3. 진로탐색학점제 활동내용 및 결과, 활동 성과 ※주요 활동 내용 및 결과에 대해 상세히 기술

1. 연구 계획 부분 (1주차 ~ 2주차)

1주차에는 지난 1학기 '영어 데이터 활용' 수업에서 수강한 내용을 바탕으로, 어떤 대상을 바탕으로 연구를 진행할지에 대한 계획을 수립하였다. 계획을 수립하는 과정에서 내 경험을 바탕으로 주제를 선정하면서, 1학년 때, 교양영어 수업을 듣고, 실제 작문을 하는 과정에서 느꼈던 작문에 대한 어려움(문법적 올바름, 어휘 선택의 적절성)을 생각하였다.

아무래도, 입시를 마치고 대학교에 입학한 신입생의 경우는 대부분 수능 영어시험을 준비하면서 읽기, 듣기 부분에 있어서는 많은 교육 경험이 있지만, 쓰기나 말하기의 영역에서는 비교적 어려움을 느낄 것이라는 생각을 했고, 이 중 작문과 관련된 데이터를 바탕으로 분석을 진행한다면 의미 있는 분석 결과가 나올 것이라는 예상을 한 뒤, 분석 계획을 수립하기 시작하였다.

아울러, 이번 연구와 관련해서 선행연구에는 어떤 종류가 있는지 분야를 탐색하였고, 크게 '코퍼스 수집 관련', '코퍼스 분석 관련', '분석 형태 선정', '저작권 관련' 이렇게 크게 4가지의 주제를 선정하였다. 이러한 주제를 바탕으로 다양한 논문들을 찾아보았고, 이 논문 안에서 어떤 방법론으로 데이터들을 분석하였는지를 중점으로 탐구를 진행하였다.

2주차에는 위에서 찾아보았던 논문들을 분석하면서, 내가 분석을 시행하기 전, 어떤 지식을 알아야 하고, 어떤 식으로 본 분석을 진행할지에 대한 세부사항을 정하기 시작했다.

우선, 한국 대학생들이 보편적으로 어떤 오류를 범하는지 체크하기 위해 「손희정. 2020. "한국인 대학생들의 영어 말하기와 쓰기에 나타난 문법 오류 분석." 박사학위논문, 단국대학교 대학원. 경기도.」 논문을 확인해 보았다. 본 논문에서는 대학생들이 주로 보이는 오류에 대해 다음과 같이 정의하고 있다.

	오류 형태	
1	명사의 복수형 생략 오류	
2	주어 - 동사의 불일치의 오류	
3	3인칭 현재 단수 -s 표시 오류	
4	관사 사용 오류	관사 생략
5	어순, 문장의 구조, 지배	부사 수식어를 목적어 앞에 위치
		정형동사 앞에 목적어를 위치시키는 오류
		장소 수식어 앞에 시간 수식어를 놓는 오류
6	부적절한 간접 목적어의 첨가	
7	간접 의문문 도치의 오류	
8	Some과 Any의 혼동	

이러한 유형을 바탕으로, 교수님과 면담한 결과, 학생들이 많이 발생하는 오류를 바탕으로 분석을 해 보자고 말씀하셨고, 결과적으로 내가 중점적으로 분석해 볼 유형은 '명사형 복수형 생략 오류', '3인칭 현재 단수 -s 표시 오류', '관사 사용 오류'에 대해 탐구해 보기로 했다.

그 다음, 「Lee, Ji Yon. 2016. "A Corpus-Based Study of the Use of Verb-Noun Collocations in Korean EFL Writing." 석사학위논문, 연세대학교 일반대학원. 대한민국.」 이 논문을 참고하면서, 학생들이 작성하는 작문 내용 중에서, 어떤 Collocation(연어 관계)가 나타나는지에 대하여 분석해 보았다.

품사적인 측면에서, Collocation은 Grammatical collocation과 Lexical collocation이 존재하는데, 각 유형에 따른 종류는 다음과 같다.

Table 1. The classification of grammatical collocations (Benson et al., 1986)

Code		Pattern	Examples
G1		noun + prep	'blockade against'; 'apathy towards'
G2		noun + *to*-infinitive	'a pleasure to do it'; 'an attempt to do it'
G3		noun + *that*-clause	'we reached an agreement that'
G4		prep + noun	'by accident'; 'in advance'; 'in agony'
G5		adj + prep	'afraid of'; 'ashamed of'; 'confident of'
G6		adj + *to*-infinitive	'it was necessary to work'; 'she is ready to go'
G7		adj + *that*-clause	'she was afraid that'; 'it was necessary that'
G8	A	verbs (trans) that allow dative movement	'he sent the book to his brother' → 'he sent his brother the book'
	B	verbs (trans) that do not allow dative movement	'they described the book to her'
	C	verbs (trans) used with *for* that allow dative movement	'she bought a shirt for her husband' → 'she bought her husband a shirt'
	D	verb + prep (+ obj)	'they came by train'; 'we cut bread with a knife'
	d	verb + prepositional phrase	'we will adhere to the plan'; 'our committee consists of six members'
	E	verb + *to*-infinitive	'they began to speak'; 'she continued to write'
	F	verb + infinitive without *to*	'we must work'; 'he had better go'
	G	verb + v-ing	'they enjoy watching television'
	H	verb (trans) + obj + *to*-infinitive	'we advised them to be careful'
	I	verb (trans) + obj + infinitive without 'to'	'she heard them leave'; 'we let the children go to the park'
	J	verb (trans) + object + v-ing	'I caught them stealing apples'; 'he kept me waiting two hours'
	K	verb (trans) + possessive (pronoun or noun) + gerund	'please excuse my waking you so early'; 'they love his clowning'
	L	verb (trans) + *that*-noun clause	'they admitted that they were wrong'; 'she believed that her sister would come'
	M	verb (trans) + obj + infinitive *to be* + complement (adj/past part/noun/pronoun)	'we consider her to be very capable'; 'we found the roads to be cleared of snow'
	N	verb (trans) + obj + complement (adj/past part/noun/pronoun)	'she dyed her hair red'; 'the police set the prisoner free'

	O	verb (trans) + obj + obj	'the teacher asked the pupil a question'
	P	verb + adverbial	'he carried himself well'; 'my brother is living in Utah'; 'she put pressure on them'
	Q	verb + interrogative word (*wh*-word)	'he asked how to do it'; 'she knew when to keep quiet'
	R	subject (it) + verb + *to*-infinitive or *that*-clause	'it puzzled me that they never answered the phone'
	S	verb (intrans) + predicate (noun or adj)	'she became an engineer'; 'she was enthusiastic'
	s	verb (intrans) + predicate adj	'she looks fine'; 'the food tastes good'

Table 2. The classification of lexical collocations (Benson et al., 1986)

Code	Pattern	Examples
L1	verb + noun/pronoun (or prepositional phrase) (creation, activation)	'come to an agreement'; 'make an impression'; 'set an alarm'; 'launch a missile'
L2	verb + noun (eradication, nullification)	'reject an appeal'; 'withdraw an offer'; 'ease tension'; 'denounce a treaty'
L3	adj + noun	'strong tea'; 'a chronic alcoholic'; 'a rough estimate'
L4	noun + verb	'alarms ring'; 'bees buzz'; 'bombs explode'
L5	$noun_1$ *of* $noun_2$	'a bouquet of flowers'; a piece of advice'; 'an act of violence'
L6	adverb + adj	'deeply absorbed'; 'closely acquainted'
L7	verb + adverb	'affect deeply'; 'appreciate sincerely'

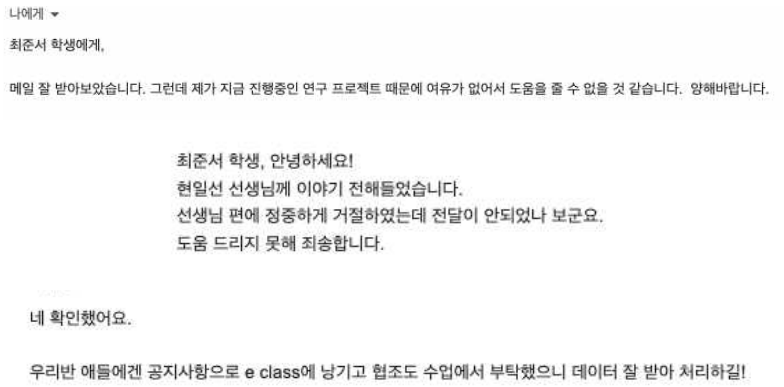
나는 이 종류들 중, 학생들이 명사를 수식하는 단어를 사용하는 데 있어서, 원어민과 차이를 보일 것이라는 가설을 설정하고, 코퍼스를 분석하겠다고 생각했다.

2. 코퍼스 수집 부분 (3주차 ~ 5주차)

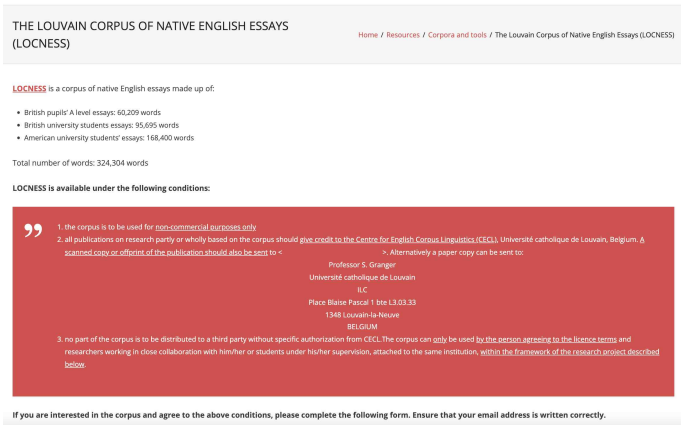
3주차에는 본격적으로 코퍼스를 만들기 위한 데이터 수집을 시작했다. 우선, 처음 주제를 '한성대학교 학생들'로 잡았기 때문에, 학생들의 데이터를 모으기 위한 작업에 도입했고, 이때 저작권 관련하여 문제가 생길 수 있어, 「김은기, Eun-Gui. 2010. "온라인 대학교육에서의 저작권문제." 정보법학 14 (3): 83-108.」 본 논문을 비롯한 다양한 자료들을 찾아보았는데, 저작권법 제30조에서 영리를 목적으로 하지 않는다면 학생 과제물을 사용이 가능하다는 점을 바탕으로, 연구를 진행해도 문제가 없을 것이라는 생각을 하게 되었다. 그러나, 데이터 수집을 할 때, 학생들에게 어떠한 목적으로 데이터를 수집하는지, 그리고 추가로 얻을 수 있는 정보들을 위하여, 설문지 및 동의서를 작성하였다.

번호	질문
1	이름
2	전화번호
3	소속 (1학년일 경우, 학부 선택 / 2학년 이상일 경우, 소속 트랙명 작성)
4	본인이 생각하는 자신의 영어 실력
5	공인영어 성적 (있을 경우에만, 없으면 생략 가능, ex. TOEIC 550)
6	2024년도 1학기 이외의 영어로 1문단 이상의 글을 써 본 적 있는지 (있다면 얼마나 써 보았는지, 없다면 X 작성) (ex. 영어 이메일 3회 작성, 등)
7	영어 작문 교육을 받은 경험이 있는지 여부(있다면 어떻게 배웠는지, 없다면 X 작성) (ex.중학교, 고등학교, 사설 학원, 과외, 등)
8	영어 작문 교육을 받은 적이 있다면, 얼마나 도움이 되었는지 여부
9	작문을 하면서 어려움을 느낄 때, 해결했던 방법 (다중 선택 가능)
10	개인정보 수집 및 활용, 과제물 사용에 관한 동의

본 내용을 작성한 뒤, 영어 커뮤니케이션 독해/작문 수업을 진행하시는 교수님들께 메일을 보냈고, 한 교수님께서만 데이터 수집에 도움을 주신다고 말씀을 주셨고, 각 수업의 e-class 공지사항 게시판에 설문지 링크(https://forms.gle/bZi6HUWuB912J5627)를 올리는 방식으로 데이터 수집 동의를 얻었다.



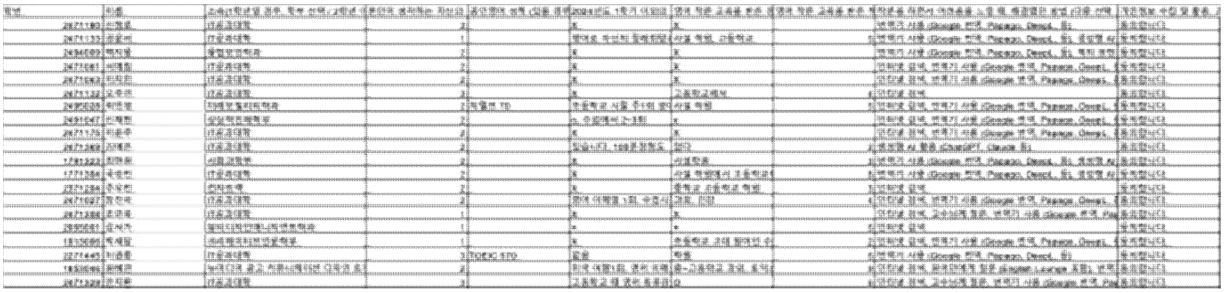
그 후, 원어민 학생들의 작문 데이터를 얻기 위해 다양한 코퍼스를 찾아본 결과, Louvain Corpus of Native English Essays (LOCNESS)를 알게 되었다. 이 데이터는 한국 학생들과 외국 학생들의 표현 능력 차이를 분석하기 위해 수집하는 코퍼스로, 영국과 미국 학생들의 작문 과제를 바탕으로 한 내용을 담고 있다. 이 데이터를 받기 위해 인터넷 사이트에서 동의서를 작성하는 과정하면, 이메일로 코퍼스 파일을 보내주는 방식을 채택하고 있어, 비교적 어렵지 않게 코퍼스를 수집할 수 있었다.



https://www.learnercorpusassociation.org/resources/tools/locness-corpus/

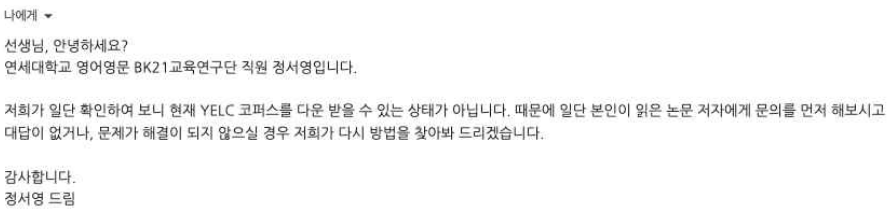
4주차에는 3주차에 한 내용을 연장하여, 한성대학교 학부생 대상으로 한 설문자를 확인하고, 다른 코퍼스의 종류를 찾아보았다.

먼저, 설문지에 응답을 한 학생들은 총 20명으로, 예상보다 적은 수의 학생들을 모집했기 때문에, 이 자체만으로 코퍼스 분석을 한다면 어려움이 있을 듯하여, 다른 학교에서 공개한 코퍼스 파일을 찾아보았다.(하단의 이미지는 설문조사에 응답한 인원 정리한 파일임. 따로 엑셀 파일 첨부함)



다른 학교의 코퍼스의 경우에는 연세대학교 학습자 코퍼스(YELC), 가천대학교 학습자 코퍼스가 있어서 본 코퍼스에 접근하기 위해 다양한 사이트를 찾아보았다.

우선, 연세대학교의 경우, 2012년부터 YELC를 연구 목적으로 공개하였다고 코퍼스 관련 사이트에는 나와있었으나, 사이트에서 접근을 할 수 있는 곳이 없어, 직접 연세대학교 BK21사업단 측과 연락한 결과, 현재 갖고 있는 코퍼스가 없다는 연락을 받았다.



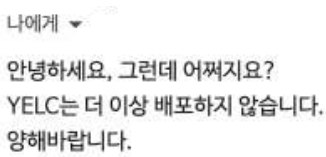
나에게

선생님, 안녕하세요?
연세대학교 영어영문 BK21교육연구단 직원 정서영입니다.

저희가 일단 확인하여 보니 현재 YELC 코퍼스를 다운 받을 수 있는 상태가 아닙니다. 때문에 일단 본인이 읽은 논문 저자에게 문의를 먼저 해보시고 대답이 없거나, 문제가 해결이 되지 않으실 경우 저희가 다시 방법을 찾아봐 드리겠습니다.

감사합니다.
정서영 드림

그래서 코퍼스 담당을 하신 교수님께 연락을 취해보았으나, 교수님께서도 더 이상 YELC는 배포하지 않는다는 답변을 주셨다.



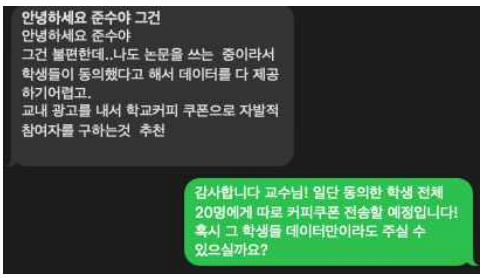
나에게

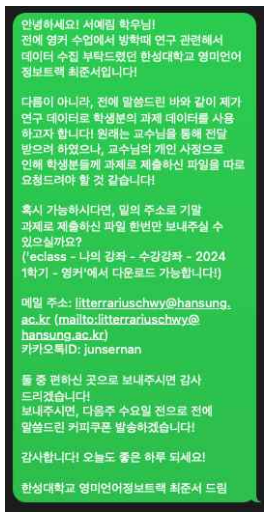
안녕하세요, 그런데 어쩌지요?
YELC는 더 이상 배포하지 않습니다.
양해바랍니다.

이러한 답변을 받고, 이 코퍼스를 활용하는 것은 어려울 듯하여, 다른 학교에서 구축한 코퍼스가 있는지 확인하였고, 가천대학교에서 코퍼스를 수집한 것을 확인하였다.

가천대학교에서는 다음과 같은 사이트 (https://koreanlearnercorpusblog.blogspot.com/p/corpus.html)에 코퍼스 파일을 업로드해 두었고, 여기에서 자유롭게 이용해도 된다는 내용이 있어, 최종본을 다운로드 받았다.

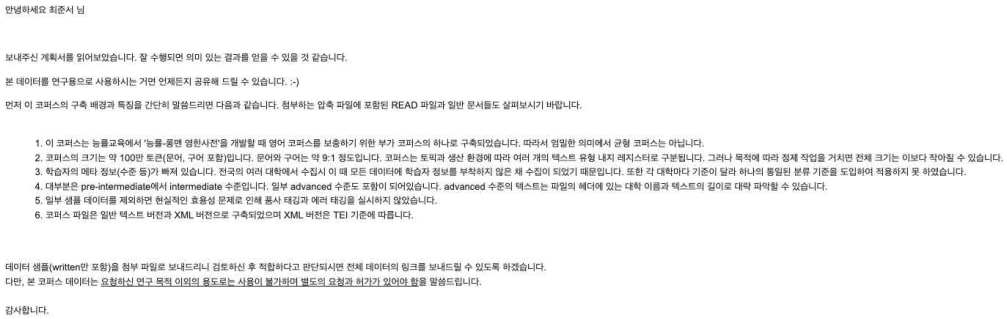
5주차에는 한성대학교 학습자 코퍼스를 수집하기 위해 교수님께 연락을 드렸다. 그러나 갑자기 교수님께서 사정이 생기셔서 학습자 코퍼스를 제공하시기 어렵다는 말씀을 주셨고, 이러한 사정 때문에, 내가 직접 설문지에 응답을 해 준 학생들에게 직접 연락을 했다.





이렇게 연락을 한 결과, 20명 중 6명의 학생들이 데이터를 보내주었고, 데이터 크기가 너무 작아, 이 데이터 자체를 분석하기보다는 다른 데이터들의 특징을 정리한 후, 개선 방안을 탐구하는 것이 더욱 효율적이라고 생각해 주제를 '한성대학교 학생'에서 '한국 대학생'으로 범위를 넓혀 진행하는 것으로 방향성을 수정했다.

가천대학교 코퍼스만을 바탕으로 분석을 실시하기에는 한국 대학생이라는 범위를 충족하기는 어렵다는 생각에 다른 코퍼스를 찾아보던 중, 능률교육에서 롱맨영어사전을 편찬하기 위해 한국 대학생들의 작문 데이터를 모아둔 코퍼스를 발견했다. 그래서, 이를 접근하기 위해 담당자에게 연락했고, 연구용으로는 데이터를 제공해 줄 수 있다는 이야기를 받았다.


안녕하세요 최준서 님

보내주신 계획서를 읽어보았습니다. 잘 수행되면 의미 있는 결과를 얻을 수 있을 것 같습니다.

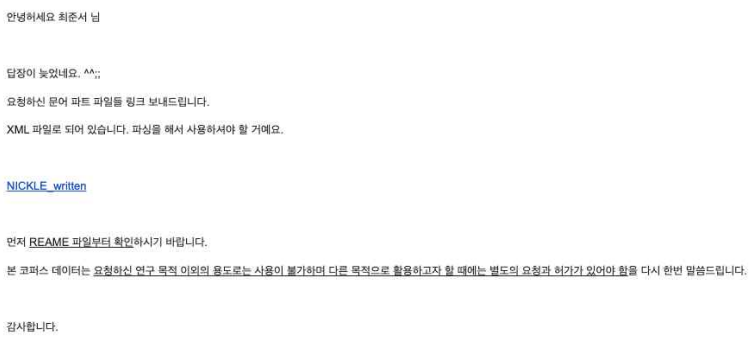
본 데이터를 연구용으로 사용하시는 거면 언제든지 공유해 드릴 수 있습니다. :-)

먼저 이 코퍼스의 구축 배경과 특징을 간단히 말씀드리면 다음과 같습니다. 첨부하는 압축 파일에 포함된 READ 파일과 일반 문서들도 살펴보시기 바랍니다.

1. 이 코퍼스는 능률교육에서 '능률-롱맨 영한사전'을 개발할 때 영어 코퍼스를 보충하기 위한 부가 코퍼스의 하나로 구축되었습니다. 따라서 엄밀한 의미에서 균형 코퍼스는 아닙니다.
2. 코퍼스의 크기는 약 100만 토큰(문어, 구어 포함)입니다. 문어와 구어는 약 9:1 정도입니다. 코퍼스는 토픽과 생산 환경에 따라 여러 개의 텍스트 유형 내지 레지스터로 구분됩니다. 그러나 목적에 따라 정제 작업을 거치면 전체 크기는 이보다 작아질 수 있습니다.
3. 학습자의 메타 정보(수준 등)가 빠져 있습니다. 전국의 여러 대학에서 수집시 이 때 모든 데이터에 학습자 정보를 부착하지 않은 채 수집이 되었기 때문입니다. 또한 각 대학마다 기준이 달라 하나의 통일된 분류 기준을 도입하여 적용하지 못 하였습니다.
4. 대부분은 pre-intermediate에서 intermediate 수준입니다. 일부 advanced 수준도 포함이 되어있습니다. advanced 수준의 텍스트는 파일의 헤더에 있는 대학 이름과 텍스트의 길이로 대략 파악할 수 있습니다.
5. 일부 샘플 데이터를 제외하면 한실적인 효용성 문제로 인해 품사 태깅과 에러 태깅을 실시하지 않았습니다.
6. 코퍼스 파일은 일반 텍스트 버전과 XML 버전으로 구축되었으며 XML 버전은 TEI 기준에 따릅니다.

데이터 샘플(written만 포함)을 첨부 파일로 보내드리니 검토하신 후 적합하다고 판단되시면 전체 데이터의 링크를 보내드릴 수 있도록 하겠습니다.
다만, 본 코퍼스 데이터는 요청하신 연구 목적 이외의 용도로는 사용이 불가하며 별도의 요청과 허가가 있어야 함을 말씀드립니다.

감사합니다.



안녕하세요 최준서 님

답장이 늦었네요. ^^;;

요청하신 문어 파트 파일들 링크 보내드립니다.

XML 파일로 되어 있습니다. 파싱을 해서 사용하셔야 할 거예요.

NICKLE_written

먼저 REAME 파일부터 확인하시기 바랍니다.

본 코퍼스 데이터는 요청하신 연구 목적 이외의 용도로는 사용이 불가하며 다른 목적으로 활용하고자 할 때에는 별도의 요청과 허가가 있어야 함을 다시 한번 말씀드립니다.

감사합니다.

3. 분석 및 보고서 작성 부분 (6주차 ~ 8주차)

6주차에는 위에서 수집한 코퍼스를 바탕으로, 실제 분석을 해 보았다. 먼저, 문법 오류 부분에서 가천대학교 코퍼스를 분석해 보았는데, 위에서 말했던 '명사형 복수형 생략 오류', '3인칭 현재 단수 -s 표시 오류', '관사 사용 오류'에 해당하는 문장의 일부분은 다음과 같다.

가천대학교 코퍼스 - 명사형 복수형 생략 오류	
1	If children see repeated violent **film** their worldview can be adhered.
2	So many **Korean** want to be pretty girl or handsome guy.
3	alcohol makes you in trance. it so goooooood. but it cause quick death. alcohol addiction, **accident** with drunk.
4	many young **person** attracts each other through body.
5	the old days, rules or **standard** was very strict from everything.

가천대학교 코퍼스 안에서 명사형 복수형 생략 오류를 위 표에서 나타난 경우를 바탕으로 선별하였다. 데이터의 사이즈를 고려하여 'many'라는 단어를 바탕으로 복수형 오류가 있는지에 대한 여부를 확인했다. 우선, accident(45회), behavior(1회), foreigner(12회), hotel(36회), many 뒤에 복수명사가 온 경우 제외), kind(109회), korean(36회, '한국인들'이라는 의미로 사용했을 때만 체크), problem(65회), room(3회)로 다수의 오류가 발견되었다.

가천대학교 코퍼스 - 3인칭 현재 단수 -s 표시 오류	
1	It **irritate** others and they from now on mind talking with you.
2	First, we talk about smoking. when you have smoke, you feels good but it is very temporary. cigarette **attack** your lung, at last you have ill, pain.
3	Although the opposite **try** to be calm and ignore the aggressive gesture, it is hard to deal one's feeling oneself.
4	I read that a fat girl **have** declared her love for one of her fave people. But He was refused. So, The girl took diet and he was after her appearance.
5	She wants to know how to move forward without life jacket. So she **catch** me and advance more and more.

가천대학교 코퍼스 안에서 3인칭 단수가 사용된 경우를 위 표에서 나타난 경우를 바탕으로 직접 파일을 눈으로 보면서 선별하였는데, 데이터의 사이즈가 커, AntConc 프로그램의 N-Gram을 추출하는 기능을 활용하였다. 검색을 할 때 3인칭 대명사 위주로 검색하였는데, 'she'의 경우에는 have(32회), want(14회), leave(6회), realize(6회)가 대표적으로 3인칭 단수 오류에 해당하였다. 'he'의 경우에는 have(47회), think(39회), realize(13회), run(13회), watch(1회)의 오류가 발생하였다. 그리고 'it'의 경우에는 have(58회)가 나타났다.

가천대학교 코퍼스 - 관사 사용 오류	
1	**Foreigh language** is essential in order to provide good service.
2	For example, **special day**, such as **birthday** or **wedding anniversary** with a special day of allows.
3	In addition to it, so that customers can arrive at **hotel** comfortably, I will provide free airport shuttle.
4	It was very thriling and frightening event.
5	At first, she can't adjust the job, because of **evil boss.**

위의 두 경우와 다르게, 관사와 관련된 부분은 각각의 명사를 파일 내에서 찾고, 거기에서 관사가 쓰였는지를 문장 구조에서 직접 찾아보는 단계를 거쳤다. 가천대학교 코퍼스에서는 위의 각 경우에서 'a'와 'an', 'the'가 쓰이지 않는 경우들을 모아보았다.

위의 내용처럼, 내가 예상한 종류들의 오류가 가천대학교 학습자들한테서 많이 나왔다는 사실을 알 수 있었다.

7주차에는 능률 교육에서 받은 코퍼스를 바탕으로, 실제 분석을 해 보았다. 능률교육 코퍼스 같은 경우에는 XML 파일로 이루어져 있어, 먼저 TXT 파일로 변환해주기 위해 파이썬 코드를 사용했다.

```python
import os
import xml.etree.ElementTree as ET


1개의 사용 위치
def extract_body_texts_from_xml(file_path):
    try:
        tree = ET.parse(file_path)
        root = tree.getroot()

        body_elements = root.findall('.//body')

        body_texts = []
        for body in body_elements:
            body_text = ''.join(body.itertext()).strip()

            if len(body_text) >= 10 and 'head:' not in body_text.lower():
                body_texts.append(body_text)

        return body_texts
    except ET.ParseError as e:
        print(f"Error parsing {file_path}: {e}")
        return []


1개의 사용 위치
def extract_and_save_all_bodies(directory_path, output_file):
    all_body_texts = []

    for filename in os.listdir(directory_path):
        if filename.endswith('.xml'):
            file_path = os.path.join(directory_path, filename)
            body_texts = extract_body_texts_from_xml(file_path)
            all_body_texts.extend(body_texts)
```

```python
directory_path = '/Users/junseo/PycharmProjects/Corpus_Analysis/xml'
output_file = 'BT_all_bodies.txt'

extract_and_save_all_bodies(directory_path, output_file)

print(f"All body texts have been extracted and saved to {output_file}")
```

그 후 마찬가지로 문법 오류 부분에서 코퍼스를 분석해 보았는데, 위에서 말했던 '명사형 복수형 생략 오류', '3인칭 현재 단수 -s 표시 오류', '관사 사용 오류'에 해당하는 문장의 일부분은 다음과 같다.

능률 코퍼스 - 명사형 복수형 생략 오류	
1	He is the lastborn in seven **brother** and **sister.**
2	Their **name** is Son Somi, Park Heehyun, Kim Seunghee.
3	Japan is compacted old **culture** and new **culture.**
4	She have many **friend**.
5	She had lots of **illness**.

능률 코퍼스의 경우, 가천대 코퍼스와 다르게 many, much, a lot of와 같이 뒤에 복수명사를 사용해야 하는 경우에 해당하는 오류의 수는 많지 않았다. 그래서 AntConc에서 오류를 분석하였을 때, many와 much와 연어 관계를 이루는 단어의 수가 적어 오류의 출현이 적었었다. 따라서, 각 파일을 열고, 하나씩 정리하였고, 그 내용은 위의 표와 같다.

능률 코퍼스 - 3인칭 현재 단수 -s 표시 오류	
1	Heehyun **major** in literature creation.
2	She **have** collected perfume since 1988.
3	She **want** to be a translator.
4	She **have** many friend.
5	She **want** to be a good English teacher.

능률 코퍼스 안에서 3인칭 단수가 사용된 경우를 위 표에서 나타난 경우를 바탕으로 직접 파일을 눈으로 보면서 선별하였는데, 데이터의 사이즈가 커, AntConc 프로그램의 N-Gram을 추출하는 기능을 활용하였다. 검색을 할 때 3인칭 대명사 위주로 검색하였는데, 가천대학교 코퍼스와의 다른 점은 오류로 발생한 단어의 다양성이었다. 'he'의 경우에는 become(1회), eat(1회), get(2회), have(5회), lead(1회), tell(1회), want(1회)의 오류가 발생하였다. 그리고 'she'의 경우에는 attend(1회), go(1회), graduate(1회), have(14회), hold(1회), laugh(1회), live(1회), prefer(1회), prepare(1회), start(1회), take(1회), teach(1회), want(9회)가 나타났다.

능률 코퍼스 - 관사 사용 오류	
1	He is student at **chonan university** with me.
2	Also, she likes see **movie**.
3	She is going to part time job in **summer vacation.**
4	Mother is **house keeper**.
5	She is going to watch **movie** 'Scream 3' with her friend tomorrow.

마찬가지로, 능률 코퍼스 내에서도 관사 사용과 관련하여 오류가 다수 발견됐다. 가천대학교 코퍼스와 마찬가지로, AntConc에서 품사 단위로 추출하는 데에는 어려움이 있기 때문에, 텍스트 파일을 보면서 관사가 빠진 단어들을 추출했고, 결과는 다음과 같다.

위에서 본 내용과 동일하게 다양한 학교 학생들이 모여 있는 능률 코퍼스에서도 위에서 말했던 '명사형 복수형 생략 오류', '3인칭 현재 단수 -s 표시 오류', '관사 사용 오류'들이 다수 나타났다.

따라서 위에서 분석한 결과에 따라, 다수의 학생들이 작문을 할 때 '명사형 복수형 생략 오류', '3인칭 현재 단수 -s 표시 오류', '관사 사용 오류'가 빈번하게 나타날 것이라는 나의 가설은 참으로 드러났다고 볼 수 있다.

이를 한성대학교 코퍼스 안에서도 학생들이 실수를 하는지에 대해서도 분석해 보았다.

한성대학교 코퍼스 - 오류 유형별 정리		
1	For example use personal tumbler instead of **plastic cups.**	관사 사용 오류
2	money problems continued to bring **stock down.**	관사 사용 오류
3	Avoiding **stres** is helpful for the function of the hippocampus and **reduce** harms memory.	스펠링 오류, 3인칭 단수 오류

한성대학교 코퍼스에서는 크기가 크지 않아 다양한 횟수의 오류가 나오지 않았지만, 위에서 나왔던 오류들이 나타났다. 각 유형별로는 다음과 같다. 3인칭 단수 오류 1회, 관사 사용 관련 오류 4회이다. 다만, 명사 복수형 관련 오류는 나오지 않았는데, 학생들의 작문 내용 중 'the number of'와 관련된 내용이 다수 출현한 것을 보면, 수업 중에 복수형 관련 내용을 다루었기 때문에 오류 패턴이 나타나지 않았다고도 볼 수 있다.

따라서, 이는 대학생들을 위한 작문 수업을 할 때, 관사 사용과 관련된 내용, 3인칭 단수 오류 관련 내용, 명사 복수형 관련 내용을 다루어야 한다는 점을 뒷받침할 수 있다.

위 조사 결과를 통틀어서 보자면, 가천대학교 코퍼스에는 동일한 범주 내에서의 오류가 다수 발견되었고, 능률교육 코퍼스에서는 다양한 형태의 오류가 발견되었음을 알 수 있다. 어느 정도 수업이 진행된 상황에서 추출된 한성대학교 코퍼스에서 보았을 때, 특정 오류에 대한 교육을 실시하였을 때, 그 방안이 개선된 것을 확인할 수 있었다.

8주차에는 위에서 분석한 내용을 바탕으로 최종 보고서 작성 및 결과물 정리를 실시하였다.

4. 최종 결과물 ※ 활동을 증빙할 수 있는 근거자료 및 최종 결과물 제시, 별도 제출 가능. 단, ppt, pdf, 웹페이지가 결과물인 경우 대표 페이지 4~6장 정도 캡쳐한 이미지를 넣어주세요.

<논문 서칭 자료>

진로탐색학점제 논문리뷰 1주차

한성대학교 영미언어정보트랙

2011051 최준서

I. 손희정. 2020. "한국인 대학생들의 영어 말하기와 쓰기에 나타난 문법 오류 분석." 박사 학위논문, 단국대학교 대학원. 경기도.

A. 결과

i. 명사구와 동사구에서 가장 높은 오류 발생률을 보임

B. 제2언어 습득과 관련된 내용

C. 오류 분석의 중요성 정리

i. 대조분석 가설: 모국어와 목표어를 대조하여 분석 한 자료를 바탕으로 연구자 및 교수자들의 관점에서 학습자들의 난점을 예측하고 이러한 오류를 사전에 피하거나 최소화하기 위해서 학습 과정 이전에 학습자를 미리 훈련하는 교수(teaching)에 주안점

ii. 오류 분석 이론: 학습 과정 중에 발견되는 오류를 자연스럽게 수용하면서 그 원인을 밝히고 분류하여 분석한 내용을 학습자 중심의 교육과정에 반영하고 이것을 다시 적용하는 학습(learning)에 초점

D. 오류 분석 과정

i. 자료 수집 단계: 어떤 학습자 언어 사용(ex. 언어 자료의 종류, 학습자의 언어 능력에 따른 수준, 학습사적 언어 학습 경험, 연구 기간(횡적 연구, 종적 연구) 등), 자연스러운 실험 견본 바람직하게 여겨짐(N. Ellis, 1996)

ii. 오류 확인 및 식별 단계: 오류를 구성하는 것이 무엇인지, 체계적 오류와 비체계적 실수를 구분, 명백한 오류와 명백하지 않은 오류 사이의 주된 차이점 구별

iii. 오류 범주화: 표층 구조 접근 방법 - 생략, 첨가, 잘못된 형태, 잘못된 어순에 따른 범주 / 의사소통에 따른 Global error, Local error 등

iv. 오류 설명 단계: 학습자들이 오류를 범할 때 어떤 책략을 사용했는지 - 언어 간 전이: 모국어의 전이로, 어휘, 문법, 화용론적 특성 / 언어 내 전이 - 과잉 일반화, 규칙 제한의 무지, 규칙 적용의 불완전함, 개념 가설화의 실패

i. 초급: 181개의 에세이, 35,631 words

ii. 중급: 136개의 에세이, 35,643 words

iii. 고급: 118개의 에세이, 35,634 words

2. LOCNESS

A. Consists of argumentative and expository essays written by British and American university students.

B. The sub-corpus compiled for this study consists of 99 argumentative essays written by students at Marquette University and University of South Carolina.

→ 비상업적 목적으로 코퍼스 사용 가능, 계약서 작성하고 코퍼스 수집 가능

3. 각 유형의 코퍼스 크기

Table 4. Sub-corpora of YELC and LOCNESS used in the study

	YELC	LOCNESS
Number of essays	435	99
Total corpus size	106,908	107,170

ii. 분석 방법

1. 사용 툴: WordSmith

A. Lemma(굴절형)를 알아서 묶어줌 (Lemmatizing 기능(

2. 과정

A. Frequency list를 추출하여 상위 5개의 동사 선택

B. Concordance line을 활용하여 동사 + 명사 조합 식별

i. 중간에 some, more 등과 같은 단어가 들어갔어도, 뒤에 직접적으로 대응하는 명사가 있는 경우도 찾아서 포함시킴

C. 사전을 통해, 결정된 사항 확정

i. Oxford Advanced Learner's Dictionary (OALD)

ii. BBI Dictionary of English Word Combinations (BBI)

<코퍼스 수집 관련 자료>

안녕하세요 최준서 님

답장이 늦었네요. ^^;;

요청하신 문어 파트 파일들 링크 보내드립니다.

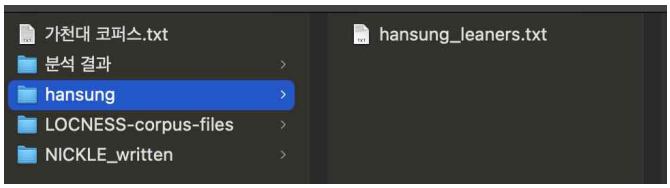
XML 파일로 되어 있습니다. 파싱을 해서 사용하셔야 할 거예요.

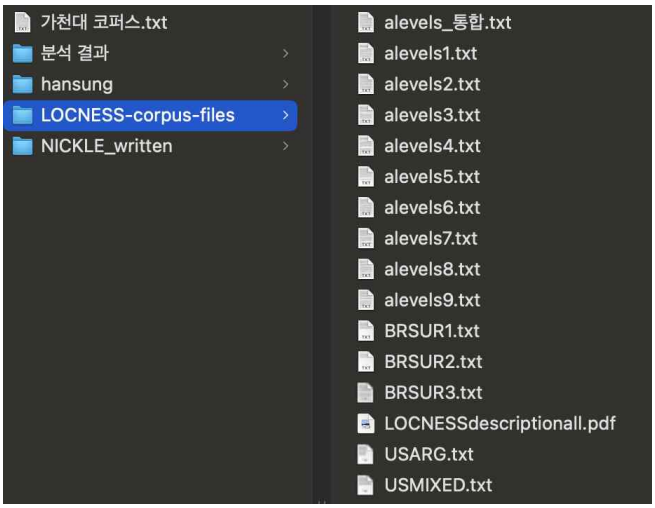
NICKLE_written

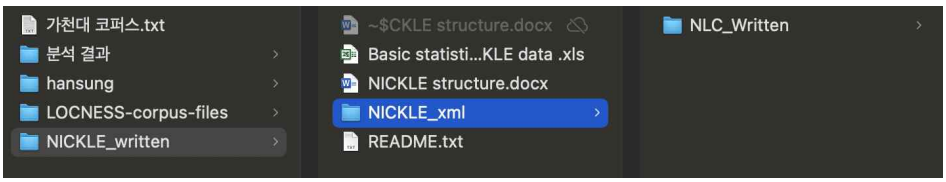
먼저 REAME 파일부터 확인하시기 바랍니다.

본 코퍼스 데이터는 요청하신 연구 목적 이외의 용도로는 사용이 불가하며 다른 목적으로 활용하고자 할 때에는 별도의 요청과 허가가 있어야 함을 다시 한번 말씀드립니다.

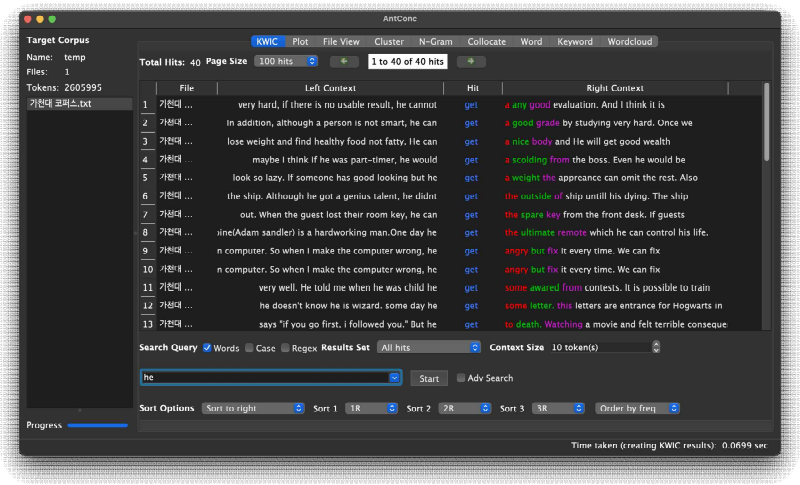
감사합니다.

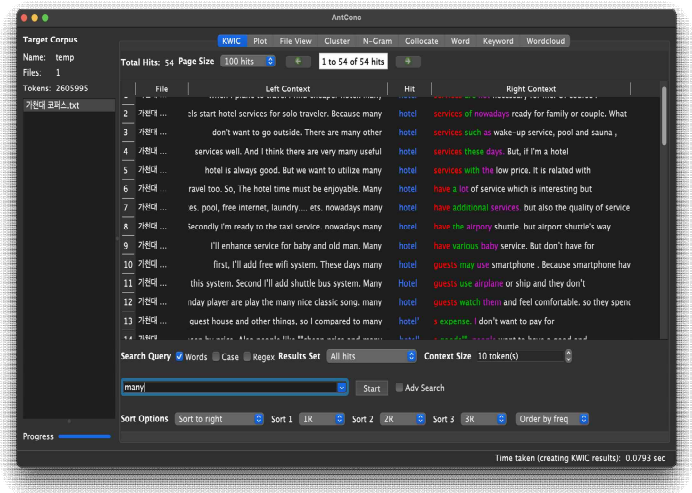



가천대 코퍼스.txt
분석 결과
hansung
LOCNESS-corpus-files
NICKLE_written
alevels_통합.txt
alevels1.txt
alevels2.txt
alevels3.txt
alevels4.txt
alevels5.txt
alevels6.txt
alevels7.txt
alevels8.txt
alevels9.txt
BRSUR1.txt
BRSUR2.txt
BRSUR3.txt
LOCNESSdescriptionall.pdf
USARG.txt
USMIXED.txt


가천대 코퍼스.txt
분석 결과
hansung
LOCNESS-corpus-files
NICKLE_written
~$CKLE structure.docx
Basic statisti...KLE data .xls
NICKLE structure.docx
NICKLE_xml
README.txt
NLC_Written

<데이터 분석 관련 자료>


AntConc
Target Corpus
Name: temp
Files: 1
Tokens: 2605995
가천대 코퍼스.txt
KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud
Total Hits: 40 Page Size 100 hits 1 to 40 of 40 hits
File Left Context Hit Right Context
Search Query Words Case Regex Results Set All hits Context Size 10 token(s)
he
Start Adv Search
Sort Options Sort to right Sort 1 1R Sort 2 2R Sort 3 3R Order by freq
Progress
Time taken (creating KWIC results): 0.0699 sec


AntConc
Target Corpus
Name: temp
Files: 1
Tokens: 2605995
가천대 코퍼스.txt
KWIC Plot File View Cluster N-Gram Collocate Word Keyword Wordcloud
Total Hits: 54 Page Size 100 hits 1 to 54 of 54 hits
File Left Context Hit Right Context
Search Query Words Case Regex Results Set All hits Context Size 10 token(s)
many
Start Adv Search
Sort Options Sort to right Sort 1 1R Sort 2 2R Sort 3 3R Order by freq
Progress
Time taken (creating KWIC results): 0.0793 sec

5. 자기평가 ※활동수행을 통해 발전, 성장한 내용을 상세히 기술

이번에 실제로 코퍼스 데이터를 수집해 보고 분석해 보면서, 데이터를 활용한 프로젝트를 하는 것에 대하여 실제적인 경험을 쌓을 수 있었다. 우선, 우리가 쉽게 접근할 수 있는 Open Access 파일이 아니라, 개인이 만든 자료를 사용할 경우, 허가를 받아야 한다는 점을 알 수 있었다. 다음으로, 내가 생각한 만큼 데이터 수집이 원활하지 못했다는 점에서, 방학 때 촉박하게 수집을 진행하는 것보다는 기간을 충분히 두고, 최대한 많은 데이터를 확보하는 것이 무엇보다 중요하다는 것을 알게 되었다. 그리고, 데이터를 수집하는 과정에서 전혀 나와 연고가 없는 사람들에게 메일을 쓰고, 전화를 하면서 실제 데이터를 얻고 나니, 다른 사람들에게 어떤 식으로 요청을 할 수 있는지에 대한 경험도 할 수 있어 좋았다. 영작문 자료를 분석하면서, 한성대학교 코퍼스는 대략적으로 어떤 수업 과정을 거치고, 어떤 주제에 대한 작문을 하였는지를 알 수 있었지만, 능률 코퍼스의 경우에는 그에 대한 정보가 부족하여 코퍼스 분석에 어려움이 존재했다. 또한, 한성대학교 학생들의 코퍼스를 수집할 때에는 학생들의 수준(영어 커뮤니케이션 반 등급 분류- 초급, 중급)으로 명확하였지만, 가천대학교 코퍼스는 토익 점수로 미루어 볼 수밖에 없었고, 능률 코퍼스의 경우에는 수준을 알 수 없었던 점이 아쉬웠다 생각하였다. 최종적으로 분석 작업을 할 때 한성대학교 학생들의 데이터가 적어 조금 더 유의미한 결과를 얻지 못해 아쉽지만, 그래도 도움을 준 학생들 덕에 대략적인 결론을 도출할 수 있었다. 마지막으로, 실제 분석을 하는 과정에서 매주 지도교수님께서 많은 도움을 주셨고, 보고서를 작성할 때 어떤 부분을 중점적으로 생각해야 하는지, 내가 놓쳤던 부분이 어떤 것인지에 대해 알려주셔서 이 또한 이번 진로탐색학점제에서 나의 능력을 성장할 수 있었던 좋은 기회였다.

5. 향후계획 ※진로탐색학점제를 통해 얻은 경험과 느낌을 바탕으로 향후 학교생활 및 진로에 대한 계획 기술

이번에 코퍼스 분석을 해 보면서, 잘 정제되어 있지 않은 데이터들을 조직화해서 분석하는 데 약간의 어려움을 느꼈고, 한 학기 수업을 들은 것으로는 많이 부족함을 느꼈다. 그래서, 이번 8월 23일 ~ 24일에 코퍼스언어학회에서 진행하는 '제 3기 코퍼스언어학학교'에 참석하여 실제로 코퍼스언어학에서 배우는 기술들, 최근 각광을 받고 있는 주제에 대한 내용을 더 배워보고자 한다. 또한, 이번 진로탐색학점제에서 진행한 프로젝트에서 경험한 것을 바탕으로 추후 영어 데이터 분석 캡스톤 디자인 수업이라던지, 개인 프로젝트를 통해 많은 연구를 해 보고자 한다.

또한, 내가 과연 연구자로서의 자질이 있을지 생각해 보았는데, 이번 진로탐색학점제를 진행하면서 내가 이러한 주제를 바탕으로 연구를 하는 것을 즐겨한다는 것을 깨달았고, 추후 많은 언어를 분석하는 언어학자가 되겠다는 마음을 굳게 다지는 계기가 되었다.