Title	TL2
Author(s)	, ; ,
Citation	. , 67(2): 27-39
Issue Date	2017-02
URL	http://s-ir.sap.hokkyodai.ac.jp/dspace/handle/123456789/8173
Rights	

Hokkaido University of Education

Designing Parallel Spoken Corpora of the English Classroom Through Embedding TL2 Elements

KATAGIRI Noriaki and OHASHI Yukiko*

Department of English Communication Studies, Asahikawa Campus, Hokkaido University of Education

*Yamazaki Gakuen University

TL2要素の埋め込みによる教室英語のパラレル発話コーパスの構築

片桐　徳昭・大橋由紀子*

北海道教育大学旭川校英語コミュニケーション学研究室

*ヤマザキ学園大学

ABSTRACT

This short paper describes one of the systems to design parallel spoken corpora of English classroom discourse. We propose one scheme of parallel classroom spoken corpora with extensible markup language（XML）annotation and extensible stylesheet language transformations（XSLT）in order to assist non-native elementary school homeroom teachers （HRTs）who are in need of improving their English language skills in foreign language activity lessons. Parallel spoken corpora include L1 as well as the target language（L2） being taught regardless of the language use ratios. By translating L1 transcripts into L2, which we define as translated L2（TL2）, and by embedding TL2 in the classroom corpora, we will be able to enhance usefulness of the conventional spoken classroom corpora. L1 and TL2 transcripts from the parallel classroom corpora will provide evidence for non-native English teachers to reflect their L1 and L2 speech in the classroom. Our parallel spoken corpus architecture and XSLT stylesheets can retrieve（Japanese）L1 and its translated English（L2）data. The findings of this study can contribute to improvement of HRTs' L2 speech in the classroom as well as providing classroom discourse researchers with opportunities to deepen understanding classroom interactions.

1. Introduction

1.1　Background

The new English education reform plan by the Ministry of Education, Sports, Science, Culture and Technology, Japan（MEXT）envisions shifting English teaching style using the target language, i.e., English（L2）instead of resorting to Japanese（L1）to teach English as a foreign language（EFL）in Japanese schools even in the lower secondary educational level in the new course of study in 2010（MEXT, 2013）. According to MEXT's reform plan, elementary schools will begin teaching English as a subject in 2020 as well. Although Japanese elementary schools started to introduce teaching English to fifth and sixth graders under the name of "foreign language activity" in 2011, homeroom teachers（HRTs）that must conduct such lessons do not feel as comfortable as they do when they teach the conventional subjects such as arithmetic and arts and crafts. The progress report concerning English education at the primary level revealed several daunting facts concerning elementary homeroom teachers that might impede the ambitions of MEXT's English educational reform plan:（1）67.3% of elementary school homeroom teachers confessed that they are not good at English; as few as 34.6% of them teach English confidently（MEXT, 2014）. Judging from these facts, it is natural that such teachers use L1 in L2 lessons, and therefore, it is of great interest to pursue ways to assist them to have a better command of L2 in the near future to come.

1.2　Review of the literature

MEXT's progress report in 2014 disclosed that HRTs need to improve their L2 skills to achieve the goals of the reform plan. The language use of non-native instructors of English is still low. Katagiri（2016）conducted a corpus-based study by collecting five middle school non-native English teachers in Japan, and showed that their L2 use is a little over 63%, which means approximately 37% of non-native classroom utterances are still in L1. Another study by Katagiri and Ohashi（2016）revealed that six non-native preservice instructors, all of whom were in the third year in a university of education in Japan, used L2 as few as 31.9% of their speech turns. As for the primary education level, Ohashi and Katagiri（2016）built corpora of elementary school foreign language activity, namely English language activity, made up of four elementary school HRTs. They found that HRTs use of L2 tallied an average of 232 tokens in contrast with L1 usage of 1,337 tokens. This may well indicate how strongly elementary school homeroom teachers depend on L1 in teaching English language.

Although middle school non-native instructors use L1 due to lack of vocabulary and for "pedagogical considerations"（Katagiri, 2015, p. 100）, it would be of pedagogical use to try to figure out whether this would really be the case, and find feasible options, i.e., to propose alternative utterances in L2 before the full enactment of the new course of study in 2020.

1.3　Parallel spoken corpora

This section proposes the notion of parallel corpora in order to introduce the essential component of this study, i.e., L1 translated into L2.

Parallel corpora can be defined as "bilingual corpora"（Ishikawa, 2012, p. 42）. According to Ishikawa（2012）, bilingual corpora contain transcripts of two languages, and they serve as research materials to show how L1 is translated

into L2. He mentions the issue of "parallelism," which questions whether translated L1 is of equal message value as well as grammatically correct linguistic value. This issue is inevitable because TL2 is produced by human translators. We need to be aware that TL2 might not represent accepted and communicatively compatible L2 in terms of its use and linguistic correctness.

The National Institute of Information and Communications Technology (NICT) created a large bilingual corpus comprised of half a million Japanese sentences and translated English versions taken from articles concerning topics in Kyoto (see Appendix A for a transcription sample). This large bilingual corpus was developed in order to support research areas such as "machine translation" and "information extraction" (NICT, 2012). However, when it comes to English classroom spoken corpora, we hardly have any examples, and still fewer examples of parallel corpora that contain Japanese transcriptions and their translated English versions. This fact gives us a good reason for seeking to build a bilingual parallel classroom spoken corpora.

1.4 Research objectives

Considering the background and the literature, it is of great interest and thus, of significance to explore what kind of assistance we can give to elementary school HRTs. Hence we, attempt to design a parallel spoken corpus architecture and examine whether it is possible to build bilingual classroom spoken corpora in a sense that it would support non-native elementary school HRTs. In this paper, we will use the notion translated L2 (TL2) to represent L2 translated from L1 speech used by Katagiri (2016). We now pose the following two research objectives:

1. Inserting TL2 in a corpus scheme which has already been made.
2. Extracting TL2 elements from parallel corpora to present them in parallel with L1 from which TL2 is translated.

We will discuss the procedure to find out whether we will be able to achieve our two research objectives in the following sections.

2.   Materials and Methods

2.1   Spoken corpora

Ohashi and Katagiri (2016) compiled classroom spoken corpora of elementary school foreign language activity (the *OK corpus* hereafter). The OK corpus contains classroom speech of four elementary school English language classes, where two Japanese HRTs, and one English-native assistant language teacher (ALT) taught four English language activity classes, two fifth grader classes and two sixth grader classes. The OK corpus is annotated with speaker tags and language tags (L1, L2, and Mix) in an extensible markup language (XML) format with transcriptions of 9,192 L1 tokens and 1,967 L2 tokens.

Figure 1 shows the base design of the OK corpus. The transcriptions with XML annotation serve as extraction purposes using extensible stylesheet language transformations (XSLT) to analyze the spoken data. Looking down from the root element ("Class" on top), the root entails four child nodes that designate speaker turns in the corpus; ALT; HRT; student (ST); students (STS). These child nodes contain their own child nodes that represent speakers' language use: English (L2); Japanese (L1); Mix (L1+ L2).
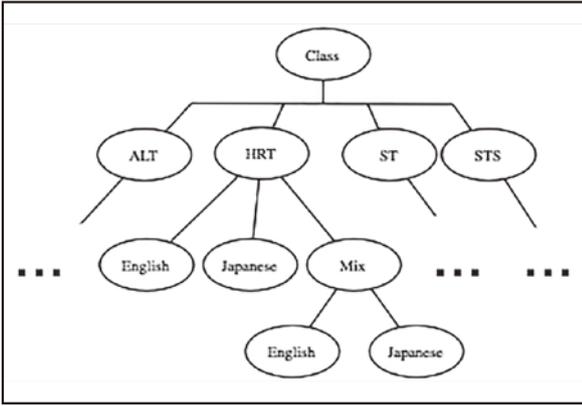
Figure 1. Tree diagram of the OK corpus.

The tree diagram in Figure 1 is converted into XML instance in Figure 2. It represents the core elements of the OK corpus. Each element displays: <class>, one lesson; <alt>, assistant language teacher; <hrt>, homeroom teacher; <eng>, English utterance; <j>, Japanese utterance; <mix>, utterances composed of English utterances and Japanese utterances.

```
<class>
    <alt> <eng></eng> </alt>
    <hrt> <j></j> </hrt>
    <alt><mix><j></j><eng> </eng></mix></alt>
    <hrt><mix><eng> </eng><j></j></mix></hrt>
    ………………..
</class>
```

Figure 2. XML representation of the OK corpus speaker turns and language use.

2.2 Architecture of the parallel corpus

We designed a parallel corpus structure using the XML annotation in the following three steps:

Step 1   Embed <TL2></TL2> tags after each <j></j> tag. Figures 3 and 4 show a tree diagram of the embedded TL2 tag and its XML representation respectively.
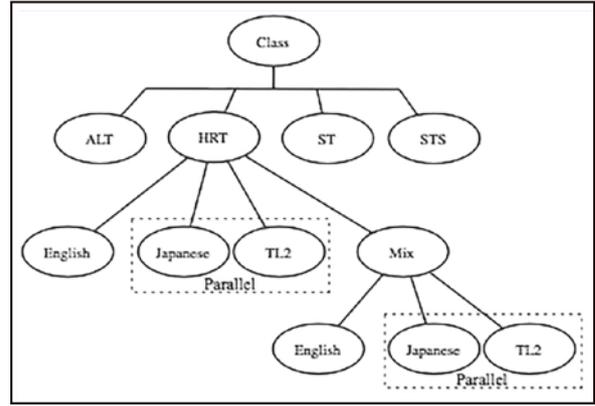

Figure 3. Tree diagram of TL2 element insertion.

```
    <hrt>
        <j></j><TL2><TL2>
    </hrt>
………………..
    <hrt>
        <mix>
            <j></j><TL2><TL2>   <eng></eng>
        </mix>
    </hrt>
```

Figure 4. <TL2></TL2> tag insertion sample.

Step 2   Search the L1 utterances in the <j></j> elements using path locations used in XML Path Language (XPath).[1] Figure 5 shows four paths that locate Japanese utterances of ALT and HRT: (1) and (2) lead to the exclusive L1 utterances, and (3) and (4) to the L1 utterances in Mix utterances.

```
(1) ALT: /class/alt/j
(2) HRT:/class/hrt/j
(3) ALT:/class/alt/mix/j
(4) HRT:/class/hrt/mix/j
```

Figure 5. XPath that locates Japanese utterances of ALT and HRT in the OK corpus.

Step 3   Translate L1 utterances in <j></j> elements into L2 within the <TL2></TL2> tags. In this step, we translated L1

utterances in <j></j> elements into the L2 called TL2, and inserted the TL2 as content in the <TL2></TL2> elements.

We underwent the above-mentioned three steps to compile parallel corpora. The next section will describe how we extract the elements that we seek from the parallel corpora.

2.3　Data extraction

After compiling parallel spoken corpora by going through the three steps in the previous section, we are now geared to have access to the parallel corpus data we intend to procure for our research and/or pedagogical purposes. The following sections will describe two sample extractions from the parallel corpora followed by motivations for linguistic research as well as pursuit for pedagogical implications.

2.3.1　Plain TL2 text output

One of the interests for researchers is to find out what kind of L1 could be translated into TL2 because researchers are curious to know whether it would be linguistically possible for elementary school HRTs to use TL2 as they use L2. Katagiri (2016) argued that it would be possible for middle school teachers to conduct English lessons in TL2 in terms of similarity of its vocabulary levels. We will use the location path in our XSLT stylesheets (Figure 6) to extract TL2 plain texts. Just as Figure 5 explained, each location path respectively leads us to TL2 of ALT and HRTs regardless of exclusive L1 or L1 in Mix language use.

```
(1)  /class/alt/TL2
(2)  /class/hrt/TL2
(3)  /class/alt/mix/TL2
(4)  /class/hrt/mix/TL2
```

Figure 6. XSLT location paths to reach TL2 elements in the parallel corpus.

2.3.2　Parallel representation of L1 and TL2

The other extraction example of utilizing TL2 is to show the extraction result paired up with its corresponding L1. Parallel listing of L1 and TL2 will enable us to compare them, and might give us educational insight for improving classroom speech of elementary school English instructors, namely HRTs. Figure 7 depicts XSLT location paths that can list parallel display of L1 and its corresponding TL2 (See Appendix D for full XSLT stylesheet sample).

```
<xsl:template match="/">

    <xsl:text> [LF]</xsl:text>
  <xsl:for-each select="body/K1/alt/mix">
    <xsl:copy-of select="j" />
      <xsl:text>[TAB]</xsl:text><xsl:copy-of
select="TL2"/>

    <xsl:text> [LF]</xsl:text>
    <xsl:text> [LF]</xsl:text>
  </xsl:for-each>
</xsl:template>
```

Figure 7. Excerpt from XSLT style sheet sample. *[LF]* and *[TAB]* respectively represent line-feeding and tab-insertion.

3.　Results and Analyses

We will look at the results through transforming the parallel corpus using our XSLT transformation style sheets. The following sections will show the results of embedding TL2 tags in the corpus (Section 3.1), and representing L1 and TL2 elements from the corpus (Section 3.2).

3.1　TL2 embedding in the corpus

Figures 8 and 9 respectively display excerpts of mix utterances of ALT and HRT in the parallel corpus. Together with L2 elements shown as <eng></eng>, the TL2 elements represented by <TL2></TL2> are embedded next to the L1 elements shown as

These figures display that <TL2></TL2> tags are embedded in the corpus.

```
<mix>
    <eng>Okay, </eng>
    <j>じゃあもう一回曜日から。</j>
    <TL2>Let's start over from days of the week.</TL2>
</mix>
<mix>
    <j>ううん、</j>
    <TL2>No, it isn't.</TL2>
    <eng>not thirty-one.</eng>
</mix>
```

Figure 8. Mix utterance excerpt of ALT in the parallel corpus.

```
<mix>
    <eng>Not thirty-one</eng>
    <j>っていうゲームをします。</j>
    <TL2>We are going to play a game called like
that.</TL2>
</mix>
```

Figure 9. Mix utterance excerpt of HRT in the parallel corpus.

Figure 10 displays an excerpt from the TL2 of HRT in the corpus (See Appendix E for a complete set of the TL2 extraction results). The result shows that the XSLT style sheet extracted the TL2 speech of the HRT from the parallel corpus.

```
<?xml version="1.0" encoding="UTF-8"?>
<TL2>We are going to play a game called like
that.</TL2>
<TL2>How many of you understood most of
them?</TL2>
<TL2>This is why the game is called...</TL2>
<TL2>All right, then.</TL2>
```

Figure 10. Excerpt of TL2 extraction from L1 utterances in Mix elements of HRT in the parallel corpus.

3.2 Parallel representation of L1 and TL2

This section will display the results of horizontal output of L1 and its corresponding TL2 from the corpus by using the XSLT style sheet (Figure 7). We aligned L1 and TL2 elements which are sibling nodes of the same parent node, i.e., either ALT nodes or HRT nodes. We excluded <eng></eng> elements, which are also the child nodes of the ALT or HRT nodes. The ALT nodes and the HRT nodes are parent nodes of <eng></eng> elements (Figure 3)

Figure 11 shows an excerpt of the horizontal display of the L1 and TL2 extraction. Each <j></j> element is horizontally aligned with its translated English, i.e., TL2. Therefore, the intended results were obtained by using our XSLT style sheets, one example of which is shown in Appendix D.

```
<j>サタデーじゃない</j> <TL2>It is not Saturday.</TL2>

<j>金曜日？</j>          <TL2>Friday?</TL2>

<j>んー、</j>           <TL2>Well, </TL2>
```

Figure 11. Excerpt of the horizontal display of L1 and TL2 extraction

4. Discussion and conclusion

4.1 Design of parallel corpora

We were able to design one example of parallel corpora using the extensible mark-up language (XML) format. As the name implies, our schema can be "extensible" because firstly additional annotation such as adding TL2 elements in the transcription was proved to be feasible without changing the entire corpus architecture. Our design of embedding additional elements in the corpus is one possible schema that is applicable to any corpus architecture that uses the XML tree structure as shown in Figure 1. Therefore,

this will expand the possibility of creating parallel corpora using one that already exists. Accumulating parallel corpus data, especially TL2 data will hopefully give evidence for HRTs and teacher trainers as well as researchers to reflect L1 speech in the EFL classroom.

4.2 Extracting corpus data

Our XSLT style sheets (Appendices B, C, and D) which used XPath locations (Figure 6) extracted the intended transcripts from the parallel corpus. The XSLT stylesheets firstly enabled us to reach TL2 data in the parallel corpus (Appendix E). The XSLT also enabled us to output L1 and TL2 horizontally (Figure 11). Therefore, building parallel corpora with XML annotation can benefit us in extracting the corpus data in manners that we intend through XSL transformations.

4.3 Implications to pedagogy

Now that we can utilize TL2 transcripts, we can propose the following possibilities that might contribute to non-native English teacher education, especially HRTs in this study, and English classroom discourse research. This section briefly describes these two perspectives.

Firstly, TL2 data can basically be used for teacher training programs to improve non-native elementary HRTs because they can reflect their classroom discourse and interactions to examine what teacher talk should be uttered in L1, L2 or a mixture of both. If HRTs are willing to improve their L2 speech in the classroom, learning to code-switch from L1 to TL2 might give them a good opportunity to achieve their ambitions.

Secondly, parallel English classroom corpora will shed new light on classroom discourse research. We are already aware of the dominant use of L1 in English activities in elementary schools (Ohashi & Katagiri, 2016). L1 and TL2 data in the parallel corpora will provide evidence to show cases where ALTs and HRTs use L1, and might provide researchers with opportunities to qualitatively analyze reasons behind such L1 uses. We believe that more researchers will take advantage of parallel corpora, and findings from parallel corpus-based research may propose practical materials for teacher education.

4.4 Limitations

In our attempts to design the parallel spoken corpus schema and extracting data from the corpora, we noticed there were at least three limitations. Our awareness is attributed to the very basic nature of our study because designing parallel spoken corpora was inspired by issues concerning non-native English teacher training programs in primary education level. We realized the need to develop schema for creating parallel spoken corpora.

The first limitation to our study concerns the quality of the TL2. Although the authors translated L1 into TL2 and proofread the TL2, TL2 still needs proofreading by native speakers of English in the first place, and then by HRTs because we need to examine whether TL2 will function in the real classroom situations. This task can be time-consuming and thus, was out of scope of the present study.

The second limitation is a question whether the need for making the parallel classroom corpora for elementary school really exists. The OK corpus we used for our study contained utterances of ALTs. This implies that HRTs co-taught the English lessons with probably full support from ALTs in presenting English to pupils as well as communicating in English. If

we assume such co-teaching style is the norm in elementary school English lessons in the next course of study that is be implemented in 2020, HRTs may not have much need for improving their L2 speech.

The final limitation might be issues of the computer operating system. The authors test-ran the XSL transformations on Mac OS X[2] and Windows 7[3] The transformations were successful on Mac OS X, however, we were unable to execute the XSLT stylesheets, which did not yield the intended outcomes. We need to see whether the XSLT would work on other versions of Windows and hopefully use a different XML parser.[4]

4.5 Future plans

Based upon the fact that we have created one example of the basic structure of the parallel corpus schema and their applications through our XSLT style sheets, we are now planning to conduct the following research to enhance the usability and applicability of the parallel spoken corpora.

The first thing we need to do is to collect more classroom spoken data. While we accumulate the spoken data, we need to proofread TL2 in the way we discussed in the previous section. In addition, we need to obtain consent of the prospective participants for open or limited access to the data since such classroom spoken data are hard to come by in the first place, and thus, pedagogically highly valuable.

The second plan is to develop more XSLT style sheets so that they can live up to the need for the in-service HRTs. We may need to survey the linguistic needs of the HRTs so they will achieve their pedagogical goals.

Finally, but not limited to, we are planning to utilize the extraction results for teacher training.

We need to examine whether findings from the parallel "spoken corpus-based" study will assist non-native elementary HRTs to develop their classroom English. We hope that our attempts will attract more researchers and non-native HRTs, and eventually will contribute to the betterment for primary English education in the years to come.

Notes

1. XPath utilizes the path location for each element. Usually the path starts with the root element (using an absolute location) until it reaches the targeted element. Each "/" represents a delimiter to separate a parent node (the node in the upper directory) from its child node (the node in the lower directory) in an XML hierarchical structure.
2. Mac OS X is an operating system that runs on Macintosh computers. It used be called Mac OS X until OS X "Mountain Lion" appeared in 2012. This study used OS X 10.11.6. (https://en.wikipedia.org/ wiki/OS_X)
3. Wondows 7 is a personal computer operating system that was released in 2009 by Microsoft. Windows 7 is an upgraded system from Windows Vista (https://en.wikipedia. org/wiki/ Windows_7). The current version (as of September, 2016) is Windows 10 (https://en.wikipedia.org/wiki/ Windows_10)
4. Editix is an XML editor downloadable for thirty-day free use at http://www.editix. com/download.html, and full license is provided upon purchase.

Acknowledgments

REFERENCES

Ishikawa, S. (2012). *Beisshiku Kopasu Gengogaku* [Basci Corpus Linguistics]. Tokyo: Hitsuji Publishing.

Katagiri, N. (2015). Analyzing Classroom English of Non-native Instructed EFL Classrooms in Japanese Middle Schools. Journal of Hokkaido University of Education (Humanities and Social Sciences) *66*(1). 91-109.

Katagiri, N. (2016). Feasibility of Using Translated Middle School Non-Native Instructor Utterances in L2 Lessons. Annual Review of English Language Education in Japan *27*. 109-124.

Katagiri, N. & Ohashi, Y. (2016). Quantitative Analyses of Non-Native Preservice Teacher Verbal Interactions at Japanese Middle Schools. Oral presentation at the 42nd Annual Convention of Japan Society of English Language Education. Saitama, Japan.

Ministry of Education, Culture, Sports, Science & Technology in Japan. (2013). English Education Reform Plan corresponding to Globalization. Retrieved from http://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2014/01/31/134370_01.pdf

Ministry of Education, Culture, Sports, Science & Technology in Japan. (2014). *Heisei 26 Nenndo Shogakko Gaikokugo Katsudo Jisshi Jyoukyou Chosa* [Progress Report on Elementary School Foreign Language Activities, Year 26 of Heisei]. Retrieved from http://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2015/09/24/ 1362168_01.pdf

The National Institute of Information and Communications Technology (2011). Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles. Author. Retrieved from https://alaginrc.nict.go.jp/WikiCorpus/index_E.html

Ohashi, Y. & Katagiri, N. (2016). *Kyoshitsu Danwa ni Okeru Intaracshon wo Mochiita Meijiteki Shitou no Yukousei* [Effects of Explicit Instructions in Classroom Discourse]. Oral presentation at the 18th National Convention of the Japan Association of English Teaching in Elementary Schools. Miyagi, Japan.

Appendix A. Sample of Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles

```
<?xml version="1.0" encoding="UTF-8"?>
<art orl="ja" trl="en">
<inf>jawiki-20080607-pages-articles.xml</inf>
<tit>
<j>龍安寺</j>
 <e type="trans" ver="1">Ryoan-ji Temple</e>
 <cmt></cmt>
 <e type="trans" ver="2">Ryoan-ji Temple</e>
 <cmt>修正なし</cmt>
 <e type="check" ver="1">Ryoan-ji Temple</e>
 <cmt>修正なし</cmt>
</tit>
<par id="1">
 <sen id="1">
  <j>龍安寺（りょうあんじ）は、京都府京都市右京区にある臨済宗妙心寺派の寺院。</j>
   <e type="trans" ver="1">Ryoan-ji is a temple in the Myoshinji branch of the
   Rinzai sect, and is located in Ukyo-ku, Kyoto.</e>
   <cmt></cmt>
   <e type="trans" ver="2">Ryoan-ji is a temple that belongs to the Myoshinji
   school of the Rinzai sect, and is located in Ukyo-ku, Kyoto city.</e>
   <cmt>妙心寺派の「派」は school の方がよく用いられている。「妙心寺派の」という表現は「妙心寺
   派に属する」という意味である。「京都市」だけを訳出してあるので、city を添えた。</cmt>
   <e type="check" ver="1">A temple belonging to the Myoshinji school of the
   Rinzai sect, Ryoan-ji Temple is located in Ukyo-ku, Kyoto city.</e>
   <cmt>フィードバックに基づき翻訳を修正しました。</cmt>
 </sen>
 <sen id="2">

中略

</par>
</art>
```

（Retrieved from https://alaginrc.nict.go.jp/WikiCorpus/index_E.html#sample）

Appendix B. XSLT Stylesheet to Extract Exclusive Japanese Utterances of ALT

```
<?Xml version="1.0" encoding="UTF-8"?>

<!-- New XSLT document created with EditiX XML Editor (http://www.editix.com) at Wed Sep 21 21:22:50 JST 2016 -->

<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

        <xsl:output method="xml" indent="yes" encoding="UTF-8" omit-xml-declaration="yes" />

        <xsl:template match="/" >

                        <xsl:copy-of select="class/alt/j"></xsl:copy-of>
                        #<xsl:copy-of select="class/hrt/j"></xsl:copy-of>
        </xsl:template>

</xsl:stylesheet>

# This line is interchangeably used to extract exclusive Japanese utterances of HRT.
```

Appendix C. XSLT Stylesheet to Extract Japanese Utterances of ALT in Mixed Use of L1 and L2

```
<?Xml version="1.0" encoding="UTF-8"?>

<!-- New document created with EditiX at Mon Feb 01 18:13:55 JST 2010 -->

<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

        <xsl:output method="xml" indent="yes" encoding="UTF-8" omit-xml-declaration="yes" />

        <xsl:template match="/" >

                        <xsl:copy-of select="class/alt/mix/j"></xsl:copy-of>
                        #<xsl:copy-of select="class/hrt/mix/j"></xsl:copy-of>
        </xsl:template>

</xsl:stylesheet>
    # This line is interchangeably used to extract Japanese utterances of HRT in mix use of L1 and L2.
```

Appendix D. XSLT Stylesheet Sample to Extract L1 and TL2

```
<?xml version="1.0" encoding="UTF-8" ?>

<!-- New XSLT document created with EditiX XML Editor (http://www.editix.com) at Wed Sep 21 21:22:50 JST 2016 -->

<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">

        <xsl:output method="xml" indent="no" encoding="UTF-8" omit-xml-declaration="no" />

        <xsl:template match="/">
                <xsl:text>[LF]
                  </xsl:text>
                <xsl:for-each select="body/class/alt/mix">

                <xsl:copy-of select="j" /><xsl:text>[TAB]</xsl:text><xsl:copy-of select="TL2"/>

                <xsl:text>[LF]
                  </xsl:text>
                <xsl:text>[LF]
                  </xsl:text>

                </xsl:for-each>

        </xsl:template>

</xsl:stylesheet>
```

Note. We inserted line-feeding meta-characters (shown as *[LF]*) and a tab key (*[TAB]*) to align pairs of L1 and T2 elements clearly.

Appendix E. Results of TL2 Extraction from Mix Utterances of HRT

```xml
<?xml version="1.0" encoding="UTF-8"?>
```

<TL2>We are going to play a game called like that.</TL2>

<TL2>How many of you understood most of them?</TL2>

<TL2>This is why the game is called...</TL2>

<TL2>All right, then.</TL2>

<TL2>Is there anybody who are not sure?</TL2>

<TL2>Yes.</TL2>

<TL2>You are safe until you say thirty.</TL2>

<TL2>Next.</TL2>

<TL2>You are out if you draw...</TL2>

<TL2>Why don't you start from there clockwise?</TL2>

<TL2>Yes.</TL2>

<TL2>It's PE isn't it?</TL2>

<TL2>Yes, Rie.</TL2>

<TL2>For example, </TL2>

<TL2>Yes, XXX. </TL2>

<TL2>Let's start with PE. Go.</TL2>

<TL2>So I heard.</TL2>

（片桐　德昭　旭川校准教授）

（大橋由紀子　ヤマザキ学園大学講師）