6 Metalanguages for source document analysis

Properties and elements[1]

Rei Miyata and Takuya Miyauchi

6.1 Introduction

A detailed analysis of a source document (SD) forms crucial sources for decision making in the core translation process. It contributes to both an improved quality of translation and a better explanation of the translation process. Although many attempts have been made to verbalise what is/should be understood about an SD for the subsequent translation process, they are not necessarily comprehensive nor well-organised. Terminologies, or metalanguages, for SD analysis that have been used in the literature are often mutually inconsistent, which may hinder their practical application as well as smooth communication among various players involved in the translation project.

To solve this problem, we have endeavoured to develop wide-ranging systematic metalanguages for the SD analysis process by examining the literature on translation studies with reference to related fields such as linguistics. We decompose the SD analysis process into two processes: one to specify the properties of an SD, and another to identify textual elements within the SD. Therefore, in this chapter, we provide two organised lists of terms as metalanguages: *document properties* and *document elements*.[2]

From a practical point of view, the way of using metalanguages is also crucially important. For example, the Multidimensional Quality Metrics (MQM) scheme (Burchardt & Lommel, 2014; Lommel et al., 2014), which is designed to assess translation quality, provides not only the detailed typology of translation issue types, but also clear, concise definitions and examples for each term. This helps users apply the scheme to their particular use cases. Annotation guidelines in the form of a decision tree have also been developed to facilitate the effective and consistent use of MQM (Burchardt & Lommel, 2014). Although the full development of the user guides for SD analysis is beyond the scope of this chapter, the ways of effective application of the formulated metalanguages will be discussed.

The remainder of the chapter is structured as follows. Section 6.2 explains the role and definition of SD analysis process and its components. Section 6.3 describes the scope and procedures of the review-based approach for metalanguage development. Sections 6.4 and 6.5 present the resulting systematic metalanguages of document properties and elements, respectively. Section 6.6

further discusses the effective application of the formulated metalanguages, and Section 6.7 presents future avenues for metalanguage development and application.

6.2 Definition and component

The SD analysis process is related to both the pre-production and production phases of translation projects. ISO 17100 (ISO, 2015, p. 9) defines source language content analysis as one of the pre-production processes, stating that the translation service provider "shall ensure that the source language content is analyzed to ensure efficient and effective performance of the translation project." ISO/TS 11669 specifically provides a list of source-content information as translation parameters which "should be used to develop preliminary project specifications during the pre-production phase" (ISO, 2012, p. 18). In this view, the focus is on an SD itself or a whole text within the SD. In contrast, Gile (2009) focuses on the production phase and models translation as a succession of two phases: a comprehension phase and a reformulation phase. In this model, what the translator directly processes is a *translation unit*, or text segment, that varies in length from a single word to multiple sentences. Here, the focus is more on the textual elements within an SD.

The above observation led us to separate properties of a document itself and elements within a document in the SD analysis process. This dichotomy is theoretically significant because we view a document as a unique and primary unit to be handled. Identification of any properties of the document and elements within the documents is premised on the proper objectification of the document. This view can be called a "documentational" approach. Formally, a document property is a pair of {**property name**: value(s)}, such as {**sender**: Government of Western Australia, Department of Mines, Industry Regulation and Safety}, {**medium**: web page}, and {**function**: informative}. Similarly, a document element can be formulated as a pair of {a text span: **element name**}, such as {"Building practitioner registration": **document title**} and {"Australian Institute of Building": **named entity**}.

Figure 6.1 illustrates the detailed role and subprocesses of the SD analysis process in relation to the core transfer process of translation, which can be summarised as follows:

- SD properties are specified to create an SD profile, i.e. a list of specified SD properties predefined for each project.
- Based on the translation brief, the SD profile is converted to a target document (TD) profile.
- SD elements within the SD are identified.
- Based on the SD/TD profiles and translation brief, SD elements are transferred into TD elements, which finally constitute a TD.
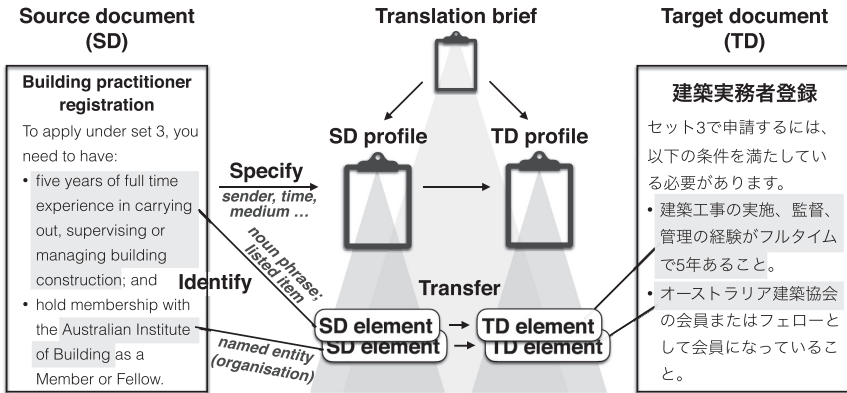
Figure 6.1 Source document analysis in the production process.[3]

In the SD property specification task, the object to be profiled, i.e. an SD, is given. In contrast, in the SD element identification task, target text spans are not given. One needs to identify a text span and element name at the same time. Even for the same text string, the way elements are identified may vary by analyser.

To effectively conduct SD analysis tasks, comprehensive, well-organised lists of property and element names are needed in advance. Therefore, in this chapter, we provide them as a core part of metalanguages for the SD analysis process.

6.3 Review-based procedure for metalanguage development

As described in Chapter 2, a metalanguage can be developed through review-based, data-driven, and/or user-focused procedures. In this chapter, we adopted the review-based procedure; we collected existing terms relating to SD analysis from the literature and systematised them as metalanguages.

To widely collect related terms, the following eight books and documents on translation were selected as the target literature: House (1977, 1997),[4] ISO (2012), Newmark (1988), Nord (2005), Reiss (1971/2000), Reiss and Vermeer (1991/2013), and Snell-Hornby (1995). These works extensively cover established textbooks on translation studies and the international standard for commercial translation services, i.e. ISO (2012).

To develop a metalanguage of document properties, we first comprehensively extracted noun phrases that refer to document properties, such as "addressee" (House, 1997; Nord, 2005), "audience" (ISO, 2012), "readership" (Newmark, 1988), "recipient" (Reiss & Vermeer, 1991/2013), and "chance receiver" (Nord, 2005). We also extracted property values for each property, such as "people listening to a panel discussion" for "chance receiver" (Nord, 2005). These values are, if provided, useful for understanding the concept and scope of a

property. We then examined the collected terms to control variations and organised them in a bottom-up manner; for example, "receiver" includes the terms "addressee" and "chance receiver" as sub-categories. Through these procedures, a hierarchical typology of property names was finally constructed.

Similarly, to develop a metalanguage of document elements, we extracted terms from the literature and hierarchically organised them. For example, the document element "idiom" (Reiss & Vermeer 1991/2013) can be categorised under the "lexis" category that is further included in the top-level category of "linguistic element." In contrast to document properties, the target literature of translation studies is not sufficient to widely cover document elements. Hence, we also referred to the literature on related fields, including linguistics (Ando, 2005), rhetoric (Abott, 1996; Peacham, 1577; Sato et al., 2006), and technical writing (University of Chicago, 2017), to add and refine elements, which improved the intrinsic values of the metalanguage, i.e. systematicity, coverage, and granularity (see Section 2.5 for details of requirements for metalanguages).

6.4 Metalanguage of document properties

A total of 57 properties were finally formulated under four major categories, namely, knowledge, communication, formation, and text properties. We present hierarchical lists of the property names with instances of values extracted from the literature.

Table 6.1 presents the knowledge properties that indicate the status of a document vis-à-vis the knowledge accumulated in society. We distinguish **(K01) subject field** and **(K02) topic** in terms of degree of abstraction; the former refers to more abstract categories of subjects. The **(K3) genre** refers to "conventional forms of texts associated with particular types of social occasion" (Hatim & Mason, 1997, p. 218). Examples of genre include patent, user manual, recipe, and weather report. Although these text classes are also called "text types" (ISO, 2012), we avoided using this term as it has a special usage in functional translation theory (Reiss & Vermeer, 1991/2013). The focus of the **(K04) difficulty** property is not on the expressions or linguistic forms but on the *content* that an SD conveys. The **(K05) background knowledge** property concerns the knowledge required to be known to properly comprehend the SD. The **(K06) resource** indicates external concrete materials such as original documents and terminologies that are related to the SD.

Table 6.2 shows the communication properties that capture the communicative situation surrounding an SD, which have been widely covered in the literature of functionalist approaches, such as Nord (2005) and Reiss and Vermeer (1991/2013). The **(C01) sending** and **(C02) receiving** properties are symmetrical, covering the basic communication factors, i.e. "who/whom" (sender/receiver), "when" (sending/receiving time), and "where" (sending/receiving place). The **(C03) sender–receiver relationship** captures the roles of the sender and receiver connected via an SD. The **(C04) communication field** indicates the domain in which an SD is communicated. The **(C05) function** and

Table 6.1 Knowledge properties.

Property name	Extracted values (options/examples)
(K01) subject field	chemical engineering, civil engineering, asphalt, street maintenance and repair, economics, poverty, family psychology, personal finance (ISO, 2012); scientific, technological, commercial, industrial, economic; literary / institutional / scientific (Newmark, 1988)
(K02) topic	a story about a bear family (House, 1997); The Age of Enlightenment (Nord, 2005)
(K03) genre	utility patent, persuasive brochure, appliance user manual, annual address to stakeholders of a public company (ISO, 2012); advertisement, summary, recipe, novel, sermon, wedding announcement, obituary, weather report; implementing rules, summaries, reviews, parodies, travesties (Reiss & Vermeer, 1991/2013)
(K04) difficulty	simple / popular / neutral (using basic vocabulary only) / educated / technical / opaquely technical (comprehensible only to an expert) (Newmark, 1988)
(K05) background knowledge	
(a) academic discipline	Cultural History, Literary Studies, Sociocultural and Area Studies, Studies of Special Subjects (Snell-Hornby, 1995)
(b) presupposition	the knowledge on the part of the receiver that this [Twelfth Night or What You Will] is the title of a play; the characters are socially classified by their names (Nord, 2005); reference to Goethe's play Faust, Part I, line 421; usual rivalry and even enmity between freshmen and sophomores at American colleges (Reiss & Vermeer, 1991/2013)
(K06) resource	
(a) origin	the French source content had originally been translated from English (ISO, 2012)
(b) terminology	

(C06) purpose of documents are closely and mutually dependent; for example, "to show that patients must have a thorough physical check-up before they start a course of drugs" (Newmark, 1988, p. 12), which is an instance of purpose, can be interpreted as informative and/or persuasive functions. The **(C07) background situation** property broadly covers how and why an SD is produced.

   Table 6.3 presents the formation properties which capture the ways of packaging content into a concrete document form, but excluding linguistic aspects. The **(F01) communication medium** and **(F02) symbol type** are basic properties to indicate the forms of information conveyance. The properties in **(F03) file**, i.e. **(a) volume**, **(b) format**, **(c) markup**, and **(d) editability**, are particularly

Table 6.2 Communication properties.

Property name	Extracted values (options/examples)
(C01) sending	
(a) sender	naturalist author, author of the Romantic period (Reiss, 1971/2000)
(i) responsible sender	company selling the product; legislative body of a state; expert; journalist (Nord, 2005)
(ii) author	
(b) sending time	18th-century, 1949, contemporary, old, modern (Reiss, 1971/2000)
(c) sending place	Syria, Jerusalem (Nord, 2005)
(C02) receiving	
(a) receiver	
(i) addressee	teenage readers, adult; Spanish, German, French; his wife, girl friend (Nord, 2005); German reader, Spaniard (Reiss, 1971/2000)
(ii) chance receiver	people listening to a panel discussion, watching a televised parliamentary debate, potential voters (Nord, 2005)
(b) receiving time	the reader who receives the text the next morning (Nord, 2005)
(c) receiving place	Germany, Austria, Switzerland (Nord, 2005)
(C03) sender–receiver relationship	addresser has de facto economic authority over the addressees (House, 1977); adults and children (House, 1997)
(C04) communication field	scholarly, philosophical, religious, aesthetic or everyday communication (Reiss & Vermeer, 1991/2013)
(C05) function	expressive / informative / vocative; aesthetic / phatic / metalingual (Newmark, 1988); referential (denotative, cognitive) / expressive (emotive) / operative (appellative, conative, persuasive, vocative) / phatic (Nord, 2005); representation / expression / pursuation (Reiss, 1971/2000); informative / expressive / operative / multimedia (Reiss & Vermeer, 1991/2013)
(C06) purpose	identify all uses of product to be patented; allow scheduled maintenance; entertainment (ISO, 2012); to show that patients must have a thorough physical check-up before they start a course of drugs (Newmark, 1988)
(C07) background situation	because he or she has fallen in love; because it is Grandfather's 70th birthday (Nord, 2005)

Table 6.3 Formation properties.

Property name	Extracted values (options/examples)
(F01) communication medium	telephones, microphones, newspaper, magazine, book, multi-volume encyclopedia, leaflet, brochure (Nord, 2005)
(F02) symbol type	text, images, audio / video recordings (ISO, 2012); visual / verbal; written / oral; text in Morse code, musical scores (Reiss & Vermeer, 1991/2013)
(F03) file	
(a) volume	characters, words, lines, pages (ISO, 2012)
(b) format	standard documents, slide presentations, databases and entire websites (ISO, 2012)
(c) markup	XML, HTML, plain text (ISO, 2012)
(d) editability	graphic is text-editable; a separate text file that accompanies the graphic (ISO, 2012)
(F04) structure	
(a) document structure	chapters, headings, sub-headings, paragraph lengths (Newmark, 1988); chapters in novels, division into sections in contracts or paragraphs in laws (Reiss & Vermeer, 1991/2013)
(b) content structure	a thesis, an antithesis, a synthesis; an introduction, an entry into the subject, aspects and examples, a conclusion; a setting, a complication, a resolution, an evaluation; a definition of the argument of the title, the pros and cons, the conclusion; a build-up, a climax, a denouement; a retrospect, an exposition, a prospect (Newmark, 1988); weather reports with their conventional sequence of general weather conditions / short-term forecast / long-term forecast (Reiss & Vermeer, 1991/2013)

important from the viewpoint of translation management and only mentioned in ISO (2012). The **(F04) structure** property refers to the internal composition of a document. In contrast to the **(a) document structure**, the **(b) content structure** may be more genre-specific; for example, weather reports have "conventional sequence of general weather conditions/short-term forecast/long-term forecast" (Reiss & Vermeer, 1991/2013, pp. 165–166).

Finally, Table 6.4 shows the text properties, which mainly pertain to linguistic aspects. Whilst some of the text properties are applicable to text spans within the document, i.e. document elements, the focus here is the document-wide characteristics of the whole text. The **(T01) language** property is an indispensable parameter when embarking on any translation projects. **(T02) register** refers to "a variety associated with a particular situation of use (including particular communication purposes)" (Biber & Conrad, 2009, p. 6). In the course of the literature review, we included **(a) mode** and **(b) formality scale** in the register

Table 6.4 Text properties.

Property name	Extracted values (options/examples)
(T01) language	Portuguese, English; Brazilian Portuguese, UK English (ISO, 2012)
(T02) register	
(a) mode	spoken / written, simple / complex (House, 1977, 1997)
(b) formality scale	officialese / official / formal / neutral / informal / colloquial / slang / taboo (Newmark, 1988)
(T03) dialect	
(a) geolect	non-marked, standard American English (House, 1977)
(b) chronolect	21st century lexis; archaic lexis (Nord, 2005); contemporary (Reiss, 1971/2000)
(c) sociolect	non-marked, educated middle class (House, 1977); higher / pedestrian (Reiss, 1971/2000)
(T04) style	
(a) stance	good, fair, average, competent, adequate, satisfactory, middling, poor, excellent; positively / neutrally / negatively (Newmark, 1988)
(b) emotional tone	intense (profuse use of intensifers) ('hot') / warm / factual ('cool') / understatement ('cold') (Newmark, 1988)
(c) literariness	
(d) peculiarity / creativity	author's creative expressions (Reiss, 1971/2000)
(T05) quality	
(a) cohesion	
(b) coherence	
(c) readability	
(d) speakability	
(e) degree of error	
(T06) representation pattern	monologue, dialogue (House, 1977); narrative / description / discussion / dialogue (Newmark, 1988)

category. **(T03) dialect** indicates linguistic variations associated with particular groups of language users (Biber & Conrad, 2009, p. 11). In our documentational approach, the dialect property is associated with the text written in a document, and not the sender of the document. The term **(T04) style** is inconsistently used in translation studies and practices. Here, based on the style perspective presented by Biber and Conrad (2009, p. 18), we define styles as text varieties attributable to the author's preferences or peculiarities. **(T05) quality** refers to the linguistic quality of a whole text. The quality of an SD is important in assessing the difficulty of translation. The detailed attributes of quality have not been sufficiently mentioned in the literature, except for ISO (2012), which presents several aspects, such as cohesion, coherence, and readability. **(T06) representation pattern** indicates ways of representing communicative acts as a text, such as monologue and

dialogue. As the focus of this property is not on the document-external communication actors but on the textual representation, it is included in the text properties.

6.5 Metalanguage of document elements

Through the review procedure, we finally formulated 342 terms regarding document elements in a five-level hierarchy, among which 258 terms are terminal elements. Tables 6.5a–c present lists of organised elements, covering the following nine top-level categories: **(DS) document structure element**, **(LO) locale**, **(TT) technical term**, **(NE) named entity**, **(TR) text referential element**, **(LI) linguistic element**, **(RH) rhetorical element**, **(FO) font element**, and **(OR) orthographic element**. Owing to space constraints, we show the top three levels of hierarchy, marking elements that have further sub-categories with an asterisk (*****).

The **(DS) document structure element** in Table 6.5a is related to the role of a text span within a document. The proper recognition of document structure elements is important because the same text string may be differently translated depending on their roles or positions in a document (Miyata et al., 2016). While the sub-categories (a)–(g) cover elements widely observed in documents of various domains, such as **(b-04) section title** and **(d-01) footnote**, the **(h) content element**, which is an indicative abstraction of given content, is genre-specific; sets of content elements have been identified or defined for particular genres, such as IMRaD (Introduction, Method, Results, and Discussion) for scientific papers and DITA (Darwin Information Typing Architecture)[5] for technical documents. DITA defines basic content elements that are used to constitute a document, such as "prerequisite," "steps," and "result" for composing a procedural document.

(LO) locale covers language- or region-specific formats and expressions. We referred to the locale category defined in the MQM framework. For example, a temperature is expressed differently depending on the region, using Fahrenheit or Celsius (**(f-03) measure**). These locale elements are particularly important in localisation projects.

(TT) technical term indicates "lexical units used in a more or less specialised way in a domain" (Kageura, 2012, p. 9). To categorise terms, we need to specify the domain, such as *medical* and *legal* domains. As a starting point, referring to the categories of specialised translation discussed by Gouadec (2007), we defined eight domains, namely, **(a) industry**, **(b) science**, **(c) information technology**, **(d) medicine**, **(e) law**, **(f) marketing**, **(g) economy**, and **(h) finance**.

The **(NE) named entity**, or proper noun, refers to the object that has a unique name, such as "Tokyo" (**(c) location**) and "Google Translate" (**(e) product**). To expand the list of named entity types, we referred to Sekine's Extended Named Entity Hierarchy,[6] which consists of fine-grained categories in three-level hierarchy, such as "Product" → "Vehicle" → "Car." We show here the top-level categories adapted from the Sekine's taxonomy. Note that the same text string may refer to different entities; for example, the word "Watson" has many possible

Table 6.5a Document elements (1/3).

(DS) document structure element	**(TT) technical term**
(a) hierarchical unit	(a) industry
(01) part	(b) science
(02) chapter	(c) information technology
(03) section	(d) medicine
(04) subsection	(e) law
(05) subsubsection	(f) marketing
(06) paragraph	(g) economy
(b) title/heading	(h) finance
(01) document title	**(NE) named entity**
(02) part title	(a) person
(03) chapter title	(b) organisation
(04) section title	(c) location
(05) subsection title	(d) facility
(06) subsubsection title	(e) product
(c) itemisation	(f) event
(01) ordered itemisation	(g) postal code
(02) unordered itemisation	(h) telephone number
(d) note	(i) url
(01) footnote	(j) artifact
(02) endnote	**(TR) text referential element**
(e) caption	(a) special string
(f) text in figure	(01) mathematical formula
(g) text in table	(02) code
(h) content element	(03) tag
(01) IMRaD (scientific paper)*	(04) foreign language string
(02) DITA (technical document)*	(05) transliteration
(LO) locale	(06) reference notation
(a) date	(b) example
(b) time	(c) annotation
(c) day	(d) proverb
(d) counters	(e) quotation
(e) number	(01) remark
(f) unit	(02) external quotation
(01) money	(03) internal quotation
(02) percent	
(03) measure	
(g) shortcut key	

referents, such as a fictional character Dr. Watson in the Sherlock Holmes stories (**(a) person**) and a question-answering system developed by IBM (**(e) product**).

The focus of the **(TR) text referential element** is not the linguistic interpretation of its meaning but the text itself. We defined five sub-categories: **(a) special string** includes elements that should be decoded by using other interpretation systems than the source language, such as mathematics (**(01) mathematical formula**) and programming languages (**(02) code**); **(b) example** is often used in the literature on linguistics, such as "Joan *is singing* well" (Quirk et al., 1985,

Table 6.5b Document elements (2/3).

(LI) linguistic element
(a) phoneme/syllable
    (01) segment*
    (02) suprasegmental features*
(b) word/morpheme
    (01) part of speech*
    (02) grammatical category*
    (03) word formation*
    (04) variation*
(c) phrase
    (01) phrase type*
    (02) complement structure*
(d) clause
    (01) functional clause type*
    (02) structural clause type*
    (03) clause pattern*
    (04) clause element*
(e) sentence
    (01) functional sentence type*
    (02) structural sentence type*

(f) discourse
    (01) discourse relation*
    (02) referential expression*
    (03) deixis*
    (04) information structure*
    (05) functional sentence* perspective
    (06) speech act*
(g) lexis
    (01) neologism
    (02) idiom
    (03) word type*
(h) mode
    (01) written language
    (02) spoken language
(i) speech style
    (01) casual speech style
    (02) honorific speech style
(j) dialect
    (01) geolect
    (02) chronolect
    (03) sociolect
    (04) idiolect

---

p. 197), which is used to show an example of the present progressive aspect; **(c) annotation** is used to meta-linguistically explain or label the text; **(d) proverb** and **(e) quotation** are the direct references to a text or utterance already produced and documented.

   (LI) linguistic element in Table 6.5b encompasses a wide variety of linguistic levels from **(a) phoneme/syllable** and **(b) word/morpheme** to **(e) sentence** and **(f) discourse**. To date, extensive and detailed metalanguages have been devised in the field of linguistics, whose core mission is to describe and explain languages themselves. When developing the current version of the linguistic element list, we mainly referred to the middle-size English grammar book by Ando (2005). Most of the elements in the table have further sub-categories that are not presented owing to space limitations; for example, **(b-01) part of speech** includes noun, verb, adjective, adverb, preposition, pronoun, conjunction, interjection, numeral, and determiner. From a practical point of view, whilst further expanding and refining elements, we need to examine the importance of the elements with the translation process in mind in order to select a reasonable number of elements to be handled. It is also notable that some linguistic elements are language-dependent and need to be adjusted to a language to be handled.

   (RH) rhetorical element in Table 6.5c is important for grasping the rhetorical effect of an expression. The rhetorical devices can be broadly categorised into two types: **(a) scheme** and **(b) trope** (Peacham, 1577). Whereas schemes involve "deviations in the patterns and arrangements of words," tropes involve

Table 6.5c Document elements (3/3).

(RH) rhetorical element	(OR) orthographic element
(a) scheme	(a) punctuation
(01) interposition	(01) period
(02) ellipsis	(02) comma
(03) hyperbaton	(03) semicolon
(04) palillogy	(04) colon
(05) parallelism	(05) question mark
(06) chiasmus	(06) exclamation point
(07) tautophony	(07) hyphen
(08) hypallage	(08) dash
(09) enallage	(09) bracket
(10) onomatopoeia	(10) slash
(11) paronomasia	(11) quotation mark
(12) antanaclasis	(12) apostrophe
(13) syllepse	(13) space
(b) trope	(14) symbol
(01) simile	(b) letter case
(02) metaphor	(01) upper case
(03) metonymy	(02) lower case
(04) synecdoche	(c) capitalisation style
	(01) sentence case
(FO) font element	(02) title case
(a) typeface	
(01) sans serif/Gothic	(03) all caps
(02) serif/Roman	(04) small caps
(03) Ming/Song	(05) all lowercase
(b) visual style	(d) character type
(01) italic/oblique	(01) alphabet
(02) bold	(02) CJKV
(03) underline	(e) typographical error/typo
(04) colour	(01) misspell
(05) size	(02) haplography
	(03) dittography
	(04) metathesis

"deviations in the meaning of words" (Abott, 1996, p. 597). Referring to the encyclopaedia of rhetoric by Sato et al. (2006), we listed 14 schemes and four tropes. These rhetorical devices may be more frequently observed in the type of documents whose communicative function is expressive or vocative, such as literature and advertisement.

(FO) font element pertains to the appearance of text. Although various font elements can be identified (e.g. Bringhurst, 2019), we selected major ones that are assumed to be referred to in the processes of writing and translation in general. For example, to emphasise a certain text span in an SD, a writer may use a **(b-02) bold** font or a different **(b-04) colour**, such as red. Such deliberate differentiation of text appearance should be recognised and considered in decision making in translation. The use of **(b-01) italic/oblique** font is used to indicate

book and movie titles, which is important information for identifying the **(NE) named entity**.

(OR) orthographic element covers conventions and norms of writing texts in a language. It is, by nature, language-dependent; in this chapter, we only provide English-based orthographic elements, whose detailed usage has been well-established in document editing practices. To investigate the range of the **(a) punctuation**, we particularly referred to *The Chicago manual of style* (University of Chicago, 2017), which is one of the most highly reputed guidelines of English editing.

6.6 Toward effective use of the metalanguages

To make effective use of the metalanguages, in this section, the following two approaches will be discussed: (1) the metalanguage scheme for supporting users and (2) proper management of the SD analysis process.

6.6.1 Metalanguage scheme for supporting users

To enhance the consistent and effective use of metalanguages, it will first be useful to provide guidelines. Nord (2005, p. 19), for example, presents detailed procedures for analysing an ST, offering detailed definitions of each aspect of an ST and a wide variety of examples. Guides on information acquisition and neatly formulated checklists are also provided. Below is an example of a guide for obtaining information about the addressed audience, which corresponds to the **(C02-a-i) addressee** property in Table 6.2:

> As in case of the sender, information about the addressees can first of all be inferred from the text environment (e.g. dedications, notes), including the title (e.g. *Bad Child s* [sic] *Pop-Up Book of Beasts*. It can also be elicited from the information obtained about the sender and his/her intention or from the situational factors, such as medium (cf. example 3.1.3./2)), place, time, and motive (cf. example 3.1.3./3). Standardized genres often raise equally standardized expectations in the receivers. (Nord, 2005, p. 61)

The following checklist is also presented to help readers obtain information relevant to the addressed audience and its expectations (Nord, 2005, p. 62):

1. What information about the addressed audience can be inferred from the text environment?
2. What can be learned about the addressees from the available information about the sender and his/her intention?
3. What clues to the ST addressee's expectations, background knowledge etc. can be inferred from other situational factors (medium, place, time, motive, and function)?

4. Is there any information about the reactions of the ST receiver(s) which may influence translation strategies?
5. What conclusions can be drawn from the data and clues obtained about the addressee regarding
    (a) other extratextual dimensions (intention, place, time, and function), and
    (b) the intratextual features?

These detailed guides will help users, including novice translators, properly make use of metalanguages.

For the SD property specification task, it is also effective to prepare *options*, i.e. a limited number (usually, fewer than 10) of distinct categories from which users can select specific one(s) as a property value. The preparation of such sharable options is important not only for the efficient specification of the SD property but also for accurate communication among the actors involved in the translation process. In the same way as the list of property names was constructed, options for property values can be collected, examined, and organised in advance through a literature review. For example, we can observe various types of options for the **(C05) function** in Table 6.2, including expressive, informative, vocative, operative, referential, phatic, aesthetic, and metalingual. We can then control the term variations and formulate a set of options. In many cases, comprehensive, well-formulated options are unavailable from the literature on translation studies. We thus need to devise sets of options with reference to the knowledge accumulated in the related fields. For instance, the subject classification schemes developed in the library and information science field would be useful for grasping the whole range of the **(K01) subject field** in Table 6.1, and the categories of genres discussed in applied linguistics could be used to develop options for the **(K03) genre**.

Another related direction for supporting users is to provide them with decision-making information to select the most appropriate option for a property or element name for a text span. As mentioned in Section 6.1, Burchardt and Lommel (2014) proposed a decision tree to classify translation issue instances into the MQM issue categories. Similarly, Fujita et al. (2017) developed a decision tree for the translation issue typology as a navigation tool for users, demonstrating its potential contribution to the consistent classification of issues through user experiments. The development and use of decision trees or classification schemes can be a promising approach for enhancing the process of SD analysis.

6.6.2 *Proper management of the SD analysis process*

In project-based translation settings, in which many players are involved, the proper task assignment is important. In particular, as mentioned in Section 6.2, the SD property specification task can largely be implemented in the pre-production phase of translation projects. Dunne (2011, p. 270), for example, claims that "the translator and translation project manager engage in a collaborative process with the requestor" to identify the necessary information for

translation, including source text information such as the author, audience, and purpose. ISO (2012; 2015) also situates the source language content analysis task within the distinctive project preparation process preceding translation. Here, project managers and requesters (clients) play important roles in specifying the SD properties. For example, ISO (2012, p. 19) states that "[t]he requester should identify the subject fields of the source content." As part of requirements collection and scope definition processes in localisation projects, Levitina (2011) explains that the following items that are particularly related to the SD properties are specified by project managers: file format of the authored content (**(F03-b) format** and **(F03-c) markup**), volume of work (**(F03-a) volume**), adherence to source language style guide (**(T05) quality**), and source language glossary (**(K06-b) terminology**).

It would also be effective to assign a part of the SD analysis task to linguists and terminologists who have specialised knowledge and skills related to SD properties and elements. For example, if the linguist identifies elements of **(LO) locale** and **(NE) named entity** comprehensively in an SD and prepares their appropriate translations, the translator would be able to conduct subsequent transfer tasks effectively and efficiently.

To achieve proper task assignment, project managers need to coordinate the roles of the various players engaged in the task (see also Chapter 5 for detailed roles of project managers). Again, well-organised metalanguages can play a crucial role in facilitating smooth communication between them.

6.7 Conclusion and outlook

As metalanguages for SD analysis process, we compiled organised lists of document properties and elements by examining the literature on translation studies and related fields. Although the current version of our metalanguages is sufficiently comprehensive and well-organised to be used in translation practices, we will further expand and refine them through data-driven and user-focused procedures. As described in Section 6.6, the development of user support guidelines is also essential to establish metalanguage schemes, which will be our next task.

Future work also includes connecting the SD analysis process and core transfer process, metalanguages for which are provided in this chapter and Chapter 7, respectively. More specifically, we will investigate the relationship between SD properties/elements and translation strategies. For example, the function of an SD affects the ways of transferring particular SD elements into a target language. Whilst some detailed examples are introduced in Chapter 10, systematic elicitation of their linkage remains to be done.

Chapter 12 describes a translation training session model that incorporates the SD property specification process and the empirical evaluation of the session model, including the effectiveness of the SD property metalanguage. Our metalanguages are currently implemented in the translation training platform MNH-TT (see also Chapter 13). Chapter 16 offers ideas not only for technical application of our metalanguages, such as explicit use of document properties

when training neural machine translation models, but also for automation of SD analysis process, i.e. automatic identification of document properties and elements. These extensive endeavours to use and validate our metalanguages will eventually lead to the improved implementation of the SD analysis process in practice.

Notes

1 Part of the content in this chapter is also presented in Miyata (2022).
2 The full specifications of the latest metalanguages are available from the following repositories: (1) https://github.com/tntc-project/document-properties; (2) https://github.com/tntc-project/document-elements.
3 The sample source text is excerpted from the following web page: https://www.commerce.wa.gov.au/building-and-energy/building-practitioner-registration. The target text is an excerpt from our translation dataset (SDset-46; 00000366-C-8-X-13-en-ja-PEed.txt) available at: https://tntc-project.github.io.
4 We used House (2014) to refer to the content of House (1977, 1997).
5 http://docs.oasis-open.org/dita/dita/v1.3/dita-v1.3-part3-all-inclusive.html
6 https://nlp.cs.nyu.edu/ene/

References

Abott, D. P. (1996). Renaissance rhetoric. In T. Enos (Ed.), *Encyclopedia of rhetoric and composition: Communication from ancient times to the information age* (pp. 594–600). Garland Publishing.

Ando, S. (2005). *Gendai eibumpou kougi* [Lectures on modern English grammar]. Kaitakusha. (in Japanese).

Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.

Bringhurst, R. (2019). *The elements of typographic style* (4th ed. version 4.3). Hartley & Marks.

Burchardt, A., & Lommel, A. (2014). *QT-LaunchPad supplement 1: Practical guidelines for the use of MQM in scientific research on translation quality*. http://www.qt21.eu/downloads/MQM-usage-guidelines.pdf

Dunne, E. S. (2011). Project as a learning environment: Scaffolding team learning in translation projects. In K. J. Dunne & E. S. Dunne (Eds.), *Translation and localization project management* (pp. 265–287). John Benjamins.

Fujita, A., Tanabe, K., Toyoshima, C., Yamamoto, M., Kageura, K., & Hartley, A. (2017). Consistent classification of translation revisions: A case study of English-Japanese student translations. In *Proceedings of the 11th Linguistic Annotation Workshop (LAW)* (pp. 57–66). ACL.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (Revised ed.). John Benjamins.

Gouadec, D. (2007). *Translation as a profession*. John Benjamins.

Hatim, B., & Mason, I. (1997). *The translator as communicator*. Routledge.

House, J. (1977). *A model for translation quality assessment*. Narr.

House, J. (1997). *Translation quality assessment: A model revisited*. Narr.

House, J. (2014). *Translation quality assessment: Past and present*. Routledge.

ISO (2012). *ISO/TS 11669. Translation projects – General guidance.* International Organization for Standardization.

ISO (2015). *ISO 17100. Translation services – Requirements for translation services.* International Organization for Standardization.

Kageura, K. (2012). *The quantitative analysis of the dynamics and structure of terminologies.* John Benjamins.

Levitina, N. (2011). Requirements collection: The foundation of scope definition and scope management in localization projects. In K. J. Dunne & E. S. Dunne (Eds.), *Translation and localization project management* (pp. 95–118). John Benjamins.

Lommel, A., Uszkoreit, H., & Burchardt, A. (2014). Multidimensional Quality Metrics (MQM): A framework for declaring and describing translation quality metrics. *Tradumàtica*, 12, 455–463.

Miyata, R. (2022). Formulating a terminology for source document profiling through a literature review: From functionalist to documentational approaches. *Perspectives.* Advance online publication. https://doi.org/10.1080/0907676X.2022.2049830

Miyata, R., Hartley, A., Kageura, K., Paris, C., Utiyama, M., & Sumita, E. (2016). MuTUAL: A controlled authoring support system enabling contextual machine translation. In *Proceedings of the 26th International Conference on Computational Linguistics (COLING), System Demonstrations* (pp. 35–39). ICCL.

Newmark, P. (1988). *A textbook of translation.* Prentice Hall.

Nord, C. (2005). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis* (2nd ed.). Rodopi.

Peacham, H. (1577). *The garden of eloquence.* https://catalog.hathitrust.org/Record/ 011405281

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language.* Longman.

Reiss, K. (1971/2000). *Translation criticism – The potentials and limitations: Categories and criteria for translation quality assessment* (E. F. Rhodes, Trans.). St. Jerome. (Original work published 1971)

Reiss, K., & Vermeer, H. J. (1991/2013). *Towards a general theory of translational action: Skopos theory explained* (C. Nord, Trans.). Routledge. (Original work published 1991)

Sato, N., Matsuo, H., & Sasaki, K. (2006). *Retorikku jiten* [Encyclopaedia of rhetoric]. Taishukan. (in Japanese).

Snell-Hornby, M. (1995). *Translation studies: An integrated approach* (Revised ed.). John Benjamins.

University of Chicago (2017). *The Chicago manual of style* (17th ed.). University of Chicago Press.