Psychological Research using Linguistic Inquiry and Word Count (LIWC) and Korean Linguistic Inquiry and Word Count (KLIWC) Language Analysis Methodologies*

Changhwan Lee¹, Kyungil Kim², Jeongsub Lim³, Yoonhyoung Lee⁴

¹Sogang University, Department of Psychology  
²Ajou University, Department of Psychology  
³Sogang University, School of Communication  
⁴Yeungnam University, Department of Psychology

Language analysis methodologies such as LIWC and KLIWC, its Korean version, have been used widely in psychology and other social sciences. Given this context, this review aims to overview previous psychological findings based on the language analysis methodologies and compare findings deriving from LWIC and KLIWC. More specifically, this review examines important results in clinical psychology, developmental psychology, personality psychology, cognitive psychology, and other social sciences. Further, this review evaluates major debates regarding the language analysis methodologies and provides a possible future direction for using the methodologies in the cognitive science field.

Keywords: Language analysis, LWIC, KLIWC, Psychology of Language, English, Korean

* This article was supported by the Sogang University Research Grant (201210070)

Corresponding Author: Yoonhyoung Lee (Department of Psychology, Yeungnam University). 280 Daehak-ro, Gyeongsan, Gyeongbuk, 712-749, Republic of Korea. Phone: +82-53-810-2231. Mail: yhlee01@yu.ac.kr
1. Introduction: Linguistic Inquiry and Word Count (LIWC) and Korean Linguistic Inquiry and Word Count (KLIWC)

Language analysis has not been considered as a major research methodology for psychology and other social sciences. Many psychological studies have employed survey and errors and reaction time to investigate various aspects of psychological characteristics and underlying representation systems.

We can explain why language analysis methodologies have played a relatively limited role as a research tool. One possible reason includes that previous analysis methodologies focus on content words. Many previous language analysis methodologies examined content words that are related to research purpose (e.g., political words, Hart, 1984). For example, if researchers want to distinguish between depressed people and normal people in a clinical diagnosed situation, a language analysis program mostly includes only positive and negative emotion words because the researchers assume that the depressed person would use less positive emotion words and more negative emotion words than the normal person do. However, an analysis of usages of the function words that is another important component of a language text could be very valuable for revealing the writers’ various psychological characteristics.

The examples of function words include articles, prepositions, auxiliary nouns, pronouns, and various morphemes. Function words have no specific meaning but the usage of the function words can reveal the complexity of a written or spoken text for a sentence to a sentence. Language users do not pay much attention to the usage of function words while they use the content words. Thus, it is highly possible for function words to reflect various psychological characteristics of the language users.

LIWC--Linguistic Inquiry and Word Count--is a language analysis program that allows researchers to analyze various function words and language structures such as articles, prepositions, pronouns, number of words in a sentence, which include psychological variables such as emotion words, cognitive words, and social words (Pennebaker, Francis, & Booth, 2001; Pennebaker, & Graybeal, 2001). LIWC has successfully represented main psychological variables for clinical, personality, developmental,
cognitive, and individual differences.

KLIWC is a Korean version of LIWC that considers Korean morphological properties and vocabulary systems (see the specific developing procedures: Lee, Shim, Yoon, 2005). The KLIWC has successfully replicated previous findings based on the LIWC. Furthermore, the KLIWC has been used in various fields of psychology and other social sciences. In this sense, we believe that it is a right time to evaluate the previous studies using the language methodologies and provide future directions.

The KLIWC is not just a translated version of the LIWC. It analyses morphological components of phrases in a text, and calculates the ratios of corresponding morphemes. In addition, it included some cultural specific content word categories. The semantic structures of languages used by Korean and Western people would be somewhat different because some words might belong to different categories of psychological variables in each language analysis program. Accordingly, developing a Korean psychological dictionary is a prerequisite step for assigning words to their corresponding categories. The KLIWC includes the variables of social status words and respect words to capture aspects of Korean culture. Social status is an important social psychological concept for every culture, but Koreans tend to be sensitive to social status words and respect words. Korean culture emphasizes respect for older people and acceptance of advice from parents and senior citizens. That is why respect words have been sophisticatedly developed in Korea. In this vein, the variable of respect words was added to the KLIWC. Honorific words have the unique morphemes in its word such as “씀/ssyeos/ye” “십/sip/” “옵/op/” variation. Respective words were detected by a focus on these unique morphemes.

The KLIWC is a user friendly software that integrates a linguistic analysis module and a variant analysis module. As shown in Figure 1, the software provides users with the same data processing function as the LIWC, but its interface is convenient and easy for users as compared with the LIWC.
Fig. 1: System Architecture of KLIWC

The KLIWC consists of three main modules: (1) linguistic analysis, (2) variant selection, and (3) variant analysis. An analysis of complex Korean morphemes has been possible because the process of modifying two natural language methodologies were developed by A. S. Yoon and her colleagues at the Korean Language Processing Laboratory of the Pusan National University. The Korean Morphological Analyzer (K-MorAn) and the Korean Grammatical Category Tagger (K-GraCat) decompose a text into sentences and divide a Korean word into...
morphemes. They also attach each constituent with POS (part-of-speech) tags and calculate the probability of each candidate analysis by using a context analysis and stochastic information. K-GraCat had the capacity to differentiate 176 hyper sub-categories in terms of syntactic, morphologic, and semantic features. Since the KLIWC does not require such a fine-grained analysis, we proposed the classification of two different upper layers in terms of coarse-grained criteria. In the KLIWC, a POS tagger divides words into nine basic grammatical categories in the first layer and 47 sub-categories through semantic-syntactic properties in the second layer. Researchers can select the fineness of grain in using and evaluating the KLIWC.

This paper reviews and compares main findings of previous studies using LIWC and KLIWC. This paper further proposes several debatable issues by considering the results of the previous findings. This paper discusses future directions and emerging topics for the usage of the two language analysis methodologies (LIWC and KLIWC). Although researchers have developed LIWC dictionaries in 11 languages, most LIWC studies have relied on data in English. This situation allows us to compare studies of English languages and Korean languages in terms of their efficiency.

2. Findings from LIWC and KLIWC

2.1. Applications to Psychological Fields

By using the two language analysis methodologies, many researchers have investigated the differences in language usage among diverse clinical groups. The researchers have relied on selecting clinical patient groups from hospitals and comparing them with normal groups chosen from students who are similar in terms of age, gender, and education.

By using the LIWC, researchers found that a depressive group used the first-person pronoun and negative emotion words more frequently than a control group did (Lee, Ahn, Kim, & Lee, 2007; Rude, Gortner, & Pennebaker, 2004; Stirman, & Pennebaker, 2001). This indicates that
depressive people are more self-focused and more negative toward their lives than normal people. The findings suggest that a clinical group can be diagnosed by the LIWC. By using the KLIWC, other researchers tested a clinical group having ADHD (Kim, Lee, & Lee, in press). As a result, the ADHD group used ‘numeral pronoun’, ‘clause per sentence’, ‘morpheme per sentence’, and ‘home related words’ less frequently than a normal group did. Simultaneously, the ADHD group used ‘TV & movie related words’, ‘sentences’, and ‘adjectives’ more significantly. The ADHD group also had few language units (i.e., clause, morpheme) in a sentence but had many sentences, indicating that the ADHD group did not use complicated linguistic structures.

The language analysis of these clinical groups reveals that clinical symptoms can induce different usage of functional words or language structure. The analysis also indicates that the language analysis program can be an effective tool for diagnosing clinical disorder when it is combined with conventional psychological tests or interviews.

Personality psychology is another subfield of psychology, which has used the language analysis program. The two language methodologies (LIWC and KLIWC) are applied to an analysis of whether different personalities lead to distinct patterns of language usages. The methodologies rely on the personality test of big five factors to examine the relationship between the personality scores and the frequencies of language variables. Consequently, the personality factor of openness is negatively correlated with first-person singular pronouns, present tense, and long words (Pennebaker & King, 1999). In addition, the extraversion and conscientiousness is negatively correlated with negations and tentativity (Lee, Seo, Kim, & Chung, 2007). These results imply that even unconscious personal factors can emerge in language uses such as writing and speaking.

The language analysis has also been applied to an examination of individual differences such as gender and social status. A representative study investigated gender differences in the frequency of word usage in a daily life (Mehl, Vazire, Ramirez-Esparza, Slatcher, & Pennebaker, 2007). The traditional assumption regarding gender differences is that women speak and use more words than males do. The LIWC study, however, detected no difference in the frequency of language usages between males
and females (Mehl, et al., 2007). This result contradicts the common belief about the word usage by males and females. The underlying fact is that females use more words in public space and males use more words in private space, and thus, differences in the total usage of words cancel out.

Other social psychology researchers have examined what aspects of a language predict social matching or social dynamics (Gonzales, Hancock, & Pennebaker, 2010; Ireland, & Pennebaker, 2010; Mehl & Pennebaker, 2003; Niederhoffer & Pennebaker, 2002). They found that partners having good social matching use language structures and employ amounts of words similar to each other (Ireland, & Pennebaker, 2010; Niederhoffer & Pennebaker, 2002). In addition, Mehl and Pennebaker (2003) found that people who stay away from a social upheaval (i.e., the September 11 terrorist attack) interact frequently and try to find the reasons of the upheaval right after the attack. It contradicts our common knowledge in that people experiencing such a social upheaval are directly related with victims and damage caused by the attack. This previous study implies that we can glean language data from social media such as Facebook, Twitter, and various news websites when important social events occur.

Developmental psychology is an important discipline to which the language analysis methodologies are applied. Traditional theories of developmental psychology argue that old people tend to think about the past more frequently than the future and that they use more negative words than positive words (Pennebaker, & Stone, 2003). General cognitive functions also decrease as people get old. In contrast to these traditional arguments, some recent theories suggest that as compared with young people, old people in developed countries focus on their retirement plans and their future; think more positively; and think in more complicated manners (Perdue, & Gurtman, 1990).

These two types of arguments about old people were tested using the language analysis program. Researchers (Pennebaker, & Stone, 2003) applied the LIWC to an analysis of the writings by old and young people. As a result, old people use positive and future-related words, and complicated sentences more frequently than young people. In particular, the finding that old people use complicated sentences indicates that some cognitive functions do not decrease as people get old. These results are not
consistent with the traditional developmental theories, because they suggest positive perspectives regarding getting old.

In a similar vein, the KLIWC was used to an analysis of television drama scripts of various actors and actresses from teenagers to more than 70-year-old senior citizens (Lee, Park, 2006). This previous study assumes that drama languages represent real language usage patterns in daily life and found patterns similar to those of the American study except for the oldest group. Specifically, as people get old, they generally use achievement oriented words, future related words, and morphemes per sentences more frequently than the other groups. However, 60-year-old people do not follow this pattern, and the oldest group uses achievement-oriented words, future-related words, and morphemes per sentences less frequently than the other groups.

This different pattern can be explained by socio-economic perspectives. South Korea is a developing country as compared to Western developed countries. Thus, the oldest group does not have sufficient economic status and welfare, which could not allow the group to make retirement plan and perform social activities while maintaining language complexities. Another possible reason is the nature of language samples. Main actors or actresses in dramas are younger people and older people might have only supporting roles. Such roles in dramas could weaken the usage of achievement orient words and future related words. This could be a confounding variable that affect the language usage of people in their daily lives. Future studies should consider actual writing or speaking among real population.

In addition to developmental psychology, cognitive psychology has not been examined substantially by the language analysis methodologies, though this discipline is the fundamental area that can explain psychological mechanism underlying the usage of the linguistic materials. In cognitive psychology, an important finding by the LIWC includes the effects of confession on working memory capacity (Klein & Boals, 2001). Klien and Boals (2001) examined the enduring effects of confession on the cognitive mechanism. Previous researchers (Pennebaker & Graybeal, 2001) found that the behavior of confessing personal trauma improves related psychological symptoms, personal problem, and even health. These researchers investigated the possible change of working memory capacity
before and after confession. They expected that concealing personal trauma would create multiple reasons for the incident and that the subject would have ambiguous attitude toward the reasons. Maintaining this ambiguity stimulates the loads of cognitive resources including working memory. Thus, confessing the possible reasons of trauma would make up a story having possible causes and effects, reducing the load of previous working memory. Klien and Boals (2001) found that when the subject wrote the confession, her or his working memory capacity increased.

This previous study has a significant impact on the field of language psychology by suggesting that some intentional activity or training can change the basic cognitive ability. Instead, other psychological researchers assumed that working memory capacity could elicit changes in clinical symptoms and mechanism.

Some researchers (Lee, Lim, Kim, & Lee, 2011) tested the effects of working memory loads on the change of language activities by using the KLIWC. They asked participants to write stories based on ambiguous pictures of the Thematic Apperception Test (TAT). The half of the participants performed a dual task of memorizing a random sequence of digits, and their writings were compared with those having no dual task. The results showed that working memory holding induced significant changes in language activities, such as usage of fewer words, simpler language structures with fewer words per sentences, and less long words. These previous studies using the language analysis methodologies show that manipulating working memory affects language usages and that psychological activity such as confession influences the important cognitive function of working memory.

Recently another researchers (Kim, Bae, Noh, & Lee, 2011) investigated the language styles of experts and novices by analyzing word profiles between expert groups and novices. They found that experts use words of conviction and restriction frequently, indicating their specialized knowledge. In addition, experts focus on complicated sentence structure and words related with causal reasoning.
2.2. Applications of LIWC and KLIWC to Social Sciences

The LIWC has been used mostly in psychology, and it has been seldom applied to other social sciences that can focus on languages as a dependent measure. Many social sciences consider languages as a main medium in the activities and the measurement of languages can provide important information for building of a theory and a model. Possible areas include mass communications, political science, library science, and education.

Researchers have started to apply the KLIWC to an analysis of topics in these fields to provide basic language profiles. For example, representative studies analyzed language usages in newspapers and propaganda from 1960 to 2000 in the mass communication fields (Park, Lee, & Park, 2007; Park, Park, & Lee, 2008). Language complexities such as morpheme and phrases increase from 1960 to 2000 (Park et al., 2007). This would signify the trend of mass communication that emphasizes micro-specific information. Another aspect of language changes includes the usage of words related to achievement. The usage of such words increased from 1960 to 1980, but decreased after that period to the lowest level in recent period. This might derive from the initiative of the Korean economic development, which the Korean government and news media encouraged people to work hard and achieve the national goal of economic prosperity. Indeed, newspapers and other news media support the corresponding stages of the economic and social development.

Other researchers have also examined the differences between South Korean languages and North Korean languages (Lee, Kim, & Park, 2010), analysis of political leaders, and criminal language analysis (Seo, Kim, Lee, & Kim, 2012).

3. Theoretical Issues and Improvement

3.1. Can Language Changes Elicit Psychological Changes?

As described earlier, the language analysis methodologies have identified psychological mechanisms in main areas of psychology successfully. Thus, a thought can induce differences in language behaviors, and psychological
processes can be revealed through an analysis of language usages, especially through the usages of function words. The assumption of the directional influence of a thought on language behaviors has a solid basis, and the influence of language behaviors on a thought should be examined. The remaining hypothesis is that when we changed or were forced to change language styles, the psychological processes could be changed.

This hypothesis also addresses an important theoretical issue in language science because the issue is related to the weak version of the Sapir-Whorf hypothesis. The Sapir-Whorf hypothesis argues that the usages of different daily life languages can induce the change in perception of the environment and the world. Several scientists are still investigating the relationship between a thought and other cognitive processes (e.g., Lupyan, & Spivey, 2010; Lupyan, & Ward, 2013). However, the original hypothesis is based on ad-hoc observations or investigations. It has not been tested by scientific experiments. The reason is that it is not practical to conduct a language change manipulation for any populations who use their native language in a real life.

The way to examine such hypothesis would be to induce certain changes in writing during several intensive training sessions and measure the effects using the LIWC or the KLIWC. Since the language analysis methodologies have discovered that psychological mechanisms can be signified by different usage of pronouns, we could design a writing format that induces the usages of first-person pronoun or third-person pronoun and investigate the possibility of psychological changes after the experiment. The writing training can be offered during many different intense sessions. Recently, Kaufman and Libby (2012) showed that the first-person pronoun focusing on writing elicited changes in participants’ behaviors and the highest levels of experience-taking.

The other way is to investigate the cognitive patterns of bilinguals. People who could speak two languages fluently could help researchers test the Sapir-Whorf hypothesis. If the bilinguals use cognitive styles differently as compared to monolinguals, we can assume that a specific language plays a role in cognitive processing.
3.2. Why do Function Words Emerge as Main Variables?

The results of from the LIWC and the KLIWC emphasize the importance of the analysis of the usage of the function words. Specifically, the usages of pronouns are as critical variables as the language structure variables such as the number of units (i.e., words, morphemes, phrases) per sentences. Then, why do the function words as a hidden mechanism play an important role in suggesting psychological variables? There are two possible reasons.

First, a writer can control the usage of contents words based on the writer’s intention and his or her tendency of being desirable to potential readers. The usage of the function words, however, is rather automatic and relatively free from the writer’s controlled intentions and other motivations. Uncontrolled behaviors are more likely to represent the individual differences in psychological research, which is similar to the traditional psychological variables such as reaction time and various psychological tests. Subjects can hardly control the reaction time and the response pattern to the questions in various psychological tests.

Second, as discussed in the cognitive psychology section of this review, if the working memory capacity plays an important role in language uses, those who have large working memory capacity might use more complicated language structures, which leads to more use of the function words. The clinical populations such as ADHD and depressive people have a lower level of working memory capacity than normal people. Therefore, it is possible that the former uses fewer function words, such as fewer number of language units and a more simple syntactic structure.

3.3. Limits of Language Analysis Program

Although the language analysis methodologies have produced meaningful findings, they have several limitations in terms of a research tool and its application. The most significant limitation includes that as compared to the traditional research tools, it cannot reveal the causal relation between interesting variables. Since it is not an experiment manipulating independent variables in laboratories, we cannot make sure that language patterns result from variables of interest. Other confounding variables can
be present in the relationship between psychological variables and the use of the language patterns.

Another limitation is related to the statistical properties of the results. The significance level of most psychological studies is 0.05, which refers to the probability of making a type I error. However, the LIWC and the KLIWC are more susceptible to a type I error. Therefore, we should interpret the results carefully and should evaluate whether the research is based on a solid theory. The dictionary-based language analysis program has another limitation, which is that the percentages of words in each category might not be meaningful without comparing a base rate. We do not have the general frequency of these categories in everyday usage.

One of the strongest remedies for these limitations would be to develop alternative hypotheses before collecting data. The alternative hypotheses make it possible to focus on specific language variables in advance, which helps increase the statistical power. Developing alternative hypotheses also increases the possibility of identifying a causal relationship between interesting variables, though it is not in a perfect manner.

4. Synthesis and General Future Directions

The KLIWC and LIWC language analysis methodologies have been successful in identifying main variables in research areas, indicating the possibility of using language behaviors as a reliable research tool. We also discuss some limitations regarding the application of the language analysis methodologies. In this section, we suggest the future directions for language analysis methodologies from more general perspectives.

4.1. Necessity for Data Accumulation Using KLIWC

While many studies have relied on the Korean analysis program, KLIWC, for their data analysis, its application remains at a low level and the program needs to be applicable to diverse research areas. First, the KLIWC is still limited to some groups of researchers and scholars. Few psychologists and mass communication researchers use it. Second, the research areas for the KLIWC are not diverse. For example, physiological psychology
and social psychology can take advantage of the Korean program. The LIWC have been used to an analysis of psychological changes after writing (Booth, Petrie, & Pennebaker, 1997; Petrie, Booth, & Pennebaker, 1999), social upheaval (Mehl & Pennebaker, 2003), lying versus truth (Newman, Pennebaker, Berry, & Richards, 2002), and literature analysis such as poem (Stirman, & Pennebaker, 2001).

The KLIWC should be applied to these unexamined topics so that it can provide in-depth data for the comparison with the LIWC. Through these processes, we can obtain the information on the similarities and differences between Korean people and Western people in terms of their language usages, which eventually suggests more general psychological models on the language usage.

4.2. Analysis of Relationships Between the RT Tests and Language Variables

As described, the reaction time has been one of main psychological dependent variables in psychological studies. Thus, to propose the language analysis as a new dependable measurement, researchers need to compare various reaction times studies that tested similar psychological phenomenon. Specifically, we can acquire extensive data on subjects as to how they use language and perform in main reaction time tasks. To this end, researchers need to calculate correlation coefficients between various indexes of language uses and subject’s reaction times of attention and examine categorization and working memory in the psychological laboratory. The main focus is on whether a high-speed cognitive performer would show efficient language usage.

An alternative way of obtaining informative data on the relationship between these two realms may be to examine how the reaction times and the language usages are changed by a manipulation of psychological variables. For example, it is possible to use a dual task paradigm and observe the effect of the manipulation on attention. In this scenario, subjects are required to perform an additional task and the experimenter measures the reaction time of the main attention task, such as engaging in the visual search task while working on certain language production as an additional
task. If certain variables of the language usage are influenced by this manipulation of divided attention, we could consider the variables to play a critical role in early human cognitive processes. This information would serve as an important basis for developing a human language information processing model that many psycholinguists seek.

4.3. Conclusions

This review examines previous language analysis studies that used LIWC and KLIWC as new research tools especially in psychology. It reviews and compares the studies on clinical, personality, social, developmental, cognitive psychology, and other social sciences.

The review discusses some limitations and future directions by considering the previous findings. Although the language analysis methodologies contain some defects and flaws, language styles can serve as new dependent variables for various psychological studies.

The language analysis methodologies can be applied to other possible future research topics and relevant fields. One good example is to use the KLIWC in humanities research such as literature and history. Humanities are based on written documents and books that can be submitted to an analysis of the program.

References


Kim, Kyungil, Lee, Sungjik, & Lee, Changhwan. in press. College students with ADHD traits and their language styles, Journal of Attention Disorder.
Lee, Changhwan, & Park, Jongmin. 2006. The analysis of linguistic styles by age in TV dramas, Psychological Reports, 90: 77-90.
Lupyan, G. & Spivey, M.J. 2010. Redundant spoken labels facilitate perception of multiple items, Attention, Perception, & Psychophysics, 72(8), 2236-2253.
Lupyan, G. & Ward, E.J. 2013. Language can boost otherwise unseen
objects into visual awareness, *Proceedings of the National Academy of Sciences*. 110(35) 1419-201


Psychology, 75: 1246-1272.