



A QUANTITATIVE STUDY OF TRANSLATION DIFFICULTY BASED ON AN ANALYSIS OF TEXT FEATURES IN JAPANESE-TO-ENGLISH SHORT-PASSAGE TRANSLATION TESTS

Diane Leighty Howard

ADVERTIMENT. L'accés als continguts d'aquesta tesi doctoral i la seva utilització ha de respectar els drets de la persona autora. Pot ser utilitzada per a consulta o estudi personal, així com en activitats o materials d'investigació i docència en els termes establerts a l'art. 32 del Text Refós de la Llei de Propietat Intel·lectual (RDL 1/1996). Per altres utilitzacions es requereix l'autorització prèvia i expressa de la persona autora. En qualsevol cas, en la utilització dels seus continguts caldrà indicar de forma clara el nom i cognoms de la persona autora i el títol de la tesi doctoral. No s'autoritza la seva reproducció o altres formes d'explotació efectuades amb finalitats de lucre ni la seva comunicació pública des d'un lloc aliè al servei TDX. Tampoc s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant als continguts de la tesi com als seus resums i índexs.

ADVERTENCIA. El acceso a los contenidos de esta tesis doctoral y su utilización debe respetar los derechos de la persona autora. Puede ser utilizada para consulta o estudio personal, así como en actividades o materiales de investigación y docencia en los términos establecidos en el art. 32 del Texto Refundido de la Ley de Propiedad Intelectual (RDL 1/1996). Para otros usos se requiere la autorización previa y expresa de la persona autora. En cualquier caso, en la utilización de sus contenidos se deberá indicar de forma clara el nombre y apellidos de la persona autora y el título de la tesis doctoral. No se autoriza su reproducción u otras formas de explotación efectuadas con fines lucrativos ni su comunicación pública desde un sitio ajeno al servicio TDR. Tampoco se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al contenido de la tesis como a sus resúmenes e índices.

WARNING. Access to the contents of this doctoral thesis and its use must respect the rights of the author. It can be used for reference or private study, as well as research and learning activities or materials in the terms established by the 32nd article of the Spanish Consolidated Copyright Act (RDL 1/1996). Express and previous authorization of the author is required for any other uses. In any case, when using its content, full name of the author and title of the thesis must be clearly indicated. Reproduction or other forms of for profit use or public communication from outside TDX service is not allowed. Presentation of its content in a window or frame external to TDX (framing) is not authorized either. These rights affect both the content of the thesis and its abstracts and indexes.

DIANE L. HOWARD

A QUANTITATIVE STUDY OF TRANSLATION
DIFFICULTY BASED ON AN ANALYSIS OF TEXT
FEATURES IN JAPANESE-TO-ENGLISH SHORT-PASSAGE
TRANSLATION TESTS

DOCTORAL THESIS

Supervised by Dr Anthony Pym and Dr Pardee Lowe, Jr.

Intercultural Studies Group



UNIVERSITAT ROVIRA I VIRGILI
Department of English and German Studies

Tarragona
2015

UNIVERSITAT ROVIRA I VIRGILI
A QUANTITATIVE STUDY OF TRANSLATION DIFFICULTY BASED ON AN ANALYSIS OF TEXT FEATURES IN
JAPANESE-TO-ENGLISH SHORT-PASSAGE TRANSLATION TESTS
Diane Howard



Professor Anthony Pym
URV. Av. Catalunya 35
43002 Tarragona, Spain
anthony.pym@urv.cat

#

November 4, 2015

I hereby certify that the present study *A quantitative study of translation difficulty based on analysis of text features in Japanese-to-English short-passage translation tests*, presented by Diane Howard for the award of the degree of Doctor, has been carried out under the supervision of myself at the Department of English and Germanic Studies of the Rovira and Virgili University, and the co-supervision of Dr. Pardee Lowe Jr., and that it fulfills all the requirements for the award of Doctor.

Professor Anthony Pym
Intercultural Studies Group
Universitat Rovira i Virgili

President
European Society for Translation Studies

UNIVERSITAT ROVIRA I VIRGILI
A QUANTITATIVE STUDY OF TRANSLATION DIFFICULTY BASED ON AN ANALYSIS OF TEXT FEATURES IN
JAPANESE-TO-ENGLISH SHORT-PASSAGE TRANSLATION TESTS
Diane Howard



#

Pardee Lowe, Jr.
3800 Powell Lane, #1013
Falls Church, VA 22041

November 20, 2015

I hereby certify that the present study *A quantitative study of translation difficulty based on analysis of text features in Japanese-to-English short-passage translation tests*, presented by Diane Howard for the award of the degree of Doctor, has been carried out under the supervision of Prof Anthony Pym at the Department of English and Germanic studies of the Rovira and Virgili University, with my co-supervision, and that it fulfills all the requirements for the award of Doctor.

X Signed Electronically
Pardee Lowe, Jr. PhD
US Department of Defense (ret.)

UNIVERSITAT ROVIRA I VIRGILI
A QUANTITATIVE STUDY OF TRANSLATION DIFFICULTY BASED ON AN ANALYSIS OF TEXT FEATURES IN
JAPANESE-TO-ENGLISH SHORT-PASSAGE TRANSLATION TESTS
Diane Howard

Acknowledgements

I am most grateful to my supervisors, Dr Anthony Pym and Dr Pardee Lowe, Jr., for their support and encouragement during the course of this research. Dr Pym has helped me refine my thinking and expression in many ways and has always proved to be amazingly generous. Dr Lowe introduced me to the Interagency Language Roundtable system, which forms the basis of this study.

Dr Natalia Lusin and Dr Ralph Howard read the manuscript and made many valuable comments, for which I warmly thank them. Dr Sergei Lusin has been an invaluable support as a sounding board, critic, and believer in the project. I could not have completed this research without him. My heartfelt thanks to Dr Esther Torres Simón for help with navigating the administrative labyrinth.

I would like to express my appreciation to the American Translators Association Executive Director and President for permission to use test materials.

At the time of the Japanese-to-English tests used in this study, one of the ATA passage selection guidelines was that the material not be available on the Internet. As a result, I have been unable to locate the authors and publishers of the six Japanese passages used in the analysis. Information about the copyright holders would gratefully acknowledged and added to any future publications based on these passages.

Abstract

Short-passage translation tests are frequently administered as a way of screening translators and in the course of professional certification examinations. However, despite the ubiquity of such tests, there has been very little research on test passage selection or on what exactly is being tested by a given text. In addition, only limited empirical research has been done on translation difficulty, with much of the published work focusing on word or phrase-level problems. At the same time, U.S. government agencies have accumulated extensive pragmatic experience in language and translation testing based on the language skill descriptions set out by the Interagency Language Roundtable (ILR). In response to the above, I have attempted to link specific text features defined as characteristic of ILR reading level 3 (general professional proficiency) to target text solutions and to then score these for difficulty using facility values.

The data set was candidate examination papers from the American Translators Association (ATA) Japanese-to-English Certification Examination (36 translations of the general passage at ILR reading level 3 and 22 translations of the technical passage at reading level 2+). The translators were self-identified professional translators meeting ATA pre-requisites for sitting the examination.

Tense of the final verb in each sentence and implicit/explicit cohesion markers were set as markers of target text cohesion thought to indicate whether the translator was following an argument requiring the use of inference, a feature of ILR level 3 texts. Tense and the other cohesion markers were dichotomously scored and facility values calculated for each item to indicate the level of difficulty in the test population. Problem points identified in earlier studies—modified nouns and culturally specific terms—were scored in the same way and facility values calculated. A close reading of the translations was also done to look for unexpected and language-specific difficulties.

The results showed that it is possible to link specific passage elements that can be objectively scored to desired target text features, in this case, the creation of a cohesive English text. They further indicated that a specific characteristic of ILR reading level 3—using inference to follow an argument—was difficult for this population. The results also confirmed that elements found to be translation difficulties in previous studies (modified nouns and official terms/culturally specific terms) were difficult for this sample of professional translators. These findings can be used to refine passage selection in short-passage translation tests.

Contents

1. Introduction	1
2. Background and literature review	5
2.1 Translation difficulty	5
2.1.1 Earlier work on translation problems and difficulty	6
2.1.2 Difficulty in Translation Studies: empirical studies	11
2.1.3 Difficulty in Translation Studies: eye-tracking studies	14
2.1.4 Difficulty in Translation Studies: summary	15
2.2 Short-passage tests	15
2.2.1 In the academic setting	15
2.2.2 In the professional setting	17
2.2.3 The question of authenticity	19
2.3 Two models of testing: communicative language ability and performance.....	22
2.3.1 The communicative language ability model.....	22
2.3.2 Proficiency tests and the performance model.....	28
2.3.3 Summary.....	30
2.4 ILR reading and translation levels.....	31
2.4.1 Description of the ILR levels.....	31
2.4.2 Published work on the ILR system.....	34
3. Aims and assumptions.....	37
3.1 Aims.....	37
3.1.2 Aims and objectives	39
3.2 Determining test items.....	41
3.2.1 Why identifying test items is important	41
3.2.2 Test items to be studied	43
3.2.2.1 ILR reading levels.....	44
3.2.2.2 Cohesion markers.....	45
3.2.2.3 Selecting problem points.....	47
3.3 Analysis.....	48
3.3.1 Measurements to be done	49
3.3.2 Earlier studies using facility values and dichotomous scoring.....	50
3.4 Summary.....	51
4. Materials and methods for analysis	53
4.1 Data sets	53
4.1.1 Test population	53
4.1.1.1 Limitations of test population	55
4.1.3 Passage selection criteria	56
4.1.3.1 Limitations of passage data set.....	57
4.1.4 Passages.....	58
4.1.4.1 General passage A.....	58
4.1.4.2 General passage B.....	60
4.1.4.3 General passage C.....	62
4.1.4.4 General passage D.....	65
4.2 Analysis method for main data set	68
4.2.1 Cohesion markers: specific verbs and other implicit and explicit markers	68
4.2.1.1 General passage A.....	68
4.2.1.2 General passage B.....	70
4.2.1.3 General passage C.....	71
4.2.1.4 General passage D.....	73
4.2.2 Problem points.....	74
4.2.2.2 Passive constructions.....	75

4.2.3 Features of ILR reading level 3	77
4.2.4 Facility values.....	77
4.3 Method of data presentation	78
4.4 Supplementary analysis based on earlier studies.....	78
4.5 Data set and analysis of the technical passages	78
5. Analysis of the general passages: results for tense, other cohesion markers, and problem points	81
5.1 Results: general passage A	81
5.1.1 General passage A: tense and other cohesion markers	81
5.1.2.1 Summary tables and facility values.....	86
5.1.3 Problem points.....	87
5.2 General passage B.....	90
5.2.1 General passage B: tense and other cohesion markers	90
5.2.2.1 General passage B: summary tables and facility values.....	95
5.2.3 General passage B: problem points	97
5.3 General passage C.....	98
5.1.2 General passage C: tense and other cohesion markers	98
5.3.2.1 Summary tables and facility values.....	103
5.3.3 General passage C: problem points	103
5.4 General passage D.....	104
5.4.1 Tense and other cohesion markers.....	104
5.4.2 General passage D: summary tables and facility values.....	109
5.4.3 General passage D: problem points.....	110
5.5 Summary of results	112
5.5.1 Summary of tense results.....	112
5.5.2 Summary of other cohesion marker results	113
5.5.3 Summary of problem point results	113
5.5.3.1 Modified nouns	113
5.5.3.2 Passive constructions.....	114
6. Results: close reading of the general passages.....	115
6.1 Discussion: general passage A.....	115
6.1.1 Discussion of individual sentences.....	115
6.1.1.1 General passage A, sentence 1	115
6.1.1.2 General passage A: sentence 2	116
6.1.1.3 General passage A: sentence 3	117
6.1.1.4 General passage A: sentence 4a and 4b.....	118
6.1.1.5 General passage A: sentences 5 and 6.....	118
6.1.2 Discussion of individual translations.....	119
6.2.1 Discussion of individual sentences.....	121
6.2.1.1 General passage B, sentence 1	121
6.2.1.2 General passage B, sentence 2	122
6.2.1.3 General passage B, sentences 3 and 4	122
6.2.1.4 General passage B, sentences 5a and 5b	123
6.2.1.5 General passage B, sentence 6	125
6.2.1.6 General passage B, sentence 7	126
6.2.2 General passage B: discussion of individual translations.....	127
6.3 Discussion: general passage C.....	128
6.3.1 Discussion of individual sentences.....	128
6.3.1.1 General passage C, sentences 1 and 2	128
6.3.1.2 General passage C, sentence 3	128
6.3.1.3 General passage C, sentences 4, 5, and 6	129
6.3.1.4 General passage C, sentence 7	130
6.3.1.8 General passage C, sentences 8 through 11.....	130
6.3.2 General passage C: discussion of individual translations.....	130
6.4 Discussion: general passage D.....	132
6.4.1 General passage D: discussion of individual sentences.....	132

6.4.1.1 General passage D, sentences 1,2, and 3	132
6.4.1.2 General passage D, sentences 4,5, and 6	133
6.4.1.3 General passage D, sentence 7	134
6.4.1.4 General passage D, summary of sentence analysis	135
6.4.2 General passage D: discussion of individual translations.....	136
6.5 Summary.....	138
7. Supplementary analysis based on earlier studies	139
7.1 Items of translation difficulty in Hale and Campbell	139
7.1.1. Translation results for culturally specific terms	140
7.2 Items of translation difficulty in eye-tracking studies	144
7.2.1 Jensen's indicators of text complexity.....	144
7.2.2 The Translation Difficulty Index	146
7.3 Summary.....	149
8. Technical passages: results and analysis	151
8.1 Technical passage A.....	151
8.1.1 Technical passage A: passage and structure.....	151
8.1.2 Technical passage A: tense, other cohesion markers, and problem points.....	153
8.1.3 Technical passage A: results of analysis of tense, other cohesion markers, and problem points.....	155
8.1.4 Technical passage A: summary of errors and comparison with general passage A	160
8.1 Technical passage B	161
8.2.1 Passage and structure.....	162
8.2.2 Tense and other cohesion markers.....	163
8.2.3 Technical passage B: results of analysis of tense, other cohesion markers, and problem points.....	165
8.2.4 Summary of errors and comparison with general passage B.....	169
9. Discussion	173
9.1 Scored items.....	173
9.1.1 Cohesion markers	173
9.1.1.1 Tense	173
9.1.1.2 Other cohesion markers.....	175
9.1.2 Problem points.....	176
9.1.2.1 Modified nouns	176
9.1.2.2 Passive constructions.....	177
9.1.3 Facility values	177
9.2 Results of close reading	180
9.3 Results from earlier studies.....	184
9.3.1 Hale and Campbell (2002).....	184
9.3.2 Eye-tracking studies	186
9.3.3 Sun and Shreve	188
9.4 Verification of ILR reading level and item difficulty	189
9.4.1 Study assumptions	189
9.4.1.1 Can elements related to inference be identified in the score text and scored in the target text?.....	190
9.4.1.2 Do the distinctive features of ILR reading level 3 make passages at that level more difficult to translate than passages at reading level 2+?	190
9.4.2 Implications for passage selection.....	191
9.5 Potential applications.....	192
9.5.1 Potential testing applications	192
9.5.2 Potential applications in translator training	194
10. Conclusion	197
10.1 Study findings.....	197
10.2 Study limitations	198
10.3 Potential applications.....	198

10.4 Avenues for future research.....	199
References	201

1. Introduction

Administration of short-passage translation tests or requiring translation of a sample passage is a common way of screening translators in the professional setting. Short-passage tests are also used in professional certification/accreditation examinations such as those administered by the American Translators Association (ATA) and the Australian National Accreditation Authority for Translators and Interpreters (NAATI). At the same time, current standards (e.g. ASTM F 2575-06) and ISO 9001: 2008 quality management procedures either require or strongly recommend translator testing as a quality control measure. Together these factors fairly well guarantee that translator testing will remain an important element in the professional venue.

However, despite the widespread use of short-passage tests, very little research has been done on fundamental test features such as effective passage length and amount of time allotted for translation (Campbell and Hale 2003). Assessment and scoring has received more attention (e.g. Williams 2004; Secară 2005; Angelelli 2009; Hale et al. 2012), but there is no consensus on what features should be marked or how they should be scored.

Almost no research has been published on passage selection or what, precisely, is being tested by a specific text. Passage selection is frequently done based on length (often the number of words translators are thought to produce in an hour) and “authenticity”, which can mean either a source-language text that has not been changed in any way or one resembling other texts translated by the testing entity. Descriptions of *what* is being tested tend to be vague (e.g. comprehension of the source language, writing skills in the target language, and translation techniques ([Angelelli 2009: 23])). The lack of a clear picture of what short-passage tests are measuring results in more testing because translation agencies are reluctant to accept the results of certification examinations and prefer to set their own exams.

In addition, there has been limited empirical research on what constitutes a *translation difficulty*, partly because this has been viewed as a purely subjective element (Nord 2005). Previous empirical studies have equated *difficulty* with *cognitive effort* (e.g. Hale and Campbell 2002; Sun 2012), but have found that effort has little relationship with accuracy. At the same time, in the process of developing tests to measure the language abilities of foreign-service officers, the U.S. State Department

and other government agencies created the Interagency Language Roundtable (ILR) skill descriptors. This system was developed to measure language proficiency; however, many U.S. government agencies use the ILR reading scale to determine passage selection for translation tests. In 2008, the ATA began mandating that the general passages for its certification test be ILR reading level 3 (Koby and Champe 2013). The assumption is that reading level difficulty can in some way be equated to translation difficulty. This connection has received little empirical study, although it has been shown to be pragmatically useful in thousands of language and translation tests (Lowe 1986; Child 1998).

In light of the above problems and as a small step toward making short-passage tests more useful, this study will focus on two main issues. First is the question of translation difficulty and whether this can be linked to features of progressive reading complexity as defined by the ILR reading level descriptors. Second is the corresponding question of whether using reading level descriptors for passage selection can help define and articulate what is being tested.

The passages examined were four general passages (36 translations) and two technical passages (22 translations) used in the ATA Japanese-to-English certification examination and their corresponding translations.

A characteristic of ILR reading level 3, “general professional proficiency” (ILR 2015), is the need to follow an argument requiring a degree of inference. In this study, inference is represented by text cohesion achieved through verb tense and implicit and explicit cohesion markers. Because tense is used very differently in Japanese and English, in order to create a cohesive English language text a translator must make tense decisions that are assumed to reflect whether an inferred argument in the source document is being followed and that also demonstrate whether a properly connected text has been created. The implicit cohesion markers are words that lack an explicitly stated antecedent, and so require inference to be resolved in translation. Explicit cohesion markers are words such as *sarani* (*furthermore*) and *shikashi* (*therefore*) that signal links between sentences, paragraphs, or ideas.

The tense of the final verb in each sentence is scored as *acceptable* or *unacceptable* and the scores for the test population—candidates taking the Japanese-to-English ATA certification examination—used to calculate the facility values (percentage of correct responses) for the population. In the same way, the implicit and explicit cohesion markers are also dichotomously scored and the facility values

calculated. Low facility values are assumed to represent translation difficulties, in the sense of failure to correctly resolve a source text feature in the target text. This measure allows accuracy to be rated at the same time as difficulty. Because the items selected for scoring are connected to ILR reading level 3 features, the assumption is that the facility scores indicate whether these traits can be mapped onto translation difficulties for a specified population of test takers.

In addition to tense and the implicit and explicit cohesion markers, problem points have been identified based on earlier studies of translation difficulty. These are modified nouns with inclusive character strings of fifteen characters or more and passive verb constructions. Because both of these can be connected to inference and the need to follow an argument and because they can be scored in a reasonably objective manner, they are also marked as *acceptable* or *unacceptable* and the facility values calculated.

Official terms/culturally specific terms are another translation problem identified in previous research (e.g. Hale and Campbell 2002; PACTE 2008). These are discussed but not scored because I could not find a way of marking the translation solutions that was not subjective. This problem has been noted before (Hale and Campbell 2002).

The main data set was taken from four general passages that were at ILR reading level 3. A secondary data set consists of two technical passages at ILR reading level 2+. Tense and implicit and explicit cohesion markers are scored in the same way as for the general passages. However, because the technical passages were informative texts rather than texts setting out an argument, the examination of cohesion was expanded to within sentences as well as between sentences. The general and technical passages are compared to investigate the difference in difficulty between ILR reading level 3 and reading level 2+.

Chapter 2 sets out some of the research background pertaining to short-passage translation tests and reviews the literature pertaining to translation difficulty in the academic and professional settings. It then discusses two of the main models currently used in testing: the communicative language ability model and the performance model. It concludes with a summary of the ILF reading levels.

The aims and assumptions of the study are given in Chapter 3. This includes a discussion of the rationale behind the study, the specific goals, and how the items to be scored were determined. The ILR reading levels are described in more detail, as are the reasons for selecting problem points from earlier studies. The only measurements done

in this research are the calculations of facility scores. Findings are presented in tabular form.

Chapter 4 describes the data sets for the general passages, including the test population, test conditions, passage selection criteria, and passages. The complete source texts of the passages are given, accompanied by a possible translation. These are followed by a description of the argument that needs to be followed in order to resolve the passage correctly. A description is given of the final verbs in each Japanese sentence, possible English translations, with notes on possible translation problems. The implicit and explicit coherence markers are set out, and the problem points—modified head nouns and passive constructions—are also described.

The results are given in Chapter 5 in tabular form for both tense and the implicit/explicit cohesion markers. I hope that having this information concerning difficulties in translating Japanese verbs to an acceptable English equivalent will prove helpful to those teaching translation and Japanese as a foreign language. The discussion of each general passage concludes with tables of overall errors for tense, the implicit/explicit cohesion markers, and the problem points. The tables allow comparison between passages and also indicate what it was in the passages that the test takers found difficult.

Chapter 6 gives the details of a close reading of the general passages. In that chapter, I essentially ask the candidates: What was so hard about the passage? The objective is to find unexpected elements that were difficult and also to analyze some of the individual translations at greater length to determine what sorts of problems each of the scored markers uncovers and whether they permit any statements about translation quality or predictions of future performance.

Translation difficulties from previous studies are discussed in Chapter 7 and facility values are calculated for items that earlier researchers identified as problem points. These are culturally specific terms and elements related to readability indices.

The semi-technical passages are given in Chapter 8, followed by a discussion of the set markers (tense, implicit/explicit facility markers, and modified nouns), as well as cohesion within sentences and unexpected difficulties.

The discussion in Chapter 9 examines the results in light of earlier research and suggests possible applications in terms of passage selection, and possible applications in testing and translation teaching. The main findings, study limitations, potential applications, and avenues for future research are described in Chapter 10.

2. Background and literature review

This study is based on earlier work on translation difficulty and identification of translation problems, short-passage translation testing, the communicative language ability and performance models of testing, the idea of authenticity, and ILR reading level descriptors, which will be reviewed below. The discussion of short-passage testing will focus on the two elements that are most pertinent for this study: passage selection and statements of what is being tested.

2.1 Translation difficulty

A typical reaction when the word *difficulty* is mentioned is to throw up one's hands and announce that difficulty is subjective and so is impossible to rate. However, this subjective element does not mean that we are reduced to saying, "it depends on who is translating what in which circumstances". By carefully defining what constitutes a difficulty in an examination (and not ontologically), by controlling for population and test setting, and by collecting and analyzing test data on the difficulties set, it should be possible to take at least some initial steps in developing translation test items that not only reflect different levels of difficulty but that can also be linked to specific text features, and so allow test designers to say exactly what is being tested and scored. The key to doing this is to look at difficulty in terms of populations, rather than focusing on individuals.

A good starting point is to look at what earlier scholars have defined as translation difficulties or translation problems. Some scholars (e.g. Nord 1997: 64) have distinguished *difficulties* from *problems*, with *difficulties* pertaining to the individual and *problems* being identifiable textual features. In terms of passage analysis and test design, however, they can be seen as the same thing with the testee population responses to a test item (a *problem*) resulting in a numerical value—a facility value—showing the *difficulty* of that problem for the given set of test takers. The goal is to identify features in the source text that can be linked to target text solutions indicating whether an item of interest—here, text cohesion—has or has not been achieved and that then allow a facility value to be calculated.

2.1.1 Earlier work on translation problems and difficulty

In the period from the mid 1950s through the mid 1960s, translation problems—without the term being explicitly defined—were generally seen as linguistic problems. The most charming example is Vinay and Darbelnet’s road trip from New York to Quebec in which they found stylistic problems in the road signs (1995: 3), leading to the conclusion that “recognizing these equivalences [in comparative stylistics] is one of the main problems of translation” (1995: 5) and ultimately to their seven translation procedures: borrowing, calque, literal translation, transposition, modulation, equivalence, and adaptation (1995: 31-41). These procedures can be viewed both as ways of solving problems posed by the source text and as possible elements for testing. Vinay and Darbelnet noted that the “degree of difficulty” provided the “ordering sequence” for the seven procedures (1995: 42), and that, in general, all seven methods can be applied “at the three planes of expression, i.e. lexis, syntactic structure, and message” (1995: 40-42). Thus, based on the seven procedures, test problems could be set at the word or text level.

In subsequent identification of textual problems, or “rich points”, as described in research by the PACTE group (2003, 2005, 2008, 2011) and as will be discussed in studies cited below, the problems set for analysis have generally been at the phrase level. What would appear to be tested then is terminology, knowledge of translation procedures (including use of parallel texts), research skills, and ability to turn a phrase. These are all elements worth testing, but if terms and phrases are not very carefully chosen, they may not provide the test evaluator with useful information about whether the person taking the test can perform similar tasks in the future. In other words, how worried should I, as a project manager, be that a candidate who has performed well in other ways has not come up with an appropriate translation of *slippery when wet*? Will this person put my project at risk? It is rather hard to say, which suggests that tests (and the related scoring) based on translation of phrases may not be helpful in the professional setting, particularly in cases when the problem could have been solved by providing a glossary. An exception to this would be translation testing in a specialized area where correct terminology is crucial, for example, in medical or legal translation.

Quine, also writing at the end of the 1950s, finds translation to be not just a linguistic challenge, but also an ontological problem for which “good” solutions

(*gavagai* = rabbit) exist only for “sentences with visibly direct conditioning” (1960: Kindle location 1036). Other translational problems are to be solved through “analytical hypotheses” or “conjectural equating” of segmented utterances that are tested, then accepted or rejected, with the goal of arriving at the “translation of coherent discourse” (1960: Kindle location 934). Ultimately, however, “radical” translation points at the “difficulty or indeterminacy of correlation” suggesting that, ontologically, some types of translation problems ultimately cannot be solved: for “theoretical sentences such as ‘Neutrinos lack mass’ [...] no hint of the stimulatory conditions of assent or dissent can be dreamed of that does not include verbal stimulation from within the language” (1960: Kindle location 1016).

Moving into the 1960s, Catford, like Quine, finds the crux of translational difficulties in the “problem” of equivalence:

Translation fails—or untranslatability occurs—when it is impossible to build functionally relevant features of the situation into the contextual meaning of the TL text. Broadly speaking, the cases where this happens fall into two categories. Those where the difficulty is *linguistic*, and those where it is *cultural*. (Catford 1965: 94; emphasis in the original)

Catford defines linguistic untranslatability as cases in which “an *ambiguity* peculiar to the SL [source language] text is a functionally relevant feature,” and offers puns as an example. He places *oligosemy* (restricted range of meaning) at the opposite end of the linguistic ambiguity scale, noting that it only becomes important when it is a “functionally relevant feature of the situation” (1965: 96). *Cultural untranslatability* refers to words that have no equivalence in the target language: his examples are *sauna* and *yukata*. A “bad translation” of such terms is one that creates not so much “cultural shock” as “collocational shock” (1965: 102).

The approach to translation problems in Quine and Catford is primarily interesting on the linguistic and philosophical levels, although it could be used to identify test problems for testing. However, translating puns and notions unavailable in the target language mainly test verbal facility and knowledge of translation procedures (a goal in some of the PACTE experiments [PACTE 2008]), with successful solutions suggesting that the candidate could handle future puns, etc. As noted above with respect

to Vinay and Darbelnet's procedures, this type of problem has limited predictive power¹ and a test designer would need to be very clear about why the information resulting from solving the problem of, for example, *chiburekki*,² would be useful for decisions about a candidate. (The EU, for example, does test for cultural knowledge in some translation tests [Lafèber 2012b: 149]). On the other hand, such problems would be useful during testing in translation training in terms of making students more aware of the types of problems they might encounter. But it is important to remember that classroom teaching for translator awareness and feedback and testing in a professional setting for the purpose of making decisions have very different goals.

Moving into the period when one can properly speak of Translation Studies, Nord, writing from a functionalist approach, expanded the discussion by differentiating between *translation difficulties* and *translation problems*:

A translation *problem* is an objective (or inter-subjective) transfer task which every translator (irrespective of their level of competence and of the technical working conditions) has to solve during a particular translation process. [...] Translation *difficulties*, on the other hand, are subjective and have to do with the individual translator and the specific working conditions. A particular translation problem which seems very difficult to the beginner will remain a translation problem, even when the student has learned to cope with it. It can turn into a difficulty again, though, if the translator has to solve it without the necessary technical resources. (Nord 2005: 166-167)

One can see how this distinction between *problem* and *difficulty* fits in with the approach being taken here in which the goal is to identify specific textual elements and then analyze test-taker response to these to obtain a group rating of difficulty, rather than Nord's individual rating. Having made a distinction between objective problems and subjective difficulty, Nord then defines four categories of translation problems:

- *pragmatic*, which is based on the contrast between the original context of the source text and the setting for which the target text is being produced;

¹ One could use puns and terms of "cultural untranslatability" to test creativity and range of cultural repertoire, but this would need to be done across a range of terms before a tester could comfortably say that a candidate was proficient in this area.

² A meat-filled fried turnover from Crimea

- *convention-based*, with differences in text-type, measurement, and translation conventions given as examples;
- *linguistic*, which occur due to structural differences between languages; and
- *text-specific*, which arise “from the particular characteristics of the source text (e.g. the translation of a play on words)” (Nord 2005: 167).

As with the linguistic approach to translation problems and difficulties, the types of problems identified, other than pragmatic problems, tend to remain at phrase level. In another work, Nord argues that there is—or should be—a “functional hierarchy of translation problems” (1997: 67) in which problems are “dealt with in a top-down way”, in the order they were presented above: pragmatic, convention-based, linguistic, then text-specific. The rationale is that once the function (documentary or instrumental in Nord’s system) of the translation is known, many of the decisions for solving the other three categories of problems will already have been made (1997: 67-68).

An interesting aspect of this hierarchy is that Nord’s functionalist approach, with its focus on pragmatic problems, would appear to have moved translation problems beyond the phrase level seen in linguistics-based theory. The overarching problem to be solved in translation becomes the creation of a target document that fulfills the translation instructions. While the instructions should govern lower-order decisions, such as how to handle measurements (conventions) and then text-specific problems (puns and proverbs, for example), it is not clear how this hierarchy would be useful in screening tests other than perhaps to reject out of hand candidates who failed to read or follow the instructions and perhaps those who viewed the exam as a test of bilingualism rather than translation. Moreover, the workplace and testing reality is that such instructions are often non-existent, or, as is the case with the certification test passages in this study, laconic to the point of uselessness (e.g. “Translate for an educated general reader”). Again, as with terminology problems that can be addressed by providing a glossary, the ability to solve problems by following translation instructions would seem to have limited predictive power about how the candidate might handle other texts. In addition, setting objective criteria for scoring would be problematic.

Nord also sets out four categories of translation difficulty in the context of classroom translation tasks: text-specific, translator-dependent, pragmatic, and technical (2005: 168-171). The first category—text-specific difficulty—is somewhat in contradiction to her previous assertion that difficulty is subjective:

Text-specific difficulties are related to the degree of comprehensibility of the source text and can be discovered by going through the intra-textual factors of text analysis, ignoring translator-dependent criteria as far as is possible. (Nord 2005: 168)

This suggests that texts have features that are inherently difficult, including:

The quantity and complexity of content, the number of presuppositions (degree of redundancy), the degree of consistency and clearness of text composition (including theme-rheme structure), the complexity of ST [source text] structure in lexis and syntax, including any defects, the number of suprasegmental features which make comprehension easier, and the number and nature of non-verbal elements. (Nord 2005: 168)

Nord goes on to say that, “[t]he degree of difficulty presented by the intratextual factors is determined by the amount of information the translator has on the extratextual dimensions”. However, this idea is somewhat contradicted by a statement under “translator-dependent difficulties” that “in translation teaching deficient linguistic competence in SL or TL is usually the biggest stumbling block” (2005: 169). In other words, while putting a text in context will help translators understand it better (and expert knowledge in the subject matter will help even more), some texts, because of their linguistic structure, are simply hard to read. While this can be controlled for in the classroom setting, in the professional setting documents are usually assigned by subject matter and not by level of difficulty (which is frequently unknown to the project manager). Therefore, knowing that a given translator can handle a fairly wide range of not just fields, but also text types at different levels is useful information for those assigning translations and suggests a specific area for testing.

Nord’s two remaining areas of translation difficulties are “pragmatic difficulties” and “technical difficulties”. The former have to do with what type of text is being translated for what defined audience, and the latter with the resources for research and documentation available.

In setting up her categories of problems and difficulties, Nord is specifically addressing translator training and how manipulating the parameters of problems and

difficulties can help develop translator skills. Messages that can be taken away for defining difficulty for professional testing include the notions that lack of context and lack of resources make the translation task more difficult and that a tester would need a defensible reason for denying these things to candidates (e.g. what would be the point of not allowing the use of spellchecker software?). Also, problems that could be solved by providing parallel documents and other references are probably not very useful as test items unless the objective is to specifically test skill in locating and using such resources.

2.1.2 Difficulty in Translation Studies: empirical studies

In 1999 Stuart Campbell and Sandra Hale described a project addressing the following questions: “What makes a text difficult to translate? Can the translation difficulty of English source texts be assessed? And does a given English text cause equal difficulty when translated into different languages?” The last question can be expanded to ask, “How universal are translation difficulties?” Another goal of the project was to separate “comprehension and production difficulties”. In the initial study, involving translation of an English text into the unrelated languages of Spanish, Arabic, and Vietnamese, they found that “a considerable portion of text difficulty is common to translation into all three languages”.

Difficulty was defined as:

Where subjects in one language provided different renditions of a chunk of text, that chunk was judged to be difficult on the basis of the processing effort needed to decide among the various choices available. If the chunk caused a similar level of difficulty in all three languages, then it was judged to be universally difficult. (Campbell and Hale 1999: 2)

The areas identified as difficult were “words low in propositional content” (or “grammatical metaphor”), “complex noun phrases” (which appear frequently in Japanese), “abstractness”, “official terms”, and “passive verbs”. Their preliminary conclusion on the separation of comprehension and production difficulties were that difficulties in comprehension were likely to be similar across languages, while the

“lexis and grammar of the target language” probably create different levels of difficulty in production (Campbell and Hale 1999: 2).

Campbell focused on source text difficulty in a 1999 article using a cognitive approach to “on-line translation” (translation under time constraints with limited reference resources available), proposing that “one way of gauging text difficulty in on-line translation is by identifying those lexical items that require higher amounts of cognitive processing” with the complexity of such processing inferred from “the mean number of alternative renditions made by a group of subjects translating the same item” (1999: 37-39). This makes a great deal of sense as a way of identifying cognitive difficulty, and the types of difficulties identified are useful items for triangulation, but, as Campbell notes, “this criterion of difficulty is not in any way entangled with the idea of correctness” (1999: 39). The picture changes a bit once correctness, a necessary element in test evaluation, is introduced. The certification test translations analyzed in this study suggest that multiple translators can mistranslate a segment of text in approximately the same way with what I suspect to be very little cognitive effort. In the present study, *difficulty* is defined by item difficulty (facility value) and not by the amount of effort on the part of any individual. However, the fact that the English source texts in Campbell’s study resulted in common difficulties when translated into Spanish Arabic, and Vietnamese, strongly suggest that the point of difficulty is inherent to the source text and that it might be possible to improve translation testing “through the selection of source texts graded for difficulty” (Campbell 1999: 57). Contributing to the possibility of such source text selection is one of the aims of this study.

Hale and Campbell (2002) addressed the issue of accuracy in what was unfortunately the final paper in their series. In that article they examine whether there is “a correlation between the number of available choices and the level of accuracy of a translation” (2002: 17). As in the earlier studies, the source text features of interest were official terms, complex noun phrases, passive verbs, and metaphors. They found that there was “not a clear correlation between choices and accuracy” and that “multiple choices” did “not necessarily mean that the item in question [was] difficult” (2002: 29). Ultimately, they called for “further empirically based research [...] to map item types and to characterize the difficulties they cause”, something that is being attempted here.

The PACTE group set “rich points” in the source text as part of a 2005 study of translation competence. The features of these were “(1) that they should provide variety in the types of translation problems studied, (2) that they do not lead to immediate and

acceptable solutions and (3) that they should be homogeneous in all the languages (so comparisons can be made)” (PACTE 2005: 614).

The problems were at the phrase level and were subjectively (it appears) scored as acceptable, partially acceptable, or unacceptable (i.e. the scoring was not dichotomous). In a 2011 study, PACTE identified rich points corresponding to what they have defined as linguistic problems (lexis and morphosyntax), textual problems (coherence, text type, etc.), extralinguistic problems (including cultural and subject matter knowledge), intentionality (including intertextuality and implicatures), and “problems relating to the translation brief and/or target-text reader” (PACTE 2011: 327). Setting these categories aside for the moment, some of the “rich points” identified in the English text used in the PACTE study were similar to the types of source text problems found by Campbell and Hale. The apposition in “‘Trojan horse’ program which could allow a hacker to take remote control of infected machine” could be seen as a grammatical metaphor and “Cheltenham-based virus filtering firm” as a complex noun phrase. While not an “official term”, “keylogger” is a specific technical term requiring approximately the same approach as an official term (PACTE 2011: 342). These similarities suggest that those features may well be useful testing problems, although the official term/technical term category would be most useful in testing a specific translation domain and only then when the candidate had access to adequate resources for solving the particular problem.

Sun and Shreve (2014), in a study on measuring translation difficulty, define it as “the extent to which cognitive resources are taken up by a translation task for a translator to achieve objective and subjective performance criteria” (2014: 99), which pertains to the translator and is therefore subjective, but they go on to discuss “a text’s level of translation difficulty” (ibid). Since they then suggest that the latter is measurable by objective means such as readability indices, the implication is that translation difficulties at the text level are objective and fixed. In their experiment they used readability indices to categorize the texts as easy, medium, and difficult, then had a group of student translators rate the degree of difficulty on a ten-point scale before and after translating the passages and also fill out a scale measuring subjective workload. The results showed that passages the translators thought would be difficult were indeed difficult for them and this correlated with the measurement of subjective workload. However, accuracy (a quality score) was not related to subjective workload.

2.1.3 *Difficulty in Translation Studies: eye-tracking studies*

Eye tracking is an empirical approach that analyzes translator attention on both source and target text during the actual translation process. Think aloud protocols have been used as another way of studying information processing and decision making during translation (e.g. Kussmaul and Tirkkonen-Condit 1995, Englund Dimitrova 2005), but these are by nature limited to data being verbalized by the subject (Jääskeläinen 1998) and have been found to slow down the translation process and change text processing segmentation (Jakobsen 2003).

Eye tracking has been used to study reading for some time (e.g. Just and Carpenter 1980) and very generally is based on the assumption that “the eye remains fixated on a word as long as the word is being processed” (Just and Carpenter 1980: 330). This idea is in line with the notion that translation difficulties requires more processing time, which was also found in the other empirical studies. Sharmin et al. (2008) found “while reading the source text for translation, translators had more fixations in complex text (whether it was the result of complex structure or difficult vocabulary) than in simple text”, but that fixation durations were “almost identical”. Similarly, “the more complex texts required readers to make more regressions in order to grasp the meaning and produce a translation” (Sharmin et al. 2008: 44-45). In other words, the additional processing time appears to be going into rereading rather than prolonged fixation on a difficult word or structure. In the study, *text complexity* was very generally defined: “Text 2 was judged to be more complex than Text 1 mainly because of the long and complex sentences. Text 3 was assessed as being more difficult mainly because it had more low-frequency items than Text 1”. In addition, possible translation difficulties could be extracted from the “heat maps” showing repeated gaze fixations (Sharmin et al. 2008: 44-45; an idea that has been questioned to a certain degree by Alves et al. 2011).

If one accepts the idea that the number of fixations indicates a reading or translation difficulty, eye tracking, linked with error analysis, could be a useful method of triangulation in identifying testing items. One of the more interesting possibilities is that, in reading, the number of regressions may increase with greater text complexity, suggesting that test items targeting whether a translator has linked elements such as antecedents and their nouns and tense throughout the passage in a coherent manner might be useful measures of the level of text a translator can successfully handle.

However, this can only be done if there is a link between text features and accuracy.

One might also note that the fact that a text is complex and requires cognitive effort to translate correctly is not a bad thing. For many translators, this is actually a pleasure and why they find the profession fulfilling. What we really want to know is not only when text complexity is too difficult for an individual,³ but also when a text is simple enough to be left to machine translation with post-editing (Howard 2011). At the same time, the relationship between cognitive effort and errors might be the opposite of what researchers have been expecting. A certain number of mistakes may be occurring in the *absence* of cognitive effort rather than through cognitive overload. This might be a promising area for eye tracking to investigate.

2.1.4 Difficulty in Translation Studies: summary

Earlier studies of translation problems or difficulties have identified possible elements that could be set as test items. Also, the fact that different methods of analysis have found similar problems suggests that items such as official names/terminology, metaphors (including grammatical metaphors), and complex noun phrases are problems that span languages and are worth piloting for item difficulty. However, in addition to determining whether a textual feature is a likely problem, one also needs to consider whether the actual solving of the problem will provide useful information to, for example, a project manager who has to decide whom among a set of translators to commission.

2.2 Short-passage tests

2.2.1 In the academic setting

While short-passage translation tests are one of the most common types of testing, both in the classroom and professionally, there are essentially no empirical studies on whether such tests are useful or informative, or on what it is they are testing. Slightly more has been written on passage selection. For example, in discussing text selection for classroom use, Nord says that only authentic materials should be used and that the

³ Although when phrased that way, surely the issue of professional ethics becomes part of the discussion. Translators are responsible for turning down assignments they think they cannot complete successfully.

material should be presented “in such a way that as much information as possible is provided on the situation in which the text ‘functioned’”. Another requirement is that the text be presented in its entirety, even if only a portion is to be translated (Nord 2005: 162-163). While Nord restricts materials for intermediate examination tests to translation problems that have been taught, she throws the gates wide open for final examinations, which can be considered somewhat equivalent to professional tests, “since a professional translator should be in a position to cope with any conceivable translation problem” (2005: 178). That last thought, plus the stipulation of authentic material, pretty much make up the basic criteria for selecting test materials by translation agencies and certifying bodies (often minus information on the context of the passage). A primary problem with this approach to passage selection is that it is very difficult to say exactly what has been tested and so to make any predictive statements about future performance.

Kiraly sees the same problem with translation school final exams that are administered under much the same conditions as some certification exams:

Can the translation of a single text without advance preparation, without computer-based tools, without access to reference works or the chance to negotiate with a client really tell us anything credible at all about a translator’s competence? (Kiraly 2000: 156)

I will argue that a single translation may be able to tell us something about translator competence, but only if we have first defined what that something is and if we can reasonably demonstrate that the something is linked to features of the source and target text.

Kelly notes another classroom exam problem that applies to certification exams, although not translation agency tests:

Professional translators rarely find themselves in situations in which they have to translate totally unknown texts in a short period of time without any form of preparation; this is a particularly important criticism if documentary resources are not available in the examination room and students have to rely on those they themselves can take. (Kelly 2005: 137)

Actually, professional translators do sometimes find themselves in precisely that situation. Nonetheless, the objection to testing without appropriate resources available is important not only in terms of fairness, but because the presence or absence of Internet access, for example, is a determinate in the types of items that can be usefully set for scoring.

2.2.2 *In the professional setting*

Critiques of certification exams tend to focus on evaluation methods (Schäffner 1998: 117-133; McAlester 1999; Secară 2005; Angelelli 2009) and rarely mention passage selection. In her critique of the British Institute of Linguists' examination for the Diploma in Translation, Schäffner notes that there is "no clear information about the source text, i.e. no information about the exact time and place of its publication. Often, the candidates do not get the complete text either" (1998: 119). The Institute now posts the source texts of past exams on their web site (these appear to be primarily from news journals such as the *The Economist* and *BBC*), but I could not find any selection guidelines or statement about what precisely is being tested (Chartered Institute of Linguists 2015).

A 2005 review of the Australian National Accreditation Authority for Translators and Interpreters (NAATI) noted that the *Examiners' Manual* provided "comprehensive guidance on the selection of test material" (Cook and Dixon 2005: 6), but the guidelines do not appear to be made public, nor can I find a statement of what is being tested other than that "[t]he finished translation must be accurate, with no margin for error. It must read fluently, and should accurately reflect the register, tone and style of the original. The finished translation is expected to be a fair copy and clearly legible" (NAATI 2015: 11). This suggests that the items being tested are accuracy, register, tone, style, and good handwriting. A recent review of the program did not mention passage selection at all, although it did note that tasks should be "authentic" (Hale et al. 2012: 46).

A very clear description of the American Translators Association (ATA) criteria for passage selection appears in Koby and Champe (2013): the general passage must meet the requirements of Interagency Language Roundtable (ILR) reading level 3 and also "present translation challenges at three levels: textual (i.e., style, register, and cohesion), sentence (grammar/syntax), and word (vocabulary/terminology)" (2013: 160-

161). However, it is not clear how passage characteristics allow the examiners “to evaluate the ability of translators to understand texts written in a source language and to transfer them into a target language text that is usable for a specified purpose in accordance with translation instructions” nor how elements such as “style, register, and cohesion” are scored (Koby and Champe 2013: 160).

Scholars working with the ILR recommend that passages be “as free as possible of the requirements of technical content” (Child 1998: 391) in order to test proficiency that is not influenced by (*supported by*, in essence) subject matter expertise (Lowe 1986: 393).

In her 2012 doctoral dissertation, Lafeber provides a very thorough description of testing practices in inter-governmental organizations (IGOs). In terms of passage selection, she found that practices varied, but that most organization tried to include “texts that contain passages that defy literal translation” (Lafeber 2012b: 149). Texts were taken from both internal and external sources, such as newspapers and specialized journals. Translation instructions, which many scholars think essential to a fair test (e.g. Schäffner 1998; Nord 2005; Koby and Champe 2013), were rarely included. Scoring appeared to be error-based. (Lafeber 2012b: 150-153). Lafeber’s conclusion after analyzing testing practices at two organizations was that “[t]he weighting of skills and knowledge types in current recruitment examinations does not correspond to that of the profile required in the IGO in question” (1012b: 178). In other words, those administering the tests appear to have only a limited idea of how to test for specific features that they would actually like to see in the translators they hire.

Lafeber asked a very basic question in reference to short-passage translation tests: “How effective are the text-based translation tests as instruments of selection?” (2012b: 5) To answer this question, she developed a text-based translation test based on a skills-knowledge profile. The desired skills were discovered via survey. Then, going against conventional practice that counsels against manipulating the source texts for testing (Kelly 2005: 119; Kiraly 2000: 101-116), Lafeber inserted features in a selected source text that would activate these skills “to produce ratable samples from which more valid inferences about abilities might be drawn” (2012b: 199). When the results were compared with those of a standard text-based translation test, it was found that the outcomes were significantly different, which could have an impact on hiring decisions (Lafeber 2012b: 261). This research suggests that it is possible to improve the

usefulness (i.e. the predictive power) of short-passages translation tests by tying text features more closely to the skills that are to be tested.

2.2.3 *The question of authenticity*

As was seen above, the main criticisms of short-passages certification tests have centered on the manner in which the test passage is presented (i.e. out of context and with inadequate translation instructions) and not on passage content. Although I have not been able to find this argument made explicitly, the justification for short-passages tests themselves as well as the passages selected appears to be that they represent an “authentic” task. A 2012 report on improvements to NAATI accreditation came very close to saying that:

Performance tests aim to be as ‘authentic’ as possible, i.e. the tasks and scoring methods aim to replicate real-world assignments. The standard practice in performance test design for translators is one or more translation passages [...], which aim to replicate real-life translations [and] [...] assignments. (Hale et al. 2012: 46)

Many of the stated objections to short-passages tests were that various aspects of the testing situation—no Internet access or computer-aided tools, out of context passages being translated for an unknown purpose and audience, no translation instructions—were not representative of a translator’s normal working environment or the usual translation task⁴ and needed to be changed so that the test task would be more authentic. “Authentic” here appears to mean more like what would happen in an actual working environment. However, in what has become a standard text on language testing, Bachman notes that:

[A]ttempts to characterize either authenticity in general, or the authenticity of a given test have been highly problematic [...] when we consider the great variety that characterizes language use—different contexts, purposes, topics, participants, and so forth—it is not at all clear how we might go about

⁴ Actually, as a professional translator, I frequently receive passages with instructions like “translate this for tomorrow at 9 am” and so consider these elements part of the standard working environment.

distinguishing ‘real-life’ from ‘nonreal-life’ language use in any meaningful way, so that attempts to characterize authenticity in terms of real-life performance are problematic. (Bachman 1990: 10)

At the same time, a major problem in the communicative language ability model is defining language ability—the test construct—in such a way that the test performance will be characteristic of language in in situations outside the test (Bachman 1990: 9). This requires clearly defining both the ability to be tested and authenticity (1990: 10).

I want to pause here and take a closer look at *authenticity*, because I suspect the concept is at the root of some of the problems in short-passage translation passage selection and testing. In a later work, Bachman and Palmer define authenticity as “the degree of correspondence of the characteristics of a given language test task to the features of a TLU [target language use] task” (1996: 23). Authenticity is what

relates the test task to the domain of generalization to which we want our score interpretations to generalize. Authenticity thus provides a means for investigating the extent to which score interpretations generalize beyond performance on the test [...] to other similar nontest language use domains. (Bachman and Palmer 1996: 23-24)

Designing an authentic test task involves, first, identifying “the critical features that define tasks in the TLU domain” and then designing or selecting “sample tasks that have these critical features” (Bachman and Palmer 1996: 24). At the same time, the test setting and content need to be realistic “in terms of expected perceptions on the part of test takers and test users”, with the caveat that if the target language domain “is very broad, or is varied, then it may be realistic to expect only a moderate level of authenticity” (Bachman and Palmer 1996: 136). As can be seen from this summary, the notion of *authenticity*, as used in testing, is far more complicated than simply finding an article of the needed length and writing translation instructions for it.

Also, while some authorities in the language-testing field (e.g. Weir 2005: 56) still advocate making test situations as realistic as possible, others who are not so grounded in Bachman’s communicative language approach have declared that

The only value of any response to any item or task is given to it by the strength of the inference we can make from the ‘response as evidence’ to the construct as *language knowledge*, or an *ability for use* within the specified parameters laid out in a framework and test specifications.

This means that the arguments over *task authenticity* that dominated the late 1970s and 1980s are no longer meaningful for us. (Fulcher and Davidson 2007: 56).

A very pertinent comment about authenticity made by Peter Skehan in 1982 sums up the problem:

A large part of the problem in testing is in sampling a sufficiently wide range of language to be able to generalize to new situations. Merely making an interaction ‘authentic’ does not guarantee that the sampling of language involved will be sufficient, or the basis for wide-ranging and powerful predictions of language behaviour. (Skehan 1984: 208)

I have gone on about the issue of authenticity at some length because I think the notion of short-passage translation as an “authentic task” may have hindered productive discussion about passage selection. For a test using short-passages to be meaningful to the end user—normally the commissioner of future translations—the test designer needs to be clear about exactly what is being tested and how that is evoked by the passage selected.

In summary, some scholars have proposed guidelines for selecting passages in translator training programs, but there has been little or empirical work on the topic. Appropriate passage selection for professional tests has received almost no attention. When it is discussed (e.g. in the description of the ATA exam) it is not matched to statements of what is being tested. The quest for authenticity may have allowed test designers to overlook the need for clear statements about what is being tested and how a selected passage contributes to that.

2.3 Two models of testing: communicative language ability and performance

Because the approach in this study has taken elements from both the communicative language ability and performance models of language testing, these will be briefly described below and major criticisms of each noted. A good starting point is to observe that, at any given point in time, the “what” to be tested in language or translation tests is very much connected to contemporary theories of second-language testing. Such links between language theory and language testing have been reviewed by Skehan 1984, Spolsky 1995, Bachmann 1990, Weir 2005, Fulcher 2010, and Green 2014, among others. The two theoretical approaches that are most often used as the conceptual underpinning of translation tests are communicative language ability, with Bachmann as the primary exemplar, and what Fulcher (2010: 113-118) terms the “performance model” (Green calls it the “proficiency movement” 2012: 43-70), as embodied in the Common European Framework of Reference (CEFR), and the level descriptors of the Interagency Language Roundtable (ILR) and American Council on the Teaching of Foreign Languages (ACTFL).

2.3.1 *The communicative language ability model*

Lyle F. Bachman’s *Fundamental Considerations in Language Testing* appeared in 1990 and has had a major impact on thinking about language and translation testing. The approach set out in that volume is based on the following assumptions:

- First, that a “clear and explicit definition of language ability is essential to all language test development and use”;
- Second, that in order to test “communicative language ability and communicative language use”, tests need to measure a “wide range of language abilities, including grammatical, discourse, sociolinguistic, and strategic competencies”; and
- Third, that tests need to be “authentic” in the sense of requiring “test takers to interact with and process both the explicit linguistic information and the implicit illocutionary or functional meaning of the test material” (Bachman 1990: 3-4).

These points can be summarized as follows:

[O]ne of the most important and persistent problems in language testing is that of defining language ability in such a way that we can be sure that the test methods we use will elicit test performance that is characteristic of language performance in non-test situations. (Bachman 1990: 9)

For Bachman, solving the above problem involves a clear definition of language ability that in turn will allow an equally clear definition of authenticity (1990: 10). Once that is done (and Bachman notes that doing it would require empirical studies of “performance on different types of test tasks”), test designers are then faced with the problem of measurement: “determining the extent to which the sample of language use we obtain from a test adequately characterizes the overall potential use of the individual” (Bachman 1990: 11).

The root assumption in establishing a measurement method is that “virtually every instance of authentic language use involves several abilities” (Bachman 1990: 11), and measurements reflecting ability thus need to go beyond unidimensionality. However, the need to define and separate out different abilities has proven to be an ongoing problem with this approach. Bachman himself notes in his discussion of language competence—which can be broken down into a dozen or more subcompetencies—that “attempts to empirically validate these various components have not been conclusive” (1990: 86) and that “clearly most instances of language use fulfill several functions simultaneously” (1990: 94). He goes on to add that, “it is the connections among these functions [the various components of language competence] that provide coherence to discourse” (ibid), leaving one to wonder what is being tested: the individual components or the ability to make connections between them.

Language Testing in Practice (Bachman and Palmer 1996), as the title suggests, sets out methods for test development based on the ideas in Bachman’s earlier book. The fundamental and very sensible premise is that “[t]he most important consideration in designing and developing a language test is the use for which it is intended, so that the most important quality of a test is its usefulness” (1996: 17). Thus, usefulness, which is made up of “reliability + construct validity + authenticity + interactiveness + impact + practicality”, is always the quality to be maximized. Doing this requires different weightings of the constituent components depending on the test purpose and

setting. While a full discussion of these elements is beyond the scope of this study, very brief definitions taken from Bachman and Palmer (1996: 19-38) are given below:

- *Reliability*: consistency of measurement in the sense of consistency of scoring between one test and set of test tasks to another (1996: 19-21)
- *Construct validity*: “the extent to which we can interpret a given test score as an indicator of the ability(ies), or construct(s) we wish to measure”, with a *construct* being defined as “the specific definition of an ability that provides the basis for a given test or test task and for interpreting scores derived from this task” (1996: 21-23)
- *Authenticity*: the extent to which characteristics of a language test task correspond to the features of a target language use task and so allow test scores to be generalized beyond the test setting (1996: 23-25)
- *Interactiveness*: “the ways in which the test taker’s areas of language knowledge, metacognitive strategies, topical knowledge, and affective schemata are engaged by the test task” (1996: 25-29)
- *Impact*: the effect a test will have on all stakeholders involved, including any washback on individuals and educational or other organizations and systems (1996: 29- 35)
- *Practicality*: “the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities” (1996: 35-37)

The components most subject to being given different weightings are authenticity and interactiveness, which are both “relative” and are to be adjusted in accordance with the test purpose and situation. For example, “[c]ertain test tasks may be relatively useful for their intended purposes, even though they are low in either authenticity or interactiveness” (Bachman and Palmer 1996: 29). Authenticity and construct validity are the most subject to debate, with the interpretation of both differing from the ILR approach discussed below. But despite differences in emphasis and definition, these elements of test usefulness provide a sound and ethical approach to test design.

Bachman and Palmer describe several possible methods of scoring, and although they show a preference for criterion-referenced ability-based scales and some reservations about global scales of ability, they set out considerations to be followed when applying most common scoring systems. The elements that dictate the scoring method to be used are “the theoretical definition of the construct to be measured” and the “test task specifications”. In other words, the “way we define the construct for a particular testing situation will determine which areas of language we need to score” and “the type of score to be reported” (1996: 194). Meanwhile, the specification of the test task determines the “type of intended response” and so suggests the scoring method to be used (Bachman and Palmer 1996: 194). Practicality is also fundamental when considering scoring methods, as are measures of reliability and the need to pilot any test before actual implementation.

As with other elements of test design in this approach, the essential element is the “theoretical definition of the construct to be measured” (Bachman 1996: 194). Translation tests in the professional realm often fail to define either what is being tested or the skill or skills that scoring reflects.⁵ Therefore, reminders that tests are actually intended to measure something and that one should know what that something is, particularly when interpreting scores, are extremely important.

The use of the word *theoretical*, as well as Bachmann’s insistence that language tests must be based on a “general theory of language ability” (1990: 9), present a potential problem with the communicative language ability approach, depending on how “theory” is being used. At times, it appears to correspond to the dictionary definition of “a system of ideas intended to explain something”, while other passages suggest theory in the sense of a scientific theory to be validated by empirical evidence. For example, in the chapter on communicative language ability (1990: 81-110), Bachman describes this ability, but with the caveat of not presuming to present the framework “as a complete theory of language abilities” (1990: 81). He then proposes three components of communicative language ability: “language competence, strategic competence, and psychophysiological mechanisms” (1990: 84), all of which can be further subdivided into subcompetencies. However, for language competency at least, Bachman notes that “[a]ttempts to empirically validate these various components have not been conclusive” (1990: 86). Alderson, in discussing reader skills and abilities,

⁵ Error marking is, by definition, aimed at identifying deficiencies.

notes that attempts to isolate the different skills thought to be necessary for reading are controversial and that

[t]here is contradictory evidence as to whether these ‘skills’ are separately identifiable. Different analyses of the same databases of skill have resulted in more or fewer factors that appear to underlie adequate understanding. [...] When judges are asked to identify which skills the [test] items are measuring, they are often unable to do so with any convincing degree of agreement. It is frequently difficult to get expert judges to agree on what skills are operationalized by which test item. (Alderson 2000: 49)

This inability to separate out individual skills while simultaneously insisting on the importance of identifying competencies and subcompetencies is a fundamental problem in any attempt to operationalize Bachman’s model in an actual test.

In Bachman and Palmer’s follow-up volume, *Language Testing in Practice*, strategic competence has become “the component that links other components within the individual, as well as providing the cognitive link with the characteristics of the language use task and setting” (1996: 62). But at the same time, they note that “we conceive of this not as a working model of language processing, but rather as a conceptual basis for organizing our thinking about the test development process” (1996: 62).

This suggests that Bachman’s statement about the necessity of a “general theory of language ability” and Bachman and Palmer’s claim to be presenting “a theoretical model of language ability” are not using the term “theory” in Karl Popper’s sense of a proposition that can be falsified (Skehan 1984: 209 made this point earlier about the microskills approach in language learning); in other words, there is no null hypothesis. As a result, saying that the Bachman and Palmer competency approach is “theory-based” is problematic when viewed in terms of Popper’s definition of a theory. Rather than proposing a theory, what they have done instead is to set out their assumptions about how language ability works which is, of course, a useful thing to know, particularly in syllabus construction. However, they have not demonstrated how such abilities actually function or how they can be isolated and identified.

The point about the non-theoretical nature of Bachman’s argument is further underscored by his choice of language, e.g. “I believe that correctly answering such

questions requires strategic competence” (1990: 105), “I consider [strategic competence] more as a general ability” (1990: 106), and “We believe very strongly that the consideration of language ability in its totality needs to inform the development and use of any language test” (Bachman and Palmer 1996: 67, my emphasis added in the three quotations).

This observation about theory is somewhat refuted in the chapter on validation (Bachman 1990: 236-295), in which Bachman writes that “in order to examine validity, we need a theory that specifies the language abilities that we hypothesize will affect test performance” (1990: 239). He further notes that “constructs can be viewed as definitions of abilities that permit us to state specific hypotheses about how these abilities are or are not related to other abilities, and about the relationship between these abilities and observed behavior” (1990: 255) and also that “[c]onstruct validation can thus be seen as a special case of verifying, or falsifying, a scientific theory” (Bachman 1990: 256).

However, what is being hypothesized, and in some way tested, is the relationship between constructs—defined as measures of language ability—and test scores, which can “be viewed as behavioral manifestations of the constructs” (Bachman 1990: 257). This would seem to omit the step of subjecting the “definitions of abilities” to some sort of validation that would show, at the least, the discrete existence of the ability in question. The limited number of empirical studies Bachman cites appears to demonstrate that certain outcomes are connected to specific pedagogical interventions more than to the existence of hypothesized language abilities (1990: 268). The Bachman model has been characterized “as overly cognitive, being concerned with the individual mental processes rather than social and interactive aspects of language use” and as offering “no clear guidance on why one language use task may be more or less demanding than another” (Green 2014: 206).

This last objection is particularly important when one objective of a test is to identify more and less demanding elements of language use, or conversely, of candidate skill. Despite its perceived flaws, the Bachman model has been the impetus for aligning several of the high-stake standardized tests with a more communicative approach (Green 2014: 202-204). It also satisfies the requirement of theory-based validity to have a “theory of the cognitive processes” involved in the skill being tested (Weir 2005: 18). The model also contributes to construct definition and validation in that it sets out “the abilities of the learner that we believe underlie their test performance, but which we

cannot directly observe” (Fulcher 2010: 96). In short, the presence of a theory of abilities is, for some, the necessary starting point in test design. However, test designers creating tests based on hypothesized abilities would seem to have some responsibility for then demonstrating that they are indeed testing the abilities that have been set out. Bachman writes that the key to solving the problem of “defining language ability in such a way that we can be sure that the test methods we use will elicit language test performance that is characteristic of language performance in non-test situations” is to specify the “characteristics of test tasks and test methods sufficiently well that we can begin to empirically examine test takers’ performance on different types of test tasks” (1990: 9, 10).

2.3.2 Proficiency tests and the performance model

Many of the high-stakes foreign language tests currently administered are proficiency tests, including the ILR-based tests used by the U.S. government, the American Council on The Teaching of Foreign Languages (ACTFL) proficiency examinations, and the Common European Framework of Reference (CEFR) tests given in Europe. Historically, the ACTFL guidelines and the CEFR descriptions have been influenced by the ILR Language Skill Level Descriptions (LSLDs), and the ACTFL guidelines were derived from the ILR descriptions and are meant to be consistent with them (Lowe 1986; personal communication November 2015). The CEFR descriptions were influenced by the ILR, but “were purposely developed for curricular design and are more open-ended than the ACTFL” guidelines (Lowe, personal communication November 2015). *Proficiency* in the ILR/ACTFL system has been defined as achievement (i.e. “mastery of ILR functions, content and accuracy when these are memorized but not used creatively in a consistent and sustained manner”) plus functional evidence (i.e. “obvious, usable skills” that are “proven by use, not merely implied”) of “internalized strategies for creativity” (i.e. “the ability to generate sentences and discourse that are new to the examinee”) given a single global rating (i.e. a single score) based on the ILR levels (Lowe 1988: 11-51). Two distinctions are important here. First, there is a line is drawn between *performance*—“whether language is *used*”—and *proficiency*—“whether language is *used creatively*” (Lowe 1988: 14, emphasis in the original, with *creativity* defined as above), with a corresponding distinction between made “performance tests” and “proficiency tests” This will be

discussed further in 2.4.1. Second, proficiency tests are not achievement tests. The latter are meant to measure how well students or others are reaching set objectives, while proficiency tests target what a person can *do* with language based on a predetermined specification (Hughes 2003: 11-13).

Lowe, one of the explicators of the ILR system, notes that the division into ILR functions, content, and accuracy may make proficiency assessment incompatible with Bachman's communicative language ability approach. This is because the position of accuracy in the latter is not clear and because communicative language ability may be "an atomistic, bottom-up view of language ability rather than a holistic, top-down view like that of AEI⁶ (Bachman, personal communication)" (Lowe 1988: 14-15).

The definition of what is being tested and the scoring system are strongly linked in proficiency tests as the descriptor says what a candidate "can do". The description defines the performance expected of a candidate, the task, and some of the evaluation criteria.

Proficiency tests would appear to have strong links between the three elements required for validation—theory, context, and scoring—in that:

- the skill being tested is defined (although without a basis in a general theory of language);
- the skill defined is reflected in the criteria for scoring; and
- the test tasks are generally reflective of the larger language domain.

However, scholars in the communicative language ability camp have objected to such tests for the following reasons, among others:

- they lack "theories of competence, knowledge, communication and skill underlying performance" (Shohamy 1996: 147)
- they are "task-based and not driven by a sound theory of communicative language" (ibid) and so rely on "purely operational definitions of what it means to know and use a language" (Fulcher 2010: 113)
- they are based on a single performance (ibid.)

⁶ "AEI" refers to features shared by the ILR and ACTFL scales (Lowe 1988: 19).

- the scales used to describe levels “have no basis in analysis of actual performance, or second language acquisition theory” (Fulcher 2010: 114)
- the “levels are flawed because they are not based on primary data” (Fulcher 2010:116).

On the other hand, in reference to the ILR levels, Skehan writes that “painstaking empirical work has been carried out to clarify the dimensions of language used [in the level descriptors] and to validate the effectiveness of such rating scales with real candidates” (1984: 217). Alderson (2007: 660-661), while defending the CEFR, agrees that more empirical research is needed and that there is “no independent estimate of the learners’ proficiency levels according to the CEFR” or on “how proficiency in the main European languages [...] develops over time”. Others involved with the CEFR illustrative scales have also noted that these are underspecified and uncalibrated (Martyniuk 2010: xiii)—objections that could also be applied to the ILR scales. However, while the above are all issues requiring further research, those developing and using proficiency tests appear to be aware of the shortcoming of their systems and are attempting to remedy them. Moreover, the problems with proficiency tests do not appear to be as unsolvable as the inability to separate out and operationalize competencies and subcompetencies in the communicative language approach. In addition, many of the objections to these tests appear to be virtues (pragmatic, performance-based, well-defined, single score) when viewed in a professional translation test setting. But at the same time, Shohamy (1996: 145-147) has observed that the pragmatic appeal of performance, coupled with the appeal of face validity, has frequently excused test designers from sorting out exactly what it is that is being tested.

2.3.3 Summary

It is not a goal of this study to evaluate theories of testing. However, some idea of the current approaches to test design is helpful when discussing views of short-passage translation tests, and particularly when focusing on *what* is being tested and *why*, as well as the ultimate *use* of the test. Without taking a stand in favor of either model, I have taken elements from each. The need to define what is being tested and how this is reflected in the score comes from the communicative language ability model. However, even though it would be possible to say that members of the test population do or do not

have the ability to produce a cohesive target language text, I am not trying to connect that to a general theory of translation competencies. Another feature I have taken from those doing research in translation skills and competencies is the idea of identifying source text problem points corresponding to target text solutions and the notion that the solutions point to the existence of another trait of interest. That trait is a competency or subcompetency for the PACTE group. In this study there are two traits of interest. The first is whether verb tense and modified nouns can be used as markers of the use of inference to follow an argument and construct a cohesive target text. The second is whether the problem points are indeed translation difficulties for the test population.

I have used the ILR reading level descriptions to identify the source text element—following an argument that requires a degree of inference—which the linked source and target text features are meant to elucidate. I will also argue that the level 3 descriptors point to textual features that make texts at this level more difficult to translate than level 2/2+ texts. But on a more general level (and assessment issues aside), short-passage translation tests fall more naturally under the umbrella of the performance model. They are a single performance that results in a score on which a decision is often made (e.g. to certify or employ), with or without feedback given to the candidate being tested. However, with careful passage selection, such tests could be used to measure proficiency. The ILR rating in its original government setting was meant to be predictive; the idea was that the sample of language rated would reveal “consistent and sustained” ability (Lowe 1986: 28) Predictive power of the absence of a desirable trait (i.e. prediction that a given translator would be likely to put future projects at risk), is easier than guaranteeing the presence of an ability (i.e. being able to say that a translator would nearly always be successful at assignments). But even this lesser power of predication is a feature that would be highly desirable in the professional test setting.

2.4 ILR reading and translation levels

2.4.1 Description of the ILR levels

Because the passages in this study were evaluated using the U.S. Government’s Interagency Language Skill Level Descriptions (LSLDs), the system will be briefly described here. The most up-to-date wording of the descriptors with a short history of

the development of the descriptions is available on the ILR website (ILR 2015) and will be the wording used in this study.

The original four LSLDs (Speaking, Listening, Reading, and Writing) were devised at the Foreign Service Institute at the U.S. Department of State in the mid 1950s and were later adopted by the U.S. Government's Interagency Language Roundtable (ILR) and renamed the *ILR Language Skill Level Descriptions*. These four original LSLDs are *proficiency* scales and differ from the eight descriptors that followed (Translation Performance, Audio-Translation Performance, Interpretation Performance, and Competence in Intercultural Communication) which are *performance* scales and are specifically designated as such (Child et al. 1993: 19-54). *Performance*⁷ skills relate to abilities that target a specific topic, area, etc., are limited to a given field or subject, and are not transferable to other areas; for example, an individual may be able to read articles on mathematics at a high level, but not on political science. *Proficiency*, on the other hand, is a set of skills that target general language, are not specific to topic, and are transferable across subjects. In other words, a person who can read one general topic article at ILR reading level 3 is expected to be able to read other articles at the same level (Child et al. 1993: 19-54).

Although the skills differ, all eight ILR LSLDs have six “base levels” and five “plus levels,” resulting in a total of 11 ranges: 0, 0+, 1, 1+, 2, 2+, 3, 3+, 4, 4+ and 5. The ILR levels of interest in this study are reading levels 2+ and 3. While there are two possible LSLDs that could be used for selecting texts—the *ILR LSLDs for Reading* and the *ILR LSLDs for Translation Performance*—the descriptors for reading current between 2004 and 2008 were used for passage selection by the ATA Japanese-to-English grading group and the 2015 version for the actual text evaluation in this study.

The ILR descriptions of reading level 3 and reading level 2+, the two levels of interest here, are given in Table 2.1.

⁷ Within the U.S. Government language community the term *performance* corresponds most closely to academia's use of the phrase *language for special purposes*.

Table 2.1. ILR Reading Levels 3 and 2+*

ILR Reading Level 3 (General Professional Proficiency)	ILR Reading Level 2+ (Limited Working Proficiency, Plus)**
Able to read within a normal range of speed and with almost complete comprehension a variety of authentic prose material on unfamiliar subjects. Reading ability is not dependent on subject matter knowledge, although it is not expected that the individual can comprehend thoroughly subject matter which is highly dependent on cultural knowledge or which is outside his/her general experience and not accompanied by explanation. Text-types include news stories similar to wire service reports or international news items in major periodicals, routine correspondence, general reports, and technical material in his/her professional field; all of these may include hypothesis, argumentation and supported opinions. Misreading rare. Almost always able to interpret material correctly, relate ideas and “read between the lines,” (that is, understand the writers’ implicit intents in text of the above types). Can get the gist of more sophisticated texts, but may be unable to detect or understand subtlety and nuance. Rarely has to pause over or reread general vocabulary. However, may experience some difficulty with unusually complex structure and low frequency idioms.	Sufficient comprehension to understand most factual material in non-technical prose as well as some discussions on concrete topics related to special professional interests. Is markedly more proficient at reading materials on a familiar topic. Is able to separate the main ideas and details from lesser ones and uses that distinction to advance understanding. The individual is able to use linguistic context and real-world knowledge to make sensible guesses about unfamiliar material. Has a broad active reading vocabulary. The individual is able to get the gist of main and subsidiary ideas in texts which could only be read thoroughly by persons with much higher proficiencies. Weaknesses include slowness, uncertainty, inability to discern nuance and/or intentionally disguised meaning.

* Material quoted from the ILR Web site (ILR 2015)

** “The ‘plus level’ descriptions are therefore supplementary to the “base level” descriptions. A skill level is assigned to a person through an authorized language examination.” (ILR 2015)

The corresponding translation performance descriptions are given in Table 2.2.

Table 2.2 ILR Translation Performance

ILR Translation Performance Level 3 (Professional Performance)	ILR Reading Level 2+ (Limited Performance, Plus)**
Can translate texts that contain not only facts but also abstract language, showing an emerging ability to capture their intended implications and many nuances. Such texts usually contain situations and events which are subject to value judgments of a personal or institutional kind, as in some newspaper editorials, propaganda tracts, and evaluations of projects. Linguistic knowledge of both the terminology and the means of expression specific to a subject field is strong enough to allow the translator to operate successfully within that field. Word choice and expression generally adhere to target language norms and rarely obscure meaning. The resulting product is a draft translation, subject to quality control.	Can render straightforward texts dealing with everyday matters that include statements of fact as well as some judgments, opinion, or other elements which entail more than direct exposition, but do not contain figurative language, complicated concepts, complex sentence structures, or instances of syntactic or semantic skewing. In these types of texts, the individual can read source language materials and render them accurately into the target language, conveying the key points and/or main ideas, supporting facts, most of the details, and some nuances. Can usually operate in more than one narrowly defined subject field, using both linguistic knowledge of the languages involved and familiarity with the subject matter. A tendency to adhere to source language structures may result in target language expressions that may appear to be correct but are awkward or perhaps unidiomatic. Such expressions may sometimes obscure meaning. The resulting product is not a professional translation and must be subject to quality control.

* Material quoted from the ILR Web site (ILR 2015)

** “The ‘plus level’ descriptions are therefore supplementary to the “base level” descriptions. A skill level is assigned to a person through an authorized language examination.” (ILR 2015)

Description of the reading levels has been further refined through the addition of text typologies (Child 1981) that can be linked to (but do not have an exact correspondence with) the reading levels: orientation mode (level 1, with a one-to-one

correspondence of language and content), instructive mode (level 2, straightforward information about the real world), evaluative mode (level 3, analysis and evaluation against “a backdrop of shared information” [Child 1998: 396]), and projective mode (level 4, extensive author input and shaping).⁸

The modes of interest for this study are the instructive and the evaluative. The first is concerned with “the transmission of factual material” in a relatively uncomplicated form. These texts “are essentially factual and hence not overly demanding in terms of culture or ideation” (Child 1998: 389). In contrast, the evaluative mode is described as one in which:

The emphasis on the transmission of facts as such is shifted to a perspective in which facts are selected and pressed into service in order to develop points of view; explain or apologize for personal conduct; state and defend past or projected policies; and for many other purposes. (Child 1998: 385-386)

As a result, the existence of an author with an individual style and/or viewpoint is more apparent in evaluative mode texts and the information is more tightly ordered and linked together (Cascallar et al. 1995: 304).

An additional feature of the ILR system is the idea of “congruity judgment” or the “ability unique to translation” and defined as “the ability to successfully match donor language features, characteristics, or forms to their most suitable receptor language equivalents” (Cascallar et al. 1995: 299). Rather simply put, the argument is that it is possible to isolate congruity judgment by first testing a person’s source language skills and then examining the individual’s translation of a text at the same reading level. In this study, the term will be used to signal a successful translation solution and not to refer to the notion of an identified skill unique to translation.

2.4.2 Published work on the ILR system

Published work on the system has primarily been by researchers connected with the United States government, e.g. Lowe 1986, Child 1987, Child et al. 1993, Child 1998,

⁸ Superficially, these modes resemble the text types described by Reiss (2000) and the text functions set out by Nord (1997), but one should be careful about comparisons because the initial aim of a system to describe reading proficiency is obviously different from typologies created for translation or descriptive pragmatics.

and Cascallar et al. 1995. The references given in the articles suggest that these scholars were working completely independently of the field of Translation Studies. In his history of the development of the ILR/ACFTL scale, Lowe notes that most of the work done on the ILR scale was done “in-house” and not published outside the government (Lowe 1986: 16). Lafeber noted a similar dearth of material related to translation testing by and concerning inter-government organizations (Lafeber 2012a: 110).

More recently, Clifford et al. (2004) used ILR reading levels to rate source-text difficulty in a study of the “effect of text difficulty on machine translation performance” across Spanish, Farsi, Arabic, Russian, and Korean. Jones et al. (2005) used machine and human translations from Arabic source texts at ILR reading levels 1, 2, and 3 to measure translation quality by testing reader comprehension. A 2007 follow-up study by Jones et al. showed a drop in comprehension accuracy for human translation with increasing ILR reading level and a large drop (from 82% for level 2 to 51% for level 3) between levels 2 and 3 for machine translation. This provides some level of empirical evidence for increasing difficulty in reading comprehension as the ILR level increases and that there is a difference between level 2 texts and level 3 texts that makes the latter more difficult to translate. This was further supported by a study of Russian oral proficiency rather than reading, which identified “linguistic features implicated in progress on the ILR scale—specifically from ILR 2 to 2+, and from 2+ to 3” (Long et al. 2012: 105) and demonstrated that tasks based on these features showed significant differences between the levels (Long et al. 2012).

3. Aims and assumptions

3.1 Aims

3.1.1 Rationale

The objective of this study is to explore specific test passage features that can be scored and then evaluated for difficulty using facility values. To do this passages that have actually been used in a professional test setting—the ATA certification examination—will be examined in order to determine the text elements that make a passage difficult for a defined population of test takers (here, people who self-identified as professional translators and met the ATA criteria for sitting the examination). The rationale for doing this is threefold. First, a general problem in testing is identifying precisely what is being tested and how it is to be measured (Angelelli 2009: 24-30; Bachman and Palmer 1996: 21; Weir 2005; Lefebvre 2012b, among others). Pinpointing text features that have caused difficulty (here signaled by facility value) could allow more focused test design, thus avoiding what Weir (2005: 18) calls the “suck-it-and-see approach” to testing. In addition, knowing what is difficult in specific types of texts would allow test designers to identify discrete text elements (e.g. cohesion markers, tense) linked to passage-level structures, rather than word-level structures, and permit these to be evaluated. Second, various empirical studies of translation difficulty have been done (e.g. Campbell and Hale 1999; Hale and Campbell 2002; Jensen 2009; Sun & Shreve 2014) and have identified what can be called “problem points” that specific populations have found difficult in a variety of language pairs.⁹ At the same time, systems such as the ILR reading and translation levels have set out elements thought to be features of progressively greater text complexity and difficulty. However, supposed difficulties and reading levels have only been linked to test passage selection in very general ways, if at all. One purpose of this study is to look at a sampling of such features and problem points to determine whether they can be found in passages used in a professional examination and whether they were difficult for the specified test population.

⁹ The PACTE group, while not addressing translation difficulty, has also identified “rich points” in source texts: PACTE 2005, PACTE 2008.

The third rationale is to provide some preliminary suggestions for evaluation items that operate on the passage, rather than the word, level. There are two elements to this. First, the product—the translation— is being evaluated and not an individual or any supposed competencies that a person might have. Because the focus here is on translation in the professional setting, the standpoint is that of the translation commissioner or translation agency, and so the translation—and whether it is essentially usable—is the main object of interest. At the same time, test sample evaluation forms for translation agencies often include some form of the question, “Would you recommend this translator for other assignments?” What is being asked, in essence, is if we assign work to this person, are we putting future projects at risk of needing extensive editing or retranslation (both of which are expensive)? As a result, the evaluator is asked to make a prediction about future performance, often based on a single sample. However, the sample frequently has been chosen because “it’s the type of material we see” (i.e. it is “authentic”) and may not contain elements that signal the level at which the translator can or cannot work.

The situation is further compounded by scoring systems that focus on word-level errors such as spelling, word choice (the rating of which can be very subjective), grammar, and punctuation (what Angelelli terms “micro-linguistic elements” [2009: 29]; see also Schäffner 1998) and often do not have a category for scoring cohesion and modification mistakes beyond “mistranslation” or perhaps “syntax”. While micro-linguistic features are important and a misspelling in a title, for example, can make a document unusable, and while such mistakes may be symptomatic of poor proofreading skills and general carelessness, they may not provide a sound basis for answering a query about whether a given translator would put future projects at risk in terms of being able to determine the type and level of texts a specific translator might be able to handle. Therefore, an idea under consideration here is that the examiner is on far more solid ground in assessing (predicting, actually) whether a translator would be a risky choice for future assignments when the sample text offers a chance to evaluate first, overall cohesion, which suggests the level at which the translation was done (word, sentence, or paragraph), and second, syntax as represented by long strings of modification. Whether successfully conveying the meaning of a source language sentence in the target language was the result of source language reading skill, target language writing ability, or congruity judgment—all of which can be further broken down into multiple subcompetencies—or some combination of the three does not matter

in this context; the only judgment required is whether the meaning of the source language was successfully conveyed in a manner that follows target language conventions and is in line with any translation instructions that may have been provided.

I am not arguing that a translation that is cohesive and has good syntax is inevitably acceptable, as it well may not be, but rather that a translation that does not have these elements has a very high probability of being unacceptable. There is also an assumption here that if a translator is having trouble with elements related to cohesion, there are very likely problems with syntactical and grammatical features at lower text levels.¹⁰

3.1.2 Aims and objectives

More specifically, the aims of this study are

1. to identify specific passage features linked to items of interest (e.g. the production of a translation with overall cohesion) that can be objectively scored and to evaluate these for difficulty by means of facility values;
2. to examine whether elements characteristic of ILR reading level 3 passages are difficult, as measured by item facility, for the test population;
3. to examine whether these items of presumed reading difficulty (e.g. features pertaining to paragraph/text level cohesion) may also be items of translation difficulty; and
4. to identify *problem points* in the examination passages to explore a method of triangulating the data from text and cohesion markers as a determination of overall translation difficulty.

These aims are linked to the larger objectives of being able to identify what *specifically* is being tested on short-passage tests and then linking the elements of interest to specific items that can be scored. The goal is to identify at least the reading level (which may also be the translation level, although that needs further verification) needed to perform successfully on a short-passage translation test comprising the set

¹⁰ This assumption will not be examined in the present study.

features (i.e. text cohesion and correct modification) and to provide a means of comparing and calibrating test passages.

Identifying specific elements to be tested would also be an aid in passage selection because scorable items of possible difficulty may not be present in texts that otherwise meet the selected reading level criteria, thus making the passage a poor candidate for test use. Additionally, the lack of such points of difficulty may also flag a passage as not being at the reading level it was presumed to be.

An area of particular interest in providing material for the validation of earlier studies is looking at the difference in difficulty between ILR reading levels 2+ and 3. A criticism of the ILR scale is that it is not “empirically based” and that the levels are based on functional criteria rather than linguistic criteria (Lowe 1986; Child 1998; Long et al. 2012: 101-102). In the same article, Long et al. note that, “it would be very useful to know which classes of problems occur across languages at these levels, and which tend to be particular typologically defined groups of languages” (2012: 107). While the present study can make only a miniscule contribution to providing such data, it will look at passage elements thought to correspond to ILR reading level 3 descriptors, particularly in relation to following an argument, to determine whether these were factors that were difficult for the test population and also consider whether some features are specific to or at least typical of the Japanese-to-English language pair.

Two secondary aims of the study are to look at items found to be translation difficulties in earlier studies (e.g. passive constructions and complex noun phrases) to see whether the earlier results can be replicated and to provide a small, but hopefully useful, data set in the form of the analyzed Japanese-to-English verb errors, which may be useful pedagogically in teaching both translation and Japanese.

It is not the aim of this study to critique existing translation testing or scoring systems or to propose anything more than elements that could be incorporated into a method of passage selection and evaluation. Also, the study focuses on the translation product, not the process, and so no attempt will be made to connect passage difficulties to translator competencies or cognitive state.

3.2 Determining test items

3.2.1 Why identifying test items is important

Determining the elements that make a specific passage difficult is important first, because it can be linked to identifying what is being tested (e.g. the production of a coherent target language text) and second, because having identified test items is part of being able to compare and calibrate passages used for examinations. Hale and Campbell wrote in 2002 that, “no reliable guidelines exist to help determine what makes a text difficult to translate”. A 2005 review of the Australian National Accreditation Authority for Translators and Interpreters (NAATI) noted that academics in the field of translation and interpretation had emphasized the “need to develop consistency in the level of difficulty of exam tests” (Cook & Dixon 2005: 26). The Standard Practice for Assessing Language Proficiency (ASTM F2889-11) mentions the need for calibration to a “common difficulty scale” (ASTM 2011: 1). One response to such requirements for setting passage difficulty has been to use passages that match descriptors of grade level or other readability indices or descriptors of foreign language reading ability, such as the CEFR or ILR reading levels. For example, the U.S. government has set ILR level 3 in all language skills, including translation, as the “professional” level and uses passages at reading level 3 for translation examinations (ILR 2015). The ATA also uses ILR level 3 passages for the mandatory general passage and level 2+/3 for the technical passages in its certification examination (Koby and Champe 2013: 161). However, other than matching the reading level (which generally means that the passage will require following an argument requiring some degree of inference and reading between the lines), there is no requirement that texts include specific elements or that the scoring system specifically score text features identified as causing a passage to be at ILR reading level 3.

An additional problem is that there has been only a limited amount of empirical research done specifically on translation difficulties (most notably by Campbell and Hale and more recently in eye-tracking experiments, e.g. Mishra et al. 2014), although several researchers have proposed textual features presumed to be “problems” (e.g.

Vinay & Darbelnet 1995: 3; Nord 2005: 168-171; PACTE 2005: 116-118). In Translation Studies there has been some reluctance to address the issue of difficulty on the grounds that difficulties “are subjective and have to do with the individual translator and the specific working conditions” (Nord 2005: 166-167). The presence of subjectivity in research on difficulty is unavoidable and can be expressed simply as the fact that the interaction between a translator and a text will always be idiosyncratic and that the working conditions are an added complication. However, by focusing on a defined test population instead of individuals, by standardizing the translation environment to the extent possible, and by identifying textual elements that can be scored dichotomously, it is possible to arrive at a numerical measure of difficulty (a facility value) that allows test designers to confirm whether a passage is at an appropriate level of difficulty, not in any absolute terms, but for the defined test population and to compare the difficulty of passages for greater consistency in passage selection.

Perhaps the major advantage of setting specific problems and linking these with both a score (acceptable/unacceptable) and a facility value is that it forces the test designer to decide on what precisely is being tested and whether this can be linked to a specific text feature. Anthony Green’s metaphor of assessment as a map is one way of describing the selection process:

As tools, language assessments are closer in nature to maps than to tape measures. [...] all tape measures are designed to measure length. In contrast, maps represent and highlight different aspects of a multidimensional world according to the needs of the user. Like maps, language assessments do not and should not all measure and represent the same things. The choice of assessment techniques should reflect the kinds of decisions that users wish to make and the kinds of consequences that they hope will follow. (Green 2014: 93)

In other words, test designers can and should decide what they are testing and how it will be measured and also consider how the results will be used.

3.2.2 Test items to be studied

The main items to be measured in this study are markers of text cohesion (tense and implicit and explicit cohesion signals) because of the assumption that these provide information about whether a translator appears to be working at the paragraph/passage level and so is able to read a source text with a running argument or hypothesis and create a cohesive target text. This presumption of some degree of predictive utility is supported by the 2003 U.S. National Center for Education Statistics (NCES) survey of adult literacy, which found that only 25% of the graduates of 4-year colleges and universities and 31% of adults with a graduate degree or some graduate courses achieved “proficiency” in prose literacy (Kutner et al. 2007: 38). Examples of tasks associated with the proficient level are “comparing the viewpoints in two editorials” and inferring “the purpose of an event described in a magazine” (Kutner et al. 2007: 4-5). The tasks at the proficient level are similar enough to ILR reading level 3, which requires inference and is typified by editorials from a newspaper such as the *New York Times* or *Washington Post* (and not *U.S.A. Today*), to suggest that using testing materials at level 3 could be helpful for those selecting freelance translators because, in the absence of some means of determining the level of texts being assigned,¹¹ knowing whether a given translator can handle a range of texts that might include documents presenting arguments or requiring inference would be useful information. That 75% of college graduates and 69% of people with some graduate study cannot read proficiently at ILR level 3 in their native language is also useful information for test design because it suggests that a failure rate of 70% to 75% is appropriate for examinations using level 3 passages unless a reading test at the same level has been previously administered.

Reading at the proficient level may—and this is speculative—also be indicative of the ability to do “careful” reading, which Green defines as nonselective reading in which “the reader tries to take in all or most of the information in the text” (translators should aim for “all”), understanding each part of the text, and appreciating “how all of the parts fit together” (2014: 99-100). Evaluating translation performance via cohesion markers does give a sense of whether the translator has “fit the parts together” and “taken in all ... of the information in the text”.

¹¹ The ASTM standard on translation quality makes determination of source text difficulty the responsibility of the requestor (ASTM 2006: 8)

As noted above, in the United States the ILR reading levels are used in the selection of test passages for the ATA, the main entity certifying translators in the country, and for positions within the government. Level 3, the first level at which inference and possible reading between the lines is required, is considered the level of “general professional proficiency” in reading and “professional performance” for translation (ILR 2015).

3.2.2.1 ILR reading levels

The ILR descriptors (described in 2.4.1) were developed to measure language proficiency; however, several U.S. government agencies demand that entry-level candidates for positions involving translation pass a short-passage translation test based on a reading level 2 (minimum working proficiency) passage and later an additional test with a passage at ILR reading levels 3 (general professional proficiency). Thus, despite the system having originally been designed to measure language proficiency, it has been used for a number of years to measure language performance (Child et al. 1993; Child 1998).

An aim of this study is to examine whether identified source text features characteristic of ILR reading level 3 can be mapped to objective errors in the target text that can then be scored and used to compute item facility. An associated goal is a preliminary look at whether and how translation difficulty differs between ILR reading levels 3 and 2/2+. Level 2/2+ texts tend to be descriptive and narrative material “predominantly in straightforward/high-frequency sentence patterns” for which the reader “is able to use linguistic context and real-world knowledge” to aid understanding. At the same time, even at level 2+, there is an “inability to discern nuance”. In contrast, at level 3 the texts “may include hypothesis, argumentation and supported opinions”, and the reader can “interpret material correctly, relate ideas and ‘read between the lines’” (ILR 2015). Table 3.1 summarizes the differences between the two levels.

Table 3.1. Differences between ILR reading levels 3 and 2+*

Features of ILR Reading Level 3	Features of ILR Reading Level 2+
Text types: news stories or international news items in major periodicals	Text types: Most factual material in non-technical prose
Text features: Hypothesis, argumentation, supported opinion; may contain “writer’s implicit intent”	Text features: Main ideas supported by subsidiary items, no nuance or disguised meaning
Text mode: Evaluative–analysis and evaluation	Text mode: Instructive–transmission of factual material
Text linkage: Tightly structured	Text linkage: Loose ordering of facts or main idea supported by fact

* Table based on ILR 2015 and Child 1998. Quoted material from ILR 2015.

In summary, the purpose here is to use ILR reading level 3 features—in particular, the need to use inference to cohesively connect parts of the text—to identify items of translation difficulty at the text level. At the same time, I hope to provide, on a very small scale, empirical data confirming progressive text difficulty from ILR reading level 2+ to level 3.

3.2.2.2 Cohesion markers

As described above, the characteristics of ILR reading level 3 are that the text include “hypotheses, argumentation and supported opinions” and require some ability to “relate ideas and ‘read between the lines’” as well has to make inferences (ILR 2015). The text typology, a category later added to the level descriptors, is that of “analysis and evaluation” (Child 1998: 386). The challenge for testing is to identify textual elements that can be scored and linked back to one or more of these features. I have chosen cohesion markers because they reflect whether ideas are being related and the argument followed, features of both ILR reading 3 and proficient readers as defined in the NCES survey. The primary cohesion marker to be examined is verb tense because it is a necessary item in every sentence, because the clear difference in how tense functions in Japanese and English requires a translator to make tense choices based on context, and because the general agreement about how tense functions in English makes it a reasonably objective error marker. While verb tense in and of itself is not necessarily proof of cohesion, it can certainly be taken as a marker in light of the following statements by Halliday and Hasan:

Cohesion is a semantic relation between an element in the text and some other element that is crucial to the interpretation of it. (1976:8)

What cohesion has to do with is the way in which the meaning of the elements is interpreted. Where the interpretation of any item in the discourse requires making reference to some other item in the discourse, there is cohesion. (1976:11)

The English tense system ... is based on two very simple principles: (1) that there is a choice of past, present and future, and (2) that this choice may be made repeatedly (within limits), each new choice taking the previous one as its point of departure. (1976:186)

Tense can be considered a marker of cohesion reflecting inference and the relating of ideas because all but the initial choice of tense must take previous decisions into consideration and because, as noted above, tense patterns create a “semantic relation” between textual elements that are “crucial to interpretation”. Tense also has merit as an evaluation item for several reasons: first, readers fairly well agree on what constitutes misuse of tense in English, which reduces the amount of subjectivity in evaluation; second, tense is a necessary sentence element and so must appear in every sentence, which increases the opportunities for scoring; and third, because the use of tense differs radically between Japanese and English, the tense choices that translators make reflect whether they are interpreting verbal elements in the text with reference to other items in the discourse, which is one of Halliday and Hasan’s definitions of cohesion (1976: 11). Making these choices at the sentence level or lower (i.e. without reference to earlier decisions), or not revising to establish an appropriate sequence of tenses, will almost inevitably result in a noncohesive English tense pattern. While two possible reasons for failing to create a coherent text have just been given (working at too low a level, not revising), because this study is focusing on what elements that make a text difficult *for a specified population*, the exact (and undoubtedly individual) processes for arriving at an error are not under consideration here.

In addition to tense, implicit and explicit cohesion markers will also be identified and marked as acceptable or unacceptable. Explicit markers are words like *が* (*ga*), *however* and *また* (*mata*), *moreover* that clearly link clauses or ideas together. In general, these should not be difficult (i.e. should have high facility values), although the presence of many explicit cohesion markers in a text points at tightly bound material that cannot be easily reordered, suggesting an ILR reading level of 3 or higher. Implicit markers are markers indicating, for example, that a sentence is a reason for the

preceding sentence but without providing an explicit pronoun signaling the connection. One would expect these to be more difficult (have lower facility values) because they require inference to be resolved. A problem with using implicit markers, however, is that they are less frequent than explicit markers and may only appear once or twice in a text if at all.

3.2.2.3 *Selecting problem points*

An additional aim is to identify *problem points* in the examination passages to explore a method of triangulating the data from text and cohesion markers as a further determination of passage translation difficulty. The selected problem points were derived from earlier work on translation difficulties in light of the type of challenges available in the passages under consideration. An additional goal in identifying problem points is to examine textual elements that may be of particular interest for this language pair. The final purpose is to offer something in the way of, if not replication, at least additional data in support of translation difficulties identified in earlier research.

In their work on translation difficulty, Hale and Campbell found complex noun phrases to be a point of difficulty for translators working from English. These consist of a noun modified by another noun for which “equivalents are constrained by grammar and semantics” (e.g. “case management”) (Hale and Campbell 2002: 25). Such phrases are further complicated when in a “matrix” of modification (e.g. “ multidisciplinary care planning and case management”). Similarly, in a study of Japanese text readability formulas, Tateishi et al. (1988: 650) found that *runs*, “a maximal string that consists of only one type of characters [sic]” reduce readability because such strings “hide the word boundaries” (i.e. they make it difficult to determine which characters belong together to form a word). I have combined the idea of complex noun phrases embedded in a matrix with “runs” in Japanese to arrive at what I think is a particular difficulty for the Japanese-to-English pair: head nouns with long strings of modification. Unlike the runs set by Tateishi et al., these may combine various elements of the Japanese writing systems and are defined as a head noun with its preceding string of modification. Somewhat arbitrarily, 15 characters were set as the minimum length. Such strings of modification are common in Japanese, and the length of a string tends to increase with text complexity.

Moreover, Japanese is a head-final language while English tends toward being head initial. Parsing the modification of a head noun can require some degree of

inference since the translator must determine what other word each word in the string is modifying. As a result, this problem point fits in nicely with the idea of elements that test inference and the ability to see and create relationships in a text.

The second problem point to be used is passive forms. Both Campbell and Hale (1999; Campbell 1999) and Tateishi et al. (1988) have identified these as difficulties. The latter noted that:

Japanese passive forms are also used for potentials. For example, *mirareru* may mean either *be seen* (passive) or *can see* (potential) and *taberareru* may have one of three meanings: *be eaten*, *can eat*, and *can be eaten*. Thus, frequent use of passives tend to make a document vague and less readable. (Tateishi et al. 1988: 651)

As with modified nouns, passive verbs offer another opportunity for examining inference as well as a chance to build up the data set for a previously identified points of difficulty.

As a note on Japanese linguistics, Japanese makes far greater use of ellipsis than does English. Recovery of the omitted elements “is made possible by knowledge of grammar, intuition (experience), and information shared by the speakers (Donahue 1998: 165). As a result, Japanese is a “high context” language in which “meaning derives more from previous mental knowledge and/or contextual cues than from verbal messages” (Donahue 1998: 13-14). Grammatically, this tendency toward ellipsis often results in dropping the subject (Kashima and Kashima 1999: 83) and other null anaphora (Tsujimura 1996: 212-215). As a result, the Japanese-to-English language pair is particularly suitable for an analysis of inference.

3.3 Analysis

This research is concerned with measurement rather than the generation of hypotheses. First, it is a *post facto* analysis of existing material rather than an experiment in which data was systematically collected to prove or disprove a hypothesis. Second, because data collection was not done with *a priori* statistical guidance it is “messy” and was not

originally intended or designed to answer a specific question. In regard to such data, Chatfield (1988), the developer of Initial Data Analysis, notes that:

It is obviously essential to start by summarizing the data, but it is not always recognized that this may be all that is appropriate, either because the data are too messy to carry out “proper” inference, or because the conclusions are obvious without further analysis. (Chatfield 1988: 1)

Third, each of the six data sets (four general passages at ILR reading level 3 and two technical passages at ILR reading level 2+) are small, consisting of twelve translations or less, and so are more appropriate for suggesting trends and possibilities for future studies than providing confirmation for an after-the-fact question. Finally, measurement and initial or exploratory data analysis are important and essential steps in both model formation (Chatfield 2004) and the potential discovery of unexpected relationships (Buja et al. 2009) and permit more freedom of examination than a strict probabilistic proving or disproving of the null hypothesis. Therefore, the analysis will be based on tabular summaries of the data, facility values, and a close reading of the translations.

3.3.1 Measurements to be done

In keeping with the primary goal of attempting to determine textual elements that can be linked to specific items of interest and scored, I have set verb tense as a primary marker of cohesion. Facility values will be calculated for the tense of the final verb in each sentence and also for the implicit and explicit coherence markers, as well as for each of the problem points (modified head nouns and passive constructions). Dichotomous scoring will be done for each verb tense, implicit and explicit marker, and problem point.

Having the facility value for each item will show whether the identified coherence marker is measuring a text feature that is difficult for the candidate population. Because facility values are a blunt instrument, they do not reveal precisely what the difficulty was or why it was difficult—those are subjective elements and may well not be consistent across the population being tested. And while tense and the other cohesion markers may allow us to make judgments about whether the distinctive

features of texts at ILR reading level 3—notably, inference and following an argument—are indeed elements that are difficult for this group of test takers, they cannot prove that such a link exists.

Facility values will also be calculated for the two types of problem points identified. Both types of problem points are related to (1) inference and (2) following an argument and have the potential to provide further evidence that these elements constitute difficulties. In addition, previous studies (Campbell and Hale 1999; Campbell 1999; Tateishi et al. 1988) have indicated that passive structures are points of difficulty. Having the facility values for passive structures could confirm this. Furthermore, resolving whether a verb is passive or potential in Japanese and correctly identifying the subject, since this is often unstated, are other elements requiring inference and following an argument.

Earlier studies on translation difficulty have focused on cognitive effort as a measure of difficulty (Hale and Campbell 2002; Mishra et al. 2014; Sun 2012; Sun and Shreve 2014). Using scored text features linked to a facility value additionally permits an examination of items that may have required time and effort to resolve, but have a high facility value because they were resolved correctly and, conversely, items with a low facility value that may not have received any particular attention from the translators.

Two technical passages at ILR reading level 2+ passages will be analyzed in the same manner using facility values for comparison with the higher reading level—and so presumably more difficult—passages and to look for types of difficulties that might be linked to texts at this level. Exploratory data analysis in tabular form as the result of close reading will also be done to look for items that were unexpectedly difficult in this population.

3.3.2 Earlier studies using facility values and dichotomous scoring

The only earlier study I have found that uses facility values to determine item difficulty is Campbell's 1991 work on translation competence. He used item difficulty (i.e. facility values) to calculate values for lexical and grammatical difficulty and concluded that "A tentative view of item difficulty analysis is that at the lexical level it must have some utility for translation testing" and suggested also that this type of "product data" from tests could usefully be used to study the translation process (Campbell 1991: 331-

332). The problems he saw were defining item boundaries for grammatical structures and the need to “make decisions about the level of delicacy with which we deal with *correctness*” (Campbell 1991: 331; emphasis in the original). I have attempted to cope with these problems by identifying discrete items (final verbs) that could be objectively scored as *correct* or *incorrect*. As in this study, Campbell’s material came from a professional accreditation examination, in his case, from Arabic to English

A 2009 article by Eyckmans et al. suggested using “text segments that have discriminating power” as identified by pre-testing a representative sample and using these “as the sole instances on which to evaluate the translation performance” on a subsequent examination (Eyckmans et al. 2009: 75). The items are to be scored dichotomously, with correctness determined by previous grader agreement. A Dutch passage of 346 words resulted in 170 items to be scored (the translation was Dutch to French). The “Calibration of Dichotomous Items” created from these was statistically shown to relate “the performance indicators to the underlying translation competence in a more reliable and consistent way” than analytic or holistic methods of evaluation (Eyckmans et al. 2009: 87). An obvious drawback of this approach is that it is extraordinarily labor intensive. Also, it is based on errors made by the pre-test population rather than a decision to test a specified item of interest.

3.4 Summary

The rationale for this study is that (1) pinpointing text features that cause difficulty could be an aid in more focused test design for evaluating translations; (2) various studies have been done on what constitutes a difficulty in translation, but the findings need further confirmation; and (3) because current methods for evaluating translations often focus on word or sentence level elements, studies of items related to evaluation at the paragraph or text level would be useful and possibly predictive.

The aims of the study are (1) to identify specific passage features linked to items of interest (e.g. the production of a translation with overall cohesion) that can be objectively scored and to evaluate these for difficulty by means of facility values; (2) to examine whether elements characteristic of ILR reading level 3 passages are difficult, as measured by item facility, for the test population; (3) to examine whether these items of presumed reading difficulty (e.g. features pertaining to paragraph/text level cohesion) may also be items of translation difficulty; and (4) to identify problem points in the

examination passages to explore a method of triangulating the data from text and cohesion markers as a determination of overall translation difficulty.

The analysis goal of the study is to provide useful measurements and not to generate and prove a hypothesis. The primary measurement of interest is the facility value of passage features being used as indicators of text cohesion and reading level. Facility values provide a measure of item difficulty within a specified population and allow a quantitative comparison of difficulty.

4. Materials and methods for analysis

This chapter describes the data sets and analysis methods for the general passages. While the analysis is based on the parameters set out in this chapter, it should be remembered that both the passages used and the test population came from a professional translation setting and were not the result of an experimental design. This has the disadvantage of precluding statistical analysis, but the advantage of allowing an essentially anthropological look at actual test practices and populations. Therefore, in addition to the analysis based on defined text features, the analysis also includes a close reading of the passages to look for interesting, but unplanned for, items and patterns, as well as a supplementary analysis based on previous studies.

4.1 Data sets

The data sets for this study are the candidate examination papers from the 2004-2007 ATA Japanese-to-English Certification Examination. For the mandatory general passage, there were 11 papers from 2004 (set A), 12 papers from 2005 (set B), 4 papers from 2006 (set C), and 9 papers from 2007 (set D). For the optional technical passage (candidates could choose between a technical or legal/business passage), there were 11 papers from 2004 (set A) and 11 papers from 2005 (set B; one candidate did not complete the technical optional passage). Because the majority of the 2006 and 2007 candidates did not translate the optional technical passage, the technical passages from those years were not included in the analysis. The legal/business passages for the same period were excerpts from a contract and a law and were not appropriate for inclusion in this study.

4.1.1 Test population

The test population consisted of people taking the ATA Certification Examination for Japanese-to-English translation between 2004 and 2007. They were self-selected and

presumably self-identified as professional translators. Everyone sitting the examination was required to meet one of the following eligibility criteria:

- Certification from another member of the Fédération Internationale des Traducteurs (FIT);
- An approved degree or certificate in translation and/or interpreting;
- Bachelor’s or advanced degree in any field with evidence of working as a translator or interpreter for at least 2 years; or
- Less than a Bachelor’s degree and evidence of working as a translator or interpreter for at least 5 years (ATA 2008).¹²

The criteria are fairly broad, but probably correspond reasonably well to the types of qualifications held by working translators and should have eliminated people with absolutely no translation experience. The criteria were also consistent over the four-year period under consideration, making facility values for text features identified as difficulties comparable across passages. However, it should be kept in mind while I refer to the “test population”, there was a different group of translators each of the four years.

The eligibility requirements went into effect in January 2004, but “they have not been notably successful either in increasing the pass rate or reducing the number of unqualified individuals who take the examination” (Koby and Champe 2013: 169). A survey of the pass rate for all language combinations for the period 2004 to 2009 by eligibility criterion was 39% for those holding a FIT certification, 17% for those with any advanced degree or a translation and/or interpreting degree, 12% for those with a Bachelor’s degree and two or more years experience, and 12% for those with at least five years of experience (Koby and Champe 2013: 171).¹³

From the NAAL survey data discussed in the previous chapter, one would expect a failure rate of at least 70% to 75% in this population. Based on that, I set 0.25 as the lowest acceptable facility value, i.e. the cut off point under which an item was considered too difficult. Items with a facility value greater than 0.70 were considered easy for the testee population. It should be noted, though, that even though all the final

¹² Since the period under consideration, the ATA has added ACTFL reading proficiency test scores of “Advanced Low” (equivalent to ILR reading level 2) in both the source and target languages as an additional option for eligibility (ATA 2015).

¹³ The percentage of people taking the Japanese-to-English exam would have been less than 5% of the total.

verbs and other implicit or explicit cohesion markers were scored, one would not expect to see a passage in which all the facility values for these were in a range between 0.25 and 0.70. The generally accepted range for most tests is 0.30 to 0.70 (Fulcher and Davidson 2007: 102-103). In terms of passage selection for this population, one would want at least one item that could be linked to ILR text features with a facility value of 0.25 to discriminate between those who could and could not handle texts at that level.

4.1.1.1 Limitations of test population

The fact that the candidates met the eligibility criteria indicates that they had some exposure to translation and had either trained as a translator or worked as one. However, that the test takers met one of the stipulated items is the only information available about this group of translators. More targeted data, such as which of the eligibility criteria was met, whether an individual had any training in translation, the amount and type of experience, and even the translation direction (B language to A language, or A language to B language) is not available. As a result, none of the evaluation items can be tested against any feature of the translators. While this is a limitation in terms of being able to, for example, link training or experience with translation product features, treating the set of translators as a test population and having criteria defining that population is the basis for calculating item facility and creating a difficulty profile for the passages relative to this defined population. It is also possible that some people took the exam in multiple years.

4.1.2 Test conditions

All translations were produced under approximately the same conditions. In a proctored examination setting, candidates were required to produce handwritten translations of two passages of 389 to 407 Japanese characters (based on the sample translations, 200-230 English words for the general passages and 190-200 words for the technical passages) within 3 hours. Paper references were allowed, but electronic resources and Internet access were prohibited. Candidates could not share references or consult other people (i.e. all test papers were the result of individual work). The standard dictionaries used would most likely have been *Kenkyusha's New Japanese-English Dictionary* and *The New Nelson Japanese-English Character Dictionary*.

The fact that all of the translations were produced in similar circumstances by a defined population means that they are comparable and can be treated as a set. At the same time, it must be noted that the test conditions were very different from the normal environment in which most translators work. First, people use computers and do not handwrite their translations. In addition, some people use handwriting so rarely that they may have trouble physically creating a document in this manner. Also, computers facilitate the reordering of text in a way that handwriting does not, which may have affected translation performance. Second, many translators depend on both electronic resources and Internet research and may not use traditional paper dictionaries often if at all. These factors may have affected the translation products, and traditional dictionary use could have led to more literal translations than an individual might have produced in another setting. Lastly, attempts on the part of candidates to meet perceived grader expectations also might have resulted in a tendency toward literal translation. However, the items selected for evaluation in this study focus on paragraph and text level cohesion markers, and the test conditions should not have greatly affected performance on these indicators.

4.1.3 Passage selection criteria

ATA passage section guidelines changed somewhat during the period under consideration. The published description of the 2004 and 2005 passages was “One passage is mandatory for all candidates. This general text is written for the educated lay reader in expository or journalistic style”. However, in their selection of passages, language workgroups were encouraged to choose passages with the following characteristics:

[The general passage] should present a clear and coherent progression of thought and reasoning in which the candidate may be required to follow an argument or supported opinion and possibly author inference. The passage should contain translation challenges in form of varied sentences patterns grammatical difficulties, and idioms. (ATA Graders 2008)

The above guidelines were drafted during 2005 and approved in 2006. During that period the idea of using ILR reading level as part of the selection criteria was introduced.

Selection guidelines for the technical passage were that:

These passages have the character of typical texts within these domains but should not contain highly specialized terminology or require mastery of a particular field. The subject matter should be such that most candidates can understand a passage without feeling the need to do research. [...] While the passages should not be overly specialized, they may be at the level of a text that would be used in a freshman college course. (ATA 2006)

As the reference to freshman textbooks suggests, there was no intention of setting the reading level of the technical passages lower than that of the general passages, but that is what happened in the set of passages under consideration. In my view, three of the general passages are clearly at ILR reading level 3; general passage C is either a weak level 3 or level 2+. The two technical passages are at ILR reading level 2+.

The passages have comparable lengths, with all texts between 389 and 407 characters.

4.1.3.1 Limitations of passage data set

The main limitation of the passages being used is that they were not specifically selected to facilitate study of text cohesion, tense as a cohesion marker, or the use of inference and following an argument. This restricts the type of quantitative analysis that can be done. The passages also contain elements (e.g. missing antecedents, statements that require more context to be fully understood, repetitious sentence structure, terms that might have needed research to be resolved) that in retrospect seem less than ideal for testing. However, they were selected based on explicit guidelines and were used in as materials for a professional translation test. Thus they are genuine “real world” materials and analyzing both the passages and the resulting translation products does provide insight into what one might term “the state of translation outside of translation training programs”.

Another limitation is the small sample size. The primary analysis is being done on a total of 36 short translations across four passages. The secondary analysis of the

technical passages has only two passages and a total of 22 translations. As a result, this research must be considered an exploratory study only.

The poor quality of some of the translations could be considered an additional limitation. However, all the translations were done by people who met the specific qualifications set out above. Also, while many of the translations could not be used without extensive editing or re-translation, they are not outside the quality range of translations I have been asked to revise professionally. The failure rate of 88% among candidates meeting the eligibility criterion of five or more years of experience also suggests that much of the translation done professionally is not of the highest quality.

4.1.4 Passages

The source texts of the four general passages, a possible translation of each (mine, for all passages), and descriptions of the argument that needs to be followed to understand the text and create a cohesive translation are given below.

4.1.4.1 General passage A

The first general passage appears to have been taken from a source such as a history of political science in Japan. It is not self-contained, since the time expression in the first sentence refers back to an earlier statement that is not provided. The text presents an argument with supporting statements, giving it a reading level of ILR 3. In addition, the material is tightly ordered both in terms of the argument and the time sequence. Some degree of author shaping is apparent.

The mode is evaluative (i.e. “facts are selected and pressed into service in order to develop points of view” [Child 1998: 386]), with an argument working toward a conclusion. The development of the topic is linear and organized in the form of problem–cause–result with supporting evidence. The sentence structure is reasonably complex, with insertion and qualification. The vocabulary is generally concrete, with several linking lexical chains. Viewed sociolinguistically, the passage does not create a wide cultural gap for this language pair, but some cultural background is helpful. The style is not highly individual or idiosyncratic, and the passage is well written.

日本人研究者による日本政治研究の水準は、これまで、少数の優れた研究があったとはいえ、全体として高くはなかった。学問の源泉が外国にあるという古来の学問的伝統に、天皇制による束縛、さらには、政治学者の近代化・民主化への関心が重なって、優れた政治学者を政治的先進国の政治の研究に駆り立ててきたからである。日本政治の研究はこれら学者の余技か、政治評論・文明批評として行われることが多かった。丸山真男の業績は、日本政治の研究水準を一挙に高めるとともに、才能のある政治学者の目を日本に向けることに貢献したことは言うまでもないが、いわゆる丸山学派の人たちの分析対象は主として思想史であって、現代日本政治の科学的研究に及ぶことが少なかった。日本政治の研究のレベルが一般的に高まり、日本政治専攻者の数が増えるようになったのは、ごく最近のことである。最近話題になった政治学書の多くは、現代日本政治を対象とした本格的な研究である。

Possible translation:

Apart from a few excellent studies, up to the time of Masao Maruyama [これまでで literally *up to now*] it could be said that the general level of research on Japanese politics by Japanese scholars was rather low. This was because restrictions resulting from the emperor system and the interest of political scientists in modernization and democratization, coupled with the long-standing scholarly tradition that the fountainhead of scholarship was overseas spurred the best scholars to study the politics of politically advanced countries. For the most part, these scholars studied Japanese politics as a hobby or in the form of political or cultural commentary. Needless to say, the achievement of Maruyama was that he raised the level of research on Japanese politics virtually overnight and also helped direct the attention of talented political scientists to Japan. However, the members of the so-called Maruyama school primarily focused on analysis of the history of ideas and rarely produced scientific studies of contemporary Japanese politics. The overall improvement in the level of research on Japanese politics and the increased number of scholars specializing in that field is an extremely

recent development. Many of the writings on political science that are current topics of discussion are genuine studies of contemporary Japanese politics.

Reading the passage requires some degree of inference in seeing the relationships between sentences. Also, there is a fairly tight linkage between sentences in terms of causality, suggesting that translating sentences in isolation might not result in a successful target text. In terms of background knowledge, knowing who Masao Maruyama (a distinguished political scientist) was and his dates in particular (1914-1996) would be helpful, but not essential. Reading comprehension of the passage requires following the argument that:

- Up until the time of Masao Maruyama, the level of research on Japanese politics by Japanese scholars had not been high.
- This was the result of domestic restrictions (i.e. the emperor system) and the desire of leading political scientists to study “politically advanced countries” (i.e. Western nations).
- Maruyama changed that situation, but not completely.
- Only recently have Japanese scholars produced genuine studies of contemporary Japanese politics.

Understanding the first point—that little serious research on Japanese politics had been done in Japan before the time of Maruyama—is complicated by the lack of an antecedent for *これまで* (*koremade*, literally *up to now*). What this refers to cannot be resolved until the mention of Masao Maruyama in the fourth sentence.

4.1.4.2 *General passage B*

The first sentence of this passage suggests that it was taken from a discussion of post-war Japanese economics, while the reference to “means of production” hints that the author may have had a Marxist orientation. The text overall lacks context and is a bit confusing in places (why, for example, does the separation of production and consumption spur the replacement of brooms with vacuum cleaners?) Nonetheless, the passage can be translated as a reasonably coherent unit. It meets the requirements of ILR reading level 3 because it contains argumentation and supported opinions, with

each sentence fairly directly referring back to the one before it. The mode is evaluative, and author shaping is apparent.

日本の高度成長期を通じて社会の基礎単位である家族の形態及び関係が変わっていった。この間増加した労働者家族においては、夫は外での勤務、妻は家事・育児という性的分業が固定されることになった。しかし、その一方で、夫婦共働きの増加によって、家庭内の性的分業を見直し、平等な夫婦関係をつくろうとする新たな動きも始まった。

家族と社会の関係にも大きな変化があらわれた。都市に集中した若い核家族と単独世帯の多くは当然労働者家族であるが、農民家族と異なって、生産の手段をもたない労働者家族は、生産と消費が分離され、居住地では消費的機能だけを担うことになる。そして、このように社会的依存度を高めた消費の場すなわち家事労働の場に対して、冷凍食品、使い捨ての消耗品等が続々とはいりこみ、洗濯・掃除器具の電化がはやく進んだ。さらに、共同体の相互扶助機能が低下したため、保育その他を社会的施設に頼らざるを得なくなった。

Possible translation:

The form of the family, the basic unit of society, and the relationships within it continued to change throughout Japan's period of high economic growth. During this period, a sexual division of labor in which the husband worked outside the home and the wife was responsible for housework and raising children became fixed in the increasing number of workers' families. However, at the same time an increase in the number of couples in which both husband and wife worked lead to a reevaluation of the sexual division of labor within the household and a new movement for an equal partnership within couples began.

Major changes were also seen in the relationship of the family and society. Most of the young nuclear families and single-person households concentrated in cities were, of course, workers' families. However, workers' families, in contrast to farming families, did not have a means of production, and so production and

consumption were separated, with the result that residential areas took on the function of consumption only. Thus, frozen food and disposable consumer goods continuously flowed into the site of household labor, that is, the site with this high degree of societal dependence, and the electrification of washing and cleaning appliances made rapid progress. In addition, the decline in the mutual support function of the community made dependence on social facilities for childcare and other services inevitable.

The links between sentences are more explicit in this passage than in general passage A and marked with conjunctions such as *しかし* (*shikashi, however*), *そして* (*soshite, thus*) and *さらに* (*sarani, in addition, furthermore*). As in the first passage, the passage is structured as a series of changes within a set period of time (here, *Japan's period of high economic growth*, roughly defined as the post-war period through the end of the 1980s). Reading comprehension of the above passage requires following the argument that:

- The structure of the family in Japan is changing and this is related to economic changes.
- Contradictory changes took place in the roles of husbands and wives in workers' families.
- Changes also took place in the relationship between families and society.
- Such changes involved a split between the means of production and consumption.
- This created social dependency.

The reader is left to infer that the separation of production and consumption somehow results in dependence on society in order to resolve the *このように* (*konoyouni, in this way*) in the sixth sentence. This proved to be a challenge in several of the translations.

4.1.4.3 General passage C

The third general passage has a simpler structure than the previous two passages and the sentences are shorter than is usual for an academic text. The argument set out does not require inference to be understood. It could be rated as either a weak ILR reading level

3 or perhaps a level 2+, statements with supporting facts, but the fact that the mode is governed by argumentation rather than the presentation of facts push the level more toward level 3. Some author shaping is apparent.

江戸初期の鎖国は、情報化に逆らうものではない。それはむしろ情報の統御を意味しており、情報というものの重要性の強い認識が、鎖国の根底をなしている。江戸初期の限られた情報処理能力では、「管理」のために、鎖国がもっとも適切な手段と考えられたのであろう。

管理能力の向上につれて近世社会は必然的に情報化する。鎖国を「悲劇」あるいは「権力の恣意」と見なすのは、情報に関する一貫した観点を欠くからに過ぎない。鎖国はむしろ、この国の置かれた状況のために、社会の情報・管理化の初期に生じた、一種の「歴史的必然」と見なすべきであろう。

実際は、現代の歴史家は、表現は異なるにせよ、すでにそれに近い視点をとっていると思われる。たとえば、山口哲二氏はいう。「鎖国という言葉が、われわれにどんなに多くの誤解を与えるか。日本が全く孤立する形で独自の世界をつくった、というように理解されがちでした。しかしそれは全くの誤解です。」

Possible translation:

The closed country policy at the beginning of the Edo period was not intended to run counter to an increasing use of information. Rather, it meant the control of information; a strong recognition of the importance of information formed the basis of the closed country policy. Given the limited ability to handle information at the beginning of the Edo period, the closed country policy was probably considered the most appropriate means of “management”.

With increased management ability, modern society necessarily becomes information oriented. Viewing the closed country policy as a “tragedy” or an

“abuse of power” is nothing more than the lack of a consistent view of information. Instead, the closed country policy should be seen as a type of “historical necessity” that arose during the early period of a society beginning to use and manage information because of the situation in which Japan found itself.

In fact, contemporary historians have already taken a similar position, although expressed differently. For example, Tetsuji Yamaguchi has written, “How many misconceptions has the term ‘closed country’ created? It has predisposed us to think that Japan created its own world in a form that was completely isolated. However, that is a complete misunderstanding.”

In terms of background knowledge, it would be helpful, but not necessary, to know something about the closed country policy in the Edo period (1603-1868). Reading comprehension of the passage requires following the argument that:

- The closed country policy was meant to control, rather than to restrict, the use of information.
- Information was recognized as being important.
- At the time, the closed country policy appears to be the only way of managing information.
- The policy should be seen as a “historical necessity” rather than as a “tragedy” or “abuse of power”.
- Some historians have already noted this.

The argument is straightforward and does not require any reading between the lines. The challenges are more in the area of vocabulary since the two terms that are at the heart of the argument—鎖国 (*sakoku*, *closed country*) and 情報化 (*jouhouka*, *information oriented*)—are a challenge for “congruity judgment” because there are no direct equivalents in English. The first refers to the closed country policy or locked door policy instituted during the Edo period (1603-1867) and requires the addition of a word like *policy* in translation. The second word, in this context, refers to the increasing use of information in a society. However, the standard meaning, as given in *Kenkyusha* is “computerization” or “the development of a computer network” (Watanabe 2003:

1307). Therefore, arriving at an appropriate English equivalent requires some thought on the part of the translator. In addition, the addition of 化 to a noun (here, 情報, *jyouhou*, *information*) maps to the English suffix -ization (e.g. 近代化 and 民主化, *modernization* and *democratization* from the first passage, and 電化 *electrification* from the second). While *informationization* is used in the translation of the same characters in government documents from China, particularly in military contexts, the word *formation* does not work in the context of Japanese history. In addition, given the special meaning of the term here, where it obviously does not refer to any aspect of computer-based information technology, I suspect it may have been defined by the author earlier in the text (if so, the definition should have been provided to the translators). However, in the passage a translator will need to find a way of coping with the suffix, probably by falling back on the implied meaning “to become more of something”.

This passage has two limitations. First, understanding it depends on resolving the two terms discussed above. While *sakoku* should not be problem for anyone familiar with Japanese history, *jouhouka* is being used in a peculiar way and could affect overall performance. Second, general passage C is a weak ILR reading level 3 because it is structured as a series of short sentences, and does not present the same challenges as the other three passages. It also represents the type of text in which word-level problem points can be found that, in some setting, could be used to test transfer skills.

4.1.4.4 General passage D

This passage is an editorial, a text type that is a classical example of ILR Reading Level 3 and the evaluative mode. The text is self-contained and is an argument for better treatment of civil servants. Author viewpoint and shaping is apparent.

日本国家の優秀な人的資源である国家公務員を誹謗中傷しても、なんら日本国民が得をするわけではない。国家のために全精力を尽くす人には、存分な報酬があつてしかるべきである。そこで次のような人事政策を提案する。

まず、国家公務員の公務の効率性を目指し、優秀な人材の確保を図るには、年功序列から能力主義・実力主義への転換が必要である。また、適

材適所および複線型人事を行う。省別採用ではない一括採用は、地方公務員の採用で取り入れられている。

そして、同期の者が局長・次官に昇任しても他の者は天下りをせず、定年まで勤めること、現在、局長職までの年俸は相対的に低いことを改善し、報酬をアップすること、定年後の年金額も例えば退職時の8割を支給するなどにより増額し、公務員在職中に培ったキャリアと本人の希望にふさわしい人生を送ることができるようにすること、これらの政策により、天下りによる民業の圧迫と、税金の浪費を回避することもできよう。

Possible translation:

The Japanese people in no way benefit from slandering and defaming civil servants, who are Japan's leading human resource. It is appropriate that those who exhaust themselves for the country should have adequate compensation. Therefore, we propose the following personnel policy.

First, aiming at the efficiency of the official duties of national civil servants and to ensure superior human resources, it is necessary to transition from the seniority system to a merit system. Moreover, the "right person in the right place" and a "multi-track personnel system" should be implemented. Collective hiring rather than hiring by each ministry is being incorporated into the employment of regional civil servants.

Next, the following policies should also make it possible to avoid wasting taxpayer money and pressuring private enterprises by having civil servants "parachuting in" to management positions: even though people who started work in the same year have been promoted to bureau chief or vice minister, others should work until retirement age and not be forced to "parachute" into private companies; increasing compensation by improving annual salaries of civil servants up to the level of bureau chief, which are currently relatively low; and making it possible for civil servants to foster their careers during their public service and to lead a life appropriate to their expectations by further increasing the amount of pension after retirement to, for example, 80% of the salary received in the final post.

Reading this passage requires more culture-specific background than the previous three passages. A translator would need to know that promotion in the Japanese civil service is based on seniority and that hiring is done by individual ministry. While those two items can be gleaned from the text, a translator might not know that civil servants start work collectively in April of a given year and are considered a “class”. When a member of the class is promoted to bureau chief or vice minister, the rest of the group is expected to retire and move to private industry. Companies foster ties with the government by allowing ex-civil servants to “parachute in” (天下り, *amakudari*). Reading comprehension of the above passage requires following the argument that:

- Japanese civil servants are not respected and do not receive adequate pay.
- The editorialist is proposing a personnel policy to remedy the situation.
- The policy requires:
 - Changing from a seniority system to a merit system
 - Making better use of human resources by using people according to their talents, having a multi-track personnel system and using an inclusive pool of employees
 - Having civil servants work until retirement age instead of moving to private industry when their “classmates” are promoted
 - Increasing salaries
 - Increasing pensions

The vocabulary is somewhat specialized. Terms include 年功序列 (*nenkou joretsu*, seniority system), 能力主義・実力主義 (*nouryoku shugi-jitsuryoku shugi*, merit system), 複線型人事 (*fukusengata jinji*, multi-track personnel system), 一括採用 (*ikkatsu saiyou*, there are several possibilities, but some phrase with the general meaning of *collective hiring*), 同期の者 (*douki no mono*, class, those who started work at the same time), and 天下り (*amakudari*, parachuting in, the appointment of a former civil servant to a management position in a private company). The first two are straightforward, but the last three phrases include examples of why evaluation methods focusing on word level translation can be a problem. Is “co-worker hired at a similar time” acceptable? or “Peers with equivalent seniority”? (both examples taken from the test translations). And for *amakudari*, does “appointed to positions of responsibility in

private corporations” actually correspond to this Japanese practice? Because answers to such questions tend to be subjective and because these types of issues can be easily solved by providing a gloss for the word, culture-specific terms without a readily equivalent can be problematic items for translation evaluation unless the test design is specifically aiming at testing these sorts of translation problems.¹⁴

4.2 Analysis method for main data set

4.2.1 Cohesion markers: specific verbs and other implicit and explicit markers

The main cohesion marker to be scored is tense of the final verb in each sentence. Sentences with an obvious breaking point (usually signaled by *が*, *ga*, *however* in mid sentence) will be treated as having two final verbs. The specific final verbs, with information about their Japanese tense and possible English translation, as well as the implicit and explicit cohesion markers are given below. The translations of the verbs and other cohesion markers are the items to be marked as either acceptable or unacceptable; the score will then be used to calculate facility value. Just the tense of the verb is being scored, which means that a sentence with an acceptable verb tense may be unacceptable in other ways (e.g. word choice, spelling, grammar) and may not even have the correct subject.

4.2.1.1 General passage A

The six (when the fourth sentence is divided) final verbs in General Passage 1 can be characterized as follows:

1. 高くはなかった (*takakuwanakatta*): base meaning, *to not be high*; inflected form, past, completed action; acceptable English translation, *was not high*, *was low* (simple past)
2. 駆り立ててきたからである (*kuritatetekita kara dearu*): base meaning of verb, *to impel*, *to spur*; base meaning of ending, *because*; non-past formal copula. The main verb in Japanese has an auxiliary verb indicating continuation of action marked as past and followed by a subordinate conjunction expressing a cause, then ending with

¹⁴ Tests for becoming a translator at the EU sometimes try “to gauge candidates’ cultural knowledge and awareness by including culture-laden concepts in the test-text” (Lafeber 2012b: 149).

- a non-past formal copula. English requires the past tense: *This was because ... spurred ...* .
3. 多かった (*ookatta*): base meaning, *many*; inflected form, adjective conjugated in past form. The syntax results in a different main verb in Japanese and English. The basic meaning is that the situation stated was a frequent occurrence. The English tense is taken from the final verbal form; the Japanese verb (行われる) that becomes the main verb in English is non-past passive.
 4. a. 貢献したことは言うまでもない (*koukenshitakoto wa iumademonai*): base meaning of verb, *contributed to*, followed by set phrase meaning *it goes without saying*; inflected form of verb, past; acceptable English translation, *It goes without saying that contributed to ...* . Because this sentence has が (*however*) in the middle, there is a natural dividing point. The tense overall comes from the final verb in 4b, which is in the past. b. 少なかった (*sukunakatta*): base meaning, *few*; inflected form, adjective conjugated in past form. One solution in English is to turn the final verbal into an adverb and take 及ぶ (*extend to*, non-past) as the main verb.
 5. である (*dearu*): base meaning, copula; inflected form, non-past, formal form; acceptable English translation, *is [an extremely recent development]*
 6. である (*dearu*): base meaning, copula; inflected form, non-past, formal form; acceptable English translation, *are [genuine studies]*

The tense structure of this passage moves from the past tense to the present, which is ideal for examining tense as a cohesion marker. The final copula of the second sentence is non-past, which could be a point of confusion in translation. The identified implicit cohesion markers are これまで (*koremade*, literally, *up to now*, *up to this time*), the time expression in the first sentence, which lacks a clear antecedent and requires using the information in the fourth sentence to be resolved, and からである (*karadearu*, *because*), indicating that the second sentence is the reason for the conditions in the first sentence. However, close reading of the translations resulted in a third implicit marker (sentence 3), which was erroneously adding *because*, signaling an incorrect inference. The sole explicit marker is が (*ga*, *however*) in the fourth sentence, which requires a sense of *he did this, but* to maintain cohesion (and coherence, in this case).

4.2.1.2 General passage B

The seven final verbs in the passage can be categorized as follows:

1. 変わっていった (*kawatteitta*): base meaning, *to change*; inflected form, continuing action in the past; acceptable English translations, *continued to change* (a past tense and a sense of continuation during the period named are needed)
2. 固定されることになった (*koteisareru koto ni natta*): base meaning, *to become fixed*; inflected form, completed action in the past. The verb 固定する (*koteisuru*) is in the non-past passive form, and the final phrase (ことになった [*koto ni natta*]), indicating a resulting situation, is in the past tense showing completed action. An acceptable English translation would be, *become fixed*.
3. 始まった (*hajimatta*): base meaning, *to begin*; inflected form, completed action in the past; acceptable English translation, *began*.
4. あらわれた (*arawareta*): base meaning, *to appear*; inflected form, completed action in the past, passive form; acceptable English translation, *appeared*
5. a. である (*dearu*): base meaning, copula; inflected form, non-past, formal form; acceptable English translation for maintaining temporal cohesion: *were*. Because this sentence has が (*however*) in the middle, there is a natural dividing point. b. 担うことになる (*ninaiu koto ni naru*): base meaning, *to take on, bear*; inflected form, non-past verb followed by a final phrase (ことになる [*koto ni naru*]) indicating a resulting situation in the non-past showing that the situation continues to exist. This can also be considered an historical present (Soga 1983: 46). To maintain passage cohesion, this needs to be translated by an English past tense: *took on*.
6. 進んだ (*susunda*): base meaning, *to progress*; inflected form, completed action in the past; acceptable English translation, *made progress*
7. 頼らざるを得なくなった (*tayorazaru o enakunatta*): base meaning, *came to be forced to depend on*. An auxiliary ending— ざるを得ない (*zaru o eanai*)— inflected in the past is attached to the negative verb stem of *tayoru* (*to depend on*), meaning that the action of the verb must inevitably happen. This is followed by *naru* (*to become*) in the past tense. An acceptable, although literal, English translation would be, *come to be forced to depend on*.

There are no implicit cohesion markers in this passage. The explicit cohesion markers are この間 (*kono aida*) *this period* referring back to 日本の高度成長期 (*nihon no koudo seichou ki*) *Japan's period of high [economic] growth*, しかし、その一方で (*shikashi, sono ippou de*) *However, at the same time; However, on the other hand*; が (*ga*), *however*; そして (*soshite*), *thus, therefore*; and さらに (*sarani*), *in addition, moreover*.

4.2.1.3 General passage C

The eleven final verbs in the passage can be categorized as follows:

1. ものではない (*mono dewa nai*): base meaning, *is not* (emphatic); inflected form, non-past (narrative present here); acceptable English translation, *was not*
2. なしている (*nashiteiru*): base meaning, *to become*; inflected form, progressive or resulting state; acceptable English translation, *formed* (resulting state)
3. 考えられたのであろう (*kangaerareta no dearou*): base meaning, *to consider*, followed by a copula; the first inflected form is past potential and the inflected copula is in the formal written presumptive form (Kaiser et al. 2001: 395), *should*; acceptable English translation, *probably could be considered*
4. 情報化する (*jyouhouka suru*): base meaning, *to become information oriented*; inflected form, non-past; acceptable English translation, *becomes information oriented*
5. 過ぎない (*suginai*): base meaning, *is nothing more*; inflected form, non-past; acceptable English translation, *is nothing more*
6. 見なすべきであらう (*minasubeki dearou*): base meaning, *should be seen/probably should be seen*, followed by a copula; the first inflected form is the non-past infinitive followed by the auxiliary *beki* (*should, ought*), followed by the sentence-ending copula in the formal written presumptive form; acceptable English translation, *should be seen as, should be viewed as*
7. とっていると思われる (*tottiru to omowareru*): base meaning, first verb, *to take*, second verb, *to think*; inflected forms, perfect and passive; acceptable English translation, *have already taken*
8. いう (*iu*): base meaning, *to say*; inflected form, non-past (narrative present); acceptable English translations, *has said, has written*

9. 与える (*ataeru*): base meaning, *to give, to provide*; inflected form, non-past followed by question marker; acceptable English translation, *has created/has caused*
10. 理解されがちでした (*rikai saregachi deshita*): base meaning, *tend to be understood*; inflected form, copula in past tense; acceptable English translation, *has tended to be understood*
11. 誤解です (*gokai desu*): base meaning, *is a misunderstanding*; inflected form, non-past copula; acceptable English translation, *is a misunderstanding*

The first tense challenge here is that several of the verbs are a non-past tense in Japanese but need to be in a past tense in English. In addition, the final verbs in the second and seventh sentences are in the *-teiru* form, which is often a progressive tense in Japanese. However, it is also used for a resulting state with verbs of change such as *nasu*, while the *すでに* (*sudeni*, already) in sentence 7 requires a perfect tense in both the source and target languages (see Kaiser et al. 2002: 489-490 for a more complete discussion).

Sentence 7 is further complicated by the sentence ending *と思われる* (*to omowareru*), which is the passive form of *思う* (*omou*, *to think*), but here indicates that the writer is expressing hesitation about expressing the stated opinion (Makino and Tsutsui 1995: 325-327; these writers say that this form is not a true passive). Given that there is no equivalent in English and that the scholarly courtesy conventions are different, this can be omitted in translation. Inclusion would suggest an overly literal approach.

In sentence 6, because *べき* (*beki*) is not used when addressing superiors (Makino and Tsutsui 1995: 14), I suspect that the form of the copula is a courtesy addition to soften the sentence. Therefore, the polite *であろう* (*dearou*, *probably*) can be safely omitted in English. However, including it would not count as a mistake.

The Japanese non-past verbs in sentences 8 and 9 need to be translated as a present perfect in English because the actions are completed but the effects remain. The verb in sentence 10 consists of the passive stem of the verb *理解する* (*rikai suru*, *to understand*) plus the auxiliary suffix *がち* (*gachi*, *tend to* [usually in used in a negative sense]), followed by a past copula, from which the sentence takes its tense. English cannot use a simple past here because the action is not complete.

There are no implicit cohesion markers in general passage C. The explicit markers are *それは* (*sorewa, this, it*) a pronoun referring to *江戸初期の鎖国* (*edo shoki no sakoku, closed country policy of the Edo period*); *むしろ* (*mushiro, instead, rather*); and *しかし* (*shikashi, however*).

4.2.1.4 General passage D

The seven final verbs in the passage can be categorized as follows:

1. 得をするわけではない (*toku o suru wake dewa nai*): base meaning, *to not profit* (emphatic); inflected form, non-past; acceptable English translation, *in no way benefit, absolutely do not profit from*
2. である (*dearu*): base meaning, copula; inflected form, formal non-past; possible English translation, *it is*
3. 提案する (*teian suru*): base meaning, *to suggest*; inflected form, non-past; acceptable English translation, *we propose* (the editorial *we* is preferable here, particularly since using it would indicate that the translator recognized the text as an editorial and was familiar with the standard form of editorials in English)
4. である (*dearu*): base meaning, copula; inflected form, formal non-past; acceptable English translation, *it is*
5. 行おう (*okanau*): base meaning, *to carry out, to implement*; inflected form, non-past; acceptable English translation, *should be implemented*. A simple present would be jarring here, because the author is making suggestions.
6. 取り入れられている (*toriirerareteiru*): base meaning, *to incorporate*; inflected form, passive progressive or passive perfect; acceptable English translation, *is being incorporated/has been incorporated*
7. できよう (*dekiyou*): base meaning, *can, to make possible*; inflected form, hortative or presumptive; acceptable English translation, *should be possible*

The tense structure of this passage is relatively simple. The first five verbs can all be rendered in the English present tense. As noted in the description of general passage C, the *-teiru* form can be either progressive or perfect. Deciding on the correct form here would require knowing whether local incorporation of a collective hiring

system was complete or in progress. Since that information is not available, both the perfect and the progressive are acceptable. The final verb is presumptive, which is consistent with the text being an editorial.

The explicit cohesion markers were *そこで* (*sokode, therefore*); *まず* (*mazu, first*); *また* (*mata, moreover, in addition*); and *そして* (*soshite, next*). While viewed individually, *まず* (*mazu, first*) and *そして* (*soshite, next*), the first words in paragraphs two and three, respectively, are explicit cohesion markers, when taken together they can be seen as showing whether the translator was conscious of the overall structure of the passage (i.e. A problem exists. To remedy it, we need first to do this, then to do that.) and so can be used as an implicit cohesion marker.

4.2.2 Problem points

The problem points in the general passages to be marked either acceptable or unacceptable and used to calculate facility values are head nouns with modification strings of 15 characters or more and passive constructions.

4.2.2.1 Modified head nouns

The modified head nouns will be scored as acceptable or unacceptable. The five nouns with long modification strings found in the four general passages are listed below.

1. General passage A: 日本人研究者による日本政治研究の水準 (*nihonjin kenkyuusha niyuru nihon seiji kenkyuu no suijun*); possible translation, “The level of research on Japanese politics by Japanese researchers”
2. General passage A: 学問の源泉が外国にあるという古来の学問的伝統 (*gakumon no gensen ga gaikoku niaru toiu korai no gakumonteki dentou*); possible translation, “The long-standing scholarly tradition that the source of scholarship was abroad”
3. General passage A: 政治学者の近代化・民主化への関心 (*seijigakusha no kindaika•minshuka he no kanshin*); possible translation, “Political scientists’ interest in modernization and democratization”
4. General passage B: 夫は外での勤務、妻は家事・育児という性的分業 (*otto wa soto de no kinmu, tsuma wa kaji • ikuji to iu seiteki bungyou*); possible translation,

“a sexual division of labor in which the husband worked outside the home and the wife was responsible for housework and raising children”

5. General passage B: このように社会的依存度を高めた消費の場すなわち家事労働の場 (*konoyouni shakai-teki izondo o takameta shouhi no ba sunawachi kaji roudou no ba*); possible translation, “The place of consumption with this sort of high degree of social dependence, in other words, the place of housework”
6. General passage D: 日本国家の優秀な人的資源である国家公務員 (*nihonkokka no yuushuuna hito-teki dearu kokka koumu'in*); possible translation, “Civil servants, who are a superior resource of the Japanese nation”

The fact that there were no long modification strings in general passage C, while there were three in general passage A further illustrates the limitations of this data set. However, modified nouns were not identified as a feature to be included in passages for the certification examination (although it would fall under a “translation challenge at the sentence level”) and working with passages used in an actual exam opens a window into how the passage selection process could be improved.

The modified nouns will be scored both individually and in tandem with the syntax of the sentence in which they appear. Scoring is accompanied by a two caveats. First, acceptable/unacceptable scoring of modification and syntax is less objective than that of tense in that the evaluator can sometimes wonder “Is this acceptable or not?” Second, because the focus is on correctly parsing the modification, grammatical mistakes within a modification string will be ignored. For example, “The overall standard of studies on the Japanese politics by Japanese scholars” (A2) is acceptable, despite the misuse of the definite article, while “the level of research in the political sciences in Japan can certainly be said to have been raised by a small number of outstanding works by Japanese scholars” (A10) is unacceptable because it misses the point that the topic of the passage is studies of Japanese politics by Japanese researchers.

4.2.2.2 *Passive constructions*

All of the general passages contain at least one passive construction. In the fourth sentence of general passage A, a present passive (*okonawareru*) appears before the nominalizer *koto*, and takes its tense from the final conjugated adjective (*ookatta*).

While the construction can be translated in the passive (“In many cases, studies of Japanese politics were done by these scholars as a hobby or as political or cultural commentary”), it can equally well be changed into an active sentence (“For the most part, these scholars studied Japanese politics as a hobby or in the form of political or cultural commentary.”)

General passage B has passive constructions in the second and third sentences. The construction in the second sentence is a present passive (*kotei sareru, is being fixed*) followed by the final phrase *koto ni natta* expressing that a situation has come about. A natural English rendering of this is a past tense in the active voice: “became fixed”. The past tense passive in the third sentence, *arawareta*, can be translated either in the passive (*were seen*) or active (*appeared*) voice and is the final verb in the sentence.

There are three passive or passive-like constructions in general passage 3. The first part of the final construction in sentence 3, *kangaerareta no dearou*, could be seen as either one. The final verb in sentence 7, *omowareru*, is the passive form of *omou, to think*. However, when used as a sentence final element it often means *it seems* or *it appears*. The test of whether it is a true passive or the writer signaling some hesitation is whether a subject marked by *-ni* (action *by someone*) can be grammatically added (Makino & Tsutsui 1995: 327). *Rikai sare-*, the passive stem in sentence 10 is a true passive: *has been understood as*, but is complicated by the auxiliary suffix *gachi*, resulting in *has tended to be understood as*.

The only passive construction in general passage 4 is the *-teiru* verb at the end of sentence 6. While the form can be either potential or passive, the potential does not fit the context. The *-teiru* ending signals action in progress when with used with an action verb, a resulting state when used with a stative verb, and either action in progress or resulting state when used with verbs of motion or change (Kaiser et al. 2001: 489-490). *Toriireru (incorporate, take in)* would seem to fall into the last category, making either interpretation possible. There is not enough context in the passage to make a determination.

Verb tense is scored as either acceptable or unacceptable for tense only. However, in order to use passive constructions to obtain a bit more information about whether the translators were following the argument in the passage, they will only be marked as acceptable if all aspects of the construction are correct, i.e. subject, object, tense, and passive or potential.

4.2.3 Features of ILR reading level 3

Specific text features typical of ILR reading level 3 are “hypothesis, argumentation and supported opinions” requiring, on the part of the reader and translator, the ability to “interpret material correctly, relate ideas and ‘read between the lines,’ (that is, understand the writers’ implicit intents)” (ILR 2015). I am summarizing that as “the ability to make inferences” as demonstrated by target language verb choice, rendering of the other implicit and explicit cohesion markers, and correct resolution of noun modification.

4.2.4 Facility values

The facility value (also called *item facility* or *item difficulty index*) is the number of correct responses, in other words, the “proportion of test takers who answer an item correctly” (Fulcher and Davidson 2007: 102). It is not a measure of objective or absolute difficulty, but rather expresses the percentage of people in a specified population who responded to an item correctly. This means that the value is dependent on both the item scored and the test population (Fulcher and Davidson 2007: 102). Although the selection criteria for the test population were rather broad, they were consistent for all examinations and so do allow for facility value comparison of passages and of passage features.

In educational testing, a facility value of 0.50 is considered ideal and a range of 0.33 to 0.67 (McNamara 2000: 60-61) to 0.30 to 0.70 (Fulcher and Davidson 2007: 102) acceptable. However, in light of the NCES survey statistics on adult literacy and because one purpose of a translation test is to find out whether the translator can read and understand the source text, I have set 0.25 as lowest acceptable facility value, but kept 0.70 as the highest. Values outside of that range could suggest various problems: a mismatch between the test population and test items, test items that do not provide useful information, or questionable passage selection.¹⁵ Having these values permits test designers to determine whether a given scored item is useful, compare passage difficulty between languages and over time, and calculate item discrimination (whether

¹⁵ For example, a passage in which a substantial part of the population got every verb tense wrong would signal a definite mismatch between the passage and the test population.

“an item discriminates between test takers with high and low scores on the test”)
(Bachman 2004: 129).

4.3 Method of data presentation

The data is presented in tabular form with tables giving tense and other cohesion marker scores. The results are summarized at the end of each general passage via tables with overall tense errors, overall cohesion marker errors, and problem point errors. These will give both the facility values for each item and the respective number of errors for each translation.

The translations are examined twice: once to score the designed cohesion markers and problems points, then a second time via a close reading in light of the objective results provided by the facility values.

4.4 Supplementary analysis based on earlier studies

With the aim of examining whether difficulties found in earlier empirical studies also appear in this data set, population, and language pair, items cited in Hale and Campbell (2002) and eye tracking studies of translation difficulty (Jensen 2009; Mishra et al. 2013) will be examined. Evaluation items in addition to those described above are culturally specific terms as a stand-in for Hale and Campbell’s “official terms”, readability indices, which Jensen uses as a measure of text complexity, sentence length, used by Mishra et al. as part of their attempt to develop a system for automatically assigning text difficulty levels (and also in many readability formula), and structural complexity, also from Mishra et al., defined as “the total length of the dependency links in the structure” (Lin 1996: 730). These items are described in detail in Chapter 7. The results are presented in tabular form.

4.5 Data set and analysis of the technical passages

Tense and other cohesion markers for the technical passages will be examined in the same way as for the general passages, but without tables for each sentence. Because

these passages are at ILR reading level 2+ and so have a lower degree of overall cohesion, cohesion within sentences will also be examined. Other items to be scored were not determined beforehand, but were discovered through close reading of the translations. Facility values will be calculated for each item. The passages and items for scoring are given in Chapter 8.

5. Analysis of the general passages: results for tense, other cohesion markers, and problem points

This chapter will present the results, in table form, for each of the cohesion markers set out in Chapter 4: tense of final verb and implicit and explicit cohesion markers. The problem points—modified head nouns and passive constructions—will also be presented in tabular form. The presentation is organized by sentence for each general passage. Tables of the overall tense, cohesion marker, and problem point errors with facility values and number of errors are given at the end of the passages to summarize the results. The chapter concludes with a brief overview of the tense errors across passages.

5.1 Results: general passage A

The results of the dichotomous scoring for tense and the implicit/explicit cohesion markers and problems points by sentence are given below in table form.

5.1.1 General passage A: tense and other cohesion markers

Table 5.1. General passage A, sentence 1: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 1		Other Cohesion Marker
	Japanese: 高くはなかった (simple past) English: simple past or past perfect		
	Acceptable	Not Acceptable	これまで (implicit)
A1		has not been high	Omitted (X)
A2		has not been high	Translated as “so far” (X)
A3		is not high	Translated as “until now” (X)
A4		has not been high	Omitted (X)
A5		has not been very high	Omitted (X)
A6		may be described ... as not high overall	Translated as “until to now” (X)
A7	was generally not very high		Translated as “until to now” (X)
A8	was generally not high		Translated as “until to now” (X)
A9	had not been		Translated as “in the past” (O)
A10	was not high		Translated as “in the past” (O)
A11	was not very high		Translated as “in the past” (O)

O indicates an acceptable solution; X, an unacceptable solution.

The fact that the passage was out of context so that there was no antecedent for *こらまで* (*koremade*, literally, *until now*) and no dates were given for Masao Maruyama are possible reasons for the poor performance on this sentence, with only three translations having acceptable solutions for both tense and the time frame cohesion marker. None of the translators picked up on the hint in sentence 4 that the pivotal point being discussed was the appearance of Masao Maruyama. Four of the six incorrect tense choices were the present perfect.

Table 5.2. General passage A, sentence 2: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 2		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 駆り立ててきたからである (verb in past + formal copula in present)		
	English: simple past		
			からであ, <i>this was because</i> (implicit)*
A1		The reason is	O
A2	... drove capable political scientists to studies on the		X
A3	was limited was restricted were driven by		X
A4		Verbs in present and past: is overseas, had set	X
A5		This is because	O
A6		has been spurring	X
A7	This was because		O
A8			O
A9			O
A10	It was because		O
A11	was held back		X

O indicates an acceptable solution; X, an unacceptable solution.

*Indicates that the situation in the first sentence is the result of the conditions in the second sentence.

Only two of the eleven translations (18%) had acceptable solutions for both the tense and the cohesion marker. That the copula *である* is in the non-past form most likely accounts for the “is because” of translations A8 and A9; this is a literal translation of the Japanese, where it indicates that the reason given remains valid. More interestingly, four of the five translations that had an acceptable tense in English at the same time omitted any reference to the second sentence being the cause of the situation in the first sentence, suggesting that *から* (*kara*, *because*) was discarded along with the copula. The implicit cohesion marker *kara dearu* (*because*) is an example of the null pronoun in Japanese, in which “[a]ny argument can be phonologically unrealized if the discourse context is sufficient to identify its reference” (Inoue & Fodor 1995: 9). Because the topic of the second sentence—the first sentence—is not represented by a pronoun and topic marker (e.g. *これは*, *kore wa*), it has to be inferred from the presence

of *kara dearu*, making this particular item a combined measure of text cohesion and a language-specific feature.

Table 5.3. General passage A, sentence 3: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 3	Other Cohesion Marker
	Japanese: 多かった (simple past) English: simple past	
	Acceptable	addition of <i>because</i> (implicit)*
	Not Acceptable	
A1	has been performed	O
A2	were often done as	O
A3	was frequently conducted	O
A4	as studied	O
A5	has been undertaken	O
A6	was commonly engaged in	O
A7	It might have been because	X
A8	Research on Japanese politics was	O
A9	was treated	O
A10	It was often the case	O
A11	Perhaps it was because	X

O indicates an acceptable solution; X, an unacceptable solution.

* Addition of *because* in the target text indicates an incorrect inference.

Sentence 3 is short (43 characters) with a clear past tense ending, and eight of the eleven translators (73%) had no difficulty with tense. The unacceptable translations of this sentence show the same preference for the English present perfect that was seen in sentence 1. Translations A7 and A11 made sentence 3 the explanation for the conditions in sentence 2, a relationship that is not in the source text.

Table 5.4. General passage A, sentence 4a: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 4a	Other Cohesion Marker
	Japanese: 貢献したことは言うまでもない (verb in simple past + final set phrase in present) English: simple past	
	Acceptable	Not Acceptable
A1	Maruyama raised	
A2	achievements of	
A3	Maruyama contributed to the work of Maruyama	
A4	Masao raised	Maruyama's achievement is
A5	the achievement of Masao Maruyama was	
A6		Maruyama's accomplishments include
A7	the achievement of Maruyama Masao was	
A8	the work of Masao Maruyama raised	
A9	brought up the standard	
A10	the achievement of Maruyama was	
A11	The achievement of Maruyama was	

The tense in translations A4 and A6 appears to have been taken from the final set phrase (言うまでもない, *iumademonai*, *it goes without saying*) echoing similar errors made by both translators in sentence 2. Given that the passage is describing the state of research on Japanese politics by Japanese scholars before, during, and after the time of Maruyama, this tense error would suggest a failure to connect the dots.

Table 5.5. General passage A, sentence 4b: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 4b	Other Cohesion Marker
	Japanese: 少なかった (simple past) English: simple past	
	Acceptable	Not Acceptable
A1	was primarily	が (however) (explicit)
A2	were mainly	omitted (X)
A3	was mainly	Moreover (X)
A4	was the main subject	O
A5		has primarily been
A6		have mainly researched
A7	was intellectual	O
A8	mainly analyzed	O
A9	remained mainly	O
A10	focus ... was	O
A11	tended to be	omitted (X)

O indicates an acceptable solution; X, an unacceptable solution.

There were tense problems in only two of the translations, with the erroneous use of the present perfect appearing again. Because the meaning of sentence 4 is that Maruyama got other Japanese scholars to focus on Japan but the result was studies of

the history of ideas and not contemporary Japanese politics, the *が* needs to be represented both to maintain cohesion and to convey the meaning.

Table 5.6. General passage A, sentence 5: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 5	Other Cohesion Marker
	Japanese:である (non-past formal copula)	
	English: present/present perfect	
	Acceptable	Not Acceptable
A1	It is only recently	
A2		It was very recently
A3	this is a very recent phenomenon	
A4	It is only recently	
A5	It has only been quite recently	
A6	are an extremely recent phenomena	
A7	It has only been very recently	
A8	Only recently has	
A9	the level ... has risen	
A10	It is only recently	
A11	is a fairly recent phenomenon	

The Japanese and English tenses are equivalent in this sentence, and only one translation had a tense error.

Table 5.7. General passage A, sentence 6: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 6	Other Cohesion Marker
	Japanese:である (non-past formal copula)	
	English: present	
	Acceptable	Not Acceptable
A1	are serious research	
A2	are full-scale studies	
A3	is the subject	
A4	are fundamental studies	
A5	have been serious works	
A6	are full-scale research	
A7	are genuine studies	
A8	are serious works	
A9	are revealing full-fledged researches	
A10	are full-fledged research works	
A11	tend to be ones	

Again, the Japanese and English sentences are both present tense, and there were no problems with tense in any of the translations.

5.1.2.1 Summary tables and facility values

Table 5.8. General passage A, overall tense errors

Translation No.	Sentence No.							No. of errors
	1	2	3	4a	4b	5	6	
A1	X	X	X	O	O	O	O	3
A2	X	O	O	O	O	X	O	2
A3	X	O	O	O	O	O	O	1
A4	X	X	O	X	O	O	O	3
A5	X	X	X	X	X	O	O	5
A6	X	X	O	X	X	O	O	4
A7	O	O	O	O	O	O	O	0
A8	O	X	O	O	O	O	O	1
A9	O	X	O	O	O	O	O	1
A10	O	O	O	O	O	O	O	0
A11	O	O	O	O	O	O	O	0
Facility value	0.45	0.45	0.82	0.82	0.73	0.91	1.00	

O indicates an acceptable solution; X, an unacceptable solution.

The facility values of the sentences with present tense in both languages and no cohesion markers (sentences 5 and 6) were notably high, with only one error in sentence 5 in translation A2. On the other hand, the second sentence, with an implicit cohesion marker and a complicated verb structure in Japanese with the verb in the past followed by the conjunctive particle *kara* (because) then a present tense copula had a facility value showing that this item was difficult for just under half the test population. The first sentence had an appropriate facility value score for tense (0.45). However, as Table 5.9 shows, the facility value for the implicit cohesion marker, which required taking information from later in the passage to be resolved, had the lowest facility value (0.27). In terms of passage selection, though, having at least one item at this facility value level is appropriate for this population in terms of differentiating the good translators from the very good translators.

Table 5.9. General passage A, overall cohesion marker errors

Translation No.	Sentence No.							No. of errors
	1	2	3	4a	4b	5	6	
	(これまで, <i>up to the time of Maruyama</i>)	(からで ある, <i>this was because</i>)	(insertion of <i>because</i>)		(<i>ga, however</i>)			
			Implicit		Explicit			
			Implicit					
A1	X	O	O	-	X	-	-	2
A2	X	X	O	-	O	-	-	2
A3	X	X	O	-	X	-	-	3
A4	X	X	O	-	O	-	-	2
A5	X	O	O	-	O	-	-	1
A6	X	X	O	-	O	-	-	2
A7	X	O	X	-	O	-	-	2
A8	X	O	O	-	O	-	-	1
A9	O	O	O	-	O	-	-	0
A10	O	O	O	-	O	-	-	0
A11	O	O	X	-	X	-	-	2
Facility value	0.27	0.64	0.82		0.73			

O indicates an acceptable solution; X, an unacceptable solution.

There is a good correspondence between the implicit cohesion markers and tense errors in sentence 1, but not in sentence 2, perhaps because the link back to the previous sentence did not involve temporal information.

5.1.3 Problem points

Four items—three modified nouns and one passive construction—were identified as problem points. A summary of the error patterns for these is given in Table 5.10, including a score for the syntax of the sentences containing the modified nouns.

Table 5.10. General passage A, problem point errors

Translation No.	Modified nouns*					Passive construction	No. of errors
	日本人研究者による日本政治研究の水準	Syntax of Sentence 1	学問の源泉が外国にあるという古来の学問的伝統	政治学者の近代化・民主化への関心	Syntax of Sentence 2	行われること Sentence 3	
A1	O	O	O	O	O	O	0
A2	O	O	O	O	O	O	0
A3	X	X	X	X	X	X	6
A4	O	O	X	X	X	O	3
A5	O	O	O	O	O	O	0
A6	O	O	O	X	X	O	2
A7	O	O	O	O	X	X	2
A8	O	O	O	O	O	O	0
A9	O	O	O	X	X	O	2
A10	X	X	X	O	X	X	5
A11	O	O	X	O	X	X	3
Facility value	0.81	0.81	0.73	0.64	0.45	0.55	

* See 4.2.2.1 for translations.

O indicates an acceptable solution; X, an unacceptable solution.

The first modified noun had a high facility value with errors in only two translations. This modification string can be translated in order working back from the head noun (i.e. from right to left) so that the literal translation “Japanese researchers by Japanese politics researcher of level” becomes the acceptable “The level of research on Japanese politics by Japanese researchers”. Because the translation does not require any decisions as to whether an adjective is modifying a noun that it immediately precedes or the more distant head noun, the modifications would be expected to be easy to resolve for the test population. The two erroneous translations were:

According to Japanese researchers, the level of Japanese political research, ... (A3)

Although the level of research in the political sciences in Japan can certainly be said to have been raised by a small number of outstanding works by Japanese scholars, overall that level was not very high in the past. (A10)

In A3 the wrong meaning for *ni yoru*, which has a range of possible uses, had been selected: *according to* instead of *by*. In A10 the noun plus modification string has been broken up, which is not grammatically possible. This has both muddled the topic

of the sentence, which is the noun *level* with its modification, and obscured the point that the issue is research on *Japanese political science* and not *research in the political sciences in Japan*.

Incorrect resolution of the modification string mapped nicely to incorrect parsing of the sentence syntax, with one additional translation showing syntax problems (A6).

The first modified noun in the second sentence had a facility index of 0.73, showing that it was relatively easy for this population. Because this string contains a verb, more reordering in English was required than in the earlier example. The literal “scholarship of source SUB foreign country in COPULA say long-standing of scholarly tradition” becomes “the long-standing scholarly tradition that the fountainhead of scholarship was overseas”. Translating this string can be done from grammatical knowledge alone and does not require any decisions about which parts of the string modify which noun. The two erroneous translations were:

The source of knowledge was limited to ancient academic traditions from abroad... (A3)

this scholarly tradition, which has foreign origins, ... (A11)

Translation A3 has broken up the modification string, which will almost inevitably distort both the syntax and the meaning. Translation A11 has omitted *long-standing*. Also, the relationship between “scholarly tradition” and “foreign origins” (an example of poor word choice) is incorrect.

The second modified noun in sentence two has a facility value of 0.45, showing that it was difficult for this population. The phrase is also the first example of a modifier at the beginning of the string modifying the head noun rather than the noun it immediately precedes. The literal *political scientists of modernization • democratization in of interest* becomes *the interest of political scientists in modernization and democratization*. The five erroneous translations were:

prominent political scientists mainly concerned with modernization and democratization... (A3)

the modernization and democratization of political scholars... (A4)

modernization and democratization among political scholars ... (A6)

the pursuit of modernization and democratization by the political scholars ...
(A9)

As the interest in modernization and democratization grew amongst political
scientists ... (A10)

The root problem appears to be a failure to connect *interest* and *political scientists*. Translation A10 comes closest to doing this, but has added a verb that is not in the source text, which confuses the meaning. None of the remaining translations have rendered the phrase as a head noun, *interest*, modified by some version of *interest of political scientists in modernization and democratization*. This type of error suggests that the translators are not recognizing the structure of the modification but also that they are not seeing the meaning of the phrase in context.

All of the translations with problems resolving the modification also failed to resolve the syntax of the sentence overall. The syntax of translation A7, in which both modified nouns were handled correctly, was incorrect because the verb 重なる (*kasanaru*, here, *added to*) was translated as *increased*, which seems to have also happened in A6.

The facility value for the passive construction in sentence 3 was 0.45, showing that this item was difficult for the test population. Translation A3 had the correct tense and form for the verb, but not the correct object (that research was done as hobby). Translation A7 had the tense correct but had an incorrect subject (“It might have been because”). Translation A10 has added an inappropriate *because* and *but for whatever reason*, resulting in the same incorrect object as in A3. Translation A11, like A7, used the English present perfect.

5.2 General passage B

The results of the dichotomous scoring for tense and implicit/explicit cohesion markers and problems points by sentence are given below in table form.

5.2.1 General passage B: tense and other cohesion markers

The tense and implicit/explicit cohesion markers for general passage B appear below.

Table 5.11. General passage B, sentence 1: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 1		Other Cohesion Marker
	Acceptable	Not acceptable	
	Japanese: 変わっていった (past tense with marker of continuing action)		
	English: past tense indicating change over time		
B1	changed as		–
B2		have changed	
B3		has undergone changes	
B4		were changing	
B5	changed through		
B6	Trough [sic] ... changed		
B7	During ... changes took place		
B8	was accompanied by change		
B9	steadily changed		
B10	During ... changed		
B11		have changed	
B12	Along with ... changed		

Unlike the first sentence in general passage A, the time frame being discussed was clearly stated here, resulting in a higher facility value (0.75) for the first sentence. Three of the four incorrect tenses were present perfect.

Table 5.12. General passage B, sentence 2: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 2		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 固定されることになった (completed action in the past)		
	English: simple past		
			この間 (during this period, during this time) (explicit)
B1	was solidified		O
B2		has become fixed	O
B3	was established		O
B4		had been fixed	Not long ago (X)
B5	became entrenched		O
B6		securing	during the period (X)
B7	became entrenched		O
B8		was established	O
B9	became established		O
B10	became established		O
B11	became fixed		O
B12	became entrenched		omitted (X)

O indicates an acceptable solution; X, an unacceptable solution.

As in the first sentence, the fixed period of time requires a verb form showing completed action. Interestingly, incorrect solutions to この間 (*kono aida*, *this period*) correlated with unacceptable verb choices.

Table 5.13. General passage B, sentence 3: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 3		Other Cohesion Marker
	Japanese: 始まった (completed action in past) English: simple past		
	Acceptable	Not Acceptable	
B1	began		O
B2		has started	O
B3	began		O
B4		have started	missing sense of an opposite action (X)
B5	began		O
B6	occured [sic]		O
B7		has given birth	O
B8		main verb omitted, sentence in present tense	O
B9	began		O
B10	began		O
B11	began		O
B12	began		O

O indicates an acceptable solution; X, an unacceptable solution.

Four of the translations had an unacceptable verb tense, with the present perfect predominating. There was only one mistake with the cohesion marker at the beginning of the sentence; this was paired with an unacceptable tense.

Table 5.14. General passage B, sentence 4: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 4		Other Cohesion Marker
	Japanese: あらわれた (completed past action, passive form) English: past passive (rewording in the active is acceptable)		
	Acceptable	Not Acceptable	
B1	emerged		–
B2		has also come about	
B3	There was also		
B4		have also appeared	
B5	There was also		
B6	chanced [sic] dramatically		
B7		have also appeared	
B8	There was also		
B9	were also seen		
B10	also appeared		
B11	also began to undergo		
B12	also appeared		

The tense errors in this sentence were again in the present perfect. Given that this sentence described a situation that was happening at the same time as that in sentence 3, it is not surprising that three of the translators made the same mistake in both sentences.

Table 5.15. General passage B, sentence 5a: tense and other cohesion markers

Translation No.	Tense of Main Verbs Sentence 5a		Other Cohesion Marker
	Japanese: である, non-past copula English: simple past, <i>were</i>		
	Acceptable	Not Acceptable	
B1	were		–
B2		is	
B3	were		
B4		are	
B5	were		
B6		are	
B7		are	
B8		are	
B9		are	
B10		are	
B11	were		
B12		are	

Table 5.16. General passage B, sentence 5b: tense and other cohesion markers

Translation No.	Tense of Main Verbs Sentence 5b		Other Cohesion Marker
	Japanese: 担うことになる (non-past expressing a situation that continues to exist, historical present) English: simple past, <i>took on</i>		
	Acceptable	Not Acceptable	が <small>が</small> (however) (explicit)
B1	resulted in		O
B2		carry the burden of	O
B3	functioned		O
B4		bear	O
B5	assumed		O
B6		take charge of	X
B7		becomes	O
B8		perform	X
B9		shoulder	O
B10		to be carried out	X
B11		come to serve	O
B12		take on	O

O indicates an acceptable solution; X, an unacceptable solution.

This sentence breaks naturally into two in translation; however, a single long sentence would not be incorrect.

Sentence 5 had the greatest number of tense errors, probably because the tense was taken from the apparent tense of the two verbs and not from overall context or the requirements of English tense. While the four translations that lacked a sense of *however* also had erroneous tenses, I do not think there was a correlation here the way there might be with time expressions. The problem seems to have been more with syntax and chunking parts of the sentence.

Table 5.17. General passage B, sentence 6: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 6		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 進んだ (completed action in the past) English: simple past		
			そして (<i>thus, therefore</i>) (explicit)
B1	progressed		and (X)
B2		has advanced	O
B3	there was progress		O
B4	made progress		omitted (X)
B5	there was progress		omitted (X)
B6	penetrating ... thus promoting*		O
B7	progressed		then (X)
B8		has made	O
B9	made an entré ... proceeded apace		omitted (X)
B10		has progressed	O
B11	became		O
B12	caught on		O

O indicates an acceptable solution; X, an unacceptable solution.

* In translation B6, the *-て/-い* (*-te/-i*) form has been taken as a causative rather than connective. That is a possible interpretation.

This sentence is interesting in that, despite the cognitive effort most likely needed to parse the modified noun (this was correct in only three translations), most of the translations had the correct tense. However, four translations (B6, B7, B9, and B12) with acceptable final verb tense had tense inconsistencies inside the sentence, with present tense verbs in one or more clauses before the final past tense verb. Once again, the unacceptable tense choices were all present perfect.

A correct representation of *そして* (*soshite, thus, accordingly*) is needed to understand the relationship between this sentence and the situation described in sentence 5.

Table 5.18. General passage B, sentence 7: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 7		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 頼らざるを得なくなった (completed action in the past) English: simple past		
			さらに (<i>in addition, moreover</i>) (explicit)
B1	it became inevitable to rely on		O
B2		has become compulsory	O
B3	it became necessary to rely on		O
B4	there could not help but increase dependence (very literal)		O
B5	it became necessary to rely on		O
B6		is no choice left but to depend on (very literal)	O
B7		has forced ... to rely on	O
B8		has to depend on	O
B9	it because necessary to rely on		O
B10		has become unavoidable	O
B11	they were forced to rely on		O
B12		they have no other choice	O

O indicates an acceptable solution; X, an unacceptable solution.

からに (*karani, in addition, moreover*) is a frequently used conjunction, and so it is hardly surprising that all of the translations used it correctly. That half of the papers had an incorrect tense for the final verb is more puzzling, since both the final verb and the verb in the first clause are in the past, as was the final verb of the preceding sentence.

5.2.2.1 General passage B: summary tables and facility values

Table 5.19. General passage B, overall tense error pattern

Translation No.	Sentence No.								No. of errors
	1	2	3	4	5a	5b	6	7	
B1	O	O	O	O	O	O	O	O	0
B2	X	X	X	X	X	X	X	X	7
B3	X	O	O	O	O	O	O	O	1
B4	X	X	X	X	X	X	O	O	6
B5	O	O	O	O	O	O	O	O	0
B6	O	X	O	O	X	X	O	X	4
B7	O	O	X	X	X	X	O	X	5
B8	O	X	X	O	X	X	X	X	6
B9	O	O	O	O	X	X	O	O	2
B10	O	O	O	O	X	X	X	X	4
B11	X	O	O	O	O	X	O	O	1
B12	O	O	O	O	X	X	O	X	3
Facility value	0.67	0.67	0.67	0.75	0.33	0.25	0.75	0.50	

O indicates an acceptable solution; X, an unacceptable solution.

While six of the seven sentences have facility values within the 0.25 to 0.70 range, four of these were at the very lower end or near the upper end of the set range. The second half of the fifth sentence had a facility value of 0.25, suggesting that verbs with a non-past form in Japanese that need to be translated in the past tense to maintain English cohesion are items of particular difficulty. That was confirmed by the tense analysis results of general passage C. B2 was the only translation to have errors in all the tenses. Six of the verbs were in the present perfect and the remaining two in the present tense. The translation is generally well written, so the tense problem does not seem to have been caused by difficulties writing English. Perhaps it is an artifact of an approach to learning Japanese as a foreign language.

Table 5.20. General passage B, overall cohesion marker* error pattern

Translation No.	この間 (sentence 1)	しかし、その一方で (sentence 3)	が (sentence 5)	そして (sentence 6)	さらに (sentence 7)	No. of errors
B1	O	O	O	X	O	1
B2	O	O	O	O	O	0
B3	O	O	O	O	O	0
B4	X	X	O	X	O	3
B5	O	O	O	X	O	1
B6	X	O	X	X	O	3
B7	O	O	O	O	O	0
B8	O	O	X	O	O	1
B9	O	O	O	X	O	1
B10	O	O	X	O	O	1
B11	O	O	O	O	O	0
B12	X	O	O	O	O	1
Facility value	0.75	0.92	0.75	0.58	1.00	

O indicates an acceptable solution; X, an unacceptable solution.

*All cohesion markers were explicit.

Only one of the five coherence markers had a facility value within the appropriate difficulty range for this population, showing that they provided little useful information. However, successful solutions for *soshite* (*thus, therefore*), which had a facility value of 0.58, do appear to be related to successful resolution of the modified noun string immediately following it. Because all of the other cohesion markers in this passage were explicit, it is not surprising that the mistakes with the explicit cohesion markers do not map to tense errors.

5.2.3 General passage B: problem points

One modified noun and two passive constructions were identified as problem points for this passage. A summary of the error patterns for these is given in Table 5.21, including a score for the syntax of the sentence containing the modified noun and the number of errors in each translation.

Table 5.21. General passage B, problem point error pattern

Translation No.	Modified nouns*		Syntax of Sentence		Passive constructions		No. of errors
	夫は外での勤務、妻は家事・育児という性的分業	Syntax of Sentence 2	このように社会的依存度を高めた消費の場すなわち家事労働の場	Syntax of Sentence 6	固定されることになった (sentence 2)	あらわれ た (sentence 4)	
B1	O	O	X	X	O	O	2
B2	O	O	X	X	X	X	4
B3	O	O	X	X	O	O	2
B4	O	O	X	X	X	X	4
B5	O	O	X	X	O	O	2
B6	X	X	X	X	X	O	2
B7	O	O	O	O	O	X	1
B8	O	O	X	X	X	O	3
B9	O	O	O	O	O	O	0
B10	O	O	O	O	O	O	0
B11	X	O	X	X	O	O	3
B12	O	O	X	X	O	O	2
Facility value	0.83	0.92	0.25	0.25	0.67	0.75	

** See 4.2.2.1 for translations.

O indicates an acceptable solution; X, an unacceptable solution.

While three translations had trouble with the first modified head noun in this passage, this only interfered with the syntax in one translation (B6).

The second modified head noun was preceded by a very long string (29 characters for noun plus modifiers). Element length has been identified as difficult in readability studies (Tateishi et al. 1988; Sun & Sheve 2014). In addition, and of particular interest here, this string in the sixth sentence comes directly after *soshite* (thus), which had a facility index of 0.58 and which also indicated that the situation in the sixth sentence was a result of the circumstances described in the fifth sentence. The initial element in the string, *konoyouni* (*in this way, such*) also refers back to the previous sentence, but in a way that is not entirely clear because it has not been explicitly stated that the separation between production and consumption results in

greater societal dependence. The structure of the phrase head noun *place (ba) of house work, that is, the place (ba) of consumption with this type of high degree of societal dependence*.

Because of the length and complexity of the phrase and the need to understand and account for the elements referring back to the previous sentence, it is not possible to determine the primary cause of the low facility value of this item, if there is a single reason. However, that both the syntactical complexity and the requirement of interpreting the sentence in context are features of ILR reading level 3 point to increased demands in terms of source text reading being at least part of the reason for the low item facility.

Not surprisingly, unacceptable resolution of the syntax of the sentence correlated with incorrect resolution of the modified head noun. Both items had a facility value of 0.25, at the bottom of the acceptability range, suggesting that this sentence is particularly useful in discriminating the translators in the population who can cope with a text of this level.

The facility values of the two passive constructions showed that resolving these was easy for this population. In both sentences the unacceptable passive constructions matched the tense errors.

5.3 General passage C

The results of the dichotomous scoring for tense and implicit/explicit cohesion markers and problems points by sentence are given below in table form.

5.1.2 General passage C: tense and other cohesion markers

Table 5.22. General passage C, sentence 1: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 1		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: <i>ものではない</i> (narrative present, context requires past tense) English: <i>simple past</i>		
C1	was not		
C2	did not		
C3		is not	
C4	was not		

This passage is self-contained, so there were no back-reference cohesion markers. Only one translation followed the literal form of the verb and used the present tense.

Table 5.23. General passage C, sentence 2: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 2		Other Cohesion Marker (explicit)
	Acceptable	Not Acceptable	
	Japanese: なしている (progressive indicating resulting state) English: simple past, <i>formed</i>		
			それは (<i>that, it</i>) pronoun referring to 江戸初期の鎖国 (<i>closed country policy of the Edo period</i>)
C1	was		O
C2		forms	O
C3		forms	O
C4		forms	O

O indicates an acceptable solution; X, an unacceptable solution.

Three of the four translations used the unacceptable present tense despite the text being a discussion of an Edo period policy. There were no errors in rendering the pronoun referring back to the previous sentence.

Table 5.24. General passage C, sentence 3: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 3		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 考えられたのであろう (past potential followed by copula in formal presumptive) English: past conditional or past passive		
			–
C1		was (no sense of conditional)	
C2	was probably		
C3		is likely to have been assumed	
C4	was probably seen		

None of the translations included the sense of the potential, the part of the conjugation suggesting an English conditional. However, the addition of *probably*, taken from the copula, was one way of working around this.

Table 5.25. General passage C, sentence 4: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 4	Other Cohesion Marker
	Japanese: 情報化する non past English: present tense	
	Acceptable	Not Acceptable
C1		became
C2	become	
C3	becomes	
C4	become	

The lack of cohesion markers suggest that this passage is at a lower reading level than the previous two general passages and that the information is not as tightly ordered. The unacceptable past tense in translation C1 may have been due to a mistranslation of “modern society” as “late medieval society”. The form is clearly non-past, here and in the rest of the paragraph.

Table 5.26. General passage C, sentence 5: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 5	Other Cohesion Marker
	Japanese: 過ぎない (non-past) English: present tense	
	Acceptable	Not Acceptable
C1	is	
C2	is	
C3	stems	
C4	is	

There were no tense problems with this sentence in any of the four translations. All of them appear to have recognized that the discussion has switched to contemporary views of the Edo period.

Table 5.27. General passage C, sentence 6: tense and cohesion markers

Translation No.	Tense of Main Verb Sentence 6	Other Cohesion Marker
	Japanese: 見なすべきであろう (non-past presumptive following auxiliary meaning <i>should, ought</i>) English: some sense of obligation in present tense	
	Acceptable	Not Acceptable
		むしろ (<i>instead, rather</i>)
C1	ought to regard	O
C2		was
C3	should be considered	O
C4	should rather be seen	O

O indicates an acceptable solution; X, an unacceptable solution.

Courtesy endings in Japanese expressing polite hesitation are frequently omitted in English; however, that was not the best choice in translation here because the author

is telling readers how the situation should be seen: *the closed country policy should be seen as a type of “historical necessity”*. Unsurprisingly, there were no problems with the simple link between the sentences.

Table 5.28. General passage C, sentence 7: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 7	Other Cohesion Markers
	Japanese: とっていると思われる (progressive indicating resulting state followed by present potential verb [<i>could be thought, seems</i> , can also be omitted]) English: present perfect	
	Acceptable	Not Acceptable
C1	have already adopted	–
C2	can be thought to have already adopted (very literal)	
C3	already seem to share	
C4		it is thought that ... have

The first three translations just about cover the range of possibilities for this verb ending. In translation C4, the present tense is correct, but has the subject wrong: “it is thought that modern historians already have viewpoints similar to this”.

Table 5.29. General passage C, sentence 8: tense and cohesion markers

Translation No.	Tense of Main Verb Sentence 8	Other Cohesion Marker
	Japanese: いう (non-past) English: present perfect	
	Acceptable	Not Acceptable
C1	has said	–
C2		says
C3		says
C4		says

While English can use the present to introduce a quotation, the use of the present here would be jarring, given that this sentence is an example of the similar position already taken by contemporary historians. The use of the present here suggests a translator that did not see the sentence in context and did not understand the varied use of tense in Japanese.

Table 5.30. General passage C, sentence 9: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 9	Other Cohesion Marker
	Japanese: 与える (non-past followed by question marker) English: present perfect	
	Acceptable	Not Acceptable
C1	has created	–
C2		provides
C3		leaves us open
C4	has caused	

Two of the four translations rendered the tense correctly. Despite the Japanese non-past, the sense is clearly of an action that has happened but continues to have an effect.

Table 5.31. General passage C, sentence 10: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 10	Other Cohesion Marker
	Japanese: 理解されがちでした (past) English present perfect	
	Acceptable	Not Acceptable
C1	has predisposed	–
C2	tended to be understood	
C3	has been perceived	
C4		was

The translation in C4 misses the sense of *がち* (*gachi*, *tends to*) as well as the notion of an ongoing state.

Table 5.32. General passage C, sentence 11: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 11	Other Cohesion Marker
	Japanese: 誤解です (non-past copula) English: present tense	
	Acceptable	Not Acceptable
		しかし (<i>however</i>)
C1	is	O
C2	is	O
C3	is	O
C4	is	O

O indicates an acceptable solution; X, an unacceptable solution.

All of the tense and cohesion marker choices for this sentence were correct.

Given the short sentences and lack of complicated syntax and modification, this passage would seem almost too easy for testing reading/translation skills at ILR level 3. The lowest number of tense errors was two (translation C1) and the greatest number was five (translation C3)

5.3.2.1 Summary tables and facility values

Table 5.33. General passage C, overall tense errors

Translation No.	Sentence No.											No. of errors
	1	2	3	4	5	6	7	8	9	10	11	
C1	O	O	X	X	O	O	O	O	O	O	O	2
C2	O	X	O	O	O	X	O	X	X	O	O	4
C3	X	X	X	O	O	O	O	X	X	O	O	5
C4	O	X	O	O	O	O	X	X	O	X	O	4
Facility value	0.75	0.25	0.50	0.75	1.0	0.75	0.75	0.25	0.50	0.75	1.0	

O indicates an acceptable solution; X, an unacceptable solution.

The facility values have been calculated, but because of the very small sample size, these should be seen as suggestive only. Whether sentences 2 and 8 would be rated as very difficult in a larger population is an unresolved question. The use of the present tense in sentence 2 is odd, given the explicit reference to the Edo period. The mistakes in sentence 8 appear to have been from taking the apparent tense from the Japanese verb rather than context and standard English usage.

Table 5.34. General passage C, overall cohesion marker errors*

Translation No.	Sentence No.			No. of errors
	2	6	11	
	それは (that, it)	むしろ (instead, rather)	しかし (however)	
C1	O	O	O	0
C2	O	O	O	0
C3	O	O	O	0
C4	O	O	O	0
Facility value	1.0	1.0	1.0	

O indicates an acceptable solution; X, an unacceptable solution.

* All cohesion markers were explicit.

Rather than being indicators of inference or maintaining source text cohesion, related, the cohesion markers—as well as the lack of such markers—are signals indicating that the test passage is too easy and is not testing features that would provide useful information about translator skill.

5.3.3 General passage C: problem points

This passage had no modified head nouns making up strings longer than 15 characters. There were three passive constructions.

Table 5.35. General passage C, problem points

Translation No.	Passive construction			Total No. of errors
	考えられた のであろう (sentence 3)	とっている と思われる (sentence 7)	理解されが ちでした (sentence 10)	
C1	X	X	O	2
C2	O	X	O	1
C3	X	O	X	2
C4	O	X	X	2
Facility value	0.50	0.25	0.50	

O indicates an acceptable solution; X, an unacceptable solution.

The passive constructions here are interesting in that the verb form in sentences 3 and 7 could indicate either a passive or potential verb. The construction in sentence 10 has an auxiliary (*gachi, tends to*), and additionally is a causative form in the passive voice. The facility values for these constructions are all in the set range indicating difficulty, but that might change in a larger population.

None of the translations used the potential, which would have resulted in an English conditional. The acceptable translations conveyed the same idea with *could*. In sentence 7, only translation C3 recognized that *omowareru* was not a true passive and translated it as “seems”.

Translations C3 and C4 both omitted any sense of *gachi, tends to* or *predisposes toward*.

5.4 General passage D

The results of the dichotomous scoring for tense and implicit/explicit cohesion markers and problems points by sentence are given below in table form.

5.4.1 Tense and other cohesion markers

The tense and implicit/explicit cohesion markers for general passage D appear below.

Table 5.36. General passage D, sentence 1: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 1		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 得をするわけではない(non-past)		
	English: present		
D1	does not benefit		
D2	will not result		
D3	are in some way profiting		
D4	do not benefit		
D5	is not benefit		
D6	do not benefit		
D7	have nothing to gain		
D8	have absolutely nothing to gain		
D9	gain nothing		

Since this is the first sentence of a self-contained text, one would not expect a cohesion maker. While there was some syntactical confusion, particularly in translation D3 (“Say what you want to libel the excellent human resource of the Japanese government official, you can’t say that they are in some way profiting from the Japanese people.”), there were no tense errors.

Table 5.37. General passage D, sentence 2: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 2		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: である		
	(non-past copula)		
	English: present		
D1	is		
D2	is		
D3		should be clamoring	
D4	is		
D5	deserves		
D6	deserve		
D7	deserve		
D8	should be compensated		
D9	is		

The tense error in translation D4 is due to breaking the *kana* incorrectly. しかるべき (*shikarubeki*) means *appropriate*, while しかる + べき (*shikaru + beki*) would mean *should be angry with*.

Table 5.38. General passage D, sentence 3: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 3		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 提案する (non-past) English: present		
	Acceptable	Not Acceptable	そこで (<i>therefore</i>) (explicit)
D1	is		That is the basis (X)
D2	suggest		O
D3	am proposing		O
D4	propose		O
D5	is proposed		O
D6	propose		O
D7	propose		O
D8	propose		O
D9	propose		O

O indicates an acceptable solution; X, an unacceptable solution.

While all the translations had the correct tense, translation D4 was the only one of the nine to use the editorial *we* and so to reproduce in English the usual style of this type of text.

Table 5.39. General passage D, sentence 4: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 4		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: である (non-past copula) English: present		
	Acceptable	Not Acceptable	まず (<i>first</i>) (explicit)
D1	must be		O
D2	is		O
D3	must switch		First of all (X)
D4	is		O
D5	is		O
D6	is		First of all (X)
D7	is needed		O
D8	is		O
D9	is		To begin, (X)

O indicates an acceptable solution; X, an unacceptable solution.

There were no tense problems in the translations of this sentence. However, the choices to use anything other than *first* to translate まず are interesting, particularly since the following paragraph begins with そして, *then, next*, suggesting that the choice was made at the sentence level and not in view of the overall structure.

Table 5.40. General passage D, sentence 5: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 5		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 行 う (non-past) English: some sense of <i>should/ought</i>		
			ま た (<i>moreover, in addition</i>) (explicit)
D1	it is important to		O
D2	we must place		O
D3		will be followed	O
D4		will be made	O
D5	it is proposed to use		O
D6	must also strive		O
D7	should be		O
D8	should be in		O
D9	should be implemented		O

O indicates an acceptable solution; X, an unacceptable solution.

The cohesion marker here is quite clear, and all of the translations offered an acceptable solution. However, translations D3 and D4 did not follow the pattern set up by the *また (moreover, also)*, which is that “we must do the things in sentence 4 and also should implement the items in sentence 5”.

Table 5.41. General passage D, sentence 6: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 6		Other Cohesion Marker
	Acceptable	Not Acceptable	
	Japanese: 取 り 入 れ ら れ て い る (passive progressive) English: can be either progress or perfect		
			–
D1		is used	
D2		can incorporate	
D3		will be made	
D4	has been adopted		
D5		is accomplished	
D6		done	
D7	has been adopted		
D8	is already being used		
D9		is now adopted	

Translations D5 and D6 had the correct tense, but had problems with syntax.

More incorrect tenses appear here than in the previous sentences. Translation D3 continued to use the future tense, while Translation D2 seems to have taken a passive form for the potential, which will not work in this context.

Table 5.42. General passage D, sentence 7: tense and other cohesion markers

Translation No.	Tense of Main Verb Sentence 7		Other Cohesion Marker
	Japanese:できよう (non-past presumptive) English: some sense of <i>should/ought</i>		
	Acceptable	Not Acceptable	そして (<i>next</i>) (explicit)
D1		will avoid	omitted (X)
D2		will make possible	Furthermore (X)
D3		shall be improved	omitted (X)
D4		can be avoided	And (X)
D5		can be prevented	Moreover (X)
D6	should be able to avoid		Further (X)
D7		should be, are aimed at	omitted (X)
D8		can save	Also (X)
D9		I think it will be possible	Moreover (X)

O indicates an acceptable solution; X, an unacceptable solution.

Translation D6 had the correct tense but the translation was problematic in other ways. None of the translations linked *そして* (*soshite*) back to *まず* (*mazu*).

While the tense structure of this sentence was relatively simple and nearly all the translations used correct tenses in the first and second paragraphs, the syntax of the extremely long sentence making up the third paragraph posed a difficulty in several of the translations. In addition, most candidates ignored the presumptive/hortative form of the final verb by using the modal verbs *can* or *will*. This could be due to a variety of factors, including limited grammatical knowledge of how the *-you* form of verbs functions in Japanese or inability to come up with an appropriate English equivalent as well as a tendency to leave off Japanese endings in English translation because they are often seen as being unnecessarily wordy. However, there are signs in each paragraph that most of the translators did not recognize a standard pattern of editorials, and this could have played a role in problems with the final verb. Only one translation (D4) used the editorial *we* in the first paragraph. The very straightforward *まず* (*mazu*, *first*) was made more complicated in two cases by being rendered “first of all” and as “to begin” in a third. The *そして* (*soshite*) at the start of the third paragraph was not translated by anyone in a manner that conveyed the meaning of the second and third paragraphs: “First we must do this, then we should do that”.

5.4.2 General passage D: summary tables and facility values

Tables summarizing the tense and implicit/explicit cohesion mark errors with facility values and total number of errors per translation are given below.

Table 5.43. General passage D, overall tense errors

Translation No.	Sentence No.							Total No. of errors
	1	2	3	4	5	6	7	
D1	O	O	O	O	O	X	X	2
D2	O	O	O	O	O	X	X	2
D3	O	X	O	O	X	X	X	4
D4	O	O	O	O	X	O	X	2
D5	O	O	O	O	O	X	X	2
D6	O	O	O	O	O	X	O	1
D7	O	O	O	O	O	O	X	1
D8	O	O	O	O	O	O	X	1
D9	O	O	O	O	O	X	X	2
Item facility	1.0	0.89	1.0	1.0	0.78	0.33	0.11	

O indicates an acceptable solution; X, an unacceptable solution.

The facility values show that sentence 7 was substantially more difficult than the other sentences, which is in line with the idea that longer sentences are more difficult to read and process. A facility value this low also suggests that the passage may not have been at an appropriate level for this population, at least in a test setting. The two facility values at 1.0 point to tense not being a useful test item for passages that do not have an argument extending over a period of time.

Table 5.44. General passage D, overall cohesion marker errors*

Translation No.	Cohesion Markers			Total No. of errors	
	そ こ で (therefore, accordingly) (sentence 3)	ま ず (first) (sentence 4)	また (moreover, in addition) (sentence 5)		そ して (next) (sentence 7)
D1	X omitted	O	O	X omitted	2
D2	X furthermore	O	O	X furthermore	2
D3	X omitted	X (First of all)	O	X omitted	3
D4	X and	O	O	X and	2
D5	X moreover	O	O	X moreover	2
D6	X further	X (First of all)	O	X further	3
D7	X omitted	O	O	X omitted	2
D8	X also	O	O	X also	2
D9	X moreover	X (To begin with)	O	X moreover	3
Item facility	0.0	0.67	1.00	0.00	

O indicates an acceptable solution; X, an unacceptable solution.

*All cohesion markers were explicit.

The item facility range was extreme (0.0-1.0) for cohesion markers, with all translations successfully rendering the straightforward *また* (*mata, moreover*), while none managed to make the connection between *first* and *next* (*mazu* and *soshite*), which could be considered an implicit cohesion marker. In addition, none of the translations used an acceptable word to represent *sokode* (*therefore, accordingly*) and so signal that the situation just described in the first paragraph resulted in the proposals given in the second and third paragraphs.

Unlike the facility value of 0.11 for the final verb of the extremely long sentence, I do not think that the facility values of zero for *sokode* (*therefore, accordingly*) and *soshite* (*next*) mark this passage as unsuitable for a population in which the members potentially have a degree in translation or a university degree in something else with two years of translation experience. Both words are very common expressions in Japanese, and a definition that would have been appropriate in context could have been taken from *Kenkyusha*. The only explanation I find at this stage is that they were translated in a state resembling autopilot in which the first standard English equivalent that came to mind was used, with no attempt to fit a specific equivalent to context. However, that is very speculative.

The three translators who omitted the cohesion markers at the beginning of the first and second paragraphs were consistent in that they omitted both of them. To speculate again, this might have been due to a tendency to omit “unnecessary elements”. Unfortunately, this passage does not have any of the verb endings like those seen in general passage C against which one could verify this.

5.4.3 General passage D: problem points

This passage contained only one modified noun and one passive verb.

Table 5.45. General passage D, problem points

Translation No.	Modified noun*	Syntax of sentence 1	Passive construction	Total No. of errors
	日本国家 の優秀な 人的資源 である国 家公務員		取り入れら れている (sentence 6)	
D1	O	O	X	1
D2	O	O	X	2
D3	X	X	X	3
D4	O	O	O	0
D5	X	X	X	4
D6	O	O	X	1
D7	O	O	O	0
D8	O	O	O	0
D9	X	O	X	2
Facility value	0.67	0.78	0.33	

* See 4.2.2.1 for translation.

O indicates an acceptable solution; X, an unacceptable solution.

The modified noun was a 20-character string with the head noun *civil servants*, preceded by the modifier string, *who are superior human resources of the Japanese nation*. The head noun is marked as an object, and two solutions to the syntax of the sentence in English are either to turn the modification into a relative clause following *civil servants* or to break the Japanese into two sentences: “Employees of the national government are outstanding human resources of the nation of Japan. Slandering and maligning them does not benefit the people of Japan in any way” (translation C1; translation C9 also adopted this approach).

The problems with the modified noun were as follows:

The excellent human resources of the Japanese government official (D3)

A national public official, considered to be a superior human resource of nation of Japan (D5)

The public servants of Japan are an excellent human resource (D9)

Translation D3 has taken “human resources” as the head noun and also omitted “of the Japanese nation”. Translation D9, which was syntactically solid, omitted the same phrase. The unnecessary “considered to be” was added in translation D5. The facility value for the set of translations was 0.67, near the top of the set difficulty range. The syntax had a higher facility value (0.78) because translation D9 simply omitted a phrase and did not have trouble with the syntax of the modified noun or the sentence overall.

The passive verb in sentence 6 had the extra complication that the *-teiru* form can be either progressive or perfect for verbs of motion or change, which include 取り入る (*toriiru*, *to incorporate*), so the translator is forced to make a choice. Translation D1 used a past passive and omitted the sense of *has adopted/is adopting*, opting for simply “is used”. That conveys most of the meaning, but loses a bit of the sense of contrast. However, whether to mark this as an error (which I did) is a subjective call. Translation D9 did the same thing, but with the addition of now: “is now adopted”; that can be marked incorrect without resorting to subjective judgment. Translation D2 used the present potential (“can incorporate”), which does not work here. Translation D3 used a future passive and had both the subject and object wrong. Translations D5 and D6 also had trouble with the subject and object and used the present tense.

5.5 Summary of results

5.5.1 Summary of tense results

A total of 33 final verbs were evaluated. The facility values for 13 verbs fell in the acceptable difficulty range of 0.25-0.70. One verb had a facility value suggesting an item that was overly difficult for the population being tested, while 19 verbs had facility values greater than 0.70, marking them as easy for this population and so as items not providing any useful information. Five of the verbs in the “easy” group had facility ratings of 1.0.

The verbs with the lowest facility values were:

担うことになる(*ninau koto ni naru*, base meaning, *to take on, bear*) (general passage B, sentence 5b (facility value 0.25)

なしている(*nashiteiru*, base meaning, *to become*) (general passage C, sentence 2) (facility value 0.25)

いう(*iu*, base meaning, *to say*) (general passage C, sentence 8) (facility value 0.25)

できよう(*dekiyou*, base meaning, *can, to make possible*) (general passage D, sentence 7) (facility value 0.11)

The most common error in translating the first three verbs was to use the present tense. This was probably due to not understanding the use of the narrative present in Japanese, which will be discussed in the next chapter. The fourth verb (*dekiyou*, base meaning, *can, to make possible*) needed to be translated as an English conditional; most translations used *will, can, or shall*. There is a possibility that the length of the sentence contributed to the low facility value for this tense decision.

All the verbs with a facility value of 1.0 were non-past in Japanese and could be translated with an English present tense. It is hardly surprising that this was easy for the test population.

Twenty-seven of the incorrect verb choices used the present perfect. All of these were in general passages A and B, which involved changes over time.

5.5.2 Summary of other cohesion marker results

The four general passages contained very few implicit cohesion markers. There were two in general passage A, plus the combination of *first ... next* in general passage D. The markers in passage A had facility values in the set difficulty range of 0.27 and 0.64, while the combined marker in passage D had a facility value of zero (as did the explicit marker *next*). Only one of the explicit markers in general passages B and C was in the set difficulty range: *sochite (therefore, accordingly)* in passage B, sentence 6 (facility value 0.58). However, correct resolution of this marker appeared to contribute, to some degree, to the successful resolution of the syntax of sentence 6.

The cohesion marker results from general passage C—only three markers with a facility value of 1.0 for each—suggest that passage selection should focus on finding texts with such markers and preferable a mixture of implicit and explicit markers. However, the unexpected difficulty with two common explicit markers in general passage D indicates that such markers are worth reviewing when evaluating a both a passage and a translation.

5.5.3 Summary of problem point results

5.5.3.1 Modified nouns

The modified nouns were reliable indicators as to whether the overall syntax of the sentence had been resolved correctly. In the second sentence of general passage B, the

syntax facility value (0.45) was lower than either of the two modified nouns in the sentence, but that was probably due to the presence of two long strings of modification in the sentence. The facility values for syntax in passage B, sentence 2, and passage D, sentence 1 (0.92 and 0.78, respectively) were higher than those for the modified noun (0.83 and 0.67, respectively), but both modification and syntax were easy in both sentences.

The facility value of 0.25 for sentence 6, general passage B, can be seen as confirming that the passage was at the correct level of difficulty for the test population.

5.5.3.2 *Passive constructions*

Of the seven passive constructions scored, five were within the set range of difficulty (facility values 0.25-0.70). In general, the score for the passive verbs mirrored those for tense. However, in sentence 3, general passage A, the facility value for tense was 0.82, while that for the passive construction was 0.45. In this instance, examining the passive construction identified problems with subjects and objects.

6. Results: close reading of the general passages

This chapter contains a close reading of the translations in light of the objective results provided by the facility values. The initial discussion will be done sentence by sentence in order to consider the cohesion markers and problem points together as well as the errors in individual translations in an attempt to identify error patterns and source or target language features that may have triggered them.

The second part of the discussion uses the facility values for tense, other cohesion markers, and problem points to select a few translations for further analysis to discover what types of problems each marker indicates, whether these overlap, and whether they allow any statements about translation quality or predictions of future performance.

6.1 Discussion: general passage A

6.1.1 Discussion of individual sentences

6.1.1.1 General passage A, sentence 1

The first sentence in general passage A had a relatively low facility value (0.45) for tense, despite the Japanese verb (高くは なかつた, *takaku wa nakatta*, was not high) being an uncomplicated past tense showing completed action. Four of incorrect tenses were the present perfect (the level ... “has not been high”) and two were present (the level ... “is not high” / “may be described ... as not high overall”). All of these choices were paired with either omitting the implicit cohesion marker *koremade* (up to that point or up to the time of Masao Maruyama) or translating it as “until now”. What appears to have happened is that the tense was taken from *koremade*. This time expression can be—and usually is—treated as a lexical unit (Martin 1975: 1067, 1070) meaning *up to now*, and *kore* by itself is cited in *Kenkyusha* as meaning *now, this time*. However, it is also a proximal deictic demonstrative referring to something close to the speaker (Martin 1975: 1067-1068) or which has been referred to previously and is now familiar to the speaker and listener (or, in this case, writer and reader) (Yoshida 2011: 35-52), which here was mostly likely something brought up in the previous paragraph,

but that can be resolved in light of the description of Murayama in the fourth sentence. As Halliday and Hasan note, using a demonstrative reference is “essentially a form of verbal pointing” with *this* (or *kore* in Japanese) pointing at something nearby (1976: 57). However, the majority of the translations used the most common meaning and seemed to not have viewed the term in the context of both the passage and the verb tense in the same sentence, resulting in a low facility value of 0.27.

The modified noun in the first sentence was a problem in only two of the translations: in translation A3, the wrong meaning of *ni yoru* (here, *by*) was selected; the modification string was broken up in translation A10. The first error, selecting the wrong meaning, echoes the mistaken interpretation of *koremade*, signaling that choosing an inappropriate meaning out of a selection of possibilities may be a common source of mistakes.

While it is impossible to determine exactly why a given erroneous choice was made, the mistakes with tense and the implicit cohesion marker in sentence one can be viewed as being connected with ILR reading level 3 features such needing to interpret the text in its overall context. On the other hand, the modified noun in this sentence could be resolved without recourse to inference or context. Resolution cannot be linked to a specific text feature outside of standard grammar; the noun string had a correspondingly high facility value (0.81).

6.1.1.2 General passage A: sentence 2

The verb structure in sentence 2 was substantially more complicated than the final verb in sentence 1, but the facility value was the same (0.45). A *-tekuru* form indicating change moving toward the speaker and in the past tense was followed by a conjunctive particle with a present tense copula, with this latter set as an implicit cohesion marker. Two of the translations (A1 and A5) included the implicit link back to the previous sentence, but took the present tense from the final copula. Conversely, three of the translations with correct past tense ignored the causal ending. This suggests difficulty in reconciling the past and present tense elements of the verb construction.

As noted in Chapter 5, the first modified noun in sentence 2 could be resolved without needing to infer which elements were modifying which and had a correspondingly high facility value of 0.73. As in sentence 1, the problems in translation were breaking up the modification string and using a noun other than the Japanese head noun as the modified noun in English.

The second modified noun was more difficult (facility value 0.45), perhaps because it did require some degree of inference to resolve. The initial noun followed by a genitive marker (*of political scientists*) directly modifies the head noun (*interest*) with the remainder of the phrase explaining what the political scientists were interested in. This problem point can be considered a marker of inference both because of the need to use some degree of interpretation to resolve the modification and because the topic of the passage being translated is the change of focus of Japanese political scientists over time, information against which decisions could be checked if the text was being seen as a unit.

In this sentence, tense, the implicit cohesion marker requiring incorporation of the null pronoun, and the second modified noun can all be considered elements connected to ILR reading level 3 features of following an argument and relating ideas. The tense mistakes also suggest some degree of cognitive dissidence in the presence of a verb construction combining past and present tense.

6.1.1.3 General passage A: sentence 3

Resolving the tense of sentence 3 was easy enough to be outside the set range for facility value difficulty (0.73), despite the need to change the final verb into an adverb and to use a present tense verb (*okonau, to do, to perform*) as the main verb in English in the past tense (*was frequently done as*). The unacceptable tense choice was the present perfect (translations A1 and A5); both translations had the same mistake in sentence 1.

Examining the passive construction revealed difficulties with the syntax of the sentence. The stumbling point seems to have been the particle *ka*, which can be either a “conjunctive particle” (*or*) or a question particle, when it can often be represented by a question mark in English (Kaiser et al. 2002: 197-198). In sentence 3 it appears between the words for *hobby* and *political criticism*. Perhaps because *ka* is most frequently used as a question particle, although its use as *or* is not rare, some of the translations had wording suggesting an attempt to incorporate some sense of questioning or doubt:

Japanese political research, which may have been a hobby of such scholars, was frequently conducted as political commentary or cultural criticism. (A3)

It might have been because the study of Japanese politics was these scholars’ hobby, a lot of it consisted of political criticism and critique of civilization. (A7)

Perhaps it was because research in Japanese politics was more of a pastime for these scholars, but for whatever reason, much of their work came in the form of political commentary and cultural critique. (A11)

This is another case in which using the most readily available meaning of a word is leading people down the garden path. There is no sense of *because* in the source text; its addition in translations A7 and A11 seems to have been part of the attempt to use *ka* as a question word.

Tense choice in this sentence cannot be linked to a decision based on inference, but the correct choice is required in order to maintain cohesion.

6.1.1.4 General passage A: sentence 4a and 4b

This was the longest sentence (120 characters) in the passage. Sentence length is often considered a feature of text difficulty in readability indices (Jensen 2009: 63-69; Mishra et al. 2013), but the explicit cohesion marker facility value (0.73) indicated that the majority of the translators in the test population did not find this sentence difficult to parse. That may be because the syntax is straightforward and clearly marked within the sentence and the break point in the middle is obvious.

The final verb of sentence 4b has the same form as the final verb of sentence 3 with the same need to change the conjugated adjective into an adverb and use the preceding verb as the main verb. As in the earlier sentence, there were very few problems with this, and the facility value showed that this was not difficult for the test population (0.82).

6.1.1.5 General passage A: sentences 5 and 6

Sentences 5 and 6 had no implicit or explicit cohesion markers or problem points. The Japanese and English tenses were both present, so no inference was needed. In addition, both sentences included a time expression (*gokusaikin*, *extremely recent* in sentence 5; *saikin*, *recent* in sentence 6). As these factors would suggest, the facility values for tense were very high (0.91 and 1.00, respectively).

6.1.2 Discussion of individual translations

When considering individual translations, it is important to remember that the goal of the tense, implicit/explicit cohesion markers, and problem points is to identify source text features that contribute to overall passage cohesion and that require following an argument and some degree of inference to resolve as well as items that can be objectively scored as acceptable or unacceptable. This contributes to creating a difficulty profile of a passage for a defined population based on the set items and so, ideally, to the possibility of creating short-passage tests in which items that can be objectively scored are linked to text features of interest.

Because the testing and evaluation of translations in professional settings is the ultimate area of interest, it is worth while to also look at some translation products to examine what the markers reveal about individual translation products. In doing this, I want to note that the markers are geared more toward looking for failure—the lack of coherence—than success and also that a cohesive translation with appropriate syntax may be unusable for other reasons, such as grammatical errors and poor word choice.

The translation with the highest number of unacceptable choices was translation A3 (10 mistakes). The majority of errors occurred in resolving the problem points, which were all incorrect. Three of the cohesion markers were mistranslated, but there was only one tense error (in sentence 1). This suggests that the translation will have both cohesion and syntax problems. Because the items of greatest difficulty cluster in the first two sentences, they should provide sufficient information about whether the text is cohesive and coherent. The first two sentences of translation A3 read

According to Japanese researchers, the level of Japanese political research up until now has consisted of a small amount of prominent research, but as a whole, it is not high. The source of knowledge was limited to ancient academic traditions from abroad and restricted by the Tenno system, and furthermore, prominent political scientists mainly concerned with modernization and democratization were driven by the political research from the politically advanced nations.

In this case, the selected markers have identified a text exhibiting lack of cohesion and problem with syntax. If asked to predict whether the translator could

handle documents of a similar reading level, an evaluator would have strong grounds for saying, “No”.

Translation A4 had errors across all the categories with three tense errors, two cohesion marker errors, and three problem point mistakes. These numbers suggest the presence of cohesion and syntax problems. Again, this is confirmed by the first two sentences:

When it comes to the level of Japanese political studies carried out by Japanese researchers, although there have been some excellent studies, on the whole the level has not been high. A combination of an ancient scholastic tradition that the source of learning is overseas, restrictions made by the Japanese Emperor system, and, furthermore, the modernization and democratization of political scholars had set the best political scholars rushing to study the politics of politically advanced countries.

This is a text that could not be used without extensive editing.

The text with the least errors was translation A8, which had one tense error, one cohesion error, and no problem point mistakes. The first two sentences read:

Up to now, the level of the research on Japanese politics by Japanese researchers was generally not high in spite of a few exceptional cases of high-quality research. This is because outstanding political scholars were encouraged to research the politics of politically advanced countries due to traditional scholarly thinking that the source of learning is in foreign countries, restrictions by the Imperial Household and the interest in modernization and democratization by political scholars.

The tense pattern of the second sentence has faithfully reproduced that of the Japanese verbs. That, coupled with the “up to now” in the first sentence create a situation in which the two sentences do not fit together in a cohesive manner. There are also internal cohesion problems in the second sentence, with verbs in both the past and present tense.

At the same time, this translation illustrates the limitations of pegging translation performance to a descriptor of reading ability alone. From reading the translation, I

would say that the translator has understood the source text. However, the person does not exhibit a corresponding ability to create a cohesive target text. Nonetheless, while this is a translation that would require editing, it could be brought up a usable level with a reasonable amount of effort.

6.2 Discussion: general passage B

6.2.1 Discussion of individual sentences

6.2.1.1 General passage B, sentence 1

The tense of the first sentence of general passage B was close to the upper limit set for the item difficulty range for this population (facility value 0.67), with most of the translations matching the idea of a fixed period of time with the verb expressing continuing action in the past. However, three of the four incorrect tenses were again present perfect, a trend that was seen in tense errors in general passage A. Therefore, these sentences merit a closer look.

Throughout the period of Japan's high growth, the forms and functions of the family, the basic unit of society, have changed. (B2)

The family, a fundamental unit of society, has undergone changes in form and relationships through Japan's high growth period. (B3)

Due to Japan's rapid economic growth, the family structure and family relationships, the basic unit of society, have changed. (B10)

The language use in the three translations suggests that they are all by native English speakers. The result of using the present perfect in translations B2 and B3 is a mismatch between a completed action and an ongoing period and additionally provides no suggestion that the period in question was in the past. In translation B10, the tense is correct in the sentence as written, but reveals that the sentence is a mistranslation (*tsuujite, throughout, via* has been mistakenly translated as "due to"). While the examination of tense here has revealed ways in which the translations have gone off track, it does not explain the choice of the present perfect. Translation B2 is particularly

puzzling, since the present perfect was used to translate almost every final verb in the passage. This resulted in all the verb tenses being incorrect and a text with no cohesion.

6.2.1.2 General passage B, sentence 2

In sentence 2, two of the translations with mistakes related to the explicit cohesion marker *this period* also made tense mistakes (translations B4 and B6), suggesting that understanding both elements helped in resolving the sentence. Translation B12 omitted the marker, reducing the cohesion between the sentences. Two of the three incorrect tense choices were the present or past perfect. In translation B4, the past perfect was appropriate with the incorrect “not long ago” for *kono aida, this period* (“Not long ago, ... gender-based responsibilities had been fixed”), but both choices point to a failure to connect this sentence with the previous sentence, where *Japan’s period of high economic growth* is explicitly stated.

The passive verb problem point had the same facility value as the tense marker (0.67), which is hardly surprising since they are scoring nearly the same item. Translation B8 had the tense and passive voice correct, but omitted any sense of “became”.

There were problems with the modified head noun in two translations. In translation 6, the modification string was broken up, which resulted in distorting the syntax of the sentence (“Within a working family which increased during the period, a husband worked outside his home while a wife took care of household matters and nursed children, thus, securing the gender-based specialization.”) The head noun (*labor*) was omitted in translation B11, but this did not affect the syntax of the sentence. Despite the need to add verbs to this phrase in English (“the husband worked outside the home”, “the wife was responsible for housework and raising children”), the facility values were quite high (0.82 for the modified noun; 0.92 for sentence syntax), showing that this problem point was easy to resolve for this population. The modification can be resolved without resorting to inference or connections to other parts of the passage and the information matches a situation the translators have most likely observed.

6.2.1.3 General passage B, sentences 3 and 4

Three translations (B2, B4, and B7) have the same tense errors in these sentences, which is again the incorrect use of the present perfect. The Japanese tenses were the straightforward past and past passive. The present perfect tense choice destroys text

cohesion and strongly suggests sentence-by-sentence translation without reference to the initial time reference which signals that the changes being discussed took place during “Japan’s period of strong economic growth”. For example, translation B4, which also omitted the contrastive sense of the explicit cohesion marker, reads, “However, with the increase of couples in which both work, gender-based divisions of labor within the home are being reconsidered, and new efforts have started to attempt to make the relationships of couples more equal.”

The error in the passive construction in the fourth sentence was use of the present perfect, which ruined cohesion: “Significant changes have also appeared in the relationships between family and society” (translation B7).

6.2.1.4 General passage B, sentences 5a and 5b

This sentence has a *ga, however*, in the middle and so easily divides into two sentences in English, resulting in two final verbs. It was the most difficult for the test population (facility values of 0.33 and 0.25 for 5a and 5b, respectively). This is almost certainly because both Japanese verbs are in the *V-ru* present form, resulting in an English present in the majority of the translations. Translation B11 had the correct tense in 5a, but changed to the present tense in 5b, despite not breaking the sentence (this might have been a typographical error):

Many of the young nuclear families and single households that were concentrated in the cities were naturally salaried-worker families, but unlike farming households, production and consumption were separate due to their lack of a means of productions; and the residential areas come to serve the function of consumption centers.

Despite the shifts from past to present tense in the Japanese, the source text is cohesive, since tense is not the cohesion marker in Japanese that it is in English. In his book on tense in modern Japanese, Matsuo Soga notes that:

The use of the non-past form *V-ru* in reference to the past is frequent both in conversation and in writing. ... Notice that in translating, it is normal to change Japanese non-past tense forms to English past tense forms. [... because] idiomatic English normally requires so-called tense agreement. Although like

Japanese English certainly uses the ‘historical present,’ we can say that the degree of temporal shift is one of the important differences between Japanese and English. (1983: 46)

Because tense is used so differently in the two languages, the translator is forced to rely on context and knowledge of target language conventions to choose the correct verb form. That doing this was difficult for this population is clear from the very low facility value of 0.25 for sentence 5b. Although I have not seen this validated anywhere, I suspect that tense changes are more frequent in Japanese texts at higher reading levels and so require a greater amount of dependence on context in translation.

Three of the translations did not represent *ga*, *however*. Although all three also used an incorrect present tense, I do not think that resolving tense and the cohesion marker in this sentence are related. Also, *ga* expresses only “a weak contrast between two clauses” (Kaiser et al. 2002: 152), so that leaving it out is usually not a major omission. That is particularly true in this sentence, as long as 異なつて (*kotonatte*, *unlike*), which appeared in the first part of 5b was somehow represented. However, omitting the particle did result in a run-on sentence in translation B10:

It is natural that most of the young nuclear families and single-member households that are concentrated in cities are working families, and in contrast to agricultural families, production and consumption are separated in working families that do not own the means of production, leading only consumer functions to be carried out in residential areas.

Among the interesting problem here is the “it is natural” at the beginning of the sentence. *Touzen* can mean *of course*, *natural*, or, *naturally*, depending on context. In this sentence it is part of a seven character string: 当然労働者家族 (*touzen roudou-sha kazoku*), literally, *of course/naturally working families*. This string of the same character type in the Japanese writing system was identified by Tateishi et al. (1988) as a reading difficulty, and it appears to have been one here. In this string, *touzen* is an adverb (*naturally* or *of course*), but perhaps because there is no obvious adverb marker, the translator has taken it as a noun and separated it from the verb it is actually modifying.

6.2.1.5 General passage B, sentence 6

This sentence was in the past tense in Japanese and so the tense was easy for this population to resolve (facility value 0.75). The incorrect choices were again the present perfect. The most common explicit cohesion marker error was omitting the marker; this appears to have been completely independent of tense choices. However, all the translations with explicit cohesion marker errors also had problems with the modified noun string in this sentence. Because noting that the conditions described in sentence 6 were the result of the situation described in sentence 5, which *soshite* (*thus* or *therefore*) does, representing this marker not only maintains cohesion between the sentences but also helps to resolve the modified noun. As a result, mistakes with this marker may have contributed to the low facility value (0.25) for the modified noun. However, as will be noted below, there were other factors contributing to the difficulty, making it impossible to point to a causal connection with the cohesion marker.

The errors with the modified noun were as follows:

in the consumptive market that had raised the degree of social dependency as such, i.e., the domestic labor market, (B1)

the areas of consumption with heightened reliance on society, that is places with housework, (B2)

In this way, in relation to this place of consumption with an increased societal dependency—that is, this place of household duties and labor (B3)

With consumption having increased the degree of social dependence in this way, specifically for household chores, (B4)

for these consumption regions of high social dependency, that is regions of domestic labor. (B5)

And to the places of consumption which has become highly dependent on society, that is, to the places of household matters, (B6)

Therefore, frozen foods, disposable goods etc. have been introduced and automation of washing machines and vacuum cleaners has made quick progress to combat this need for consumption which has increased social interdependence in other words to meet the demand for performance of household chores. (B8)

Thus, in these consumption centers or working family residential areas where there was a high social dependency, (B10)

Under these circumstances, frozen foods and other throw-away consumables have been incorporated into the lifestyle in quick succession in order to serve the needs of the sites of consumption that have a high degree of social reliance—in other words, the working household. (B11)

While the above phrases can be analyzed in various ways, a combination of the omission of *soshite* and not understanding the antecedent of *konoyou-ni* (*in this way*) are part of the problem in translations B1, B2, B4, B5, and B6. Translation B3 has *konoyou-ni* modifying the verb meaning *continuously flowed into* rather than the phrase of which it is a part. Translation B8 has some of the modifiers, but not the head noun. *Konoyou-ni* was omitted in translation B11, as was the sense of *household labor*. *Household labor* (which has the sense of “housework” or “household chores”) is also missing from translation B12, which has replaced it with “working household”.

An addition factor appears to be *ni taishite* (*toward*), the preposition immediately after the head noun. This was translated as “in relation to” in translation B3 and as “to combat” in B8. Both errors seem to have been triggered by using other possible meanings of the *kanji* 対 (*tai*), which has a range of meanings, including *versus*, *opposite*, and *to* (in the sense of a ratio).

All of the translations with mistaken modification also had problems with sentence syntax overall.

6.2.1.6 General passage B, sentence 7

Half of the translations had an incorrect tense in this sentence, with four of the mistakes being the present perfect and the fifth in the present tense. It is odd, and at this point, inexplicable, that most of the translations took an English present tense from the Japanese present tense in sentence five, while five of them ignored the obvious past tense here. However, even without knowing what motivated the choices, the tense analysis does show that only one of the twelve translations (B5) met the requirements for tense cohesion in English.

6.2.2 General passage B: discussion of individual translations

As just noted, there were no tense errors in translation B5. The text had one cohesion error (*soshite, thus/therefore* in sentence 6) and an unacceptable rendering of the modified noun string in the same sentence. The omitted cohesion marker may have been significant in this case, since there are problems in both sentences 5b and 6 (the sentences with the lowest facility values), which *soshite* links:

However, unlike agricultural families, in the case of working families that did not have any means of production, production and consumption separated-these families assumed only the function of consumption at their places of residence.

Frozen food, disposable goods and the like appeared, and there was rapid progress in the electrification of laundry and cleaning appliances for these consumption regions of high social dependency, that is regions of domestic labor.

The bulk of the problems resulted from the ordering of the information. While that could be a target language writing problem rather than a source language reading issue, the distinction does not really matter in a non-pedagogical (i.e. professional) setting where the focus is on the product. What is interesting here is that no inferences about how the information might be connected appear to have been made at appropriate places in the text. For example, the ending of the verb *rikaisare* (were separated) is in the causative passive conjunctive form, implying that the lack of a means of production caused the separation between production and consumption. While the conjunctive ending can often be translated as *and*, here the sense of a result is more appropriate: *with the result that*. The choice of a hyphen in the translation suggests that this link was not made. In addition, the adverb *konoyou-ni* (*in this way*) has been turned into a demonstrative pronoun modifying “consumption regions” rather than “high degree of societal dependence”. The types of linkages being missed are typical of ILR reading level 3 and suggest that while this translator was successful in creating a coherent English tense pattern, the individual still might not have been processing how discrete pieces of information fit together.

Translation B2 is interesting in that *all* of the tense decisions were incorrect but *all* of the explicit cohesion markers were resolved correctly. This testifies to the

independence of tense decisions and explicit cohesion marker resolution. However, the use of the present perfect to translate five of the six final verbs is a mystery.

The profile of translation B9, with only two tense errors (sentence 5a and 5b) and one explicit marker error (omission of *shoshite, therefore* in sentence 6) coupled with no problem point errors suggest a very good translation, which reading confirms. Interestingly, though, even for an obviously skilled translator, the pull of the present tense Japanese verbs in sentence 5 was too strong to resist: “Unlike farming families, however, these working families do not possess the means of production so for them production and consumption become separated, and in the residential areas they shoulder only consumption-related functions” (B9).

6.3 Discussion: general passage C

6.3.1 Discussion of individual sentences

6.3.1.1 General passage C, sentences 1 and 2

The only tense error in sentence 1 (translation C3, “The seclusionism early in the Edo period is not again information orientation”) is interesting in light of the time phrase *at the beginning of the Edo period*, suggesting that the form of the Japanese verb had a stronger pull than meaning. In sentence 2, three translations used the present tense (C2, C3, and C4). Interestingly, this sentence was split into two sentences in C4, with the verb in the resulting first sentence in the past tense (“was [...] meant”), but that of the second sentence in the present (“forms”). This violates both the English cohesion pattern and the general Japanese rule that tense is taken from the final verb: “It was instead meant to manage control information. A deep understanding of the importance of information forms the foundation of national isolation” (C4).

There was an explicit cohesion marker (*that, it*) marker in the second sentence referring back to the first sentence, but this was not a problem in any of the translations.

6.3.1.2 General passage C, sentence 3

The verb construction in sentence 3 is rather complicated—a past potential followed by a presumptive copula. As noted earlier, the potential and passive forms are the same in Japanese and so must be resolved from context. However, in the construction here, the

could that would account for the potential in English can be subsumed in a *probably* representing the presumptive. Two of the translations (C2 and C4) used that solution, while C1 omitted any sense of a conditional (“the closed-door policy was the most appropriate means to ‘manage’ information”). Translation C3 chose both the present tense (“is”) and the awkward “likely to have been assumed as”. The results for the passive construction mirrored those for tense. Because of the presence of the presumptive copula, it is impossible to say whether the verb form was seen as passive or presumptive.

6.3.1.3 General passage C, sentences 4, 5, and 6

The time frame switches to modern society (近世社会, *kindaishakai*¹⁶) in the fourth sentence with a corresponding change to the present tense. The only tense error was in translation C1, in which *kindaishakai* was incorrectly translated as “late medieval society”; this mistake was probably a contributing factor. By sentence 5, all of the translations were using the present tense. Neither of these sentences had a cohesion marker.

The verb construction in sentence 6 includes the auxiliary *-beki*, indicating obligation and is then softened by the presumptive copula, which would result in a literal English translation of *should probably be seen as*. The three acceptable translations correctly omitted the copula, which functions as a way of politely softening the statement in Japanese, while the incorrect translation (C2) went too far in that direction and turned the statement into a direct assertion: “isolation was rather a type of ‘historical necessity’”. However, this rendition, which omits sense of *to view as*, is also associated with an error in identifying the correct subject of the sentence, which is, implicitly, the reader. We, as readers, are being told how we should view the closed country policy.

There were no mistakes with the straightforward explicit cohesion marker *moshiro* (*instead, rather*).

¹⁶ “近世社会” is sometimes used to denote the “early modern period”, but I think it is being used here in the broader sense of “modern society”. This interpretation is consistent with the definition given in *Kenkyusha* (Watanabe et al. 2003: 771).

6.3.1.4 *General passage C, sentence 7*

Sentence 7 has an explicit subject—modern historians—and that, coupled with the presence of *sudeni* (*already*), requires an English past tense despite the Japanese non-past verb. While three of the translations had the correct tense, only one managed to parse the verb construction correctly. The stated subject is important in resolving the passive construction as well as the tense because, as noted in Chapter 5, this sentence final element is not a true passive, but rather an expression meaning *it seems, it appears*. Translation C1 omitted any sense of hesitation (“Actually, contemporary historians have already adapted a view resembling this, albeit differently phrased.”), and C2 used the very literal “can be thought to have already adopted”. Translation C4 added a dummy subject that threw the meaning off: “it is thought that modern historians already have viewpoints similar to this”.

6.3.1.8 *General passage C, sentences 8 through 11*

Sentence 8, like sentence 2, has a facility value of 0.25 (although in a very small sample). The difficulty appears to have been the same as in sentence 2—a Japanese present tense that needed to be translated with something other than a simple present (in this case, the present perfect) in order to maintain cohesion in English. Sentence 9 has the same pattern, but a higher success rate. Neither sentence had a cohesion marker.

The agglutinative elements in the verbs, *-gachi* (*predispose to*) in sentence 10 and *-beki* (*ought to, should*), caused only minor problems, with the two mistakes being omission.

The passive construction in sentence 10 is also a causative, but the causative sense can be incorporated into that of *-gachi*: *predisposes us to think that*. Translations 3 and 4 failed to do this. Translation 3 reads, “It has often been perceived that...” and translation 4, “A common interpretation was that...”.

There were no mistakes in tense or with the explicit cohesion marker in the short final sentence. The tense pattern was Japanese non-past to English present.

6.3.2 *General passage C: discussion of individual translations*

If one were pre-testing a text for a short-passage translation test and obtained the facility value results shown for general passage C, the passage would probably be deemed unsuitable. While the two lowest values are acceptable predicated on a failure rate of

about 75% and with the idea of identifying the superior translators, having five facility values of 1.0 and three more at 0.75 suggest that the passage is too easy for the test population and that a number of the identified test items will not result in any useful information.¹⁷ Nevertheless, the sentence analysis in 7.3.1 did show that tense cohesion and passive constructions were difficulties for this very small population.

There were no head nouns with long strings of modification in this passage, with the result that none of the items being scored were connected with syntax.

Translation C1 had the lowest number of tense errors (two) plus two passive construction errors. The first of these, omitting the conditional sense in the verb construction in sentence 3 was the same as the tense error, so having the passive construction problem point did not add any information. This may be because no inference is needed to resolve the verb form. The error in the passive construction in sentence 7 was more interesting since the tense was correct, but the passive-like construction was omitted completely. Taken together, these two omissions might give a reviewer pause in terms of the translator's ability to capture nuance, but they do not provide information about reading level or maintaining target text cohesion. The tense error in sentence 4 is points to one of the two serious mistakes in the text: translating *kindai shakai* (*modern society*) as "late medieval society", which is good to have found even though it is not the type of error that is being addressed here. None of the markers found the other serious mistake, which was translating *no tame ni* as "in order to" rather than *because of*: "[...] born in the early stages of information management intensification of society, in order to address the situation in which this nation was placed" for "...that arose during the early period of society beginning to use and management information because of the situation in which Japan found itself." This can be viewed both as a mistake of misinterpretation in context and of selecting the wrong meaning for a frequently used word, something seen in the earlier passages.

Translations C2, C3, and C4 all have enough tense mistakes to signal difficulty in creating a cohesive text in English. The passive construction problem points add information about the degree of literal translation, but because none of the constructions required inference to be resolved, they are not connected with the reading level elements being studied here. What seems to be happening is that rather than identifying problems with the translations, the items set for analysis are pointing to problems with

¹⁷ As a note of caution, it is very easy to arrive at very low or very high facility values in a population of four translators.

the test passage, suggesting that it lacks a level of complexity that would require problem solving through inference.

6.4 Discussion: general passage D

6.4.1 General passage D: discussion of individual sentences

6.4.1.1 General passage D, sentences 1, 2, and 3

The modified noun string was the only item in this sentence to provide useful information. In translation D3, failure to resolve the modification pointed to overall problems with the syntax of the sentence: “Say what you want to libel the excellent human resource of the Japanese government official [sic], you can’t say that they are in some way profiting from the Japanese people.” In translation D5, it uncovered missing text: “There is no benefit in criticizing a national public official, considered to be a superior human resource of nation of Japan” (*citizens of Japan* is missing).

The tense error in sentence 2 (translation D3) is interesting in that it involved a mistake in breaking a string of *hiragana*. Tateishi et al. (1988) found long strings of any single type of Japanese character to be a reading difficulty, and the current string, with the preceding subject marker, is 12 *hiragana*. The resulting *shikaru* (verb meaning *be angry with, scold*) + *beki* (auxiliary meaning *ought to, should*) was forced into matching the context by being rendered as *should be clamoring for* (when the characters are parsed correctly, the phrase reads *it is appropriate, shikarubeki [appropriate] + dearu [copula]*).

There were no mistakes in tense in sentence 3. The single cohesion marker error (“That is the basis” instead of *therefore* in translation D1) is interesting because the marker has been turned into the subject of the sentence. Subjects are a particular challenge in Japanese to English translation because, as Martin, relying on data from the Japanese National Language Research Institute, notes, “[t]he frequency with which a subject is NOT explicitly stated ... may be as high as 74 percent of the sentences in a discourse” (1975: 185, emphasis in the original). Another solution was making the sentence passive (“Therefore, the following personnel reform policy is proposed.”

translation D5), while six of the translations used *I* as the subject. Only one (translation D4) followed the correct English convention for this type of article and used the editorial *we*. Translation D3 used *we* as the subject of the first sentence in the next paragraph, after using *I* in sentence 3. Unstated subjects are another form of cohesion in Japanese (Baker 1992: 185), which was broken by that choice.

6.4.1.2 General passage D, sentences 4, 5, and 6

There were no tense problems in this sentence. The verb construction is the noun *hitsuyou* (*necessary*) followed by a non-past copula: *It is necessary to*. That obvious solution was the most common choice. The proposed solutions to the explicit error marker were more interesting because in terms of overall passage structure, the *mazu* (*first*) at the beginning of this sentence pairs with the *soshite* at the beginning of the third paragraph. The pattern is typical of editorials in both languages: *this is the problem, to solve it we need to do first this, then that*. Changing the plain *first* to *first of all* (translations D3 and D6) or *to begin* (translation D9) breaks this pattern, particularly since the respective pairing are *First of all ... Even if* (*shoshite* omitted), *First of all ... Further*, and *To begin ... Moreover*.

The short (18 words) sentence 5 reads almost as an addendum to sentence 4. As a result, even though the verb is a plain non-past form, it is appropriate to carry over the sense of something being necessary, which most of the translations did, using *must* or *should*. Translations D3 and D4 both chose the English future tense, despite using the present tense in the previous sentence. This sentence also illustrated problems with the unstated subject. Three possible solutions were offered:

A dummy subject (e.g. “It is important”, “it is proposed”)

We (e.g. “we must place”)

Personnel or (e.g. “Personnel must”, “the appropriate personnel should be”, “personnel policies should be”)

In terms of cohesion, the second choice is the most disruptive, particularly if an editorial *we* was appropriately used in the first paragraph. Because the editorial is proposing changes to government personnel policies, *personnel policies* is the probably the best choice.

While all of the translations rendered the explicit cohesion marker *mata* (*moreover, in addition*) correctly, because of the use of the future tense in translations 3 and 4, the sense of *policies that do this and, moreover, that are necessary* was lost:

Moreover, the assignment of the proper person to the proper location and a double-track style of personnel management will be followed. (translation D3)
Also, appointments will be made placing the right man in the right position as well as supporting multiple career tracks. (translation D4)

The tense choice here, in combination with the cohesion markers, affects not only cohesion, but also the meaning.

As noted in Chapter 4, the verb tense in sentence 6 is tricky because the final verb *toriirerareteiru* (*is being incorporated/has been incorporated*) can be seen as either progressive or perfective. I think that the progressive *is being incorporated* makes more sense in the context of the national government being encouraged to take similar action, but would mark either choice correct. Two translations (D1 and D9) used a simple past passive and omitted the sense of *incorporated/adopted*, which reduced the impact of the sentence. The remaining three translations had trouble resolving the subject and object of the sentence:

Appointments will be made from regionally appointed officials and will not made by ministry but from a general pool. (Translation D3)
General employment as opposed to ministry-level employment is accomplished through employment of regional public officials. (Translation D5)
Recruiting is not by separate department but done collectively, accepting them in the local public service employee recruiting. (Translation D6)

Many of the problems in the above sentence appear to stem from not understanding some of the terms. This will be discussed in Chapter 8 while considering difficulties identified by previous studies.

6.4.1.3 General passage D, sentence 7

Because the major difficulty of this sentence—its length—was not set as a problem point, it will be discussed in the next chapter. However, one can assume that sentence

length contributed to the facility value of 0.11 for the verb and 0.00 for the explicit cohesion marker.

The final verb is in the presumptive form (or presumptive form of the hortative Kaiser et al. 2002: 191, 394), and means *it should be possible to* (*dekiyou*). That was correctly reproduced in only one translation (D6). Translation D7 had *should* but it was not connected with *be possible*. The incorrect solutions to the verb construction included “will avoid”, “will be avoided” (D1, D3), “will make possible”, “will enable” (D2, D8), “can be avoided”, “can be prevented” (D4, D5), and “I think it will be possible” (D9).

In this case, I think the tense errors and the cohesion marker errors were linked. As noted above, the cohesion markers set out the structure of the passage: first this needs to be done, then that needs to be done, which should make it possible not to pressure private industry or waste taxpayer money. If the second and third paragraphs are not linked by a clear expression of *soshite* (*next*), it becomes more difficult to see this connection.

An additional hurdle is the structure of the sentence. The translator is likely to run into serious problems with English syntax if they start at the top and simply work through to the end of the sentence, which appears to be what most people did. The target language text will flow easier if one starts with *by means of these policies* (i.e. starting at character 154 in the sentence; given that employment policies are the topic of the editorial, that would be the most cohesive choice) or with the final verb, *It should be possible to... by means of*. Translations D1 and D8 used the first structure and translation D2 the second.

The link between *mazu* (*first*) and *soshite* (*next*) could be considered an implicit cohesion marker that was missed or misunderstood by this test population.

6.4.1.4 General passage D, summary of sentence analysis

The analysis of individual sentences showed that the major problems in this passage were recognizing and reproducing structure and conventions of an editorial and the unstated subject, a cohesion marker not set for analysis, but one that requires inference. The most difficult verb was the final presumptive; properly representing that could be another indication of recognizing the format of an editorial. The modified noun problem point was not difficult for most of the population. The passive construction pointed to

further problems identifying the correct grammatical subject as well as to some difficulties in including all the information present in the Japanese construction.

6.4.2 General passage D: discussion of individual translations

The facility value profile for this passage includes three values of 1.0 for tense and one for cohesion markers. However, these are balanced by the tense difficulties in sentences 6 and 7 and by two cohesion markers facility value of 0.0. What should be noted though, is that scoring tense is being trialed as a way of objectively looking at cohesion and that while the facility values give us information about the population and passage, the total number of tense errors provides information about the translation and translator. Therefore, it should be possible to say that translation D3 is less cohesive than translation D8. A look at the two texts confirms this (only the first paragraph is given):

Say what you want to libel the excellent human resource of the Japanese government official [sic], you can't say that they are in some way profiting from the Japanese people. We should be clamoring for as much compensation as desired for people who puts [sic] all their energy towards the good of the country. For that reason I am proposing the following human resource policies. (Translation D3)

The citizens of Japan have absolutely nothing to gain from the baseless disparagement of their public employees who are a great resource for the country of Japan. Naturally, people who give their all for the benefit of the country should be compensated appropriately. Therefore I propose the following policies in regards to matters involving personnel. (Translation D8)

While translation D8 requires editing, D3 would need to be retranslated. This overall picture is further supported by both the number of cohesion marker errors (3 for D3 and 2 for D8) and problem point mistakes (3 for D3 and 0 for D8).

The tense pattern of this passage is fairly simple in that there is no progression over time or use of the historical present, and this probably contributed to the generally high facility values and makes the low values (0.33 and 0.11, respectively) in the last

two sentences even more interesting. The final verb in sentence 6 is a passive construction, but examination of the sentences suggest that failure to successfully convey the meaning of the verb may have been rooted in an inability to parse the four character phrase 省別採用 (*shou betsu saiyou*, *employment by ministry*) and the expression 一括採用 (*ikkatsu saiyou*, *collective hiring*). Neither of these was set as a problem point, since I was focusing on long strings of modification, but it is worth looking at how failure to resolve these short character strings affected treatment of the verb:

Integrated implementation without implementation by each ministry can incorporate the implementation of regional governmental officials. (Translation D2)

Appointments will be made from regionally appointed officials and will not [sic] made by ministry but from a general pool. (Translation D3)

General employment as opposed to ministry-level employment is accomplished through employment of regional public officials. (Translation D5)

Recruiting is not by separate department but done collectively, accepting them in the local public service employee recruiting. (Translation D6)

Viewed solely in terms of the passive construction that was being scored, translation D2 has made “integrated implementation” (i.e. *collective hiring*) the subject of an active verb rather than the object of a passive construction, perhaps partly motivated by seeing the verb as a potential rather than a passive form. Translation D3 also has an active verb with an explicit subject, which is a noun (*saiyou*, *hiring*, *employment*) marked as a location (i.e. followed by the particle *de*) in the source text. The same particle (*de*) was misunderstood in translation D5, where it is treated as making the means of action, a function it sometimes has. Translation D6 has gone off the rails by inserting a verb in the modification string preceding *collective hiring* (*collective hiring and not hiring by individual ministry*, a 12-character modification string) and then dropping the final verb.

In this sentence, looking at the passive construction identified modification difficulties that I would not have originally seen as being a problem. The modified noun set as a problem based on length turned out to be much less of a difficulty (the 12-

character string has a facility value of 0.56 as opposed to 0.67 for the longer modification). Unsuccessful resolution did not affect sentence syntax in one translation. In the other two translations, the problem was not using the Japanese head noun as the head noun in English (translation D3) and inserting an unnecessary verb in the modification string (translation D5).

Because the cohesion markers in this text explicitly set out the structure of the entire passage and not just relationships between sentences, it is striking that none of the translations reproduced them correctly in English. While in many of the translations, this does appear to be a sign that the translator was not working at the text level, it would appear to be less so in the case of a translation like D1 that had reasonably good cohesion and could be used with minimal editing. Other possibilities might be a lack of familiarity with editorial text conventions or a tendency to drop “empty words” with a corresponding loss of nuance.

6.5 Summary

Tense and cohesion markers tended to correlate with the need for inference more in the two passages in which there were changes over time, and mistakes with time expressions were linked to some degree with tense errors. This was particularly true of the one implicit (general passage A) time marker. Incorrect meaning choices for common function words such as *ka* (*or*), *ni yoru* (*here, by*), and *no tame ni* (*because*) was a recurring problem. Mistakes with modified nouns included breaking up the modification string and inserting pieces of it throughout a sentence and not keeping the head noun as the head noun in translation. There was a tendency to omit explicit cohesion markers and elements of verb constructions, which could lead to some loss of nuance and also disrupt cohesion. The use of the narrative present in Japanese triggered tense errors in English. Finally, an additional cohesion element—unstated subjects—was a problem.

7. Supplementary analysis based on earlier studies

In the conclusion to their 2002 paper, Hale and Campbell wrote that, “further empirically based research is necessary to map item types and to characterise the difficulties they cause” and that “this work needs to be done with multiple language combinations to test the universality of these difficulties” (Hale and Campbell 2002: 29). To respond to this request in a small way, this chapter will look at some items of difficulty that have been previously identified in Translation Studies through empirical research to see whether early findings can be confirmed here. None of the earlier studies dealt with the Japanese-to-English language pair.

The items to be examined are taken from Hale and Campbell (2002) and eye tracking studies of translation difficulty (Jensen 2009; Mishra et al. 2013). Throughout the discussion it should be remembered that *difficulty* in the earlier studies refers to subjective difficulty experienced by the translator in the form of cognitive effort while in this study it points to text elements and their resolution.

7.1 Items of translation difficulty in Hale and Campbell

Hale and Campbell define “text difficulty for the purpose of translation” as “a function of the cognitive effort required to process the item in question and convert it into the target language” (2002: 15). In an earlier study they suggested that counting the number of different translation solutions to a source text item was a possible way of assessing difficulty (Campbell and Hale 1999), but found this that this metric was unrelated to accuracy because the a group of translators could render the item in the same way, but incorrectly, or in a variety of ways that were all acceptable. In the 2002 study they found that there was “not a clear correlation between choices and accuracy” (albeit with caveats as to the definition of “accuracy”) (2002: 29). At the same time, they set “Official Terms, Complex Noun Phrases, Passive Verbs and Metaphors” as the source text items to be studied because these had generated a large number of variants in their earlier studies (Hale and Campbell 2002: 17).

Complex noun phrases in the form of head nouns with long strings of modification have been examined in the main analysis and found to be a difficulty in

terms of accurate representation of the Japanese modification in English. Why some of the phrases were more difficult than others will be discussed in Chapter 9.

Passive verbs have been covered in the results section and found to generally measure the same thing as the tense marker, although examining the entire passive construction, including the subject and object, also revealed interesting syntax problems in some cases.

The current set of passages does not contain any metaphors or official terms. However, it does have a number of culturally specific nouns that resulted in a variety of English translations. These will be examined in lieu of Hale and Campbell’s “official terms”. Most of the words fall into their category of “no equivalents, multiple options”.

7.1.1. Translation results for culturally specific terms

The terms scored were 天皇制 (*tenno sei, emperor system/imperial system*) from general passage A, 鎖国 (*sakoku, closed country*) and 情報化 (*jyouhou-ka, to become more information oriented*) from general passage C, and 一括採用 (*ikka saiyou, collective hiring*), 同期の者 (*douki no mono, people from the same class*), and 天下り (*amakudari, parachute into the private sector*) from general passage D. Of these, it could be argued that *jyouhou-ka* and *ikka saiyou* are not words specific to Japan, but I have included them because of the variants in translation and their close link to other culturally specific words in the passage.

Table 7.1. Facility values for culturally specific terms

	Culturally specific term					
	天皇制 (<i>tenno sei</i>)	鎖国 (<i>sakoku</i>)	情報化 (<i>jyouhou-ka</i>)	一括採用 (<i>ikka saiyou</i>)	同期の者 (<i>douki no mono</i>)	天下り (<i>amakudari</i>)
	O	O	O	O	O	X
	X	O	O	X	X	X
	X	O	O	X	X	X
	O	X	O	O	O	O
	O			X	X	X
	O			X	O	X
	O			O	O	X
	X			X	X	X
	O			O	X	X
	O					
Facility value	0.72	0.75	–*	0.44	0.44	0.11

O indicates an acceptable solution; X, an unacceptable solution.

Facility values were calculated with different denominators.

* Because the acceptability of this item could only be evaluated subjectively, no facility value was calculated.

Tenno sei resembles Hale and Campbell's "official term" (a word with a corresponding set expression in English) and eight of the translations used one of the two standard English versions: *emperor system* or *imperial system*. Two translators opted for transliteration ("Tenno system"), which is used in scholarly works, but could be considered a mistake here because the translation instructions were to "translate for a general reader". One translation had "Imperial Household", which is the translation of 宮内 (*kunai*) and not 天皇制 and so clearly a mistake.

There were three translations for *sakoku* (closed country): "closed-door policy", "national isolation" (twice), and "seclusionism". The last is incorrect since it is not an accepted English word. "National isolation" may have been the preferred choice in one translation because it obviates the need to add a word like *policy*. "Closed-door policy" is the standard term used in most textbooks, bringing this word close to the definition of an official term, and so could be considered to be the best choice.

The difficulties with the term *jyoukou-ka* (*information oriented*) were described in Chapter 4. While the word is not necessarily specific to Japanese culture, it is being used in a special sense in general passage C, which makes including the term in the *culturally specific* category appropriate. There were a variety of translations, even within the same text. These included "the acquisition and dissemination of information", "information-intensive", "information orientation", and "information oriented". As was the case with one of the items in the Hale and Campbell study, evaluating whether these solutions are acceptable is subjective, making a rating essentially impossible.

Ikka saiyou (*collective hiring*) was discussed in the previous chapter in terms of sentence syntax. There was no variation in the correct version of the term, which was consistently translated as: "collective hiring". The unacceptable versions were "integrated implementation", "general employment", "recruiting is ... done collectively", "employing new recruits generally", and one translation in which it was omitted. This set of characters could be considered a "complex noun phrase" as defined by Hale and Campbell (2002: 25) rather than a culturally specific term, although some knowledge of Japanese civil service hiring practices might have aided in its resolution. Taking it as a noun phrase lends support to the idea that such phrases are a specific difficulty with a resulting low rate of accuracy and that the difficulty is on the level of syntax (Hale and Campbell 2002: 26). A root problem appears to be in grasping the idea

that the head noun really does need to remain the head noun in translation and should be faithfully attended by its modifiers.

Douki no mono refers to the Japanese governmental practice of hiring a “class” of new recruits at the same time, normally in April. The variants for it were “others who entered the civil service at the same time”, “person of the same class”, “peers”, “the remainder of those hired in the same year”, “others”, ““hiring class””, “peers with equivalent seniority”, “co-worker”, and “contemporaries”. None of the translators came up with the same solution. An acceptability rating depends on whether the solution conveys the necessary information to an English reader who knows nothing about Japanese hiring practices. “Others who entered the civil service at the same time” succeeds best at this, although the quotation marks in ““hiring class”” signal that something is going on that the reader might not understand. That this expression cannot be translated directly marks it not only as an item of potential difficulty, but also as a possible feature testing “congruity judgment”.

Amakudari originally referred to the descent of the Shinto gods from heaven to earth and is now a set phrase used to describe senior Japanese bureaucrats moving to lucrative private company positions when they retire. While similar to the U.S. “revolving door” between government and industry, the movement between government positions and private sectors jobs is more institutionalized in Japan. It is not the same as a “golden parachute”, which refers to high-level retirement packages, but *Kenkyusha* does give “parachuting in” as a possible translation (Watanabe et al. 2003: 75). As with *douki no mono*, the translation should convey the necessary information to a reader who is unfamiliar with the Japanese practice. The solutions offered were transliteration (including one mistransliteration: *tenkudari*), “transferring to the private sector”, “parachuting out of government”, “retired and appointed to positions of responsibility in private corporations”, “have favorable appointments to the private sector”, “jumping to the private sector”, “leave the public sector”, and “golden parachute”. While “retired and appointed to positions of responsibility in private corporations” lacks the flavor of the Japanese, it does get the point across.

Based on the proposal in Campbell and Hale (1999) that the difficulty of a source text item can be assessed by counting the different versions in a specific population, the last two terms discussed here could be considered particularly difficult. In this study, that notion of difficulty was supported by the facility values calculated, which were 0.44 and 0.11, respectively. In their 2002 paper, Hale and Campbell asked

whether “there is a correlation between the number of available choices and the level of accuracy of translation” and did not find a clear correlation (2002: 29). In this analysis, the terms with restricted meanings—or perhaps better phrased as terms with existing English equivalents—had a much higher rate of accurate translation than the terms for which the translators needed to essentially coin a word or provide an explanation. As a result, the wide range of possible solutions could be considered an element in both reducing accuracy and increasing difficulty.

In general, the examination of modified nouns in the main study and of culturally specific terms (as stand-ins for “official terms”) here support Hale and Campbell’s propose that these are points of difficulty across language combinations. The next question then is, Are these good candidates for items to be scored on short-passage translation examinations?

I think the answer to that question in regard to the modified nouns is Yes. We saw in the analysis of the results that resolving the modification (and keeping the head noun as the head noun!) was a problem in many of the translations, with the facility value dropping to 0.25 in one case. In addition, the discussion of *ikka saiyou* (*collective hiring*), both here and in the previous chapter suggests that much simpler modification strings than I originally suspected can be items of difficulty.

However, I am not convinced that official terms or culturally specific terms make good test items. Hale and Campbell found that one of their official terms could not be scored for accuracy because the evaluation was “subjective”. One item here—*jyouhou-ka* (*information oriented*)—also could not be scored because there was no objective basis for determining whether the solutions were acceptable. That brings up the questions of setting items to be scored and identifying what, as precisely as possible, is being tested. When an examination passage contains a word like *jyouhou-ka* should the term be scored at all? Would a score provide any information other than whether an evaluator was satisfied with the solution? And would a better solution in the test setting be to simply provide a set translation for the word? Expanding the discussion a bit further, on a general test, what is the point of testing terminology? In an examination that allows Internet searches, the solutions to the sorts of culturally specific terms examined could be viewed as an indicator of research skills, something that would be of interest to a potential user of the translator’s services. But in an environment in which electronic resources are unavailable, testing dictionary use (or worse, whether the right

dictionaries were brought to the test venue) seems hardly worthwhile, particularly if there are no objective grounds for scoring the translation results.

7.2 Items of translation difficulty in eye-tracking studies

7.2.1 Jensen's indicators of text complexity

Jensen is interested in the level of text complicity for the purpose of evaluating the “amount of both production effort and comprehension effort needed during a translation process” (2009: 61-62). While he notes that, “a text’s level of complexity should not be mistaken for a text’s level of difficulty” he also writes that “we shall assume that complexity in general-purpose texts is a strong indicator of difficulty”, which is very close to the proposition under examination in this dissertation.

The indicators of text complexity he selects are “readability indices, calculations of word frequency, and calculations of the number of non-literal expressions, i.e. idioms, metaphors and metonyms”. The only term close to a metaphor is the set of texts under consideration here is *amakudari*, which was discussed in the section above. Calculation of word frequency is difficult in Japanese for two reasons: it is an agglutinative language and “the definition of *word* is not clear ... In fact, a dozen different definitions of *word* have been proposed in Japanese linguistics” (Sato 2014: 2811). Various readability formulas have been proposed (Tateishi et al. 1988; Pichl and Narita 2007; Sato et al. 2008; and Lee and Hasabe 2015), but to date there does not seem to be a commonly accepted standard.

A fairly complete list of the types of elements (but not their relative weighting) used to evaluate Japanese texts is given in Pichl and Narita:

- Coefficient of variation of sentence length
- Special kanji over 15 strokes
- Frequently used kanji (ratio of)
- Percentage of roman letters in the text
- Percentage of katakana letters in the text
- Percentage of kanji letters in the text
- Percentage of hiragana letters in the text

- Tooten to kuten ratio (commas to dots)
- Relative frequency of runs for alphabet
- Relative frequency of runs for hiragana
- Relative frequency of runs for kanji
- Relative frequency of runs for katakana
- Average number of letters per sequence (Pichl and Narita 2007: 133)

Additional markers for systems with a method for calculating words in Japanese are the number of high frequency words and the *wago* to *kango* ratio (words of Japanese origin to words of Chinese origin; this is analogous to Anglo-Saxon based words and Latin based words in English). While the reading index formulas for English differ somewhat from each other (see Jensen 2009: 64), none of them are vaguely comparable to the Japanese systems.

However, a way to make the evaluation more comparable is to calculate the readability statistics for the English translation rather than the source text. Using the editcentral.com website recommended by Jensen (2009: 64), the Flesch-Kincaid grade level for general passage A was 16.7 and the Gunning fog index was 22.2. General passage B had a Flesch-Kincaid grade level of 14.3 and Gunning fog index of 19.1 and general passage C a Flesch-Kincaid grade level of 11.8 and Gunning fog index of 16.1. The same scores for general passage D were 14.1 and 17, respectively. These figures are compatible with the earlier item facility analysis (e.g. all of the explicit cohesion markers in general passage C were 1.0 and there were no modified noun strings), which suggested both that general passages A, B, and D had relatively high text complexity and that general passage C was at a lower reading level than the other passages. Despite this compatibility being based on analysis of the translations rather than the source text, it does support the idea that tense and the other cohesion markers as well as the problem points are identifying what I, but not Jensen, would call features of text difficulty. Conversely, my analysis supports his idea that “we can go some way towards predicting the probable degree of difficulty of a text by employing a battery of objective measures”, including readability indices (Jensen 2009: 77).

7.2.2 *The Translation Difficulty Index*

The *Translation Difficulty Index* is a trial method for automatically assigning difficulty levels through a combination of the “time taken to translate the sentence” and fixation and saccade measured by eye tracking. Translation difficulty in this system is thought to be the result of sentence length, the degree of polysemy of a sentence (defined as the “sum of senses possessed by each word in the Wordnet normalized by the sentence length”), and sentence structural complexity (greater complexity is defined as attachments units being at a greater distance from each other). Correlation coefficients were calculated for the Translation Difficulty Index (i.e. the combination of time taken for translation and the eye tracking measurements) and each of the assumed difficulties and positive correlation coefficients obtained. Currently, the automatic system has reached a maximum accuracy of 67.5% (Mishra et al. 2013).

Obviously, this system is still in the developmental stage, but the defined items of difficulty are additional text features that the items in this study can be compared against. Polysemy had the lowest correlation coefficient (0.41) and needs to be calculated mathematically, so I will ignore it. Looking at sentence length, the longest sentences in passages A, B, and D were:

丸山真男の業績は、日本政治の研究水準を一挙に高めるとともに、才能のある政治学者の目を日本に向けることに貢献したことは言うまでもないが、いわゆる丸山学派の人たちの分析対象は主として思想史であって、現代日本政治の科学的研究に及ぶことが少なかった。(Needless to say, the achievement of Maruyama was that he raised the level of research on Japanese politics virtually overnight and also helped direct the attention of talented political scientists to Japan. However, the members of the so-called Maruyama school primarily focused on analysis of the history of ideas and rarely produced scientific studies of contemporary Japanese politics.) General passage A (120 characters)

都市に集中した若い核家族と単独世帯の多くは当然労働者家族であるが、農民家族と異なって、生産の手段をもたない労働者家族は、生産と消費が分離され、居住地では消費的機能だけを担うことになる。(Most of

the young nuclear families and single-person households concentrated in cities were, of course, workers' families. However, workers' families, which, in contrast to farming families, did not have a means of production, and so production and consumption were separated, with the result that residential areas took on the function of consumption only.) General passage B (91 characters)

そして、同期の者が局長・次官に昇任しても他の者は天下りをせず、定年まで勤めること、現在、局長職までの年俸は相対的に低いことを改善し、報酬をアップすること、定年後の年金額も例えば退職時の8割を支給するなどにより増額し、公務員在職中に培ったキャリアと本人の希望にふさわしい人生を送ることができるようにすること、これらの政策により、天下りによる民業の圧迫と、税金の浪費を回避することもできよう。(Next, the following policies should also make it possible to avoid wasting taxpayer money and pressuring private enterprises by having civil servants “parachuting in”: even though people who started work in the same year have been promoted to bureau chief or vice minister, others should work until retirement age and not be forced to parachute into private companies; increasing remuneration by improving annual salaries up to the level of bureau chief, which are currently relatively low; and making it possible for civil servants to foster their careers during their public service and to lead a life appropriate to their expectations by further increasing the amount of pension after retirement to, for example, 80% of the salary received in the final post.)

General passage D (193 characters)

The first sentence did not cause this test population particular difficulty. The *ga* in the middle signals a clear breaking point (just like a semi-colon followed by *however*). Interestingly, the second sentence has the same *ga* structure and was found to be more difficult. A purely Japanese measure of sentence difficulty, the *tooten* to *kuten* ratio (commas to periods) was the same for each (4:1), so the difference may have been because of difficulties with tense (facility values of 0.33 and 0.25, respectively, for sentences 5a and 5b of general passage B; the verbs were in the present narrative), or the structure of the second sentence which had the *ga* break much closer to the beginning than often happens, with some of the translations then putting an additional,

and disruptive, break in the sentence later on and others not breaking the sentence at all. My intuition is that misinterpretation of the particles in the second sentence (needing to see *wa* as *in* and *dewa* as making *residential area* plural and general) also played a role. Also, while the test population met the same selection criteria each year, the translations were done by different sets of translators.

The concluding sentence of general passage D is long enough to be a problem in any language. However, the syntax and punctuation should make the structure quite clear. Each *こと* (*koto*) followed by a comma is a policy being proposed; if this were not an editorial, these segments could very easily be made into bullet points. Nonetheless, the tense and explicit cohesion marker facility values were the lowest in the study (0.11 and 0.0, respectively), indicating that this sentence was extremely difficult for this population.

It has been argued that “the length of a sentence is not a reliable indicator of its readability” (Lin 1996: 729), and even that longer sentences may be easier to read and process because of the possible inclusion of clarifying elements (Koda 2004: 109). However, the results of this very small study supported conventional wisdom: general passage C, in which all the sentences were less than 100 characters long, appeared to have been much easier for this population than general passage D, which had an extremely long sentence. But the sample size here was far too small to allow for any definite statements.

The definition of “structural complexity” in Mishra et al. (2013: 348) is based on Dekang Lin’s earlier work on sentence structural complexity. Lin defines structural complexity in a dependency structure (i.e. a head and modifier) as “the total length of the dependency links¹⁸ in the structure”, with the idea that “shorter dependency links are easier to establish than longer ones” (Lin 1996: 730). This notion of structural complexity is quite useful. Looking back at the modified nouns, it explains why the modified noun 日本人研究者による日本政治研究の水準 (*the level of research on Japanese politics by Japanese researchers*, general passage A), which can be processed in the order the characters appear, would be easier to process than このように社会的依存度を高めた消費の場すなわち家事労働の場 (*frozen food and disposable consumer goods continuously flowed into the site of household labor, that is, the site*

¹⁸ Lin defines a “dependency link” as “one plus the number of words between the head and the modifier” (1996: 730).

with this high degree of societal dependence, general passage B), in which the adverb *kounoyouni* (*this, this type of*) has been separated from its verb *takameta* (*high*; Japanese adjectives become verbs in the past tense) by seven characters (correspondingly facility values, 0.81 and 0.25). In line with the premise in this study, connecting separated elements would also require a certain degree of inference and ability to sort out the relationships in the sentence.

In conclusion, the findings here support Mishra et al.'s claim that structural complexity in the sense of distance between related items is a feature of translation difficulty, and suggest that sentence length as an independent item may be supported (again, the sample size is too small for meaningful inferences). Interestingly though, one of the shortest sentences had the most syntactical errors: sentence 6 of general passage D (省別採用ではない一括採用は、地方公務員の採用で取り入れられている。 [*Collective hiring rather than by each ministry is being incorporated into the employment of regional civil servants.*]). I should add that this sentence does not have structural complexity as defined above. The item of difficulty is instead a complex noun phrase as identified by Hale and Campbell.

7.3 Summary

The modified noun strings set as problem points in this study can be viewed as corresponding to the complex noun phrases identified as difficulties by Hale and Campbell (2002). These were problem points for the translators in both studies. The combined studies involved three languages pairs (English to Arabic, English to Spanish, and Japanese to English) across three unrelated language families, suggesting that the difficulty of such phrases is language independent. While I set the minimum length of the problem point modification strings at fifteen characters, the close reading of the translations described in Chapter 6 showed that phrases as short as four characters could present difficulties. That finding was consistent with the data in the Hale and Campbell article.

One of Hale and Campbell's research questions was whether the terms with a wider range of possible translations were more difficult. They found that this was generally not the case. However, when I analyzed culturally specific terms as stand-ins for their official terms, I found that terms without an existing set translation and hence a

wider range of translation possibilities had lower facility values than terms with accepted English translations. While this makes intuitive sense, it should be kept in mind that the sample size was very small. Both the Hale and Campbell study and this study found that some terms could not be evaluated because rating of the acceptability of the translation was subjective.

The items from the eye tracking studies that were investigated were readability indices, sentence length, and structural complexity defined as distance between related elements in a sentence. Comparison with the passages and translations here showed that readability indices corresponded with the difficulty level of the passages in that both the indices and the facility value profile showed that general passage C was the easiest of the four passages. Sentence length may be a useful parameter, particularly in the sense of avoiding passages with no long sentences such as general passage C as well those with extremely long sentences, such as general passage D. The facility values for each of these passages suggest that they were too easy and too difficult, respectively.

Structural complexity as defined by Lin (“the total length of the dependency links in the structure” 1996: 730) fits in nicely with the argument being made here that the need to connect dispersed textual elements (i.e. make inferences) signals both text complexity and difficulty (i.e. a higher failure rate in a defined test population).

8. Technical passages: results and analysis

ATA selection criteria during the period of the tests studied here did not require the technical passages to be a specific ILR reading level, but the technical passages tended to have a linguistic structure that was simpler than the general passages. As a result, one would expect errors to be more in the areas of style conventions or vocabulary particular to the subject matter. An additional assumption is that the translation product quality will be better for these passages. Because of the loose connections between sentences in texts at this level, the discussion of cohesion will be extended to cohesion within sentences.

8.1 Technical passage A

The first passage is a discussion of global warming that self-contained with a fairly simple structure. It can be seen as a collection of facts, with two definitions. The factual statements are not tightly connected and could be reordered to some extent. These are characteristics of ILR Reading Level 2; the appearance of chemical names in this text push the level toward 2+, as do some of the syntactical structures. The text mode is instructive.

8.1.1 Technical passage A: passage and structure

Technical passage A:

地球の気温は、平均すると 15°C程度に保たれている。地球を取り巻く大気の層が、地上で反射した太陽熱が宇宙に逃げていかないように作用するため、これを「温室効果」という。もし大気の層がなければ、地球の平均温度はマイナス 18°C程度になり、生物の生育はおぼつかなくなる。

。

地球温暖化とは、人間社会が生み出した温室効果ガスによって「温室効果」が過度になる現象である。

人為的に排出される温室効果ガスには、二酸化炭素・メタン・一酸化二窒素・フロン類がある。特に二酸化炭素の人為的排出は、18世紀中ごろから始まった産業革命以来、増大化の一途をたどってきた。

現在、人為的に排出される二酸化炭素は年間約 240 億トンにもものぼり、その 7 割が石油などの化石燃料から放出されるといわれている。人為的な二酸化炭素の排出が、地球温暖化の元凶とされる。しかも、排出量の 3 分の 2 は先進国によるもので、日本の場合は温室効果ガス排出量の約 9 割を占めている。

Possible translation:

The temperature of the earth is maintained, on average, at around 15°C. Because a layer of air surrounding the earth functions to keep the solar heat reflected by the ground from escaping into space, this arrangement is called the “greenhouse effect”. If that layer of air were not there, the average temperature of the earth would drop to around minus 18°C, and the continuation of life would become essentially impossible.

Global warming is the phenomenon in which the “greenhouse effect” becomes excessive as the result of greenhouse gases produced by human society.

Anthropogenically emitted greenhouse gases are carbon dioxide, methane, nitrogen monoxide, and chlorofluorocarbons. Emissions of carbon dioxide in particular have increased steadily since the industrial revolution that began around the middle of the eighteenth century.

At present, the amount of man-made carbon dioxide emitted has reached approximately 24 billion tons annually; of that amount, 70% is said to be released by fossil fuels such as petroleum. Emission of man-made carbon dioxide is considered to be the main culprit in global warming. Moreover, two thirds of carbon dioxide emissions are generated by the advanced nations; in the case of Japan, carbon dioxide accounts for approximately 90% of greenhouse gas emissions.

Reading comprehension of the above passage requires following the presentation of the facts that:

- The average temperature of the earth is 15°C.
- The fact that this is maintained by a layer of air is known as the “greenhouse effect”.
- Global warming is an anthropogenic phenomenon that creates an excessive “greenhouse effect”.
- Of the greenhouse gases being emitted, carbon dioxide emissions in particular are increasing, making up two-thirds of the emissions by the advanced nations and 90% of Japan’s greenhouse gas emissions.

Understanding the passage does not require inference or reading between the lines. The terms are the four chemicals, but given the one-to-one correspondence to the English names of the materials, translating those should not be difficult. As will be discussed below under sentence 3, おぼつかなくなる (*obotsukanakunaru*, *to become uncertain*) is the main phrase that requires deciding between multiple meanings, while 生物の生育 (*seibutsu no seiku*, see below for meaning) in the same sentence requires finding a meaning that is not in the dictionary.

8.1.2 Technical passage A: tense, other cohesion markers, and problem points

The nine final verbs in the passage can be categorized as follows:

1. 保たれている (*tamotareteiru*): base meaning, *to keep*; inflected form, passive progressive indicating a resulting state (Kaiser et al. 489-490); possible English translation, *is kept*, *is maintained*
2. という (*to iu*): base meaning, *to say*, *to be called*; inflected form, non-past; possible English translation, *is called*
3. おぼつかなくなる (*obotsukanakunaru*): base meaning, *to become uncertain*; inflected form, non-past; possible English translation, *would become uncertain*
4. である (*dearu*): base meaning, copula; inflected form, formal non-past; possible English translation, *it is*
5. ある (*aru*): base meaning, *exists*, *is*; inflected form, non-past; possible English translation, *are*

6. たどってきた(*tadottekita*): base meaning, *to follow*; inflected form, past (the form is past; however, the meaning is the speaker's "realization of a present state" (Kaiser et al. 464-465); possible English translation in context, *have increased steadily*
7. いわれている(*iwareteiru*): base meaning, *to say*; inflected form, progressive; possible English translation, *it is said that, supposedly*
8. される(*sareru*): base meaning, *to do*; inflected form, passive non-past; possible English translation (with particle と), *is considered to be*
9. 占めている (*shimeteiru*): base meaning, *to account for*; inflected form, progressive indicating a resulting state; possible English translation, *accounts for*

While the tense structure of the English translation is reasonably straightforward and is primarily in the simple present with two perfect verbs and one conditional, the Japanese is a bit more complicated. The *-teiru* ending of the main verbs of the first and ninth sentences is often described as a progressive, but as noted in the discussion of the general passages, it can also be a perfect, in which case the verb can be translated using the simple present in English. Which tense to use is dependent on whether the verb in question is a change of state verb, an action in progress verb, or a verb that can fall into either category, at which point context becomes important (McClure 2000: 160-165).

おぼつかなくなる(*obotsukanakunaru, to become uncertain*), the final verb of the third sentence, is non-past, but the conditional verb in the first clause (もし大気の層がなければ, *moshi taiki no sou ga nakereba, if the layer were not there*) needs to be represented by a conditional in second clause in English. The main verb of the sixth sentence consists of the たどって (*tadotte*, connective verb form), followed by the verb くる (*kuru, to come*). -てくる (*-tekuru*) has several meanings (e.g. action directed toward the speaker,), but in this context gives a sense of change or development (Kaiser et al. 2001: 504-505). It is in the past tense, but refers to the "realization of a present state" (Kaiser et al. 2001: 464-465) and can be appropriately translated using either the English present perfect with its sense of change over time or a present progressive. いわれている (*iwareteiru*, sentence 7) is primarily an expression to keep the author from sounding overly assertive and could be omitted in translation.

As would be expected from a passage that is essentially a loosely ordered collection of facts, cohesion markers do not play much of a role in this text. However, because of the tendency of Japanese to omit nouns and pronouns that are required in English, some inference is required on the level of finding antecedents. A certain amount of knowledge of global warming and the chemicals thought to be responsible for it would be helpful for translation but is not necessary.

The sentences are all reasonably short, with no unusual syntax. However, some degree of inference is required to connect the definition of the greenhouse effect given in the second sentence with the stable temperature of the earth noted in the first sentence. The final sentence requires inference to connect *mono (it)* back to *carbon dioxide* in the previous sentence. Without a dictionary, translation of the chemical names might be difficult, but with the appropriate references they should not be a problem. There are no idioms. Two phrases—おぼつかなくなる (*become uncertain, be all but impossible*) and 元凶とされる (*be seen as the main culprit, be seen as the principal cause*)—could have a range of translations, and so these could be considered a point of possible difficulty (Hale and Campbell 2002).

8.1.3 Technical passage A: results of analysis of tense, other cohesion markers, and problem points

Because the error patterns in the technical patterns differ from those of the general passages, tables for each marker will be not used. Instead, a sentence-by-sentence analysis will be done.

Sentence 1: There were no problems with tense in the translation of the first sentence, which is a straightforward statement of fact.

Sentence 2: Again, all the verbs in this sentence in both languages were non-past/present tense. However, there were an interesting problems with the *ため* (*tamede, because*) connecting the two clauses in this sentence. Four of the twelve of the translations (A2, A5, A9, and A10, facility value 0.64) took it as referring to the first sentence, which is incorrect, and began the second sentence with some version of “This is because”. The core meaning of the second sentence is that the structure that prevents solar heat from escaping into space is called the greenhouse effect because—the next part is implied—it resembles a greenhouse. The difficulty is compounded by the need in

English to provide an antecedent for *これ* (*this, it*), which refers to the system described in the first clause, by adding a noun. Translation A1 solved this by adding “action” (“Because ..., this action is called the “greenhouse effect.” Translations A4, A8, and A11 solved the problem by breaking the sentence in two with the first sentence describing how solar heat is kept around the earth, then beginning the second sentence with a “this” referring to the first sentence overall and omitting “because”. It should be noted that the test situation may have affected the translation by making people reluctant to add a word not explicitly in the source text. However, this sentence is a case of inference being required to resolve the structure of the sentence. The facility value for internal cohesion for sentence 2 was 0.64, confirming that the item was somewhat difficult.

Sentence 3: None of the translations had a tense mistake in the final verb of this sentence, and there was laudable usage of the English subjunctive in the initial clause. Use of the gerund in two papers (A4, A6) was clever and nicely incorporated the causative sense of the connective form final verb in the preceding clause (*nari, would become*).

The variety of translations possible for *obotsukanakynaru* mark this word as a potential difficulty. The base adjective is *obotsukakai*, which, according to *Kenkyusha*, means “uncertain, doubtful, nearly hopeless; insecure, unsteady; vague, indistinct”, which certainly qualifies as a broad set of potential meanings (Watanabe et al. 2003: 418). In the *-ku* form, it functions as an adverb. The final *naru* is a verb meaning *to become*. *Seibutsu no seiiku* is a problem for the opposite reason: it does not have a wide range of possible meanings and essentially represents “growth of living things” or “development of living things”, neither of which will quite work in context. As a result, some degree of transfer skill is required, and the phrase could be seen as a test of congruity judgment.

In this sentence, the phrase with a restricted number of definitions appears to have been more problematic than the one with a wide range of meanings—the opposite finding of what was seen for culturally specific terms in the previous chapter. For *obotsukanakynaru*, five of the eleven translations had either “uncertain” or “no longer be certain,” and two used “unlikely”. An additional two had “hopeless” or “little hope”. Translation A9 reduced *obotsukanakynaru* to *barely*, which does not catch the sense of the term or the idea of *become* represented by *naru*.

In the context of the passage, *development* is not an appropriate choice for *seiiku* (three translations had *development* and five, some form of *grow*). The sentence is describing that would happen if the atmospheric layer holding in solar heat was not there (the Earth's temperature would drop to around -18°C), and the result would be that most life forms would disappear. This makes the issue survival, not development. *Seibutsu* here is best translated simply as *life* (chosen by four people). While most of the translations stayed very close to the equivalents given in *Kenkyusha*, translation A4 exhibited good congruity judgment (coupled with unfortunate grammar and spelling: “making the continuance life uncertain”) by moving away from the literal meaning. However, the end result was that, subjectively speaking, none of the translations of this sentence could be used without editing.

Sentence 4: This sentence was a definition given in standard syntax. It is not surprising that there were no mistakes in tense and that all the translations provided a definition of global warming.

Sentence 5: There were no problems with tense in this sentence, which is quite straightforward. Interestingly, however, eight of the eleven translations had the verb *include*. This verb is often used to express that a list is not comprehensive because *など* (等, etc.) is placed after the last item, but *など* is not present here and *are* could be safely used.

Sentence 6: This sentence has a time expression (18 世紀中ごろから始まった産業革命以来, “since the industrial revolution that began around the middle of the eighteenth century”) as well as a verb form indicating movement over time. These two elements together call for an English perfect construction, and all of the translations had some form of this. The one mistake (A9: “has continuously been increased”) may well have been a slip of the pen.

Sentence 7: There was only one tense mistake for this sentence (translation A1, “with 70% of the total is released from”). Three of the translations had a more definite statement in English than is present in the source text and omitted the final *iwareteiru*, *it is said that*. However, this could be considered to be in line with standard conventions in both languages.

Sentence 8: Tense was not a problem in this sentence. Some of the translations omitted the final verb, the main function of which is to soften an assertion of fact. Whether than that counts as an error (*emission of man-made carbon dioxide is*

considered to be the main culprit in global warming as opposed to *emission of man-made carbon dioxide is the main culprit in global warming*) is a subjective call, so I have not marked this as incorrect.

Sentence 9: This was another present tense sentence, and again, there were no tense mistakes. However, understanding sentence 9 required linking it back to sentence 8, and the majority of the translations failed to do this (seven translations were clearly incorrect and two were borderline, depending on how the pronoun *it* is interpreted). The eighth sentence says that the man-made emission of carbon dioxide is the main culprit in global warming. Because omission is a coherence pattern in Japanese (Baker 1992: 185), that no new topic or subject has been given means that such emissions are the topic of the next sentence. Therefore, since we have not been told otherwise, the type of emissions in sentence 9 remain the same type as those in sentence 8: man-made carbon dioxide.

Part of the problem in seeing this link is that it is implicit. There are no pronouns signaling an antecedent; the translation must depend on context as well as on knowledge of how cohesion functions in Japanese. The only other aid is real-world knowledge, and according to the ILR reading level descriptions, using “real-world cues to understand the text” is a characteristic of reading level 2 (ILR 2015). Does it make sense to say that Japan is responsible for 90% of the world’s greenhouse gases? Where does that leave the U.S.? And what about the great, spewing factories of China?

While there were infelicities in several of the translations before the final sentence, at this point we have run up against a major error suggesting a lack of ability to connect the dots when the train of thought is longer than a sentence as well as problems with re-reading the translation for meaning. This merits a closer look. A table of the translations of the final sentence is given below.

Table 8.1. Translations of sentence 9, technical passage A

Translation No.	Syntax Problems in Final Clause
	しかも、排出量の 3 分の 2 は先進国によるもので、日本の場合は温室効果ガス排出量の約 9 割を占めている。
A1	Moreover, two-thirds of emissions volume are due to advanced nations, and Japan accounts for about 90% of greenhouse gas emission volumes.
A2	Furthermore, two thirds of the discharge are by the advanced nations, Japan discharging approximately 90% of the green house effect gases.
A3	Furthermore, since 2/3 of the exhaust volume is due to the advanced nations, Japan accounts for approximately 90% of the greenhouse gas exhaust volume.
A4 *	Furthermore, 2/3 of those emissions are caused by the developed countries. In the case of Japan it accounts for approximately 90% of greenhouse gas emissions.
A5 *	Moreover, two thirds of the amount discharged comes from advanced countries. In the case of Japan it amounts to about 90% of the amount of greenhouse gases discharged.
A6	Moreover, 2/3 of emissions come from developed countries, and Japan makes up about 90% of greenhouse gas emissions.
A7	In addition, since two-thirds of the quantity of emissions comes from advanced countries, Japan ranks in the 90 th percentile in terms of quantity of greenhouse gases emitted
A8	Furthermore, developed nations account for two-thirds of this emission, and about 90 percent of greenhouse gases are emitted by Japan
A9 **	Furthermore, two thirds of such emissions come from industrialized nations. In Japan, the artificial carbon dioxide emission makes up about 90% of the total gas emission having the green house effect.
A10 **	Furthermore, two thirds of such emissions come from industrialized nations. In Japan, such emissions account for about 90% of the total emissions of greenhouse effect gases.
A11	Moreover, two thirds of the amount of discharge is produced by industrially advanced nations. This means that Japan is responsible for approximately 90% of the discharge of greenhouse gases.

*The sentence has a correct reference to the previous sentence if the *it* refers to carbon dioxide. However, the pronoun usage here is not acceptable in English.

** The sentence is correct.

The uniformity of the unacceptable translations saying that Japan is the world's major polluting nation is striking and appears to have come from a collection of the features that were being examined for the ILR reading level 3 texts. The explicit cohesion marker, *shikamo* can be, and often is, translated as *moreover*, *furthermore*. However, unlike other words meaning *moreover* (e.g. *mata*, *sarani*), it has the additional sense of *to boot*, *on top of that* (Watanabe et al. 2003: 1125). While one probably would not use either of those in the translation, seeing that there are two problems with carbon dioxide (it is mainly responsible for global warming and, on top of that, it makes up the bulk of the greenhouse gases omitted by the developed nations) is important in resolving the sentence. In the Japanese cohesion structure, there is no need to repeat

carbon dioxide or a demonstrative pronoun before *emissions* in sentence 9. However, a demonstrative pronoun is required in English. This was handled nicely in the correct translations A9 and A10 with “such”. Translation A4, which has a potentially correct reference to carbon dioxide, used “those”. Translation A8 had a correct “this”, but then lost the connection in the final clause. Since no new subject (the *ba'ai* after *Japan* marks that noun as an example—*in the case of Japan*) has been stated, the subject of the final verb *shimeteiru* (*to account for, to make up*) has to be *carbon dioxide*, the subject of the previous sentence. However, all of the incorrect translations mistakenly took *Japan*, which is not even the head noun in its phrase, as the subject. Three translations (A3, A7, and A11) additionally translated the connective copula *de* as a causative (*since, this means that*), which is grammatically, but not semantically, possible.

I have discussed this sentence at length for two reasons. First, it confirms the utility of examining cohesion markers and structures requiring inference in evaluating test passages and their resulting translations even at ILR reading level 2/2+. Second, the sentence points toward recurring language specific errors that were also seen in the earlier passages, notably, identifying head nouns and null subjects, showing that these are items of difficulty even in texts with simpler structures and arguments.

In addition, I think the mistakes in this sentence point to an interesting problem that will be discussed in the next chapter, that of difficulty, in terms of making a mistake, without cognitive effort. I strongly suspect that the translators who had Japan producing 90% of the world’s greenhouse emissions did not view this sentence as one requiring particular effort, because they did not see the various elements pointing to the correct solution.

8.1.4 Technical passage A: summary of errors and comparison with general passage A

The following summary table shows that evaluating tense did not provide useful information about the translations. This is in contrast with general passage A, in which tense identified the first two sentences of that passage as difficult. The lack of tense errors indicates that technical passage A had a reasonably simple structure with loosely organized information and so contributes to the passage difficulty profile. The facility values identified the internal cohesion of sentence 2 as a problem point and that *tamede* (*because*) might be an item to flag in other potential passages. At 0.09, the facility value of the phrase involving wording that with choices not in the standard dictionary was too

difficult for this population (0.25 is the cut off point for acceptability).¹⁹ However, less challenging versions of this type of item might be useful in testing congruity judgment. In the table, the error in sentence 9 is reduced to having the correct subject. Depending on how one evaluates translations A4 and A5, this sentence may have been too difficult for this population. However, given that the main problem appears to have been inadequate grasp of Japanese grammar, this may point to issues with the eligibility criteria for sitting the exam.

Table 8.2. Technical passage A: errors and facility values

Translation No.	Sentence No.									No. of errors
	1	2	3	4	5	6	7	8	9	
		Internal cohesion	Final clause						Subject	
A1			WC				V		S	3
A2		CE	WC						S	
A3			WC						S	2
A4									?	1/0
A5		CE	WC						?	3/2
A6			WC						S	2
A7			WC						S	2
A8			WC						S	2
A9		CE	WC			V				3
A10		CE	WC							2
A11			WC						S	2
Facility value		0.64	0.09			0.01	0.91		0.18/0.36*	

CE is a cohesion error; WC is a word choice error; V is a tense error; and S is a subject error. ? indicates that the choice is not clearly correct or incorrect.

* The facility value was calculated first with translations A4 and A5 being marked incorrect and then with the two translations being marked correct.

8.1 Technical passage B

The second technical passage is a self-contained text, and while the information is a bit more tightly ordered than in the previous passage, the text is still a collection of facts and is at ILR Reading Level 2+. The translation sample size is 11; translator B6 did not complete the technical passage. The subject matter—flow—is more technical than in the previous passage and is not as accessible to a general reader. The mode is instructional.

¹⁹ Another possibility is that people were unwilling to take risks in a test situation.

8.2.1 Passage and structure

Technical Passage B:

流れの制御

流れを“制御する”ということは自然に学び、これにひとの知恵を加えて人為的に流れを望ましい方向に変えるということであろう。一般的には乱れが起こると抵抗が増すためにこれを防止又は制御するのが望ましいが、目的によっては、例えば実験現場でトリップワイヤを使用する場合のように逆に乱流を促進するのも乱れ制御のひとつである。しかし、工学的に重要かつ難しいのは前者の乱れを制御する方法であり、古くから数多くの研究がなされてきたが実用化されたものは少ない。

ところが近年、可視化試験やその他の流れ計測技術の発展に伴い乱流の研究が飛躍的に進展し、そのメカニズムが明らかになってきた為、これらの知見に基づいた新しい制御法が見出されてきた。船舶や航空機では摩擦抵抗を僅か2 – 3%でも低減できれば年間相当量の燃費を節約できるため、低抵抗への要求は強く、乱れの制御に関する関心が高まってきている。

Possible translation:

Flow Control

“Controlling” flow means to artificially change a flow to a preferred direction by learning from nature and then applying human knowledge to this. In general, it is desirable to prevent or control flow because the occurrence of turbulence increases resistance; however, based on the objective, conversely promoting flow as in the case of using a trip wire in an experimental site, for example, is also another way of controlling turbulence. However, the first method, in which turbulence is controlled, is the one that is both important and difficult in engineering. Therefore, numerous studies of this have been done over the years, but few of these have resulted in practical applications.

However, in recent years, with developments in visualization experiments and other flow measurement technologies, research in flow turbulence has made dramatic progress. Because the mechanisms of flow have been clarified, new

methods of control have been discovered based on these findings. If the friction resistance of ships and airplanes could be reduced by even as little as 2%-3%, an equivalent amount of fuel could be saved annually; therefore, demands for low resistance are increasing and interest in flow control is growing.

Reading comprehension of the above passage requires following the presentation of the following facts:

- The definition of flow control
- A description of the usual type of flow control
- An example of another type of flow control
- The fact that the first type is the most important, but that studying it has not resulted in practical applications
- New technologies are leading to better results
- Interest in flow control is increasing because of its connection with fuel reduction

8.2.2 Tense and other cohesion markers

The five final verbs in the passage can be categorized as follows:

1. であろう (*dearou*): base meaning, copula; inflected form, formal written presumptive form (Kaiser et al. 2001: 395); this could be omitted in translation. The sentence is a definition of *flow control*, and in English the main verb is *means*.
2. である (*dearu*): base meaning, copula; inflected form, formal non-past; possible English translation, *it is*
3. Omitted copula. 少ない (*sukunai*, *few*), the final word in the sentence, is a nominal adjective (McClure 2000: 227) and requires a following form of *desu*, which has been omitted here. As a result, the tense comes from the last verb in the sentence, 実用化された, *jitsuyou-ka sareta* (*have resulted in practical applications*), which is past tense, passive voice.

4. 見出されてきた (*miidasaretekita*): base meaning, *to discover*; inflected form, past passive; possible English translation, *have been discovered*
5. 高まってきている (*takamattekiteiru*): base meaning, *to increase, to grow*; inflected form, progressive; possible English translation, *is growing*

Because the Japanese verbs in a sentence take their tense from the final verb, the copula at the end of the first sentence is important. However, in addition to providing a non-past tense, its main function is softening the statement in which the definition of flow control is given. The other verbs in the sentence are also non-past. In contrast to the formality of the first sentence, the copula in the third sentence has been omitted. The tense of the last verb given (*jitsuyou-ka sareta, have resulted in practical applications*) is past passive, suggesting that the English present perfect is appropriate. The structure appeared previously in sentences 3 and 4 of general passage A: the final word needs to be treated as an adverb and the verb brought forward as the main verb in the sentence. This was done correctly for both sentences with this structure in the general passage A translations.

The verb in the fourth sentence is in the past passive, but is also in the *-tekuru* form discussed in the previous passage and, as in the earlier text, has a sense of change or development. The final verb is quite agglutinative, with *-tekuru* then *-teiru* added to the base verb (高まる, *takamaru, to increase, to grow*). The two *-te* forms provide a sense of change/development plus a progressive marker. English can only respond with the progressive tense.

While the above sounds rather complicated, the verb tenses remain within the range of normal Japanese usage. The main non-verbal cohesion markers are various forms of *however*, as well as the repetitions of *flow* and *turbulence*.

The passage does not require inference or reading between the lines, and while some background knowledge of flow and how it works would be helpful, all the information needed to translate the passage is provided in the text. Successful translation does require the ability to see connections between elements in different sentences.

8.2.3 Technical passage B: results of analysis of tense, other cohesion markers, and problem points

As with technical passage A, tables for each marker will be not used. Instead, a sentence-by-sentence analysis will be done.

Sentence 1: Tense was not a problem in this sentence, but an overly literal approach to the final copula turned out to be. Of the five unacceptable translations, two used “probably means”, one, “It may be said that”, and one, “can perhaps be said to be”, none of which are appropriate for the text type or consistent with English usage. More interestingly, 自然に学び (*shizen ni manabi, learn from nature*) turned out to be an unexpected problem. Rather than some form of *learn from nature*, several of the translations had errors such as “is learned naturally” (twice), “what is learned naturally”, and “we grew up learning”. The difficulty appeared to have been not differentiating between a noun followed by を学ぶ (*wo manabu, to learn something*) and a noun followed by に学ぶ (*ni manabi, to learn from someone/something*). The particle *wo* is a direct object marker, while the particle *ni* markers *from/to* an object. A confounding factor is that when *ni* is added to the noun *shizen* (nature), it forms the adverb *naturally*. This suggests that people failed to parse the *kana* correctly and made the familiar adverb construction their first choice. We saw problems dividing character strings correctly and choosing the wrong definition of familiar words in the translations of the general passages. This suggests that examining familiar words used in different and/or incorrect ways might be a fruitful area for further research into Japanese-specific errors.

Sentence 2: All of the translations used an *it is* construction, which was appropriate. The source sentence was 95 characters long, the longest sentence in the passage, and splitting the sentence at *ga* (*however*) allowed the translators to keep the syntax and modification under better control. As discussed in Chapter 7, the fact that many long sentences have a ready break point at *ga*, keeps sentence length from being too much of an indicator of difficulty in Japanese.

Sentence 3: As noted in the discussion of the verbs, the tense of this sentence needed to be taken from the verb *jitsuyou-ka sareta*, which is in the past tense. While this was correct in nine of the translation, two (B2 and B4) used a present tense.

The twelve-character noun modification string 前者の乱れを制御する方法 (*the former method of controlling turbulence*) was a problem point, generally because of

problems with 前者 (*zensha*, *the former*). As noted in Chapter 7, distance between related elements in a sentence increases text complexity (Lin 1996), and the analysis of the general passages showed that modification strings did not have to be long (defined as 15 characters or more) to be difficult. Six of the eleven translations did not have an acceptable solution.

前者 (*zensha*) functions in Japanese in the same way as *the former* in English and so would not seem to be much of a challenge. In this passage, it points back to the information given in the first sentence describing the usual goal of flow control. However, the problem for the translators appears to have been unpacking the modification of 方法 (*houhou*, *method*): 前者の乱れを制御する方法 (interlinear translation: *former of turbulence object marker to control method*; translation: *the former method, in which turbulence is controlled*). Rather than seeing that *zensha* no modified *houhou*, the most common error (five of eleven people) was to see it as modifying the closest noun (乱れ, *midare*, *turbulence*), despite the fact that two methods (a former and a latter) had been implicitly discussed in the previous sentence. The implicitness—in addition to the distance—may have been part of the problem, since the word 方法 (*houhou*, *method*) is not used in the preceding sentence. Another factor may be an inability to break the *kana* and characters into appropriate units. Three of the six translations in which the modification in this sentence was not correctly resolved, also showed problems in parsing the *kana* for *to learn from nature* correctly in the first sentence (B2, B8, and B11).

Sentence 4: While there were assorted problems in the translation of sentence 4, the translations generally did not show problems with the tense of the main verb. Translation B8 strayed a bit from the source text and missed the *have* in front of the final verb to complete the past perfect (“Recently, there has been rapid development in the studies of turbulence accompanying possible tests in flow measuring technology which clearly explain their mechanism and based on this knowledge caused new control methods to capture headlines.”)

Internal tense consistency was good in this sentence, with only one translation (B1) failing to take the tense of the verb in the connective form from the two conjugated verbs. How the translations handled the 為 (*tame*, *because*) in the middle of the sentence was interesting in that while the majority of the translations used some form of *because*, which then required some reordering of the information in English,

translations B3 and B5 used *hence* and *therefore*, respectively. This choice let the translators reproduce the order in which the information was given in Japanese. Both solutions can be seen as correct.

Translation B7 omitted any sense of cause and result: “Recent years, however, have seen dramatic progress in research into turbulence with the development of visualization testing and other flow measurement technologies. This research has clarified the mechanism of turbulence, and new control methods based on this knowledge have started to appear.” While the text reads well, the relationship between the elements in the sentence has been simplified.

Sentence 5: There was only one error in tense for sentence 5 (no sense of ongoing action expressed), and that was in translation B8, which continued to be an outlier. Three translations (B5, B7, and B12) showed an interesting decision to use a noun phrase rather than a verb (some variant of “increased interest in”); a good reminder that, given the variety possible in language, setting a pattern that must be matched is a dubious idea in test construction.

While the tense choices were appropriate in this sentence, the structure of the translations suggested possible problems in understanding the syntax, as well as the meaning. The first hint of trouble was that all eleven translators mistranslated 年間相当量 (*nenkan soutou ryou*, *an annual equivalent amount*, i.e. 2%–3% annually), which, like 省別採用 (*shoubetsu saiyou*, *hiring by ministry*) in general passage D, I would not have identified as a problem point. Few of the translators showed evidence of realizing that the author was setting up a hypothetical situation: “If friction resistance could be reduced by even as little as 2%–3%, then an annual fuel savings of the same amount would be possible. Because of that, people want such a reduction and have increased interest in turbulence control [as a method of achieving it].” Two translations came close to this structure (B2 and B5), but only one of these (B5) used a conditional verb: “Even reducing the friction resistance of a ship or aircraft by a mere 2 or 3% would make it possible to annually reduce fuel consumption by a considerable amount. Therefore there is strong demand for low resistance, and increased interest in turbulence control.”

As in the technical passage A, the last sentence, which often has the most important information in Japanese, caused the most problems and again resulted in solutions in which the translation was contrary to fact. As noted in the sentence above,

the punch line to the discussion of controlling flow and turbulence is that there is the practical result of fuel economy. The problem is determining how much fuel can be saved. The source text says (*nenkan soutou ryou no nenpi*), “fuel consumption of an equivalent amount yearly”, pointing back to the 2% to 3% reduction in friction resistance. However, none of the eleven translations made this connection. A table of the translations of the first part in sentence 5 follows.

Table 8.3. Translations of first half of sentence 5, technical passage B

Translation No.	船舶や航空機では摩擦抵抗を僅か2-3%でも低減できれば年間相当量の燃費を節約できるため、
B1	Vessels and airplane can save a significant amount of fuel annually by reducing frictional resistance [sic] by just 2-3%.
B2	If a mere 2% to 3% decrease in frictional resistance can be accomplished in ships or airplanes, a considerable amount of fuel can be saved annually,
B3	Since an amount of fuel equivalent to a year can be saved by reducing the resistance due to friction by a mere 2 to 3% for ships and airplanes,
B4	With ships and airplanes, because annual amounts of fuel consumption can be economized if friction resistance can be reduced by merely 2-3%,
B5	Even reducing the friction resistance of a ship or aircraft by a mere 2 or 3% would make it possible to annually reduce fuel consumption by a considerable amount.
B6	Passage not translated
B7	A reduction of a mere 2 to 3% in friction resistance in ships and aircraft enables considerable savings in fuel costs each year,
B8	is because a reduction of friction resistance as little as 2-3% in ships and aircraft can lead to a great annual saving in fuel costs.
B9	For ships and airplanes, if friction resistance can be reduced by just 2-3%, a substantial amount of savings in annual fuel cost can be achieved.
B10	Because a mere 2-3% reduction in friction resistance in a ship or airplane can make a substantial savings in annual fuels costs possible,
B11	Because a tremendous amount of fuel consumption volume is saved with only a 2 to 3 percent decrease in the friction resistance of ships and planes,
B12	Since in the shipping and airline industries a significant yearly fuel costs savings can be achieved from even a slight 2 to 3% reduction in frictional force,

While it is impossible to be sure, the triggers for the mistakes in this sentence appear to have been not translating 低減できれば (*teigen dekireba, could be reduced*) as a conditional and the verb of the following clause 節約できる (*setsuyaku dekiru, would/could be saved*), which is non-past, as something that could happened, but has not happened yet. The other complication is *nenkan soutou ryou no nenpi, fuel consumption of an annual equivalent amount*, another short modification phrase similar to those that have been problems in earlier passages. The main problem here appears to

be the word 相当 (*soutou*, *equivalent to*), which is causing difficulties in two ways that have appeared before. First, it has a range of meanings, including *equivalent to* (the first definition in *Kenkyusha*), and *fairly, rather, considerable* (Watanabe 2003: 1511). The first of these was the one needed to correctly translate the phrase, but nine of the translations used some form of *rather* or *considerable*, with B11 pumping this up to “a tremendous amount”. (A bit of basic physics might help here.) Translation B3 used “equivalent”, but attached it to “year” instead of the 2% to 3% given earlier in the sentence, leading to the second difficulty: sorting out the modification.

In the phrase, *nenkan* (*annually*) is an adverb modifying *saving*; most of the translations made this link, despite the distance between the two words (7 characters). Translation B3 made the amount “equivalent to a year”, B4 had “annual amount”, and B11 omitted it altogether. None of the translations connected *soutou ryou* (*equivalent amount*) back to the percentage of reduction in friction resistance while four of them (B7, B8, B10, and B12 broke *soutou* away from its head noun and used it to modify *savings*. As a result, the cohesion of the sentence was broken and at the same time, the meaning was lost.

The mistakes with this phrase provide further support for considering shorter modification phrases than I originally anticipated as problem points in short-passage tests as well as further evidence that “complex noun phrases” as identified by Hale and Campbell (2002) are language independent difficulties. The phrase also brings up questions about one of the stated abilities of ILR reading level 2+: the ability to use “real-world knowledge to make sensible guesses about unfamiliar material” (ILR 2015). The problems with the final sentences in the two technical passages at ILR Level 2+, in which the mistranslations resulted in sentences that were contrary to fact, lead to questions about whether the translators were simply not reading at Level 2+ or whether the additional cognitive challenge of translation in addition to reading made processing more difficult.

8.2.4 Summary of errors and comparison with general passage B

There were more tense errors in technical passage B than in the previous technical passage, with the facility value for the final verb in sentence 1 at 0.55. When inclusion of a sense of a conditional rather than tense of the final verb is scored in the final sentence, the facility value for the verb drops to 0.09. This suggests that scoring verbs

does result in useful information for this passage and also that the text is more complex than technical passage A.

The two scored cohesion markers were *zensha* (*the former*) in sentence 3 and the link of *amount equivalent to* back to 2%-3%. The facility value for the first was 0.45 (the mistranslation in B11—“as in the latter case”—had the modification correct and so was not counted as a mistake). The facility value for the second was zero. These values indicate that scoring cohesion markers provided useful information in this passage. A summary table of the verbs and identified errors is provided below.

Table 8.4. Technical passage B: errors and facility values

Translation No.	Sentence No.								Total No. of errors
	1	1	2	3	4	5	5	5	
	自然に 学び、 learn naturally	Verb	Verb	前者 (<i>the former</i>)	Verb	Verb	Sense of a conditional	年間相当 量、an annual equivalent amount	
B1		V		CM			C	CM	4
B2	P		V	CM			C	CM	4
B3							C	CM	2
B4		V	V	CM			C	CM	5
B5								CM	1
B7							C	CM	2
B8	P			CM	V	V	C	CM	6
B9		V					C	CM	3
B10		V		CM			C	CM	4
B11	P						C	CM	3
B12	P	V		CM			C	CM	5
Facility value	0.64	0.55	0.81	0.45	0.91	0.91	0.09	0.0	

P is a parsing error; V is a tense error; CM is a cohesion marker error; and C is a conditional tense error.

This group of translators had more trouble with the cohesion markers in the technical passage, which was at a lower reading level, than in the ILR reading level 3 general passage. I suspect this was because the markers in the general passage were all explicit, and could be produced in the target language nearly automatically. The two cohesion markers in this passage were not only implicit, but also were part of modified noun phrases, resulting in a double whammy. In addition, both can be seen as elements requiring inference and following an argument, features that are characteristic of a reading level 3 text. As with technical passage A, the facility values suggest that the reading level 3–like elements caused the most difficulty.

Looking at the passage, rather than the translations, in light of the facility values, it would appear to be rather easy up until the final sentence. Of course, since only a

limited range of markers is being examined, problems with word choice (e.g. translating *turbulence* as “confusion” [translation B2]), syntax, and micro-linguistic elements have been ignored. Conversely, facility values for the last sentence are notably lower than the set limit of 0.25. While that suggests that the passage was too difficult for this population, on the other hand it provides useful information about the translators’ ability to follow an argument and reproduce it in the target language, which would be very much of interest to a translation agency or employer.

One also might wonder whether the test environment contributed to the extremely low facility values for sentence 5. One could claim that the real world knowledge needed to resolve the sense of the sentence is simply comparing the idea of a very small reduction in drag with saving a large amount of fuel and thinking, “probably not”. An Internet search would help confirm that intuition but that resource was not available to the test population. However, some perceived problem would have been necessary to trigger a web search, and I am not convinced that the majority of the translators realized that they had gotten off track.

9. Discussion

This chapter discusses the results for the scored items, problem points, and facility values, as well as the results of the close reading and connections with earlier studies. It concludes with a discussion of the results related to the differences between ILR reading levels 3 and 2+.

9.1 Scored items

The items selected for scoring are discussed in terms of whether they could be linked with a text feature of interest, whether they provided useful information, and whether they might be predictive. Facility values are discussed in light of the types of information they supply and possible uses

9.1.1 Cohesion markers

9.1.1.1 Tense

One of the aims of this study was to determine whether it is possible to identify text characteristics that can be linked to translation product evaluation results that a user of the product would be interested in and, if so, whether such items can be objectively scored. Source text cohesion is very much connected with usability. Halliday and Hasan provocatively noted that one of the rare examples of a non-text, i.e. a text without consistency of register and internal coherence due to lack of cohesion, is “bad translations” (1978: 24-25). Therefore, one of the first features a translation commissioner or user might want to look for is cohesion, or whether the translation being evaluated is *text* or *non-text*.

While native and near-native speakers generally have a good notion of whether a text is cohesive (Halliday and Hasan 1978: 1)²⁰, translation testing and evaluation should be based on something other than reviewer intuition. The argument in this research has been that because tense is a necessary cohesion element in English and

²⁰ I wonder if this is true of texts at a higher reading level. What does a complex text look like to those not reading at the NAAL proficient level?

there is good agreement on what constitutes a tense error, tense can be used to evaluate passage cohesion.

Evaluation of the texts supported this contention for the general passages, where there were a significant number of tense errors—only four of the 36 translations had all the final verbs tenses correct. It could be argued that this represents a deficiency in grammatical competence and that some of the translators may have been suffering from such a lack of ability. This is particularly possible since the population included people working from their B to A language. However, for the purposes here, it doesn't matter *why* the tense was incorrect, but simply that the incorrect verbs reduced the *usability* of the translation product.

Three of the four translations with no tense errors were of general passage A, and the remaining one was of general passage B. Both passages included time references and were also about changes over a defined period of time, and both these elements may have helped tense choices. At the same time, the incorrect choices in this type of passage would give a reviewer an objective basis for saying that the translator did not understand or was unable to represent important elements in the text.

General passage C also had a time reference (“beginning of the Edo period”), but none of the four translations had all of the verb tenses correct. In addition to signaling a lack of cohesion, the examination of tense also showed that the use of the present narrative in Japanese was a problem for this small group of translators, who tended to reproduce it literally in English with a present tense.

Do the tense errors in the general passages allow us to make objective judgments about the cohesiveness of the translations? I think that they do. A text in which the majority of the main verbs are in a tense that is not taken from context and the tense of the preceding verb is obviously in need of editing. Furthermore, seeing such a product, a reviewer would begin to think that the translator would not be a good bet for future projects. In addition, the evaluation of tense and other cohesion markers shows that many of the translators in this population were translating sentences—and sometimes phrases—and then not fitting these together to create a cohesive text. Whether that stemmed from a lack of “textual competency” or proofreading/editing skills is not of concern in the professional setting.

The utility of tense markers was not as clear in the two technical passages, which had simpler tense and cohesion structures than the general passages. The tense choices were almost all correct in technical passage A, indicating that examining tense resulted

in no useful information. Verbal markers revealed some interesting problems in technical passage B, but these were not related to tense (overly literal translations of the copula in sentence 1 and failure to fully represent the conditional sense in sentence 5).

That tense of the final verbs could not used to evaluate cohesion in the technical passages, which were at ILR reading level 2+, is useful in creating a profile of the test passages. First, it tells us that the tense structure of the texts is relatively simple and suggests that information presented is comparatively loosely ordered. Second, it goes some way in confirming (in this small sample of passages) that the sort of inference needed to make cohesive tense decisions comes into play at reading level higher than 2+. That will be discussed further in section 9.4 below.

This finding also points to the double use of scored items: they can be used either to evaluate a translation product or, in connection with pre-testing, create a profile of a potential test passage.

9.1.1.2 Other cohesion markers

The time-related implicit cohesion marker in general passage A, sentence 1, was loosely correlated with tense decisions, but in general scores for the other cohesion markers and tense items were unrelated.²¹ This suggests that the two items were indicating different aspects of cohesion, making it worthwhile to look at both. While implicit cohesion markers might be expected to be more difficult than explicit ones, that was not the case in the general passages, in which two of the explicit markers in general passage D had facility values of zero. That seems to have been related to problems in seeing the structure of the editorial. It also suggests that the implicit and explicit cohesion markers are pointing to structural and syntactical cohesion, an element separate from the temporal cohesion targeted by tense. This makes the use of both sets of markers valuable in evaluating translation products evaluation and in creating passage profiles.

Scoring the other cohesion markers can be more subjective than scoring tense. One can question whether omitting a marker was an error or not and there can be doubts about word choice for some markers. In general though, cohesion markers remain reasonably objective items that can be marked as correct or incorrect.

In general passage D, the selected cohesion markers also function to identify whether the translators had recognized the standard form of an editorial.

²¹ As an extreme example, in translation B2, all of the tense markers were incorrect and all of the cohesion markers were correct.

Cohesion markers played less of a role in the technical passages, but the mistakes were perhaps more interesting than whether or not a marker had been omitted. Passage A had a cohesion problem in the internal structure of one sentence (see Halliday and Hasan 1978: 7-9 for a discussion of cohesion within a sentence). A failure to grasp the full range of meanings of *shikamo* (*on top of that*) appears to have contributed to the low facility value for resolution of the final sentence. The errors in technical passage B, sentence 3, involved a combination of back reference to an unstated noun—making it an implicit noun—and resolving a modification string. This suggests that looking for similar items that are less obviously cohesion markers to score as indicators of cohesion might be worthwhile in texts at this level.

Seeing several cohesion marker errors in a short-passage would suggest that a translator was not seeing how the parts of the text fit together and would not be a safe choice for similar assignments, particularly if there were also tense errors.

9.1.2 Problem points

9.1.2.1 Modified nouns

Modified nouns with inclusive strings of 15 characters or more provided an objective (i.e. a clear choice between correct and incorrect) item for scoring syntax, helping to create a fuller picture of the translation product. In the case of the longer strings in which elements modifying the head noun were some distance away from it, it was also possible to gain some insight into whether the translator was following relationships within the passage and making the correct inferences. To my surprise, *shorter* modification phrases were a point of difficulty. One of the main reasons for modified nouns being scored as unacceptable was the failure to keep the head noun in Japanese as the head noun in English, hinting that many of the translators did not understand the concept of “head noun”.

Modified noun constructions might be predictive of future performance in regard to syntax and the ability to connect text elements at a distance from each other, but any item selected would need to be checked to ensure that it required inference for resolution. In this set of passages, some of the long strings of modification could be resolved without reference to the rest of the text. They also had facility values showing that they were easy test items in this population.

The syntax of the sentences in which the modified nouns appeared was also scored. In most cases, the facility values were either the same as that of the modified noun or slightly higher. Therefore, including sentence syntax as an item to be scored did not add much in the way of additional useful information. At the same time, I found that determining whether or not syntax was correct was difficult to do independently of other sentence features and also that some decisions about correctness bordered on being preferential. For all three of these reasons, I think sentence syntax may not be a good candidate for an item to be dichotomously scored.

9.1.2.2 Passive constructions

While passive constructions have been identified in earlier studies as a potential source of translation difficulty (Campbell 1991; Campbell and Hale 1999; Campbell 1999; and for reading difficulty, Tateishi et al. 1988), in this study they did not add a notable amount of information to the analysis of the tense of the final verb. Also, except in cases in which resolving the passive can be explicitly connected to a feature that requires inference or a decision linking the solution to another text element (such as resolving the subject of the last sentence in technical passage A), the score would be linked to a grammatical element and so be at a lower text level than the type of text elements I am trying to score here.

However, in terms of providing supporting data for previous research, the facility value range for the passive constructions in this study (0.25²² to 0.78) does suggest that passive constructions are difficult in the Japanese-to-English language pair.

9.1.3 Facility values

Calculating facility values was a major part of this study. It should be kept in mind that these values were calculated to provide information about the test items and passages and not about the translators or translations, although they do allow qualified statements to be made about performance and the translation product. Facility values permit quantified statements to be made about the difficulty of text features in a defined population. For example, it is possible to say that in a population meeting the prerequisites for taking the ATA certification examination and sitting the exam in

²² This value was in general passage 3, which had only four translations. The next lowest value was 0.33 (general passage 4, with nine translations).

comparable conditions, the task of translating Japanese present tense verbs to a corresponding English present tense was easier than translating a Japanese verb in the narrative present to a English past tense verb that preserved text coherence. One can also say that for this restricted population and small passage set, tracing back a null pronoun or subject was difficult, as was correcting matching a modifier to a head noun when these were separated by other modifiers. Thus, having facility values is a way of avoiding the “difficulty is subjective” problem, of identifying text features that provide useful information, in the current case about whether a cohesive text has been created, and—even though I am not focusing on performance here—of pinpointing lacunae in translator training or performance. They would also be helpful in the creation of a bank of items that would be useful to test in this language pair.

Facility values also offer a way of comparing text difficulty over time and across languages as long as the test population is consistent. In this study, the facility values showed that general passage C was easier than the other three general passages. Conversely, the value of 0.09 for the expression “making the continuation of life uncertain” (technical passage A), which involved coming up with wording that was quite different than the Japanese source, suggests that the item, and so perhaps the passage, was too difficult for the test population, particularly in an examination environment in which people tend to be cautiously literal. However, the value also permits one to identify this as a test item that could be linked to congruity judgment or transfer skill in another test setting.

While facility values can be calculated after the fact, as I have done here, the best practice would be to calculate them in a pre-test population mirroring the test population, then again in the test population. However, for organizations that do not have the resources to do pre-testing, periodic post-test checks using facility values would still be beneficial as this would allow monitoring not only the passage difficulty level, but also the appropriateness of the match between test passages and test population, potentially signaling the need to reconsider test pre-requisites and who will be tested.

One of the most useful features of facility values is that they force the test designer to select items to be scored and to link these with an identified element that is being tested. Ideally, the *what* to be tested should be based on a needs analysis (Lafeber 2012b: 5), job task analysis (Koby and Melby 2013), or even on a running list of “problems we are seeing in recent translations”. The skill or problem to be invoked can

then be connected to specific text features. If that is not possible, then the testing entity should reconsider whether a short-passage translation test is the best means to its evaluation ends.

A commonly cited problem with the short-passage examination is that it is often unclear exactly what is being tested, with the result that success or failure is ultimately determined by whether or not a candidate meets grader expectations—expectations that are frequently unclear or even unknown to the test taker (Nord, 2005; Schäffner 1998; Kiraly 2000; Campbell and Hale 2003; Williams 2004). Currently, the two most common forms of evaluation are rubrics (holistic methods) and error point systems. The first normally focuses on supposed translation abilities. The method is grounded in Bachman's (1990) work in language testing and is designed to provide feedback as well as a score. The various error points methods (either points off or points accumulated), such as the systems used by the ATA and many translation agencies, set various categories (e.g. mistranslation, wrong term, omission, addition, overly literal, syntax, grammar, spelling, etc.) and count the number of errors, which may or may not be weighted, in each. This method is often used for decision-making (i.e. to certify or not, employ or not) rather than feedback.

While both systems have merits, neither links specific source-text features with specific target-text characteristics that the potential user of the translation would like to see. In other words, they do not set individual items for testing.²³ As a result, it is not possible to calculate facility values, with the corresponding loss of the various advantages cited above. As I hope that I have demonstrated in this research, general source text features such as syntax or cohesion can be signaled by specific features in the source text and objectively marked. I am not proposing doing this as a complete method of evaluating translations, but rather as a feature that can be added to other methods.

Error point deduction/addition evaluation systems tend to focus evaluator attention on elements within sentences rather than on the links between sentences. Any element that helps turn the attention to the larger picture would assist in countering this, particularly if a cohesion/syntax review were to be done before a review of grammatical elements.

²³ Categories like “terminology”, “false friend”, and “mistranslation” do target the source and target texts, but the first two are usually errors at the word level and so can be considered a micro-linguistic or grammatical feature (important to note and score, but perhaps not very revealing in terms of predictive performance), while the later does not explain the *how* of the mistranslation, making the source-target link quite broad.

9.2 Results of close reading

The close reading of passages was an indirect way of asking the translators about their manner of working and what they found difficult (in the sense of places where they made mistakes) as well as of looking for text features that might be good candidates for future test items.

The translation unit has been an ongoing topic of interest in Translation Studies (e.g. Gile 1995; Kussmaul 1995; Krings 2001; Toury 2006), and has usually been considered as a unit in the source text. Eye tracking studies (Dragsted and Hansen 2008: 9-29; Alves et al. 2010: 109-142) mark translation unit boundaries instead by pauses during the translation process, which of course vary by person (and probably by time for the same person). While it is impossible to infer the translation unit with confidence from a set of translation products, in the translations under examination here, the largest unit appears to have been the sentence, as evinced by the lack of tense cohesion and mistakes with the implicit and explicit cohesion markers. This was particularly obvious in the universal failure to see the structure of the editorial in general passage D. Problems in sorting out the noun modification strings suggested that the translation unit was at the sub-sentence level for some of the translators.

While a few of the translators may have read the test passage prior to starting their translation, the translations generally read as though they were done segment by segment. That was particularly obvious in the long final sentence of general passage D. Only two of the nine translators started their English text with the information that came at the end of the Japanese sentence, showing that they had read the single-sentence paragraph through before starting to translation. This is also consistent with eye-tracking data showing that translators often do not read the source text before starting work (Sharmin et al. 2008: 36).

Text features in addition to those set for marking that were frequency resolved incorrectly were shorter modified noun phrases than I originally anticipated: 省別採用, *shoubetsu saiyou*, *hiring by ministry* and 一括採用, *ikkatsu saiyou*, *collective hiring* from general passage D were notable examples. The sentence in which both terms appeared had a facility rating of 0.33 for tense. This was mirrored by a facility rating for the passive problem point of 0.33, but in this instance the passive seems to have been a genuine difficulty, apart from its tense, because several of the translators did not

correctly identify the subject of the sentence (*ikkatsu saiyou, collective hiring*). These elements in addition to the modified nouns may have a synergistic effect on the difficulty of the sentence overall.

The incorrect choice of a definition for common particles and conjunctions was also a recurring problem (e.g. か, *ka, or* being used as a question word and 通じて, *tsuujite, throughout* being translated as “due to”). This suggests that in some cases particles and conjunctions could be set as items to be scored as indicators of syntax. They are certainly markers of whether the translator understood and was able to reproduce the relationships between units of meaning in a sentence.

Whether a character string has been parsed correctly is another possible marker of both syntax (in terms of resolving modification) and of overall understanding of context.²⁴ A good example was “learn from nature” being turned into “learn naturally” in technical passage A.

Agglutinative elements added to verbs were not a problem for this population. However, none of the passages had verbs of giving or receiving; parsing the endings for such verbs might have resulted in subject/object difficulties.

The last sentence in each of the technical passages was a problem in that the majority of the translations had a sentence that was contrary to fact that one might have expected the translators to catch on re-reading. The fact that this did not happen is interesting in light of the connection between difficulty and cognitive effort. Text difficulty has been defined as “a function of the cognitive effort required to process the item in questions and convert it into the target language” (Hale and Campbell 2002), various eye-tracking studies have equated gaze fixations with cognitive effort (Jakobsen and Jensen 2008; Jensen 2009; Alves et al. 2010), and a recent study on measuring translation difficulty defined it as “the extent to which cognitive resources are taken up by a translation task for a translator to achieve objective and subjective performance criteria” (Sun and Shreve 2014: 99). However, some of the errors seen in this set of translations, and particularly in the final sentence of technical passage A, suggest that people are sailing over features constituting text/translation difficulties and resolving them incorrectly with minimal expenditure of cognitive effort. At the same time, the facility values for the items quantitatively show these features to be difficult for this population. Empirical studies in psychology offer explanations for this phenomenon and

²⁴ Of course, it is also a marker of the translator’s grasp of grammar.

also suggest that equating *difficulty* with *cognitive effort* will inevitably be a problem when studying translation accuracy.

First, there is the Dunning-Kruger effect based on studies done at Cornell University and well supported by a series of follow-up papers (Dunning et al. 2003; Ehrlinger et al. 2008, among others) in which the fact that an individual is unskilled in a particular area is the very thing that keeps the person from recognizing their own incompetence. An example that Kruger and Dunning give is the ability to write English with acceptable grammar:

The skills that enable one to construct a grammatical sentence are the same skills necessary to recognize a grammatical sentence, and thus are the same skills necessary to determine if a grammatical mistake has been made. In short, the same knowledge that underlies the ability to produce correct judgment is also the knowledge that underlies the ability to recognize correct judgment. To lack the former is to be deficient in the latter. (Kruger and Dunning 1999: 1121)

Rephrased in terms of translation, one could say that the skills needed to recognize an argument in the source text are the same as those needed to see deficiencies in the target text.²⁵ According to Kruger and Dunning, the solution to this is to raise metacognition skills overall. Therefore, an interesting question that could be tested is whether raising source language reading skills improves both target text accuracy and the ability of the translator to recognize errors. However, the presence of this combined state of ignorance and unawareness suggests that a certain number of translation errors will happen without there being much cognitive effort.

Perfetti et al. found a similar phenomenon in the acquisition of reading comprehension skill:

Readers who strive for coherence in their representation of a text must be able to monitor their comprehension. Monitoring allows the reader to verify his or her understanding and to make repairs where this understanding is not sensible. Skilled readers can use the detection of a comprehension breakdown (e.g., an

²⁵ In my experience, translators who fall into this unfortunate state of unknowing are the ones likely to argue, “But that’s what it says” when a error is pointed out to them.

apparent inconsistency) as a signal for rereading and repair. Less-skilled readers may not engage this monitoring process. (Perfetti et al. 2005: 234)

The second argument for not connecting accuracy and difficulty defined as cognitive effort is the work done by Dietrich Dörner on why people fail when faced with complicated problems—and translation does count as a complex task. Based on a series of experiments in which participants responded to complex situations via computer simulations, Dörner noted several characteristics that distinguished successful from unsuccessful participants. For example, successful subjects proposed hypotheses about the effects of their actions that they then went on to test (i.e. expended cognitive effort) while unsuccessful subjects considered the first proposal they generated as “truth” (conserved cognitive effort) (Dörner 1996: 24). One can see a bit of the same thing in the use of the most readily available—but wrong—definitions of words in some of the translations. Successful participants considered causal relations while the unsuccessful ones saw events as unrelated (Dörner 1996: 24), which sums up the difference between the better and worse translations in this study.

Looking at Dörner’s experiments in terms of translation, one could consider some of more egregious mistakes—having Japan responsible for 90% of the world’s greenhouse emissions or translating “learning from nature” as “learning naturally”—to result from producing a final version from the start in which sentences, and even parts of sentences, were viewed as unrelated units rather than generating hypotheses about possible relations in the text. Dörner notes that: “[t]o deal with a system as if it were a bundle of unrelated individual systems is, on the one hand, the method that saves the most cognitive energy. On the other hand, it is the method that guarantees failure” (Dörner 1996: 88).

This view from a leading psychologist counters the idea that items of difficulty can be identified by pinpointing places in a text where cognitive effort increases and supports instead the notion of another type of difficulty in which mistakes occur because a minimum amount of effort is being expended. It would be interesting to have eye-tracking studies of this hypothesis.

In his experiments, Dörner found that most failing participants did not achieve an overall view of the system they were presented with, nor did they see the interactions within a given system (1996:87). The analysis of tense in this study suggests translators making multiple errors similarly lacked a larger vision of the text they were translating.

For example, they do not appear to have analyzed the source texts for cohesive features or thought about how an initial choice would govern later tense decisions.

Studies in second language reading have noted similar problems with clarifying “the semantic relations among the sentence elements” (Koda 2004: 109):

To benefit from the information provided by connective devices, the reader must have not only a sound grasp of their functions but also full understanding of the clauses to be connected. If what is to be connected, and in what way, is not clear to the reader, the relational information, however explicit, is unlikely to be incorporated in ongoing sentence interpretation. (Koda 2004: 109)

9.3 Results from earlier studies

Although the passages examined in this study were not originally selected to contain a particular set of features that could be considered general or language specific problems, it was possible to look at some elements similar to features identified in earlier empirical studies on translation difficulty to see if they were problem points here.

9.3.1 Hale and Campbell (2002)

In viewing items similar to those described in Hale and Campbell (2002), I found that that nouns with long modification strings and culturally specific terms corresponded in some degree to their categories of complex noun phrases/modified complex noun phrases and official terms. I did not find that having more available options for an item increased difficulty as measured by facility value, although it could be seen as increasing difficulty in terms of cognitive effort, which is their definition. With one exception, discussed below, having a number of possible choices did not reduce accuracy. However, the presence of a range of possibilities was seen as having the potential for increasing evaluation subjectivity both here and in the Hale and Campbell study. For me, that sends up a red flag in terms of marking such items (the example here was the Japanese word *amakudari*, which refers to a particular relationship between the senior government bureaucracy and private industry in Japan). In test situations without Internet access it is questionable whether error points should be deducted or added for

such terms. If they are scored, the evaluators should articulate precisely what predictive trait or competency (depending on the test model) is being evaluated (e.g. in some settings testing cultural awareness would be beneficial, if that were the type of future content to be translated). An argument such as “They (the candidates) should know this” is not valid unless someone can say *why* knowledge of that particular term is important. Also, it would be difficult to argue that someone who came up with a good solution to *amakudari* will be equally clever with other culturally specific words. However, in a translation agency test done in normal working conditions (web access and the ability to include notes), such words could be set as items to test research skills and adaptation (Vinay and Darbelnet 2000: 90-92) and scored specifically for these items.

The clause *seibutsu no seiku no obotsukanakunaru* (“the continuation of life would become impossible”) from technical passage 1, although not an official or culturally specific term, was another element with a wide range of possible translations and, in this case, was linked to low accuracy (facility value 0.09). Any solution to this that would read like natural English involved moving away from the equivalents given in the *Kenkyusha* dictionary. Such a clause could be used to score congruity judgment on a translation agency test with clear instructions about the target audience and use of natural English, as well as permission to include translator’s notes. However, error point systems, including the ATA Framework, tend to have categories for either “too free” or “too literal”. In the current ATA Framework for Standardized Error Marking: Explanation of Error Categories, candidates are told that:

A **faithfulness** error occurs when the target text does not respect the meaning of the source text as much as possible. Candidates are asked to translate the meaning and intent of the source text, not to rewrite it or improve upon it. The grader will carefully compare the translation to the source text. If a “creative” rendition changes the meaning, an error will be marked. (ATA 2015, emphasis in the original)

Faced with that definition, translators would have a quite understandable tendency to be conservative. Therefore, unless such structures can be explicitly related to a desirable and preferably predictive element being tested, they should not be scored,

since any score would be based on a subjective evaluation (which might further involve discretion over the number of points to subtract or award).

Hale and Campbell noted that complex noun phrases were an example of “constrained equivalent” with “multiple options”. The constrain comes from the necessity of the head noun in the source language remaining the head noun in the target language, something that several translators in this study found difficult. They also observed that the “real” difficulty ... is found at the level of the syntax and semantics of the whole matrix phrase”, a finding that was reproduced here. I agree with their observation that the “inaccuracy was obviously produced by a lack of comprehension of the source text rather than a problem of production into the target text” (Hale and Campbell 2002: 25-27).

Hale and Campbell further noted that in order to progress in the understanding of translation difficulty:

further empirically based research is necessary to map item types and to characterise the difficulties they cause. This work needs to be done with multiple language combinations to test the universality of these difficulties, and to provide richer cognitive explanations for them. (Hale and Campbell 2002: 29)

I have tried to make a bit of progress toward achieving that goal in this dissertation. However, to make mapping difficulties significant in improving “the effectiveness of the teaching and assessing of translation” (2002: 30) I think that simply mapping them is insufficient. For assessment in particular, we need to go further and connect source text difficulties with features that translator users want to see in the translation product.

9.3.2 Eye-tracking studies

Because readability indices, which were a major part of the difficulty studies using eye tracking, are not comparable in Japanese and English, readability ratings of the translations of the general passages was done. This provided supporting evidence that the general passages had a high degree of text complexity and that general passage C was easier than the other three passages. The readability rates also confirmed the ILR reading level evaluation and the pattern seen in the facility values.

Sentence length was most likely a difficulty in general passage D, where a single sentence constituted the entire paragraph. One of the features that probably made general passage C easier than the others was that it had roughly the same number of characters but nearly twice as many sentences. In general though, this study did not contain enough passages of sufficient word count to be able to evaluate sentence length as an item of difficulty.

Knowing whether sentence length affects difficulty would be useful for consistency in passage selection, but it is not clear that setting sentence length, as opposed to problem points within the sentence, as a separate test item would contribute to an evaluation. The only evaluation items set for sentence 7, the very long sentence in general passage D, were final verb tense and the explicit cohesion marker at the beginning of the sentence, which had facility values of 0.11 and 0.0, respectively, successfully flagging this part of the text as extremely difficult. Both values can be linked to translator failure to see the editorial structure, but there is, of course, a strong possibility that the length of the sentence also contributed to the very low facility values.

In terms of evaluation, scoring the tense of the verbs within the sentence would provide more information about the structure of the sentence and its usability than would a “long sentence acceptable/unacceptable” score and could also be linked to other interesting features such as syntax and cohesion.

One of the most fruitful difficulty indicators used in the eye-tracking studies was “structural complexity” cited by Mishra et al. (2013: 348) and based on earlier work by Dekang Lin. This fit very nicely with Hale and Campbell’s identification of “complex nouns in matrices” as not only a difficulty, but also as an item with low accuracy (2002). The idea of “the total length of the dependency links in the structure” (Lin 1996: 730) is useful in explaining why some modification strings are more difficult (have lower facility values than others). In addition, Campbell and Hale (1999) found the complex nouns to be a difficulty in the English to Arabic, Vietnamese, and Spanish language pairs, Lin between Dutch and German, and here, from Japanese to English, suggesting that structural complexity as defined above is a difficulty across languages. Setting this structure as an item to be scored could be linked to syntax, modification, and inference.

9.3.3 Sun and Shreve

A recent study by Sun and Shreve carried out empirical research on translation difficulty, which they defined as “the amount of cognitive effort required to solve a problem”. Translation difficulty is connected to “mental workload” and “cognitive load” (Sun and Shreve 2014: 99). The study considered some of the questions that were under examination in this study, including “whether translation quality scores (i.e., accuracy) was a performance indicator that can be used to represent translation difficulty”, whether readability formulas predict level of translation difficulty, and whether translation difficulty for a translator can be known beforehand (Sun and Shreve 2014: 102). It should be noted that “translation difficulty” in all the above questions equates to “cognitive load”.

Sun and Shreve found that accuracy (scored holistically and with errors points, and not with item facility values) and translation difficulty had essentially no relationship (2014: 108-109). Hale and Campbell, in their study defining *translation difficulty* as the cognitive effort needed to choose between multiple options and *accuracy* as appropriateness in terms of pragmatic equivalence (i.e. rater opinion) also saw no “clear correlation between choices and accuracy” (2002: 29). These results support the argument made in 9.2 that cognitive effort may not be related to translation difficulty, defined here as a facility value of 0.70 or less on a test item that can be objectively scored for a specified population (a definition that would be considered “accuracy” in the other two studies), because failure to come up with a correct solution to a problem is often the result of not seeing that there is a problem in the first place (Kruger and Dunning 1999) and strategies that conserve mental effort (Dörner 1996).

Sun and Shreve conservatively concluded that “the hypothesis that a readability formula can be used to measure translation difficulty could not be falsified”, and that “average sentence length, word difficulty, and other statistically measurable factors might be useful to measure or predict the translation difficulty of a text” (2014: 112-113). Translation difficulty was measured by translator perception of difficulty as measured on a subjective task load index, with the result that objective and subjective values were being compared. Subjective perception of test passage translation difficulty

after a first reading²⁶ and then after translation was considered possibility predictive of the translation difficulty (i.e. amount of cognitive effort required) of a passage.

9.4 Verification of ILR reading level and item difficulty

9.4.1 Study assumptions

This study has involved two assumptions related to ILR reading levels. First, that the distinctive features of ILR reading level 3 passages make them more difficult (i.e. there will be fewer correct translation solutions in a defined population) to translate than ILR reading level 2/2+ passages and second, that text elements related to these features can be identified in the source text and scored as correct or incorrect in the target text. The specific feature chosen for examination was inference, the ability to “relate ideas and ‘read between the lines’”, which is not present in reading level 2. At level 2, the reader is “insufficiently experienced with the language to draw inferences directly from the linguistic aspects of the text”, which is precisely what readers or translators are being asked to do at level 3 (ILR 2015).

I assumed that target text cohesion signaled whether or not ideas and linguistic structures in the text had been correctly related. I realize that some of the translators may have had a clear view of the relationships but were not able to reproduce these in English. However, since this study does not examine individual abilities and skills, but rather translation product features and because the research questions involve the relationship of the source text to the target text, *why* and *how* questions concerning the translators have been put aside.

Tense, which is recognized as a necessary element of cohesion in English, and which can be objectively scored as correct or incorrect, was used as an indicator of whether the text cohesion present in the source text had been maintained in the target text. Implicit and explicit cohesion markers were also scored as correct or incorrect and long modification strings were examined to see whether the source text connections were present in the target text.

²⁶ The study limitations did not note that a methodology requiring the translators to read the passage before translation—which most translators do not seem to do—may have introduced noise into the results.

9.4.1.1 Can elements related to inference be identified in the source text and scored in the target text?

To begin with the second assumption, was the idea that it is possible to identify text elements related to inference in the source text and score these as correct or incorrect in the target text supported? I think the answer is yes. The first paragraph of a translation of general passage B in which all the tense decisions were unacceptable follows:

Throughout the period of Japan's high growth, the forms and functions of the family, the basic unit of society, have changed. The number of working families has increased during this period and a division of labor by sex has become fixed whereby the husband works outside the home and the wife engages in housework and raising the children. However, due to the increase in both spouses working, at the same time a new movement has started that attempts to reconsider the sexual division of labor within the household and to create an equal relationship between the spouses. (B2)

This is a collection of sentences, rather than a cohesive text. A similar loss of cohesion occurs in a text with incorrect cohesion markers:

The overall standard of studies on the Japanese politics by Japanese scholars so far has not been high, even though there have been a small number of outstanding studies. An antiquated academic tradition that the origin of the academics belongs to the foreign countries, combined with the restrictions by the Tenno system as well as political scientists' interest in modernization and democratization drove capable political scientists to studies on the politics of the politically advanced nations. Studies of the Japanese politics were often done as an avocation of these scholars, or as political journalism or a cultural critique. (General passage A, translation A2)

9.4.1.2 Do the distinctive features of ILR reading level 3 make passages at that level more difficult to translate than passages at reading level 2+?

What is striking about the translations cited in 9.4.1.1 in terms of supporting the first assumption, that the features that define texts as ILR reading level 3 make them more difficult to translate than texts at reading level 2/2+, is that in translation, the tight links

between sentences and strictly ordered information that are typical features of reading level 3 have essentially been stripped away, reducing the texts (insofar as the first example is a text) to reading level 2/2+. Conversely, the facility values for the sentences in the technical passages with reading level 3 features were low (the final sentence in each case), ranging between 0.0 and 0.36. This suggests that there is a genuine difference in difficulty between the levels and that it is connected to cohesion and linking aspects of an argument together.

However, only a very small sample has been examined here and the utility of the texts was further reduced by not having been specifically selected to answer the study questions. Nevertheless, one could also say that finding evidence of the difference between the two levels “in the wild”—in an actual professional test setting—rather than as the result of an experimental design, further strengthens the argument there is a real distinction in difficulty between the two levels. This is an area requiring further research.

9.4.2 Implications for passage selection

One of the ongoing problems in translation testing is how to select passages with consistent levels of difficulty. The U.S. government has used the ILR reading level system for decades, and it was adopted by the ATA in 2008 (Koby and Champe 2013: 161). As this study suggests, there seems to be real differences between passages at reading levels 2/2+ and 3, and most people commissioning or using translations would probably like some assurance that the translators they employed could function at level 3.²⁷ The use of the ILR reading levels is a step in the right direction in terms of test passage uniformity, but as can be seen even in the small sample here, the range and types of difficulties can still vary (e.g. the short sentences in general passage C, the lack of tense variety in general passage D). Obviously, not all passages, even passages at the same reading level, will present the same opportunities for identifying text elements that can be linked to specific items of interest and objectively scored. Therefore, test designers need to first decide exactly what it is that they want to test (e.g. inference, congruity judgment, research skills) and then determine what source text elements can be reasonably thought to represent the item of interest and whether these link to target

²⁷ Assuming of course, that the commissioners or users were aware of how text types and difficulty levels can differ.

text elements that can be objectively scored as correct or incorrect. Then, they need to find a passage at the set reading level and with the necessary text elements. This makes passage selection onerous, but, in addition to ensuring greater passage uniformity, it also opens up possibilities for making short-passage tests useful by defining exactly what is being tested. Based on the facility values obtained for several of the translations in this study, as a reviewer I would feel I had sufficient grounds for saying that some of the translators in this population would be unlikely to be able to create a coherent and acceptable English text if assigned a Japanese passage that required any degree of inference to read and translate.

9.5 Potential applications

9.5.1 Potential testing applications

The use of facility values as a measure of translation difficulty as described in this study is not being proposed as stand-alone evaluation system for short-passage translation tests. The passage and translator profiles visible in the overall tense, cohesion marker, and problem point tables in Chapter 5 provide summary information of both passage difficulty (based on the defined items) and translator performance. In a translation agency setting, one might be able to use them to eliminate some of the translators from further consideration. However, additional evaluation of features such as word choice, grammar, and register would need to be done in most cases, and most definitely in a high-stakes test situation.

A primary benefit of linking specific source text features to a larger trait of interest and identifying target text elements representing this is that it forces test designers to identify precisely what it is that they are testing and perhaps even to think about why they are testing it. In an ideal testing situation, that should lead back to the creation of a design statement (Bachman and Palmer 1996: 86-90) or test specification (Fulcher & Davidson 2007: 52-61).

Another advantage is that there is an objective basis for scoring the items being tested. That scoring is frequently subjective has been mentioned by many researchers (e.g. Nord, 2005; Schäffner 1998; Kiraly 2000; Campbell and Hale 2003; Williams 2003). Examples of problems with subjectivity in scoring items with a wide range of

possible solutions and official/culturally specific terms were given in Chapter 7 and in Campbell (1991) and Hale and Campbell (2002). While subjectively in scoring short-passage tests can never be completely eliminated,²⁸ having some items that can be clearly scored acceptable or unacceptable—particularly items at the sentence level or above—provides something of a touchstone on which to ground an overall evaluation. It might also serve to reduce the “halo effect” in which especially good writing or a few clever solutions make the evaluator less likely to see errors in the text.

In addition, identifying source text items linked to target text solutions turns some of the evaluator’s attention back to the source text and language. Campbell and Hale (2003: 209) noted that most error marking schemes are heavy focused on target language errors (the ATA Framework for Standardized Error Marking is a good example [ATA 2015]), and consequently tend to ignore source language-related competencies or cohesion errors. Holistic scoring systems are perhaps less prone to this, since they may include a category for scoring “Source Text Meaning”. Interestingly, in the rubric Angelelli created for professional associations to assess translation ability, “Source Text Meaning” was the only category not connected to a specific competency (Angelelli 2009: 39-40). I do not want to enter into enumeration of various competencies here, but I will say that getting a sense of source language reading ability through evaluation of specific reading level indicators could be considered a first step in creating a predictive translation evaluation system. Also, the U.S. adult literacy statistics demonstrate that reading at higher text levels, even in the A language, is not something that can be taken for granted (Kutner et al. 2007).

As discussed in **9.4**, one of the drawbacks of the type of error analysis and facility value calculation done here is that it makes passage selection very labor intensive. Another possible drawback is language specificity, although this could also be seen as an advantage. Because the expression of tense as well as the conventions governing use of it are very different in Japanese and English, tense worked well as a sign of whether translators were using context and the flow of the argument to make tense choices. It might not work as well between other language pairs. Incorrect parsing of character strings, particularly when they could be broken up in a way to create a familiar, but incorrect word in context, occurred often enough in this short study that I would include this as an item to be scored in future evaluations, particularly since it can be linked to

²⁸ In English, even grammatical correctness can be a matter of debate.

coherence. This would also work in Chinese, but not in languages that have clear breaks between words. The search for language specific items is another area with potential for future research. However, the problem point analysis in this research confirmed that at least one item that had been identified in previous studies—noun modification strings—not only operates across a variety of language pairs, but can also signal whether inference is being used in the translation.

Another application of passage analysis using facility values is in pre-testing, which should be done if at all possible for high-stakes tests. If we look at the four general passages examined here as four candidate passages for a short-passage translation test,²⁹ the summary tables of tense and other cohesion marker errors and problem point errors can be used to examine passage suitability for actual exam use. Looking at the results for general passage C, two values at 0.25 for tense suggest that an interesting problem has been identified, but the lack of problem points and consistent facility values of 1.0 for the implicit/explicit cohesion markers suggest that the passage is too easy for the test population. Conversely, the fact that general passage D contains a sentence with a facility value of 0.11 for tense and 0.00 for an explicit cohesion marker flag this passage as too difficult for the test population. The facility values for general passages A and B are comparable, although the values for tense indicate that passage A is somewhat easier. Both have a figure close to the bottom of the acceptable range (0.27 and 0.25, respectively), suggesting that these passages may be able to separate the very good translators from the good translators.

9.5.2 Potential applications in translator training

The items selected for scoring have identified not only translation difficulties but also elements of possible focus in translator training. These include cohesion conventions in different languages, recognizing and reproducing links between sentences, and modification patterns. Kruger and Dunning noted that in order for “incompetent individuals” (their term) to both realize that they are incompetent and to improve, they “must come to an accurate understanding of why ... failure has occurred” (1999: 1130). Error marking systems point to mistakes, but do not necessary give a reason for them. The explicit links between source and target text and an element contributing to

²⁹ For consistency, one would want to have a unified test population, rather than four separate populations, when pre-testing passages.

successful translation (e.g. cohesion, correct modification) used for facility value calculation could equally well be used to provide feedback directed at making the translator consciously aware of how, for example, cohesion functions in both languages.

Exercises directed at such awareness could be as simple as circling words and structures indicating links between sentences. Awareness of modification could be increased by classic sentence diagramming and creating parse trees. From examining the translations, I had the impression that many of the translators did not realize that cohesion is created in different ways in different languages (e.g. by tense in English and with unstated subjects in Japanese) or that tense patterns do not function in the same way in the two languages (temporal cohesion need to be maintained in English, varied tense can be a sign of good writing in Japanese). Saying these things out loud, in the classroom and even in venues such as the ATA conference sessions, might cause some pennies to drop that would eliminate the types of errors seen in this population. Such knowledge, even though it is linguistically based, could be viewed as part of the declarative knowledge about translation that is part of the PACTE group model (PACTE 2011: 319).

A secondary aim of this study was to provide a data set that might be pedagogically useful in teaching both translation and Japanese. The tense analysis has shown that the Japanese narrative present is a problem in translation, with verbs in this form having consistently low facility values. There was also an inexplicable tendency for some translators to use the English present perfect in the translation of general passages A and B, both of which described changes over time. Referring back to an unstated subject or an implied relationship was a problem, as was resolving modification when the modifier was separated from its noun by other elements.

An interesting area of follow-up research, particularly in light of the NAAL statistics, would be trying to get a better understanding of reading problems I suspect are involved in the weakest translations and whether improved (improved via the sort of explicit knowledge mentioned above) source language reading would improve target language reading.

10. Conclusion

This study was prompted by the fact that short-passage tests are a frequent means of translator selection and certification, but that there has been almost no research on what is being tested, how scoring is linked to the objective of the test, or whether the test itself is useful for its purported purpose.³⁰ In addition, earlier studies have identified various elements thought to be points of translation difficulty. At the same time, translation tests based on passages selected according to ILR reading levels have been used to distinguish “limited translation performance” from “professional performance” (ILR 2015), on the assumption that specific text features signal greater text complexity and so difficulty in terms of translation as well as reading. However, the identified problem points and specific features differentiating ILR reading levels have not been linked to specific passage elements to refine passage selection for short-passage tests nor have they been evaluated for item difficulty in specified populations. Finally, the fact that assessment of short-passage tests has been criticized for focusing on micro-linguistic and grammatical elements rather than text-level features such as coherence (Angelelli 2009: 29) suggested a need for identifying passage features that signal higher-level text elements.

10.1 Study findings

This has been an exploratory study and so any discussion of study findings is suggestive only. With that caveat, the analysis of the translated passages has shown that:

- It is possible to link specific passage features—in this study, tense and implicit/explicit cohesion markers—to a desired trait, which in this work was the creation of a coherent target text.
- The elements signaling coherence could be objectively scored as acceptable or unacceptable and used to calculate facility values for the test population.

³⁰ The surveys done by Lafeber (2012a and b) are a laudable exception.

- The facility values could be used to create a profile of comparative passage difficulty and to quantitatively identify items of difficulty within a passage for a specific population.
- A feature characteristic of ILR reading level 3, the need to use inference to follow an argument, appears to have been a translation difficulty for the test population based on the facility values obtained for the cohesion markers and modified nouns.
- Features that earlier studies found to be translation difficulties, including modified nouns and official terms/culturally specific terms were also difficult for the Japanese-to-English language pair and in this test population. Passive verbs were not found to be a particular difficulty.
- Potential language-specific difficulties included incorrectly parsing strings of characters, recognizing head nouns, null anaphora, and translating the Japanese present narrative in a way that maintains temporal cohesion in English.

10.2 Study limitations

This research had several limitations. First, it was a *post facto* review of test passages that were not chosen to answer specific study questions. Second, while members of the study population all met the eligibility requirements to take the ATA certification exam, I lacked targeted data about translation training, amount and type of experience, and the translation direction (B language to A language, or A language to B language). Third, the test passages were produced under similar circumstances for the population, but test conditions may have caused the candidates to be more literal than usual in some of their choices and the fact that the exam was handwritten may have restricted re-ordering of information. Finally, the analysis was done on a small sample set: 36 translations for the general passages and 22 translations for the technical passage. For these reasons, this research should be considered an exploratory study only.

10.3 Potential applications

The primary application I envision for this research is use as a tool to refine passage selection for short-passage translation tests. Facility values have a long history of

application in test design and validation, but using them requires having discrete items that can be scored. That requirement has the advantage of forcing the test designer to identify what is being scored in light of both the discrete item (e.g. tense) and the trait to which it is linked (e.g. text cohesion, inference). For short-passage translation tests, it also involves analyzing potential passages in terms of exactly what scorable traits can be invoked by the text and what can then be said about future performance.

A secondary application of calculating facility values for discrete items is to monitor the match between test passages and test population and to confirm whether items that the designer thought to be difficult (or easy) are appropriate for the target group being tested. A mismatch could indicate that either the test material or the population descriptors might need to be refined. In addition, close reading of candidate translations should allow test designers to discover new and potentially useful translation problems.

In this study cohesion and inference were found to be difficulties in a population that self-identified as professional translators. This has implications for translator training and suggests that focusing on elements such as cohesion conventions in different languages, recognizing and reproducing links between sentences, and modification patterns could be fruitfully addressed in training programs.

10.4 Avenues for future research

More study is needed to uncover translation difficulties that are present across languages as well as language-specific difficulties. We also need additional information about the type of items that are difficult in different populations. For example, would a group of very experienced translators find tense and other forms of cohesion difficult to resolve? Or perhaps a better way to phrase the question would be: When does cohesion and inference stop being a translation difficulty? There is also work to be done on what text elements correspond to inference and cohesion in other language pairs.

In terms of translator training, areas of possible research are whether more explicit focus on cohesion conventions and linking elements within texts improve performance on ILR level 3 texts, as well as whether knowing about text types (i.e. being able to sort a set of texts into their respective ILR levels) would be of benefit to students.

What users of short-passage translation tests, both those being tested and those using the tests to make decisions want short-passage tests to actually signal—in Chan’s (2009) sense of separating the good translators from the bad—is whether a translator is good or might put future projects at risk. In light of that, an important area for future research is verifying whether well-designed short-passage tests have predictive power.

References

- Alderson, J. Charles. 2000. *Assessing Reading*. Cambridge: Cambridge University Press.
- Alderson, J. Charles. 2007. "The CEFR and the need for more research". *Modern Language Journal* 91(9): 659-663.
- Alves, Fabrio, Adriana Pagano, and Igor A.L. da Silva. 2010. "A new window on translators' cognitive activity: methodological issues in the combined use of eye tracking, key logging and retrospective protocols". In I.M. Mees, F. Avles, and S. Göpferich (eds) *Methodology, Technology and Innovation in Translation Process Research*. Frederiksberg: Samfundslitteratur. 267-291.
- American Translators Association (ATA). 2008. ATA Certification Program.
<http://www.atanet.org/certification/index.php/>. Entry point for information about the certification examination. Visited February 2008.
- American Translators Association (ATA). 2015. "About ATA Certification".
<http://www.atanet.org/certification/>. Accessed October 2015.
- Angelelli, Claudia V. 2009. "Using a rubric to assess translation ability". In C.V. Angelelli and H.E. Jacobsen (eds) *Testing and Assessment in Translation and Interpreting Studies*. Amsterdam and Philadelphia: John Benjamins.
- ASTM International. 2006. "Standard guide for quality assurance in translation". Designation: F 2575-06.
- ASTM International. 2011. "Standard practice for assessing language proficiency". Designation: F2889-11.
- Bachman, Lyle F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, Lyle F. 2004. *Statistical Analyses for Language Assessment*. Cambridge and New York: Cambridge University Press.
- Bachman, Lyle F., and Adrian S. Palmer. 1996. *Language Testing in Practice*. Oxford: Oxford University Press.
- Baker, Mona. 1992. *In Other Words: A Coursebook on Translation*. London and New York: Routledge.
- Buja, Andreas, Dianne Cook, Heike Hofmann, et al. 2009. "Statistical inference for exploratory data analysis and model diagnostics." In *Philosophical*

- Transactions, Series A, Mathematical, Physical, and Engineering Sciences*.
367(1906): 4361-4383.
- Campbell, Stuart. 1991. "Towards a model of translation competence". *Meta* 36(2-3):
329-343.
- Campbell, Stuart. 1999. "A cognitive approach to source text difficulty in translation".
Target 11(1): 33-63.
- Campbell, Stuart and Sandra Hale. 1999. "What Makes a Text Difficult to Translate?"
In Proceedings of the 1998 ALAA Congress. goo.gl/glnlgl/. Accessed
February 2008.
- Campbell, Stuart and Sandra Hale. 2003. "Translation and Interpreting Assessment in
the Context of Education Measurement". In G. Anderman and M. Rogers
(eds) *Translation Today: Trends and Perspectives*, Clevedon: Multilingual
Matters. 205-224.
- Campbell, Stuart, and Sandra Hale. 2003. "Translation and interpreting assessment in
the context of educational measurement". In G. Anderman and M. Rogers
(eds) *Translation Today: Trends and Perspectives*. Clevedon, Buffalo,
Toronto, Sydney: Multilingual Matters. 205-224.
- Cascallar, Eduardo C., Marijke I. Cascallar, Pardee Lowe, Jr., James R. Child. 1995.
"Development of new proficiency based skill level descriptors for translation:
theory and practice". In M. Milanovic and N. Saville (eds) *Performance
Testing, Cognition and Assessment: Selected Paper from the 15th Language
Research Testing Colloquim*. Cambridge: Cambridge University Press. 248-
258.
- Catford, J.C. 1965. *A Linguistic Theory of Translation*. London: Oxford University
Press.
- Chan, Andy Lung Jan. 2009. *Information Economics, the Translation Profession and
Translation Certification*. Doctoral thesis. Tarragona: Universitat Rovira i
Virgili. <http://goo.gl/yp26KL/>. Accessed October 2015.
- Chartered Institute of Linguists. "IoLET Level 7 Diploma in Translation (QCF)".
goo.gl/xl0xfN/. Accessed October 2015.
- Chatfield, C. 1988. "The initial examination of data". *Journal of the Royal Statistical
Society: Series A* 148(3): 214-253.
- Chatfield, C. 2006. "Initial data analysis" in *Encyclopedia of Statistical Sciences*.
Online: John Wiley & Sons. goo.gl/AXTXLq/. Accessed October 2015.

- Child, James R. 1998. "Language skill levels, textual modes, and the rating process".
Foreign Language Annals 31(3): 381-391.
- Child, James R., Ray T. Clifford, Pardee Lowe, Jr. 1993. "Proficiency and Performance
in Language Testing". *Applied Language Learning* 4(1/2): 19-54.
- Clifford, Ray, Neil Granoien, Douglas Jones, et al. 2004. "The effect of text difficulty
on machine translation performance: a pilot study with ILR-rated texts in
Spanish, Farsi, Arabic, Russian and Korean" in *Proceedings of the Language
Resources and Evaluation Conference (LREC) 2004, Lisbon, Portugal*. No
page numbers given.
- Cook, John, Hilary Dixon. 2005. *A Review of NAATI Administrative Processes Related
to Testing Including Quality Control Processes: Final Report*. Access MQ.
<http://goo.gl/JvtwKx/>. Accessed November 2015.
- Donahue, Ray T. 1998. *Japanese Culture and Communication: Critical Cultural
Analysis*. Lanham, MD: University Press of America.
- Dörner, Dietrich. 1996. *The Logic of Failure: Recognizing and Avoiding Error in
Complex Situations*. New York: Basic Books. Translation of *Logik des
Misslingens*. 1989. Hamburg: Rowholt Verlag. Rita and Robert Kimber,
translators.
- Douglas, Dan. 2010. *Understanding Language Testing*. London: Hodder Education.
- Dragsted. Barbara, and Inge Gorm Hansen. 2008. "Comprehension and production in
translation: a pilot study on segmentation and the coordination of reading and
writing process". In S. Göpferich, A.L. Jakobsen, I.M. Mees (eds) *Looking at
Eye: Eye-Tracking Studies of Reading and Translation Processing*.
Frederiksberg: Samfundslitteratur. 9-29.
- Dunning, David, Kerri Johnson, Joyce Ehrlinger, and Justin Kruger. 2003. "Why people
fail to recognize their own incompetence". *Current Directions on
Psychological Science* 12(3): 83-87.
- Englund Dimitrova, Birgitta. 2005. *Expertise and Explicitation in the Translation
Process*. Amsterdam and Philadelphia: John Benjamins.
- Ehrlinger, Joyce, Kerri Johnson, Matthew Banner, et al. 2008. "Why the unskilled are
unaware: further explorations of (absent) self-insight among the incompetent".
Organizational Behavior and Human Decision Processes 105(1): 98-121.
- Eyckmans, June, Phillipe Anckaert, and Winibert Segers. 2009. "The perks of norm-
references translation evaluation". In C.V. Angelelli and H.E. Jacobsen (eds)

- Testing and Assessment in Translation and Interpreting Studies*. Amsterdam and Philadelphia: John Benjamins.
- Fulcher, Glenn. 2010. *Practical Language Testing*. London: Hodder Education.
- Fulcher, Glenn, and Fred Davidson. 2007. *Language Testing and Assessment: An Advanced Resource Book*. London and New York: Routledge.
- Gile, Daniel. 1995. *Basic Concepts and Models for Interpreter and Translation Training*. Amsterdam and Philadelphia: John Benjamins.
- Green, Anthony. 2014. *Exploring Language Assessment and Testing: Language in Action*. London and New York: Routledge.
- Hale, Sandra, Ignacio Garcia, Jim Hlavac, et al. 2012. "Improvements to NAATI testing", report for the National Accreditation Authority for Translators and Interpreters (NAATI). <https://goo.gl/0n9G2P/>. Accessed October 2015.
- Hale, Sandra and Stuart Campbell. 2002. "The interaction between text difficulty and translation accuracy". *Babel* 48(1): 14-33.
- Halliday, M.A.K. and R. Hasan. 1976. *Cohesion in English*. London and New York: Longman.
- Howard, Diane. 2011. "Machine Translation as an Aid in Identifying Translation Difficulties for Texts Used in Introductory Translation Courses". In M. Forstner and H. Lee-Jahnke (eds) *CIUTI-Forum new needs, translators & programs: on the translational tasks of the United Nations*. Bern and New York: Peter Lang: 157-177.
- Hughes, Arthur. 2003. *Testing for Language Teachers, Second Edition*. Cambridge: Cambridge University Press.
- Inoue, Atsu, and Janet Dean Fodor. 1995. "Information-paced parsing of Japanese" in R. Mazuka and N. Nagai (eds) *Japanese Sentence Processing*. Hillsdale, NJ and Hove, UK: Lawrence Erlbaum Associates. 9-63.
- Interagency Language Roundtable (ILR). 2015. "Interagency Language Roundtable language skill level descriptors—reading". Available at: <http://goo.gl/YDg0RK>. Accessed October 2015.
- Interagency Language Roundtable (ILR). 2015. "ILR skill level descriptions for translation performance." <http://goo.gl/k3xI5J/>. Accessed October 2015.
- Jääskeläinen, Riitta. 1998. "Think-aloud protocols". In M. Baker (ed) *Routledge Encyclopedia of Translation Studies*. London: Routledge. 99-120.

- Jakobsen, Arnt Lykke. 2003. "Effects of think aloud on translation speed, revision, and segmentation". In F. Alves (ed) *Triangulating Translation*. Amsterdam and Philadelphia: John Benjamins. 69-95.
- Jensen, Kristian T.H. 2009. "Indicators of text complexity". In S. Göpferich, A.L. Jakobsen, I.M. Mees (eds) *Behind the Mind: Methods, Models and Results in Translation Process Research*. Frederiksberg: Samfundslitteratur. 61-80.
- Jones, Douglas, Martha Herzog, Hussny Ibrahim, et al. 2007. "ILR-based MT comprehension test with multi-level questions". *Proceedings of NAACL HLT 2007*. Rochester, NY: Association for Computational Linguistics. 77-80.
- Jones, Douglas, Wade Shen, Neil Granoien, et al. 2005. "Measuring translation quality by testing English speakers with a new Defense language proficiency test for Arabic". In *Proceedings of the 2005 International Conference on Intelligence Analysis*, May 2-6, 2005, McLean, VA: MITRE. No page numbers given.
- Just, Marcel Adam, and Patricia A. Carpenter. 1980. "A theory of reading" from eye fixations to comprehension". *Psychological Review* 87(4): 329-354.
- Kaiser, Stefan, Ysuko Ichikawa, Noriko Kobayashi, and Hilofumi Yamamoto. 2002. *Japanese: A Comprehensive Grammar*. London and New York: Routledge.
- Kashima, Yoshihisa, and Emiko S. Kashima. 1999. "Culture, connectionism, and the self". In J. Adamopoulos and Y. Kashima (eds) *Social Psychology and Cultural Context*. Thousand Oaks, CA: Sage Publications.
- Kelly, Dorothy. 2005. *A Handbook for Translation Trainers*. Manchester and Northampton: St. Jerome.
- Kiraly, Donald C. 2000. *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. Manchester: St. Jerome.
- Koby, Geoffrey S., Gertrud G. Champe. 2013. "Welcome to the real world: professional-level translator certification". *Translation & Interpreting* 5(1): 156-173.
- Koby, Geoffrey S., and Alan K. Melby. 2013. "Certification and job task analysis (JTA): establishing validity of translator certification examinations". *Translation & Interpreting* 5(1): 174-210.
- Koda, Keiko. 2004. *Insights into Second Language Reading: A Cross-Linguistic Approach*. Cambridge: Cambridge University Press.
- Krings, Hans P. 2001. *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Process*. Kent, OH, and London: The Kent State University Press.

- Kruger, Justin, and David Dunning. 1999. "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments". *Journal of Personality and Social Psychology* 77(6): 1121-1134.
- Kusssmaul, Paul. 1995. *Training the Translator*. Amsterdam: John Benjamins.
- Kusssmaul, Paul, and Sonja Tirkkonen-Condit. 1995. "Think-aloud protocol analysis in translation studies". *TTR : traduction, terminologie, rédaction* 8(1): 177-199.
- Kutner, Mark, Elizabeth Greenberg, Ying Jin, et al. 2007. *Literacy in Everyday Life: Results From the 2003 National Assessment of Adult Literacy (NAAL)*. Washington, DC: National Center for Education Statistics. <http://goo.gl/yRCNy9/>. Accessed October 2015.
- Lee, Jae-ho, and Yoichiro Hasabe. 2015. "Automatic assessment of Japanese text readability based on a level corpora". Publication information unavailable. <http://goo.gl/pl3W7C/>. Accessed October 2015.
- Lafeber, Anne. 2012a. "Translation skills and knowledge—preliminary findings of a survey of translators and revisers working at Inter-governmental Organizations". *Meta* 57(1): 108-131.
- Lafeber, Anne. 2012b. *Translation at Inter-Governmental Organizations: The Set of Skills and Knowledge Required and the Implications for Recruitment Testing*. Doctoral thesis. Tarragona: Universitat Rovira i Virgili. <http://goo.gl/BStWfR/>. Assessed October 2015.
- Lin, Dekang. 1996. "On the structural complexity of natural language sentences". In (no editors given) *Proceedings of the 16th International Conference on Computational Linguistics*. . Stroudsburg, PA: Association for Computational Linguistics.
- Long, Michael H., Kira Gor, and Scott Jackson. 2012. "Proof of concept with ILR 2-3 in Russian". *Studies in Second Language Acquisition* 34: 99-126.
- Lowe, Pardee, Jr. 1986. "The unassimilated history". In P. Lowe, Jr. and C.W. Stansfield (eds) *Second Language Proficiency Assessment: Current Issues*. Englewood Cliffs, NJ: Prentice-Hall, Inc. 11-51. <http://eric.ed.gov/?id=ED296612/>. Accessed October 2015.
- Makino, Seiichi, and Michio Tsutsui. 1995. *A Dictionary of Intermediate Japanese Grammar*. Tokyo: The Japan Times.

- Marais, Kobus. 2013. "Constructive alignment in translator education: reconsidering assessment for both industry and academy". *Translation & Interpreting* 5(1): 13-24.
- Martin, Samuel E. 1975. *A Reference Grammar of Japanese*. Rutland, VT and Tokyo: Charles E. Tuttle.
- Martyniuk, Waldemar. 2010. "From the editor of this volume". In W. Martyniuk (ed) *Aligning Tests with the CEFR*. Cambridge: Cambridge University Press. xxi-xxiii.
- McAlester, Gerard. 1999. "The source text in translation assessment" in G. Anderman and M. Rogers (eds) *Word, Text, Translation*. Clevedon, Buffalo, Toronto, Sydney: Multilingual Matters. 169-175.
- McClure, William. 2000. *Using Japanese: A Guide to Contemporary Usage*. Cambridge and New York: Cambridge University Press.
- McNamara, Tim. 2000. *Language Testing*. Oxford and New York: Oxford University Press.
- Mishra, Abhijit, Pushpak Bhattacharyya, and Michael Carl. 2013. "Automatically predicting sentence translation difficulty". In (no editor given) *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, PA: Association for Computational Linguistics.
- National Accreditation Authority for Translators and Interpreters (NAATI). 2015. *Accreditation by testing: information booklet*. <https://goo.gl/mz2ZQd/>. Assessed October 2015.
- Nord, Christiane. 1997. *Translating as a Purposeful Activity*. Manchester, UK and Northampton, MA.
- Nord, Christiane. 2005. *Text Analysis in Translation. Theory, Methodology and Didactic Application of a Model for Translation-Oriented Text Analysis*, 2nd ed. Amsterdam and New York: Rodopi.
- PACTE Group. 2003. "Building a translation competence model." In F. Alves (ed) *Triangulating Translation*. Amsterdam and Philadelphia: John Benjamins. 41-66.
- PACTE Group. 2005. "Investigating translation competence: conceptual and methodological issues" *META* 50(2): 609-619.
- PACTE Group. 2008. "First Results of a Translation Competence Experiment: 'Knowledge of Translation' and 'Efficacy of the Translation Process'". In J.

- Kearns (ed) *Translation and Interpreter Training. Issues, Methods and Debates*. London: Continuum International Publishing Group. 104-126.
- PACTE Group. 2011. "Results of the validation of the PACTE translation competence model: translation problems and translation competence". In C. Alvstad, D. Hild, E. Tiselius (eds) *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*. Amsterdam and Philadelphia: John Benjamins. 317-342.
- Perfetti, Charles A., Nicole Landi and Jane Oakhill. 2005. "The acquisition of reading comprehension skill". In M.J. Snowling and C. Hulme (eds) *The Science of Reading: A Handbook*. Malden, MA, and Oxford: Blackwell Publishing. 227-247.
- Pichl, Lukáš, and Joe Narita. 2007. "Readability factors of Japanese text classification" In *Proceedings of the 5th International Conference on Databases in Networked Information Systems*. Heidelberg and New York: Springer-Verlag. 132-133.
- Quine, Willard V.O. 1960. *Word and Object*. Cambridge, MA: The MIT Press. Kindle edition.
- Sato, Satoshi. 2014. "Text readability and word distribution in Japanese". In N. Calzolari, K. Choukri, T. Declerck et al. (eds) *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, Reykjavik, Iceland. Luxembourg: European Language Resources Association. 2811-2815.
- Satoshi Sato, Suguru Matsuyoshi and Yohsuke Kondoh. 2008. "Automatic assessment of Japanese text readability based on a textbook corpus". In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. European Language Resources Association (ELRA).
- Schäffner, Christina. 1998. "Qualification for professional translators. Translation in language teaching versus teaching translation". In *Translation and Language Teaching. Language Teaching and Translation*, Kristen Malmkjaer, ed. Manchester: St. Jerome. 117-133.
- Secară, Alina. 2005. "Translation evaluation—a state of the art survey". In *eCoLoRe/MeLLANGE Workshop*. Leeds, UK: Published Proceedings, 39-44.
- Sharmin, Selina, Oleg Spakov, Kari-Jouko Räihä, and Arnt Lykke Jakobsen. 2008. "Where on the screen do translation students look while translating, and for

- how long?” In S. Göpferich, A.L. Jakobsen, I.M. Mees (eds) *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing*. Copenhagen: Samfundslitteratur. 31-51.
- Shohamy, Elana. 1996. “Competence and performance in language testing”. In G. Brown, K. Malmkjaer, and J. Williams (eds) *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press. 138-151.
- Skehan, Peter. 1984. “Issues in the testing of English for specific purposes”. *Language Testing* 1(2): 202-220.
- Soga, Matsuo. 1983. *Tense and Aspect in Modern Colloquial Japanese*. Vancouver: University of British Columbia Press.
- Spolsky, Bernard. 1995. *Measured Words: The Development of Language Testing*. Oxford: Oxford University Press.
- Sun, Sanjun. 2012. *Measuring Difficulty in English-Chinese Translation: Towards a General Model of Translation Difficulty*. Doctoral thesis. Kent: Kent State University. <https://goo.gl/WVPvf3/>. Assessed October 2015.
- Sun, Sanjun, and Gregory M. Shreve. 2014. “Measuring translation difficult: an empirical study”. *Target* 26(1): 98-127.
- Tateishi, Yuka, Yoshihiko Ono and Hisao Yamada. 1988. “A computer readability formula of Japanese texts for machine scoring”. In *Proceeding of COLING 1988*. 649-654. <https://aclweb.org/anthology/C/C88/C88-2135.pdf/>. Accessed October 2015.
- Toury, Gideon. 2006. “Conducting research on a ‘Wish-to-Understand’ basis”. In J. Ferreira Duarte, A. Assis Rosa and T. Seruya (eds), *Translation Studies at the Interface of Disciplines* (pp. 55-66). Amsterdam and Philadelphia: John Benjamins.
- Tsujimura, Natsuko. 1996. *An Introduction to Japanese Linguistics*. Malden, MA, and Oxford: Blackwell.
- Vinay, Jean-Paul, and Jean Darbelnet. 1995. *Comparative stylistics of French and English: a methodology for translation*. Amsterdam and Philadelphia: John Benjamins.
- Vinay, Jean-Paul, and Jean Darbelnet. 2000. “A methodology for translation”. In L. Venuti (ed) *The Translation Studies Reader*. London and New York: Routledge. 84-93.

- Watanabe, Toshiro, Edmund R. Skrzypczak, Paul Snowden (eds). 2003. *Kenkyusha's New Japanese-English Dictionary, Fifth Edition*. Tokyo: Kenkyusha.
- Weir, Cyril. 2005. *Language Testing and Validation: An Evidence-Based Approach*. Basingstoke, Hampshire and New York: Palgrave MacMillan.
- Williams, Malcolm. 2004. *Translation Quality Assessment: An Argumentation-centred Approach*. Ottawa: University of Ottawa Press.
- Yoshida, Etsuko. 2011. *Referring Expressions in English and Japanese*. Amsterdam and Philadelphia: John Benjamins.