Article

Journal of Language and Social Psychology
2017, Vol. 36(3) 343–355

DOI: 10.1177/0261927X16657855
journals.sagepub.com/home/jls

Development and Examination of the Linguistic Category Model in a Computerized Text Analysis Method

Yi-Tai Seih[1,2], Susanne Beier[3], and James W. Pennebaker[2]

Abstract
The linguistic category model (LCM) seeks to understand social psychological processes through the lens of language use. Its original development required human judges to analyze natural language to understand how people assess actions, states, and traits. The current project sought to computerize the LCM assessment based on an idea of language abstraction with a previously published data set. In the study, a computerized LCM analysis method was built using an LCM verb dictionary and a part-of-speech tagging program that identified relevant adjectives and nouns. This computerized method compared open-ended texts written in first-person and third-person perspectives from 130 college students. Consistent with construal-level theory, third-person writing resulted in higher levels of abstraction than first-person writing. Implications of relying on an automated LCM method are discussed.

Keywords
LIWC, linguistic category model, LCM, imagery perspective

[1]National Taiwan University of Science and Technology, Taipei City, Taiwan
[2]The University of Texas at Austin, TX, USA
[3]University of Heidelberg, Heidelberg, Germany

Corresponding Author:
Yi-Tai Seih, Department of Business Administration, National Taiwan University of Science and Technology, No.43, Keelung Rd., Sec.4, Da'an Dist., Taipei City 10607, Taiwan.
Email: yitaiseih@gmail.com

Language provides a window to understand social behavior and psychological processes. Using language analyses, it is possible to explore facets of the psychological states of a speaker or writer. For example, people who are focused on others tend to use more second-person or third-person pronouns in conversations, whereas people who are self-aware are more likely to use first-person pronouns (for a review, see Tausczik & Pennebaker, 2010).

People communicate through language, and language itself carries rich information about human thoughts. Due to this, language becomes a rich resource for psychologists to study social processes. One of the prominent psychological theories about language use is linguistic category model (LCM; Semin & Fiedler, 1988, 1992). Semin and Fiedler distinguished among four types of words that are used to describe interpersonal behavior and communication and to have a better understanding of social psychological processes. These four types include adjectives and three types of verbs. The underlying logic was that each of these linguistic categories identified how abstractly the speaker or writer was understanding their world.

Adjectives, which are often used to describe personality traits, are the most abstract category in the LCM because traits that refer to personality cannot be objectively examined (e.g., *charming*). Compared with adjectives, verbs are assumed to be less abstract and more concrete in the LCM. Semin and Fiedler (1988) distinguished three types of verbs with which interpersonal descriptions can be made. The first type of verb is the state verb (SV). SVs are used to describe emotional or mental states but do not contain a clear beginning and end (e.g., *admire*). In the LCM, SVs are less abstract than adjectives, but SVs are the most abstract type among the three types of verbs. The second type of verb is the interpretative action verb (IAV). IAVs are used to characterize more general behavior that share a physically invariant feature (e.g., help). Thus, conceptually, IAVs are more concrete than SVs in the LCM. The most concrete type of verb is the descriptive action verb (DAV). DAVs refer to verbs that describe a specific and observable action with a clear beginning and end (e.g., walks). Although functions of various verbs are similar, the three types of verbs that Semin and Fiedler (1988) identified reflect different representation of physical or mental activities.

Later, Semin, Görts, Nandram, and Semin-Goossens (2002) added nouns to the LCM because nouns can describe characteristics of a person in the same manner as adjectives in some circumstances. For instance, if we say someone is an athlete, we may describe that person as athletic. Therefore, nouns are classified as an abstract category like adjectives in the LCM. These five language categories constitute a linguistic continuum in the LCM that can reflect levels of abstraction in language, and this continuum distinguishes words on a scale from abstract to concrete predicates (nouns, adjectives, SVs, IAVs, and DAVs; for a review, see Semin & Fiedler, 1991). With the development of the LCM, language abstraction becomes an indicator that can assess to what degree people reveal their abstract or concrete thoughts to others in language usage.

The LCM has implications for practical use. One such example is construal-level theory (CLT; Liberman & Trope, 1998; Trope & Liberman, 2003, 2010). The CLT examines abstractness with which people mentally represent objects, events, and people as a function of subjective closeness or distance in different dimensions. According to the CLT, people construe entities that are more distant in any of those dimensions in a more abstract way. Thus, they tend to use more abstract language to describe things that are psychologically distant, whereas they tend to use more concrete language to describe something that is psychologically near. Applying the LCM framework, CLT researchers can assess psychological distance via language usage (e.g., Freitas, Gollwitzer, & Trope, 2004; Fujita, Henderson, Eng, Trope, & Liberman, 2006).

The LCM examines levels of abstraction in people's language use. Most LCM research has relied on human-based coding strategy. However, there are two crucial limitations that emerge when manual coding is used: a highly demanding cognitive load for the LCM coders and, as a consequence, small sample sizes of texts coded in LCM studies. Traditionally, LCM scores have been computed based on several different human coding schemes. In a standard LCM coding procedure, coders first learn definitions and instructions for each LCM category, then they go through texts sentence by sentence to identify words that match the definitions. After the identified words are coded, an LCM score is calculated based on LCM weighted coefficients (Semin et al., 2002). It is a big challenge to ask coders to evaluate thousands of words in a short period of time. From a methodological perspective, this coding procedure is time consuming and demanding for judges.

An alternative approach for studying the LCM is to adapt computer-based rating systems. Louwerse, Lin, Drescher, and Semin (2010) first integrated the LCM coding system into a computational implementation model. They identified verbs and adjectives by the LCM criteria and sent identified words to an English lexical database (Baayen, Piepenbrock, & van Rijn, 1993) to obtain derivations and inflections (e.g., come, comes, coming, and came). Their final LCM list consisted of 31,444 words. Louwerse et al. (2010) used this computational approach to analyze a big data set with 255,637 e-mail messages from a corporate social network. Their findings suggest that levels of abstraction in the LCM can be an indicator of fraudulent events in social networks. This computational approach is an important first step in understanding how to use computer-based systems to substitute for traditional LCM coding procedure. More important, Louwerse et al.'s (2010) computational approach suggests the possibility of overcoming the methodological limitations of the LCM.

Building on Louwerse et al.'s (2010) computational idea, the goal of this research was to adopt a simple but efficient computer-based system to assess levels of language abstraction in the LCM. Currently, there are several computer-based systems for the investigation of language use. One of those systems is Linguistic Inquiry and Word Count (LIWC, pronounced "Luke"), which was developed by Pennebaker, Booth, and Francis (2007). LIWC is a computerized text-analysis software program used to count words in defined dimensions. In particular, one of the advantages of LIWC is to allow users to develop their own dictionaries for analyzing any given dimension of language use. This advantage provides us an opportunity to create a dictionary to study the LCM.

The purpose of this article was not only to propose a computerized text analysis method for the LCM but also to empirically test this method with writing samples. In the present study, we followed the definitions of each category in the LCM (Semin &

Fiedler, 1988) to develop a word dictionary for the LCM verb categories and used a part-of-speech (POS) tagging program for adjective and noun categories in the LCM. The word dictionary could identify LCM-related verbs, whereas the POS tagging program can identify adjectives and nouns based on grammatical algorithms.

Because there are three types of verbs in the LCM, we created a word dictionary specifically to distinguish each verb type. The dictionary for verb use was assessed by the coders' rating. Adjectives and nouns were identified by a POS tagging program. Our computerized method was examined on the theoretical approach to language abstraction derived from CLT and the personal perspective literature.

The Idea of Language Abstraction

CLT states that psychological distance can lead people to reveal abstract or concrete thinking (Liberman & Trope, 1998; Trope & Liberman, 2003). Psychological distance has been defined in several dimensions, such as temporal, spatial, social, and hypothetical distance (Trope & Liberman, 2003, 2010). One of the dimensions of psychological distance is social distance, which can be induced by taking different person perspectives and leads to distinguishable mind-sets. For example, previous research has found that using self-immersed versus self-distanced perspectives to recall or describe experience causes different construal levels (Kross & Ayduk, 2011; Kross, Ayduk, & Mischel, 2005). When people recall from a self-immersed perspective, they are self-centered and describe more concrete features of their experience. On the other hand, when people recall from a self-distanced perspective, they become ego-decentered and use more abstract terms in their descriptions (Kross et al., 2005).

Similarly, Libby, Shaeffer, and Eibach (2009) manipulated visual perspectives (owner's first-person vs. observer's third-person) and found that participants in a first-person perspective interpreted human actions on a concrete level. However, when participants took a third-person perspective, they described human actions on an abstract level. Eyal, Liberman, and Trope (2008) also found that people are more attentive to contextual details when making decisions from a first-person perspective than from a third-person perspective. If visual perspectives indeed reflect different construal levels, the levels should appear to be different outcomes on a certain concept, such as language production. Thus, these relevant findings about visual perspective suggest an idea that a third-person perspective might lead to more use of abstract language, whereas a first-person perspective might lead to more use of concrete language.

The ultimate goal of this article was to examine our computerized method for the LCM. If the computerized method is promising, it should distinguish different levels of language abstraction while participants used different perspectives to narrate. Thus, we analyzed a data set from Study 2 in Seih, Chung, and Pennebaker (2011). In the data set, participants were instructed to describe a negative event in first-person, second-person, and third-person perspectives. The order of personal perspectives was counterbalanced across participants. Because our assumptions in this study specifically referred to a difference in abstraction between writing in first-person and third-person perspectives, we only focused on these two personal perspectives in our analysis.

The Present Study

There were three goals in the present study. The first and the second were to develop a dictionary for the LCM verbs and to test a POS tagging program for the LCM adjectives and nouns. The LCM dictionary for verb use and the POS tagging program constitute our computerized text analysis method. The last goal was to examine this computerized method with the idea of language abstraction.

This study first developed a word dictionary that consists of various verbs according to the definitions of the LCM, and examined whether this dictionary can identify verbs that can be classified into the three verb categories of the LCM coding scheme. Adjectives and nouns were directly identified by a POS tagging program. In order to determine if our computerized method is promising to differentiate levels of language abstraction, we used the LCM dictionary and the POS tagging program to examine the idea of language abstraction with a published data set.

Method

Procedure of the Development of the LCM Dictionary for Verb Use

To select commonly used verbs, our verbs were sourced from two places in two ways. The first place was from a giant corpus of texts we compiled, and other place was from a developed computer-based dictionary, General Inquirer LCM dictionary. In order to detect commonly used LCM verbs and classify those verbs into appropriate LCM categories, we took several steps to develop an LCM dictionary. To identify LCM-relevant words from a broad representation of natural language, we compiled a corpus of texts distributed across a variety of genres, including transcribed daily conversations, college admissions essays, writing samples from various undergraduate Psychology class exercises, blog entries, scientific articles, and inaugural addresses. The final corpus consisted of over 74,000 texts of approximately 600 words each, for a total of approximately 44 million words.

The second step was to select commonly used verbs from our corpus. To identify verbs in our corpus, the texts were processed by TreeTagger (Schmid, 1994), which is a POS tagging software with an accuracy of 96% per word (Schmid, 1995).[1] We then used WordSmith (Scott, 1996), a software package that calculates word frequency, to identify the frequency of texts in our corpus that included verbs. Verb's base rate of occurrence in all words is 24.4%, thus, we selected 1% to set criterion. The selection criterion for verbs was set at 0.2, indicating that a selected verb was used in more than two texts per 1,000. The selected verbs were considered to be common verbs, and unselected verbs were characterized as obscure verbs. Because obscure verbs were rarely used and required time and manpower to process, we did not include them in our dictionary due to efficiency considerations. The top 3,000 verbs were selected for inclusion. Among the 3,000 verbs, some verbs were in present tense and some were in other tenses. To get an exact number of verbs, we derived present-tense verbs from the 3,000 verbs, resulting in a total of approximately 900 commonly used present-tense verbs.

Steps 1 and 2 used a bottom-up approach to compile a verb dictionary. However, previous research has classified verbs on a theoretical basis (e.g., Stone, Dunphy, Smith, & Ogilvy, 1966). To construct a dictionary that used such bottom-up construction methods as well as theoretically based dictionaries, we also collected verbs from previously published dictionaries that were related to the LCM.

The third step was verb collection from the General Inquirer LCM dictionary. The General Inquirer, a computer-based dictionary for content analysis, was released in 2002 (Stone et al., 1966). The General Inquirer dictionary contains a variety of language categories, several of which are derived from the LCM. The LCM categories include three types of verbs: DAV (540 verbs), IAV (1,947 verbs), and SV (102 verbs). In total, 2,589 verbs were collected from the LCM dictionary of the General Inquirer. Eight native English speakers judged the 2,589 LCM verbs from the General Inquirer for the degree to which each verb is common or obscure (interjudge reliability was .53). Verbs rated as obscure by four or more judges were removed from the list because obscure verbs required additional manpower to process and only made a marginal contribution to our method. The remaining verbs were processed with our corpus described above for their usage frequencies. If fewer than 10 texts in our corpus included a particular verb, or if the absolute frequency was less than 0.001% of total words used, then the verb was removed from the list. This procedure resulted in 1,000 commonly used present-tense verbs from the General Inquirer LCM dictionaries.

Because we had two resources for verb collection, the fourth step was to merge present-tense verbs from our corpus and the General Inquirer LCM dictionary. The 900 verbs derived from our corpus were compared with the 1,000 verbs from the General Inquirer. There was an overlap of a 100 verbs between our verb corpus and the LCM verbs from the General Inquirer.

The last step was to classify the 800 unlabeled present-tense verbs from our corpus. Following the coding scheme of the LCM (Semin & Fiedler, 1988), three coders classified the 800 present-tense verbs into three categories: DAV, IAV, and SV. Cronbach's alpha for the internal reliability among the three well-trained coders with LCM manual coding procedure was .68. The primary researcher resolved the inconsistencies among the three coders based on the coders' majority decision. Our final LCM dictionary includes 1,000 common verbs from the LCM category of the General Inquirer and 800 frequently used verbs that were coded according to the LCM coding schemes from our corpus. In total, there were 1,800 present-tense verbs with LCM labels in the dictionary. The text analysis program, LIWC, searched for LCM-related verbs based on these 1,800 verbs. To take different verb tenses into account, we included verb inflections for all those verbs in our LCM dictionary. For instance, the verb *make* was in our LCM dictionary. We added other verb forms of *make* (makes, made, making) to our dictionary. The final LCM dictionary consisted of 7,489 verbs and was imported to LIWC for detecting LCM verbs.[2]

Procedure of Using a Part-of-Speech Tagging System

To detect adjectives and nouns, we used a POS tagging system, TreeTagger,[1] to process our text files. TreeTagger is an open source system and its program packages are free to use. TreeTagger is able to identify adjectives and nouns at rates from 93% to 97% accuracy compared with human judges and other POS methods (Schmid, 1995). In the current study, LIWC was used to convert POS tags into count ratios of adjectives and nouns.

The LCM Algorithm

The LCM assumes five language categories that are ordered on a continuous dimension of concreteness–abstractness (from DAVs, IAVs, SVs, to adjectives and nouns). Based on those categories, an algorithm to calculate an LCM score reflecting the difference in abstraction of the LCM concepts was developed by Semin et al. (2002). In their algorithm, they scored DAVs, IAVs, and SVs with the weights of 1, 2, and 3, and scored adjectives and nouns with the weight of 4. However, Carnaghi et al. (2008) conducted six studies on person perception and indicated that nouns are conceptually more abstract than adjectives. Their findings suggest that nouns should be weighted more heavily than adjectives in the LCM coding system. Consistent with Carnaghi et al.'s (2008) findings, we assigned the weight of 5 to score nouns. The revised LCM algorithm is as follows:

$$\text{LCM score} = \frac{\left[(\text{DAV}\times 1)+(\text{IAV}\times 2)+(\text{SV}\times 3)+(\text{adjective}\times 4)+(\text{noun}\times 5)\right]}{\left(\text{DAV}+\text{IAV}+\text{SV}+\text{adjective}+\text{noun}\right)}$$

In this algorithm, each language category was assigned a theoretical weighted coefficient to differentiate levels of abstraction among the five language categories. The final LCM score ranges from 1 (*very concrete*) to 5 (*very abstract*).

The Data Set

The data set examined whether people perceive more valuable meaningfulness or emotionality from different perspectives (see Study 2; Seih et al., 2011). One hundred and thirty undergraduate students at the University of Texas at Austin participated in the study. One participant did not follow the instructions, resulting in 129 participants for final analysis. Seih et al. (2011) used a within-subjects design to ask participants to recall a recent negative event and then describe the event in assigned personal perspectives, including first-person, second-person, and third-person perspectives. They found that participants perceived the writing to be more valuable and emotional while writing in the first-person perspective compared with other perspectives.

The LCM dictionary was imported into LIWC to process the texts. Three categories, DAV, IAV, and SV, were provided in LIWC's processing results for each data set. Because LIWC has neither adjective nor noun categories, the texts were also analyzed by TreeTagger which provided POS tags for the adjective and noun categories. According to our revised LCM algorithm, LCM scores were computed using the five categories from the text analysis. The final computerized derivation included the five LCM categories and the computerized LCM scores for the data set. The LCM categories and the LCM scores were analyzed.

Table 1. Paired *t* Tests of Differences and Correlations Between Writings in First-Person and Third-Person Perspectives on Linguistic Categories.

Perspective	First-person, *M* (*SD*)	Third-person, *M* (*SD*)	$t_{(128)}$	*p*	*r*
Category					
DAV	8.03 (2.60)	8.45 (2.89)	−1.49	.14	.34***
IAV	9.55 (2.50)	10.22 (3.00)	−2.47	.02	.39***
SV	6.32 (2.58)	5.49 (2.70)	3.27	<.001	.41***
ADJ	5.25 (1.70)	5.57 (2.09)	−1.58	.12	.26**
Noun	13.46 (3.62)	15.67 (4.90)	−6.26	<.001	.59***
LCM score	3.15 (0.19)	3.20 (0.26)	−2.52	.01	.47***

Note. DAV = descriptive action verb; IAV = interpretative action verb; SV = state verb; ADJ = adjective; LCM = linguistic category model. Degrees of freedom = 1,793.
$^{**}p < .01$. $^{***}p < .001$.

Manipulation Check

The idea of language abstraction derived from CLT hypothesized that participants would use more abstract language in a third-person perspective than in a first-person perspective. To test this assumption, 50 text files were randomly selected from the first-person and third-person conditions, respectively. Two coders blind to the hypothesis received LCM human coding training and counted LCM words for five categories. Interrater reliability on the five categories ranged from .79 to .87, and the average reliability was .84, suggesting good interjudge reliability.

Using the LCM algorithm, composite LCM scores were calculated for first-person and third-person perspectives based on coders' rating. As predicted, participants in the third-person perspective used more abstract language ($M = 3.76$, $SD = 0.20$) than the first-person perspective ($M = 3.35$, $SD = 0.17$), $t_{(99)} = 11.01$, $p < .001$.

Results

The present study developed a computerized text analysis method for the LCM and sought to investigate if the method could distinguish different levels of abstraction. To examine differences in language abstraction between the two writing tasks, paired *t* tests were conducted on the five LCM categories and the LCM scores. The results are shown in Table 1. Students used more DAVs, IAVs, adjectives, and nouns when they were writing in a third-person perspective than when writing in a first-person perspective, and they used more SVs when writing in a first-person perspective than when writing in a third-person perspective. Results on the adjective and noun categories supported previous findings that people use more abstract language while writing in a third-person than in a first-person perspective. To test an overall LCM concept, we tested for the differences between the composite LCM scores for the two writing perspectives. As predicted, students had higher LCM scores when they were writing in a third-person perspective than in first-person perspective. These results suggest that overall, they used more abstract (concrete) language when writing in a third-person (first-person) perspective. In addition, the participants' LCM scores while writing in a first-person perspective were positively correlated with their LCM scores while writing in a third-person perspective ($r = .47$, $p < .001$), suggesting that our composite LCM scores can serve as an indicator of individual differences in regard to language abstraction.

Discussion

Building on the LCM, the present research developed a dictionary to identify verbs according to the LCM categories and used a POS tagging program to identify adjectives and nouns, which are the remaining categories of the LCM. These two approaches constitute a computerized text analysis method as an alternative approach to human coding of LCM categories.

In the study, we tested our computerized text analysis method with two writing tasks, including writing in first-person and third-person perspectives. As predicted, participants used more abstract language while writing in a first-person perspective and more concrete language while writing in a third-person perspective. These results not only replicate previous relevant findings about language abstraction but also indicate the practicability of using the computerized text analysis technique to study language styles.

The LCM is a theoretical framework that was developed to understand individuals' language use on social psychological processes (Semin & Fiedler, 1988, 1992). If there is a computerized method for the LCM coding system that can substitute for human coding, researchers can analyze their textual data more efficiently and process a large amount of texts at one time. Users should be cautious when using any computerized method for LCM because virtually all computerized methods are not context sensitive. Our word-search approach occasionally misclassifies words into an incorrect LCM category. For example, the word *love* is a state verb when it is used to describe individuals' mental states in the LCM; but *love* is also a noun when it is used to describe human affection. In the LCM coding procedure, coders can identify ambiguous words, such as *love*, based on the content in which the word occurs. However, with our computerized text analysis method, words are classified into one certain category according to our initial classification. Therefore, there is a practical limitation when we transform the LCM coding procedure to a computerized technique.

Is it acceptable to adopt a computerized technique to analyze texts? There are two reasons suggesting the promise of our computerized method. First, our method has been tested with a published data set in this research. The results of this study suggest that our method could be an alternative approach for studying language use in the LCM framework. Second, the purpose of developing this computerized text analysis method was to measure language abstraction in a fast and reliable way. According to our findings, this computerized method could distinguish language abstraction between different writing tasks. A subtle misclassification caused by ambiguous words did not affect the final results.

Our computerized text analysis method for the LCM categories has two advantages. First, it saves time and human resources for users. The traditional LCM coding procedure requires extensive training with multiple judges. With a computerized method, thousands of texts or millions of words can be processed in seconds. Because of this efficiency, researchers can increase sample sizes and boost statistical power for their studies. Second, composite LCM scores from our computerized method provide an opportunity to explore individual differences. In our correlational findings, people who used more abstract language in a first-person perspective were more likely to use abstract language in a third-person perspective, suggesting individual differences in chronic levels of language abstraction.

It is worth noting that evidence from the present study made a theoretical contribution to the LCM. In the LCM literature, few researchers draw conclusions from any individual category in the LCM instead of the composite LCM scores (e.g., Watson & Gallois, 2002). However, the direction of the SV category in our study was not what the LCM predicted. This may reflect the writing samples themselves. When people adopt a more natural first-person perspective, they might spontaneously demonstrate more cognitive processing, resulting in more use of SV words (e.g., think).[3] Interestingly, the composite LCM scores corresponded to what the LCM would predict. This result implies that the LCM is built on a broader concept that integrates five language categories, not just one particular category. Therefore, composite LCM scores may be more appropriate for representing the level of abstraction rather than one particular category of the five. Future researchers who adopt the LCM could focus on composite LCM scores instead of one LCM category. Of course, focusing on an individual LCM category might be also legitimate if researchers have specific assumptions to test.

There are two limitations in this research. Initially, some researchers tested the LCM without considering a noun category. In our study, the noun category was examined. Due to a smaller amount of discussion about the noun category in the LCM literature, future researchers should address the noun category more carefully if they focus particularly on nouns in the LCM. Second, the LCM was developed for testing interpersonal issues. However, verbs from our LCM dictionary, and adjectives and nouns from the POS tagging program, include human and nonhuman-related words. The direction of our findings might be the result of human issues in our study. To understand the external validity of this technique, it is important for future researchers to test this computerized text analysis method on nonhuman topics. In addition, we had limited human power to manually code texts, leading to a small sample size (N = 50). This is why we did not directly compare manual and automated LCM coding results. Future studies with sufficient human power could manually code more texts and compare manual and automated LCM coding results. There is a potential psychometric debate about the weights of the LCM coding schemes in this article. Unlike the original coding schemes (1: DAVs, 2: IAVs, 3: SVs, 4: adjectives, and 4: nouns; Semin et al., 2002), we reassigned a weight of 5 to the noun category to be compliant with previous research (e.g., Carnaghi et al., 2008). The purpose of our article was to provide a computerized text analysis method for the LCM, not to resolve the psychometric issue of the LCM coding schemes. The debate of the weights of the LCM coding schemes should be examined and discussed in the future.

The approach and findings are directly relevant to the current literature on personal perspectives and language abstraction. First, Libby et al. (2009) used visual perspectives to induce different levels of abstraction on verbal description. Similar to their strategy, we used writing perspectives to induce language abstraction. It is not easy to ensure if people indeed produce an assigned visual perspective with Libby et al.'s (2009) methods. With our writing strategy, we could investigate how deeply participants became personally involved in their writing by examining their personal pronoun use. Second, Fiedler, Jung, Wänke, Alexopoulos, and de Molière (2015) used linguistic terms to trigger mental constructs on psychological distance. Our research shows an opposite pattern on the relationship between linguistic terms and mental constructs. We demonstrated that manipulating mental constructs (e.g., personal perspective) can induce different levels of abstraction on linguistic terms.

Taken together, the findings underscore the potential of a computerized text analysis method to study LCM. The computerized text analysis method, which is an algorithmic computational approach employing natural language processing, might be a fruitful avenue for future research. Perhaps most exciting is that labs using a computerized approach will be able to directly benchmark their participants' language abstraction using a single measure and not relying on judges who have been trained independently across very different labs. Particularly exciting is that a computer-based approach will allow LCM researchers to test abstraction ideas in natural speech in conversations, books, speeches, and other text samples across cultures and languages.

Acknowledgments

We thank Kate Blackburn, Erin Burgoon, Ryan Boyd, and two anonymous referees for their helpful comments.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported in part by the Army Research Institute (W5J9CQ12C0043) and the Ministry of Science and Technology (MOST 103-2410-H-011 -038 -MY3).

Notes

1. The website link for TreeTagger: http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/

2. The website link for LIWC: http://www.liwc.net/
3. The correlation coefficients between our SVs and cognitive mechanism words from LIWC were .45 ($p < .001$) on the first-person perspective and .51 ($p < .001$) on the third-person perspective.

References

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX lexical data base on CD-ROM*. Philadelphia, PA: Linguistic Data Consortium.

Carnaghi, A., Maass, A., Gresta, S., Bianchi, M., Cadinu, M., & Arcuri, L. (2008). Nomina sunt omina: On the inductive potential of nouns and adjectives in person perception. *Journal of Personality and Social Psychology*, *94*, 839-859.

Eyal, T., Liberman, N., & Trope, Y. (2008). Judging near and distant virtue and vice. *Journal of Experimental Social Psychology*, *44*, 1204-1209.

Fiedler, K., Jung, J., Wänke, M., Alexopoulos, T., & de Molière, L. (2015). Toward a deeper understanding of the ecological origins of distance construal. *Journal of Experimental Social Psychology*, *57*, 78-86.

Freitas, A. L., Gollwitzer, P., & Trope, Y. (2004). The influence of abstract and concrete mind-sets on anticipating and guiding others' self-regulatory efforts. *Journal of Experimental Social Psychology*, *40*, 739-752.

Fujita, K., Henderson, M. D., Eng, J., Trope, Y., & Liberman, N. (2006). Spatial distance and mental construal of social events. *Psychological Science*, *17*, 278-282.

Kross, E., & Ayduk, O. (2011). Making meaning out of negative experiences by self-distancing. *Current Directions in Psychological Science*, *20*, 187-191.

Kross, E., Ayduk, O., & Mischel, W. (2005). When asking "why" does not hurt: Distinguishing rumination from reflective processing of negative emotions. *Psychological Science*, *16*, 709-715.

Libby, L. K., Shaeffer, E. M., & Eibach, R. P. (2009). Seeing meaning in action: A bidirectional link between visual perspective and action identification level. *Journal of Experimental Psychology: General*, *138*, 503-516.

Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology*, *75*, 5-18.

Louwerse, M. M., Lin, K., Drescher, A., & Semin, G. (2010). Linguistic cues predict fraudulent events in a corporate social network. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (pp. 961-966). Austin, TX: Cognitive Science Society.

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic Inquiry and Word Count: LIWC [Computer software]. Austin, TX: LIWC.net.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In D. B. Jones & H. L. Somers (Eds.), *Proceedings of the International Conference on New Methods in Language Processing* (pp. 44-49). Manchester, England: University of Manchester Institute of Science and Technology.

Schmid, H. (1995). Improvements in part-of-speech tagging with an application to German. In H. Feldweg & E. Hinrichts (Eds.), *Lexikon und Text* (pp. 47-50). Berlin, Germany: Walter de Gruyter.

Scott, M. (1996). *WordSmith*. New York, NY: Oxford University Press.

Seih, Y.-T., Chung, C. K., & Pennebaker, J. W. (2011). Experimental manipulations of perspective taking and perspective switching in expressive writing. *Cognition and Emotion*, *25*, 926-938.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, *54*, 558-568.

Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. In W. Stroebe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 2, pp. 1-30). Chichester, England: Wiley.

Semin, G. R., & Fiedler, K. (1992). The configuration of social interaction in interpersonal terms. In G. R. Semin & K. Fiedler (Eds.), *Language, interaction and social cognition* (pp. 58-78). London, England: Sage.

Semin, G. R., Görts, C., Nandram, S., & Semin-Goossens, A. (2002). Cultural perspectives on the linguistic representation of emotion and emotion events. *Cognition and Emotion*, *16*, 11-28.

Stone, P. J., Dunphy, D. C., Smith, M. S., & Ogilvy, D. M. (1966). *The General Inquirer: A computer approach to content analysis*. Cambridge: MIT Press.

Tausczik, Y., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*, 24-54.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*, 403-421.

Trope, Y., & Liberman, N. (2010). Construal-level theory of psychological distance. *Psychological Review*, *117*, 440-463.

Watson, B., & Gallois, C. (2002). Patients' interactions with health providers: A linguistic category model approach. *Journal of Language and Social Psychology*, *21*, 32-52.

Author Biographies

Yi-Tai Seih is an assistant professor for the Department of Business Administration at National Taiwan University of Science and Technology. He received his PhD at the University of Texas at Austin in 2013. His research focuses on the interplay between language and interpersonal relationships. His most recent research focuses on how employers perceive information from potential employees on workplace.

Susanne Beier received her PhD at the University of Heidelberg in 2013. Her research focused on aggression as well as social cognition.

James W. Pennebaker is a regents centennial professor for the Department of Psychology at the University of Texas at Austin. His most recent research focuses on the nature of language and social dynamics in the real world. The words people use serve as powerful reflections of their personality and social worlds.