

한국어 구어의 억양단위와 발화 경계의 상관에 관한 고찰

서상규(연세대학교 국어국문학과 교수)
백해파(연세대학교 대학원 박사과정)

1. 서론

문어에서 문장에 대한 정의는 논란의 여지가 있지만 문장의 경계는 대체적으로 확립되어 있는 데 비해, 문어의 “문장”에 대응하는 구어에서의 통사적 단위는 아직 명시적인 기준이 확립되지 않고 있다. 발표자가 “21세기 세종계획”의 일환으로 대량으로 구축해 온 한국어 구어 (전사) 말뭉치에서는 발화에서 ‘억양이 바뀌기 이전 까지를 하나의 연속된 발화 단위로’ 보아, 그 기본단위를 설정한 바 있다. 이때의 기본단위는 이른바 ‘억양 단위’를 기준으로 하고 있다고 할 수 있다. 그런데 이 기준은 문법 연구에서 통사적 구조를 규명하거나 특정 어휘의 용법을 연구하는 데에는 적합하지 못한 여러 가지 문제점이 드러나고 있으며, 최근 이에 대한 연구들이 발표되고 있다

한국어 문법에서 문장은 문어에서의 문장으로서 『우리말본』(1937)에서의 정의를 대체적으로 따르고 있다. 『우리말본』에서는 문장(월)에 대한 정의를 아래와 같이 하고 있다.

월갈에서 월(文)이란 것은 한 통일된 말로 들어낸 것이니: 뜻으로나 꼴(形式)로 나 온전히 다른 것과 따로선(獨立한) 것이니라.

즉, 월이 되기 위해서는 낱말이 모여서 그 속에 통일된 생각이 들어있어야 된다는 것과 다만 하나의 끝남을 소용한 따로섬이 있어야 한다는 것이다. 문장의 이런 정의에서 전자는 “의미적 완결성”으로 후자는 “형식적 자립성”으로 정의할 수 있는 바 문어에서의 문장은 서술어를 중심으로 “의미적 완결성”을 이루는 단어들로 이루어지며 종결어미로 종결된다.

문어에서 문장은 마침표나, 느낌표, 물음표와 같은 문장부호로 그 경계를 확인하기 어렵지 않지만 구어에서는 문장부호나 종결어미와 같은 형식적인 특징에 의하여 문장의 경계를 확인하기 어렵다. 전영옥(1993)의 연구결과는 이런 사실을 입증할 수 있는데 이 논문에 의하면 한국어에서의 발화는 비어말어미의 실현이 65.5%로서 어말어미(34.5%)보다 더 많이 실현된다고 하였다. 따라서 이 연구에서 사용한 “21세기 세종계획”의 구어 말뭉치는 “억양 단위”를 기본 단위로 하여 구축되어 있다.

그러나 이런 “억양 단위”는 “의미적 완결성”을 이루고 있다고 할 수 없다. 비록 서은아, 남길임, 서상규(2004)에서 제기한 “문장을 구성하는 여러 성분 가운데 한 성분으로 문장을 끝맺지만 의사소통에는 전혀 지장을 주지 않는 구어의 문장”인 “조각문”은 독립적으로 볼 때 “의미적 완결성”이 없다고 할 수 없지만 많은 발화 상황에서는 “의미적 완결성”이 있다고 볼 수 없는 발화가 많다. 예를 들어:

1. <P2_M30s>: 근데 이호선으로 딱 넘어가면은,
2. <P2_M30s>: 운이 좋아야지 앓아 갈 수 있지.

위 발화에서 발화1과 발화2는 “P2”라는 한 사람이 한 말로서 “약한 상승이나 하락 억양,’”에 의하여 두 개의 발화로 나뉘어져 있다. 그러나 “P2”는 두 가지 의미를 담아 발화한 것으로 보이지 않는다. 한 가지 의미를 표현함에 있어서 하나의 “억양”에 의하여 두 개의 발화로 나뉘어진 것이다. 이런 발화를 하나의 단위로 묶어 놓지 않으면 문법연구에 어려운 점이 많다.

따라서 이 연구는 이런 “억양단위”들이 “의미적 완결성”에 의하여 하나의 단위로 묶이는 데 어떤 특징이 있는지를 살펴보고 이렇게 묶인 “문장”에서 “억양단위”들 사이의 관계를 고찰하여 “억양단위”와 “문장”과의 상관관계에 대하여 살펴보고자 한다.

제2장에서는 대화(사적/공적)와 독백(사적/공적)에서 “억양 단위”로 나뉘어진 발화들이 한 단위로 묶일 때의 특징에 대하여 살펴보고 제3장에서는 이렇게 묶여진 “문장”에서 “억양단위”특징에 대하여 살펴보기로 한다.

2. 말뭉치의 재주석

이 연구에서는 “21세기 세종계획”에서 배포한 구어말뭉치에서 대화(사적/공적)와 독백(사적/공적)의 4개의 샘플에 대하여 분석하였다. 그 말뭉치의 구성을 보이면 아래와 같다.

	파일명	어절수	발화수	"문장"수
사적대화	5CT_0013	3,730	1,069	876
공적대화	4CT_0001	5,063	1,322	875
사적독백	6CT_0010	2,435	493	333
공적독백	5CT_0046	3,783	529	197
총계		15,011	3,413	2,281

<표1>: 말뭉치의 구성

2.1. 말뭉치의 재주석 과정

“21세기 세종계획”의 구어말뭉치에서 발화는 “차례말기”와 “억양단위”에 따라 나누고 있다. “차례말기”는 발화자의 발화순서에 따라 “<u>”태그로 묶어 주고 있고, “억양단위”는 “<s>”태그로 묶어 주고 있다.

```
<u who=P2><s n=00057>그 아저씨가 딱 내리고 나면,</s>
<s n=00058>자리가 많이 비좁아,</s></u>
<u who=P1><s n=00059>어.</s></u>
<u who=P2><s n=00060>거기 딱 앉아서 편안하게 가는 거야,</s></u>
<u who=P1><s n=00061>그렇구나,</s></u>
<u who=P2><s n=00062>나만의 로하우야.</s>
<s n=00063>근데 어찌다가 실패하는 경우가 있어.</s>
<s n=00064>어찌다 실패하면 학생인데 안 내리고 끝까지 가는 거야 이게,</s></u>
<u who=P1><s n=00065><vocal desc='웃음'></s></u>
```

<표2>: 말뭉치의 예

이런 말뭉치를 Python 스크립트 언어를 사용하여 프로그램을 만들어 재주석에 용이한 형태로 변환하여 엑셀에서 작업하였다.

	A	B	C	D	E	F
1	일련번호	발화자	발화번호	문장번호	발화	발화수
119	118	P1_F30s	118	285	그때는 그까 [[[않어]]] [[[가는]]] 일은 별로 없어.	
120	119	P2_M30s	119	288	음.	
121	120	P1_F30s	120	291	별로 애들이 거기도 뭐~ 애들이 아한 거기쯤에 [[[내리더라.]]]	
122	121	P1_F30s	121	292	학생들이 그쪽 학교 있지?	
123	122	P1_F30s	122	293	너두 거기 [[[나왔잖아?]]]	
124	123	P2_M30s	123	294	음.	
125	124	P2_M30s	124	294	이들이 뭐지?	
126	125	P1_F30s	125	297	환영.	
127	126	P2_M30s	126	300	어이구 [[[아네.]]]	
128	127	P1_F30s	127	303	<vocal desc='웃음'>	
129	128	P1_F30s	128	303	그래서.	
130	129	P2_M30s	129	306	음.	
131	130	P1_F30s	130	303	거기 애들 [[[내리고]]] 나면 자리가 [[[생기는데.]]]	
132	131	P2_M30s	131	312	음.	
133	132	P1_F30s	132	303	그래도 막 이렇게 굳이 막 [[[않을려고]]] 노력을 안 [[[하니까]]] 별로 [[[않는]]] 일은 없는 것 같애.	
134	133	P1_F30s	133	315	오늘두 막 아침에 거기 막 [[[서]]] 있는데.	
135	134	P2_M30s	134	318	음.	
136	135	P1_F30s	135	315	사람이 딱 [[[일어났는데.]]]	
137	136	P2_M30s	136	324	음.	
138	137	P1_F30s	137	315	나 그 옆에 아가씨가 한 명 서 있었고.	
139	138	P1_F30s	138	315	내가 서 있었고.	
140	139	P1_F30s	139	315	내 오른쪽에 아줌마가 서 있었거든?	
141	140	P2_M30s	140	330	음 역시 아줌마.	
142	141	P1_F30s	141	333	아니 이제~ 아줌마가 이제~ 좀<phon>좀</phon> 약간 머뭇거리는 듯해서 그냥 가만히 서 있었어.	
143	142	P2_M30s	142	336	음.	
144	143	P1_F30s	143	339	이렇게 [[비켜]] 줬어.	
145	144	P2_M30s	144	342	음 [[[그랬더니]]] 누가 [[[앞다]]]	
146	145	P1_F30s	145	345	아줌마.	
147	146	P2_M30s	146	348	음.	
148	147	P1_F30s	147	351	근데 아줌마가 이렇게 막 노력한 아줌마는 아니고.	
149	148	P2_M30s	148	354	음.	
150	149	P1_F30s	149	351	또 이렇게 그 여자 그 아가씨도 나두 [[[않을려구]]] 막 [[[그러지]]] 않구.	
151	150	P2_M30s	150	360	음.	
152	151	P1_F30s	151	351	나두 아줌마 앞에 서 있는데 [[[않을려니까]]] 좀 그래서 [[비켜]] 줬어.	
153	152	P2_M30s	152	366	음.	
154	153	P1_F30s	153	369	그래서 나는 뭐~ 맨날 심부스 [[[타고]]] [[[가니까]]] 별로 없는데 근데 [[[음]]] 때가 힘들어.	
155	154	P1_F30s	154	370	거기서 다시 학교로 [[가잖아.]]]	
156	155	P2_M30s	155	372	음.	
157	156	P1_F30s	156	375	근데 진짜 좀거든?	
158	157	P1_F30s	157	375	[[[말지]]]	

<그림>: 말뭉치의 재주석 과정

그림1에서 보면 B열은 발화자의 정보이고 C열은 “역양 단위”에 의한 발화의 일련번호이고 D열은 “문장”의 일련번호인데 작업은 의미적으로 연결성을 가지는 발화들은 같은 일련번호를 부여하는 방식으로 진행하였다(그림1의 네모 상자를 참조). 이렇게 만들어진 말뭉치는 또 프로그램을 만들어서 말뭉치의 “<s>” 태그에 “GR” 속성을 추가하여 일련번호를 부여하는 것으로 재주석 작업을 진행하였다. 그 결과를 보이면 아래의 표3과 같다.

<u who="P1_F30s"><s n="00135" GR="00108">사람이 딱 일어났는데,</s></u>
<u who="P2_M30s"><s n="00136" GR="00110">음.</s></u>
<u who="P1_F30s"><s n="00137" GR="00108">나 그 옆에 아가씨가 한 명 서 있었고,</s>
<s n="00138" GR="00108">내가 서 있었고,</s>
<s n="00139" GR="00108">내 오른쪽에 아줌마가 서 있었거든?</s></u>

<표3>: 말뭉치 재주석 결과

이런 처리과정을 거친 말뭉치는 “GR” 속성의 일련번호를 통하여 특정 어휘가 쓰인 “문장”을 추출하여 문법연구에 사용할 수 있다.

2.2. “역양단위”가 “문장”으로 묶이는 경우

이 절에서는 위의 2.1에서 재주석 과정에서 발화가 “문장”으로 묶이는 데 나타나는 현상을 살펴보기로 한다. 발화가 “문장”으로 묶는 것은 크게 발화 중에 다른 발화자가 끼어들어 “발화”가 나누어진 경우와 발화 중 역양이 바뀌면서 “발화”가 나누어진 경우가 있다.

먼저 발화 중에 다른 발화자가 끼어든 경우를 살펴보면, 발화 중에 다른 발화자가 웃거나, 간투사를 발하는 경우이다.

예1:

- 1.1 <P1_F30s>: 그래서,
- 2.2 <P2_M30s>: 음.
- 3.1 <P1_F30s>: 거기 애들 내리고 나면 자리가 생기는데,
- 4.3 <P2_M30s>: 음.
- 5.1 <P1_F30s>: 그래도 막 이렇게 굳이 막 앉으려고 노력을 안 하니까 별로 앉는 일은 없는 것 같애,

예2:

- 1.1 <P1_F50s>: 뭐~ 순간 순간으루 미운 사람은 있어두,
- 2.2 <P3_F50s>: <vocal desc='웃음'>
- 3.1 <P1_F50s>: 한 한 시간 두 시간 지나믄,
- 4.1 <P1_F50s>: 그냥 잊어 버리거든요?

위의 예1은 P1이 말하고 있는 과정에 P2가 연속적으로 간투사 “음.”을 발하는 예이고, 예2는 P1이 말하는 과정에 P3이 웃음으로서 발화가 나누어진 경우이다. 이런 경우는 대화에서 많이 생기는 현상으로서 독백에서는 거의 나타나지 않는다. 또 발화 중에 다른 발화자가 미리 짐작하고 끼어들어 먼저 말하는 경우가 있다.

예3:

- 1.1 <P2_M30s>: 그러면 나는,
- 2.2 <P1_F30s>: 눈치를 보면서,
- 3.1 <P2_M30s>: 어 눈치를 보면서 인제~ 탈려고 그래,

위 예3에서 P1은 P2의 말할 내용을 미리 짐작하고 끼어들어 먼저 발화하고 P2는 계속해서 발화를 마치는 경우이다. 그러나 발화자가 발화 중에 다른 발화자의 질문을 받아 먼저 간단하게 답하고 계속 원래 발화를 계속하는 경우에는 서로 다른 문장으로 하였다.

예4:

- 1.1 <P1_F20s>: 그래서,
- 2.1 <P1_F20s>: 너무 당황스러워서
- 3.2 <P2_F30s>: 너네 학교가 어딘데?
- 4.3 <P1_F20s>: 동덕여대요.
- 5.4 <P1_F20s>: 그랬는데,
- 6.4 <P1_F20s>: 그 얘기를 딱 듣고 난 다음에,
- 7.4 <P1_F20s>: 너무 황당해서,
- 8.4 <P1_F20s>: 내가 대 뭐라고 말을 해야 되나,
- 9.4 <P1_F20s>: 이럴 때 내가 뭐라고 말을 해야 되나,
- 10.4 <P1_F20s>: 나 정말 아무 생각이 안 스치드라구요,

위 예4에서와 같이 P1은 발화 중에 P2의 질문을 받아 그 질문에 답하고 계속하여 원래 발화를 계속하는 경우인데 이런 경우에는 두 “문장”으로 처리하였다.

다음, 발화 중에 억양이 바뀌어 “억양단위”로 나누어진 경우를 살펴보면, 일반적으로 앞 발화가 연결어미로 끝나는 경우가 많다.

예5:

- 1.1 <P2_M30s>: 아니면은 이걸 정말 잘 골라야 되는데,
- 2.1 <P2_M30s>: 한번 도박을 하는 거야.

예6:

- 1.1 <P1_F30s>: 그 답에 또:: 번개 치며 달리는 물이기도 쉽고,
- 2.1 <P1_F30s>: 바다를 메꾸고,
- 3.1 <P1_F30s>: 그 답에 강을 메꾸고,
- 4.1 <P1_F30s>: 그 답에 또 남으면 우물이 되고,
- 5.1 <P1_F30s>: 그래도 남으면,
- 6.1 <P1_F30s>: 뭐가 돼요,

또한, 질문을 반복하여 질문하는 경우가 있는데 일반적으로 두 번째 발화는 첫 번째 발화에서 특정 내용을 반복하여 강조하여 질문한다. 이런 경우는 하나의 문장으로 하였다.

- 예7:
- 1.1 <P2_M40s>: 그런 얘기 많이 하시지 않아요?
 - 2.1 <P2_M40s>: 주 주변 분들이?

- 예8:
- 1.1 <P1_F30s>: 근데 어떻게 토할 수가 있어?
 - 2.1 <P1_F30s>: 버스<phon>버스</phon>에서?

그러나 한 발화자의 발화에 질문이 연속하여 나오는 경우에도 아래와 같은 경우도 있다.

- 예9:
- 1.1 <P1_F30s>: 아니 커서,
 - 2.2 <P2_M30s>: 최근에?
 - 3.3 <P2_M30s>: 그래?

예9에서 P2는 연속하여 두 개의 질문을 하는데 발화2와 발화3 사이에 P1이 “머리 끄덕임”과 같은 비언어적인 제스처를 취한 것으로 보인다. 이런 경우에는 두 개의 문장으로 처리하였다.

또한 아래와 같은 예문들도 하나의 문장으로 처리하였다.

- 예10:
- 1.1 <P3_F50s>: 지금두 가수 한 것,
 - 2.1 <P3_F50s>: <vocal desc='들이마시는숨,스'>
 - 3.1 <P3_F50s>: 참 잘 했다 그래요.

- 예11:
- 1.1 <P1_F50s>: 제 아주 직계 선배입니다,
 - 2.1 <P1_F50s>: 중고등학교 대학교,

- 예12:
- 1.1 <P1_F20s>: 그래서 그냥 그렇게 얘기하는 게,
 - 2.1 <P1_F20s>: 조금:: 뭐라고 해야 되지,
 - 3.1 <P1_F20s>: 위선은 아닌데 위선이라고 해야 되나,
 - 4.1 <P1_F20s>: 아니면 가진 자들이 부리는 뭐::라고 해야 되나,
 - 5.1 <P1_F20s>: 그냥 정확하게 딱 용어는 모르겠는데,
 - 6.1 <P1_F20s>: 좀 그런 생각이 들었어요.

- 예13:
- 1.1 <P1_F50s>: 그리고,

- 2.1 <P1_F50s>: 저희 고등학교 일학년 때?
- 3.1 <P1_F50s>: 어~ 고등학교 삼 학년 아니야,
- 4.1 <P1_F50s>: 중학교 삼학년 때 고등학교 삼학년이셨어요.

예14:

- 1.1 <P2_M40s>: 예.
- 2.1 <P2_M40s>: 그렇게 열심히 그~ 가수 활동도 인제 하시구요 또,
- 3.1 <P2_M40s>: <vocal desc='목청가다듬는소리'>
- 4.1 <P2_M40s>: 죄송합니다.
- 5.1 <P2_M40s>: 프로그램,
- 6.1 <P2_M40s>: 열심히 진행하셔 가지고 최근에 뭐~ 큰 상도 받으셨다구요.

예10은 발화 중에 비언어적 요소가 개입되면서 발화가 끊긴 경우이고, 예11은 소위 말하는 도치문의 경우이고, 예12~14는 발화 중에 다른 문장이 삽입된 경우인데 예12는 다음에 말할 단어가 생각나지 않아서 더듬으면서 말하는 경우이고, 예13은 앞에 말한 말의 내용을 고쳐서 다시 얘기하는 경우이고, 예14는 발화 과정이 매끄럽지 않아 사과하고 계속하여 말하는 경우이다. 이상의 경우의 발화는 모두 하나의 문장으로 묶었다.

이상 구어말뭉치 재주석 작업에서 “의미적 완결성”을 고려할 때 “억양단위”가 “문장”으로 묶이는 경우에 대하여 살펴보았다. 아래 3장에서는 이렇게 주석된 말뭉치에서 “문장”에 포함된 “억양단위”의 특징에 대하여 살펴보기로 한다.

3. “억양 단위”과 “문장”

3.1. “문장”에 포함된 “억양단위”

먼저, 제2장에서 재주석을 거친 말뭉치에서 “문장”에 포함된 “억양단위”의 개수를 살펴보면 아래의 표4와 같다.

	공적대화		사적대화		공적독백		사적독백		총계	
	빈도	점유율	빈도	점유율	빈도	점유율	빈도	점유율	빈도	점유율
s1	620	70.8%	733	83.7%	225	67.6%	74	38%	1652	72.4%
s2	159	18.2%	107	12.2%	73	21.9%	38	19%	377	16.5%
s3	52	5.9%	25	2.9%	25	7.5%	33	17%	135	5.9%
s4	22	2.5%	8	0.9%	5	1.5%	16	8%	51	2.2%
s5	8	0.9%	3	0.3%	3	0.9%	17	9%	31	1.4%
s6	8	0.9%	0	0.0%	2	0.6%	9	5%	19	0.8%
s7	4	0.5%	0	0.0%	0	0.0%	4	2%	8	0.4%
s8	3	0.3%	0	0.0%	0	0.0%	2	1%	5	0.2%
s9	0	0.0%	0	0.0%	0	0.0%	1	1%	1	0.0%
s10	0	0.0%	0	0.0%	0	0.0%	1	1%	1	0.0%
s11	0	0.0%	0	0.0%	0	0.0%	1	1%	1	0.0%
총계	876	100%	876	100.0%	333	100.0%	196	100%	2281	100.0%

<표4>: “문장”에 포함된 “억양단위”의 개수의 빈도표

위 표4를 보면 “억양단위”를 1개만 포함한 “문장”이 대부분을 차지한다. 장르별로 보면 “대화”에서 1개의 “억양단위”를 포함한 “문장은” 70%이상을 차지하고 “독백”에서는 “공적독백”인 경우에는 67.6%를 차지하지만 “사적독백”은 38%밖에 차

지하지 않는다.

남길임(2007)에서는 “억양단위”와 통사단위와의 관련성에 대하여 살펴본 결과 “억양단위”의 통사적 층위는 절 이상이 82.5%를 차지한다는 결론을 제기하였다. 남길임(2007)에서 사용한 말뭉치는 “공적독백”에 속하는 “강의전사말뭉치”를 사용한 결과로서 “대화”와 “공적독백”의 결과와 비슷한 결과를 보인다.

“대화”에서는 “차례말기”에 따라 자기의 차례에 가급적 하나의 “억양단위”에 “완결된 의미”를 전달하려고 하는 노력에서 “억양단위”가 바로 “문장”으로 되는 경우가 많이 나타나는 것 같다. 독백에서는 “공적독백”은 발화자가 강연을 하는 동안 거의 끼어들지 않아 “억양단위”가 바로 “문장”으로 되는 경우가 많은 반면 “사적독백”은 발화자가 자신의 경험담을 이야기하는 동안 청자는 수시로 간투사를 연발하여 여러 개의 “억양단위”가 하나의 “문장”으로 되는 경우가 많이 나타나는 것 같다.

하나의 “억양단위”로 이루어진 “문장”의 특징을 살펴보면, 하나의 어절로 된 것은 724번으로 44% 차지하고 둘 이상의 어절로 된 것은 928번으로 56%를 차지한다. 한 어절로 된 것은 아래와 같은 것들이다. 그중 “<vocal>”는 “웃음”과 같이 비언어적인 요소를 가리킨다.

음(182), <vocal>(112), 어(90), 예(52), 네(46), 응(27), 그래(26), 아(14)...

이 외에 둘 이상의 어절로 이루어진 “문장”은 아래와 같은 문형들이 많이 쓰인다.

동사문(157), 형용사문(82), 기타(58), 동사문+동사문(52), 동사+보조용언문(35), 지정사(긍정)문(33), 동사문+형용사문(24), 형용사문+동사문(18), 동사+지정사(긍정)문(16), 명사+동사과생접미사문(16), 형용사+지정사(긍정)문(14), 동사문+동사문+동사문(12), 명사+동사과생접미사문+동사문(10), 형용사+보조용언문(10)

예:

[동사문]

1. P1_F50s: 확실히는 몰라요. <4CT_0001_GR:00343>
2. P2_M30s: 갑자기 넓어지는군. <5CT_0013_GR:00705>
3. P1_F30s: 김윤정 어땠어요? <5CT_0046_GR:00279>
4. P1_F20s: 그냥 알고 지내라고. <6CT_0010_GR:00058>

[형용사문]

5. P2_M40s: 실천이 어렵다구요? <4CT_0001_GR:00016>
6. P2_M30s: 저런 황망스러웠겠구나, <5CT_0013_GR:00360>
7. P1_F20s: 나 정말 바쁘다. <6CT_0010_GR:00159>
8. P2_M40s: 저 저는 참 기분이 묘해요. <4CT_0001_GR:00041>

[동사문+동사문]

9. P3_F50s: 일본에서 발매되고 그 다음에 엘레이로 가고. <4CT_0001_GR:00582>
10. P2_M30s: 나는 이제 신문을 불러고 노력을 해. <5CT_0013_GR:00062>

[동사+보조용언]

11. P1_F30s: 이런 말 들어 보셨어요? <5CT_0046_GR:00052>
12. P2_M30s: 그러다가 요행히 자리를 차지하고 앉어. <5CT_0013_GR:00069>

[지정사(긍정)문]

13. P1_F30s: 그게 이젠 정해진 코스지, <5CT_0013_GR:00495>
14. P2_M40s: 어 저 딸만 들어요. <4CT_0001_GR:00527>

[기타]

- 15. P2_M30s: 음 역시 아줌마, <5CT_0013_GR:00111>
- 16. P2_M40s: 요즘 요즘 신세대 분들이 <4CT_0001_GR:00213>

둘 이상의 “억양단위”를 포함하는 “문장”이 쓰이는 예는 아래와 같다.

[부사+ 절]

- 17. P1_F50s: 그런데, @그때 대대장을 하셨어요. <4CT_0001_GR:00118>
- 18. P1_F20s: 그래서, @너무 당황스러워서 <6CT_0010_GR:00165>

[간투사+ 절]

- 19. P2_M40s: 네, @만우절이지요? <4CT_0001_GR:00051>
- 20. P1_F30s: 어, @들켰어? <5CT_0013_GR:00077>

[절+ 절]

- 21. P1_F50s: 마음 예쁘고, @얼굴 예쁘고. <4CT_0001_GR:00571>
- 22. P1_F30s: 계속 자? @앉아서? <5CT_0013_GR:00061>

[절+ 보충어]

- 23. P2_M40s: 그런 얘기 많이 하시지 않아요? @주 주변 분들이? <4CT_0001_GR:00064>
- 24. P1_F30s: 너무 안타까워. @학원에서 특히, <5CT_0013_GR:00254>

[인용문+ 절]

- 25. P2_M30s: 아 그래? @하는 사람이 있단 얘기네, <5CT_0013_GR:00751>
- 26. P1_F20s: 너 빠졌지? @그리는 거예요:: <6CT_0010_GR:00039>

3.2. “문장”에서 동사의 쓰임

위 2장에서 재주석을 한 말뭉치에 출현한 359개의 동사의 전체 빈도는 2,686번으로 전체(형태소 전체: 10,158)의 약 8.91%를 차지하였다. 그중 출현빈도가 10이상인 40개 동사의 전체 빈도는 1,982로서 전체 동사 출현빈도의 73.8%를 차지한다. 그 동사의 목록을 보이면 아래와 같다.

하다(505), 되다(194), 가다(136), 그러다(132), 보다(115), 알다(60), 만나다(55), 쓰다(53), 오다(49), 나오다(45), 모르다(44), 들다(43), 가지다(40), 나다(40), 타다(37), 읽다(37), 먹다(33), 듣다(32), 맞다(26), 앓다(25), 들어가다(22), 내리다(20), 살다(20), 주다(17), 찾다(17), 받다(17), 다니다(15), 대하다(15), 서다(14), 뛰다(14), 관하다(13), 이리다(12), 만들다(12), 생기다(12), 내다(11), 보이다(10), 떨어지다(10), 넘다(10), 지나다(10), 벌다(10)

이상의 동사를 『연세한국어사전』에서 제시한 격틀정보를 참조하여 사용된 갈래 뜻에 따라 필수로 하는 논항의 출현 여부에 근거하여 격틀 정보에서 필수로 하는 논항의 수에 따라 3자리 동사, 2자리 동사, 1자리 동사로 나누었다. 아래에 각 유형의 동사들의 논항의 실현 양상을 보이면 아래의 표5~7과 같다.

동사	전체빈도	3자리 동사 빈도	비율	NP1	비율	NP2	비율	NP3	비율
받다	17	17	100.0%	3	17.6%	15	88.2%	2	11.8%
주다	17	17	100.0%	5	29.4%	13	76.5%	8	47.1%
만들다	12	9	75.0%	3	33.3%	9	100.0%	9	100.0%

보이다	10	3	30.0%	1	33.3%	3	100.0%	0	0.0%
듣다	32	3	9.4%	1	33.3%	3	100.0%	3	100.0%
내다	11	1	9.1%	0	0.0%	1	100.0%	1	100.0%
하다	505	1	0.2%	0	0.0%	1	100.0%	1	100.0%

<표5>: 3자리 동사의 쓰임

동사	전체빈도	2자리 동사 빈도	비율	NP1	비율	NP2	비율
하다	505	504	99.8%	114	22.6%	407	80.8%
가다	136	129	94.9%	23	17.8%	72	55.8%
보다	115	97	84.3%	21	21.6%	74	76.3%
알다	60	60	100.0%	21	35.0%	48	80.0%
쓰다	53	53	100.0%	18	34.0%	45	84.9%
만나다	55	52	94.5%	8	15.4%	16	30.8%
모르다	44	44	100.0%	14	31.8%	30	68.2%
나오다	45	42	93.3%	25	59.5%	18	42.9%
듣다	43	42	97.7%	5	11.9%	42	100.0%
가지다	40	40	100.0%	21	52.5%	38	95.0%
읽다	37	37	100.0%	7	18.9%	21	56.8%
타다	37	36	97.3%	5	13.9%	30	83.3%
먹다	33	33	100.0%	2	6.1%	29	87.9%
듣다	32	29	90.6%	7	24.1%	22	75.9%
오다	49	29	59.2%	4	13.8%	19	65.5%
들어가다	22	22	100.0%	7	31.8%	14	63.6%
찾다	17	17	100.0%	4	23.5%	13	76.5%
맞다	26	13	50.0%	2	15.4%	13	100.0%
내다	11	10	90.9%	4	40.0%	10	100.0%
벌다	10	10	100.0%	2	20.0%	9	90.0%
서다	14	10	71.4%	6	60.0%	9	90.0%
내리다	20	9	45.0%	5	55.6%	7	77.8%
다니다	15	8	53.3%	5	62.5%	7	87.5%
넘다	10	6	60.0%	2	33.3%	4	66.7%
생기다	12	6	50.0%	2	33.3%	4	66.7%
떨어지다	10	5	50.0%	1	20.0%	2	40.0%
앉다	25	4	16.0%	0	0.0%	4	100.0%
되다	194	2	1.0%	1	50.0%	2	100.0%
만들다	12	2	16.7%	0	0.0%	2	100.0%
살다	20	2	10.0%	0	0.0%	2	100.0%
뛰다	14	1	7.1%	0	0.0%	1	100.0%
보이다	10	1	10.0%	0	0.0%	1	100.0%

<표6>: 2자리 동사의 쓰임

동사	전체빈도	1자리 동사 빈도	비율	NP1	비율
되다	194	76	39.2%	76	100.0%
가다	136	7	5.1%	2	28.6%
보다	115	1	0.9%	1	100.0%
만나다	55	3	5.5%	3	100.0%
오다	49	20	40.8%	10	50.0%
나오다	45	3	6.7%	3	100.0%
나다	40	40	100.0%	40	100.0%
타다	37	1	2.7%	0	0.0%
맞다	26	13	50.0%	3	23.1%

앉다	25	21	84.0%	5	23.8%
내리다	20	11	55.0%	5	45.5%
살다	20	18	90.0%	7	38.9%
다니다	15	7	46.7%	5	71.4%
뛰다	14	13	92.9%	3	23.1%
서다	14	4	28.6%	1	25.0%
생기다	12	6	50.0%	6	100.0%
넘다	10	4	40.0%	4	100.0%
떨어지다	10	5	50.0%	1	20.0%
보이다	10	6	60.0%	5	83.3%
지나다	10	10	100.0%	10	100.0%

<표7>: 1자리 동사의 쓰임

위 표5~7에서 볼 수 있듯이 동사들의 논항의 실현은 매우 높게 나타난다. 3자리 동사, 2자리 동사의 목적어(NP2)의 실현은 몇 개의 동사 외에는 80%이상으로 실현되었고 1자리 동사의 주어(NP1)의 실현 빈도가 일부 동사에서 매우 높게 나타났다. 이런 사실은 타동사의 경우 목적어로 쓰이는 명사의 종류에 따라 동사의 의미 분석도 가능하기 때문에 구어말뭉치에서 동사의 의미주석 작업도 가능할 것으로 예상된다.

4. 결론

이상 이 논문에서는 기존의 “억양단위”를 기본단위로 하여 구축된 “21세기 세종 계획”의 “구어형태소분석말뭉치”가 문법연구에 적합하지 않다는 판단 하에 “의미적 완결성”을 기준으로 구어에서 “문장”이라는 단위를 설정하여 말뭉치의 재주석을 진행하고, 이렇게 “문장”으로 묶여진 “억양단위”들의 특징을 살펴보았으며 마지막으로 이런 “문장”을 단위로 동사의 쓰임을 논항의 실현여부에 대한 검토를 통하여 살펴보았다.

결과 대화(공적/사적)와 “공적독백”의 경우 80%이상이 “억양단위”와 “문장”이 일치하다는 것을 확인하였고 “사적독백”의 경우 여러 개의 “억양단위”가 “문장”으로 되는 경우도 많다는 것을 확인하였다. 또한 구어에서 동사가 쓰일 때 그 논항의 실현 정도가 매우 높게 나타난다는 것을 확인하였고 구어말뭉치에서 동사의 의미주석 작업의 가능성을 확인하였다.

시간적인 제한으로 둘 이상의 “억양단위”가 “문장”으로 될 때 “억양단위”들 사이의 관계를 구체적으로 확인하지 못하였다. 이는 앞으로의 과제로 남기기로 한다.

참고문헌

- 강상호, 1989, 『조선어입말체연구』, , 평양 : 사회과학출판사
 구현정, 2000, 입말뭉치 구성과 구어의 특징, 21세기 세종계획 특수자료 구축 분과 전문가 초청 특강 발표요지, 연세대 언어정보개발연구원.
 권재일, 2004, 『구어 한국어의 의향법 실현방법』, , 서울대 출판부
 김미형, 2004, 한국어 구어와 문어의 특징 연구, 한말연구 15, 한말연구학회
 김순자·장경희, 2005, 구어자료의 정보화: 화행 주석 방법 연구, 한국언어문화 28, 한국언어문학학회
 김정자, 1999, 잡지 기사의 구어성 분석 - 10대 후반에서 20대 초반의 독자를 대상으로 한 잡지를 중심으로, 텍스트언어학 7, 한국텍스트언어학회
 김홍규 외, 1998, 21세기 세종계획 국어 기초자료 구축 (한국어 입말뭉치의 구축 방법), , 국

립국어원

- 남길임, 2007, 구어 억양 단위의 통사적 상관성 연구 -구어 독백 말뭉치를 중심으로-, 어문학 96집, 한국어문학회
- 노대규, 1996, 『한국어의 입말과 글말』, , 국학자료원
- 박동근, 2005, 구어 음운 전사 말뭉치 구축의 필요성과 방법, 겨레어문학 35, 겨레어문학회
- 서상규 외, 2006, 21세기 세종계획 국어 특수자료 구축, , 국립국어원
- 서상규·구현정, 2002, 『한국어 구어 연구 (1) -구어 전사 말뭉치와 그 활용-』, , 연세대 언어정보개발연구원
- 서상규·구현정, 2005, 한국어 구어 연구(2) -대학생 대화 말뭉치를 중심으로-, , 한국문화사
- 서상규·구현정, 2005, 코퍼스에 반영된 구어와 문어, 한국언어학회 2005년 겨울학술발표대회 자료집, 한국언어학회
- 서상규·김형정, 2005, “구어 말뭉치 설계의 몇 가지 조건”, <언어정보와 사전편찬> 제 14~16합집, 연세대 언어정보연구원
- 성연숙, 2005, 한국어 구어의 특징 -TV 토론 방송 대본과의 비교를 중심으로-, 제34차 한국어학회 전국학술대회 논문집, 한국어학회
- 안의정, 1998, 「한국어 입말뭉치 전사 방법 연구」, , 연세대학교 대학원
- 안의정, 2008, 말뭉치를 이용한 어휘의 구어성 측정과 활용, 어문논집 57, 민족어문학회
- 안정오, 1997, 언어의 문어성과 구어성, 인문대 논문집 16, 고려대학교
- 이진희, 2003, 언어의 구어성과 문어성에 관한 연구, 독일문학 85, 한국독어독문학회
- 임성규, 1989, 글말과 입말의 문체 분석, 국어국문학 102, 국어국문학회
- 장경현, 2003, 문어/문어체·구어/구어체 재정립을 위한 시론, 한국어 의미학 13호, 한국어 의미학회
- 장경희, 2009, 구어 연구의 현황과 전망, 한국어학 vol. 45, 한국어학회
- 장소원, 1995, 국어학에서의 구어성, 「한일어학논총」, 남학 이종철 선생 회갑기념논총 간행위원회
- 전영옥, 1993, 「한국어 발화단위 종결방식의 화용론적 분석」, , 상명대학교
- 전영옥, 1997, “담화분석(1) : 발표 요지와 실제 발표의 상관 관계”, 담화인지언어학회 제3회 학술대회 논문집,
- 전영옥, 2003, “한국어 억양단위 연구”, 담화와 인지 10-1, 담화인지언어학회
- 전영옥, 2006, 구어의 단위 연구, 한말연구 19, 한말연구학회
- 최명식, 1988, 『조선말구두어문법』, , 료녕민족출판사
- 최명식, 1985/1996, “『조선말 입말체 문장연구』, 해외 우리문화학연구총서95”, , 한국문화사
- 최현배, 1937, 『우리말본』, , 정음문화사
- 小磯花絵, 1996, 対話コーパス作成法, 『音声データベース』, 強誠社
- 小磯花絵, 2008, 音声談話における発話の終了性を備えた単位に関する一考察, 「単位」としての文と発話, ひつじ書房
- 沖裕子, 2008, 談話論からみた「文」と「発話」「単位」としての文と発話, ひつじ書房
- 金田純平, 2008, 発話内における単位認定の問題点について, 「単位」としての文と発話, ひつじ書房
- 国立国語研究所, 2006, 『日本語話し言葉コーパスの構築法』, 国立国語研究所報告書124, 国立国語研究所
- 国立国語研究所&情報通信研究機構, 2004, 『日本語話し言葉コーパス』(DVD), 国立国語研究所
- 前川喜久雄, 2004, 日本語話し言葉コーパス」の概要, 日本語科学 15, 国立国語研究所
- Michael McCarthy, 1998, Spoken language and applied linguistics, , Cambridge University Press.