

발행인: 엄재일

편집인: 이한정

발행처: 한국언어정보학회

발행일: 2014년 10월 8일

URL: <http://society.kisti.re.kr/~ksli/>

110-745

서울 종로구 성균관로 25-2

성균관대학교 문과대학 영어영문학과 이한정 교수 연구실내

전화: (02) 760-0251

e-mail: hanjung@skku.edu

## 1. 월례 논문 발표회 및 강독회

- ▶ 10월 월례 논문 발표회가 10월 18일 숙명여대 진리관 306호에서 열립니다. 이번 월례회는 ‘언어와 통계’를 주제로 한 다음의 특강으로 진행됩니다. 일정 및 초록은 아래를 참조하시기 바랍니다.

최재웅(고려대): 통계언어학: 가벼운 탐색

김미란·최재웅·홍정하(고려대): 군집화 기법을 이용한 한국어 초-중성 간 음소배열 제약 분석 연구

이한정(성균관대): 문법적 변이 연구를 위한 로지스틱 회귀분석

신효필(서울대): 중첩된 출처(nested source)의 SVM 분석을 통한 문장의 주관성 연구

- ▶ 기획의도: 이번 발표회에서는 언어 연구에 통계적 방법론을 적용한 사례를 일부 살펴보게 될 것입니다. 잘 알려져 있듯이 언어통계는 가장 기본적인 어휘빈도 제시 방식에서부터 매우 정교한 수준의 통계기법을 이용한 연구에 이르기까지 다양합니다. 언어 현상 별로 어떤 통계기법이 적합한가라는 의문은 정답이 있다기 보다는 계속 모색해 보아야 할 쟁점이라 생각됩니다. 본 발표회는 이와 관련한 몇 가지 연구 사례를 비교적 평이한 방식으로 보여주고자 하는 의도에서 기획되었습니다.

첫 번째 발표 "통계언어학--가벼운 탐색"은 이번 발표의 도입에 해당되며, 통계언어학적 접근법의 활용과 관련한 몇 가지 사례를 비교적 가벼운 방식으로 제시합니다. 두 번째 발표 "군집화 기법을 이용한 한국어 초-중성간 음소 배열 제약 분석 연구"는 한국어에서 초성과 중성이 결합할 때 나타나는 연쇄 빈도를 살펴보고, 그 분포적 특성을 기술합니다. 한국어대사전의 표제어 전체를 자료로 하고 계층적 군집 분석(Hierarchical clustering)을 이용한 분석입니다. 세 번째 발표 "문법적 변이 현상 연구를 위한 로지스틱 회귀분석"은 구조적 변이형의 분포를 설명하기 위하여 그러한 변이에 영향을 미치는 다양한 요인들을 어떻게 식별해 내야 하는지, 또 왜 통계분석을 시도해야 하는지를 논합니다. 아울러 도구로써 로지스틱 회귀분석의 특성과 적합성 등을 설명합니다. 네 번째 발표 "중첩된 출처(Nested source)의 SVM 분석을 통한 문장의 주관성 연구"는 언어 자원의 감정분석에 대한 논의로 하나의 문장이 드러내는 태도를 밝히는 데는 화자/작가뿐만 아니라 다른 관련 인물의 눈까지 고려하여 주관성 정도를 판단해야 한다는 점을 제기합니다. 이를 위한 코퍼스 구축 방법을 비롯해서, 구축된 코퍼스, 그리고 분석도구인 감독기반 SVM의 사용 등이 소개됩니다.

- ▶ 월례회 후 식사에 이어 강독회가 열립니다. 이번 강독회는 한상훈 선생님(서강대)께서 ‘Aggressive verbal-acts: Impoliteness and disagreement’를 주제로 진행하실 예정입니다. 강독 논문은 본 소식지와 함께 배포해 드리며, 강독회와 월례회 일정 및 내용에 관한 자세한 사항은 아래를 참조하시기 바랍니다.
- ▶ 월례회 또는 강독회에서 다루고자 희망하시는 주제 또는 논문이 있거나, 발표를 희망하시거나 발표자를 추천하고자 하시면 연구이사들에게 알려주시기 바랍니다. 연락처는 아래와 같습니다.

논문 발표회: 유은정 (ejyoo@snu.ac.kr), 전영철 (ycjun@snu.ac.kr), 양정석 (yjsyang@yonsei.ac.kr), 최승권 (choisk@etri.re.kr)

강독회: 김종복 (jongbokkim@gmail.com), 위혜경 (hkwee@dankook.ac.kr),  
정소우 (swchung@sungshin.ac.kr)

## 2. [언어와 정보] 논문 모집

- ▶ [언어와 정보] 편집위원회는 상시 투고 시스템을 운영하고 있습니다. [언어와 정보]에 논문을 게재할 분들은 소정의 논문투고신청서 (홈페이지에서 내려 받음)와 심사용 논문(영문 요약 필수 포함)을 편집위원장인 류병래 교수(ryu@cnu.ac.kr 혹은 kslijournal@gmail.com)에게 보내시면 됩니다. 심사용 논문은 아래야 한글이나 MS Word, LaTeX 등 어떠한 조판 시스템을 사용하셔도 되며, 스타일은 자유입니다. 그 외 논문 투고나 심사에 대한 규정은 학회 홈페이지 (<http://society.kisti.re.kr/~ksli/>)에 자세히 소개되어 있으니 참조하시기 바랍니다.
- ▶ 상시 투고 시스템을 운영하고 있지만, 게재하고자 하는 특정 호수의 [언어와 정보]에 게재를 원하시면 발간일 기준 적어도 60일 이전에 투고를 하셔야 합니다. 발간 일정과 원고 마감일은 다음과 같습니다. 원고 마감일을 넘겨 투고한 논문은 심사에 소요되는 시간적인 이유로 해당 호수의 [언어와 정보]에 게재가 어려울 수도 있습니다.
  - 제18권 2호 (2014년 12월 31일 발간): 2014년 10월 30일 원고 마감
- ▶ [언어와 정보]는 일반 논문(general paper) 외에도 소고(squib), 서평 또는 논평(review), 또 이에 대한 답변(reply) 등 여러 형태의 투고를 받습니다. 논문 투고 시 투고 형태를 알려 주시기 바랍니다.
- ▶ 논문 심사료와 게재료는 다음과 같습니다.
  - 심사료: 60,000원 (3인 심사 기준, 추가 심사시 20,000원씩 추가)
  - 논문 게재료 (3인 심사, 20쪽 분량 기준):
    - 1) 연구비 수혜 논문: 270,000원 (심사료 포함 총 330,000원)
    - 2) 연구비 수혜를 받지 않은 일반논문:
      - (i) 저자 중 한 사람 이상이 전임인 경우 170,000원 (심사료 포함 총 230,000원)
      - (ii) 모든 저자가 전임이 아닌 경우 20,000원 (심사료 포함 총 80,000원)
  - 참고
    - 1) 논문 출판 시 라텍 편집비를 따로 받지 않음.
    - 2) 논문의 분량은 최종 조판 기준 20쪽을 기준으로 하되, 30쪽까지는 별도의 추가 금액을 받지 않으나, 31쪽부터는 쪽당 10,000원을 게재료에 추가.
- ▶ 논문 투고에서 게재까지의 절차는 대략 다음과 같습니다.
  - [1] 투고신청서 작성 (투고자, 서식은 홈페이지) -> [2] 투고 (투고자) -> [3] 심사료 입금 (투고자) -> [4] 심사 의뢰\*3인 (편집위원회) -> [5] 심사 결과 취합 (편집위원회) -> [6] 심사 결과 통보 (편집위원회) -> (게재 확정된 경우 다음 절차 진행, 게재 불가인 경우 [6]에서 종료) -> [7] 수정 보완요구서 (편집위원회) -> [8] 수정보완이행확인서 (투고자) -> [9] 게재료 납부 (투고자) -> [10] 수정 논문 제출 (투고자) -> [11] LaTeX 조판 작업 (편집위원회) -> [12] 교정\*3회 (투고자) -> [13] 출판 (편집위원회/출판사)
- ▶ 논문 투고나 게재에 관하여 기타 문의사항은 메일을 통하여 편집위원장인 류병래 교수 (ryu@cnu.ac.kr 혹은 kslijournal@gmail.com)에게 연락해 주시면 감사하겠습니다.

### 3. 회원 가입 및 회비 납부 안내

- ▶ 우리 학회는 회원 여러분들께서 납부하신 회비로 운영됩니다. 원활한 학회 운영을 위해 연회비를 납부해 주시기 바랍니다. 회비는 월례 발표회장에서 또는 아래의 학회계좌로 납부하실 수 있습니다. (계좌번호: 신한은행 110-382-598367, 예금주: 김미숙)

연회비: 정회원 40,000원, 준회원 20,000원, 기관회원 100,000원  
평생회비: 400,000원

**2003년 12월 이전 평생회비 납부 회원님들께서는 올해부터는 일반회원이 되어 4만원씩 연회비를 납부하시면 됩니다. (아래 명단 참조)**

- ▶ 회원주소록은 연중 관리하고 있습니다. 주소 및 이메일, 전화번호가 바뀌신 분은 변경된 내용을 총무이사 (hanjung@skku.edu)에게 알려주시면 많은 도움이 되겠습니다. 아울러 주변의 교수님들과 신진학자 및 대학원생들이 신입회원으로 많이 가입할 수 있도록 적극적으로 홍보해 주시기를 부탁드립니다.

### 4. 2014년도 회비 납부 현황

- ▶ 평생회원: 강범모(고려대), 곽은주(세종대), 권희상(한밭대), 김용범(광운대), 김은주(이화여대), 김종복(경희대), 김지영(이화여대), 노용균(충남대), 박명관(동국대), 박병수(경희대), 박소영(상명대), 박은주(오사카대), 유은정(서울대), 윤영은(이화여대), 이기용(고려대), 이영현(조선대), 이익환(연세대), 이한정(성균관대), 이현우(인하대), 전영철(서울대), 정대호(한양대), 정재현(충북대), 조세연(강원대), 조영순(전남대), 최재웅(고려대), 최혜원(이화여대), 홍성심(충남대)

[2003년 12월 이전 평생회비 납부회원: 연회비 선납 처리 ~2012년]

김경애(한경대), 김경학(수원대), 김랑혜윤(고려대), 김미경(대덕대), 김성목(SKT), 김용환(양산대), 김창익(호서대), 류병래(충남대), 윤애선(부산대), 윤평현(전남대), 이남근(조선대), 이민행(연세대), 이예식(경북대), 정소우(성신여대), 정찬(동서대), 채명희(조선이공대), 채희락(한국외대), 최규련(연세대), 최기선(KAIST), 최현숙(영남대), 한영균(연세대)

- ▶ 정회원: 박유경(서울대), 송상현(고려대), 염재일(홍익대), 임서현(서울대), 이영주(서울여대), 이정민(서울대), 이주익(경북대), 장석진(서울대), 홍민표(명지대)
- ▶ 준회원: 박은영(충남대), 신재연(서울대), 윤혜림(홍익대), 이혜연(서울대)
- ▶ 명단에 착오나 누락 사항이 있으면 총무이사에게 알려주시기 바랍니다.

### 5. 월례회/강독회 장소 오시는 길

- ▶ 이번 월례회와 강독회는 숙명여대 진리관 306호에서 열립니다.
- ▶ 숙명여대 제1캠퍼스 오시는 길: 4호선 숙대입구 역 10번 출구에서 효창공원 방면으로 왼쪽으로 돌아와 갈월지하도를 건너 5분 정도 도보로 직진하면 오른쪽에 제1캠퍼스 정문이 있습니다 (지하철역에서 정문까지 도보로 약 10-15분). 버스와 자가운전으로 오시는 법은 다음 사이트를 참조하십시오: 버스: <http://www.sookmyung.ac.kr/contents/contents.jsp?cmsCd=CM0071>  
자가운전: <http://www.sookmyung.ac.kr/contents/contents.jsp?cmsCd=CM0072>
- ▶ 주차: 정문 입구에 있는 지하주차장을 이용하시고, 할인권을 (4시간 주차 2천원) 총무이사를 통해 구입하시면 됩니다.



- ▶ 제1캠퍼스 진리관은 아래 지도에서 4번 건물입니다 (10번이 정문).
- ▶ 월례회 후 점심식사는 숙명여대 근처 ‘더함’ 한정식(02-707-3692)에서 있습니다.



**한국언어정보학회 2014학년도 10월 월례회 및 강독회 일정표**  
**월례회 장소: 숙명여대 진리관 306호    일시: 10월 18일 9:30 – 12:00**

시간	발표자	논문제목
09:30-10:10	최재웅 (고려대)	통계언어학: 가벼운 탐색
10:10-10:50	김미란, 최재웅, 홍정하 (고려대)	군집화 기법을 이용한 한국어 초-중성 간 음소배열 제약 분석 연구
Coffee Break (10분)		
11:00-11:40	이한정 (성균관대)	문법적 변이 연구를 위한 로지스틱 회귀분석
11:40-12:20	신효필 (서울대)	중첩된 출처(nested source)의 SVM 분석을 통한 문장의 주관성 연구

**강독회 장소: 숙명여대 진리관 306호    일시: 10월 18일 14:00 – 16:00**  
**주제: Aggressive verbal-acts: Impoliteness and disagreement**

발표자	한상훈 (서강대)
논문	Culpeper, J. 1996. Towards an anatomy of impoliteness. <i>Journal of Pragmatics</i> 25, 349-367. Sifianou, M. 2012. Disagreements, face and politeness. <i>Journal of Pragmatics</i> 44, 1554-1564.

**2014년 10월 월례회 발표논문 초록**

통계언어학: 가벼운 탐색

최재웅 (고려대)

본 발표는 언어학 탐구에 통계적 접근법을 활용한 몇 가지 사례를 '호기심', '끌림', '매력'으로 나누어 비교적 가벼운 형식으로 제시해 본다. 다루는 소재는 '어휘-저자관별', '어휘-주제분류', '어휘-복합 변인', '어휘-구조' 등이다. 본 발표는 이번 이어지는 발표의 도입부 역할도 하는 것으로, 통계적 관점의 특징과 통계 활용도에 대하여 몇 가지 최근 연구 사례를 중심으로 개략적으로 살펴보는 것이 목적이다. 또한 실제 작업 절차에 대한 구체적인 사례도 일부 제시하여 언어에 대한 통계적 연구 방법 및 절차에 대한 일단을 엿보는 기회로 삼는다. 아울러 통계적 분석 결과를 다양한 시각적 도표로 정리해 보여줄 수 있다는 점도 일부 예시로 보여준다.

## 군집화 기법을 이용한 한국어 초-중성 간 음소배열 제약 분석 연구

김미란, 최재용, 홍정하 (고려대)

본 연구에서는 한국어에서 초성과 중성이 결합할 때 나타나는 연쇄 빈도를 살펴보고, 그 분포적 특성을 기술한다. 자료는 한국어대사전에 등록된 510,032개 단어에서 총 1,761,214 음절을 분석하였고, 자료 분석은 계층적 군집 분석(Hierarchical clustering)을 이용하였다. 군집분석은 개체/범주 사이의 관계를 거리(distance) 계산을 통해 군집으로 분류하는 통계기법이며, 이러한 통계적 방법론은 통계적 관찰 또는 기술을 통해 관찰 자료에 내재하는 특성을 도출하여 새로운 가설을 생성해 내는 접근 방식이다(Gries 2013). 즉, 관찰 자료에 나타나는 특성을 미리 전제된 기준이나 주관적 해석에 의존하여 판단하기보다는 관찰적 관점에서 자료의 객관적 특성을 기술하고 이를 통해 미지의 가설을 도출할 수 있도록 하는 접근 방법이다. 계층적 군집 분석은 역사비교언어학, 이론언어학, 코퍼스언어학 등 다양한 분야에서 활용되어 그 유효성이 입증되어 왔다 (Dyen et al. 1992, Gries 2008, Hilpert and Gries 2009, 최운호 2010, Gries and Otani 2010). 군집 분석 결과는 통계 패키지 R을 이용하였다 (R Core Team 2014: 버전 3.0.3).

본 연구는 한국어 초성과 중성 결합에 나타나는 통계적 분포 정보이며, 이를 통해 발견되는 초성-중성 결합의 분포적 특성은 세 가지로 정리해 볼 수 있다. 첫째, 초성이 중성과 결합하는 빈도적 분포를 중심으로 한국어 모음 군집을 살펴보면 단모음 계열과 이중 모음 계열이 서로 다른 결합적 특성을 보인다. 즉, 비교적 다양한 초성과 결합하는 단모음 계열과 상대적으로 초성과의 결합적 제약이 많은 이중 모음이 서로 다른 군집으로 분석된다. 둘째, 한국어 모음 체계 분류에서 단모음/이중모음 구분에 아직도 논의가 되고 있는 '니/기'의 경우, 초성과의 결합 분포성이라는 측면에서 이 두 모음은 이중모음과 상대적으로 높은 유사성을 보이는 것으로 나타난다. 마지막으로 현대 한국어 발음형에서 중화된 '히/기' 모음의 경우 모음 '히'가 적어도 초성+모음의 결합 분포라는 측면에서 '기'보다 상대적으로 다양한 자음군과 결합하고 있음을 확인하였다. 이는 문어체에서는 '히/기'가 어휘적 변별성을 가지고 있고, 이러한 분포적 특성이 실제 새로운 어휘 생성이나 실제 발음형 변화에 어떠한 영향을 미치게 되는지 향후 연구가 흥미롭다고 할 수 있다.

## 문법적 변이 연구를 위한 로지스틱 회귀분석

이한정 (성균관대)

자연언어에는 두 가지 이상의 형태 또는 구조가 동일한 언어 환경에서 명제적 의미 변화 없이 교체하여 쓰이는 변이 현상이 빈번하다. 변이 현상 연구에 있어서의 문제는 변이형(variant form) 또는 교체하는 구조(alternating structure)의 선택에 영향을 미치는 요인이 다수 있으며, 이 요인들 간에는 빈번히 상관관계가 존재하는데, 언어적 직관이나 코퍼스 자료 분석만으로는 어느 요인이 독립적 영향을 미치는지 그리고 어느 요인이 더 중요하고 덜 중요한지 알기 어렵다는 점이다. 따라서 변이의 패턴을 정확히 기술하고 설명하기 위해서는 변이형의 선택에 영향을 미치는 다수의 요인들의 유의성, 독립적 영향력 여부와 상대적 중요성을 검증할 수 있는 통계적 분석이 요구된다. 본 발표에서는 다른 학문 분야에서 이분형 종속변수의 사건 발생 확률 예측을 위해 널리 쓰이는 로지스틱 회귀분석의 개념, 실제

분석 과정과 결과 해석법을 수량적 변이에 관한 최근 연구와 함께 소개하여 로지스틱 회귀분석이 어떻게 변이 현상 분석을 위해 유용하게 적용될 수 있는지 보이고자 한다.

### 중첩된 출처(nested source)의 SVM 분석을 통한 문장의 주관성 연구

신효필 (서울대)

현재 컴퓨터언어학 분야에서는 의견분석(opinion analysis), 감정분석 (sentiment analysis), 그리고 주관성(subjectivity)에 관한 연구가 활발히 이루어지고 있다. 그간의 연구들은 긍정어, 부정어와 같은 어휘들의 극성 연산에 초점을 맞추어 왔다. 문장의 주관성은 실제 화자의 태도(speaker's attitude)와 같은 양상 정보, 담화와 맥락과 같은 화용론적 정보에 더 영향을 받을 수 있다. 이 분야에 화용적 정보를 포함시키려는 시도로 화자의 태도가 반영되는 한 양상인 중첩된 출처(nested source)를 주관성 분석에 도입한다. 발화 내에서 어떤 표현이 화자 자신이 그 표현의 출처(source)가 되느냐 아니면 다른 사람의 눈을 통한(nested source) 것이냐에 따라 주관성의 정도가 달라진다고 가정하고 이를 실제 자료에서 확인해보도록 한다. 영어의 의견분석의 대표적인 코퍼스인 MPQA(Multi Perspective Question Answering )와 한국어 감정분석 코퍼스 KOSAC(Korean Sentiment Analysis Corpus)에는 개별 감정표현에 중첩된 출처에 관한 정보가 주석되어 있다. 이 자료의 기본적인 통계를 baseline 으로 하여 현재 기계학습에서 지도학습(supervised learning) 방법으로 많이 사용되고 있는 SVM(Support Vector Machine)을 사용하여 중첩된 출처가 문장의 주관성에 어떻게 기여하는지를 실험하는 과정을 간략히 소개하도록 한다.