
말뭉치 문체론의 관점에서 본 한국어 발생 문법 양상

- '거든'의 공시·통시의 실증적 연구-

제조현(QIZHAOXIAN)
연세대학교 국문과 박사과정

목차

- ‘거든’의 공시·통시의 실증적 연구-

1. 연구 목적

2. 예비적 검토

- <1> 발생 문법: ① ‘GGUU’와 ‘GUUG의 논쟁’에서 ‘사용 기반 언어학’으로
② 발생 문법: 개념과 유형(한국어) ⇄ 매체, 문체, 사용역, 장르
- <2> 말뭉치 문체론: ① ‘문체론’에서 ‘말뭉치 문체론’으로
② 한국어의 구어 문법 연구

3. 연구 방법

- <1> 연구용 말뭉치의 구성: ① 현대국어 균형 ② 시기별·장르별·유형형
- <2> 연구 방법:
 - ① ‘전통적인 말뭉치언어학’에서 ‘다변량분석 말뭉치언어학’으로
 - ② 문체 통계학에서의 통계 방법

4. 공시적인 실험

- <1> 연구 질문: 텍스트 장르에 따른 접속과 문말 위치의 ‘-거든’의 담화 기능 차이
- <2> 연구 방법: 다중 대응분석(MCA), 혼합효과 로지스틱 회귀분석 등

5. 통시적인 실험

- <1> 연구 질문: ‘-거든’의 통시적인 의미변화와 문체 간의 상관관계
- <2> 연구 방법: 다차원적 척도분석(MDS analysis)

6. 맺음말

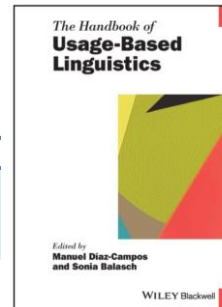
01
02
03
04
05

1. 연구 목적

이 글은 크게 두 가지 목적을 목표로 한다.

첫째, 발생 문법(잠정 문법, Emergent Grammar, 浮現文法)의 정의와 유형을 귀납적으로 정리하고, 그 유형 중 하나인 국어학계에서 주목받아 온 ‘종결어미화된 연결어미’(혹은 탈종속화) 현상의 연구사와 주요 쟁점을 밝히는 것이다.

둘째, ‘문체 변이(Stylistic Variations)’가 사회언어학과 변이언어학에서 하위 변인으로 작용하며, ‘생략, 도치, 문법화 등 다양한 요인으로 형성된 연결어미의 의미 확장 과정’에 실제로 영향을 미쳤는지를 탐구하는 데 있다. 이를 위해 ‘-거든’을 대상으로 공시적·통시적 말뭉치에서 추출한 용례를 분석하며, 문체 변인과 연결어미의 의미확장 간의 통계적 상관관계 및 교호작용을 다변량분석과 혼합효과 로지스틱 회귀분석을 통해 실증적으로 규명하고자 한다.



2.1. 한국어에서의 발생문법 양상

발생 문법(잠정 문법, Emergent Grammar, 浮現文法)은 기능주의 문법론자인 Hopper가 「Emergent Grammar」(1987)에서 제안한 개념으로, 언어 문법이 본질적으로 불확정성을 지니며 고정되거나 최적화된 체계가 아니라, 끊임없이 변화하고 진화하는 과정 속에 존재한다는 점을 강조한다. 이 이론은 문법(grammar)의 동적 성격과 사용 기반(usage-based)을 부각시켜 사용-기반 언어학(usage-based linguistics)의 형성과 발전에 중요한 영향을 미친다.

(1) 변화와 진화 ⇨ 대화분석 이론과 문법화론: 화청자 사이에 상호작용을 통해 문법이 발생, 발전, 변화됨

(2) 문법(grammar)와 용법(usage)의 논쟁 ⇨ Newmeyer(2003) vs. Bybee(2006)

Newmeyer, F. J. (2003). Grammar is grammar and usage is usage. *Language*, 79(4), 682-707. ⇨ GGUU

Bybee, J. L. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82(4), 711-733. ⇨ GUUG

2.1. 한국어에서의 발생문법 양상

(1) 변화와 진화 → 대화분석 이론과 문법화론

문법화는 특정 어휘나 구문이 화자들 상호작용의 과정에서 특정 기능을 수행하기 위해 형태의 변화와 새로운 화용적 의미를 갖게 되는 동적이고도 단일 방향의 역사적 변화라고 주장함. (Traugott, E. C, & König, E, 1991:189, 김해연, 2022:246)

(2) 문법(grammar)와 용법(usage)의 논쟁 → Newmeyer(2003) vs. Bybee(2006)

Newmeyer(2003): 정신 문법이 언어 사용에 기여하지만 **사용, 빈도 등은 문법 자체에 포함되지 않는다는 관점**

Bybee(2006): 문법을 **언어 경험의 인지적 조직화**로 간주하고 **사용 기반 문법(usage-based grammar)**을 주장함

GRAMMAR IS GRAMMAR AND USAGE IS USAGE

FREDERICK J. NEWMYER
University of Washington

FROM USAGE TO GRAMMAR: THE MIND'S RESPONSE TO REPETITION

JOAN BYBEE
University of New Mexico

01
02
03
04
05

2.1. 한국어에서의 발생문법 양상

한국어에서의 발생 문법 현상 👉 문체(제3 문체인 메신저 대화, 통신언어)와의 관련성

- | | |
|-----------------------------|--|
| (1) 담화표지 의 기능과 문법화 | 이기갑(1995): 이제, 참, 진짜; 이원표(1994, 1995): 그러니까, 왜냐하면 |
| (2) 종결어미 의 기능과 문법화 | 손성옥(1994): -지 않아(요) |
| (3) 연결어미 의 기능변화와 문법화 | 이효상(1999): -는데; 김명희(2013): -거든, -더라;
박미정·손성옥(2002): -거든, 손성옥(2003): -니까 |
| (4) 의존명사 의 종결 구성의 양상 | 구현정(2020): ‘듯’ 종결형, ‘줄’ 종결형 |

(1~4) 이성하(1998/2016), 김해연(2022)을 참고. (5) 구현정(2020) 참고.

이성하(2016:56-62) 참고 바람.

2.1. 한국어에서의 발생문법 양상

한국어에서의 발생 문법 현상 📖 문체(제3의 문체인 메신저 언어, 통신언어)와의 관련성

(4) 의존명사의 종결 구성의 양상 구현정(2020): ‘듯’ 종결형, ‘줄’ 종결형 [사용빈도 ↑]

의존명사는 통사적으로 홀로 쓰이지 못하고 통합관계의 제약이 있다. 그리고 선행하는 관형 성분의 유형과 후행하는 요소를 통해 그 의미적 특성을 규명할 수 있다. 구현정(2020)에서는 다양한 매체에서 사용된 의존명사를 조사한 결과, 후행 요소가 생략되고 의존명사만으로 문장이 종결되는 사례가 다수 발견됨.

- (1) 가. 대전 갈까 용인에서 할까 고민 중 (카톡 개인)
나. 자물쇠를 재설정한 것에 대한 분노뿐 (카톡 개인)
다. 지금 나랑 장난하는 것? (카톡 개인)
르. 개재미있을 듯 (카톡 개인)
미. 어디 좋은 불족발집 알아낸 줄 (카톡 개인)

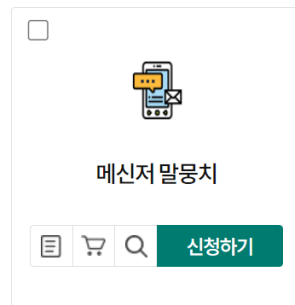
→ 서술어가 없이 종결되는 현상에서 일부는 탈종속화(insubordination) 현상으로 설명될 수 있음.

탈종속화: 기존에는 종속절의 일부로 사용되어 왔기에 종속절로 보이는 것이 관습적으로 주절로 사용되는 것.
(Evans, 2007; 문숙영, 2015, 2023; 박주원, 2020/2022, 한지수, 2020)

2.1. 한국어에서의 발생문법 양상

한국어에서의 발생 문법 현상 👉 문체(제3의 문체인 메신저 언어, 통신언어)와의 관련성

연구의 자료는 사용자들의 도움을 받아 수집한 실제 카톡을 활용한 대화 자료⁴⁾와 인터넷 블로그와 카페 등에 사용된 자료들을 검색하여 활용하였다. 자료의 출처는 모두 개인적이고 사생활이 드러나는 자료이기 때문에 개인 자료는 '카톡(/트위터/페이스북), 개인'으로, 블로그와 카페 등은 '블로그/카페 게시날짜'와 같이 표기하였다.



매체
문체
사용역
장르

매체: 통신언어, 전자 담화와 관련 연구

(1) 전반적 소개: 박용예(2012), 구현정(2016)

(2) 통신언어의 언어학적 특성: 이정복(2012), 이진성(2013)

(3) 통신언어의 비대면 의사소통: 안예림·양명희(2018), 강현석·김민지(2018)

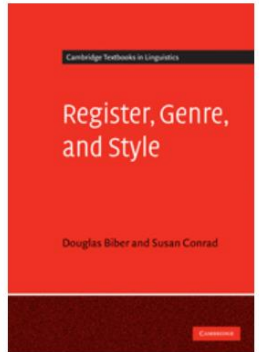
(4) 통신언어와 차별어: 이정복(2010, 2016), 박동근(2012), 김규훈(2019)

口語와 文語의 二分을 넘어 개별 使用域 資質에
기반하는 문법 연구 제안*

제3의 매체 - 메신저 언어 연구

01
02
03
04
05

문체(style)	언어학, 문학, 수사학, 말뭉치언어학	Biber
사용역(register)	사회언어학, 변이언어학, 담화분석	Labov
매체(media)		
장르(genre)	(체계) 기능 언어학, 제2 언어 습득, 글쓰기	Halliday, Hyland



<표 1> Biber&Conrad(2009)의 장르/사용역/문체의 연구 경향 비교

	장르	사용역	문체
대상 텍스트의 범위	전체	텍스트에서 발췌한 샘플	텍스트에서 발췌한 샘플
언어학적 특성	특정 표현, 수사적 구조화, 포맷	모든 어휘-문법적 자질	모든 어휘-문법적 자질
언어학적 특성의 분포	텍스트에서, 또는 텍스트의 특정 위치에 서 보통 1회 출현	변이에 따라 텍스트 전체에 자주, 전반적으로 나타남	변이에 따라 텍스트 전체에 자주, 전반적으로 나타남
해석	자질은 관습적으로 장르와 연관됨 : 예상되는 포맷이지만 종종 기능적이지 않음	자질은 사용역에서 중요한 의사소통적 기능을 제공	자질은 직접적으로 기능적이지 않음 : 이들은 미학적인 가치가 있어 선호된다.



안의정(2014:112)을 참고.

2.2. 한국에서의 문체학 연구와 구어문법 연구

차례

발간사

1부

1-1	김홍수 선생의 문체 연구	장소원	12
1-2	언어학적 문제론의 위상과 과제	김홍수	43
1-3	詩의 言語 分析 試論 -金祿暎의 詩에 대한 語學的 接近-	김홍수	59
1-4	국어의 통사현상과 문체	김홍수	88
1-5	문체 의미론 -어휘·형태와 통사 총위를 중심으로-	김홍수	124
1-6	문학 텍스트와 문체론	김홍수	165
1-7	국어문체론 연구의 현 단계와 어학적 문제론	김홍수	191
1-8	이론과 개화기의 표기체 유형과 양상	김홍수	214
1-9	소설의 방언에 대하여	김홍수	239

2부

국한문체

국한문체	김주필	272
현대국어와 역사상: 문체	장소원	279
近代啓蒙期 國漢混用文의 類型·文體 特性·使用 樣相	한영균	323
대한제국 시기 國漢文의 형성과 기원		367
-諺解와 관련하여-	김주필	
조선 전기 이두 번역문의 문체와 어휘	박상홍	403

계량적 문체 연구

계량적 문체 연구	임근석	440
-----------	-----	-----

한국 현대소설의 기계적 문체 분석 가능성을 위한 계량적 방법론	447	
-1930년대 작가들 중심으로-	문한별	
지자 판별을 위한 전산 문제론	481	
-조기 현대소설을 대상으로-	김일환·이도길	
계량적 문체 분석 시론	박진호	519
광고 언어에 대한 코퍼스 언어학적 분석 시고	임근석	587

구어 담화

구어 담화	이동운	620
구어 연구의 현황과 전망	장경희	629
구술발화의 전사와 분석	이기갑	659
말뭉치와 언어연구 -외국의 사례와 경향-	최재용	689
한국어 구어 담화 연구를 위한 전사와 주석에 대한 논의	이동운	727

문학 속의 언어

문학 속의 언어	정선태	750
유감준의 국한문체 기획과 문체의 전환		755
-신체호 최남천과의 비교연구-	임삼석	
서구시 번역을 통한 근대 한국어 문장 형성 연구		785
-김억의 베를렌스와 보들레르 번역을 중심으로-	배수한	
백석 시의 방언과 그 미학적 의미	전봉관	811
표준어의 집형, 지역어의 내부식민지화		861
-원기영의 「술이삼촌」을 중심으로-	정선태	

김홍수 교수님 논저 목록

《글의 무늬를 찾아서(김홍수 선생 정년 기념논총)》

- (1) 국한문체 『역사, 표기법』
- (2) 계량적 문체 연구 『말뭉치언어학』
- (3) 구어 담화 『구어, 문어』
- (4) 문학 속의 언어 『문학』

이를 통해 문체학과 관련 분야를 확인할 수 있음.

01
02
03
04
05

2.2. 한국에서의 문체학 연구와 구어문법 연구

그리고 한국어의 문체학의 연구사에 대해 참고할 만한 내용은 다음과 같다.

문체, [bookchapter](#)

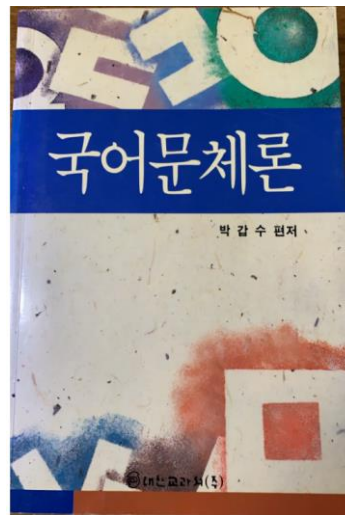
박갑수(1990), 文體, 서울대학교 大學院 國語研究會 편, 《國語研究 어디까지 왔나》, 東亞出版社, 582-593.

김홍수(1992), 국어문체론 연구의 현 단계와 어학적 문체론, 국어국문학회 편, 《국어국문학 40년》, 집문당,

장소원(1998), 문체, 서태룡·민현식·안명철·김창섭·이지양·임동훈 공편, 《문법 연구와 자료(이익섭 선생 회갑 기념 논총)》, 태학사.

양오진(1999), 문체론 연구사, 김민수 편, 《현대의 국어연구사: 1945~1998(수정 증보판)》, 박이정, 211-226.

장소원(2024), <문체의 시대적 고찰>의 이해, 《이승녕 현대국어학의 선구자: 회상, 학술 활동》, 논저의 이해》, 700-703.



2.2. 한국에서의 문체학 연구와 구어문법 연구

구어문법의 연구

일찍이 <우리말본>(1937/1971)의 일러두기에서 입말(구어)과 글말(문어)에 대한 논의가 있었음.

- (1) a. “어느 나라의 말에든지 제각기 일정한 본(法)이 있나니, 그 본을 말본(語法)이라 하며, 그 말본을 닮는 학문을 말본갈(語法學), 더러는 줄이어서 말본(語法)이라 하느니라”(최현배, 『우리말본』 1937:36)
- b. “이 책은, 오늘날에 實際로 쓰히는 입말(口語)과 글말(文語)의 본(法)을 풀이한 것이다.”(최현배, 『우리말본』 1937:일러두기1)
- c. “우리말에 입말(口語)과 글말(文語)과의 다름이, 어떤 경우, 特히 풀이씨(用言)의 씨끝(語尾)에서, 얼마큼 나타남이 없지 아니하지마는, 이 책에 서는 그것을 區別하지 아니하고 섞어 썼노니: 이는 하나는, 그 區別이 必要가 없다고 생각한 때문이요; (하략) (최현배, 『우리말본』 1937:일러두기3)

서상규(2008)를 참고.

2.2. 한국에서의 문체학 연구와 구어문법 연구

그리고 구어체와 문어체의 특성과 관련 연구는 장소원(1986), 민현식(1994), 노대규(1996) 등이 있음.
1999년 이전의 연구.

言語	口語	口語體 口語(1a)	口語體 言語
		文語體 口語(2a)	
	文語	口語體 文語(1b)	文語體 言語
		文語體 文語(2b)	



- ㄱ. 구어체 구어: 일상 대화.
- ㄴ. 문어체 구어: 강단어, 연설어, 보도어처럼 글이 전제되는 구어이다.
- ㄷ. 구어체 문어: 소설, 시나리오, 희곡, 광고문처럼 입말 중심의 문어이다.
- ㄹ. 문어체 문어: 논설문, 기사문처럼 읽거나 씀을 전제로 한 문어이다.

이후, 김미형(2004), 장경현(2006), 안의정(2009), 전영옥(2009), 홍종선(2022, 2023) 등이 있음.

01
02
03
04
05

2.2. 한국에서의 문체학 연구와 구어문법 연구

구어 문법과 말뭉치 언어학

Biber(1991): 다차원분석과 인자분석을 통해 영어의 구어와 문어의 특성을 밝혔다.

1999년 이전의 연구: 다차원분석, 인자분석, G2값 등 방법의 활용

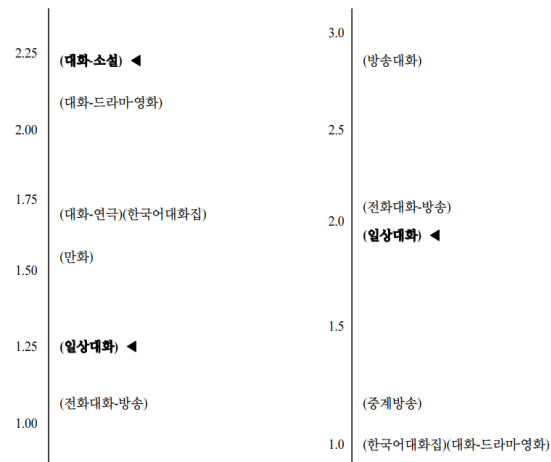


최근에 강혜림·백재파(2020), 안소진(2024) 등이 있음
인자분석을 통해 소설 자료의 특성을 밝혔다.

Variation
across speech
and writing

Douglas Biber

(6) 가. [차원 1]의 인자점수 평균 플롯 나. [차원 2]의 인자점수 평균 플롯



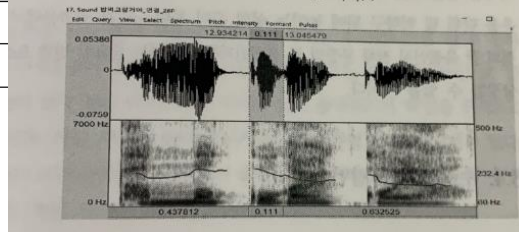
2.2. 한국에서의 문체학 연구와 구어문법 연구

서론에서 제시한 바와 같이, 발생문법은 대화분석 이론과 문법화론과 밀접한 관련이 있으며, 네 가지 유형으로 분류될 수 있다. 이 중 종결어미화된 연결어미, 즉 탈종속화와 관련된 현상에 대해 살펴보고자 한다.

	연구	연구 방법론
1	김태엽(1999)	문법화(문장 구조의 축소→문법 기능의 어전→ <u>끊어짐 수행</u> → <u>억량</u> <u>없힘</u> →문장 종결 기능의 획득)
2	구현정·이성하(2001)	문법화와 대화분석(담화 전략의 생략)
3	유현경(2003)	도치와 생략, 문법화의 정도성
4	한길(1991, 2006)	어찌마디(부사절)의 도치
5	조민하(2011), Mee-Jeong Park(2012)	실험 음성학
6	문숙영(2016), 김주원(2020), 한지수(2020)	<u>탈종속화</u>

‘-(으)니까’, ‘-거든’, ‘-(으)려고’, ‘-다고’ 등이 있음

(11) 접속문의 연결어미 ‘-고’ (“밥 먹고 갈 거야.”)



2.2. 한국에서의 문체학 연구와 구어문법 연구

선행연구에서 제시한 선정 기준

- (1) 빈도: 하지선(2006), 조민하(2011), 임현정(2012), 손마정(2014) 등
- (2) 기능 변화: 김태엽(2000), 이종희(2004), 손옥현(2009) 등
- (3) 의미 및 기능 변화: 박재연(1998), 전영전(2001), 김유진(2008) 등
- (4) 기타(선행연구+의미분류+목록화): 이은경(2000), 유현경(2003), 손옥현·김영주(2009), 목지선(2015), 박주원(2020/2022), 고우미(2021), 이후인(2022) 등

(1) 연결 어미의 의미 범주의 연구사: 주시경(1910), 최현배(1937), 권재일(1985), 채연강(1985), 정경덕(1985), 최재희(1989), 전혜영(1989), 윤평현(1989/2005), 이은경(2000, 2015, 2016, 2018), 박진희(2012)

(2) 개론서: 허용(1983, 1999, 1995/2000), 김석득(1992), 임흥빈·장소원(1995), 서경수(1994, 1996), 이익섭·채완(1999), 남기심·고영근·유현경·최형용(2019), 이관규(1999/2002/2012/2023), 홍종선(2023), 유현경·한재영·김홍범·이정택·김성규·강원화·구본관·이병규·황화상·이진호(2018/2023)

(3) 사전: <한국어 교육 문법_자료편>, <한국어 유사 문법 항목 사전>, <한국어 구어 빈도 사전 1, 2>, <현대 한국어 어휘 빈도 사전 1, 2>, <외국인을 위한 표준 한국어 문법> 등

예: 손옥현·김영주(2009)에 따르면, 종결기능의 연결어미의 목록의 연구사는 다음과 같다.

<1> <표준국어대사전>: -거니, -거들랑, -거든, -게, -고, -다고, -는데, -는지, -나, -니, -도록, -을세라, -려고, -어, -지

<2> 유현경(2003): -고, -어/아서, -고/고서, -(으)면, -(은)는데, -지만, -(으)니까, -(으)려고, -(으)러, -든지, -느라고, -거든, -게, -도록

<3> 권재일(2003): -거든, -는데, -고/구, -으니까, -어 가지구, -어서, -으려고, -으면서, -지만, -게, -느라고, -더니, -어도, -으라고, -으면

<4> 하지선(2006): -고, -는데, -거든, -다고, -어서, -다니까, -면서, -다면서, -는지, -다며, -면, -려고, -지만, -든지

<5> 손옥현 외(2009): -거든, -게, -고, -는데, -다고, -다니까, -다면서, -려고

3. 연구 방법

3.1. 연구용 말뭉치의 구성

(1) 현대국어 균형 말뭉치: 새연세말뭉치4

검색 도구: 말뚝 2017

현대국어 균형 말뭉치는 연세대학교 서상규 교수가 배포한 ‘새연세말뭉치 4(20231010버전)’이며, 용례 검색은 미국 벤더빌트 대학의 장석배 교수가 제작하여 제공한 주석 말뭉치 용례 검색프로그램인 ‘말뚝2017’을 사용하였다.

[표2] 새연세말뭉치4의 구성

구분	세부 장르	특성	표본수	마디수(어절수) 합	표본 크기
구어	공적대화	공적 장면, 상호작용	30	150,014	448~13,756
	공적독백	공적 장면, 비상호작용	44	150,048	450~7,343
	사적대화	사적 장면, 상호작용	65	150,004	297~5,295
	사적독백	사적 장면, 비상호작용	59	150,013	270~7,906
	국회회의록	회의 속기록	33	150,014	1,142~14,638
문어	학술	논설과 설명 텍스트	40	150,009	559~5,080
	신문	보도와 논설 텍스트	34	150,010	396~5,087
	잡지	보도, 논설, 상상적 텍스트	36	150,088	688~4,903
	소설	상상적 텍스트	37	150,010	208~4,995
	준구어	희곡, 대본, 시나리오, 신문 잡지의 인터뷰와 좌담 기사 텍스트	50	150,000	459~5,006
합계			428	1,500,210	270~14,638

새연세 말뭉치4

[표1] 배포용 말뭉치의 특성과 개요

특성	이름	nYsc4 <새연세말뭉치 4>
구성 및 주석		서상규 교수(연세대학교 국어국문학과)
언어		현대 한국어 구어, 문어
말뭉치 성격		문구어 10개 장르 비교 균형 말뭉치
규모		150만 마디(어절)
바탕 말뭉치		연세 말뭉치, 21세기세종계획 말뭉치 등
구성 방법		1) 문어와 구어를 대표하는 장르 각각 5개를 선정하여, 장르별 표본 크기를 같게 구성함 2) 모든 낱말과 형태의 동형어를 구분하여 세밀한 형태 주석 말뭉치로 만들 (<연세한국어사전>(1998), <표준국어대사전>(1999), <고려대한국어대사전>(2011)의 동형어 구분을 순차적으로 따름)

국어역사말뭉치

KOHICO 말뭉치 업데이트 일자(형식)

2024년5월(txt통합파일, 유니코드용)

2024년5월(2b통합파일, 깜짝새용)

3. 연구 방법

(2) 역사 자료 말뭉치의 구성 방법

- 21세기세종계획(1998~2007)
- 역사자료종합정비(2015, 2016)
- KOHICO 말뭉치(2024)
- 모두의 말뭉치, 국어 역사 말뭉치(2024)

검색 도구: Emeditor, 파이썬, uniconc, 깜짝새

문체는 국어학과 국문학의 중요한 주제로 여겨져 왔다. 국어학 학계에서 박갑수(1994), 김홍수(1988, 1992, 1994, 1996) 등이 문체와 어휘, 통사, 번역의 관계를 논의하였다. 아울러 역사 자료의 문체 특성을 주목해서 문법 현상의 사용과 문체와 관련성을 탐구한 박사 학위 논문은 서상규(1991), 김영수(2001) 서형국(2005), 강용택(2007) 등이 있다.

<표>

	논저	역사 자료의 문체 유형
1	서상규(1991)	<1> 역학서; <2> 의학서; <3> 불경언해; <4> 교과서-1, 교과서-2; <5> 간찰; <6> 시가
2	김영수(2001)	<1> 산문체; <2> 시가체; <3> 대화체.
3	서형국(2005)	<1> 경전류: ① 향가, ② 고가요, ③ 정음으로 창작된 첫 제품, ④ 번역 시가, <2> 열전류; <3> 안내서; <4> 역학서; <5> 인간; <6> 운문류
4	강용택(2007)	<1> 시가체: ① 향가, ② 고가요, ③ 정음으로 창작된 첫 제품, ④ 번역 시가, ⑤ 시조, ⑥ 가사, ⑦ 잡가; <2> 산문체: ① 번역 산문체, ② 이야기 소설, ③ 판소리 소설, ④ 규방 산문; <3> 대화체.

① 15세기 문헌

분류	문헌명	간행 연도	총 어절 수
안내서	訓民正音	1446	989
불경	釋譜詳節	1447	34730
운문류	龍飛御天歌	1447	2101
불경	月印釋譜 권1~10	1459	41796
불경	楞嚴經諺解 권1	1461	9528
불경	圓覺經諺解 권1-1	1465	13653
안내서	救急方	1466	17802
불경	蒙山和尚法語略錄諺解	1467	2651
열전류	三綱行實圖	1471	10346
운문류	分類杜工部詩諺解 권3~11	1481	30788
불경	南明泉禪師繼頌諺解	1482	23886
안내서	救急簡易方 권1~2	1489	17752
합계			206,022

<표 2> 15세기 자료 문헌

(2) 연구용 시기별·장르별·유형별 역사자료 말뭉치의 구성

시기별 유형		15세기	16세기
장르와 하위 유형			
(가) 시가체	(a) 용비어천가 ^{1447(후기)}	—	—
	(b) 두시언해 ^{1489(후반)}	—	—
	(c) 훈민정음(언해본) ^{1447(년월불문)}	—	—
	(d) 석보상절 ^{1447(약간)}	—	—
	(e) 월인석보 ^{1438(약간)}	—	—
	(f) 농암경언해 ^{1440(후반)}	—	—
	(g) 금강경언해 ^{1446(후반)}	—	—
	(h) 내훈 ^{1447년경(내훈)}	—	—
	(i) 구급방언해 ^{1448(후반)}	—	—
	(j) 번역소학 ^{1511(후반)}	—	—
(나) 산문체	(k) 소학언해 ^{1517(초반)}	—	—
	(l) 맹자언해 ^{1592(명인)}	—	—
	(m) 번역 박물사 ^{1517(명인)}	—	—
	(n) 춘전 견제 묘 출도 한글 편지 ^{16세기(초반)}	—	—
(다) 대화체		—	—
(라) 서간체(書翰體, 諺簡)		—	—

② 16세기 문헌

분류	문헌명	간행 연도	총 어절 수
역학서	翻譯老乞大	1517 이전	11325
역학서	翻譯朴通事 권상	1517 이전	6073
안내서	簡易辟瘟方	16세기 초	1619
열전류	續三綱行實圖	1514	3494
유경	翻譯小學 권6~10	1517	16372
유경	呂氏鄉約諺解	1518	3279
열전류	二倫行實圖	1518	3684
유경	正俗諺解	1518	3352
안내서	牛馬羊猪染疫病治療方	1541	668
안내서	分門瘟疫易解方	1542	1816
언간	순천김씨묘출도간찰	16세기 중반	19085
불경	七大萬法	1569	2739
운문류	百聯抄解	1576	2827
불경	初發心自警文	1577	3122
불경	禪家龜鑑	1579	7237
유경	小學諺解	1586	29547
언간	김일근(1991) 소수 언간	16세기	656
합계			116,895

<표 3> 16세기 자료 문헌

③ 17세기 문헌

분류	문헌명	간행 연도	총 어절 수
언간	현풍곡씨 언간	17세기 전반	13285
언간	김일근(1991) 소수 언간	17세기	9180
안내서	諺解胎產集要	1608	7140
안내서	諺解痘瘡集要	1608	12933
열전류	東國新續三綱行實圖	1617	24538
안내서	新傳煮硝方諺解	1635	1521
안내서	火炮史諺解	1635	1770
안내서	辟瘟新方	1653	1465
역학서	老乞大諺解	1670	11447
역학서	원간활자본 捷解新語	1676	6603
역학서	朴通事諺解	1677	16838
안내서	馬經抄集諺解 권상	1682	8972
합계			115,692

<표 4> 17세기 자료 문헌

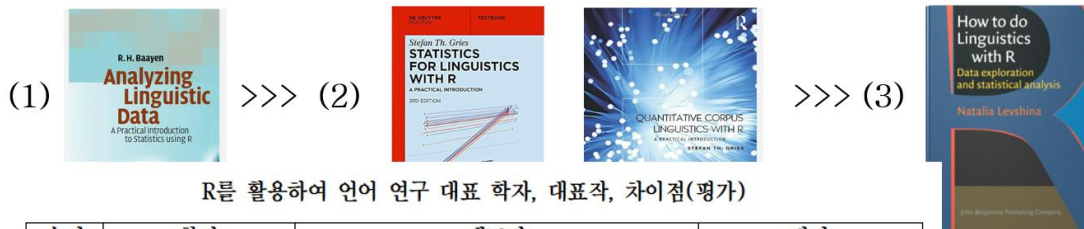
서형국(2005)을 참고.

<중세국어 의미빈도 사전 편찬용> 말뭉치
2023년 사전학회 발표문 참고.

01
02
03
04
05

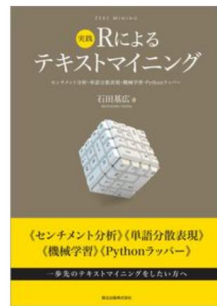
3. 연구 방법

3.2. R를 활용한 다변량분석 방법론



R를 활용하여 언어 연구 대표 학자, 대표작, 차이점(평가)

순서	학자	대표작	평가
1	Baayen, R. H.	Analyzing linguistic data a practical introduction to statistics using R(2008)	초창기
2	Gries, S. T.	Quantitative corpus linguistics with R(2009/2016) Statistics for linguistics with R(2013/2021)	(1) 통계 (2) Rling 등 패키지를 사용하지 않음 (3) 한국어 번역본: 홍정하·최재웅(2013)
3	Levshina, N.	How to do linguistics with R(2015)	(1) 통계 (2) Rling 등 패키지를 사용하였음 (+) 본 연구의 사용 방법(언어 모델링)



3. 연구 방법

3.2. R를 활용한 다변량분석 방법론

‘전통적인 말뭉치언어학’에서 ‘다변량분석 말뭉치언어학’으로

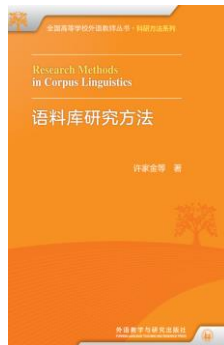
단일 요소의 특징만을 주목한 ‘전통적인 연구 방법’

(1) 언어 특징을 계량적으로 분석하기

- ① 어휘의 빈도를 계산하고 빈도표를 작성하기
- ② 연어의 빈도를 계산하고 빈도표를 작성하기
- ③ 품사별로 어휘나 구성의 빈도를 계산하고 비교하기
- ④ 담화 기능, 의미 기능, 은유나 환유를 나타낸 형태의 빈도를 계산하고 비교하기
- ⑤ 통계적인 기법(카이제곱, logistic regression 등)을 통해 언어 특징의 차이를 계산하거나 키워드(주제어) 표를 작성하기

(2) 어휘 문법의 공선택을 통해 문맥 의미를 파악하기

- ① 색인 분석(t-검정, MI-값 상호정보, 로그-우드 측정 등)을 통해 어휘의 결합 관계를 묘사하고 통계적으로 해석하기
- ② 부분 문법(local grammar)을 통해 교육적인 측면에서 문맥 의미를 해석하기
- ③ 결합 강도 분석
- ④ 다차원분석



다변량의 특징을 고려한 ‘새로운 연구 방법’ (후 뉴노멀 시대)

① **문장 내부**의 특징(어휘나 연어의 의미와 통사)과 **문장 외부**의 특징(담화, 사회언어학적 변인)을 종합적으로 고려해서 **공선택**을 통해 **형식-의미** 간의 관계를 파악하기

② **다변량 분석**: 주성분 분석, 인자분석, 대응분석, 다차원분석, 군집분석 등

③ 통계 방법의 활용: 로지스틱 회귀분석, 베이지안 회귀분석, 의사 결정 나무, 무작위 숲, 선형 판별분석 등



01
02
03
04
05

머신러닝 알고리즘 유형

	범주형	수치형
지도학습 정답 있음	분류 classification	회귀 regression
비지도학습 정답 없음	군집화 Clustering	차원축소 Dimensionality reduction

연구 방법



박조은·송영숙(2023)을 참고.

한국텍스트언어학회 2023 여름 북 콘서트
모두의 한국어 텍스트 분석
with 파이썬

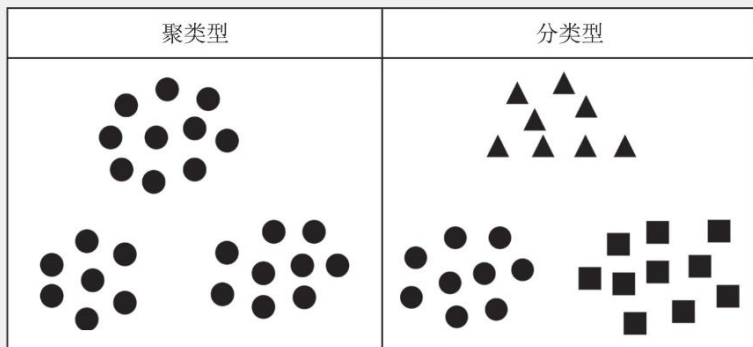
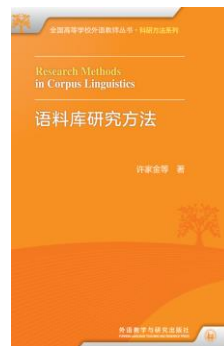


图1.1 聚类型和分类型方法的示意图²



4. 공시적인 실험

4. 공시적인 실험

<1> 연구 질문: 텍스트 장르에 따른 접속과 문말 위치의 ‘-거든’의 담화 기능 차이

<2> 연구 방법: 다중 대응분석(MCA), 혼합효과 로지스틱 회귀분석 등

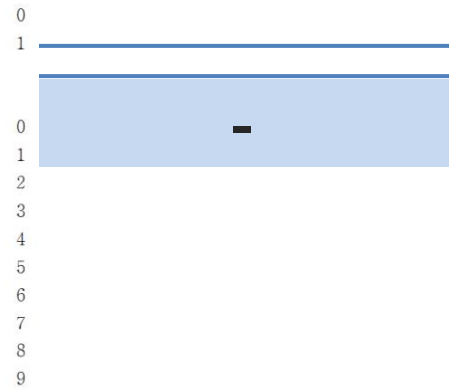
방법	탐색적 분석	확인적 분석
다중 대응 분석 (MCA)	✓	
구문 결합 분석	✓	
주성분 분석 (PCA)	✓	
요인 분석 (Factor Analysis)	✓ (탐색적 요인 분석)	✓ (확인적 요인 분석)
다차원 척도 분석 (MDS)	✓	
군집 분석 (Cluster Analysis)	✓	
의사결정나무 (Decision Tree)	✓	✓
랜덤 포레스트 (Random Forest)	✓ (변수 중요도)	✓
로지스틱 회귀		✓
혼합효과 로지스틱 회귀		✓

사용역	총 빈도수	유형	빈도 수(절대 빈도)	장르(공공성, 상호작용성 등)	빈도 수(절대 빈도)	유형	빈도 수(절대 빈도)
구어	909	연결	80	사적대화	453	연결	39
						중결	414
		중결	829	사적독백	356	연결	30
						중결	326
				공적대화	90	연결	10
						중결	80
문어	95	연결	22	공적독백	10	연결	1
						중결	9
		중결	93	국회회의록	0	연결	0
						중결	0
				준구어	44	연결	14
						중결	30
				소설	33	연결	3
						중결	30
		연결		잡지	10	연결	3
						중결	7
		중결		신문	0	연결	0
						중결	0
		연결		학술교양	8	연결	2
						중결	6

- 01
- 02
- 03
- 04
- 05

<1> 사용역(register): Regi
구어(spoken language): Oral
문어(written language): Writ

<2> 장르(genre): Genre
공적 대화 (public conversation): PuCo
공적 독백 (public monologue): PuMo
사적 대화 (private conversation): PrCo
사적 독백 (private monologue): PrMo
준구어 (semi-spoken language): SeSp
소설 (novel): Novl
잡지 (magazine): Magz
학술 교양 (academic and educational materials): AcEd
국회 회의록 (parliamentary minutes): Parl
신문 (newspaper): News



<3> 담화 기능(discourse function): DFnc	
DiscFA	0
DiscFB	1
DiscFC	2
DiscFD	3
DiscFE	4
DiscFF	5

<4> 결합어(construction collocation): CoCo

<5> 의미운율(semantic prosody): SePr	
적극적(positive): Posi	0
중립적(neutral): Neut	1
소극적(negative): Nega	2

<6> 인접 범주(colligation): Collig	
동사(verb): Verb	0
보조용언(auxiliary predicate): AuxP	1
지정사(copula): Copl	2
형용사(adjective): Adjv	3

<표 1> 언어학적 층위에 따른 어휘 점화의 유형

	정의	관련 용어
A.형태·통사	단어나 구가 공기하는 문법범주, 문법 기능, 또는 특정 어휘 연쇄에서의 특정 위치를 선호하거나 회피하는 것.	연접범주
B.어휘·의미	단어나 구가 특정 어휘의 형태나 의미의 부류와 공기하는 것을 선호하는 것.	연어, 연어관계, 의미적 선호
C.담화 화용	단어나 구가 특정한 화용론적 기능들과 함께 나타나도록 연관되어 있는 것.	의미적 운율

<7> 구어 변이형(spoken variant form): SpVa

<8> 위치(position): Posi

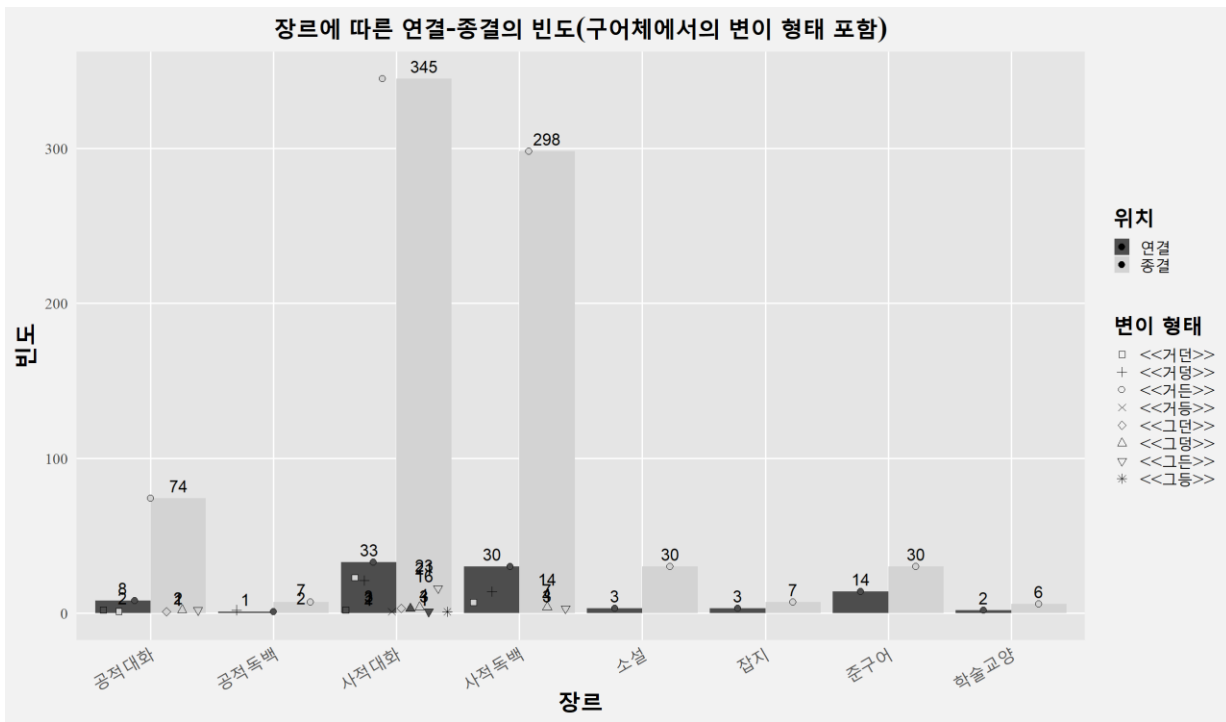
01
02
03
04
05

4. 공시적인 실험

		← →	🔍 Filter							
	1002	사용역	장르_통계용	담화_기능	종결_연결	구어형	결합어	인접범주	의미운율	
1	거든	구어	사적대화	A 조건이나 가정을 제시하기	연결	<<NULL>>	오_w_y1	동사	b 중립적	
2	거든	구어	사적대화	C 이유이나 원인 제시하기	종결	<<그던>>	낫_w_y1	동사	a 적극적	
3	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<거당>>	겔리_w_y1	동사	c 소극적	
4	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<NULL>>	있_vx_y3	보조용언	c 소극적	
5	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<NULL>>	않_va_y0	형용사	b 중립적	
6	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<NULL>>	갔다오_w_y0	동사	b 중립적	
7	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<그당>>	이_vc_y3	지정사	b 중립적	
8	거든	구어	사적대화	D 배경이나 전제 제공하기	연결	<<NULL>>	이_vc_y3	지정사	b 중립적	
9	거든	구어	사적대화	C 이유이나 원인 제시하기	종결	<<NULL>>	되_w_y1	동사	b 중립적	
10	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<NULL>>	이_vc_y3	지정사	b 중립적	
11	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<그든>>	있_vx_y3	보조용언	b 중립적	
12	거든	구어	사적대화	C 이유이나 원인 제시하기	종결	<<그든>>	하_w_y1	동사	a 적극적	
13	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<NULL>>	보_vx_y2	보조용언	b 중립적	
14	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<NULL>>	있_va_y1	형용사	a 적극적	
15	거든	구어	사적대화	D 배경이나 전제 제공하기	종결	<<NULL>>	이_vc_y3	지정사	b 중립적	

01
02
03
04
05

4. 공시적인 실험



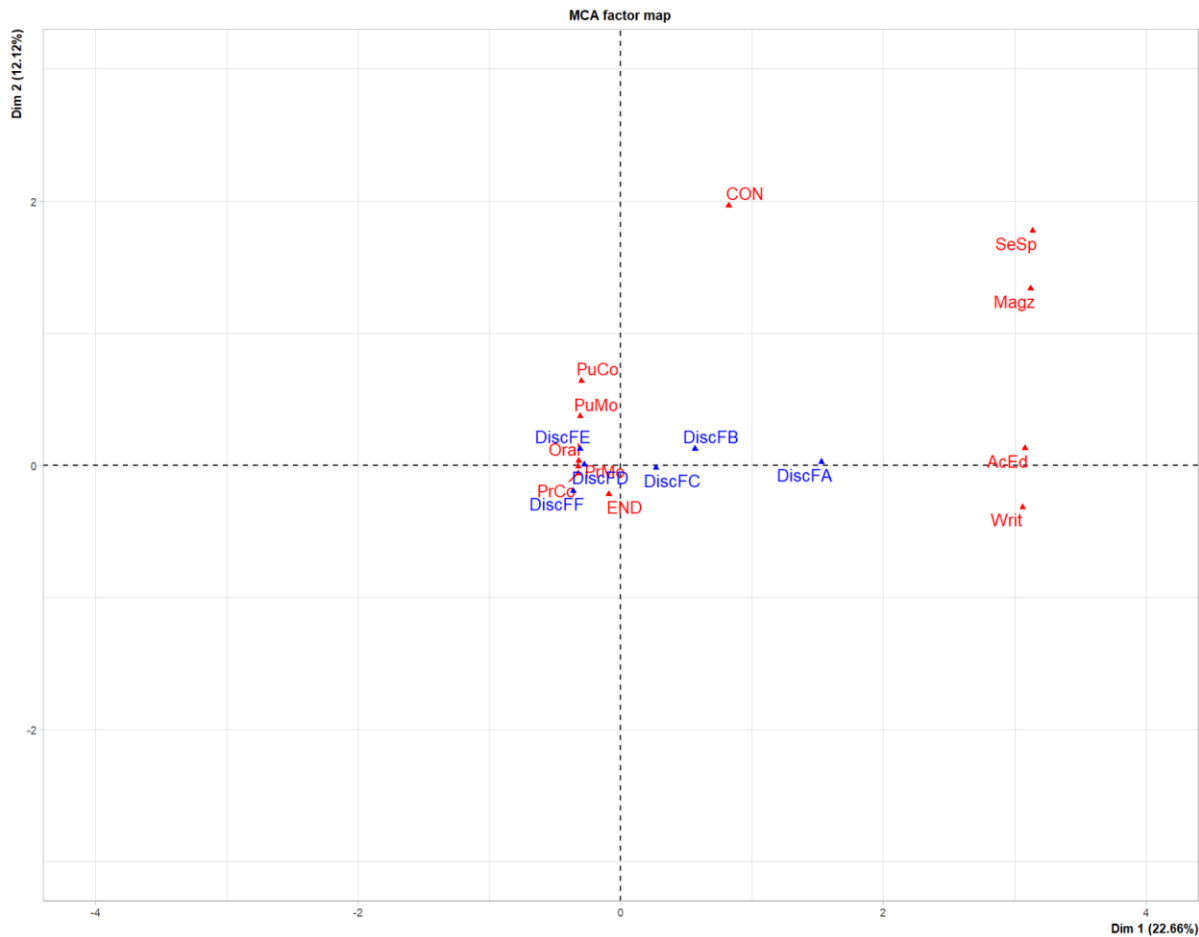
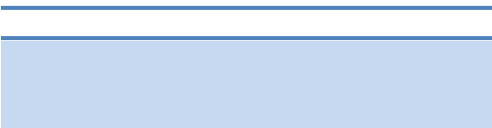
```
> df %>%
+   count(종결_연결)
# A tibble: 2 x 2
  종결_연결     n
<chr>     <int>
1 연결       100
2 종결       902

> df %>%
+   count(사용역)
# A tibble: 2 x 2
  사용역     n
<chr>   <int>
1 구어    907
2 문어     95

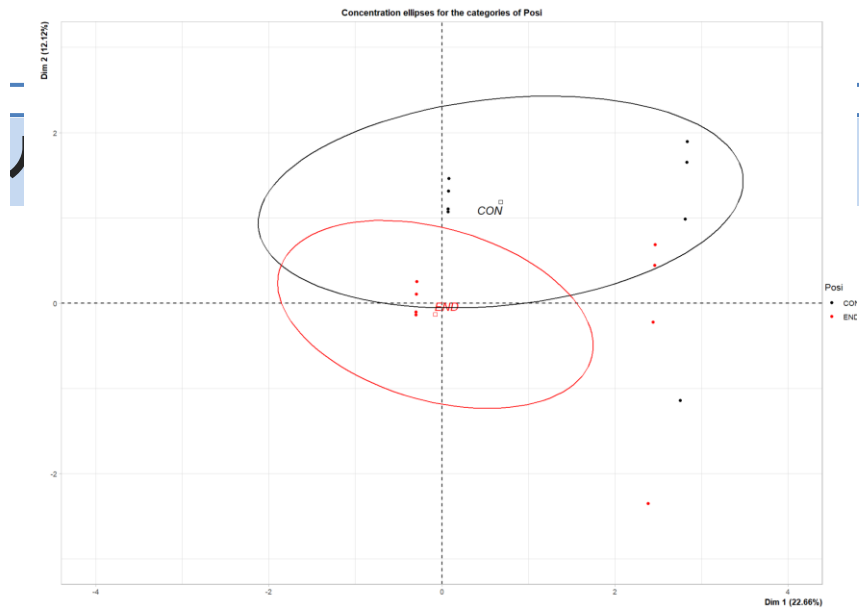
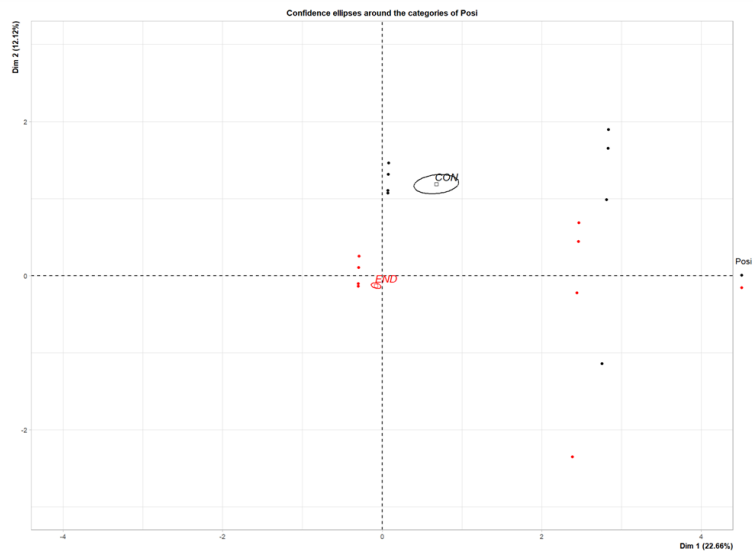
> df %>%
+   count(df$사용역, df$종결_연결)
# A tibble: 4 x 3
  df$사용역 df$종결_연결     n
<chr>     <chr>     <int>
1 구어      연결       78
2 구어      종결      829
3 문어      연결       22
4 문어      종결       73

> df %>%
+   count(장르_통계용)
# A tibble: 8 x 2
  장르_통계용     n
<chr>     <int>
1 공적대화     90
2 공적독백     10
3 사적대화    451
4 사적독백    356
5 소설        33
6 잡지        10
7 준구어      44
8 학술교양     8
```

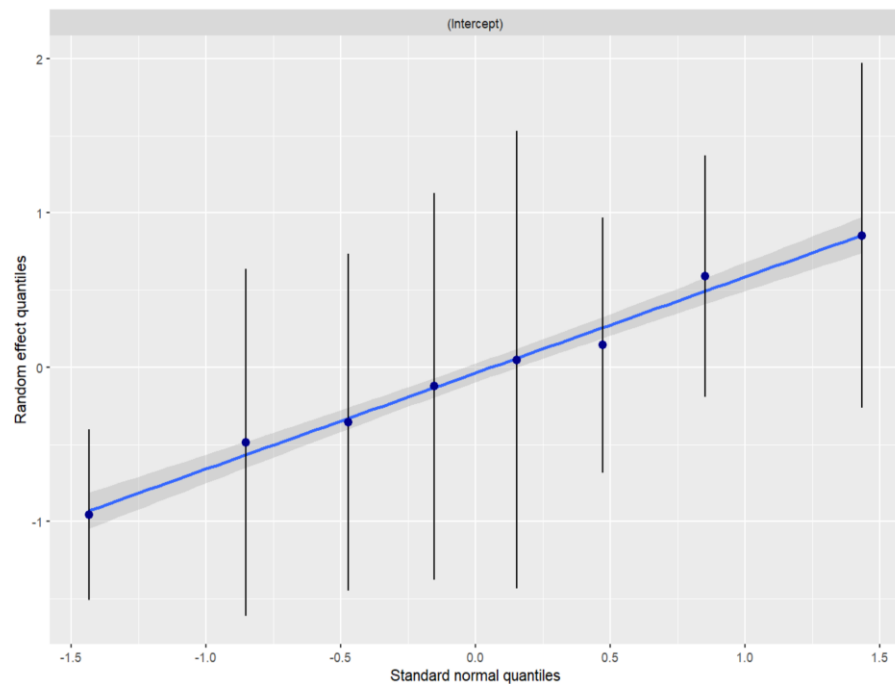
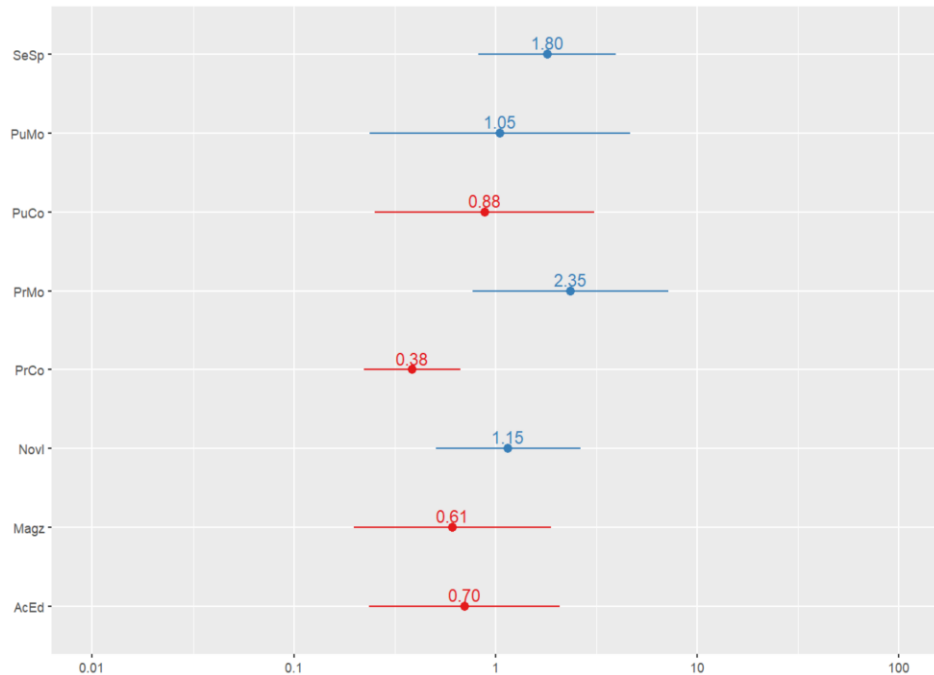
01
02
03
04
05



- 01
- 02
- 03
- 04
- 05



Random effects



01

02

03

04

05

5. 통시적인 실험

5. 통시적인 실험

<1> 연구 질문: ‘-거든’의 통시적인 의미변화와 문체 간의 상관관계

<2> 연구 방법: 다차원적 척도분석(MDS analysis)

6. 맺음말

1. 발생문법의 정의와 유형: 4가지 >>> 실현방법: 매체
2. 매체, 문체(구어, 문어, 제3의 문체인 메신저 언어), 사용역, 장르
3. 한국어의 문체 연구, 한국어의 구어문법과 말뭉치언어학
4. R를 활용한 다변량분석 방법론
5. 실증적 연구 및 결과