

자연발화상에 나타난 단음절 단일간투사의 길이특성 분석¹⁾

김기호*, 송민석**, 장태엽***, 정현성***

*고려대, **관동대, ***한국외대

Analysis of the Durational Characteristics of Monosyllabic Interjections in Natural Spoken Language

Kee-Ho Kim*, Minsuck Song**, Tae-yeoub Jang***, Hyunsong Chung***

*Korea Univ., **Kwangdong Univ., ***Hankuk University of Foreign Studies

요약문

본 논문은 자연발화상에 포함되어, 음성언어 인식에 장애를 초래하는 간투사(군소리)와 음성적 특성 중, 가장 뚜렷이 구별되는 길이특성을 분석하여 음성언어 인식에 도움을 주는 것을 목적으로 한다. 이 연구에서는 간투사의 대부분을 차지하는 단음절 단일 간투사에 한정하여, 실제 대화의 녹음자료에서 나타나는 간투사의 빈도수와, 그 길이특성을 신분별, 성별, 간투사 유형별로 분석하였다. 또 간투사를 위치에 따라, 음운구조 간투사, 음운구말 간투사로 나누고, 그 길이를 음절 평균, 음운구초 음절이나 음운구말 음절의 길이와 비교하여 간투사의 증가율을 측정하였다. 분석결과, 가장 높은 빈도수를 보이는 단음절 단일 간투사는 '어'이며, 간투사 길이 증가율은, 음절평균에 대해서는 '그'가, 음운구초 평균에 대해서는 '응'이 가장 큰 증가율을 나타낸다. 전체적으로 음운구초 음절 길이에 대한 간투사 길이 증가율이 음절평균 길이에 대한 간투사 길이 증가율보다 더 크게 나타났다. 이러한 분석결과를 통해 하위레벨에서 제거할 수 있는 간투사와, 통사적 또는 의미론적 분석이 필요한 상위레벨에서 처리해야 할 간투사를 구별할 수 있다. 이와 같은 길이특성 외에 간투사에 대한 다양한 음성적 특성과, 다음절 단일 간투사와, 이중 간투사에 대한 연구가 진척된다면, 음성언어 인식에 장애가 되는 간투사의 효과적 배제가 가능할 것으로 보인다.

1. 서론

이제까지 개발된 음성언어 인식 시스템에서는 '어', '아', '저' 등의 간투사 개념으로 인해 음성언어 인식에 많은 오류가 나타나거나, 인식불능 상태가 되는 문제가 있다. 자연발화상에는 전달하고자 하는 뜻과는 상관없는 아래와 같은 간투사(군소리)들이 많이 포함되어 있으며, 이것은 음성언어 인식에 장애를 초래한다.

- (1) 어, 떠블름으로요?
아, 팔릴 한 십, 십사일 정도요.
저, 연기가 되었거든요.

이러한 간투사들은 일반 음절과 대비되는 여러가지 음성적 특성을 띠 수 있다. 즉 길이, pitch변화, energy변화, 비음화의 정도, formant구조 등의 음향적 특징을 중심으로 간투사의 음성적 특성을 분석할 수 있다.

만일 간투사들이 갖는 음성적 특성을 밝힐 수 있다면 하위 레벨에서 바로 간투사를 제거하고 상위 레벨로 올

라갈 수 있으므로 인식에 많은 도움을 줄 수 있으므로 바로 이런 간투사의 특성을 찾는 것이 본 연구의 목적이다.

이러한 특성 중, 현재로서는 간투사를 일반음절들과 구별해 낼 수 있는 가장 뚜렷한 단서가 길이 특성이라 할 수 있기 때문에, 본 논문에서는 이러한 길이 특성을 중심으로 간투사와 음절의 상관성을 알아보기 위하여 음절 평균길이, 음운구초 음절 길이, 음운구말 음절 길이에 대한 간투사의 길이 증가율을 신분별, 성별, 간투사의 종류별로 측정하였다.

2. 간투사의 정의와 분류

2.1. 간투사의 정의

본 연구에서는 기존 국어문법의 개념을 바탕으로, 감탄사의 하위부류를 이으면서 문장에서 아무런 문법적 기능을 담당하지 못하고 의미적으로도 독립성을 가지지 못하는 '임버릇 및 더듬거림'을 연구 대상으로 삼고 특별히 이들을 '간투사(Interjection)'라고 정의했다. 기존의 언어학 연구에서는 간투사를 독립적인 주요 연구과제로 삼은 적이 없었고, 단지 감탄사의 한 부류로 연구하여 왔다. 국어문법에서 감탄사에 관해 김민수(1985)에서는 의미적 기준에 따라 감정을 나타내는 감탄사와 의지를 나타내는 감탄사를 위주로 분류하고 있으며, 남기심, 고영근(1985)에서는 감탄사를 감정감탄사, 의지감탄사, 임버릇 및 더듬거림의 세 가지로 나누어서 상세하게 분류하였다. 신지연(1989)에서는 감탄사를 진정한 감탄의 의미물 지니는 것과 그 외의 것으로 나누었는데, 감탄사가 문장과 문법적 관계를 맺지 않으며, 나타날 수 있는 위치가 자유롭다는 점에 주목하여 영어의 간투사(Intenection)에 대응하는 간투사로 부를 것을 제안했다.

2.2. 간투사의 종류와 빈도수

간투사는 단음절 단일 간투사와 다음절 단일 간투사, 이중 간투사로 나누어 볼 수 있다. 단음절 단일 간투사는 '어', '아', '저' 등의 한음절로 이루어진 간투사이며, 다음절

1. 본 연구는 '한국통신 연구개발단'의 1994년도 '장기 기초 연구과제' 연구비의 지원을 받아 이루어졌다.

자연발화상에 나타난 단음절 단일간투사의 길이특성 분석

단일 간투사는 '워아, '거기' 등 2음절 이상으로 된 독립 간투사를 가리킨다. 또한 이중 간투사가 나타나기도 하는데, 이것은 독립된 두개의 간투사가 한 덩어리를 이루며 연속적으로 나타나는 것이다. 본 논문에서는 이러한 간투사중 단음절 단일 간투사를 연구대상으로 삼았다.

본 연구에서는 73개의 실제 대화상황을 통해 590개의 간투사를 발견할 수 있었다. 남자직원 1명, 여자직원 2명, 그리고 73개의 대화중 간투사를 사용한 남자손님 46명, 여자손님 13명이 이 분석의 대상이 되었다. 이들은 총 13종류의 단음절 단일간투사를 사용하였다. 종류별로 빈도수를 살펴보면, 아래 표와 같이 '어'는 362회 사용되어 61.4%의 가장 높은 빈도수를 보였다. '아'는 92회로 15.6%, '저'는 41회로 6.9%, '그'는 33회로 5.6%, '응'은 19회로 3.2%, '뭐'는 14회로 2.4%, '거'는 8회로 1.4%의 빈도수를 보였으며, 기타 '영', '에', '고', '오', '음', '한' 등의 단음절 단일 간투사가 사용되었지만 낮은 빈도수를 나타내었다.

표 1. 각각의 간투사의 사용빈도

종류	어	아	저	그	응	워	거	기타	합계
빈도수	362	92	41	33	19	14	8	21	590
비율(%)	61.4	15.6	6.9	5.6	3.2	2.4	1.4	3.5	100

3. 실험환경 및 분석방법

이 분석을 위해서 수집된 실제대화의 녹음자료로 글로 옮겨 전사 하였다. 그 전사자료를 기준으로 실제대화를 듣고 각각의 간투사를 음운구초 간투사, 음운구말 간투사로 구분하였다. 음운구란 실제 발화된 문장에서 나타나는 pause와 pause 사이의 음절 연속체로서, 쉬지 않고 한숨에 말하는 단위이다. 간투사를 음운구를 단위로 분석한 이유는 하나의 간투사가 나타난 이후 다음 간투사가 나타날 수 있는 최소한의 단위가 음운구이기 때문이다.

음운구초 간투사는 그 간투사 바로 뒤에 나오는 음운구의 평균 음절 길이와 그 음운구의 첫음절 길이를 측정하여 그 간투사의 길이와 각각 비교하였고, 음운구말 간투사는 간투사 바로 앞에 나오는 음운구의 평균 음절 길이와 그 음운구의 끝음절 길이를 측정하여 각각 간투사의 길이와 비교하였다.

예를 들어, "어, 잠시만요"라는 예문에서 음운 구초에 나타나는 간투사 '어'의 길이를 측정하고, 그 뒤의 음운구 "잠시만요"의 전체길이를 측정한 다음, 그 길이를 음절수로 나눈 음절 평균길이에 대한 간투사의 증가율을 측정하고, 음운구초 음절 '잠'에 대한 간투사 길이 증가율을 측정하였다.

다음과 같은 음운구말 간투사의 경우, 음운구말에 나타나는 간투사 '뭐'의 길이를 측정한 다음, 그 간투사 앞에 나타나는 음운구 '없으면'의 전체길이를 음절수로 나눈 음절 평균길이에 대한 간투사 길이 증가율과, 음운구말 음절 '면'에 대한 간투사의 길이 증가율을 측정하였다.

- (2) 음운구말에 나타나는 '뭐'의 예: 없으면 뭐 (그렇게라도 해야지)

이 외에 바로 앞에 문장으로 끝나는 문말 간투사도 분석의 대상이 될 수 있지만, 본 연구에서는 관찰할 수 있는 자료가 부족하였다.

분석장비로는 길이측정을 위해 KAY사에서 개발한 CSL(Computerized Speech Lab: Kay model 4300B)을, 통계분석을 위해서는 Microsoft Excel 4.0을 사용하였다.

4. 단음절 단일 간투사의 종류별 길이 특성

음절 평균 길이, 음운구초 길이, 음운구말 길이 각각에 대한 간투사 길이의 증감율을 신분별, 성별, 간투사 종류별로 측정한 분석결과는 표 2와 같다.

표 2. 간투사의 길이 증가율

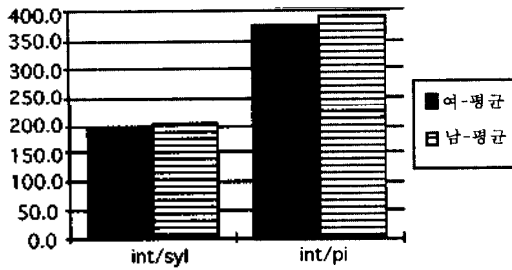
		간투사/음절	간투사/구초	간투사/구말
신분별	손님	201.8	398.2	156.7
	직원	199.2	371.6	
성별	남자	205.4	392.5	156.7
	여자	197.5	375.9	
간투사종류별	아	164.2	376.7	
	어	202.1	382.3	
	그	275.4	386.5	
	저	188.9	346.8	
	거	131.8	199.2	
	응	244.1	507.5	
	워	112.8	167.6	156.7
	기타	272.3	506.3	
전체		200.2	381.5	156.7

기타항목: 영, 에, 고, 오, 음, 한

4.1. 성별 분석

성별로는 아래 그림에서 나타난 바와 같이 여자와 남자의 간투사 길이 증가율이 비슷하다. 음절 평균길이에 대한 간투사 길이 증가율이 남자의 경우 205.4%, 여자는 197.5%, 음운구초 길이에 대한 간투사 길이 증가율은 남자 392.5%, 여자 375.9%, 음운구말 길이에 대한 간투사 길이 증가율은 남자는 156.7%인 반면, 여자는 관찰할 수 있는 자료가 없었다.

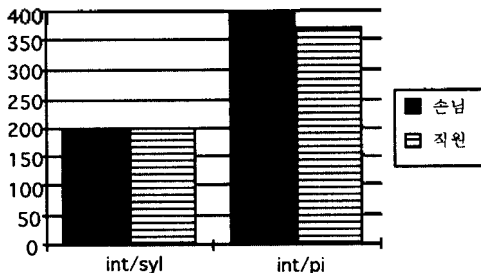
그림 1. 성별 길이 증가율 비교



4.2. 신분별 분석

이 분석에서 신분별로는 직원들과 손님들의 간투사 길이 증가율에 큰 차이가 없다. 아래 표와 같이 음절 평균 길이에 대한 간투사 길이 증가율이 손님의 경우는 201.8%인 반면, 직원의 경우에는 199.2%로 나타나고, 음운구조 길이에 대한 간투사 길이 증가율은 손님 398.2%, 직원 371.6%, 음운구말 길이에 대한 간투사 길이 증가율은 손님은 156.7%인 반면, 직원은 관찰할 수 있는 자료가 없었다.

그림 2. 신분별 길이 증가율 비교



4.3. 간투사 종류별 분석

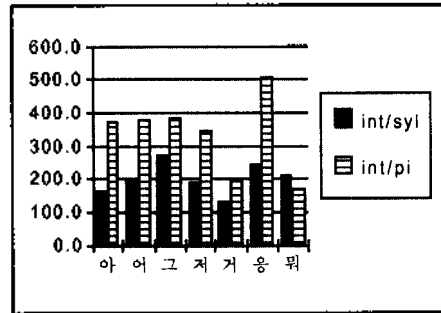
아래 표에서 나타난 바와 같이 간투사 종류별로는 사용빈도가 33회로 5.6%를 보이는 '그'와 19회로 3.2%의 사용빈도를 보이는 '옹'이 평균적으로 가장 큰 간투사 길이 증가율을 보이고, 14회로 사용빈도가 2.4%인 '뽀'는 가장 작은 증가율을 보인다. '뽀'를 제외한 모든 간투사가 음운구조에 나타나는 반면, '뽀'는 음운구조에 10회, 음운구말에 4회 나타났다.

'그'의 경우 음절 평균길이에 대한 간투사 길이 증가율이 275.4%로 다른 간투사의 길이 증가율에 비해 가장 큰 증가율을 보이고 음운구조 길이에 대한 간투사 평균 길이 증가율은 386.5%로 '옹'에 이어 두번째로 큰 증가율을 보인다.

'옹'의 경우 음운구조 길이에 대한 간투사 길이 증가율은 507.5%로 다른 간투사의 길이 증가율에 비해 월등히 큰 증가율을 보이고, 음절 평균길이에 대한 간투사 길이 증

가율이 244.1%로 '그'에 이어 두번째로 큰 길이 증가율을 나타낸다.

그림 3. 간투사 유형별 길이 증가율 비교



'뽀'의 경우 음절 평균길이에, 음운구조 길이에 대한 간투사 길이 증가율이 각각 112.8%, 167.6%으로 다른 간투사들의 증가율에 비해 가장 낮은 증가율을 보인다. 또 '뽀'는 다른 간투사들과는 달리 음운구말에도 4회 나타났는데, 이때 음운구말 길이에 대한 간투사 길이 증가율은 156.7%이었다.

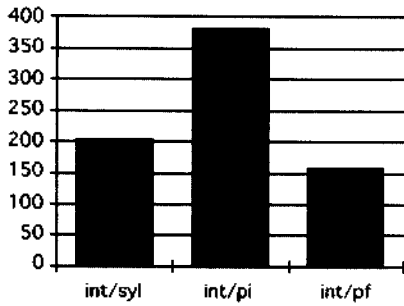
전체적으로는 음절 평균길이에 대한 간투사 길이 증가율은 '그' > '옹' > '어' > '저' > '아' > '거' > '뽀'의 순으로, 음운구조 길이에 대한 간투사 길이 증가율은 '옹' > '그' > '어' > '아' > '저' > '거'의 순으로 나타났다. 기타 6가지 단 음절 단일 간투사는 분석이 가능한 만큼의 자료가 충분하지 못하기 때문에 분석대상에서 제외하였다.

4.4. 전체평균의 분석

전체적으로 음절 평균길이에 대한 간투사 평균길이 증가율은 200.2%, 음운구조 길이에 대한 간투사 길이 증가율은 381.5%, 음운구말 길이에 대한 음운구말 간투사 '뽀'의 길이 증가율은 156.7%이었다. 아래 표에서 알 수 있듯이 음절 평균 길이에 대한 간투사 길이 증가율보다 음운구조 길이에 대한 간투사 길이 증가율이 더 큰 것은 음운구조의 경우 간투사 뒤에 새로운 음운구조가 이어지면서 음운구조 음절의 길이가 평소보다 짧아지기 때문에 결국 그것에 대한 간투사 길이 증가율이 커지는 것으로 이해할 수 있다.

자연발화상에 나타난 단음절 단일간부사의 길이특성 분석

그림 4. 전체 평균 길이 증가율 비교



5. 결론

이상 살펴본 바와 같이 간부사의 길이증가율은 신분과 성별로 큰 차이를 보이지 않으며, 간부사 종류별로는 음절 평균길이에 비해서는 '그'가, 음운구조 음절의 평균길이에 비해서는 '옹'이 가장 큰 증가율을 나타내고, '뫼'가 가장 낮은 증가율을 나타낸다. 전체적으로는 음절 평균길이에 대한 간부사 길이 증가율에 비해서 음운구조 길이에 대한 간부사 길이 증가율이 더 크다. 현재까지의 연구 결과로서는 간부사의 길이가 200%(2배)이상의 증가율을 보이는 것들은 하위레벨인 음성언어 레벨에서 구분하여 처리할 수 있을 것이지만, 길이 증가율이 그다지 뚜렷하지 않은 것들, 특히 150%이하의 증가율을 가진 것들은 간부사가 아닌 정상음절과의 길이 구분이 어려울 것이므로 상위레벨인 통사론적 또는 의미론적 분석에 의해 처리되어야 할 것으로 보인다. 한편, 이상과 같은 길이 특성 이외에, pitch 변화, energy 변화, 비음화의 정도, formant 구조 등의 음향적 특징에 대한 간부사의 음성적 특성과, 단음절 단일 간부사 외에 다음절 단일 간부사나, 이중 간부사에 대한 연구가 앞으로 더 구체적으로 이루어 진다면, 음성언어 인식에 장애를 초래하는 간부사를 효과적으로 추출, 제거할 수 있을 것으로 보인다.

참고문헌

김기호. 1990. "Revisiting Distinctive Feature Approach in Speech Recognition." 어학연구 27권 2호, 서울대학교 어학연구소.
 김기호. 1993. "연속 음성인식에 있어서 음운론의 역할을 재고함." 「음성 음운 형태론 연구」 음운론 연구회 pp. 49-63. 한국문화사.
 김민수. 1985. 「신국어학, (개정 증판), 서울: 일조각.
 남기삼, 고영근. 1987. 「표준국어문법론」 (4쇄) 서울: 탑출판사
 신지연. 1989. "간부사의 화용론적 특성," 「주시경학보」 3, 서울: 탑출판사.
 이익섭, 임홍빈. 1985. 「국어문법론」 서울: 학연사.

유길준. 1909. 「대한문전」, 역대한국문법대계 1-6. 서울: 탑출판사.
 이익섭, 임홍빈. 1985. 「국어문법론」 서울: 학연사.
 유길준. 1909. 「대한문전」 역대한국문법대계 1-6. 서울: 탑출판사.
 최현배. 1971. 「우리말본」 (네번째 고침), 서울: 정음문화사.
 허웅. 1989. 「국어학」 (증판), 서울: 샘문화사.