Vocabulary Usage and Intelligibility in Learner Language

Emi Izumi,[1] Kiyotaka Uchimoto[1] and Hitoshi Isahara[1]

1. Introduction

In verbal communication, the primary purpose of which is to convey and understand messages, speakers need to make intelligible utterances. The intelligibility of an utterance depends on the sentence structure, discourse, and the vocabulary usage. In foreign-language utterances, "errors" can reduce the intelligibility. The correct and appropriate use of vocabulary is essential for successful message conveyance especially in foreign language communication where people often have difficulty in constructing sentences precisely, and where they rely more on vocabulary because of their lower grammatical competence. Vocabulary skill development has become one of the high-priority issues in recent foreign language education.

We analysed on the correlation between vocabulary usage and the intelligibility of utterances made by Japanese learners of English by investigating to what extent lexical errors interfere with the intelligibility of utterances. We did this based on the error-coded learner corpus in which each sentence is labelled with its levels of intelligibility. More precisely, we calculated lexical semantic relatedness between an erroneous word and a correct word, and then observed how the relatedness values are distributed across different levels of intelligibility and different proficiency levels as well. The remainder of this paper is as follows. In Section 2, we discuss how intelligibility is positioned in foreign language learning and teaching. Section 3 explains the corpus data used in the experiment, focusing especially on data annotation of intelligibility as evaluated by humans and the analysis of the relationship between intelligibility and learner errors. Section 4 describes our further investigation into the correlation between lexical errors and intelligibility conducted by focusing on lexical semantic relatedness between an erroneous word and a correct word. In Section 5, we draw some general conclusions.

2. Intelligibility of learner language

First we would like to consider how intelligibility is positioned in foreign language learning and teaching, especially in recent language education based on the communicative approach.

Improving communicative competence is one of the major goals in a communicative approach to foreign language teaching as stated by Ellis (2003) in the following quote. "Learners need the opportunity to practice language in the same conditions that apply in real-life situations - in communication, where their primary focus is on message conveyance rather than linguistic accuracy". To successfully convey messages by producing "intelligible" utterances that can be understood by

[1] Computational Linguistics Group, National Institute of Information and Communications Technology, Japan
e-mail: emi@nict.go.jp, uchimoto@nict.go.jp, isahara@nict.go.jp

others is important. Similarly, according to Skehan (1998), meaning and task-completion are primary factors in communication task activities, and are often employed in a communicative approach.

It is true that too much concentration on accuracy sometimes prevents learners from acquiring free language production and fluency, especially in speech communication because it often introduces more time pressure than does writing. However, this does not mean that learners can hold accuracy in low account in language production because obviously if linguistic components such as grammar, lexis, or phonemes that constitute the bedrock of languages are completely inaccurate, language communication does not occur. Accuracy, especially of grammar, is often contrasted with communicability, but Canale and Swain (1980) confirm that grammatical competence is one of the important elements for building communicative competence. Since accuracy and communicability (intelligibility) are complementary, we need to know the extent to which accuracy should be taken into account in communicative foreign language production. In other words, if we could describe what kind of factors can change the level of intelligibility explicitly and could recognize the necessary degree of accuracy for making communication successful, this would effectively help improve communicative competence.

3. Intelligibility of Japanese learner English

In order to describe the level of intelligibility of the learner language explicitly, we first decided to add level-of-intelligibility information to the learner corpus.

3.1 Human judgment of intelligibility

We asked two native English speakers to check the corpus data and measure the intelligibility of sentences by labelling each sentence with either “intelligible”, “unclear” or “unintelligible” (Table 1). Although the labelling was done sentence by sentence, the checkers decided the level of intelligibility based on the “contextual” intelligibility of each sentence. A sentence that could easily be understood was labelled “intelligible”, even if it contained errors. A sentence would be labelled “unclear”, if it made sense, but was sometimes unclear or did not sound like native speech. If the checkers could not understand the meaning of a sentence at all, they labelled it “unintelligible”. If errors were found in a sentence, the checkers rewrote it.

Level of intelligbility	Description
intelligible	There is no difficulty in understanding the meaning of the sentence.
unclear	It is possible to understand the meaning of the sentence, but the sentence is sometimes unclear or sounds unnatural.
unintelligible	The sentence does not make sense at all.

Table 1: Level of intelligibility

The sentences we used are part of the NICT Japanese Learner English (JLE) Corpus (Izumi *et al.* 2003). This corpus consists of transcriptions of an oral proficiency test, the *Standard Speaking Test* (*SST*). The *SST* is a face-to-face interview of a test-taker conducted by an examiner. This 15-minute interview test is comprised of an informal chat and three task-based activities: picture description, role-playing, and story telling. Two or three raters judge the proficiency level of each examinee (Levels 1 to 9. Level 9 is the most advanced.) based on an *SST* evaluation scheme. The entire corpus contains 1,281 interviews, which amount to 325 hours and two million words. The results of the human judgment, including the numbers of "intelligible", "unclear" and "unintelligible" sentences, the number of words, and the average sentence length (mean length of utterance: MLU) are presented in Table 2.

Level of intelligibility	# of sentences	# of words	MLU (words)
intelligible	5,774	30,530	5.28
unclear	1,282	15,058	11.74
unintelligible	238	2,198	9.23
total	7,294	47,786	6.55

Table 2: Results of human judgment

From a total of 7,294 sentences, 5,774 sentences were labelled "intelligible", 1,282 were "unclear", and 238 were "unintelligible". The MLU was 5.28 words for "intelligible", 11.74 for "unclear", and 9.23 was for "unintelligible".

The numbers of intelligible, unclear and unintelligible sentences per 100 sentences across different proficiency levels are presented in Figure 1.

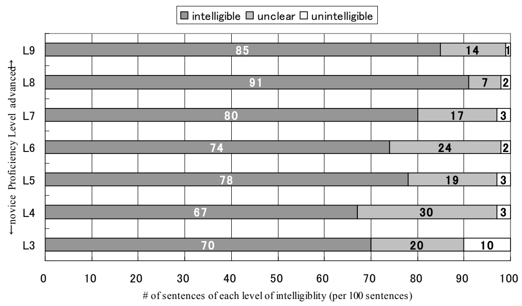

Figure 1: Number of sentences of each level of intelligibility across proficiency levels

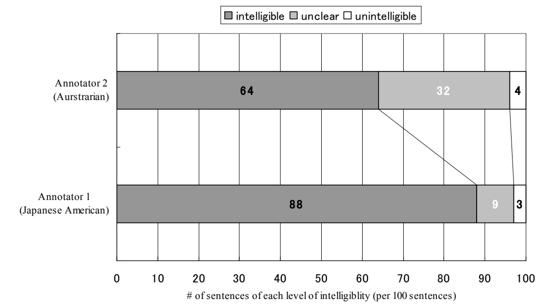

Figure 2: Results of human judgment on a per-annotator basis

Intelligible sentences accounted for 67–70 percent of Level 3 and 4 data. In Level 5 and 6 data, this rose to 74–78 percent. At advanced levels (Levels 7, 8 and 9), this increased to around 80–90 percent. The number of uclear sentences did not always correlate with the proficiency level. This category accounted for 7–30 percent of all the texts. The number of unintelligible sentences in Level 3 data was more remarkable (10 percent) than those in other proficiency levels (1–3 percent).

One of the reasons why the number of these three levels of sentence intelligibility does not completely correlate with proficiency levels might be that two people checked the data, and their judgment might have been disparate. Twenty-seven texts were labelled by Annotator 1, a Japanese American, and 22 texts were labelled by Annotator 2 from Australia. Figure 2 shows the result of human judgment on a per-annotator basis. Annotator 1 judged 88 percent of the sentences as intelligible, while Annotator 2 judged 64 percent of the sentences as intelligible. The gap between the annotators' evaluations becomes bigger for unclear sentences. The sentences labeled by Annotator 1 as unclear account for only 9 percent, while for Annotator 2, this goes up to 32 percent. On the other hand, no big difference was found in their judgment of unintelligible sentences. This accounted for 3-4 percent of the data in the evaluations of both annotators. Guessing from their background, Annotator 1 might be more familiar with English spoken by Japanese people than Annotator 2 because Annotator 1 is Japanese American and has some knowledge of Japanese language.

3.2 Error tagging and extraction of feature quantity of each error type

To study the relationship between intelligibility and errors, we added error tags to the data by hand. Errors were localized and categorized by referring to the corrections made by the native speakers. We used the error tags that were already implemented as part of the NICT JLE Corpus. The error tagset consists of 46 tags. Most of the tags are related to morphological, grammatical and lexical errors, which are, in most cases, local errors, but some are special tags that involve global errors such as incorrect word order. We then clustered the error-tagged sentences into three groups depending on their intelligibility ("intelligible", "unclear" and "unintelligible"), and then extracted the feature quantity of each type of error for each cluster. The feature quantity is the proportion of frequency of a certain type of error in a cluster compared to the frequency of the same type of error in all of the data (normalized per 1,000 words). This information can be used to help estimate the gravity of each type of error. Table 3 shows the feature quantity of major types of errors in three clusters.

As shown in Table 3, errors in morphological inflection of nouns, verbs and adjectives were distinctively frequent in unclear sentences. Some of them appear in intelligible sentences, too, but in unintelligible sentences, they are not distinctively frequent at all. In this type of error, an erroneous word appears in a non-existing form and sounds quite unnatural; however, this error does not really interfere with understanding because in most cases, a listener is able to guess which word the speaker intended to produce.

Major grammatical errors such as errors in noun number, verb tense, compliment of verbs and articles are also distinctively frequent in unclear sentences. Some of them appear in intelligible and unintelligible sentences, too, so some grammatical errors appear not to interfere with understanding while others make sentences unintelligible.

Lexical errors for content words are distinctively frequent in unclear and unintelligible sentences. Special types of lexical errors such as Japanese English, erroneous collocational expressions, had a certain degree of influence in making sentences unclear and unintelligible, and the use of Japanese words can greatly interfere with understanding.

4. Correlations between vocabulary usage and intelligibility

On the basis of the results shown in Table 3, we further investigated the correlation between lexical errors and intelligibility, and the relationship between lexical errors and proficiency levels as well. It is widely recognized that lexical competence is essential for being able to communicate in a foreign language. One might be able to speak using just a few grammar rules and might still be understood, but without using appropriate vocabulary, communication can hardly be successful (Kormos 2006). The aim of the investigation was to learn more about the broad pattern of how lexical errors can change the level of intelligibility and can vary across different proficiency levels.

Error type	Level of intelligibility	intelligible	unclear	unintelligible
Morpheme	(1) noun inflection	0.00	3.17	0.00
	(2) verb inflection	0.52	2.11	0.00
	(3) adjective inflection	0.78	1.58	0.00
	(4) countability of noun	0.60	1.95	0.00
Grammar	(5) number of noun	0.71	1.61	0.75
	(6) subject-verb agreement	0.68	1.69	0.64
	(7) verb tense	0.59	1.90	0.44
	(8) complement of verb	0.41	2.23	0.62
	(9) position of adverb	0.73	1.58	0.67
	(10) article	0.63	1.77	0.73
	(11) verb form	0.78	1.28	2.03
	(12) verb negation	0.46	1.86	2.55
	(13) number/gender agreement of pronoun	0.32	0.93	1.82
Lexis	(14) noun	0.51	1.85	1.93
	(15) verb	0.50	2.01	0.87
	(16) adjective	0.47	2.00	1.42
	(17) adverb	0.52	1.75	2.40
	(18) normal preposition	0.57	1.83	1.15
	(19) dependent preposition	0.50	2.09	0.36
	(20) conjunction	0.33	2.31	1.21
	(21) collocation	0.50	1.97	1.15
	(22) Japanese English	0.73	1.49	1.27
Others	(23) word order	0.31	2.32	1.44
	(24) global errors	0.33	2.09	2.78

Table 3: Feature quantity of major types of errors in three levels of intelligibility

4.1 Analysis

We analysed the relationship between lexical errors, intelligibility and proficiency levels by measuring semantic relatedness between the pairs of concepts (of an erroneous word and a correct word) using the concept hierarchy described in WordNet. The details of the criterion used for measuring relatedness will be stated in 4.2.2. It should be noted that we tried to obtain semantic relatedness between not the lemmas of two words, but the "senses" of them in a particular context. Therefore, we first decided in which sense each word was used in each context. It is difficult to set the criterion of deciding the sense of an erroneous word. In this analysis, we chose the sense which is the most similar to the sense of a correct word. If an erroneous word didn't have a similar sense, the first sense was chosen.

We used two types of data in the analysis. For the analysis of the relationship between lexical errors and intelligibility, we used the same data (50 files. The total number of words is 47,786 words.) as we used in the analysis described in Section 3. To analyse the relationship between lexical errors and proficiency levels, another 167 files (The total number of words is 131,195.) without intelligibility information were used.

4.2 Lexical semantic relatedness

4.2.1 Definition of lexical semantic relatedness

First we would like to confirm the definition of lexical semantic relatedness. According to Budanitsky and Hirst (2006), when discussing the relationship between the concepts of two different words, it is necessary to distinguish clearly among the following three terms: semantic relatedness, semantic similarity, and semantic distance. Resnik (1995) distinguishes the first two terms by saying, "*Cars* and *gasoline* would seem to be more closely related than *cars* and *bicycles*, but the latter pair is certainly more similar". From this perspective, we could assume that semantic similarity is a type of semantic relatedness. On the other hand, semantic relatedness includes not only similarity, but also other kinds of relations such as meronymy, antonymy, functional association, and so on. The third term, "semantic distance" can be considered as the inverse of semantic relatedness". Budanitsky and Hirst (2006) claim that two concepts are "close" to one another if their similarity or their relatedness is high, and otherwise they are "distant". In this analysis, we tried to measure "semantic relatedness" between the senses of an erroneous word and a corrected word.

4.2.2 Measures of lexical semantic relatedness

Many kinds of criterions to measure lexical semantic relatedness have been proposed mainly for applications in Natural Language Processing (NLP) such as word sense disambiguation, automatic detection of errors in texts, *etc*. The most popular approach in this field would be the measures based on semantic taxonomy (networks/hierarchies) such as WordNet. Table 4 is Pedersen *et al.*'s (2007) list of the major taxonomy-based measure of semantic relatedness or similarity (partly updated for Patwardham and Pedersen (2006)). Rada *et al.*'s (1989) "path length" measure is the simplest and most straight forward way. In most semantic hierarchies, the related concepts are linked by nodes. In this measure, semantic similarity between two concepts is determined by tracking the path from one node to another. The shorter the path is, the more similar they are. However, the results which relied only on path length can be biased by the variability in depth of hierarchies. The measures proposed by Wu and Palmer (1994) and Leacock and Chodorow (1998) are also based on path length, but call this problem into account by including the global or maximum depth of the hierarchy in their metrics. All of three measures explained so far rely only on

IS-A relation. Hirst and St-Onge (1998) is the only path-length-based measure which takes meronymy and other relations beyond IS-A.

Type	Name	Principle	Advantages	Disadvantages
Path Finding	Rada *et al.* (1989)	Count of edges between concepts	-Simplicity	-Requires a rich and consistent hierarchy -no multiple inheritance -WordNet nouns only -IS-A relations only
	Wu and Palmer (1994)	Path length to subsumer, scaled by subsumers path to root	-Simplicity	-WordNet nouns only -IS-A relations only
	Leacock and Chodorow (1998)	Finds the shortest path between concepts and log smoothing	-Simplicity -Corrects for depth of hierarchy	-WordNet nouns only -IS-A relations only
	Hirst and St-Onge (1998)	Relies on sysets in WordNet	-Measures relatedness of all POS -More than IS-A relations	-WordNet specific -relies on synsets and relations not available in UMLS
Information Content	Resnik (1995)	Information Content (IC) of the least common subsumer (LCS)	-Uses empirical information from corpora	-Does not use the IC of individual concepts, only that of the LCS -WordNet nouns only -IS-A relations only
	Jiang and Conrath (1997); Lin (1998)	Extensions of Resnik; scale LCS by IC of concepts	-Accounts for the IC of individual concepts, only that of the LCS	-WordNet nouns only -IS-A relations only
Gloss Vector	Patwardham and Pedersen (2006)	Combining the information of WordNet with context vectors which represent the meaning of concepts derived from co-occurrence statistics of the glosses in WordNet	-Measures relatedness of all POS -Uses empirical knowledge implicit in a corpus	-Definitions (glosses) can be short and inconsistent -Computationally intensive

Table 4: Major measures of semantic relatedness (based on Pedersen *et al.* 2007)

The measures proposed by Resnik (1995), Jiang and Cornarth (1997) and Lin (1998) are based on not only the information from ontology but also the information from a corpus to measure how two concepts share information in common, that is, word co-occurrence information in actual texts. Patwardham & Pedersen (2006) also use the empirical knowledge from a corpus, but what the "corpus" is called here is the glosses for all of the concepts in WordNet.

In this analysis, we used the measure by Leacock and Chodorow (1998) which marks the high value of the coefficient of correlation with human rating. The reason why we chose this measure is that when people try to understand the speaker's intention form his/her utterance which contains lexical errors, they would estimate a correct word which has a similar meaning to the erroneous word. Since human rating which was used for evaluation is based on semantic similarity, the measure may be close to the semantic representation of concepts in humans. For comparison, we also used the measures by Hirst and St-Onge (1998) and Patwardham and Pedersen (2006). As we stated, the measure by Hirst and St-Onge (1998) deals with not only IS-A reations but also meronymy. Language learners often use related words such as hypernyms, hyponyms, synonyms, and even meronyms when they do not know or cannot retrieve an appropriate word. This is one of the learners' important communication strategies. Patwardham and Pedersen's (2006) measure treats the word co-occurrence statistics from a corpus of glosses in WordNet. This means that the measure takes related words beyond IS-A relations and their co-occurrence patterns into account. For actual measurement, we used the freely-available software package, WordNet::Similarity (Pedersen *et al.* 2004). The version of WordNet built in this system is 2.1.

4.3 Results of comparison across the levels of intelligibility

Table 5 shows the mean values of semantic relatedness which was obtained with three measures across three levels of intelligibility. It can be seen that semantic relatedness decreases as the level of intelligibility goes down in all measures. Since a word sense is determined by the context, and the same lexical errors in different contexts can have different degree of influence to understanding of an entire utterance. Therefore, we know the analysis with a single-word basis like this is not sufficient to capture the entire picture. However, the results shown in Table 5 indicate that we can catch a glimpse of the relationship between semantic relatedness and intelligibility.

Intelligiblity / Measure	intelligible	unclear	unintelligible
Leacock and Chodorow	2.63	2.23	1.72
Hirst and St-Onge	6.27	4.34	3.62
Patwardham and Pedersen	0.54	0.43	0.38

Table 5: Mean semantic relatedness across levels of intelligibility

The results are shown from a different point of view in Figure 3, 4 and 5. These figures are the scatter plots obtained by correspondence analysis. Correspondence analysis is a descriptive/exploratory technique designed to analyze simple two-way and multi-way tables containing some measure of correspondence between the rows and columns. The results provide information which is similar in nature to those produced by factor analysis techniques, and they allow us to explore the structure of categorical variables included in the table.

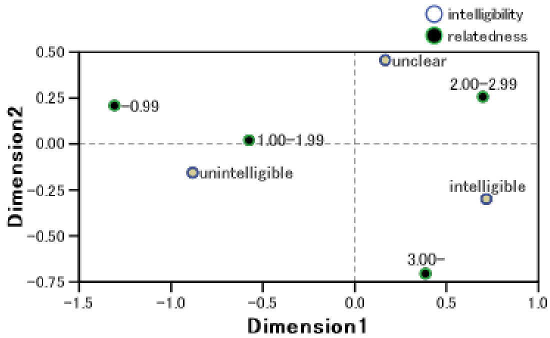

Figure 3: Relationship between semantic relatedness based on Leacock and Chodorow (1998) and levels of intelligibility

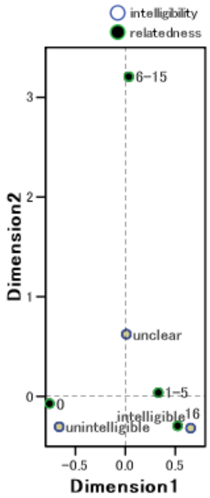

Figure 4: Relationship between semantic relatedness based on Hirst and St-Onge (1998) and levels of intelligibility

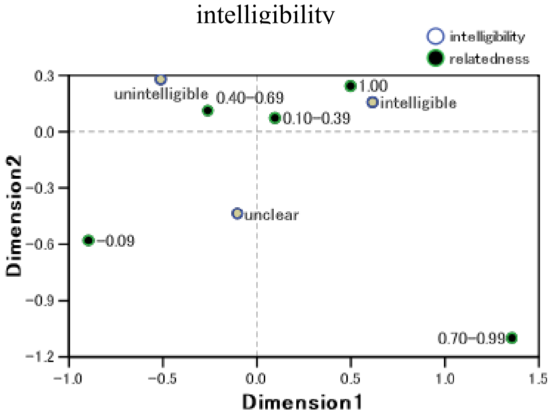

Figure 5: Relationship between semantic relatedness based on Patwardham and Pedersen (2006) and levels of intelligibiltiy

The points "○" stand for the level of intelligibility, and the points "●" stand for the values of semantic relatedness. From these figures, it is revealed that the level of intelligibility is reflected on two dimensions which marked the highest contribution rate. From the positional relations between the level of intelligibility and the variables (semantic relatedness values), we could assume that the higher the relatedness values are, the level of intelligibility goes up.

4.4 Results of comparison across proficiency levels

Table 6 shows the mean semantic relatedness which was obtained with three measures across different proficiency levels (L2-9: L9 is the most advanced). Unlike the results across three levels of intelligibility shown in Table 5, we cannot find a perfect mutual relation between the relatedness values and proficiency levels. Fluctuations can be found especially among L4, 5 and 6.

Proficiency level / Measures	L2	L3	L4	L5	L6	L7	L8	L9
Leacock and Chodorow	1.73	2.03	2.19	2.05	1.82	2.08	2.33	2.36
Hirst and St-Onge	0.18	0.43	0.31	0.25	0.28	0.34	0.40	0.37
Patwardham and Pedersen	2.00	2.75	3.25	2.47	2.79	4.34	4.36	6.14

Table 6: Mean semantic relatedness across proficiency levels

Figure 6, 7 and 8 are the scatter plots obtained by correspondence analysis. Again, unlike the results across three levels of intelligibility shown in Figure 3, 4 and 5, proficiency levels are not perfectly reflected on two dimensions, and there is less correlation between the values of semantic relatedness and proficiency levels. Hirst and St-Onge's (1998) measure reflects proficiency levels and correlation between relatedness and proficiency levels in some degree compared to other two measures although some fluctuations can still be seen among L3, 4, 5, and 6. This measure assigns all weakly-related pairs the value of zero. Because of this cut-off, the measure might fail to describe the details of the lower proficiency levels (L2-6), while it succeeded in describing the advanced learners' (L7-9) error pattern where weakly-related pairs were hardly found.

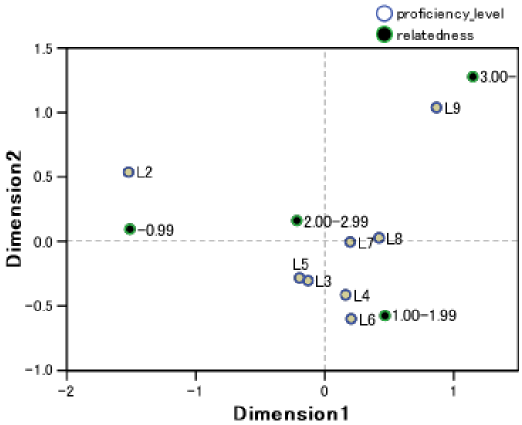

Figure 6: Relationship between semantic relatedness based on Leacock and Chodorow (1998) and proficiency levels

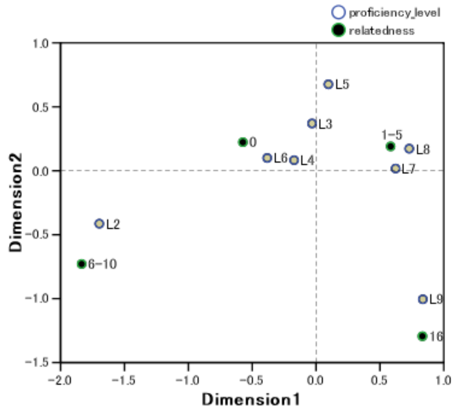

Figure 7: Relationship between semantic relatedness based on Hirst and St-Onge (1998) and proficiency levels

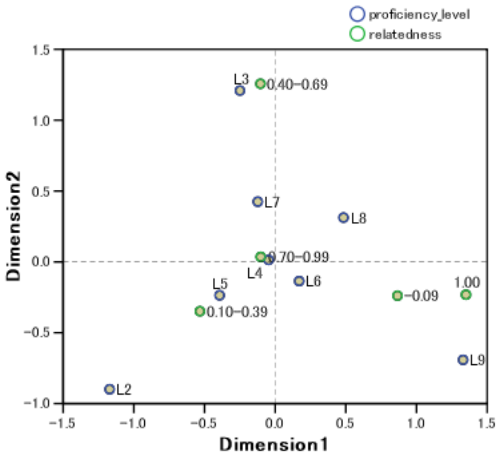

Figure 8: Relationship between semantic relatedness based on Patwardham and Pedersen (2006) and proficiency levels

The reason why fluctuation often occurs among L4, 5, and 6 might be that learners in these proficiency levels are in the period of growth in their vocabulary size. Table 7 shows the transition of the standardized Type-Token Ratio (TTR) across proficiency levels which has been extracted from the NICT JLE Corpus. After the high rates of increase can be seen from L3 to L4, from L4 to L5, from L5 to 6 and from L6 to L7, the vocabulary size remains steady in the upper levels. We assume that although learners' vocabulary size itself grows dramatically in L4, 5 and 6, and they try to use the newly-learned words, but it takes some time to use them properly.

Proficiency levels	Standardized TTR (per 200 words)	Rate of increase (percent)
L2	42.03	-
L3	42.48	1.07
L4	44.2	4.04
L5	46.22	4.57
L6	47.58	2.94
L7	49.02	3.02
L8	49	-0.04
L9	49.35	0.71

Table 7: Transition of standardized TTR across proficiency levels

4.5 Findings from individual cases

When an erroneous word and a correct word belong to the same synset, path length is "1", and in most cases, the maximum value is assigned by all three measures we used. In such case, the meaning of the utterance can be easily understood or guessed, but the correct word is more appropriate because it is more frequently and idiomatically used, or collocates with the adjacent words better (e.g. 1).

e.g. 1) *sunset **scene**/sunset **view***, *have a **dialect**/have an **accent***,
*private **matter**/private **issue***, *accident **situation**/accident **site***
*crowded **situation/**crowded **place***

Even if the word pairs mark high values of semantic relatedness, they can make the utterances awkward because their register does not appropriate to the situation where the utterances occur (e.g. 2). Although, in most cases, the sentences which include this kind of error were categorized as unclear sentences, it might be better to consider them as "unnatural" sentences.

e.g. 2) (in interview situation) **my **mom**/my **mother***

Most erroneous words in intelligible sentences are, of course, semantically-related to correct words, and they are similar in their pronunciation as well (e.g. 3).

e.g. 3) *bag/baggage, blackboard/board, hometown/town*

Errors in unclear sentences often involve the words whose meanings can change across domains and contexts (e.g. 4).

e.g. 4) (in the business situation) **have an **engagement**/have an **appointment***
(in a restaurant) **a **servant** served wine/a **waiter** served wine*
he is in his second **grade in university/he is in his second **year…***
he paid the **fee at the restaurant/he paid the **bill** at the restaurant*

We found some cases where the correct word can be associated with the erroneous word although the pair of words do not have high relatedness values or even their relatedness cannot be measured because they have different parts of speech. For example, "*cook-a-doodle*" was used for "*chicken*", and "*eat*" was used for "*food*". Although it is difficult to connect these pairs of words with the existing conceptual hierarchies, human can do it. As stated in Maera (1996), native speakers have a broad network of word association which plays an important role in communication as real-world knowledge. It is important for learners to broaden their word association network because it makes it possible to retrieve an alternative word when they cannot retrieve an appropriate word, which is one of the most effective communication strategies.

Concerning the errors in unintelligible sentences, no general findings could be obtained because of the limited amount of data. Most of them are global errors including discourse errors. In most cases, they are grammatically correct as a single sentence, but do not make sense within a context. To analyze these errors, context information is needed and the error analysis with single-word basis cannot cover them.

5. Conclusions

In this paper, we carried out the analysis on vocabulary usage in Japanese learner English mainly by focusing on the relationship between lexical semantic relatedness of an erroneous word and a correct word. In the analysis, we found some correlation between them. Although our analysis was with s ingle-word basis and dealing with them separately, there are many sentences which contain multiple errors and it is necessary to examine which of them has a major impact on changing the level of intelligibility of the sentence. This means that even if two errors are categorized as the same type, their impact can change depending on what kind of context they appear in. As future work, we will continue to find out the correlation between vocabulary usage and intelligibility in learner language not only by analyzing errors locally, but also by examining relationship between individual errors and the context such as how impact of errors can change depending on the context and how different kinds of errors in one sentence or across sentences interact each other.

References

Budanitsky, A. and G. Hirst. (2006) 'Evaluating WordNet-based measures of lexical semantic relatedness'. *Computationla Linguistics*, Volume 32, Issue 1, 13–47.

Canale, M., and M. Swain (1980) 'Theoretical bases of communicative approach to second language teaching and testing'. *Applied Linguistics*, Volume 1, 1–47.

Ellis, R. (2003) Task-based Language Learning and Teaching. Oxford: Oxford University Press.

Hirst, G. and D. St-Onge (1998) Lexical chains as representations of context for the detection and correction of malapropisms, in C. Fellbaum (ed.) (1998) WordNet: An electronic lexical database. pp. 305–321, Cambridge MA: MIT Press.

Izumi, E., K. Uchimoto, H. Isahara (2004) 'Standard Speaking Test (SST) Speech Corpus of Japanese Learners' English and automatic detection of learners' errors'. *International Computer Achieves of Modern and Medieval English (ICAME) Journal*, Volume 28, 31–48.

Jiang, J., D. Conarth (1997) 'Semantic similarity based on corpus statistics and lexical taxonomy'. in *Proceedings of the 10th International Joint Conference on Research in Computationa Linguistics*, 19–33, Taipei, Taiwan.

Kormos, J. (2006) Speech Production and Second Language Acquisition, Mahwah, NJ: Lawrence Erlbaum Associates, Inc..

Leacock, C. and M. Chodorow. (1998) Combining local context and WordNet similarity for word sense identification, in C. Fellbaum (ed.) (1998) WordNet: An electronic lexical database. pp. 265–83, Cambridge MA: MIT Press.

Lin, D. (1998) 'An information-theoretic definition of similarity', in *Proceedings of the 15th International Conference on Machine Learning*, 296–304, Madison, WI.

Maera, P. (1996) The dimensions of lexical competence, in G. Brown, K. Malmkjær, and J. Williams (eds.) (1996) Performance and Competence in Second Language Acquisition, pp. 35–53, Cambridge: Cambridge University Press.

Patwardhan, S. and T. Pedersen. (2006) 'Using WordNet-based context vectors to estimated the semantic relatedness of concepts'. in *Proceedings of the EACL 2006 Workshop, making sense of sense: Bringing computational linguistics and psycholinguistics together*, 1–8, Trento, Italy.

Pedersen, T., P. Siddharth, M. Michelizzi (2004) 'WordNet::Similarity – Measuring the reatedness of concepts'. in *Proceedings of the 5th Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-2004)*, 267–70, Boston, MA. System available on www.d.umn.edu/~tpederse/similarity.html

Pedersen, T., S. Pakhomov., S. Patwardhan. and C. Chute (2007) 'Measures of semantic similarity and relatedness in the biomedical domain'. *Journal of Biomedical Informatics*, Volume 40, Issue 3, 288–99.

Rada R., E. Bicknell., and M. Bletttner. (1989) 'Development and application of a metric on semantic nets'. in *IEEE transactions on systems, man and cybernetics*, Volume 19, Issue 1, 17–30.

Resnik, P. (1995) 'Using information content to evaluate semantic similarity'. in *Proceedings of the 14th International Conference on Artificial Intelligence*, 448–53, Montreal, Canada.

Skehan, P. (1998) A Cognitive Approach to Language Learning. Oxford: Oxford University Press.

WordNet, http://wordnet.princeton.edu

Wu, Z. and M. Palmer. (1994) 'Verb semantics and lexical selection'. in *Proceedeings of the 32nd Annual Meeting of the Association for Computational Linguistics*, 133–38, Las Cruces, New Mexico.