

О КОРПУСЕ ИЗУЧАЮЩИХ КОРЕЙСКИЙ ЯЗЫК

Ховрова Сахаяна Николаевна

магистрант, Северо-Восточный федеральный университет им. М.К. Аммосова, РФ, г. Якутск

ABOUT KOREAN LEARNER'S CORPUS SEARCH ENGINE

Sakhaiana Khovrova

Master student, Northeastern Federal University named after M.K. Ammosov, Russia, Yakutsk

Аннотация. Данная статья посвящена корпусу изучающих корейский язык как иностранный (ККИ - Korean Learner's Corpus Search Engine). В статье рассматривается структура сайта корпуса, описываются материалы корпуса, принципы поиска словоупотреблений на сайте.

Abstract. This article focuses on the Korean Learner's Corpus Search Engine (ККИ - Korean Learner's Corpus Search Engine). The article examines the structure of the corpus site, describes the corpus materials, principles of word usage search on the site.

Ключевые слова: корпусная лингвистика, корпус корейского языка, корейский язык.

Keywords: corpus linguistics, corpus of Korean language, Korean language.

В данной статье рассматривается корпус изучающих корейский язык как иностранный язык (ККИ - Korean Learner's Corpus Search Engine), который создан на базе Национального института корейского языка (НИКЯ). Данный корпус доступен в сети интернет по адресу: <https://kscorpus.korean.go.kr/> и открыт для свободного доступа.

Корпус изучающих корейский язык (ККИ - Korean Learner's Corpus Search Engine) – это языковой материал, который был собран из текстовых и разговорных материалов иностранцев, изучающих корейский как второй иностранный язык.

Формирование корпуса. Национальный институт корейского языка разработал данный корпус в 2015 году и в течение 5 лет исследовал материалы корпуса. Главным источником текстов для корпуса стали тексты, созданные иностранцами, изучающие корейский язык как второй иностранный язык. Тексты корпуса были собраны и компьютеризованы в данный корпус. В текстах допустимы грамматические и орфографические ошибки, так как тексты были созданы иностранцами, которые изучают корейский язык.

Данный корпус, собранный на основе текстов иностранцев, изучающих корейский в Корее и за рубежом, служит материалом в исследованиях и разработке учебных материалов для корейских и зарубежных исследователей и преподавателей корейского языка. Также

созданный корпус может использоваться в систематизации словаря корейского языка и приложений для изучения корейского языка.

Сайт корпуса состоит из двух вкладок, в которых представлена поисковая система корпуса (□□) и статистика корпуса (□□). Также сверху представлена информация о корпусе (□□□) и вкладка загружаемых разработчиками инструкций и открытых материалов (□□□□) (Рис.1).

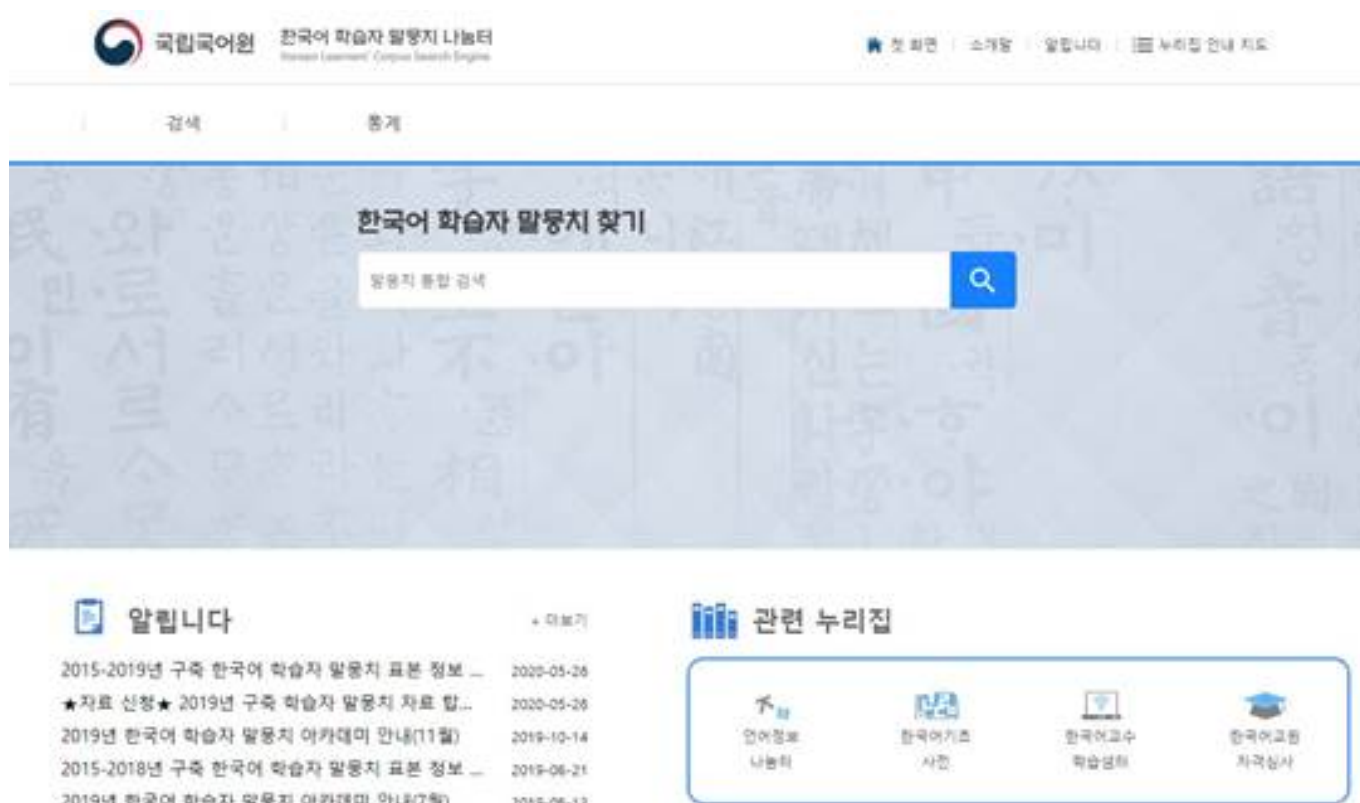


Рисунок 1. Внешний интерфейс корпуса изучающих корейских язык

Статистика. Для создания данного корпуса были собраны материалы представителей 142 стран, которые изучают корейский как иностранный язык. Представители данных 142 стран говорят на 93 языках.

На текущий момент в базе данных корпуса присутствует массив текстов 3 784 091 словоупотреблений, из них текстовые – 2 952 566 слов, разговорных – 831 525 слов. Общее количество документов – 26 152 документов, из них текстовых – 24 342 (93,1%) документов, разговорных – 1 810 (6,9%) документов.

В разделе статистика разработчики распределяют словоупотребления по представителям стран (количество в документах): Китай – 10 243 (39,2%), Япония – 3 407 (13,0%), Вьетнам – 1 945 (7,4%), Гонконг – 1 091 (4,2%), Тайвань – 877 (3,4%), США – 799 (3,1%), другие страны – 7 790 (29,8%) (Рис.2).

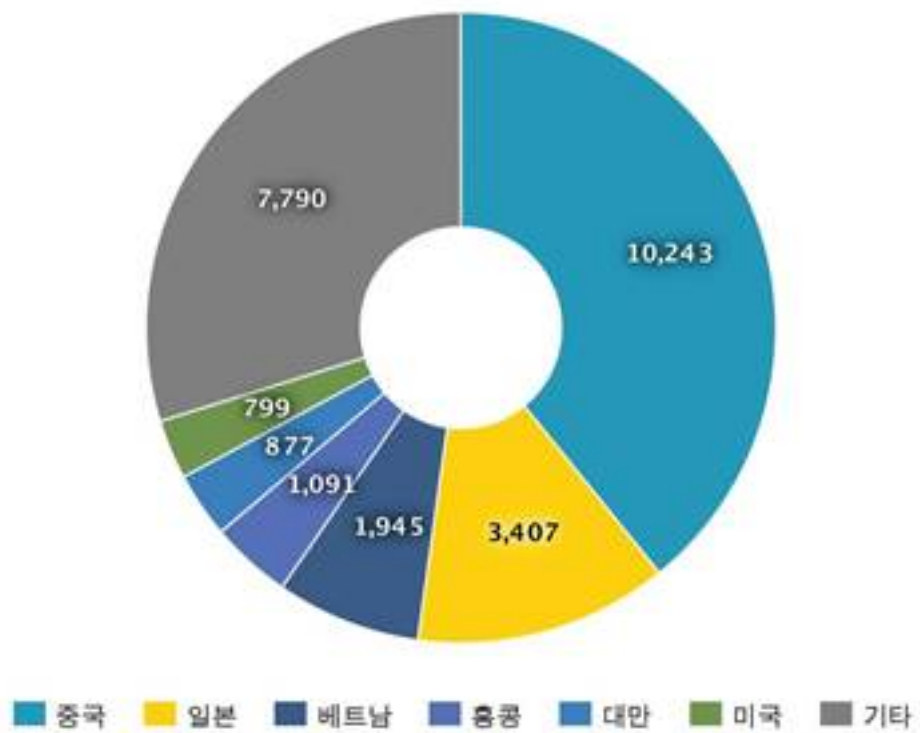


Рисунок 2. Распределение словоупотреблений по национальности

Также предоставлено процентное соотношение словоупотреблений по родному языку, на котором говорят иностранцы, изучающие корейский язык (количество в документах): китайский – 11 495 (44,0%), японский – 3 500 (13,4%), вьетнамский – 1 942 (7,4%), английский – 1 682 (6,4%), китайский (кантонский) – 1 074 (4,1%), русский – 1 043 (4,0%), другие языки – 5 416 (20,7%) (Рис.3).



Рисунок 3. Распределение словоупотреблений по родному языку

Национальный институт корейского языка также распределил количество словоупотреблений по уровням владения корейского языка (количество в документах): 1 уровень – 5 763 (22,0%), 2 уровень – 5 327 (20,4%), 3 уровень – 5 187 (19,8%), 4 уровень – 4 412 (16,9%), 5 уровень – 3 148 (12,0%), 6 уровень – 2 125 (8,1%), выше 6 уровня – 169 (0,6%) (Рис. 4).

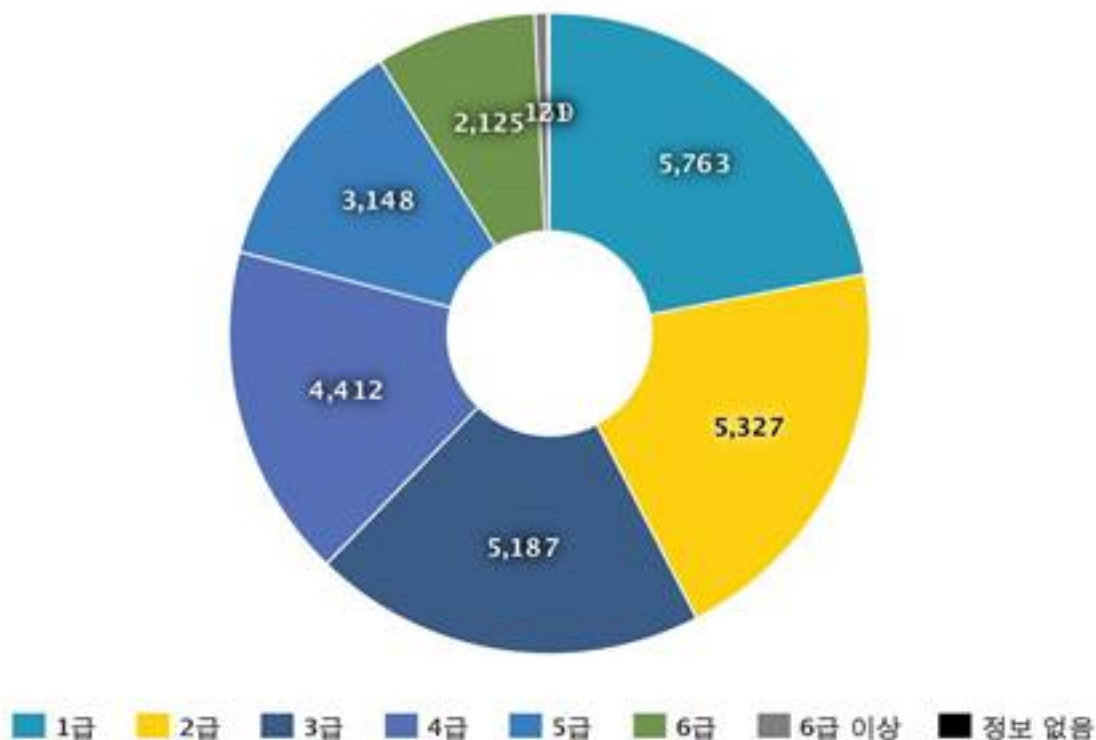


Рисунок 4. Распределение количества словоупотреблений по уровню владения корейским языком

Так, из детализированной таблицы можно узнать количество словоупотреблений, изучающих корейский язык из России (количество в документах): 1 уровень – 134, 2 уровень – 118, 3 уровень – 144, 4 уровень – 102, 5 уровень – 78, 6 уровень – 26, выше 6 уровня – 7. Также владеющих русским языком (количество в документах): 1 уровень – 203, 2 уровень – 214, 3 уровень – 238, 4 уровень – 180, 5 уровень – 144, 6 уровень – 51, выше 6 уровня – 9.

В разделе статистики по морфологической системе (○○ ○○○ ○○○ ○○) представлена система обозначений грамматических категорий данного корпуса.

형태 주석된 형태소 (4,854,943개)			(단위: 형태소 수, 비율)		
NNG(일반명사)	983,869	20.27%	MAJ(접속부사)	48,106	0.99%
VV(동사)	431,291	8.88%	JKG(관형격조사)	47,677	0.98%
EC(연결어미)	338,098	6.96%	XSA(명물사파생접미사)	43,622	0.90%
EF(종결어미)	305,196	6.29%	JC(접속조사)	26,415	0.54%
JKB(부사격조사)	237,866	4.90%	ETN(명사형전성어미)	25,735	0.53%
SF(마침표, 물음표, 느낌표)	231,378	4.77%	SN(숫자)	14,906	0.30%
ETM(관형형전성어미)	230,162	4.74%	XR(어근)	14,093	0.29%
JX(보조사)	206,039	4.24%	NR(수사)	13,435	0.28%
MAG(일반부사)	197,307	4.06%	SP(명표, 가운뎃점, 줄표, 빗금)	10,995	0.21%
JKO(목적격조사)	167,368	3.45%	SS(마음표, 향표표, 줄표)	9,032	0.19%
NNB(의존명사)	152,300	3.14%	JKC(보격조사)	8,367	0.17%
JKS(주격조사)	150,957	3.11%	VCN(부정지정사)	7,158	0.15%
VA(형용사)	147,276	3.03%	XPNI(재언접두사)	2,626	0.05%
XSV(동사파생접미사)	123,097	2.54%	SL(화국어)	2,533	0.05%
NP(대명사)	111,383	2.29%	JKQ(언용격조사)	1,529	0.03%
EP(선어말어미)	110,309	2.27%	SW(기타 기호(논리 수학기호, 화 제 기호 등))	622	0.01%
NNP(고유명사)	103,769	2.14%	SH(한자)	568	0.01%
VX(보조동언)	82,322	1.70%	SE(활영표)	310	0.01%
IC(감탄사)	77,318	1.59%	SO(불영표(물결, 숨김, 빠짐))	303	0.01%
VCP(공정지정사)	68,317	1.41%	JKV(포격조사)	130	0.00%
XSN(명사파생접미사)	63,630	1.31%	E(어미)	5	0.00%
MM(관형사)	53,504	1.21%			

Рисунок 5. Система обозначений грамматических категорий корпуса

Также в морфологической системе можно сформировать таблицу отдельно по грамматической категории, где выходит таблица, первый столбец – слова (в рис. существительные), второй столбец – обозначение грамматической категории, последующие столбцы - количество словоупотреблений по уровням владения корейским языком и общее количество словоупотреблений.

При нажатии поискового слова выводится контекстная информация и информация об авторе текста: национальность, родной язык, возраст, профессия, цель изучения корейского языка,

уровень владения корейским языком и тип корпуса (текстовой или разговорный).

В данном поиске (рис.7) было введено поисковое слово «неологизмы -[신조어]». Из результата поиска можно увидеть, что слово «неологизмы -[신조어]» употребляют студенты с высоким уровнем знания корейского языка. Данный поиск показал, что самое используемое словосочетание это «[신조어] [신조어] - неологизмы и популярные выражения». Все результаты можно скачать в Excel версии.

검색 조건

검색어

신조어

+

검색어 포함 조건

또는

▼

상세 조건

검색

표본 검색 결과

(중심어 선택시 상세정보 표출)

검색된 표본 : 14개(0.05%)

검색된 어절 : 33건(0%)

모국어 가나다순

20개씩 보기

연번	모국어	급수	원문 문맥	중심어	오른쪽 문맥
1	러시아어	5급	...기 때문에 거기에서 유행어와	신조어가	되게 잘 사용된다.

[문맥정보]

어전 문장

우즈베키스탄에는 8가지의 사투리가 있다. 한국과 달리 우즈베키스탄의 수도에서 이야기하는 언어는 표준어가 아니라 타슈켄트 사투리다.

현재 문장

타슈켄트는 겨울 발칸고원 연구가 조밀한 도시이기 때문에 거기에서 유행어와 신조어가 되게 잘 사용된다.

다음 문장

다른 지역에서 타슈켄트에 가는 사람들이 말하는 스타일을 타슈켄트 사투리로 비꾸려고 노력한다. 우즈베키스탄 표준어에 뭐 하고 있냐고 물어볼 때 '니마 울리얌산?'이라고 한다.

[표본 상세 정보]

국적	우즈베키스탄	모국어	러시아어	나이	22	직업	학생
학습 기관	비공개	학습 목적	거주	한국어 급수	5급	말뭉치 유형	문어

2	영어	6급	요즘 시대 사람들은	신조어를	많이 사용한다
3	영어	6급	컴퓨터나	신조어에	너무 깊이 발을 담그지 말고 표...

Рисунок 7. Результаты поиска

Поиск по морфологической системе ([신조어] [신조어]). В этом подразделе поиск возможен по морфологической системе. Также можно выбрать национальность, родной язык, уровень владения, возраст, цель изучения, тип корпуса и место изучения корейского языка. Данный тип поиска может быть использован для изучения и составления учебного материала.

복합 주석 검색		단순 검색															
검색 조건																	
검색어 ○ 세종장태표지 ○ N(재언) ○ NN(영사) ☒ NNG(일반영사) ☐ NNP(고유영사) ☐ NNB(외환영사) ☐ NR(대명사) ☐ NR(수사) ○ V(용언) ○ VC(지정사) ☐ VCP(공칭지정사) ☐ VCN(부정치정사) ☐ VI(동사) ☐ VA(형용사) ☐ VX(보조용언)		학생 <table border="1"> <tr> <td>국적</td> <td> 국적을 선택하세요 </td> </tr> <tr> <td>모국어</td> <td> 모국어를 선택하세요 </td> </tr> <tr> <td>한국어 등급</td> <td> ☐ 1급 ☐ 2급 ☐ 3급 ☐ 4급 ☐ 5급 ☐ 6급 ☐ 6급~ ☐ 정보 없음 </td> </tr> <tr> <td>나이</td> <td> ☐ ~10대 ☐ 20대 ☐ 30대 ☐ 40대 ☐ 50대~ </td> </tr> <tr> <td>학습 목적</td> <td> ☐ 진학 ☐ 취업 ☐ 거주 ☐ 취미 ☐ 결혼 ☐ 기타 </td> </tr> <tr> <td>말뭉치 유형</td> <td> ☐ 문어 ☐ 구어 </td> </tr> <tr> <td>학습 환경</td> <td> ☐ 국내 ☐ 국외 </td> </tr> </table>		국적	국적을 선택하세요	모국어	모국어를 선택하세요	한국어 등급	☐ 1급 ☐ 2급 ☐ 3급 ☐ 4급 ☐ 5급 ☐ 6급 ☐ 6급~ ☐ 정보 없음	나이	☐ ~10대 ☐ 20대 ☐ 30대 ☐ 40대 ☐ 50대~	학습 목적	☐ 진학 ☐ 취업 ☐ 거주 ☐ 취미 ☐ 결혼 ☐ 기타	말뭉치 유형	☐ 문어 ☐ 구어	학습 환경	☐ 국내 ☐ 국외
국적	국적을 선택하세요																
모국어	모국어를 선택하세요																
한국어 등급	☐ 1급 ☐ 2급 ☐ 3급 ☐ 4급 ☐ 5급 ☐ 6급 ☐ 6급~ ☐ 정보 없음																
나이	☐ ~10대 ☐ 20대 ☐ 30대 ☐ 40대 ☐ 50대~																
학습 목적	☐ 진학 ☐ 취업 ☐ 거주 ☐ 취미 ☐ 결혼 ☐ 기타																
말뭉치 유형	☐ 문어 ☐ 구어																
학습 환경	☐ 국내 ☐ 국외																
		검색															

Рисунок 8. Поиск по морфологической системе

Поиск по ошибочной системе (□□ □□ □□ □). В данном поиске выводятся тексты, где изучающие допустили грамматические и орфографические ошибки. В данном примере было введено поисковое слово «□□ – студент» и были выведены результаты текстов с грамматической и орфографической ошибкой в этом слове. При нажатии на поисковое слово выводится объяснение грамматической ошибки с исправлением в виде формулы, например □□/NGG+□/XSN+□/JKO.

17	카자흐어	5급	..을 오게 된 외국>	<u>학생들에게</u>	학생/NNG	도와주며 입학과 ..	주석: 띄어쓰기	주석: 띄어쓰기	주석: 띄어쓰기
----	------	----	--------------	--------------	--------	-------------	----------	----------	----------

교정정보	학생/NNG+들/XSN+을/IKO				
	<name> 한국에 센터에서 외국 대학장으로 한국어를 공부한 경험도 있기 때문에 한국에 처음 오게 된 외국 학생들에게 도와주며 입학과 관련한 일을 할 수 있다.				
교정결과	<name> 한국에 센터에서(센터+에서) 외국 대학장으로 한국어를 공부한 경험도 있기 때문에 한국에 처음 오게 된 외국 학생들에게(학생+들+을) 도와주며(도와주+고) 입학과 관련한 일을 할 수 있다.				
상세 정보	원형태	교정형태	오류 위치	오류 영상	오류 순위
	학생	학생	주석: 띄어쓰기	주석: 띄어쓰기	주석: 띄어쓰기
	들	들	주석: 띄어쓰기	주석: 띄어쓰기	주석: 띄어쓰기
	에게	을	부사격조사	대치	주석: 띄어쓰기

Рисунок 9. Поиск по ошибочной системе

Таким образом, на базе Национального института корейского языка существует корпус изучающих корейский как иностранный язык (한국어 학습자용 텍스트 데이터베이스 - Korean Learner's Corpus Search Engine), который является интересным и уникальным корпусом. Рассмотрев систему данного корпуса, можно сказать, что данный корпус действительно служит хорошим материалом для исследователей и преподавателей корейского языка при составлении исследовательских работ и учебных материалов.

Преподаватели корейского языка как второго иностранного языка могут использовать данный корпус при составлении учебных материалов по корейскому языку, а также приводить примеры правильных и ошибочных словоупотреблений. Студенты, которые изучают корейский язык как второй иностранный язык, могут использовать данный корпус при изучении языка, подготовке к экзамену TOPIK (Test of Proficiency in Korean) и при написании выпускной квалификационной работы, связанной с педагогической темой. Также данный корпус может послужить примером для создания корпуса изучающих русский язык как иностранный. В настоящее время иностранцы, изучающие русский язык, сдают экзамен ТРКИ – Тест по русскому языку как иностранному (TORFL – Test of Russian as a Foreign Language), в состав которой входит часть «Письмо». Из работ данной части можно составить корпус изучающих русский язык. В результате исследователи и преподаватели русского языка как иностранный могут использовать данные тексты в разработке учебных материалов русского языка для иностранцев.

Список литературы:

1. 한국어 학습자용 텍스트 데이터베이스 - Korean Learner's Corpus Search Engine. Режим доступа: <https://kcorpus.korean.go.kr/> (Дата обращения: 18.11.2020)