International Journal of Educational Research and Development
www.allinnovationjournal.com
Online ISSN: 2664-7095
Received: 13-02-2026, Accepted: 16-03-2026, Published: 31-03-2026
Volume 8, Issue 1, 2026, Page No. 75-78

Corpus-based DDL instructional design: A case study on semantic differences between the Korean near-synonyms jamkkan and jamsi

Meixuan Li
Qingdao Hengxing University of Science and Technology, Qingdao, China

Abstract
This study aimed to examine the semantic distinctions and differences in usage tendencies between the Korean temporal adverb near-synonyms jamkkan and jamsi through a corpus-based analysis and, based on these findings, to propose vocabulary teaching methods employing Data-Driven Learning (DDL). To this end, collocates of the two lexical items were extracted from a large-scale corpus, and their combinational patterns, usage contexts, and semantic characteristics were investigated through an analysis of authentic examples. The results showed that, although both jamkkan and jamsi denoted a short period of time, they displayed different semantic tendencies in actual usage. Jamkkan tended to occur in discourse-oriented and interactional contexts, frequently collocating with lexical items related to conversation, communication, and brief actions, whereas jamsi more often co-occured with words associated with emotional, cognitive, and psychological states, thereby expressing a temporary pause or suspension. On the basis of these analytical findings, this study proposed a semantic differentiation framework and a five-stage instructional design that enables learners to observe and analyze authentic language data and to infer the semantic differences between near-synonyms independently. This study was significant in that it strengthens the empirical foundation of near-synonym instruction by applying corpus-based collocational analysis to Korean vocabulary education and presented a learner-centered and sustainable approach to vocabulary learning.

Keywords: Corpus-based analysis, Korean near-synonyms, semantic differentiation, data-driven learning (DDL), temporal adverbs

Introduction

Vocabulary is a core component of the linguistic system that is responsible for conveying meaning and plays a crucial role in determining both the accuracy of communication and the diversity of expression. In particular, near-synonyms often share a similar basic meaning while differing in such respects as usage context, collocational relations, and pragmatic function. Although these differences are relatively intuitive for native speakers, they often make it difficult for foreign language learners to judge the appropriate context of use. Therefore, research that systematically analyzes the subtle semantic differences and actual usage patterns of near-synonyms and proposes criteria for their differentiation is of great significance for both vocabulary education and semantic studies.

In educational settings, learners often experience difficulty in distinguishing and appropriately using lexically similar items. In many cases, when near-synonymous expressions are introduced, their semantic differences are explained primarily through dictionary definitions; however, dictionaries tend not to provide sufficiently clear distinctions between the meanings of near-synonyms [1]. The meanings of near-synonyms are revealed more clearly in actual language use and become more evident through collocational relations and contextual features. Therefore, explanations based solely on definitions are insufficient to account fully for the subtle semantic differences between near-synonyms. For this reason, an approach that analyzes the meanings and usage tendencies of near-synonyms on the basis of authentic usage data is required.

This study focuses on the temporal adverb near-synonyms jamsi and jamkkan, both of which denote “a short time”, with the aim of clarifying the semantic differences and usage tendencies between the two lexical items and, furthermore, of designing pedagogical methods based on learner-centered Data-Driven Learning (DDL). Such an approach enables learners to directly observe and analyze authentic language data, thereby allowing them to discover and construct linguistic rules and meanings for themselves and to engage in sustainable vocabulary learning.

Theoretical Background and Significance

1. The Concept and Characteristics of DDL

DDL is a learning approach in which learners independently discover linguistic rules and meanings through the direct observation and analysis of authentic language data [2]. Unlike traditional teacher-centered instructional methods, DDL requires learners to explore language data themselves and to induce rules through an inductive process, thereby assuming the role of “language detectives”. Biber *et al*. explained that analyzing authentic language-use data plays an important role in understanding actual language use patterns and emphasized that corpus data can also be effectively utilized in language education [3]. This perspective is closely connected to the DDL approach, which highlights the importance of authentic data in language learning. Lee argued that the analysis of large-scale corpus data makes it possible to examine the collocational relations of lexical items as they appear in actual contexts of use and, through this, to identify more clearly their combinational patterns and usage tendencies. Accordingly, she emphasized that collocational analysis and the observation of usage contexts play a crucial role in enabling learners to distinguish semantic differences between near-synonyms on their own. Therefore, language education using corpus data is regarded as an effective method for helping learners understand language patterns as they occur in real usage environments [4].

2. Advantages and Significance of Data-Driven Near-Synonym Instruction

DDL-based instruction is an approach in which learners construct meaning by directly observing and analyzing authentic language data, and in this respect it differs fundamentally from traditional vocabulary teaching centered on teachers' explanations. In particular, the pedagogical advantages and significance of DDL in near-synonym instruction may be discussed in the following respects.

First, it deepens semantic awareness through learner-centered observation activities. Whereas traditional learning methods rely on one-way transmission of meaning from the teacher, data-driven instruction enables meaning to be formed through learners' own observation and cognition. Learners do not simply memorize meanings; rather, by directly observing authentic contexts of use, they recognize for themselves the combinational patterns, emotional tone, and pragmatic information associated with the target lexical items. Through this process, meaning is internalized and transformed into long-term memory [5].

Second, it promotes the recognition of near-synonym meaning in terms of "tendencies" rather than fixed rules. In actual language use, the selection of near-synonyms is determined not by strict rules but by frequency, collocational tendencies, and discourse conventions. Data-driven instruction in near-synonym meaning guides learners to understand semantic differences not as absolute rules but as relative tendencies. As a result, it fosters the ability to apply lexical knowledge more flexibly in authentic discourse situations.

Third, it strengthens ability of learners to judge pragmatic appropriateness. A central issue in the use of near-synonyms is contextual appropriateness. In particular, pragmatic differences are difficult to convey through explanation or definition alone. DDL instruction enables learners to observe discourse patterns that recur across large amounts of authentic language data, thereby allowing them to intuitively acquire the pragmatic functions of near-synonyms [6].

Fourth, it contributes to the enhancement of learner autonomy and metalinguistic awareness. DDL-based instruction involves a process in which learners formulate questions, establish hypotheses, and verify them independently. Such learning experiences not only strengthen learner autonomy but also promote metalinguistic awareness regarding lexical usage. This awareness can be transferred into a learning strategy through which learners are able to explore semantic differences independently even when encountering new near-synonyms. In this sense, DDL is significant in that it cultivates sustainable vocabulary learning abilities that go beyond short-term vocabulary acquisition.

This study seeks to compare and analyze the Korean temporal adverb near-synonyms jamsi and jamkkan in order to clarify the semantic differences and differences in usage tendencies between the two lexical items and, furthermore, to design DDL-based pedagogical methods. To this end, the study first collects the contexts in which jamsi and jamkkan occur and examines their usage patterns by analyzing their collocational relations and contextual features. These two lexical items are frequently used in everyday discourse and often appear to be interchangeable, making them a group of near-synonyms that foreign learners find difficult to distinguish. By systematically presenting the actual usage patterns of jamsi and jamkkan through corpus-based analysis, this study is expected to contribute to expanding the empirical foundation of Korean near-synonym research. The findings of the study may also be utilized as pedagogical materials to help learners of Korean recognize the semantic differences between the two lexical items more accurately and clearly and to use them appropriately in relevant contexts.

Research Method

This study aims to extract collocates on the basis of the most recently compiled spoken and written data from the National Institute of Korean Language. The method of collocate extraction involves downloading the corpus data, entering the target lexical items, and extracting all relevant sentences. Subsequently, the lexical analysis program AntConc 3.4.9 was used to extract collocates occurring within a span of six spans on either side of the target word and appearing with a frequency of 30 or more occurrences. In order to confirm the semantic relevance of the extracted collocates to the target lexical items, the relevant sentences were further analyzed so as to ensure the accuracy of collocate extraction. Through corpus analysis, this study seeks to secure an empirical basis for designing pedagogical methods grounded in semantic differentiation and tendencies in lexical usage. In what follows, the collocates extracted from the corpus are presented in order of frequency, and the tendencies in semantic usage are examined in order to clarify the semantic distinction between the two lexical items.

Semantic Differentiation of Near-Synonyms

The frequencies of the collocates of jamkkan extracted from the corpus are presented in Table 1 below.

Table 1: Rankings of collocates of jamkkan

No.	Lexical Item	Frequency	No.	Lexical Item	Frequency
1	talk	3267	31	meet	150
2	during	2302	32	sleep	140
3	remark	2204	33	body	134
4	see	1838	34	film	124
5	listen	1212	35	utterance	105
6	we	1112	36	text message	94
7	hesitate	872	37	greeting	92
8	thought	832	38	mobile phone	86
9	wait	823	39	seat	85
10	story	690	40	move on	78
11	explanation	658	41	conversation	77
12	mention	560	42	come out	74
13	head	477	43	mood	73
14	eye	452	44	content	72
15	between	431	45	appearance	71
16	enter	420	46	silence	71
17	screen	418	47	sound	70
18	rest	396	48	photograph	69
19	time	322	49	restroom	67
20	work	300	50	facial expression	59
21	video	274	51	question	56
22	introduction	261	52	rest	55
23	organize	240	53	stop by	51
24	home	230	54	lie down	51
25	stop	211	55	tea	50
26	face	203	56	evening	49
27	interview	202	57	these days	47
28	telephone	168	58	company	46
29	song	164	59	position	46
30	phone call	154	60	nearby	45

An examination of the frequency and distribution of the collocates of jamkkan first reveals that it occurs with high frequency in discourse-oriented and communicative situations. It appears in interpersonal conversational contexts frequently. It is commonly used in situations in which an ongoing action or circumstance is temporarily interrupted, or in which a certain action is carried out for a short period of time. In other words, it tends to occur in external, actual conversational situations or events. The following examples are presented in order to examine its meaning in greater depth.

1. If you have time, let's meet for a moment.
2. I had a brief conversation with a friend.
3. After taking a short nap, I felt better.
4. I'll go to the restroom for a moment.

As can be seen from the examples above, expressions such as jamkkan iyagireul hada (have a brief conversation), jamkkan mannada (meet briefly), jamkkan natjameul jada (take a short nap), and jamkkan hwajangshireul gada (go to the restroom for a moment) all denote performing an action for a short period of time in concrete and dynamic event situations. Thus, jamkkan may be interpreted as conveying the meaning of temporarily stepping away in the midst of an ongoing event. Next, the frequency ranking of the collocates of jamsi will be examined (Table 2).

Table 2: Rankings of collocates of jamsi

No.	Lexical Item	Frequency	No.	Lexical Item	Frequency
1	after	15675	31	composure	148
2	thought	2887	32	greeting	134
3	during	2371	33	interruption	73
4	later	1185	34	stillness	76
5	breath	870	35	hesitate	272
6	silence	1293	36	glance	73
7	speech	3667	37	distance	53
8	stop	1517	38	rest	594
9	hesitate	916	39	mood	191
10	facial expression	897	40	memory	70
11	time	1038	41	video	42
12	head	1092	42	step	254
13	appearance	327	43	alone	164
14	story	318	44	leg	79
15	mind	520	45	sky	118
16	home	299	46	surroundings	110
17	seat	758	47	mention	35
18	telephone	444	48	laughter	35
19	eye	331	49	crying	35
20	gaze	343	50	situation	35
21	step	254	51	walk	48
22	smile	261	52	prayer	49
23	mind/spirit	363	53	bed	47
24	explanation	177	54	movement	42
25	conversation	272	55	outside the window	39
26	answer	171	56	break away	38
27	question	175	57	falter	37
28	ear	128	58	mistake	28
29	face	131	59	hospital	27
30	breathing	106	60	caution	26

The collocational characteristics of jamsi indicate that it is frequently associated with psychological or emotional states, while also co-occurring with external phenomena. In collocations such as jamsi chimmukhada (remain silent for a moment) and jamsi pyojeongi gudeo itda (one's expression stiffens for a moment), jamsi may be regarded as favoring abstract domains and as being closer to a static meaning. Its collocational features reveal an abstract domain involving a temporary emotional or cognitive pause on the part of a human experience. The following examples are examined in greater detail.

1. It seemed that he had briefly regained control of his emotions, but then it started again.
2. I am trying to calm my mind for a moment.
3. I was taking a brief rest.
4. It seems that the shop was closed for a while.

As can be seen in Examples (1) and (2), jamsi occurs in sentences that express emotional states. In Examples (3) and (4), it collocates with hyusik (rest) and gage (shop), conveying the meaning of a temporary suspension of a situation; in these contexts, a more [+extended] semantic feature can be observed in comparison with jamkkan. Therefore, jamsi can be distinguished from jamkkan in that it carries more abstract and temporally extended semantic features. The results of the analysis of the collocates and contexts of the two lexical items may be summarized in Table 3 below.

Table 3: Semantic differentiation of the near-synonyms jamkkan and jamsi

Category	Jamkkan	jamsi
Domain of Use	External actions	Internal states
Context of Use	Conversational situations	Emotional and psychological situations
Type	Interaction	Temporary pause
Duration	Short	Relatively longer
Expression	talk	think

A comparison of the collocates extracted from the corpus revealed that jamkkan and jamsi exhibit different semantic tendencies in their collocational patterns. Jamkkan frequently co-occurs with discourse-related lexical items such as yaegi (talk), malsseum (remark), iyagi (story), seolmyeong (explanation), and jilmun (question), and it tends to be used primarily in communicative and interactional contexts with words such as jeonhwa (telephone), tonghwa (phone call), boda (see), and deutda (listen). In contrast, jamsi combines with lexical items such as saenggak (thought), maeum (mind), sum (breath), chimmuk (silence), meomchuda (stop), and mangseolida (hesitate), and is often used to indicate psychological states or the temporary suspension of an action. These differences in collocational patterns demonstrate that, although both lexical items denote a short period of time, they display different semantic tendencies in actual contexts of use. That is, jamkkan more often denotes a brief action performed in a discourse situation, whereas jamsi shows a stronger tendency to express a temporary pause occurring within emotional or cognitive states.

Design of Pedagogical Methods

This study analyzed the collocates and usage contexts of the near-synonyms jamsi and jamkkan using corpus data and, on the basis of the analytical results, proposed a DDL-based instructional approach through which learners could independently discover semantic differences by observing and analyzing authentic language data. Since its introduction by Tim J., DDL had developed into a learning approach in which learners explored corpus data and independently identified patterns of language use. Subsequent studies by scholars such as Guy Aston and

Sylviane Granger presented concrete vocabulary and grammar learning activities using corpus data and reported that such approaches were effective in enhancing learners' active participation and their understanding of actual language use patterns [7]. However, these previous studies mainly focused on the exploration of corpus data and the discovery of patterns and thus showed limitations in providing concrete analytical criteria for systematically distinguishing the subtle semantic differences between near-synonyms. Accordingly, this study organized near-synonym learning into the following five stages, as shown in Table 4 below.

Table 4: DDL-based instructional procedure for near-synonym education

Stage	Purpose of the Stage	Teacher's Role	Learner's Role
Presentation Stage	To recognize the problem of near-synonym usage	Collect materials	Recognize the problem
Observation Stage	To observe and explore authentic language data	Present corpus data	Observe usage examples
Analysis Stage	To analyze the data systematically and derive meaning	Guide the analysis	Classify collocates and infer meaning
Verification Stage	To verify the derived semantic hypotheses and revise them	Provide additional data	Verify and revise semantic hypotheses
Application Stage	To understand semantic differences and apply them to actual language use situations	Present extension activities	Create sentences and practice vocabulary use

The presentation stage was the stage in which learners became aware of the usage problem related to the target lexical items. At this stage, example sentences containing the near-synonym pair were presented so that learners were guided to judge the semantic difference between the two lexical items intuitively. This was the stage in which learners began to develop an awareness of the linguistic phenomenon.

The observation stage was the stage in which learners observed authentic examples extracted from the corpus and explored the usage contexts of the target lexical items. Through authentic language data, learners were able to recognize naturally the usage patterns, situational contexts, and pragmatic information associated with the lexical items.

The analysis stage was the stage in which learners, following the teacher's guidance, extracted collocates and classified them according to semantic categories in order to analyze and understand meaning [8]. By observing the words that co-occurred before and after the lexical items in authentic language data and by drawing inductive generalizations, learners attempted to derive meaning. In addition, the teacher provided a large number of sentences so that learners could analyze semantic differences within sentence contexts.

The verification stage was the stage in which learners confirmed and verified the semantic hypotheses derived in the previous stage by using new data. In the process of applying the results of their analysis to actual sentences, learners were able to examine the validity of the semantic rules they had formulated.

The internalization stage was the stage in which learners applied and used the semantic differences between near-synonyms in actual language situations, thereby strengthening their language-use ability.

Conclusion

This study attempted to distinguish the meanings of the Korean near-synonyms jamkkan and jamsi through the use of corpus data and, in addition, designed a DDL-based instructional approach. Unlike traditional methods that rely on teachers' explanations, this approach enables learners to analyze and explore meaning independently through authentic language data, and in the process of naturally understanding usage contexts, usage tendencies, and pragmatic information associated with lexical items, it can realize learner-centered learning. Therefore, the DDL-based approach is pedagogically significant in that it encourages learners to actively explore and understand lexical meaning, improves their ability to distinguish meanings, and enables sustainable vocabulary learning.

References

1. Kim K. A Study on strategies for refining lexical definitions through semantic comparison of synonyms. The Journal of Korean Language and Literature,1998:26:5-40.
2. Johns T. Should you be persuaded-Two samples of data-driven learning materials. English Language Research Journal,1991:4:1-16.
3. Biber D, Conrad S, Reppen R. Corpus linguistics: investigating language structure and use. Cambridge University Press, 1998.
4. Aston G. Enriching the learning environment: Corpora in ELT. Teaching and Language Corpora, 1997.
5. Aston G. Learning with corpora. Athelstan Press, 2001.
6. Granger S. The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. TESOL Quarterly,2003:37:538-546.
7. Gang H. The theory and practice of Korean vocabulary education research. The Language and Culture,2013:9:1-38.
8. Li M. A research on Education of Korean emotional adjective synonyms for Chinese learners. PhD thesis of Inha University, 2024.