

# “Linguistic Distance” and Its Influence on TESOL (Teaching English to Speakers of Other Languages)

Masako Nishikawa-Van Eester

## I. Introduction

Since the last century, English teachers in Japan have always faced enormous difficulties with the poorly performing students in spite of all the measurements and policy settings by the Ministry of Education, Culture, Sports, Science and Technology (MEXT, hereafter). Gottlieb (2005, pp. 67 - 68) mentioned about this situation by describing the fact that Japan recorded itself as the 180th among the 189 countries in the TOEFL ranking of the year 1998. Further she mentioned that 1) English is one of the major compulsory subjects learned over years, and 2) teaching English as a foreign language is not limited just to the schools and universities but it is also a major private industry.

Why are the Japanese such poor learners of English then? A number of aspects we do (or do not) experience in the process and context of language learning have been discussed as the factors causing this lamentable phenomenon in Japan. One of the major issues that is often brought up is “Linguistic Distance.” As known generally, English is considered to belong to the group of Germanic languages such as German and Dutch, whilst Japanese is considered to be a totally different language, not yet completely identified to be in any language family. A number of English learners/teachers in Japan accredit the Linguistic Distance between these two languages to be the major obstacle when they account for the hardship of learning English and so does the rest of the world. Many established researchers and educators such as Hughes (1999) have referred to this point.

Despite all those accusations to the Linguistic Distance, Yoshida (2009) - on the other hand - showed his skepticism on the Linguistic Distance issue as the major obstacle when the Japanese learn English, discussing that there might be something even more essential than that. In order to have a proper discussion on this topic, we need to raise the question: what actually is “Linguistic Distance”? Even the name is not really clear. I will use “Linguistic Distance” in this article, but it is also referred to as “Language Distance.”

In this article, I raise two main research questions focusing on the literature related to this issue. Here are the two research questions:

Research Question 1: What is the definition of “Linguistic Distance”?

Research Question 2: Is there any measurement-system that determines the concept known as “Linguistic Distance” or/and “Language Distance”?

In order to answer those two questions, I will explore four articles, “Linguistic Distance,”

“Disadvantages of Linguistic Origin: Evidence from Immigrant Literacy Scores,” “Linguistic Distance: A Quantitative Measure of the Distance between English and Other Languages,” and “Learning Complex Features: A Morphological Account of L2 Learnability.”

## II. Distances between languages

According to “Longman’s Dictionary of Language Teaching & Applied Linguistics” (Richards & Schmidt), “Language Distance” is defined as below:

The relative degree of similarity between two languages. Some languages have similar linguistic features and are said to be “close”. Others have very different linguistic features and are said to be “distant”. For example, two languages may have similar word order rules and similar rules for certain syntactic or phonological structures. There is said to be a greater degree of Linguistic Distance between English and French, for example, than between French and Spanish. Language Distance is thought to be one factor which influences the ease or difficulty with which learners acquire new languages. (p. 288)

According to Elms (2008), the relations of several language groups from the lexical point of view can be described as in Figure 1 below.

This chart indicates the degree of overall vocabulary divergence. The size of each circle represents the number of speakers for that language. The circles of the same color belong to the same language group. All the groups except for Finno-Ugric are members of the Indo-European language family. Although we see here only languages in Europe, this enables us to realize the distances between languages/language groups visually.

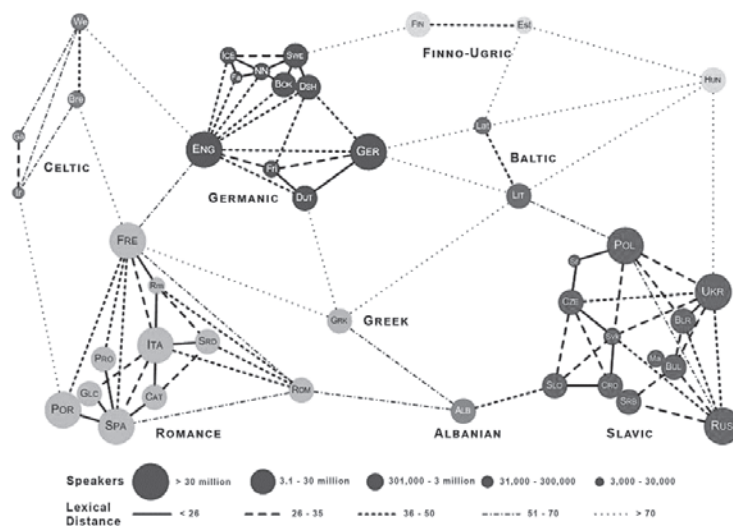


Figure 1: Lexical distance among the language of Europe (posted by Elms in 2008)

The Language Distance is considered to be strongly related to the difficulty the learners face when learning a new language. The farther the distance between the target language and the learner’s mother tongue is, the more difficult it becomes for the learner to acquire the target language. In other words, the degree of the distance could be a robust index for estimating the needed time for a learner to acquire a new foreign language.

The Foreign Service Institute that belongs to the United States Department of State is the Federal Government’s primary training institution for officers and support personnel of the U.S. foreign affairs community, preparing American diplomats and other professionals to advance U.S. foreign affairs interests overseas and in Washington. FSI created a list to indicate the approximate time a well-educated American would need to learn a specific language as an English native speaker (1973). The table below is an excerpted list of the languages based on their categorization showing the difficulty to learn foreign languages, indicating the length of time that the learner would need. This is called “the ACTFL (American Council of Teachers of Foreign Languages) scale.”

Table 1: How long will it take to learn LANGUAGE X ?

## How Long Does it Take?

**Group I Languages:** *Afrikaans, Danish, Dutch, French, Haitian Creole, Italian, Norwegian, Portuguese, Romanian, Spanish, Swahili, Swedish*

---- Aptitude for Language Learning ----			
Length of Training	Minimal Aptitude	Average Aptitude	Superior Aptitude
8 weeks (240 hours)	Intermediate Low	Intermediate Mid	Intermediate High
16 weeks (480 hours)	Intermediate High	Advanced Low	Advanced Mid
24 weeks (720 hours)	Advanced Mid	Advanced High	Superior

**Group II Languages:** *Bulgarian, Dari, Farsi, German, Greek, Hindi, Indonesian, Malay, Urdu*

---- Aptitude for Language Learning ----			
Length of Training	Minimal Aptitude	Average Aptitude	Superior Aptitude
16 weeks (480 hours)	Intermediate Low	Intermediate Mid High	Intermediate High
24 weeks (720 hours)	Intermediate High	Advanced Low Mid	Advanced Mid High
44 weeks (1320 hours)	Advanced Mid High	Advanced High Superior	Superior

**Group III Languages:** *Amharic, Bengali, Burmese, Czech, Finnish, Hebrew, Hungarian, Kinner, Lao, Nepali, Filipino, Polish, Russian, Serbo-Croatian, Sinhali, Tai, Tamil, Turkish, Vietnamese*

---- Aptitude for Language Learning ----			
Length of Training	Minimal Aptitude	Average Aptitude	Superior Aptitude
16 weeks (480 hours)	Novice High	Intermediate Low Mid	Intermediate Mid High
24 weeks (720 hours)	Intermediate High	Advanced Low	Advanced Mid High
44 weeks (1320 hours)	Advanced Mid High	Advanced High Superior	Superior

**Group IV Languages:** *Arabic, Chinese, Japanese, Korean*

---- Aptitude for Language Learning ----			
Length of Training	Minimal Aptitude	Average Aptitude	Superior Aptitude
16 weeks (480 hours)	Novice High	Intermediate Low	Intermediate Low Mid
24 weeks (720 hours)	Intermediate Low Mid	Intermediate Mid High	Intermediate High
44 weeks (1320 hours)	Intermediate High	Advanced Low	Advanced Mid High
80-92 weeks (2400-2760 hours)	Advanced High	Superior	Superior

The concept of “the Language Distance” is, thus, used as a measuring tool for estimating the length of time for a learner to reach a certain level in acquiring a specific foreign language.

### III. Article 1: Linguistic Distances

by Nerbonne, J. & Hinrichs, E. (2006)

Stating that the notion of Linguistic Distance is frequently used in many theoretical and applied areas of computational linguistics, Nerbonne and Hinrichs further described that it has received little focused attention in reality and they made an attempt to explain some common issues while categorizing it in three major approaches, “pronunciation,” “syntax” and “semantics.” Also, in this study, the notion “Linguistic Distance” is defined as “Linguistic Similarity.” In other words, these two terminologies are compatible in their article.

In order to estimate the Linguistic Distance, or Linguistic Similarity, various measures are employed to demonstrate the learner's ability in the targeted language, such as consistency, validity, stability, etc. Besides those abstract properties, furthermore, it is important for researchers to be able to exchange information on how to analyze distance information for further investigation.

#### Phonetics

In the area of phonetics, Nerbonne and Hinrichs first explained, by citing John Laver, the author of the most used textbook in phonetics, “Principles of Phonetics” (1994), that phonetic similarity is one of the least discussed issues despite the fact that it is one of the most basic concepts. Laver went on to sketch the research on phonetic similarity, or phonetic distance, in particular the confusability in which the likelihood is indicated “with which people get confused to tell one sound from another.” However, the confusability matrices are often asymmetric, which suggests that, although the notion of confusability is considered to be a reasonable reflection of phonetic similarity, there is something more significant and complex at play, which is still unknown. Psycholinguists have been paying a great deal of attention to the problem of word recognition related to the notion of the degree of phonetic similarity among the words, but, in order to clarify the idea of phonetic distances, there has risen a strong need to establish more sophisticated computational models of pronunciation distances in the future.

#### Syntax

Scholars are even less interested in syntactical theories. The authors find only a very limited number of studies of this problem although it could play an important role in syntactic typology. Additionally, there are two interdisciplinary linguistic studies in which similarity and/or distance plays a great role. One is “language contact” which seeks to identify the elements of one language adopted in the other, in a situation where two or more languages are used in the same community. As part of this field, the studies about “syntactic contamination” are of significance as well from the viewpoint of language similarity. The other to remark is

“second-language learning” in which syntactic patterns of the dominant (often, first) language are imposed on a second. The language similarity’s measure is used, furthermore, in memory-based learning strategies although the system is complex, for example, in the selection of the target structures to compare.

### Semantics

Concerning the study of “lexical similarity,” more attempts have been done to identify regularities of and systematic relations among word meanings than compare natural language semantics. With the recent increase of availability of large electronic corpora, there has been more research on capturing the notion of “context similarity” in computational linguistics. There is another approach to lexical semantics developed by linguists and cognitive psychologists, which relies on the intuition of lexicographers for capturing word meanings. The authors value two semantic resources: the Princeton WordNet (Fellbaum, 1998) and the Berkeley Framenet (Baker et al., 1998). Both were developed originally for the English language, but are generalized to other languages.

### Conclusion

It was reviewed how the notion of “Linguistic Distance,” or “Linguistic Similarity,” are employed in the three major areas of phonetics, syntax and semantics. There are diverse strategic patterns and models developed for analyzing the relation between/among languages based on a number of theories and philosophies. However, there has been very little research conducted about this issue and its own theoretical background whilst many researchers tend to use the notion of “Linguistic Distance.” Nerbonne and Hinrichs successfully introduced readers to the diversity of studies in this field, but at the same time, they state their assumption that there is always a “hidden variable,” which remains unclear.

## IV. Article 2: Disadvantages of Linguistic Origin: Evidence from Immigrant Literacy Scores by Isphording, I. E. (2013)

### Study

Based on a research program by the Study of Labor (IZA) in Bonn and together with the University of Bonn, Isphording (the University of Bochum) published a paper on the literacy test in German for immigrants (2013). The major argument in this study is that the disadvantages immigrants face in the German speaking community arise from the Linguistic Distance between the learner’s mother tongue and the host country’s language.

According to Isphording, the study utilizes data from the public use file of the International Adult Literacy Survey (IALS), which presents a data source on adult’s literacy skills and socio-economic characteristics over the period of 1994 to 1998. The advantage of this data set is

that it originates from the direct measurement of individual scores. Three different dimensions are assessed independently in this test: prose literacy such as the capability to understand texts, document literacy such as skills to use the information contained in the texts, and quantitative literacy such as the skills to locate numbers found in printed materials and to apply them to simple arithmetic operations.

To identify linguistic barriers in the formation of literacy skills, the scores are regressed on a measure of Linguistic Distance between mother tongue and host country language. The measurement of Linguistic Distance is from the Automatic Similarity Judgment Program developed by the German Max Planck Institute of Evolutionary Anthropology to explain geographical distribution and historical development of languages.

Analyzing the utilized data resulted in two interesting and somewhat expected remarks. First, it indicated that the literacy acquisition in the target language is crucially influenced by the linguistic origin of an immigrant, which means that it is more difficult for immigrants with linguistically distant backgrounds to reach a sufficient level of a command of a targeted language. The other is that the Critical Period Hypothesis got theoretically confirmed, which means that there is a significant difference, in their achievement in the literacy test, between immigrants arriving in a host country before the age of twelve or later. The table below shows linguistics distances among diverse languages.

## Conclusion

Ispording analyzed and reported that significant differences were found in literacy scores among immigrants, suggesting that linguistically distant immigrants tend to face a strong disadvantage in literacy of a target language. Moreover, late arrivers (after the age of twelve) in

Table 2: Matrix of Linguistic Distance – excerpted from the study of Ispording (2013).

Test lang.	Dutch	French	English	German	Finnish	Hungarian
Dutch	0.00	91.06	63.22	51.50	99.00	99.16
French	91.06	0.00	91.02	95.87	98.08	100.65
English	63.22	91.02	0.00	72.21	102.27	95.22
German	51.50	95.87	72.21	0.00	96.31	98.43
Finnish	99.00	98.08	102.27	96.31	0.00	84.53
Arabic	100	97.20	97.95	98.96	98.15	98.68
Greek	96.02	95.08	97.15	97.25	100.20	96.76
Japanese	101.92	101.94	99.39	100.14	96.98	99.16
Korean	99.04	102.74	99.12	104.30	100.18	100.92
Turkish	102.33	98.12	101.04	99.91	96.70	94.55
Vietnamese	100.81	101.81	104.06	96.14	97.80	98.86

Source: IALS' own calculations using programs for calculating ASJP distance matrices (Version 2.1)

a host country have enormous hardship to acquire the reading ability of a target language. The linguistic origin is thus considered to have a very powerful influence upon the success of immigrants in the host country.

However, there still remains a question even after reading this article. Isphording explained that the research project was carried out by utilizing the available data set presented by the International Adult Literacy Survey. IALS states that they use their own calculations using the programs for calculating distance matrices (version 2.1), yet, there is very little research on this calculating method itself and there is no explanation anywhere, in this article, what sort of theoretical background the calculating system/formula is based on. In other words, it is practically unknown where these matrices stem from. A number of studies in the field of Linguistic Distance, or Linguistic Similarity, discuss the problem of having disadvantages of coming from distant linguistic backgrounds, and almost all of them employ the same data source that does not account for its origin and theoretical model. Because it does not tell us how these matrices were created and there is no other equivalent as the presenter of the same sort of database, there is a reasonable chance that some of the readers might raise their skepticism towards the validity of the whole procedural process in this complex research issue.

## V. Article 3: Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages

by Chiswick, B. R. & Miller, P. W. (2004)

In this quantitative type of study, Chiswick and Miller (2004) first discussed the concept of Linguistic Distance while adopting it to their own research context in which a great diversity is observed in the way different immigrant groups acquire the target language in their SLA situation. The study started witnessing the phenomena that immigrants from some countries of origin appear to be less proficient in the dominant language of the destination than other immigrants. To a certain extent, however, this might depend on some other variables such as the personal situation of each immigrant – whether s/he is likely to stay only temporarily or settle down in the new environment. Or, some tendencies of the attitude and some featuring traits of an immigrant’s characteristics towards the behavior, “learning a new language,” might have to be taken into consideration to a great extent.

Even if those factors would be incorporated in the analysis, there is another reasonable variable that can be possibly considered to be one of the causes of the phenomena mentioned above: the differences in the “distance” between the various immigrant languages and the destination language. Chiswick and Miller introduced a hypothesis, citing the example by Corder:

If English is linguistically “closer” to Western European languages (such as French and German) than it is to East Asian languages (such as Korean and Japanese), it would be expected that Western European immigrants in the U.S., UK, Canada and Australia would attain a higher level of proficiency in English, and would attain any

given level of proficiency sooner, than immigrants from East Asia (see, for example, Corder, 1981, pp. 95-102).

At the same time in the article, it is also explained that languages are complex, differing in many aspects on, for instance, on the lexical, syntactical and semantic level. The phonological aspect should not be overlooked either. They also raised a question, “If the difference is ‘large,’ how large is ‘large’? We can imagine at ease that English is ‘closer’ to French than to Chinese, but by how much is it closer? It would be also more complicated when we start considering the closeness in the relation between Arabic and Russian or Chinese and Japanese.

Under the circumstances, linguists have developed models of the origins of languages, which are called “language trees” as properly introduced by Crystal (1987, p. 292). Furthermore, Crystal wrote that “the structural closeness of languages to each other has often thought to be an important factor” in Foreign Language Learning and if a learner’s L1 (mother tongue) is structurally similar to L2 (the target language), learning should be easier than otherwise (p.371).

How can the Linguistic Distance be measured then? Hart-Gonzalez and Lindemann reported language scores for 43 languages for English-speaking Americans of average ability after periods (16 weeks and 24 weeks) of foreign language training (1993). The table below indicates the “direct codes” labeled for the languages for which “the Ethnologue Language Family Index” was utilized, which was published by Grimes and Grimes in 1993. The scores are set from low (harder to learn) to high (easier to learn). Therefore, this measure is based on the difficulty that Americans have when learning other languages than English.

Based on this table, we can visually confirm that, for example, Japanese is one of the hardest language to learn and Afrikaans is one of the easiest to learn among these languages. These languages are reported to be the ones spoken by foreign born and native born segments of the population in the United States, whose scores are considered to be usable for statistical analyses of language issues.

Table 3: Index of difficulty of learning a foreign language (language scores) and codes for languages reported in the U.S. Census – excerpted from the study of Chiswick and Miller (2004)

Language	Score	Language	Score
Afrikaans	3.00	Swedish	3.00
Dutch	2.75	French	2.50
Spanish	2.25	German	2.25
Finnish	2.00	Greek	1.75
Vietnamese	1.50	Arabic	1.50
Mandarin	1.50	Cantonese	1.25
Japanese	1.00	Korean	1.00



This matrix is then used in analyzing the English language proficiency of adult immigrants as well in order to observe the immigrants’ acquiring the English language in the United States and Canada, after arriving from non-English speaking origins. Chiswick and Miller at once described that what is empirically found is the greater the distance between an immigrant’s L1 and English, the lower is the level of the immigrant’s English proficiency. It is to some extent possible, therefore, for researchers and teachers to predict a learner’s starting point of learning and reachable level of English proficiency beforehand, which enables the English educational system to establish some more policies and strategies in the framework of ESL in their countries.

VI. Article 4: Learning Complex Features:  
A Morphological Account of L2 Learnability

by Schepens, J., Van der Slik, F., & Van Hout, R. (2013)

Schepens, J., van der Slik, F., and van Hout, R. reported their study on the L2 (second language) learnability. They first stated that adults often struggle when learning to understand a second language (L2) and express themselves in it while children seem to learn languages easily in a natural way. It is noticed, in particular, from the viewpoint of morphological complexity in L2. The study of Schepens et al. focused on this point, by investigating the Linguistic Distance (language differences) and complexity of the learners’ first languages and Dutch as the target L2 language.

According to Schepens et al., there has been very little research so far in this field, concentrating specifically on the morphological aspect in complexity. “Linguistic Distance” is often to be blamed for the obstacles in the case of adult language learning, however, major factors of the obstacles, in other words, what hinders the fluency in the language learning process, has not been discussed in details and major factors contributing to the learnability of target languages have not been defined clearly. Schepens et al. paid a special attention to the morphological aspects in language learning and the morphological differences in Dutch and other languages. The table below explains the distinctions between morphologically complex

Table 4: Dimensions in which morphologically more and less complex languages are assumed to differ

Dimension	Morphologically less complex	Morphologically more complex
Restrictedness	Ambiguous	Overspecified
Linguistic Strategy	Lexical / word order	Inflectional / conjugational
Learning Mechanism	Selection (facilitates L2)	Redundancy (facilitates L1)
Linguistic Type	Isolating	Synthetic
Cultural Type	Exoteric	Esoteric
Population	High, many adult learners	Low, many child learners

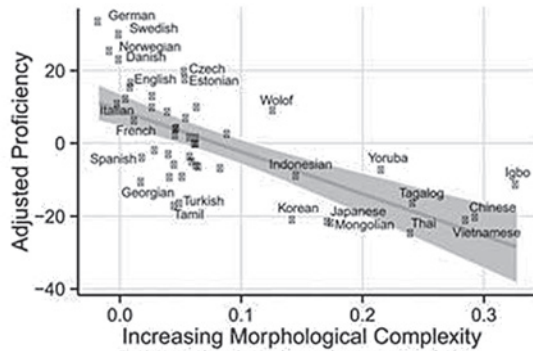


Figure 2: Correlation between the degree of morphological complexity and the proficiency of learners of the Dutch language (L2)

and less complex languages.

Schepens et al. investigated a set of 28 morphological features, looking at both of individual features and the total set of features. The table above explains the dimensions of morphologically more and less complex languages and the assumption of how they differ. Further they looked into the class of increasing and decreasing morphological complexity and found out through observation and analyses that the proficiency of adult Dutch learners shows correlation with features based on increasing morphological complexity. They concluded that the decisive property of L2 learnability seems to be “increasing complexity” of morphological features of the target language, in this case, Dutch.

Based on their statistic data in the article, they claim that the more morphological complexity increases, the more difficulty adult learners seem to face in acquiring the L2. Figure 2 indicates the correlation between the degree of morphological complexity and the proficiency of learners of the Dutch language as their L2.

In this study, they made an investigation on the relation between proficiency measures of adult language learning and cross-linguistic differences in morphological similarity and complexity between 49 different L1s and Dutch as L2. Also they expounded that the concept of L2-learnability overlaps with cross-linguistic influence, and the L2 learnability depends on the L1 across learners. Based on these theoretical backgrounds, the study reached a conclusion: L2 learnability cannot be determined by simple measures of so-called “Linguistic Distance.” Instead, it is the morphological complexity of the target language (L2) that plays the major role to determine the learner’s L2 learnability.

## VII. Conclusion and Discussion

By carefully reviewing and holistically considering those four articles, dealing with the issue of the relation between the degree of L2 learnability and “Linguistic Distance,” I will attempt to

answer the research questions raised in the beginning.

Research Question 1: “What is the definition of “Linguistic Distance”?”

The answer to this question is: “There has not been any definition of “Linguistic Distance” clearly described in the related literature of the field. Furthermore, I have not been able to find – so far – any literature in which the authors have given the difference between “Linguistic Distance” and “Language Distance”; instead they have been using those two expressions occasionally interchangeably, or all the way not mentioning about one – without giving any definition.

Research Question 2: Is there any measurement-system that determines the concept known as “Linguistic Distance” or/and “Language Distance”?

In spite of the fact that a number of researchers have attempted to measure, or to establish the measurement system for determining the accurate “Linguistic Distance” or/and “Language Distance,” there has been very little literature dealing with this topic in the holistic approach. I have not been able to find any article that firmly refers to a measurement-system that could function as a robust and accurate tool to measure the distance between languages. In other words, no study has established any solid standard or criteria to measure it yet.

Finally, there is a new perspective I try to cast. Back to the very first discussion that was brought up in the Introduction-section, we mentioned about poor L2 learners. It is usual that the difficulty or the degree of L2 learnability is heavily dependent upon the distance between the languages (the learner’s L1 and L2). However, there seems to be another major factor that should be seriously taken into consideration: “Motivation” to learn the L2. Yoshida (2009, p. 375) sharply pointed out, about this issue, that we often miss a few more significant factors in L2 learning/teaching. One of them is the role the society is playing in “providing the necessary external motivation” for the L2 learners to want to learn the target language. Instead of only keeping blaming the “distance” for not being able to learn the language, we should also focus on researching other factors contributing to the poor result we get in learning/teaching new languages.

## References

- ASJP: The Automatic Similarity Judgment Program. Wordlist & meanings. Retrieved from <http://asjp.cld.org/> (wordlist & meanings) Retrieved, February, 2015.
- Chiswick, B. R. & Miller, P. W. (2004). Linguistic Distance: A Qualitative measure of the Distance Between English and Other Languages. *Discussion Paper Series. Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labor)*.
- Crystal, D. (1987). *The Cambridge Encyclopedia of Language*. Cambridge: Cambridge University Press.
- Elms, T. (2008). Lexical Distance Among the Languages of Europe.  
Retrieved from:  
<https://elms.wordpress.com/2008/03/04/lexical-distance-among-languages-of-europe/>  
Retrieved, February, 2015.
- Hughes, H. (1999). Cultivating the walled garden: English in Japan. *English Studies*, 80, 6:556-568.
- Ispording, I. E. (2013). Disadvantages of Linguistic Origin: Evidence from Immigrant Literacy Scores. *Discussion Paper Series. Forschungsinstitut zur Zukunft der Arbeit (Institute for the Study of Labor)*.  
Language Difficulty Ranking. Effective Language Learning.

- <http://www.effectivelanguagelearning.com/language-guide/language-difficulty> Retrieved, February, 2015.
- Nerbonne, J. & Hinrichs, E. (2006). Linguistic Distances. *Proceedings of the Workshop on Linguistic Distances, Sydney, 2006, Association for Computational Linguistics*, 1- 6.
- Richards, J. C & Schmidt, R. (2002). *Longman's Language Teaching & Applied Linguistics*. (3<sup>rd</sup> ed.). Essex: Pearson Education Limited.
- Schepens, J., Van der Slik, F., & Van Hout, R. (2013). Learning Complex Features: A Morphological Account of L2 Learnability. *Language Dynamics and Change*, 3, 218-244.
- Yoshida, K. (2009). Factors to Improve English Language Education in Japan. *Sophia Linguistica: working papers in linguistics*, (57), 371-376.