

위의 <표 5>는 어말어미의 빈도를 보인 것이다.⁴⁾ 종결어미와 연결어미를 가리지 않았다는 점에서 위의 <표 5>는 다시 둘로 나누어질 수도 있을 것이다. 그럴 경우에는 연결어미와 종결어미를 구분하지 않은 '고'라든가 '게' 또는 '나'와 같은 형태에 대한 재분석이 이루어져야 할 것이다.⁵⁾ 하지만 본 연구가 목적하고 있는 한국어 교육에서의 활용이라는 측면에서 본다면 그들이 대한 별도의 분석 작업은 보류되어도 일단은 무방할 것으로 보인다. 교육의 장면에서는 하나의 형태가 제공될 경우 같은 자리에서 다양한 의미를 제공할 필요도 있을 것으로 보이기 때문이다. 즉 피교육자의 혼동 방지와 교육 효과를 높이기 위하여 기본적인 의미나 기능을 나누어 제공한다고 하더라도 해당 형태의 비중을 파악하기 위한 빈도 조사에서는 같은 자리에서 파악하는 것이 온당한 것이라고도 할 수 있기 때문이다. 아무튼 <표 5>에 반영된 결과는 어말어미인 '다'와 'ㄴ'의 빈도가 압도적인 것임을 보여 주고 있다. 그러나 위의 <표 4>나 <표 5>에서와 같이 개별 형태의 출현 빈도수는 표면적인 것이라 할 수 있다. 교육적인 측면에서의 활용도를 높이기 위해서는 2차 해석이 필요하기 때문이다. 형태 목록을 문법 범주별로 재분류하는 것이 그것이다.

아울러 위에서 따로이 살펴본 선어말어미와 어말어미의 출현 빈도는 함께 묶어 보는 것도 유용할 것이다. 문법 형태 측면에서의 학습 진행 과정을 구성하는 데에는 함께 살펴 순서를 정하는 것이 필요할 것이기 때문이다. 본고에서는 선어말어미와 어말어미를 아우른 전체 빈도 양상을 [부록1]로 보이기로 한다. [부록2]에는 본 연구의 대상이 된 기초 자료를 소개하기로 한다.

3. 결론

3.1 활용 방안

본 연구의 결과는 다음과 같은 측면에서의 활용 가능성을 가진다.

첫째, 외국인 한국어 학습의 실러버스를 합리적으로 재조정하는 기준 제공한다. 이를 위하여 현재 각 기관에서 간행한 외국인을 위한 한국어 교재에서 교차된 어미 형태의 제시 순서가 어떻게 되어 있는지 검색해 볼 필요가 있다. 그에 따라 본 연구의 결과를 통하여 그 개선 방안을 제시할 수

4) 번호에 '-'를 붙인 형태는 분석과정에서 추가된 목록임을 뜻한다.

5) 어미 '이'의 빈도 1,018에는 부사파생접사도 포함되어 있다. 부사형어미 '게'를 염두에 둔 처리이기는 하지만, 다른 형태들과의 형평을 이루어야 한다는 점에서는 다시 다듬어질 필요가 있다고 하겠다.

있기 때문이다.

둘째, 국내의 초급 국어 교재에서 각 형태소의 도입 순서도 재검토해 볼 기준 자료로 활용할 수 있다. 앞서와 마찬가지로 국어교육의 초급교재들에서도 용언부 연쇄 형태의 학습을 재검토하여 가장 자연스런 배열을 권장할 수 있는 것이다.

셋째, 기계번역의 (특히 영한, 일한 부분의) 개발에 큰 지표로써 어떤 국어 표현을 더 선호해서 번역해야 하는가 하는 기준 목록을 보여 준다. 이 공계 주도로 이루어지고 있는 기계 번역에 기초 언어정보 제공이 갈망되고 있는 실정¹⁾이므로, 계량언어학적, 즉 자연과학적 방안에 의한 언어 정보의 연구 결과를 비인문계통에 넘겨주는 적절한 기여를 이 연구의 결과에서 기대해도 좋을 것이다.

네째, 위에서 살핀 검토의 결과는 이형태와 기본형 선정이라는 면에서도 유용할 수 있다. 이형태들 가운데 기본형을 정하는 기준으로 설명이 자연스러운 방향, 역사적인 측면의 고려 등을 들기도 하지만, 그러한 판단 기준 적용이 쉽지 않을 경우에 또다른 중요한 기준으로 빈도수를 들기 때문이다. 하지만 그동안 문법형태의 빈도에 대한 조사 결과가 없었기 때문에 그러한 기준을 드는 것 자체가 공허한 것이기도 하였던 것이다. 이를 테면 과거시제형태소 '았/었/였' 가운데에서 기본형을 정하는 경우에 우리의 자료에 따르면 압도적인 출현 양상을 보이는 '었'으로 정할 수 있는 것이다.

3.2 남은 문제

이렇듯 본 연구의 결과가 시사하는 내용과 그 활용 가능성이 많다고는 하지만, 아직은 이와 같은 작업이 시작에 불과한 것이라는 점에서 여전히 많은 문제를 남기고 있다고 할 수 있다. 우선 들 수 있는 것은 대상 자료의 확충이다. 보다 정확한 이야기를 하기 위해서는 보다 많은 자료를 통한 확인 작업이 필요하기 때문이다. 위에서 살핀 결과도 그 효용성을 높이기 위해서는 보다 다양하게 분류될 필요가 있다. 앞서 과거시제의 압도적인 출현 양상이 대상 자료의 성격에 의한 것일 가능성을 시사한 바 있다. 그와 같은 가능성의 확인을 위해서는 자료별로 그 분포 양상을 갈라서 살펴볼 필요가 있다는 것이 그것이다. 앞서 지적한 바 있듯이 결과에 대한 2차적인 해석 작업 즉 범주별 재분류 작업 및 그에 따른 해석 작업도 본 연구의 후속 작업으로 수행되어야 할 것이다.

아울러 본 연구에서는 다루지 않았으나 조사의 분포 및 빈도 확인 작업도 본 연구와 평행적인 성격의 작업이라는 점에서 우리의 손을 기다리고 있는 부분이라고 할 수 있다.

참고문헌

- 고영근(1989) 국어형태론연구, 서울대학교 출판부.
- 남기심, 고영근(1993) 표준국어문법론, 개정판, 탑출판사.
- 김영석, 이상익(1993) 현대형태론, 학연사.
- 안병희(1965) 문법론, 국어학개론, 수도출판사. pp.108-146.
- 안병희(1967) 문법사, 한국어 발달사, 한국문화사대계 V, 고대 민족문화연구소. pp. 165-261.
- 이상익(1986) 계량언어학(어휘론부분): 어휘부의 음운, 형태, 의미론적 구조. 언어 11.2. pp.323-339.
- 이상익(1989a) 국어 어휘 목록의 형태, 음운론적 구조 연구: 계량언어학적 표준조사. 어학연구 25.1. pp.111-128.
- 이상익(1989b) 음절구조의 변화 원인에 대한 이런 설명은 어떨까요? 정연찬 선생 회갑기념논총, 태학사. pp.203-210.
- 이상익(1989c) A Glottometrical Study of Korean Lexicon. Harvard Studies in Korean Linguistics III, Hanshin Publishing Co. pp.159-166.
- 이상익(1990a) 현대국어 음변화규칙의 기능 부담량. 어학연구 26.3. pp.441-467.
- 이상익(1990b) 국어 어휘부의 계량언어학적 연구. 강신함 선생 회갑기념논총, 태학사. pp.331-346.
- 이상익(1993) 현대 국어의 통사적 현상에 대한 기능 부담량. 안병희 선생 회갑 기념논총, 신구문화사. pp.475-487.
- 김영택 교수 주관 서울대 공대 컴퓨터공학과 자연언어처리연구실 보고 총서.
- Cf. Altmann 교수 편집 Quantitative Linguistics 총서, 독일 Studienverlag Dr. N. Brockmeyer, Bochum 출간.

001	FE127	다	: 6039	036	FE223	라	: 256
002	FE042	ㄴ	: 5790	037	FE120	니	: 239
003	FE021	고	: 3843	038	FE166	도	: 223
004	FE093	는	: 3056	039	FE696	조	: 218
005	PFE015	었	: 2454	040	FE647	지만	: 209
006	FE371	어	: 2341	041	PFE007	더	: 196
007	PFE016	였	: 1498	042	FE567	음	: 193
008	PFE014	았	: 1300	043	FE071	나	: 180
009	FE017	게	: 1288			(연, 종결 구분)	
		(연결, 종결 구분)		044	FE062	ㄴ 데	: 177
010	FE643	지	: 1109	045	FE386	여	: 174
011	FE355	아	: 1070	046	FE379	어요	: 170
012	FE581	이	: 1018	047	FE171	듯	: 165
		(접사포함)		048	PFE001	거	: 163
013	FE326	바니다	: 870	049	FE076	냐	: 156
014	FE499	은	: 858	050	FE395	오	: 154
015	FE403	요	: 842	051	PFE009	리	: 147
016	FE376	어서	: 674	052	FE307	리라	: 134
017	FE367	야	: 656	053	PFE005	느(ㄴ/냐/ㄴ지):127	
018	FE320	면	: 562	054	FE022	고는	: 118
019	PFE006	느(다)	: 557	055	FE035	군	: 117
020	FE036	기	: 531	055	FE167	도록	: 117
021	FE319	며	: 480	057	FE681	으면	: 115
022	FE347	습니다	: 475	058	FE226	라는	: 114
023	FE046	ㄴ 다	: 454	059	FE484	으며	: 112
023	FE360	아서	: 454	060	FE410	으나	: 111
025	PFE002	겠	: 453	061	FE224	라고	: 105
026	FE153	던	: 449	062	FE150	더니	: 103
027	FE131	다고	: 440	063	FE340	세요	: 99
028	FE593	이다	: 411	064	FE043	ㄴ가	: 91
029	FE633	자	: 361	065	FE633	이지	: 86
030	FE512	을	: 359	066	FE048	ㄴ다는	: 85
031	FE321	면서	: 357	067	FE635	인데	: 82
032	PFE013	시	: 352	068	FE121	니까	: 76
033	FE115	는데	: 312	069	FE087	느냐	: 74
034	FE123	다가	: 284	070	FE068	ㄴ지	: 72
035	FE133	다는	: 271	070	FE103	는다	: 72

070	FE117	는지	: 72	109	FE187	르라고	: 36
073	FE030	구나	: 70	110	FE670	래도	: 35
073	FE230	라도	: 70	111	PFE019	으시	: 34
075	FE138	다면	: 66	112	FE077	냐고	: 32
075	FE375	어라	: 66	113	FE232	라면	: 31
075	FE600	이라	: 66	113	FE636	인지	: 31
078	FE025	고서	: 61	115	FE361	아야	: 30
079	FE414	으니	: 60	116	FE072	나니	: 29
080	FE341	소	: 59	116	FE388	여라	: 29
081	FE170	든지	: 58	118	FE169	든	: 28
081	FE359	아라	: 58	118	FE254	려	: 28
081	FE325	버니까	: 57	118	FE515	을까	: 28
084	FE007	거든	: 55	118	FE591	이니	: 28
085	FE001	지나	: 54	118	FE682	이면	: 28
086	FE377	어야	: 53	123	FE148	더군	: 26
086	FE589	이나	: 53	123	FE231	라며	: 26
088	FE257	려고	: 52	125	FE174	르	: 25
088	FE374	어도	: 52	126	FE045	고	: 24
090	FE409	우	: 51	127	FE134	다니	: 22
091	FE648	지요	: 49	127	FE644	지는	: 22
091	FE663	다며	: 49	129	FE094	는가	: 21
091	FE667	다며	: 49	129	FE664	대	: 21
094	FE690	이요	: 48	131	FE123	니깐	: 20
094	FE095	는걸	: 47	131	FE168	되	: 20
096	FE172	듯이	: 46	131	FE477	으리라	: 20
097	FE037	기로	: 45	131	FE503	은데	: 20
098	FE364	아요	: 43	131	FE572	웁니까	: 20
099	FE041	가에	: 41	131	FE650	절	: 20
099	FE261	려는	: 41	131	FE692	이지만	: 20
099	FE370	야지	: 41	138	FE170-1	든가	: 19
102	FE245	러	: 40	138	FE366	았습니다	: 19
102	FE335	예요	: 40	138	FE666	든가	: 19
104	FE030	네	: 39	141	FE154	더라	: 18
104	FE346	습니까	: 39	141	FE389	여서	: 18
106	FE010	거라	: 38	143	FE098	는구려	: 17
107	FE044	고걸	: 37	143	FE102	는군	: 17
107	FE628	이며	: 37	143	FE669	르래	: 17

146	FE630	이오	: 16	180	FE322	먼은	: 10
147	FE434	으리	: 15	184	FE204	르세	: 9
147	FE573	옵니다	: 15	184	FE227	라니	: 9
147	FE691	이어서	: 15	184	FE358	아도	: 9
150	FE023	고도	: 14	187	FE018	제나	: 8
150	FE088	느냐고	: 14	187	FE089	느노	: 8
150	FE105	는다는	: 14	187	FE097	는구나	: 8
150	FE212	르자	: 14	187	FE239	란다	: 8
150	FE393	였습시다	: 14	187	FE317	마	: 8
150	FE415	으니까	: 14	187	FE422	으라	: 8
150	FE417	으니깐	: 14	187	FE435	으려	: 8
150	FE501	은걸	: 14	187	FE649	진	: 8
158	FE011	건	: 13	187	FE668	래	: 8
158	FE023	곤	: 13	187	FE683	이건	: 8
158	FE073	냐는	: 13	197	FE065	느바	: 7
158	FE155	더라도	: 13	197	FE083	노라	: 7
158	FE164	던지	: 13	197	FE092	느라고	: 7
158	FE694	시우	: 13	197	FE130	다간	: 7
164	FE079	너라	: 12	197	FE176	르걸	: 7
164	FE081	노니	: 12	197	FE228	라니까	: 7
164	FE243	랴	: 12	197	FE264	려는데	: 7
164	FE485	으세요	: 12	197	FE423	으라고	: 7
164	FE660	는단	: 12	197	FE599	이든지	: 7
169	FE090	느니	: 11	197	FE640	자마자	: 7
169	FE139	다면서	: 11	197	FE645	지를	: 7
169	FE140	다시피	: 11	208	FE129	다가는	: 6
169	FE149	더냐	: 11	208	FE173	디	: 6
169	FE162	던데	: 11	208	FE207	르수룩	: 6
169	FE183	르께	: 11	208	FE266	려니	: 6
169	FE255	려거든	: 11	208	FE269	려다가	: 6
169	FE339	세	: 11	208	FE275	려면	: 6
169	FE604	이라면	: 11	208	FE332	바시오	: 6
169	FE652	느데다	: 11	208	FE432	으라	: 6
169	FE673	으므로	: 11	208	FE500	은가	: 6
180	FE104	는다고	: 10	208	FE672	려면	: 6
180	FE159	던가	: 10	208	FE674	마는	: 6
180	FE297	로군	: 10	208	FE688	이니까	: 6

208	FE695	시요	: 6	253	FE142	단다	: 3
221	FE137	다만	: 5	253	FE205	르세라	: 3
221	FE141	다오	: 5	253	FE265	러는지	: 3
221	FE233	라면서	: 5	253	FE268	려다	: 3
221	FE237	라오	: 5	253	FE285	련만	: 3
221	FE238	락	: 5	253	FE288	럽니까	: 3
221	FE281	려오	: 5	253	FE353	시거든	: 3
221	FE306	리다	: 5	253	FE392	여요	: 3
221	FE514	을걸	: 5	253	FE447	으려니	: 3
221	FE524	올라고	: 5	253	FE474	으리니	: 3
221	FE556	을지	: 5	253	FE523	올라	: 3
231	FE015	건만	: 4	253	FE569	음에도	: 3
231	FE135	다마는	: 4	253	FE590	이네	: 3
231	FE156	더라면	: 4	253	FE625	이로다	: 3
231	FE160	던걸	: 4	253	FE642	잘	: 3
231	FE201	르밖에	: 4	253	FE651	ㄴ답니까	: 3
231	FE277	려무나	: 4	253	FE656	냐며	: 3
231	FE323	므로	: 4	253	FE665	답니다	: 3
231	FE363	아야지	: 4	253	FE675	면야	: 3
231	FE368	야만	: 4	276	FE004	거니와	: 2
231	FE373	어다가	: 4	276	FE008	거들랑	: 2
231	FE412	으냐	: 4	276	FE012	전대	: 2
231	FE438	으려고	: 4	276	FE031	구려	: 2
231	FE442	으려는	: 4	276	FE073	나마	: 2
231	FE582	이거나	: 4	276	FE096	는고	: 2
231	FE585	이거든	: 4	276	FE132	다느니	: 2
231	FE588	이기에	: 4	276	FE145	더구나	: 2
231	FE592	이니깐	: 4	276	FE173-1	디	: 2
231	FE653	ㄴ데도	: 4	276	FE175	르거나	: 2
231	FE655	보담	: 4	276	FE184	르꼬	: 2
231	FE689	이라든가	: 4	276	FE198	르망정	: 2
231	FE693	자는	: 4	276	FE199	르바	: 2
231	PFE008로(구나/구려)	: 4		276	FE229	라든지	: 2
253	FE003	거니	: 3	276	FE258	려나	: 2
253	FE027	고자	: 3	276	FE267	려니와	: 2
253	FE033	구먼	: 3	276	FE287	럼	: 2
253	FE058	ㄴ대	: 3	276	FE300	로라	: 2

276	FE304	리니	: 2	314	FE295-1	로구만	: 1
276	FE356	아다	: 2	314	FE298	로다	: 1
276	FE357	아다가	: 2	314	FE316	口직하다	: 1
276	FE382	언만	: 2	314	FE318	매	: 1
276	FE390	여야	: 2	314	FE362	아야만	: 1
276	FE436	으려거든	: 2	314	FE372	어다	: 1
276	FE449	으려다	: 2	314	FE400	올시다	: 1
276	FE472	으리	: 2	314	FE411	으나마	: 1
276	FE476	으리다	: 2	314	FE421	으되	: 1
276	FE605	이라야	: 2	314	FE428	으라면	: 1
276	FE621	이로구나	: 2	314	FE430	으락	: 1
276	FE624	이로군	: 2	314	FE450	으려다가	: 1
276	FE626	이로되	: 2	314	FE454	으려도	: 1
276	FE658	냐가	: 2	314	FE462	으려오	: 1
276	FE659	누나	: 2	314	FE466	으련만	: 1
276	FE662	다네	: 2	314	FE482	으마	: 1
276	FE671	래야	: 2	314	FE483	으매	: 1
276	FE677	이므로	: 2	314	FE509	은지	: 1
276	FE685	이긴	: 2	314	FE517	을까말까	: 1
276	FE687	이노	: 2	314	FE522	을는지	: 1
314	FE002	거늘	: 1	314	FE587	이기로	: 1
314	FE026	고야	: 1	314	FE617	이려니와	: 1
314	FE032	구료	: 1	314	FE629	이므로	: 1
314	FE034	구면	: 1	314	FE646	지마는	: 1
314	FE050	ㄴ다니까	: 1	314	FE654	ㄴ테서	: 1
314	FE053	ㄴ달	: 1	314	FE657	냐면서	: 1
314	FE059	ㄴ대서	: 1	314	FE661	는도다	: 1
314	FE066	ㄴ즉	: 1	314	FE673	로구만	: 1
314	FE084	노라고	: 1	314	FE679	으래	: 1
314	FE086	노라면	: 1	314	FE680	쇼	: 1
314	FE108	는달	: 1	314	FE684	이거니	: 1
314	FE116	는바	: 1	314	FE686	이길래	: 1
314	FE202	ㄹ뿐더러	: 1				
314	FE222	ㄹ진저	: 1	363	FE005	거드면	: 0
314	FE235	라야	: 1	(363 순위 이후 361개 항목은 출현 빈도가 역시 0이므로 생략, 따라서 원래 총 항목 수는 724개.)			
314	FE241	랍니다	: 1				
314	FE263	려는고	: 1				

[부록2] <대상 자료(원고, 인쇄물, 녹음) 일부 예>

<TV뉴스>

오늘의 주요뉴스

박태중, 김현철 자금제공 했다

김철, 한상수 의원 소환키로

카메라 활동, 한강의 심야 불법어로

미스코리아는 기형인가?

이인용 앵커:

안녕하십니까? MBC 뉴스 데스크입니다.

(중략)

타고난 美 아니다

이인용 앵커:

우리나라에서 가장 아름다운 여자를 꼽으라면 아무래도 미스코리아를 떠올리게 됩니다. 그런데 한 한의사가 역대 미스코리아의 체형을 분석해 봤는데 키는 평균에 비해서 너무 크고 체중은 평균치에 모자라는 비정상 체형이다 이런 결론을 내렸습니다.

이진숙 기자입니다.

이진숙 기자:

지난 95년에 미스코리아, 키는 그 나이 또래 천 명 가운데 18번째로 큰 170cm였고, 가슴둘레도 101번째로 큰 편이었습니다. 그러나 몸무게는 50kg 밖에 나가지 않았습니다. 92년부터 96년까지 미스코리아 최종 후보들의 키를 분석해 봤습니다. 평균키가 171.7cm 인데 반해서 몸무게 평균은 51kg에 불과했습니다.

옥도훈 한의사:

체중만 보더라도 표준 체중보다 10kg이 모자란 상태인 저체중이거나 영양 결핍인 상태인데 가슴은 일반인보다 기형적으로 큰 편으로 나타납니다. (후략)

현대국어 용언 형태의 계량언어학적 연구¹⁾

이상익·이현희

(서울대학교)

Lee, Sang-Oak and Lee, Hyeon-Hie. 1997. A Quantitative Study on the Verbal Endings in Korean. *Studies in Phonetics, Phonology and Morphology*, Vol. 3, 203-222. A reliable investigation on the frequency of grammatical elements in the Korean predicate is not yet available. Occurrence of prefinal and final endings should be measured statistically for many purposes. In order to facilitate the learning process of Korean, it is desirable to introduce them in parallel with their frequency in Korean corpus. For example, in the case of foreigner's learning Korean as well as the case of Korean children's learning Korean should be guided by these statistics in their textbooks and other reference material such as dictionaries and thesauri. In machine translation, the frequency affects the design of such program in considering favorably high-frequency elements. In selecting the basic form out of Korean past tense markers(-ot-, -at- and -yot-) the most frequent -ot- may be chosen because of its high frequency. (Seoul National University)

1. 서론

1.1 연구목적

본 연구는 현대 국어 용언 형태론의 복잡한 교착(膠着 agglutination) 상태, 즉 선어말어미 및 연결어미, 종결어미 등의 각개 출현 및 연쇄 출현 빈도를 계량적으로 파악하여, 외국인이나 초학자들이 한국어를 배울 때 직면하는 가장 큰 난점을 합리적으로 해결하는 방법을 마련하고, 기계번역 등에서의 절차 설정에도 통계적 바탕을 제공하려는 데에 목적을 두고 있다.

현재 세계 각처에서 한국어를 외국인에게 가르치고 있고 그 확장세가 뚜렷하지만 정작 그 교육 방법은 다른 주요 언어의 그것에 비해 아직 잘 개발되지 못한 부분이 많다. 심지어는 국내외의 초급 국어 교재에서 각 형

1) 본 연구는 1996년도 학술진흥재단의 국제협력연구과제 학술연구조성비 지원에 의한 것이다. 본 연구는 용언 형태 중에서도 특히 선어말/어말 어미에 대한 연구이며, 한신대 국문과 한재영 교수팀의 전산처리에 힘입은 바 크다.

태소의 도입 순서도 마련되어 있지 못한 것이 현실인 것이다.

본 연구의 주목표로는 다루지 않으려 하지만, 본 연구의 부차적 효과로서 최근 시도되고 있는 한영, 영한, 한일, 일한 언어간의 기계 번역에 대한 내용 확충과 효율성 증대를 위한 개선 노력에도 본 연구의 결과는 많은 도움을 제공하게 될 것이다. 번역의 질적 개선이나 외국인 한국어 교육 방법의 개선은 자기 다른 목표를 가지는 것처럼 보이나 유사한 방법으로 성취할 수 있는 내용이 많다. 두 분야는 또 연구의 필요성도 현대 사회에서 국제적 교섭이 빈번해 지는 데 비례해서 더더욱 증대해 가고 있다.

이러한 현실에 비추어 볼 때, 본 연구가 행하려는 현대국어 선어말어미 및 연결어미, 종결어미 등의 각개 및 연쇄 출현 빈도를 계량적으로 파악하려는 작업은 다른 작업에 앞서 이루어져야 할 기초적인 연구라고도 할 수 있는 것이다.

1.2 연구사

그동안 이루어진 국어의 계량적 연구로는 과거 연차적 연구 계획에 의해 '참고문헌'에 보인 이상역(1989a,b,c,1990a,b,) 등을 들 수 있다. '국어 어휘 목록의 형태, 음운론적 구조 연구:계량언어학적 표준조사'(1989a,b,c)와 '현대국어음변화규칙의 기능 부담량'(1990 a) 및 '현대 국어의 통사적 현상에 대한 기능 부담량'(1993)이 그것으로 그 배경에는 이상역(1986)에 소개된 것과 같이 독일 보쿰(Bochum)의 Altmann교수등이 이끄는 계량언어학(quantitative linguistics)이 자리잡고 있다. 계량언어학적 연구는 현재도 독일 트리어(Trier) 대학을 중심으로 국제학회가 구성되어 많은 연구업적을 쌓고 있으며, 이와 유사한 관점의 다른 연구 집단의 출현도 볼 수 있다.

한편 안병희(1965,1967)는 이러한 연구들을 위한 국어의 형태론적 기술의 틀을 잘 보여 주고 있다. 그러나 종결어미 및 연결어미 출현 빈도의 조사 집계는 기이하게도 아직까지 조사가 본격적으로 수행 공개된 바가 없다. 1955년에 실시된 문교부 주관 "우리말 찾기 조사"에서도 용언의 기본(비활용) 형태로만 조사해 놓았을 뿐 활용어미가 붙은 형태는 묵살해 버렸다. 근래에야 활용어미(종결, 연결어미 등)에 대한 조사가 관심을 끌고 있지만 그 방대한 작업량때문에 아직 사용 가능한 결과 보고가 전무한 실정이다.

국어는 선어말 및 어말 어미의 교착 내지 첨가로 피동 및 사동태, 시상, 존비법 및 겸양법 등을 다양하게 덧붙이는 언어로서, 굴절(inflexion)의 양상이 복잡해서 외국인이나 어린아이들이 한국어를 배워가는 초기 단계에서 큰 고비를 느끼게 한다.

가령 ‘깨-뜨리-시-었-겠-습니-다-만’ 같은 표현을 놓고 외국인이 분석하려면 어디에 형태소 경계가 있는지 쉽게 알아채기 어렵다. 보통 사전에는 각 형태소 별로 ‘깨-’ ‘-뜨리-’ ‘-시-’ 등만이 등재되어 있어서 만약 ‘깨뜨-’, ‘-리시-’ 등으로 잘못 경계를 끊었을 때는 그 뜻을 찾아 보기도 어려운 형태이다.

이런 난점은 ‘깨뜨-’, ‘-리시-’ 같이 잘못 자른 형태까지 사전에 등재하는 편법으로 해결할 수 있을지는 모르겠으나, 그런 연쇄형태란 그 수가 엄청나게 많을 뿐만 아니라 잘못 분석된 형태를 자꾸 보이게 된다는 점에서도 바람직한 조치는 아니라 하겠다. 그 보다는 초기에 잘못 자르는 시기가 좀 있더라도 학습자가 각 형태소의 경계를 스스로 알아챌 수 있도록 점차 훈련 시켜 가는 것이 문제에 대한 온당한 접근 방법이라고 할 수 있다. 이를 위해서는 어떤 연쇄 형태가 가장 많이 쓰이는가를 통계적으로 조사해서, 학습과정에서 적절히 순서대로 노출시켜 주는 것이 필요하다. 예컨대 ‘-었겠-’은 가능한 연쇄 형태로서 실제 어떤 빈도로 우리가 쓰고 있는지 조사해서, 만약 그것이 자주 쓰이는 것이라면 학습 초기에 도입시켜야 할 것이다. 이렇게 모든 내용을 계량적으로 파악하면 모든 형태의 적절한 도입 순서를 설정할 수 있다.

본 연구에서는 종결어미의 연구와 함께 연결어미의 출현 빈도도 조사 집계해 볼 계획이다. 기이하게도 아직까지 이들 어미의 사용빈도에 대한 조사가 본격적으로 수행 공개된 바가 없다. 1955년에 실시된 문교부 주관 “우리말 찾기 조사”에서도 용언의 기본(비활용) 형태로만 조사해 놓았을 뿐 활용어미가 붙은 형태는 묵살해 버린 이후 근래에 와서 활용어미(종결, 연결어미 등)에 대한 조사가 관심을 끌고 있지만 그 방대한 작업량때문에 아직 사용 가능한 결과 보고가 전무한 실정이다.

2. 본론

2.1 연구방법

본 연구의 연구방법은 두 가지 측면에서 이야기할 수 있다. 그 하나는 대상자료에 관한 것이고, 다른 하나는 자료의 검색방법에 관한 것이다. 먼저 본 연구의 대상이 현대 국어라는 점과 개별 형태의 빈도에 관심을 가지고 있다는 점에서 본다면, 현대 국어의 전형적인 모습을 담고 있는 모든 자료가 검토의 대상이 되어야 할 것이나 그와 같은 작업은 불가능한 것일 뿐만 아니라 비경제적인 것이기도 한 것이다. 더구나 우리의 목적 달성에 적절한 양의 전형적인 자료를 통한 표본 검사로 족하다 할 수 있다. 검

토 대상 자료의 양은 많으면 많을수록 좋은 것이고, 앞으로 연구가 진행되어 감에 따라 검토 대상 자료의 양을 확대하여 결과에 대한 수정, 보완 및 정밀화 과정이 뒤따라야 할 것이나, 본 연구에서는 1차적인 대상 자료의 양을 약 100,000 어절 정도로 잡았다. 물론 작업의 편의 또는 수행 가능성이라는 기준 이외에 100,000이라는 수가 가지는 객관적인 근거를 제시하기는 어렵다. 그러나 구성 자료의 성격이 될 수 있으면 어느 한 방면에 치우쳐서는 안된다는 점과 현실성이 있는 자료로 구성이 되어야 한다는 점 그리고 그들의 내용을 어느 정도는 반영할 수 있는 하나의 단위가 될 수 있는 자료이어야 한다는 점을 자료 선택의 큰 기준으로 삼았다는 데에서 타당성을 부여할 수 있을 것이다.

자료명	어절수	전체에 대한 비율
신문	31,759	30.25%
성경(창세기)	24,793	23.62%
TV뉴스	4,998	4.76%
소설(깃발..)	22,727	21.65%
드라마(짹)	5,393	5.14%
PD수첩	3,331	3.17%
시(생명의 서)	3,696	3.52%
희곡(광랑의 전설)	8,279	7.89%

<표 1> 대상 자료의 내용

위의 <표 1>에 보인 자료의 전체 어절수는 104,976개이다. 신문은 1996년 8월 17일자 '조선일보'와 '중앙일보' 및 '한국일보'에서 취한 것이고, 성경은 공동번역 가운데 '창세기'를 가져온 것이다. TV 뉴스는 1997년 4월 22일에 방송된 'MBC 뉴스 데스크'의 것이다. 소설은 선우 휘의 '깃발 없는 기수'이며, 드라마는 1997년 5월 11일에 방송된 드라마 '짹 : 제129회 라이벌 죽이기'를 취한 것이다. PD수첩은 1997년 5월에 방송된 '굶주린 복녀 두만강 접경지대를 가다'의 내용이다. 시는 1947년 행문사에서 나온 유치환의 '생명의 서'이며, 희곡은 인문평론 1940년 10월호에 실린 '螳螂의 傳說'이다. 시사성이 있는 자료들은 현대국어의 사용어휘의 양상을 잘 반영하고, 문학

작품들은 이해어휘의 내용을 보여준다. 뉴스와 드라마, PD수첩 및 희곡은 구어적인 모습을 반영하는 자료로 약 21%를 차지한다. 쉽게 생각하기로는 구어와 문어의 비율이 각각 50%인 것이 이상적인 것으로 생각할 수도 있겠으나, 구어자료의 경우에는 자료 처리상의 어려운 점 이외에 자료의 질이 덜 정제된 모습을 보인다는 점에서²⁾ 20% 이상을 많이 넘는 것은 오히려 피해야 할 것으로 판단한 것이다. 시를 대상 자료에 포함시킨 것은 일상 생활에서는 잘 쓰이지 않으나 문학 작품 등에서 여전히 생명력을 가지고 있는 형태들을 놓치지 않으려는 배려에서이다. 1940년대의 희곡을 선택한 데에 대해서도 또다른 이견이 있을 수 있겠으나 이는 현대국어의 사용자가 젊은이들에 국한되는 것이 아니라는 점을 기억하려 한 조치이다. 국어사용자들의 연령을 고려한 것이다. PD수첩의 경우는 여러 가지 내용 가운데 특히 '끓추린 복넙 두만강 접경지대를 가다'를 취한 것은 대화 가운데 북한어 언어 모습을 살필 수 있다는 점을 취한 것이다. 이러한 배려들이 결과에 반영되었는지의 여부는 오히려 2차적인 것이라 할 수 있다. 대상 자료를 선택하는 과정에서 특히 적은 량의 자료를 통하여 전형적인 모습에 다가서기 위한 작업에서는 개별 자료가 가지는 다양한 성격을 염두에 두는 것이 온당한 태도라 할 것이다.

이상의 자료를 통하여 개별 문법 형태의 사용 양상을 살피는 가장 정확한 방법은 하나하나 눈으로 확인하고 카드 작업화하여 정리하는 것이라 할 수 있다. 그러나 그와 같은 자료 처리 방식이 가지는 한계는 자명하다. 많은 양의 자료를 감당할 수 없다는 것이 그것이다. 그렇다고 하여 기계적인 방식으로 일률적으로 적용할 수도 없는 것이 현실이다. 아직 우리의 '형태소 분석기'는 문맥에 따른 다양한 가능성은 제시할 수 있되 각각의 문맥에서의 용법 판단에는 이르지 못하고 있기 때문이다. 따라서 본 연구에서는 두 가지 방식을 병용하기로 하였다. 그를 위하여 기존의 업적을 통한 문법 형태소 목록의 확인 작업으로부터 출발하였다. 다음에 보이는 <표 2>와 <표 3>이 그것이다.

■ 신어말어미

거	겠	ㄴ(다)	나이(다)	느(ㄴ/냐/ㄴ지)	는(다)	더
로(구나/구려)	리	사오	사읍	삼	시	았
였	오	읍	으시	으오	으읍	자오
						자읍
						잡

<표 2>

2) 자료의 정제성이라는 점에서 본다면 방송 자료는 그래도 나은 편이다. 일상적인 구어를 대상으로 삼을 경우에는 정제의 정도는 이야기로 삼기조차 어렵다 할 것이다.

■ 이말어미

거나 거늘 거니 거니와 거드면 거드면은 거든
 거들랑 거들랑은 거라 건 건대 건댄 건마는
 건만 것다 게 게나 게시리 게끔 고 고는
 고도 고달고 고서 고야 고자 곧 관테 구나
 구려 구료 구면 구면 군 기 기로 기로서
 기로서니 기로선들 기에 ㄴ ㄴ가 ㄴ걸 ㄴ고
 ㄴ다 ㄴ다고 ㄴ다는 ㄴ다느냐 ㄴ다니 ㄴ다마는
 ㄴ단다 ㄴ달 ㄴ담 ㄴ답니다 ㄴ답디다 ㄴ답시고 ㄴ대
 ㄴ다서 ㄴ대서야 ㄴ대야 ㄴ데 ㄴ들 ㄴ만큼 ㄴ바
 ㄴ죽 ㄴ죽슨 ㄴ지 ㄴ지고 ㄴ지라 나 나니 나
 마 나йка 나이다 나 나고 나는 너라
 네 노 노다 노라 노라고 노라니 노라면 느냐
 느냐고 느뇨 느니 느니라 느라고 는 는가 는
 걸 는고 는구나 는구려 는구료 는구면 는구면
 는군 는다 는다고 는다는 는다마는 는단다 는
 달 는담 는답시고 는대 는대서 는대서야 는대야
 는데 는바 는지 는지고 는지라 니 니까 니
 까는 니깐 니라 니만차 니만큼 다 다가 다가는
 다간 다고
 다느니 다는 다니 다마는 다마다 다만 다면
 다면서 다시피 다오 단다 담 답시고 더구나 더
 구려 더구면 더군 더냐 더니 더니라 더니마는
 더니만 더라 더라도 더라면 더면 던 던가 던
 걸 던고 던데 던들 던지 데 도 도록 되
 든 든지 듯 듯이 디 리 리거나 리걸
 리게 리까 리까마는 리까말까 리까보냐 리까보다
 리께 리꼬 리는지 리라 리라고 리라치면 리래
 리러니 리러라 리런가 리런고 리레 리레라 리말로
 는 리말로야 리망정 리바에 리바에야 리밖에 리
 뿐더러 리새 리세 리세라 리세말이지 리수룩 리시
 리쏘냐 리이만큼 리작시면 리지 리지나 리지니 리
 지니라 리지라 리지라도 리지어다 리지언정 리진대
 리진랜 리진저 라 라고 라느니 라는 라니
 라니까 라든지 라도 라며 라면 라면서 라서
 라야 라야만

라오	락	란다	람	랍니다	랍시고	랴	라마
는	려	려니	려니라	려니이까	려니이다	려이까	
려이다	런가	런들	려	려거든	려기에	려고	려나
려네	려느냐	려는	려는가	려는고	려는데	려는지	
려니	려니와	려다	려다가	려더니	려더라	려던가	
려도	려며는	려면	려면야	려무나	려서는	려서	
야	려야	려오	런	런다	런마는	런만	렐
렐니까	렐니다	렐다	로고	로구나	로구려	로구	
료	로구면	로구면	로군	로다	로되	로라	로세
리	리카	리니	리니라	리다	리라	리로다	리만치
리만큼	리요	리	리세	리에도	리에라	리직하다	
마	대	며	면	면서	면은	므로	비닐다
비닐다	비디까	비디다	비딘다	비쇼	비시다	비시오	
사오이다	사웁니까	사웁니다	사웁디까	사웁디다	사웁디까	사웁디다	
사외다	세	세요	소	소서	소이다	쇠다	습닐다
습니까	습니다	습디까	습디다	습딘다	시거늘	시거니	
와	시거든	시어요	아	아다	아다가	아도	아라
아서	아야	아야만	아야지	아요	아지이다	왔습니	
다	야	야만	야요	야지	어	어다	어다가
어라	어서	어야	어야만	어요	어지이다	언마는	
언만	언정	엇다	에요	여	여도		
여라	여서	여야	여야만	여요	였습니다	였습니	
다	오	오리카	오리다	오이다	올습니다	올시다	
웁나이다	웁나이다	웁니까	웁니다	웁디까	웁디다		
외다	요	우	으나	으나마	으나	으냐고	으니
으니까	으니까는	으니깐	으니라	으니만치	으니만큼		
으되	으라	으라고					
으라느니	으라니	으라며	으라니까	으라면	으라면		
서	으락	으람	으랴	으라마는	으려	으려	으
려거든	으려기에	으려고	으려나	으려네	으려느냐	으	
려는	으려는가	으려는고	으려는데	으려는지	으려니		
으려니와	으려다	으려다가	으려더니	으려더라	으려		
던가	으려도	으려며는	으려면	으려면야	으려무나		
으려서는	으려서야						
으려야	으려오	으런	으런다	으런마는	으런만	으	
렐	으렐	으렐니까	으렐니다	으렐다	으리	으리카	

으러니	으러니라	으리다	으리라	으리로다	으리만치
으리만큼	으리요	으마	으메	으며	으세요
으소서	으시어요	으오	으오리카	으로리다	으오이다
으옵니까	으옵니다	으옵디까	으옵디다	으외다	으이
은	은가	은질	은고	은대	은들
은즉	은즉슨	은지	은지고	은지라	을
을까	을까마는				을거나
					을걸
을까갈까	을까보냐	을까보다	을까	을까	을는지
라	을라고	을라치면	을락	을랑	을랑은
을러니	을러라	을런가	을런고	을레	을레라
로는	을말로야	을망정	을바에	을바에야	을밖에
을새	을세	을세라	을세말이지	을소냐	을수룩
나	을이만큼	을소냐	을수룩	을쏘냐	을이만큼
면	을지	을지나	을지니	을지니라	을지라
을지이다	을지언정				을지라도
을진디	을진댄	을진저	음	음세	음에도
은넌디	옵니까	옵니다	옵디까	옵디다	옵딘다
쇼	을시다	옵시오	의	이	이거나
와	이거든	이저들랑	이기로	이기에	이나
이니까	이다	이더니	이든가	이든걸	이든고
이든지	이라	이라니	이라니까	이라든지	이라면
이라야	이라야만	이라지	이람	이라	이러니
라	이러니까	이러니이다	이러이까	이런가	이런들
이려니와					
이런마는	이런만	이로고	이로구나	이로구료	이로구
면	이로군	이로다	이로되	이로라	이며
이오	이을시다	이옵나이다	이지	이지마는	인테
인지라	자	자꾸나	자마자	잔	잘
지를	지마는	지만	지요	진	질
					지는

3) 실제 투석 과정에서 이들 목록에 몇몇 형태가 보태어지고, 보태어질 형태 목록

이들 목록 가운데에서 일차적인 검토 대상 목록을 선정하였다. 일차적인 검토 작업은 1992년 한글과 컴퓨터에서 만든 '문자열 검색 · 사전 검색 프로그램(HGREP)'을 가지고 행하기로 하였다. 따라서 일차적인 검토 대상 목록에서 'ㄴ, ㄹ, ㅁ'과 그를 포함한 형태들은 배제된다. '문자열 검색 · 사전 검색 프로그램(HGREP)'이 문자열에 대한 검색 프로그램이어서 음절말의 기음 'ㄴ, ㄹ, ㅁ' 등을 분석하지는 못하기 때문이다.

그에 대한 보완을 위하여 취한 것은 강승식 교수(한성대)가 만든 'HAM(한국어 형태소분석기)'이다. 전체 대상 자료를 HAM으로 돌려 얻은 결과 가운데 'ㄴ, ㄹ, ㅁ'을 가지고 있는 분석 내용을 취하고 그들에 대하여 하나하나 판별하는 과정을 거친 것이다. 처음부터 HAM을 사용하지 않고 일차적인 도구로 사용한 것은 HAM이 모든 분석 가능성을 제시한다는 점에서 많은 노동량을 요구하기 때문이었다. 이와 같은 과정을 통하여 얻어진 결과는 다음 절에서 살펴기로 한다.

2.2 분석의 결과

2.2.1. 선어말어미

아래 보인 <표 4>는 선어말 어미의 출현 빈도와 빈도순위를 보인 것이다. 과거시제형태소인 '었'의 출현빈도가 압도적임을 알 수 있다. 또다른 과거시제형태소인 '였'과 '았'의 빈도를 합할 경우 선어말어미 가운데에서 과거시제 형태가 차지하는 비율은 72%를 조금 넘는다. 선어말어미에 대한 접근의 출발점에 과거시제가 있어야 한다는 점을 강력하게 시사하는 것이라고 이해해도 좋을 것이다. 이와 같은 결과에 대하여 우리가 취한 자료의 성격때문이라는 지적도 가능하다. 일종의 역사자료라 할 수 있는 성경과 지난 일들에 관한 진술을 내용으로 하는 신문과 TV 뉴스 등의 비중이 크다는 점에서 그와 같은 지적은 어느 정도는 타당한 면도 가지고 있다고 할 수 있다. 하지만 다른 측면에서 본다면, 우리 언어 생활 속에서의 과거에 대한 진술이 차지하는 비중이 그만큼 것이라는 이해도 가능하다고 할 것이다. <표 4 참조>

가운데 몇몇은 작업의 편의를 위하여 목록 추가를 보류하고 관련 목록에서 함께 처리하기도 하였다. 물론 함께 처리하는 것이 나으리라는 판단이 가능한 경우에 한한다.

2.2.2 어말어미

다음은 <표 5>의 첫 부분이다. 원래 FE696번까지 계속되는 표가 마련되어 있으나 인쇄 형편상 여기에 가나다 순에 의한 <표 5>의 전부를 실지 않고, 본 원고 끝의 부록 부분에 보일 [부록 1]에서 선어말어미와 어말어미를 적은 빈도순 목록으로 대신한다. 그 이유는 실제로 가나다순보다 빈도순 목록이 더 쓸모가 있기 때문이다.

<표 4>

번호	형태소목록	출현빈도	빈도순
PFE001	거	163	009
PFE002	갸	453	006
PFE003	ㄴ(다)	454	005
PFE004	나이(다)	0	013
PFE005	느(ㄴ/냐/ㄴ지)	127	011
PFE006	느(다)	557	004
PFE007	녀	196	008
PFE008	로(구나/구려)	4	013
PFE009	려	147	010
PFE010	샤오	0	014
PFE011	샤읍	0	014
PFE012	삽	0	014
PFE013	시	352	007
PFE014	왓	1300	003
PFE015	엿	2454	001
PFE016	엿	1498	002
PFE017	오	0	014
PFE018	읍	0	014
PFE019	으시	34	012
PFE020	으오	0	014
PFE021	으읍	0	014
PFE022	자오	0	014
PFE023	자읍	0	014
PFE024	잡	0	014

<표 5>

번호	형태목록	출현 빈도	빈도 순	번호	형태목록	출현 빈도	빈도 순
FE001	거나	54	075	FE041	기에	41	089
FE002	거늘	1	302	FE042	ㄴ	5790	002
FE003	거니	3	241	FE043	ㄴ가	91	054
FE004	거니와	2	264	FE044	ㄴ걸	37	097
FE005	거드면	0	351	FE045	ㄴ고	24	115
FE006	거드면은	0	351	FE046	ㄴ다	454	019
FE007	거든	55	074	FE047	ㄴ다고	0	351
FE008	거들랑	2	264	FE048	ㄴ다는	85	056
FE009	거들랑은	0	351	FE049	ㄴ다느냐	0	351
FE010	거라	38	096	FE050	ㄴ다니까	1	302
FE011	건	13	147	FE051	ㄴ다마는	0	351
FE012	건대	2	264	FE052	ㄴ단다	0	351
FE013	건텐	0	351	FE053	ㄴ달	1	302
FE014	건마는	0	351	FE054	ㄴ담	0	351
FE015	건만	4	220	FE055	ㄴ답니다	0	351
FE016	것다	0	351	FE056	ㄴ답디다	0	351
FE017	게	1288	006	FE057	ㄴ답시고	0	351
FE018	게나	8	176	FE058	ㄴ대	3	241
FE019	게서리	0	351	FE059	ㄴ대서	1	302
FE020	게끔	0	351	FE060	ㄴ대서야	0	351
FE021	고	3843	003	FE061	ㄴ대야	0	351
FE022	고는	118	044	FE062	ㄴ데	177	037
FE023	고도	14	139	FE063	ㄴ들	0	351
FE024	고말고	0	351	FE064	ㄴ만큼	0	351
FE025	고서	61	068	FE065	ㄴ바	7	186
FE026	고야	1	302	FE066	ㄴ죽	1	302
FE027	고자	3	241	FE067	ㄴ죽순	0	351
FE028	곤	13	147	FE068	ㄴ지	72	060
FE029	관테	0	351	FE069	ㄴ지고	0	351
FE030	구나	70	063	FE070	ㄴ지라	0	351
FE031	구려	2	264	FE071	나	180(연	036
FE032	구료	1	302	FE072	나니	29	105
FE033	구먼	3	241	FE073	나마	2	264
FE034	구먼	1	302	FE074	나이까	0	351
FE035	군	117	045	FE075	나이다	0	351
FE036	기	531	016	FE076	냐	156	041
FE037	기로	45	087	FE077	냐고	32	101
FE038	기로서	0	351	FE078	냐는	13	147
FE039	기로서니	0	351	FE079	너라	12	153
FE040	기로선들	0	351	FE080	네	39	094