaker envisions the listener as being engaged in a parallel activity in real time. Thus, the primary focus may not always be the listener (*you*) but can also imply both the speaker and the listener (*we*). Moreover, the 1st person singular *I* seems to be used in some utterances in the HandPIC in the same way that *you* is in the HI-CORE. In other words, in the above examples, any of these three pronouns could be interchangeably used by the speaker. In some cases, there was a blurring of 1st and 2nd person perspectives even within the same utterance, such as: *We want to take a slap shot only when you have the time.*

Results also suggest that this sense of the listener's parallel engagement is implied in the way that speakers reference the progress of procedural steps. This is seen in the PRONOUN + *am/are* + *going to* construction, which, via the present continuous, reflects the concept that the listener is engaged and being guided through the steps in real time. The same may be true regarding the use of modals, as the noted constructions suggest the importance of completing a procedural step properly (MODAL + *see*) and then moving on to the next step (MODAL + *go ahead and* + VERB).

As previously stated, the underlying goal of exploring PM by native speakers in this study is to establish a baseline for a future comparison of similar texts produced by Japanese undergraduates studying English as a foreign language. Towards this goal, two additional future studies are planned. Currently in early analysis, the first will compare the HandPIC to an originally compiled corpus of 50 similarly themed how-videos produced by Japanese university students. By taking a discrepancy view to compare the two sets of data, this new investigation aims to identify what communicative functions Japanese students typically do not include in PM. Once these deficiencies are better understood, a second study aims to explore how they may be addressed on a more practical level in the EFL classroom by reporting on prototype learning materials focused on producing communicatively effective PM for a particular ESP context, with examples from my own focus on PM embedded in demonstrative artist talk genre in the field of ceramic arts.

6. Conclusion

The study supports a conclusion that several linguistic constructions associated with personal pronouns and modal verbs are frequently found in PM produced by native-level speakers of English. Pronouns were often found in two patterns: PRONOUN + *am/are* + *going to* + VERB, and PRONOUN + *have* + NOUN; modal verbs occurred in three patterns: MODAL + *see*, MODAL + *be* + LOCATION, and MODAL + *use*. Additionally, in terms of communicative functions unique to the situational contexts of PM, the study shows that pronouns occur often when signaling a forthcoming action and introducing objects to be used in the task. Similarly, modal verb usage was associated with both the confirmation of properly executed procedural steps and as a signal to move on to a new step, as well as in utterances that suggested alternative objects or technique. While the study is preliminary and exploratory in nature, it may support future comparative analysis of spoken procedural discourse by non-native speakers of English.

References

- Bhatia, A. (2018). Interdiscursive performance in digital professions: The case of YouTube tutorials. *Journal of Pragmatics*, 124, 106–120. <https://doi.org/10.1016/j.pragma.2017.11.001>
- Biber, D. (1992). On the complexity of discourse complexity: A multidimensional analysis. *Discourse Processes*, 15(2), 133–163. <https://doi.org/10.1080/01638539209544806>
- Biber, D., & Egbert, J. (2018). *Register Variation Online*. Cambridge University Press. <https://doi.org/10.1017/9781316388228>
- Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes*, 51, 84–97. <https://doi.org/10.1016/j.esp.2018.03.006>
- Coxhead, A., Parkinson, J., Mackay, J., & McLaughlin, E. (2020). *English for vocational purposes: Language use in trades education*. Routledge. <https://doi.org/10.4324/9780429449932>
- Davies, M. (2016). *Corpus of Online Registers of English (CORE)*. Available online at <https://www.english-corpora.org/core/>
- Eiriksdottir, E., & Catrambone, R. (2011). Procedural instructions, principles, and examples: how to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human Factors*, 53(6), 749–770. <https://doi.org/10.1177/0018720811419154>
- Farkas, D. K. (1999). The logical and rhetorical construction of procedural discourse. *Technical Communication*, 46, 42–52. <https://eric.ed.gov/?id=EJ581382>
- Filliettaz, L., de Saint-Georges, I., & Duc, B. (2010). Skiing, cheese fondue and Swiss watches: Analogical discourse in vocational training interactions. *Vocations and Learning*, 3(2), 117–140. <https://doi.org/10.1007/s12186-010-9035-4>
- Hammond, M. (Principal Investigator). (2019–2022). *Demonstrative Artist-talks in English as a Foreign Language: A Needs Analysis for Contexts in Japanese Higher Education of the Ceramic Arts*. (Project No. 19K00758) [Grant]. Japan Society for the Promotion of Science, Grant-in-

- Aid for Scientific Research (C). <https://kaken.nii.ac.jp/en/grant/KAKENHI-PROJECT-19K00758/>
- Hammond, M. (2020). ESP for ceramic artists: Exploring the demonstrative artist talk genre. In P. Clements, A. Krause, & R. Gentry (Eds.), *Teacher efficacy, learner agency*. Tokyo: JALT. <https://doi.org/10.37546/JALTPCP2019-28>
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1(1), 7–36. <https://doi.org/10.1007/s40607-014-0009-9>
- Morain, M., & Swarts, J. (2012). YouTutorial: A framework for assessing instructional online video. *Technical Communication Quarterly*, 21(1), 6–24. <https://doi.org/10.1080/10572252.2012.626690>
- Parkinson, J., Coxhead, A., Demecheleer, M., Mackay, J., Matautia, L., McLaughlin, E., & Tu'amoheloa, F. (2017). *The language in the trades education project*. Ako Aotearoa. <https://hdl.voced.edu.au/10707/457589>
- Ulatowska, H. K., Allard, L., & Chapman, S. B. (1990). Narrative and procedural discourse in aphasia. In Y. Joannette & H. H. Brownell (Eds.), *Discourse ability and brain damage*. (1st ed., pp. 180–198). Springer. https://doi.org/10.1007/978-1-4612-3262-9_8