

# 학습자 말뭉치 - 형태 주석 말뭉치 네트워크 분석

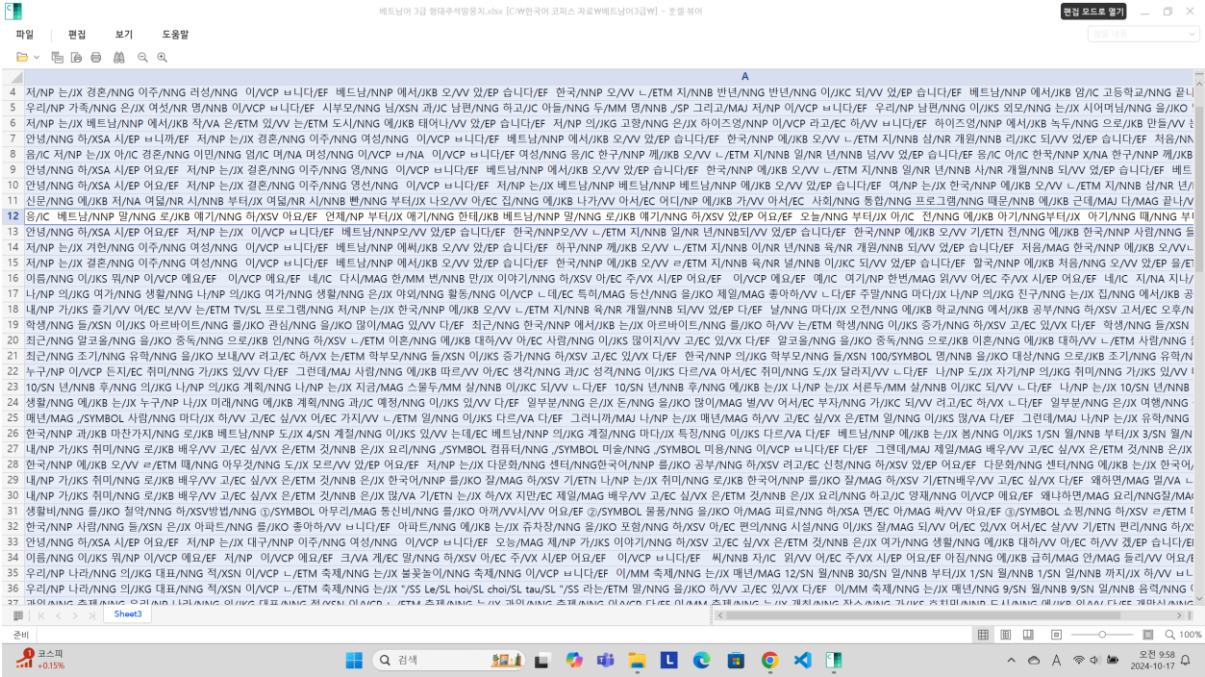
목적 : 베트남어 사용 3급 한국어 학습자의 내용어휘 사용 양상 분석

대상 : 베트남어 사용 3급 한국어 학습자, 100명

방법 : 텍스트마이닝 키워드네트워크 분석

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
69	ST	ST	SP																					
70	ST	MDV	ST	DC	POS	ST	ST	POS	SC	SC	ST	ST	DC	ST	ST									
71	ST	MDC																						
72	MCJ	MCJ	ST	MDV	POS	MCJ	MCJ	SN	MCJ	MCJ	WO	SN	ST	ST	MDV	POS								
73	MCJ	MCJ	MCJ	SN	ST	MCJ	MCJ																	
74	DC	WO	WO	SH	MDC	POS	SC	MCJ	MCJ	MCJ	MDV	MCJ												
75	MDC	ST	ST	DC	MDC	SH																		
76	SH	ST	SN	SP	WO	WO	MCJ	SP																
77	MCJ	ST	ST	ST	DS	ST	ST	SP	MCJ															
78	MCP	MDC	MCJ	MCJ	MCJ	MCJ																		
79	WO	MCJ	MCJ	MCJ	MCJ	MCJ																		
80	ST	ST																						
81	ST	MDV	ST	SN	DS																			
82	SH	SH	SH	SH	ST	SH																		
83	DS	ST	ST	POS	SH	DC	DS																	
84	DS	SH	SH	SH	SH	DS	SH	SH	SH	MCP														
85	MDV	POS	MCJ	MCJ	DS	ST	MDV	POS	ST	ST	MCJ	MCJ												
86	MDC	DS	ST	ST	ST																			
87	DS	DS	SC	SC																				
88	SC	DS	MDV	POS	MCJ	POS	MCJ																	
89																								
90	MCJ	MCJ	MCJ	MCJ	MCP	MCP	ST	SH	MCJ	MCJ	ST	SH	ST	POS	MCJ	MCJ								
91	DC	DS	WO	MDV	POS	ST	ST	ST	ST															
92	ST	ST	ST	SN	DR																			
93	DS	POS	ST	MDV	ST	MCJ	ST	MDV	ST	MDV	ST	WO	WO	WO	ST	ST	ST	DS	ST	ST	SH	MCJ		
94	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP
95	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP	PP
96	ST	SH	DS																					
97	MDC	DS	POS	DS	DS	ST	MDC	ST	DC	ST	MCJ	WO	POS	WO										
98	SH	WO	SC	ST	DR																			
99	ST																							
100	ST	POS	MCJ	MCJ	ST																			
101																								
102																								

<베트남어 3급 오류주석말뭉치>



## <베트남어 3급 형태주석말뭉치>

특별/NNG 하/XSA ㄴ/ETM 손님/NNG 빨리/MAG 께/VV 어/EF 빨리/MAG !/SYMBOL 엄마/NNG ,/SYMBOL 지금/MAG 다섯/NR 시/NNB 이/VCP ㄴ다/EF 나/NF JX 모르/VV ㄴ다/EF 그냥/MAG 이/MM 동네/NNG 예/JKB 있/VV 다/EF 다른/MM 사람/NNG 주/VV ㄴ/ETM 음식/NNG 이나/JC 땀/NNG 을/JKO 먹/VV 고/EC 다 /G/MM 이모/NNG 의/JKG 죽음/NNG 때문/NNB 예/JKB 씨/NNB 가/JKS 그/MM 글/NNG 을/JKO 쓰/VV 게/EC 되/VV 있/EP 째/NA 아요/EF 예/IC 이모/NNG NNB 랑/JKB 쫘/MAG 다크/VA ㄴ/ETM 마음/NNG 도/JX 있/VV 어요/EF 네/IC 저/NNP 저/NP ㄴ/JX 슬프/VA ㄴ데/EC 뭐/NP 가/JKS 슬프/VA ㄴ데/EF 그냥/MAG 의/JKG 기분/NNG 이/JKS 어떻/VA 있/EP 는지/EF 음/IC 그래서/MAJ 이모/NNG 의/JKG 죽음/NNG 이랑/JKB 씨/NNB 가/JKS 삶/NNG 의/JKG 흔적/NNG 이/V M 공부/NNG 랑/JKB 관련/NNG 되/XSV 는/ETM 것/NNB 도/JX 씨/NNB 가/JKS 살아오/VV ㄴ/ETM 삶/NNG 이/VCP 째/NA 아요/EF 맞/VV 아요/EF 그래서/M /JKB 는/JX 뭐/NP 이/VCP 예요/EF 그림/MAJ 나/NP 는/JX 열심히/MAG 공부/NNG 하/XSV 있/EP 어요/EF 가끔/MAG 은/JX 애기/NNG 도/JX 울/VV 고/EC 남편 /VA 고는/EC 쓰/VV 먼/EC 안/MAG 되/VV 지/EF 그쵸/IC 이모/NNG 는/JX 돌아가/VV 하/EP 있/EP 는데/EC 그/MM 이모/NNG 는/JX 자기/NP 가/JKS 하/VV 고 것/NNB 처럼/JKB 나/NP 는/JX 결혼/NNG 하/XSV 고/EC 이럴/VA 게/EC 열심히/MAG 살/VV 있/EP 어요/EF 얼마/NNG 전/NNG 예/JKB 이모/NNG 가/JKS 죽/VJ 저/NNP 는/JX 갯혼/NNG 이주/NNG 여성/NNG 애요/EF 저/NNP 는/JX 아/IC <name2>/NNP 에서/JKB 오/VV 있/EP 어요/EF 한국/NNP 예/JKB 오/VV ㄴ/ETM 저 /IC NNB 있/VV 는데/EC 한국/NNP 는/JX 한복/NNG 한복/NNG 이/VCP 라고/EC 작/NA 음/IC 뭐/NP 이/VCP 지/EF 음/IC 화영/XR 하/XSA 고/EC 이쁘/VA 고/EC 찬/NNG 뭐/NP 가/JKS 있/VV 을까요/EF 우리/NP 애기/NNG 든/JX 특히/MAG 케란말이/NNG 케란마니/NNG 니나/JX 안/MAG 그/VV 먼/EC 주욱/NNG 주욱밤/ 복잡/XR 하/XSA 고/EC 문장/NNG 도/JX 만/VA 꼬/EC 또/MAG 정이/MAG 하/XSV 는/ETM 것/NNB 도/JX 많/VA 아요/EF 음/IC 우리/NP 나라/NNG 는/JX 이사 우리/NP 가족/NNG 은/JX 어섯/NR 땀/NNB 이/VCP ㄴ다/EF 시부모/NNG 남/XSN 고/JC 남편/NNG 하고/JC 아들/NNG 두/MM 땀/NNB ,/SP 그리고/MAJ 저/ 저/NNP 는/JX 베트남/NNP 에서/JKB 작/VA 은/ETM 있/VV 는/ETM 도시/NNG 예/JKB 태어나/VV 있/EP 습니다/EF 저/NNP 의/JKG 고향/NNG 은/JX 하이즈영/NNF 저/NNP 는/JX 오랫동안/NNG 고향/NNG 예/JKB 못/MAG 가/VV 있/EP 는데/EC 도/JX 좋/VA 은/ETM 추억/NNG 이/JKS 많이/MAG 있/VV 고/EC 부모/NNG 남/X 안병/NNG 하/XSA 시/EP 어요/EF 저/NNP 는/JX 결혼/NNG 이주/NNG 영/NNG 이/VCP ㄴ다/EF 베트남/NNP 에서/JKB 오/VV 있/EP 습니다/EF 한국/NNP 예

## <포함어 사전1>

특별/NNG  
하  
ㄴ  
손님/NNG  
빨리/MAG  
깨/VV  
어  
  
엄마/NNG  
  
지금/MAG  
다섯/NR  
시/NNB  
이/VCP  
ㄴ 다  
  
나/NP  
는  
줄리/VA  
고서  
를  
도와주/VV  
우리/NP  
준비/NNG  
하  
아야  
하/VX

#### < 포함어 사전2>

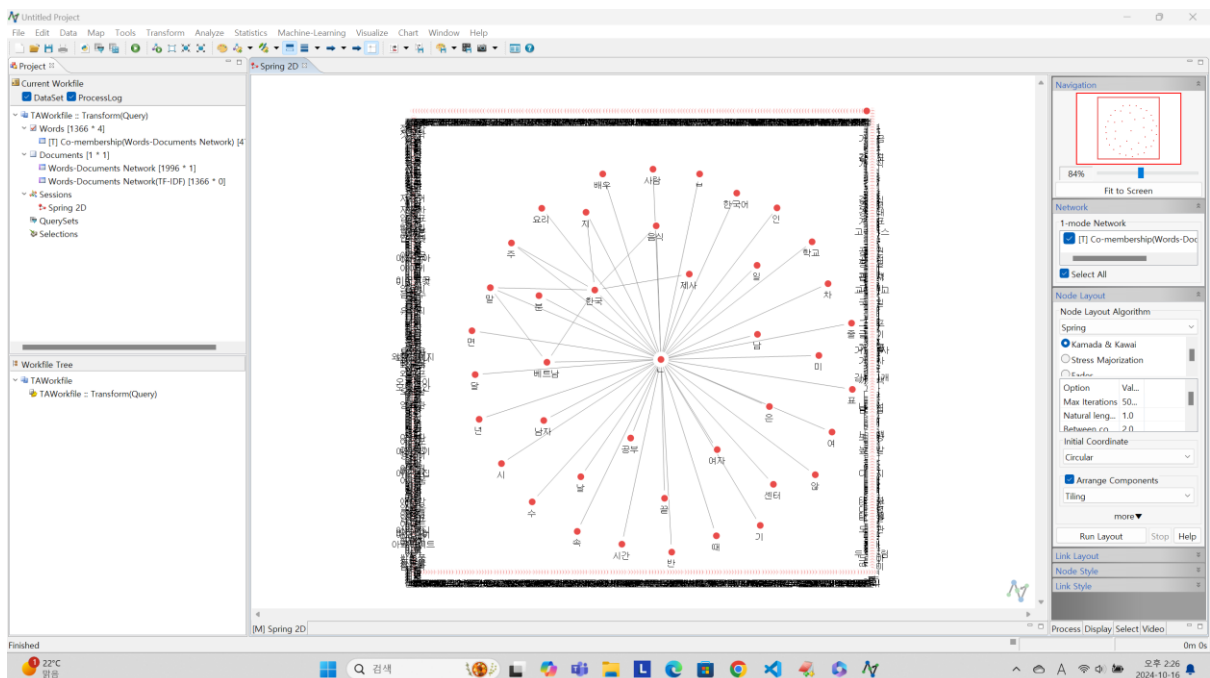
포함어 사전 제작 과정을 정리하면 다음과 같습니다:

1. **텍스트 합치기:** '베트남어 3급 오류주석말뭉치'와 '베트남어 3급 형태주석말뭉치' 두 파일의 텍스트 데이터를 통합했습니다. 이 과정에서 두 파일의 내용을 병합하여 사전을 구성했습니다.
2. **오류 파일 제거:** 데이터 통합 후 오류가 포함된 파일이나 항목은 분석하여 제거했습니다. 이를 통해 잘못된 정보가 포함된 데이터를 걸러냈습니다.
3. **중복 항목 제거:** 통합된 텍스트 내에서 중복된 항목을 찾아 삭제하였으며, 이를 통해 사전에 고유한 항목만 남기고 중복 데이터를 제거했습니다.

이러한 과정을 거쳐 최종적으로 오류 없이 중복되지 않는 포함어 사전을 완성하게 되었습니다.

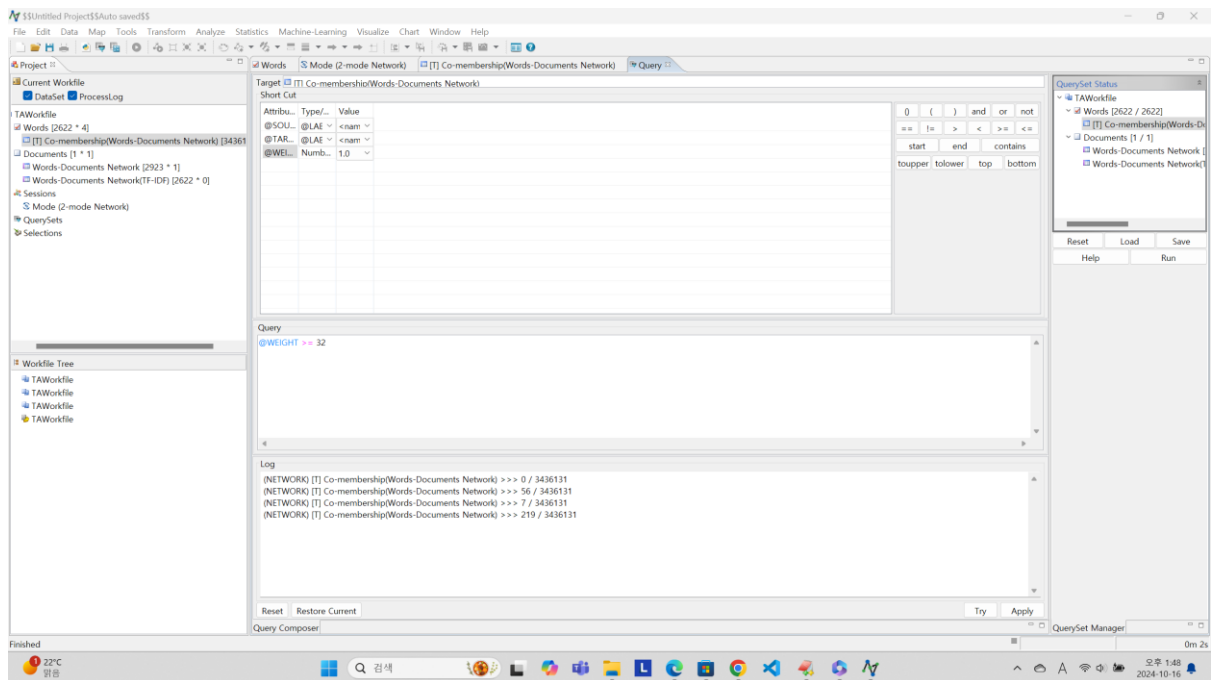
id of Speech	Frequency	Word length	Name Type
1	1	1.0	1.0
2	1	1.0	1.0
3	1	1.0	1.0
4	1	1.0	1.0
5	1	1.0	1.0
6	1	1.0	1.0
7	1	1.0	1.0
8	1	1.0	1.0
9	1	1.0	1.0
10	1	1.0	1.0
11	1	1.0	1.0
12	1	1.0	1.0
13	1	1.0	1.0
14	1	1.0	1.0
15	1	1.0	1.0
16	1	1.0	1.0
17	1	1.0	1.0
18	1	1.0	1.0
19	1	1.0	1.0
20	1	1.0	1.0
21	1	1.0	1.0
22	1	1.0	1.0
23	1	1.0	1.0
24	1	1.0	1.0
25	1	1.0	1.0
26	1	1.0	1.0
27	1	1.0	1.0
28	1	1.0	1.0
29	1	1.0	1.0
30	1	1.0	1.0
31	1	1.0	1.0
32	1	1.0	1.0
33	1	1.0	1.0
34	1	1.0	1.0
35	1	1.0	1.0
36	1	1.0	1.0
37	1	1.0	1.0
38	1	1.0	1.0
39	1	1.0	1.0
40	1	1.0	1.0

<1,366개 어휘 빈도순으로 나열하기>



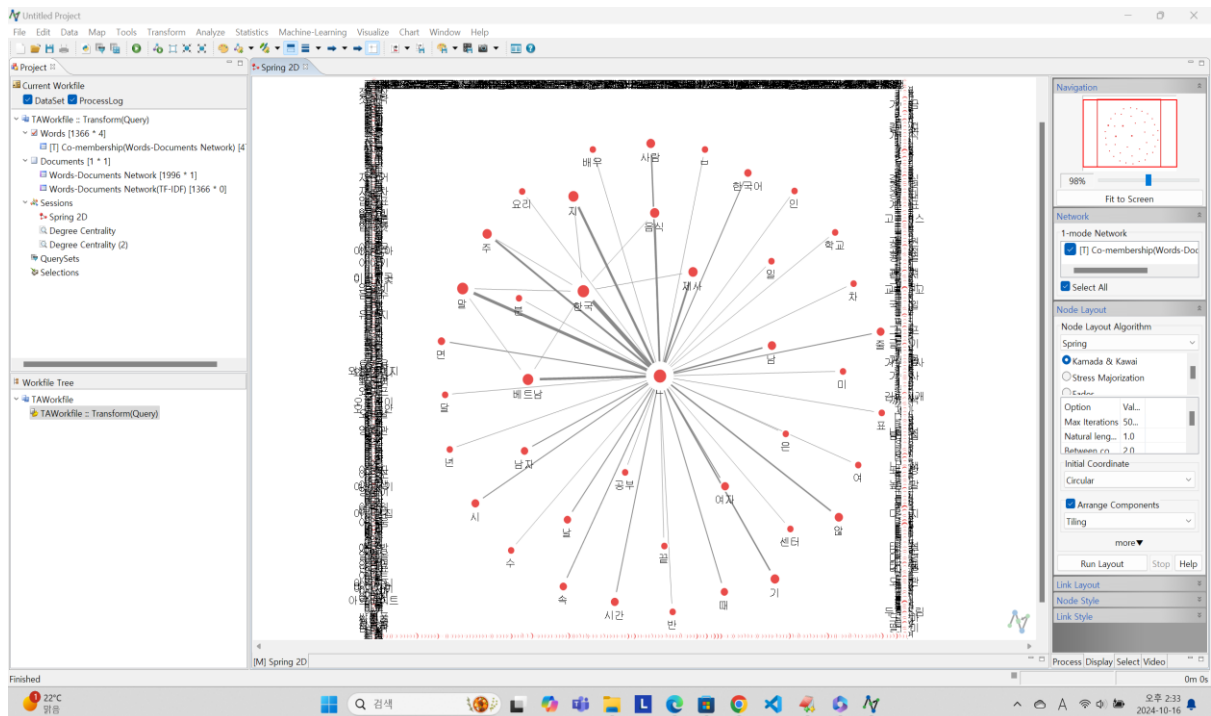
<자료 가공 : 3,436,131개의 어휘쌍 분석>

어휘간의 유사도(동시출현 정도)를 바탕으로 상관성 분석 -> 수가 너무 많음!!



<자료 가공 : 3,436,131 -> 219개로 축소>

유사도 32이상의 노드 219개로 네트워크 축소



<자료 가공 : 유사도 32이상의 노드 219개로 네트워크 지도 그리기>

이 네트워크 지도는 베트남어를 사용하는 3급 한국어 학습자들이 자주 사용하는 주요 내용어들이 어떻게 연결되어 있는지 시각화한 결과입니다. 주로 빈도 높은 단어들을 중심으로 단어들 간의 연결 강도가 나타나며, 군집화를 통해 학습자들이 어떤 주제에 대해 이야기했는지 파악할 수 있습니다.

1. **중심 노드:** 중앙에 위치한 큰 빨간색 노드는 가장 빈도가 높은 단어를 나타내며, '하다'와 같은 고빈도 동사일 가능성이 큼니다.
2. **주요 연결:** 중심 노드에서 뻗어나가는 연결선의 굵기는 단어들 간의 연결 강도를 나타내며, 굵을수록 자주 함께 사용된 단어들입니다.
3. **주요 단어:** '한국어', '베트남', '사람', '음식', '학교', '시간', '공부' 등이 네트워크 상에서 중요한 단어로 나타나며, 이들은 학습자들이 자주 사용한 내용어휘입니다.
4. **군집화:** '한국어', '공부', '학교'와 같은 단어들이 가까운 위치에 군집화되어 있으며, 관련성이 높은 단어들끼리 묶여 있는 것을 보여줍니다.
5. **연결 구조:** 방사형 패턴으로 단어들이 하나의 그룹으로 묶여 있으며, 학습자들이 한국과 베트남, 혹은 그와 관련된 주제를 중심으로 자주 사용한 어휘들이 밀접하게 연결되어 있습니다.