Construction of the Chinese Learners' Parallel Corpus of Japanese and Its Preliminary Analysis

Masaaki Shimizu, Fenggang Du [†], and Masatake Dantsuji [‡]
Open University, Tokyo Metropolitan University, Japan
[†]School of Foreign Languages, Dalian University of Technology, China
[‡]Academic Center for Computing and Media Studies, Kyoto University, Japan

Abstract: This study aims to introduce the project to construct the Chinese learners' corpus (LC) of Japanese at Dalian University of Technology (DUT), and detail the LC construction, development of DUT Corpus Linguistics Tools, and contribution to the education of Japanese as a second language. The outstanding characteristic of the LC is its parallel form with learners' Japanese texts and their Chinese translation, which enables us to make comprehensive analysis of the influence of Chinese (L1) to Japanese (L2). We have made a preliminary analysis of the errors contained.

Key words: Chinese learners' corpus of japanese; parallel corpus; tagging tool, Sino-Japanese

Recognizing the significance of specifying the learners' first languages as well as the target languages in constructing LC, several L1 and L2-specified LC have been constructed. Especially in China, several projects to construct the Chinese learners corpora of English (Gui &Yang, 2003), including the Chinese portion of *International Corpus of Learner English*, *HKUST Corpus*, etcetera are under way (Yang, 2002). Meanwhile, the construction of learners' corpora of Japanese has been carried out mainly in Japan (Ooso & Takizawa, 2003), but their learners' first languages are fairly diversified. Given these backgrounds, this study focuses on the learners' corpus of Japanese written by the Chinese students to clarify the aspects of L1 to L2 interference.

METHODOLOGY

Japanese compositions written by 412 Chinese university students of 3rd or 4th year were collected, which contain about 13,800 sentences. The compositions were of two styles: one narrative and the other expository.

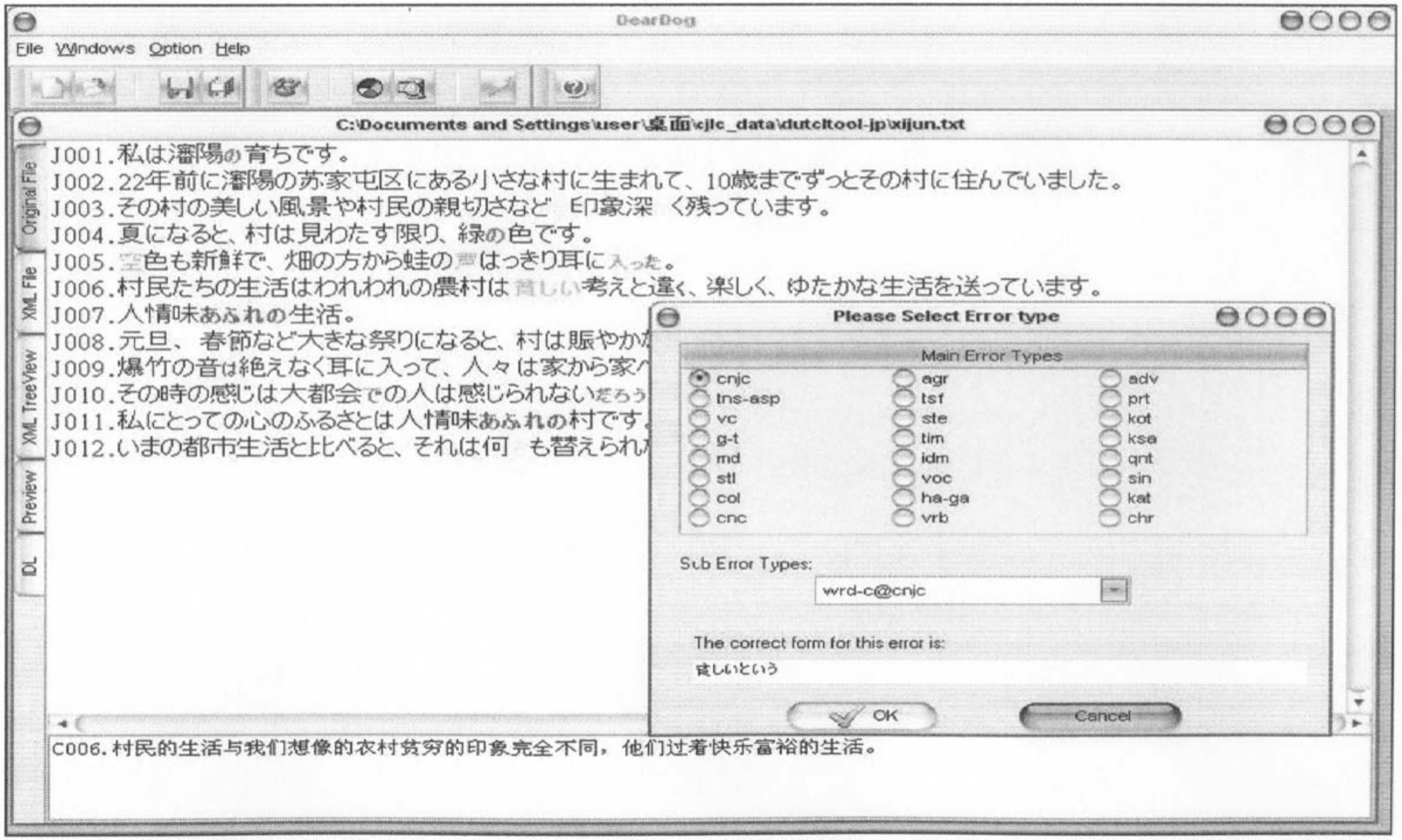
Selecting one of the topics: (1) *Japanese for Me*, (2) *My Hometown in Mind*, (3) *Computers and Language Learning*, and (4) *Economic Development and Environmental Problems*, the students wrote Japanese sentences first, and then translated them into Chinese, so that the parallel learners' corpus could be identified. From a pedagogical point of view, the order of writing Japanese sentences first and then translating them into Chinese is important for avoiding the unexpected transfer from L1. Translation of the compositions is also meaningful for avoiding the misunderstanding of the learners' intended meaning of each sentence, and only this point was explained to the students as the reason for their translation.

All the compositions were handwritten so that we could acquire the data of character errors, especially the use of Simplified Chinese Characters (*jiantizi*) or the different forms of characters which are not used in Japanese. Digitalizing of this data was undertaken using the following procedure: 1. to separate the whole body of compositions into sentences, 2. to number each sentence with initial 'J' for Japanese and 'C' for Chinese, giving the same number for equivalent sentences, and 3. to save each composition as text files (.txt).

Development of DUT Corpus Linguistics Tools

The next procedure in constructing LC is tagging to attribute the background information of the learners and the error information to each sentence. An original set of tools (DUT Corpus Linguistics Tools) was developed to tag the error information and conduct a preliminary statistical analysis. In the tagging window (Fig.1), we can input the information about learners' basic background and the data concerning each error after selecting the composition files and store them in the database file (.mdb). After completing these processes the XML files (.xml) describing all the information are automatically generated for each composition.

Based on the information stored in the database file, a preliminary statistical analysis (Fig.2) and simple query of errors can be carried out with these tools. We can acquire a general tendency of errors through this analysis and easily retrieve samples of each error type.



C:\Documents and Settings\user\桌面\cjlc_data\dutcltool-jp\xijun.txt

J001.私は瀋陽の育ちです。
J002.22年前に瀋陽の苏家屯区にある小さな村に生まれて、10歳までずっとその村に住んでいました。
J003.その村の美しい風景や村民の親切さなど 印象深 く残っています。
J004.夏になると、村は見わたす限り、緑の色です。
J005.空色も新鮮で、畑の方から蛙の声はっきり耳に入った。
J006.村民たちの生活はわれわれの農村は貧しい考えと違く、楽しく、ゆたかな生活を送っています。
J007.人情味あふれの生活。
J008.元旦、 春節など大きな祭りになると、村は賑やか
J009.爆竹の音は絶えなく耳に入って、人々は家から家へ
J010.その時の感じは大都会での人は感じられないだろう
J011.私にとっての心のふるさとは人情味あふれの村です。
J012.いまの都市生活と比べると、それは何 も替えられ

C006.村民的生活与我们想像的农村贫穷的印象完全不同，他们过着快乐富裕的生活。

Please Select Error type

Main Error Types

● cnjc	○ agr	○ adv
○ tns-asp	○ tsf	○ prt
○ vc	○ ste	○ kot
○ g-t	○ tim	○ ksa
○ md	○ idm	○ qnt
○ stl	○ voc	○ sin
○ col	○ ha-ga	○ kat
○ cnc	○ vrb	○ chr

Sub Error Types: wrd-c@cnjc

The correct form for this error is:
貧しいという

Figure 1. A tagging window of DUT Corpus Linguistics Tools

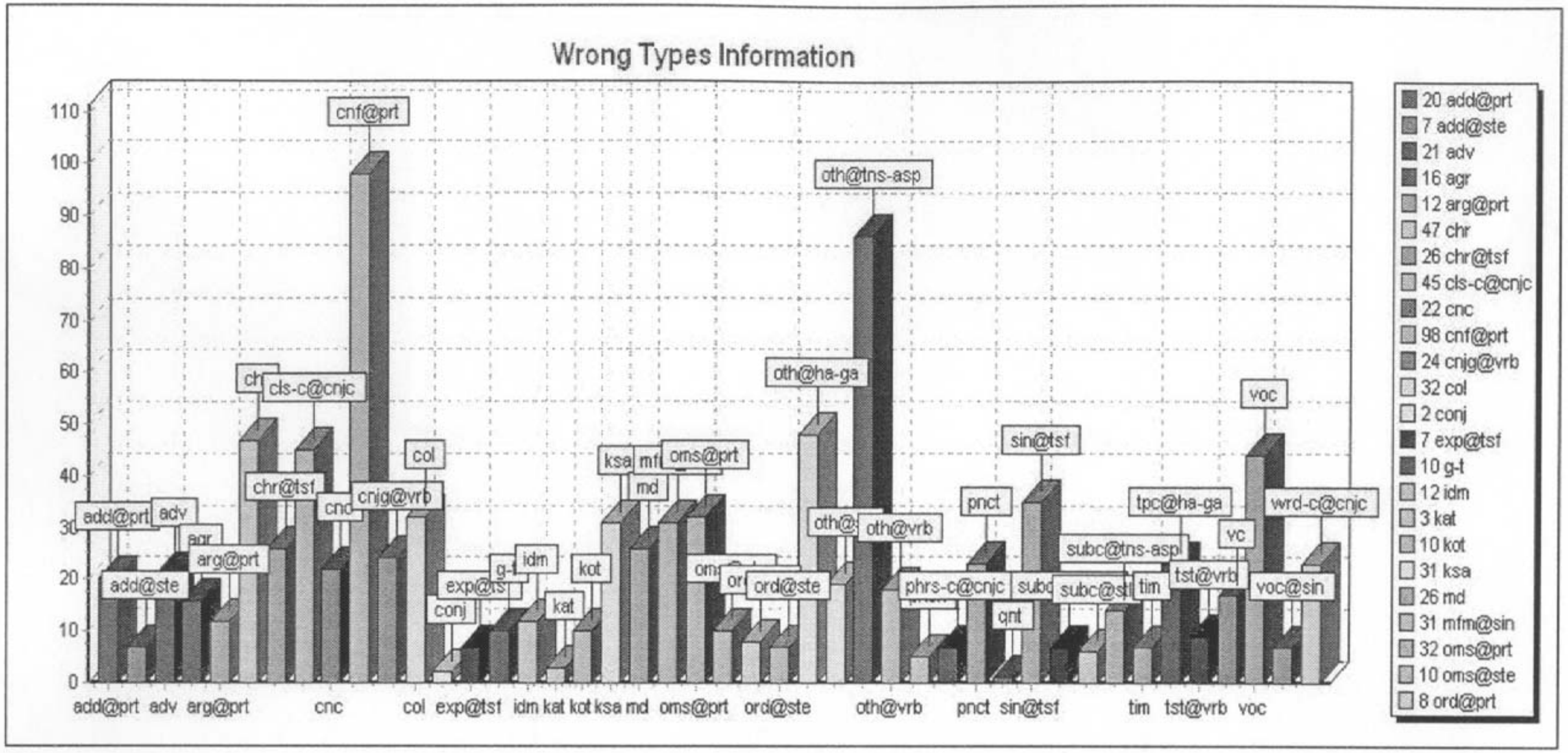


Figure 2. A chart of error types

Tagging

To carry out the tagging process, we have constructed a preliminary error tagset for Chinese Learners Japanese compositions (Error Tagset for DUT CJLC, ver.2.01). In general, we can think of two ways of constructing the error tagset: one is based on the existing grammatical framework of Japanese (Ooso et al., 1998), and the other is to extract the error types from a certain quantity of samples (Ichikawa 1997, 2000). The former will be applicable for the LC construction with L1-diversified learners, and the latter for the L1-specified learners. Therefore, adopting the latter method, we have analyzed about 80 samples of compositions containing about 1,890 sentences to extract error types specific to Chinese learners of Japanese. The extracted error types and the constructed tagset are shown in Table 1.

Table 1. Error Tagset for DUT CJLC (ver.2.01)

Conjunctive	Word-Level [*wrd@cnjc*]
	Phrase-Level [*phrs@cnjc*]
	Clause-Level [*cls@cnjc*]
Tense and Aspect	Subordinate Clause [*subc@tns-asp*]
	Other [*oth@tns-asp*]
Voice [*vc*]	
Verbs of *yari* (Giving) and *morai* (Taking) [*g-t*]	
Mood [*md*]	
Style [*stl*]	
Collocation [*col*]	
Concord [*cnc*]	
Logical Agreement between Subjects and Predicates [*agr*]	
Transfer	Character [*chr@tsf*]
	Sino-Lexicon [*sin@tsf*]
	Other Lexicon [*vcb@tsf*]
	Expression [*exp@tsf*]
Sentencial Elements (verbs, arguments, adjuncts, etc.)	Omission [*oms@ste*]
	Addition [*add@ste*]
	Ordering [*ord@ste*]
Time Expression [*tim*]	
Idiomatic Expression [*idm*]	
Vocabulary [*voc*]	
-ha (topic marker) and *-ga* (nominative marker)	Topicalization [*tpc@ha-ga*]
	Subordinate Clause [*subc@ha-ga*]
	Other [*oth@ha-ga*]
Verbals	Conjugation [*cnjg@vrb*]

	Transitivity [*tst@vrb*]
	Other [*oth@vrb*]
Adverbials [*adv*]	
Particles	Confusion [*cnf@prt*]
	Omission [*oms@prt*]
	Ordering [*ord@prt*]
	Addition [*add@prt*]
	Argument Marker [*arg@prt*]
Conjunction [*conj*]	
Formal Nouns (*-koto*, *-no*, etc.) [*kot*]	
Ko-so-a Words (deictic pronouns) [*ksa*]	
Quantifiers [*qnt*]	
Sino-Lexicon	Vocabulary [*voc@sin*]
	Misformation [*mfm@sin*]
Katakana Words (Western loan words) [*kat*]	
Chinese character [*chr*]	
Phonetic [*phon*]	
Punctuation [*pnct*]	

Note: The abbreviations in brackets ([ ]) are used in the tagging tool.

As the tagging process has tentatively commenced, the tagset proposed here (ver.2.01) will be revised in future process.

FINDINGS AND DISCUSSION

Through the analysis of samples done for extracting the error types and the preliminary tentative tagging work, we have come across several predominant error types, such as tense and aspect errors, confusion, omission, and addition of the particles, confusion of *–ha* (topic marker) and *–ga* (nominative marker), etcetera, among which one of the most significant errors specific to the Chinese learners of Japanese is that of the usage of Sino-lexicon (or Sino-Japanese). Historically speaking, the Sino-Japanese lexes were the loan words from Chinese, especially from the Chang'an dialect in the Tang dynasty (618-907), as one of the most important cultural borrowings from China. In the Meiji era (1868-1911), the Japanese invented their own Sino-lexicon (the words formed on the Chinese originated morphemes) for translating the Western cultural items, and some of which were re-borrowed to China and other East Asian countries (Shin, 1994). The point here is that the forms of Sino-lexicon are often identical between modern Japanese and Chinese but their semantic and syntactic usages are not always the same, which cause various types of errors by the Chinese learners of Japanese. As a preliminary consideration, we have analyzed some samples of each linguistic level, and in this section, we focus on the samples of syntactic level and point out the significance of special pedagogical consideration to this problem and the necessity of compiling the educational lexicon of Sino-Japanese.

The typical characteristic of this error type is its correctness in the lexical selection and its error in the syntactic usage. These errors can be categorized into two groups: the first are the errors caused by word-for-word translation, and the second, also more important, are those of overlapping the syntactic properties of L1 lexicon to that of L2.

The examples of the former type are as follows: ***the first line*** of each example is the form corrected by the Japanese native speaker, ***the second line*** is the learners' original writing, and ***the third line*** is the learners' own translation into their mother tongue (Chinese). All the examples are transcribed into Latin characters for convenience of printing.

(1) *nihongo* *sonomno* *-ni* *kyoomi* *-o* *kanzi* *-nai*
Japanese itself DAT[1] interest ACC feel NEG
nihongo *sonomono* ***-ni*** ****kyoomibukaku*** *-nai*
be interested
dui *riyu* *benshenbing* *bu* *tai* ***ganxingqu***
in Japanese itself at all not so be interested
'not so interested in Japanese itself' [003_017][2]

The expression *ganxingqu* in (1) can be analyzed as the verb *gan* 'feel' with its complement noun phrase *xingqu* 'interest', and be translated into Japanese as *kyoomi-o* (interest-ACC) *kanziru* (feel). In this case, the noun phrase meaning the object of interest can be marked with *-ni*, an equivalent to the Chinese *dui*. However, it is always the case that the whole expression *ganxingqu* is used to translate the Japanese equivalent expression *kyoomibukai* 'be interested in.' And the Japanese adjective

[1] Abbreviations of grammatical terms used here are: ACC(usative) , DAT(ive), NEG(ative), NOM(inative), PERF(ective), TOP(ic), P(arti)CL(e).

[2] Examples quoted here are all from the writings entitled *Japanese for Me* and *Economic Development and Environmental Problems,* composed by the DUT learners, which were used to extract the error types and the ID number given in [ ] denotes: [(the writer's number)_(sentence number)].

kyoomibukai takes an argument NP marked with *-ga* which means the object of interest. In this way, the error occurred by using the adjective with an unaccepted marker *-ni* which is equivalent to the Chinese *dui*.

(2) *aru teido zibun -no nooryoku -o arawasu*
some extent oneself of ability ACC represent
aru*teido ***ni*** *zibun -no nooryoku -o hyookasuru*
to evaluate
zai *yiding chengdu* ***-shang*** *daibiao -le ziji de nengli*
on some extent above represent PERF oneself of ability
'represent one's ability to some extent' [004_006]

(3) *aruteido ... keikoo -ga deteki*
tendency NOM appear
aru*teido ***-de*** ... *keikoo -ga* *miidasi
on find out
zai *mo zhong chengdu* ***-shang*** *chuxian -le ... qingxiang*
on some kind extent above appear PERF ... tendency
'there appeared a tendency ... to some extent' [036_027]

As for (2) and (3), the Japanese expression *aru teido* 'to some extent' is a noun phrase which is used as an adverbial without being marked by adverbial suffixes (*-ni*, *-de*). The learners added *-ni* and *-de* to the noun phrase to give equivalents to the Chinese *zai*.

(4) *yuugai-na kagakugenso*
harmful chemical element
*yuugai ***-no*** *kagakugenso*
harm of
youhai ***de*** *huaxueyuansu*
harm of chemical element
'harmful chemical elements' [029_016]

(5) *zibun -ni yuueki-na hookoo*
oneself DAT beneficial direction
zibun -ni *yuueki ***-no*** *hookoo*
benefit of
dui ziji youyi ***de*** *fangxiang*
to oneself benefit of direction
'the direction beneficial to oneself' [037_024]

These examples can also be explained as the errors caused by translating the boldface element *de* of Chinese to the Japanese equivalent -*no*, which form the unacceptable Sino-lexicon in standard Japanese.

The second group contains the errors overlapping the syntactic properties of the L1 lexicon to that of L2. To identify the part of speech of each Chinese lexicon, we follow the corpus-based categorization proposed in *The Grammatical Knowledge-base of Contemporary Chinese – A Complete Specification* (*Xiandai Hanyu Yufa Xinxi Cidian Xiangjie*) and the word lists of each part of speech given in *Hu* (2004). Some examples include the misformed *na*-ending adjectives as follows:

(6) *Kintyoo-si* *tari* *tadotadosikat* *tari*
be tense$_{(v)}$ and falter and
Kintyoo-ninat *tari* *tadotadosikat* *tari*
(tense$_{(adj)}$) become
jinzhang *jieba*
tense$_{(adj)}$ faltering
'be strained and faltering' [004_007]

(7) *keizai* *-wa* *sonnani* *hattatu-site-i* *-nai*
economy TOP not so be progressing$_{(v)}$ NEG
keizai *-wa* *sonnani* ****hattatu-de*** *-wa* *-nai*
(progressive) TOP
jingji *haimeiyou* *fazhan* cf. ***fada***
economy not yet progress progressive$_{(adj)}$
'not making so much progress in the economy' [011_033]

(8) *mainiti* *zyuuzitu-site* …
everyday fill up
mainiti ****zyuuzitu-de*** …
(full)
mei yi tian *dou* *guo* *-de* *chongshi*
everyday all live PCL full
'everyday live a full life' [021_024]

(9) *seikoo-suru* *hito*
succeed$_{(v)}$ person
****seikoo-na*** *hito*
(successful)

chenggong *de* *ren*
successful of person
‘successful person’ [023_022]

In (6) – (9), the correct form of all the examples is the *suru*-ending verb (*kintyoo-suru*, *hattatu-suru*, *zyuuzitu-suru*, and *seikoo-suru*). However, in the learners’ writing, they are all misformed as *na*-ending adjectives (**kintyoo-ni*, **hattatu-de*, **zyuuzitu-de*, and **seikoo-na*), because of the corresponding Chinese adjectives (*jinzhang*, *fada*, *chongshi*, and *chenggong*). The following examples are concerned with the argument structure of L1 verbs.

(10) *keizai* *-o* *hatten-s* *-aseru* *tameni*
economy ACC develop $_{(vi)}$ CAUSE in order to
keizai *-o* ****hatten-suru*** *tameni*
develop
weile ***fazhan*** *jingji*
to develop economy
‘in order to develop the economy’ [011_006]

(11) *zinsei* *-ni* *kantan-site*
life DAT admire
zinsei* *-o*** *kantan-site*
ACC
gantan *ta* ... *yisheng*
admire his … life
‘admire his life …’ [001_018]

(12) *watasi* *-no* *tisiki sisutemu* *-o* *kenzen-ni* *suru*
I of knowledge system ACC sound-DAT make
watasi *-no* *tisiki sisutemu* *-o* ****kenzen-suru***
(make sound)
jianquan *-le* *wo* *de* *zhishi tixi*
make sound PERF I of knowledge system
‘made my knowledge sound’ [015_018]

(13) *bukatu* *-no* *koto* *sika* *kansin* *-ga* *nakat* *-ta*
club activity of matter only interest NOM not possess PAST
bukatu *-no* *koto* *sika* ****kansin-si*** *-nakat* *-ta*
(be interested in) NEG
zhi ***guanxin*** *sheduan huodong*
only be interested in club activity
‘be only interested in club activities’ [023_016]

In (10) – (13), all the Chinese equivalents of misformed Japanese lexicon are the verbs which take two arguments, and the internal argument of them is realized as a NP following the verb without any grammatical marker: *fazhan* 'develop *something*' in (10), *gantan* 'admire *something/someone*' in (11), *jianquan* 'make *something* sound' in (12), and *guanxin* 'be interested in *something*' in (13). Meanwhile, in the standard Japanese, *hatten-suru* in (10) is a verb taking only a single nominative argument, *kantan-suru* in (11) takes a dative NP (NP-*ni*) as the internal argument, *kenzen-na* in (12) is a *na*-ending adjective which must be used in a causative construction to express the intended meaning, and *kansin* in (13) can only be used as a noun. The correct forms of all these examples are varied this way, but the common characteristic among them is that all of them are misformed as verbs followed by the accusative NPs (NP–*o*) as the internal arguments. These examples can also be regarded as taking place through the interference of the syntactic property of L1 lexicon, that is, 2-argument verb, to L2.

(14) *dokuritu-sita* *seikatu*
be independent$_{(v)}$ life
****dokuritu*** *seikatu*
independence
duli *shenghuo*
independence life
'independent life' [017_019]

In (14), both Japanese *dokuritu-suru* and Chinese *duli* are verbs with their corresponding noun forms: *dokuritu* and *duli*. When forming the compound nouns including these nouns, they can be used as bare nouns like *dokuritu kokka* and *duli guojia* 'independent nation.' However, in the case of the compound meaning 'independent life,' the similar construction *duli shenghuo* is acceptable in the standard Chinese, while *dokuritu seikatu* is unacceptable in Japanese, the correct form for which is *dokuritu-sita seikatu* with the *suru*-ending verb form.

(15) *eien* *-no* *kadai*
eternity of problem
****eienna*** *kadai*
(eternal)
yongyuan *de* *wenti*
eternally of problem
'eternal problem' [026_012]

We can suppose several ways to explain this error, among which we can still think about the possibility of L1 interference. Chinese _yongyuan_ is described as an adverb, while the adverb form of Japanese _eien-ni_ cannot be a direct modifier to the noun _mondai_ as in *_eien-ni mondai_. In this case, we may think of two ways to avoide this problem, one is to give a word-for-word translation as in _eien no mondai_, and the other is to give an adjective form meaning 'eternal' as in *_eien-na mondai_. Unfortunately, the learner may have selected the latter way and provided the misformed expression. If this analysis is not the case, another possibility might be the learner's avoidance of word-for-word translation which turned out to give the correct form.

The following are the error samples of logical structure and collocation which can also be analyzed as the interference of Chinese to Japanese. Errors of logical structure can be seen in the following samples.

(16) _Shoorai no seikatu-o soozoosuru-no mo tanosii mono dearu._
Future of life-ACC imagine-to also pleasant thing be
*_shoorai no seikatu-o soozoosuru-no-ha_ **_issyu-no omosirosa da_** _to omou_
Future of life-ACC imagine-to-TOP a kind of pleasure be that think
Xiangxiang jianglia de shenghuo ye **_shi yizhong lequ_**.
Imagine future of life also be a kind of pleasure
'It is interesting to think about the future.' [005_027]

(17) _nihongo-ga daiitini hansyatekini dete kita_
Japanese-NOM firstly reflectively have come out
*_nihongo-ha_ **_daiiti hannoo-ni natta_**.
Japanese-TOP first reaction-DAT have become
riyu **_chengwei-le diyi fanying_**.
Japanese become-PERF first reaction
'Japanese words came out as the first reaction.' [006_018]

The following samples of collocation errors can also be explained from the view point of L1 influence.

(18) _kyuusokuna hatten-o toge-ta_
Rapid progress-ACC accomplish-PERF
kyuusoku hatten-o ***tot**-ta_
Rapid progress-ACC take-PERF

qude *feisu* *fazhan*
acquire rapid progress
'made fast progress' [035_006]

(19) *kyodaina* *sinpo-o* *toge-te*
huge progress-ACC accomplish-and
kyodaina *sinpo-ga* ***tore**-te*
-NOM can take-and
***Qude**-le* *juda* *de* *jinbu*
acquire-PERF huge of progress
'made huge progress' [036_002]

(20) *tekidoni* *gensoku-o* *mamori*
moderatelyrule-ACC obey
tekidono *gensoku-o* ***moti***
moderate have
Jianchi *shidu* *yuanze*
Keep moderate rule
'moderately obey the rules'[036_040]

Most of the examples analyzed so far, especially (6) – (15), can be explained by the lack of learners' attention to the syntactic properties of Sino-Japanese lexicon. Several course books and materials on the Sino-Japanese lexicon, especially designed for Chinese learners, have been published so far (Peng 2003, Wu 1999, Qu 1996, etcetera), but they often pay more attention to the problems of semantic level than those of syntactic level. Here we point out the necessity of paying more attention to the syntactic differences and constructing the Sino-Japanese lexicon with basic syntactic information (parts of speech and argument structure) designed from the pedagogical point of view.

CONCLUSION

In this study, we have introduced our project to construct the Chinese learners' corpus of Japanese compositions and to develop the DUT Corpus Linguistics Tools, and made a preliminary analysis of the typical errors contained there, especially those that can be explained by the interference from Chinese to the Sino-Japanese lexicon.

As the next step, besides continuing to analyze our corpus, we need to study the same phenomenon in relation to other L1 speakers of a Sino-cultural background, such as Korean and Vietnamese, to see if the same thing happens. We also need to look at the problem of lexical acquisition to see if the syntactic property of L1 lexicon, such as parts of speech and argument structure, influence acquisition of L2 lexicon and syntax (White, 2003; Wei, 2004).

As syntactic properties of L1 lexicon could possibly be the most important point of L2 lexical and syntactic acquisition, we would like to point out the urgent necessity of constructing the Sino-Japanese lexicon with basic syntactic information, especially for learners from a Sino-cultural background.

ACKNOWLEDGEMENT

We are grateful to all the contributors at Dalian University of Technology and Dalian College of Foreign Languages. We are also grateful to Prof. Liu Liguo at Dalian College of Foreign Languages for his support in collecting the data at his college, and Mr. Li Rui and Mr. Fu Yifan at Department of Computer Science at Dalian University of Technology for their support in developing the DUT Corpus Linguistics Tools.

REFERENCES

Gui, S., & Yang, H. 2003. *Zhongguo xuexizhe yingyu yuliaoku (Chinese learner English corpus).* Shanghai: Shanghai Waiyu Jiaoyu Chubenshe.

Hu, M. 2004. *Cilei wenti kaocha xuji (Consideration of the problems of parts of speech – continued).* Beijing: Beijing Yuyan Daxue Chubenshe.

Ichikawa, Y. 1997. *Nihongo goyoo reibun syoojiten (A dictionary of Japanese language learners' errors).* Tokyo: Bonzinsya.

Ichikawa, Y. 2000. *Zoku-nihongo goyoo reibun syooziten (A dictionary of Japanese language learners' errors II).* Tokyo: Bonzinsya.

Ichikawa, Y. 2001. *Nihongo no goyoo kenkyuu (Japanese error analysis): Nihongo kyooiku tuusin.* Tokyo: The Japan Foundation.

James, C. 1998. *Errors in language learning and use: exploring error analysis.* London: Addison Wesley Longman.

Ooso, M., et al. 1998. *Nihongo gakusyuusya no sakubun koopasu: densika ni yoru kyooyuu sigen ka (A learners' corpus of Japanese compositions: digitalizing and sharing the data).* Poster Presentation at the 15th Meeting of Japanese Cognitive Science Society.

Ooso, M., & Takizawa, N. 2003. Koopasu ni yoru Nihongo kyooiku no kenkyuu – korokeesyon oyobi sono goyoo o tyuusin'ni (A corpus-based study on Japanese education), *Nihongogaku*, *22(5)*, Meizisyoin, 234-244.

Peng, F. 2003. *Gaikokuzin o nayamaseru Nihongo kara mita Nihongo no tokutyoo (Characteristics of Japanese - from examples which trouble foreigners).* Tokyo: Bonzinsya.

Qu, W. 1996. *Riben de Wenzi (Japanese writing system)*, Jilin: Jilin Renmin Chubanshe.

Shimizu, M., & Du, F. 2003. Guanyu zhongguo riyu xuexizhe de zhongri duiyi zuowen yuliaoku jianshe de yanjiu – fuma fuzhu gongju de yanzhi – (A study on the construction of parallel learners corpus of Japanese composition writtern by the Chinese learners – a development of tagging tools). *Waiyu Jiaoxue Waiyu Yanjiu*, *3*, *(Internal Communication)*, 21-26.

Shimizu, M., Du, F., & Dantsuji, M. 2004. *Tyuugokuzin Nihongo gakusyuusya aiyaku sakubun koopasu no kootiku ni okeru ayamari tagu no kootiku ni tuite (Construction of the error tagset in the development of Chinese learners' Japanese writing parallel corpus)*, Paper presented at the International Symposium on Education and Study of Japanese Language and Culture, Beijing, October, 20-23, 2004.

Shin, K. 1994. *Kindai nittyuu goi kooryuusi (A history of Sino-Japanese lexical exchange in the modern era)*. Tokyo: Tikuma Syoin.

Wei, L. 2004. *Second language production of English dative constructions: a lexicon-driven approach to target syntax.* Paper presented at the Sixth International Symposium on Applied Linguistics and Language Teaching, Beijing-Shanghai, August 7-11, 2004.

White, L., 2003, *Second language acquisition and universal grammar.* Cambridge: Cambridge University Press.

Wu, K. 1999. *Riyu cihui yanjiu (A study of Japanese lexicon)*, Shanghai: Shanghai Waiyu Jiaoyu Chubanshe.

Yang. H. 2002. *Yuliaoku yuyanxue daolun (An introduction to corpus linguistics).* Shanghai: Shanghai Foreign Language Education Press.

Yu, S., et al. 2003. *Xiandai hanyu yufa xinxi cidian xiangjie, di er ban (The grammatical knowledge-base of contemporary Chinese - a complete specification). 2nd* ed. Beijing: Qinghua Daxue Chubanshe.