Automatic Classification of Language Learner Sentences into Native-Like or Non-Native-Like Based on Word Alignment Distribution

Katsunori Kotani[1], Takehiko Yoshimi[2], Takeshi Kutsumi[3] and Ichiko Sata[3]
[1]Kansai Gaidai University, [2]Ryukoku University, [3]Sharp Corporation
Japan

1. Introduction

The recent advancement of natural language processing techniques has brought about significant development in computer-assisted language learning and teaching. For instance, Lee et al. (2007) proposed a computer-based evaluation system for writing proficiency of English as a second language (ESL) learners. The basic technique employed is statistical linguistic models for machine translation (MT) evaluation (Corston-Oliver et al. 2001, Kulesza & Shieber 2004, Gamon et al. 2005). The evaluation method of Lee et al. (2007) automatically classifies English sentences produced by EFL learners either into native-like or non-native-like sentences by analyzing morphological and syntactic features. This evaluation method is intuitively correct, because sentences judged as native-like should be adequate and fluent, but, on the contrary, sentences regarded as non-native-like must involve some unnatural expressions.

The classifier of Lee et al. (2007), however, is difficult to identify what makes sentences judged as non-native-like unnatural. Since this classifier just examines the morphological and syntactic features such as dependency relation of subject-verb, we could find typical morphological and syntactic patterns in non-native-like sentences. From pedagogical viewpoint, this identification of (non-)native-likeness may help language teachers identify what problems a learner has. A classification-based evaluation method could more directly reveal a learner's problems if a classifier examines specific linguistic features concerning with a learner's problems. Then, we developed an automatic classifier that examines the existence/absence of linguistic problems in learner sentences. Among various learner problems, the first language influence is seen as a critical problem that language learners and teachers have to face (Ellis 1994). Therefore, we focused on the influence of learners' first language in the target language.

Since the influence of the first language often takes the form of word-for-word translation, our classifier examines whether learner sentences involve unnatural literal translation based on word alignment distribution. Word alignment distribution shows word-level correspondence between learner sentences and sentences conveyed the relevant meaning in a learner's first language. Word alignment technique is one of the natural language processing techniques, and word alignment distribution is available with a word aligner, e.g., GIZA++ (Och & Ney 2003).

The goal of this paper is to address a writing proficiency evaluation method for learners of Japanese as a second language (JSL) whose first language is English. As we will see below, there are various linguistic differences between Japanese and English. The differences can be seen in lexical, syntactic and discourse levels. The validity of our classifier will be examined. It will be shown that our system can correctly classify approximately 80% of JSL learner sentences.

We will further examine the adequacy of our classifier. First, we will examine the adequacy of our word-alignment-based classifier by comparing our classifier with a classifier using syntactic features. The result showed that our classifier achieved higher classification accuracy than classifiers based on parsing information features. Secondly, we will investigate whether the classification results reflect the learner proficiency. The proper classification results should exhibit the decrease of the classification accuracy for sentences produced by learners with higher writing proficiency, because these learners can write native-like sentences. Thus, the classification task becomes more difficult for sentences written by proficient learners. The experiment result showed that the classification results exhibited statistic significant difference ($p<.05$) between learners who marked more than 50 points of human evaluation scores and learners with less than 50 points. Given these findings, we concluded that word alignment-based classification techniques can be used for evaluating the writing proficiency of foreign language learners.

2. Related Studies

In this section, we review related studies on (i) classification-based MT evaluation (Corston-Oliver et al. 2001, Kulesza & Shieber 2004, Gamon et al. 2005, Paul et al. 2007), (ii) word alignment-based MT evaluation (Blatz et al. 2004, Lin & Gidea 2007) (iii) classification-based evaluation of foreign language learner sentences (Lee et al. 2007) and (iv) Computer-assisted language assessment (Chapelle 2008).

First, let us review studies on classification-based MT evaluation (Corston-Oliver et al. 2001, Kulesza & Shieber 2004, Gamon et al. 2005, Paul et al. 2007). These studies constructed classifiers for MT that distinguish between MT-like sentences and human translation (HT)-like sentences, assuming that good MT sentences should be similar to HT sentences. This idea is intuitively understandable, as poor MT sentences can be easily distinguished from HT sentences, but good MT sentences are often mistaken for HT sentences. Hence, these studies treated evaluation of quality of MT sentences as a classification problem.

Corston-Oliver et al. (2001) used decision trees (Quinlan 1992) with both perplexity and linguistic features of Spanish-to-English MT sentences and HT sentences. A perplexity-based classifier yielded an accuracy of 74.7%, a linguistic feature-based classifier showed an accuracy of 76.5%, and a classifier using both features achieved the best classification accuracy of 82.9%.

Kulesza & Shieber (2004) constructed a classifier for Chinese-to-English MT sentences using well-known machine learning algorithms, Support Vector Machines (SVMs) (Vapnik 1998). This classifier examined the following classification features: (i) n-gram precision of MT sentences compared with HT sentences, the length of MT sentences and HT sentences, and (iii) the word error rate of MT sentences. Their classifier yielded an accuracy of 64.4%.

Gamon et al. (2005) developed an SVM classifier based on linguistic features. Classification features included subcategorization properties and semantic properties such as finiteness or argument structures. This classifier showed a classification accuracy of 77.6% for English-to-French MT sentences.

Paul et al. (2007) developed a classifier with decision tree algorithms employing evaluation scores obtained with other automatic MT evaluation metrics including BLEU (Papineni et al. 2001), NIST (Doddington 2002) and METEOR (Banerjee & Lavie 2005). Unlike the other classifiers (Corston-Oliver et al. 2001, Kulesza & Shieber 2004, Gamon et al. 2005), this method requires a lot of manual evaluation results for MT evaluation scores.

By contrast, our classifier as well as the other classifiers (Corston-Oliver et al. 2001, Kulesza & Shieber 2004, Gamon et al. 2005) needs parallel corpora, which are more easily obtained than manual MT evaluation results. As we mentioned above, our classifier could identify learner problems more directly than classifiers based on the general linguistic features (Corston-Oliver et al. 2001, Kulesza & Shieber 2004, Gamon et al. 2005, Paul et al. 2007)

Secondly, we will review research on word alignment MT evaluation (Blatz et al. 2004, Lin & Gildea 2007), which is similar to our classifier. As we will see below, our classifier employs word alignment distribution between learner sentences and sentences written in a learner's first language. Our classifier uses alignment features differently from the way these previous methods employed.

Blatz et al. (2004) used word alignment results for evaluating MT sentences. Under their method, contiguities of aligned words are taken as classification features. Liu & Gildea (2007) also constructed a classifier that employs source sentence-related features. Under this approach, alignment features were used to identify overlapping words for counting in their metric: words were counted only if the words were aligned to corresponding words in source sentences. By contrast, our approach uses aligned pairs and non-aligned words directly as classification features, as explained in the following section.

Thirdly, we will review research on classification-based evaluation of learner sentences. Lee et al. (2007) constructed a classifier for sentences written by EFL learners. Interestingly, this study employed MT sentences for training a classifier instead of EFL learner sentences, because MT sentences are more cheaply available than learner sentences. Lee et al. (2007) showed that MT sentences can be used as an alternative language data to learner sentences. As an experiment result showed, a classifier trained with EFL learner sentences yielded a classification accuracy of 66.4%, while a classifier based on MT sentences achieved the similar classification accuracy of 59.0%. From this experiment result, we determined to employ MT sentences in training a classifier for sentences written by JSL learners, as explained in Section 4.1.

Last, we will introduce research on the advantage of computer-based evaluation for language learners. Chapelle (2008) pointed out major contributions of computer-based evaluation methods. According to Chappelle (2008), one of the advantages of computer-based evaluation methods can be seen in that it is a computer-adaptive test that can assess learner proficiency more effectively than non-adaptive tests can. Given this advantage, we will modify our classifier so that it assesses learner sentences more adoptively.

3. Classification Features

Good, natural sentences should differ from word-for-word translation, whereas unnatural learner sentences often causes from word-for-word translation. For instance, an English nominal modifier "some" can convey the existential meaning of an entity as in Sentene (1a). A literal translation of this nominal modifier makes a sentence unnatural as seen in a JSL learner's sentence (1b). In Japanese, the existential meaning of an entity is often expressed using an existential verb "i-ta (exisited)" as in Sentence (1c). Sentence (1c) is clearly natural, while Sentence (1b) is perfectly grammatical but less natural. The unnaturalness of Sentence (1b) is due to the word-for-word translation of the English nomial modifier "some" to Japanese nominal modifier "ikuraka-no (some)."

(1) a. Some students came.
b. Ikuraka-no gakusei-wa ki-ta
some-GEN student-TOP come-PST
'Several students came.'
c. Ki-ta gakusei-mo i-ta
come-PST student-also exist-PST
'Some students came.'

(GEN: genitive Case marker, PST: past tense marker, TOP: topic marker)

As Example (1) illustrates, the influence of the first language can be seen in the literal translation. In order to identify words derived from word-for-word translation, we decided to examine word alignment distribution between JSL learner sentences and sentences conveying inteded meaning in a learner's first language, i.e., English. This is because literally translated words are more easily aligned than non-literally translated words. Literal translation maintains lexical features such as part of speech, but, on the other hand, non-literal translation usually lacks parallel lexical features, as seen in Example (1).

Let us consider another example of unnatural word-for-word translation. Sentence (2a) is literally trasnlated into Sentence (2b). Sentence (2b) is grammatical but unnatural as a Japanese sentence. Though Sentence (2c) is not literal translation, this sentence is perfectly natural.

(2) a. Today the sun is shining. (original sentence)
b. Kyoo taiyoo-wa kagayai-teiru (MT)
today the-sun-TOP shine-BE-ING
'Today the sun is shining.'
c. Kyoo-wa seiten-da (HT)
today-TOP fine-BE
'It's fine today.'

(BE: copular verb, ING: gerundive verb form)

Here, we examined whether word alignment distribution differs between a literally traslated sentence and a non-literally translated sentence. The word alignment distribution of Sentence (2b) and Sentence (2c) is automatically derived with our experimental word aligner, described in Section 4.

Table 1 shows the word alignment distribution of Sentence (2b) and Sentence (2c). Note that "align(A, B)" means that an English word "A" and a Japanese word "B" compose an aligned pair, "non-align_eng(C)" means that an English word "C" remains unaligned, and "non-align_jpn(D)" means that a Japanese word "D" remains unaligned. As shown in Table 1, the rate of alignment and non-alignment varies based on whether or not a sentence is literally translated. That is, more aligned words are observed in Sentence (2b), i.e., a non-native-like sentence, and more non-aligned words appear in Snentece (2c), i.e., a native-like sentence. Thus, non-aligned words should create a sense of naturalness, while aligned words would make a sentence unnatural.

Classification-based writing evaluation should not only classify learner sentences into native-like or non-native-like sentences, but also identify linguistic problems that arise from the influence of a learner's first language in the form of literal translation. The difference between a non-native-like sentence (2b) and a native-like sentence (2c) can be drwan with the alignment distribution of these sentences as shown in Table 1. The English words "today" and "is" are aligned with the relevant Japanese words in both sentences. By contrast, the other words exhibit the different alignment distribution. Whilie the words "sun" and "shining" are aligned with Japanese words in Snetnece (2b), these words are remained unalinged in Sentence (2c). From this result, examining alignment distribution could reveal linguistic problems in learner sentences arising from literal translation.

Unnatural sentence (2b)	Natural sentence (2c)
align(today, kyoo [today])	align(today, kyoo-wa [today-TOP])
align(is, teiru [BE-ING])	align(is, da [BE])
align(sun, taiyoo [sun])	non-align_jpn(seiten [fine])
align(shining, kagayai [shine])	non-align_eng(the)
non-align_jpn(wa [TOP])	non-align_eng(sun)
non-align_eng(the)	non-align_eng(shining)

Table 1. Alignment Distribution of Example (2)

4. Experiment

4.1 Designs

The goal of this experiment is to validate our classification-based evaluation method for JSL learners. In this experiment, we constructed the following two types of classifiers using word alignment distribution as classification features: (i) a classifier based on non-aligned words and (ii) a classifier based on both aligned and non-aligned words.

Classifiers were constructed with SVMs that have high generalization performance (Vapnik 1998). SVMs were carried out with TinySVM (a software implemented as a packaging tool available at the following URL: http://chasen. org/~taku/software/TinySVM/). The first order polynomial was taken as the type of kernel function, and the other settings were taken as the default settings.

The classifiers were trained with a parallel corpus of HT and MT instead of a learner corpus. Although various learner corpora have been distributed (The National Institute for Japanese Language 2001, Izumi et al. 2004), the amount of data they contain is not as large as first language corpus. In addition, what we need for training a classifier based word alignment distribution is a parallel corpus. The data size of parallel learner corpus further decreases. Then, we decided to employ MT data as learner data, because MT sentences are similar to learner sentences, in that, both sentences involve some linguistic problems. This alternative use of MT data as learner data was investigated in Lee et al. (2007), and it is suggested that MT data can be used as an alternative data to learner data.

Training data consisted of Reuters' articles in English and expert Japanese translations of the original articles, i.e., HT, (Utiyama & Isahara 2003). Duplicates of sentences and translations were deleted if they appeared repeatedly in the corpus. MT data was obtained by operating machine translation systems over this corpus. The translation systems are state-of-art systems commercially supplied in Japan (MT-α and MT-β). A total of 25,800 example sentences were obtained (12,900 HT sentences and 12,900 MT sentences).

Word alignment distributions were provided with an experimental word aligner between the source English sentences and Japanese translation sentences (MT and HT). This tool segments Japanese sentences into word-units before aligning English and Japanese words. Word alignment process is performed using a bilingual dictionary/thesaurus and a dependency analyser. The alignment results consisted of aligned pairs and non-aligned words. Each alignment instance was taken as a classification feature.

Our classifiers were evaluated with a learner corpus. Test data was taken from a JSL corpus (The National Institute for Japanese Language 2001). This corpus consists of essays written by JSL learners and the corresponding English sentences. Classification accuracy of our classifiers was examined with 689 sentences. Then, the validity of our classifiers was further investigated by examining the relation between human evaluation results and classification accuracy. This examination was carried out with 279 sentences out of the 689 learner sentences. The fluency of the 279 sentences was evaluated by three native Japanese speakers. This evaluation was carried in 100-point scale. Then, the test sentences were divided into higher and lower at the 50-point. The higher group consists of 63 sentences (22.6%), and the lower group is made up of 216 sentences (77.4%).

In this experiment, we first examined the robustness of our classifier against MT of training data by comparing the classification accuracies of classifiers trained with MT-α and MT-β. Secondly, we examined our classifiers with JSL learner data. Classifiers tested are constructed with either aligned features or both non-aligned & aligned features. Thirdly, we compared our word-alignment-based classifiers with classifiers based on syntactic features. Last, we investigated whether the classification results can reflect the learner proficiency.

4.2 Results and Discussion

Before reporting the experiment results, we shall briefly report the word alignment distribution in MT and HT data. Table 2 shows the distribution of aligned pairs and non-aligned words in the training data. The alignment rate of MT data was higher than that of HT data, as more aligned pairs appear in MT data. The alignment rates between HT (the control sample) and both MT data were analyzed with a one-way ANOVA test and Tukey's test, and a statistic significant difference was observed between HT and MTs. ($F(3, 56)=616.10$, $p<.01$). Therefore, it is evident word alignment distribution differs between HT and MT.

	N	Aligned pairs (%)	Non-aligned words (%)	Alignment rate (%)
MT-α	518894	37.1	62.9	59.0
MT-β	537460	36.4	63.6	57.3
HT	568259	24.1	75.9	31.7

Table 2. Alignment Distributions of MT & HT

Hereafter, we will report the experiment results. First, we examined the robustness of our classifier against MT of training data by comparing the classification accuracies of classifiers trained with MT-α and MT-β. We evaluated the classifiers in a five-fold cross validation test. Table 3 shows the mean classification accuracy in the five trials. As both the classifiers marked rather high classification accuracy, it is suggested that our classifier is robust against machine translation systems.

	Mean classification accuracy (%)
MT-α-based classifier	99.7
MT-β-based classifier	99.8

Table 3. Mean Classification Accuracy

Secondly, we examined our classifiers using JSL learner sentences as test data. Classifiers tested are constructed with either aligned features or both non-aligned & aligned features. Table 4 shows the classification accuracy of the two types of classifiers. Any type of classifiers marked high classification accuracy (more than 70%). The highest classification accuracy (approximately 80%) was yielded by non-aligned-based classifier trained with MT-α. Given this classification accuracy, our classifier is tenable for classifying learner sentences.

	MT-α (%)	MT-β (%)
Non-aligned-based classifier	79.1	71.8
Aligned & non-aligned-based classifier	72.5	77.6
Mean classification accuracy	75.8	74.7

Table 4. Classification Accuracy

Thirdly, we compared our word-alignment-based classifier with a classifier using syntactic features. Lee et al. (2007) employed parsing results such as (i) context-free grammar rules used for parsing sentences, (ii) parsing scores, and (iii) co-occurrence relations between a verb and its subject/object noun. Then, we constructed a classifier using Japanese syntactic parsing information extracted with a dependency parser Cabocha. We regarded phrase dependency relation as classification features. The dependency relation shows part-of-speech information of a modifier and a modifiee. Although this syntactic dependency relation does not precisely reproduce the experimental conditions of Lee et al. (2007), we consider that this comparison would suggest the validity of our method because these features correspond to the classification features (i) and (iii) used by Lee et al. (2007).
Table 5 shows the classification accuracy of parsing information-based classifiers for 279 learner sentences. The classification accuracy of our classifiers (shown in Table 4) is much higher than the accuracy of the parsing information-based classifiers. From this result, it is

suggested that learner sentences can be identified more properly by word-alignment-based classifiers than the parsing information-based classifiers.

	Classification accuracy (%)
Classifier trained by MT-α	43.7
Classifier trained by MT-β	37.6

Table 5. Classification Accuracy of Parsing Information-based Classifier

Last, we investigated whether the classification results can reflect the learner proficiency. If the classifiers are tenable, the classification accuracy should decrease for sentences written by learners with high writing proficiency. This is because a proficient learner can write sentences similar to sentences written by native speakers.

In this experiment, the classification accuracy was compared between sentences written by less proficient learners and the ones written by more proficient learners. The learner proficiency was determined by human evaluation score (more/less than 50 points).

Table 6-9 show the classification accuracy of non-aligned-based classifiers and that of aligned & non-aligned-based classifiers trained with MT-α and MT-β. The classification accuracy decreases for more proficient learner sentences, except for the accuracy of a non-aligned-based classifier trained with MT-β. The difference of classification accuracy was statistically analyzed by the chi-square test, and the statistically significant decrease ($p<.05$) was observed for the classification results of an aligned & non-aligned-based classifier trained with MT-α. (Table 7) Therefore, it is clear that our classifier can account for learner proficiency. This is another piece of supporting evidence that our classifier is tenable as a writing evaluation method.

Learner sentence	Correct	Incorrect	Classification accuracy (%)
Less proficient learner sentence	58	5	92.1
More proficient learner sentence	198	18	91.7

Table 6. Human Assessment and Non-aligned-based Classification Accuracy (MT-α)

Learner sentence	Correct	Incorrect	Classification accuracy (%)
Less proficient learner sentence	61	2	96.8
More proficient learner sentence	188	28	87.0

Table 7. Human Assessment and Aligned & Non-aligned-based Classification Accuracy (MT-α)

Learner sentence	Correct	Incorrect	Classification accuracy (%)
Less proficient learner sentence	56	7	88.9
More learner sentence	198	18	91.7

Table 8. Human Assessment and Non-aligned-based Classification Accuracy (MT-β)

Learner sentence	Correct	Incorrect	Classification accuracy (%)
Less proficient learner sentence	58	5	92.1
More proficient learner sentence	189	27	87.5

Table 9. Human Assessment and Aligned & Non-aligned-based Classification Accuracy (MT-β)

5. Conclusion

In this paper we constructed classifiers for JSL learner sentences based on word alignment distribution for evaluating writing proficiency from viewpoint of the first language influence. Our classifiers were trained with MT data, and evaluated with JSL learner sentences. Though the classification results vary depending on word alignment features, any type of classifiers marked more than 70% classification accuracy. The highest classification accuracy (approximately 80%) was yielded with a classifier based on non-alignment feature trained by MT-α.

We further examined the adequacy of our word-alignment-based classifier by comparing our classifiers with classifiers using syntactic features. Our classifiers achieved higher classification accuracy than parsing information-based classifiers.

In addition, we evaluated our classifier by examining whether classification results reflect learner proficiency. Our classifiers were tested with 279 JSL learner sentences that were evaluated as more/less than 50 points in human evaluation scores. The classification accuracy showed the statistic significant decrease ($p<.05$) for more proficient learner sentences.

From these results we conclude that a word alignment-based classifier is tenable as an evaluation method for JSL learner writing proficiency.

Even though the experiment results demonstrated the validity of word alignment-based classifiers, there are remaining problems to be solved. First, we have to examine to what extent MT and learner sentences are similar. In this paper, the similarity of MT and learner sentences is assumed based on research of the previous study (Lee et al. 2007). Second, we will compare classification-based evaluation results not only in two classes (more proficient learner sentence & less proficient learner sentence), but also in more classes, e.g., 5 classes for examining the distinctiveness of our classifier. Thirdly, we will examine to what extent our classification-based evaluation method can reveal the first language influence. Last, our classification-based evaluation method has to be implemented on a computer language learning system.

6. References

Banerjee, S. and A. Lavie. 2005. METEOR: An automatic metric for MT evaluation with improved correlation with human judgement. In Proceedings of the 43th Annual Meeting of the Association of Computational Linguistics (ACL) Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization. 65-72.

Blatz, J., E. Fitzgerald, G. Foster, S. Gandrabur, C. Goutte, A. Kulesza, A. Sanchis, and N. Ueffing. 2004. Confidence estimation for machine translation. Final Report, JHU /CLSP Summer Workshop.

Chapelle, C. A. 2008. Utilizing technology in language assessment. In Encyclopedia of Language and Education: Language Testing and Assessment Shohamy, E. and N. H. Hornberger (eds.) Heidelberg: Springer-Verlag.

Corston-Oliver, S., M. Gamon, and C. Brockett. 2001. A machine learning approach to the automatic evaluation of machine translation. In Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL01). 148-155.

Doddington, G. 2002. Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In Proceedings of the second international conference on Human Language Technology Research (HLT). 138-145.

Ellis, R. 1994. The Study of Second Language Acquisition. Oxford University Press, Oxford.

Gamon, M., A. Aue and M. Smets. 2005. Sentence-level MT evaluation without reference translations: Beyond language modelling. In Proceedings of the 10th the European Association for Machine Translation Conference (EAMT05). 103-111.

Izumi, E., K. Uchimoto, and H. Isahara. 2004. The overview of the SST speech corpus of Japanese learner English and evaluation through the experiment on automatic detection of learners' errors. In Proceedings of the International Conference on Language Resources and Evaluation (LREC-04), 1435-1439.

Kulesza, A. and S. M. Shieber. 2004. A learning approach to improving sentence-level MT evaluation. In Proceedings of the 10th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI04). 75-84.

Lee, J., M. Zhou, and X. Liu. 2007. Detection of non-native sentences using machine-translated training data. In Proceedings of the 2007 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT07). 93-96.

Liu, D. and D. Gildea. 2007. Source-language features and maximum correlation training for machine translation evaluation. In Proceedings of the 2007 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT07). 41-48.

Och, F. J. and H. Ney. 2003. A systematic comparison of various statistical alignment models. Computational Linguistics 29(1): 19-51.

Papineni, K. A., S. Roukos, T. Ward, and W.-J. Zhu. 2001. Bleu: A method for automatic evaluation of machine translation. Technical Report RC22176 (W0109-022), IBM Research Division, Thomas J. Watson Research Center.

Paul, M., A. Finch, and E. Sumita. 2007. Reducing human assessment of machine translation quality to binary classifiers. In Proceedings of the 11th International Conference on Theoretical and Methodological Issues in Machine Translation (TMI 2007). 154-162.

Quinlan, J. 1992. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA.

The National Institute for Japanese Language. 2001. Contrastive Linguistic Database for Japanese Language Learners' Written Language.

Utiyama, M. and H. Isahara. 2003. Reliable measures for aligning Japanese-English news articles and sentences. In Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL). 72-79.

Vapnik, V. 1998. Statistical Learning Theory. Wiley-Interscience, New York.

Advances in Technology, Education and Development
Edited by Wim Kouwenhoven

ISBN 978-953-307-011-7
Hard cover, 474 pages
Publisher InTech
Published online 01, October, 2009
Published in print edition October, 2009

From 3rd to 5th March 2008 the International Association of Technology, Education and Development organised its International Technology, Education and Development Conference in Valencia, Spain. Over a hundred papers were presented by participants from a great variety of countries. Summarising, this book provides a kaleidoscopic view of work that is done, all over the world in (higher) education, characterised by the key words 'Education" and 'Development'. I wish the reader an enlightening experience.