

북한 고빈도어의 어휘 체계와 양상*

이 성 연
(조선대학교)

Lee, Sungyun. 1999. The North Korean High-frequency Word System and Aspects. *Linguistics*, 7-1, 227-242. This paper is a study of North Korean high-frequency word system and its aspects. In this study, the top 500 North Korean words were selected and analyzed for type and part of speech. The South Korean vocabulary was compared with the North Korean vocabulary for basic words, distribution gap features, and linguistic features. This study will help the understanding of the North Korean language and will be useful source following the unification of Korea.(Chosun University)

1. 서론

어휘 체계와 양상은 비교적 제한된 범주를 다루는 음운과 통사 체계와는 달리, 언어의 역사와 사회 및 문화 전반과 연관이 있으므로 어휘론은 다양한 연구 영역을 가진다. 어휘론의 하위 영역에 속하는 '어휘 계량'은 어휘소들의 분포, 사용률 같은 수량적인 면을 주 연구 대상으로 삼고 있으므로 어휘 빈도의 조사 분석은 빠뜨릴 수 없는 연구 항목이 된다.

어휘 빈도의 조사 결과인 '어휘 빈도수'는 기초어휘, 기본어휘, 이해어휘, 사용어휘 등의 양상을 보여주는 실용적인 통계의 성격을 지니기 때문에 실제적인 언어 생활의 모습을 파악하는 데 큰 도움을 준다. 또한 언어 표현은 사물 지시적 의미만 지니는 것이 아니라 사용되는 사회 환경에 따라 여러 가지 사회적인 의미를 지니며, 의미도 어떤 단어가 내재적으로 가지고 있지 않고 사회적 환경의 쓰임에 따라 결정되므로, 어휘 빈도수를 통한 언어 사회의 특징 유추도 가능하게 된다.

본고는 북한의 고빈도어(高頻度語)²의 어휘 체계와 양상에 대한 고찰을

* 이 논문은 1997년도 조선대학교 학술연구비의 지원을 받아 연구되었음

1. 근래의 사회언어학적 입장에 의한 것임

2. 북한의 「조선어 빈도수 사전」의 자료에 의거함

함에 있어서 상위순 500개의 고빈도어를 일차적으로 택하여 어종과 품사별 분류를 하겠다. 그리고 남한 기초 어휘와의 관계 및 분포 간극의 특징, 북한 언어 사회의 특징 등의 순서로 살펴 보겠다. 이러한 연구는 북한의 언어 생활에 대한 이해를 도우며 통일 후 국어 문제 해결을 위한 자료로서도 활용될 수 있을 것이다.

2. 빈도수 조사의 방법

본고의 일차적인 분석 대상인 북한 고빈도어 500개는 북한의 『조선어 빈도수 사전』에서 택했는데, 그 사전에서는 어휘 부분에 대한 본문 선택의 원칙을 다음과 같이 제시하고 있다.

① 어휘를 선정하기 위한 본문은 “문학예술”, “사회정치”, “신문보도”, “과학기술”로 대별한 후 ‘분과 주제’, ‘장르’, ‘시기’와 ‘항목’별로 확정한다.

② 구체적인 본문의 양은 현대 출판물의 발행, 배포 및 보존의 비율에 따라 분과별, 주제별, 시기별, 장르별로 확정한다. 그러나 어휘 사용과 기능의 특성을 고려하여 “문학예술” 분야는 44%, “사회정치” 분야는 25%, “신문보도” 분야는 20%, “과학기술” 분야는 11%의 비율³⁾로 본문을 선택한다.

③ 본문의 선택은 도서, 신문, 잡지, 교과서, 각종 문건 등을 기본으로 하고 일부 본문의 특성에 따라 필사본을 이용한다.

④ 어종은 순조선말, 조선글로 된 것만으로 한한다.

⑤ 각종 사전류들은 본문 대상에서 제외한다.

그리고, 통계 분석을 위해서는 ‘다단계 표본 선택법’을 적용하고 있다. 즉 본문을 부문별로 나누고 그것을 다시 문체로 주제와 장르, 시기와 항목별로 단계적으로 세분화하는 방법을 택한 것이다. 다음과 같이 정리할 수 있다.

1) 문학예술본문

① 소설 : 장편소설, 중편소설, 단편소설, 고전소설, 번역소설

② 희곡 : 장·중편, 단막극, 소품 ③ 시 : 서사시, 서정시, 동요, 동시

3. 어휘 빈도수 조사를 위하여 결정된 본문량의 부문별 묶은, 문학예술본문 441,229개, 사회정치본문 275,710개 신문보도본문 225,285개 과학기술본문 105,152개임

2) 사회정치본문

- ① 노작 ② 문헌 ③ 사상해설 ④ 혁명역사 ⑤ 회상기 ⑥ 덕성실기

3) 신문보도본문

- ① 신문 - 노동신문, 민주조선, 노동청년 ② 보도 - 조선중앙통신

4) 과학기술문

- ① 사회과학 - 교과서, ② 자연과학 - 잡지, ③ 기술부문 - 단행본

다음에, 평가 방법으로는 빈도 크기순 목록에서 그 상대 빈도수 값이 0.00019보다 작지 않은 부분의 절대 빈도수들의 합이 분석 대상으로 취한 본문량의 75% 이상을 차지하는 경우를 신뢰성을 가진 것으로 인정했다.

빈도수 조사 본문으로 선택된 주요 문헌과 작품의 연대는 사회정치 부문 1946~1988년, 문학예술 부문 1955~1987년, 신문보도 부문 1965~1987년, 과학기술 부문 1980년 초중반이다. 문헌과 작품은 다음과 같다.

1) 사회정치 부문

- ① 김일성 발표 문헌(12편) ② 김정일 발표 문헌(5편)
③ 혁명전통 교양 자료(7편)

2) 문학예술 부문

- ① 소설(장편소설 : 「피바다」 등 26편, 중편소설 : 「전사들」 등 6편,
단편소설 : 「평양행 급행열차」 등 8편)
② 희곡(재담 9편, 만담 2편, 벽소설(1편)
③ 시(서사시·아동시·동시·동시극·동요·동요극 각1편, 동화 4편, 동화극 1편
④ 민족고전(「춘향전」, 「피설집」) ⑤ 번역소설(7편)

3) 신문보도 부문(「노동신문」, 「민주조선」, 「노동청년」)

4) 과학기술 부문

- ① 사회과학 논문(논문집 3편)
② 기초과학 : 교과서 14종(고등중학교 3~5학년용 일부, 공업대학용 일부, 대학 예비과용 일부, 대학용 일부) 잡지 2종(「정자공학」, 「체육 과학」)

3. 고빈도어의 어휘 체계

빈도 크기에 따른 고빈도어(高頻度語) 500개를 빈도순대로 먼저 제시한

다음에 이들의 어휘 체계에 대해서 살펴 보기로 한다.⁴

1. 하다(46612) 2. 있다(26562) 3. 것(18118) 4. 되다(10816) 5. -적(8682) 6. 이(-)(7783). 7. 우리(7766) 8. 같다(6324) 9. 없다(5645) 10. 주다(5135) 11. 가다(4769) 12. 그(4549) 13. 인민(4177) 14. 위하다(3965) 15. 사랍(3823) 16. 말(언어)(3776) 17. 당(3671) 18. 한(3514) 19. 혁명(3350) 20. 그러다(3317) 21. 나라(3194) 22. 수(합~있다)(3137) 23. 투쟁(3106) 24. 아니다(2908) 25. 대하다(2741) 26. 나(2550) 27. 지다(그늘~)(2470) 28. 때(시간)(2371) 29. 잘(2353) 30. 우(上)(2248) 31. 가지다(2161) 32. 오다(2083) 33. 알다(2033) 34. 크다(1923) 35. 일(1893) 36. 자기(1834) 37. 나다(1794) 38. 앞(1781) 39. 힘(1724) 40. 눈(眼)(1715) 41. 다(1707) 42. 생활(1706) 43. 내다(1682) 44. 늙(1653) 45. 놓다(1652) 46. 더(1646) 47. 두(1613) 48. 건설(1610) 49. 중요하다(1606) 50. 동무(1598) 51. 둘다(당에~)(1597) 52. 나가다(1565) 53. 집(1558) 54. 못(~마땅하다)(1516) 55. 어쩔다(1503) 56. 나오다(1491) 57. 많다(1472) 58. 다음(1437) 59. 내(1424) 60. 지금(1424) 61. 때문에(1409) 62. 앓다(1402) 63. 좋다(1390) 64. 길(1383) 65. 제(~집)(1383) 66. 같이(1365) 67. 그러나(1358) 68. 문제(1356) 69. 통일(1337) 70. 동지(1332) 71. 서다(1332) 72. 또(1330) 73. 사상(1314) 74. 시키다(1313) 75. 누구(1282) 76. 오늘(1281) 77. 여러(1277) 78. 전(1269) 79. 따르다(1268) 80. 그렇다(1264) 81. 이렇다(1260) 82. 다르다(1254) 83. 모든(1251) 84. 위대하다(1248) 85. 요구(1235) 86. 치다(1229) 87. 마음(1227) 88. 날(1221) 89. 모르다(1199) 90. 다시(1190) 91. 더욱(1175) 92. 말다(그만두다)(1161) 93. 적(1144) 94. 그것(1111) 95. 밤(1100) 96. 속(1082) 97. 끝(1072) 98. 운동(1059) 99. 말하다(1058) 100. 잡다(1054) 101. 의하다(1052) 102. 무엇(1040) 103. 받다(1038) 104. 머리(1026) 105. 민족(1024) 106. 곳(1013) 107. 뒤(996) 108. 찾다(993) 109. 못하다(984) 110. 안(內)(974) 111. 만들다(973) 112. 높다(971) 113. 사이(960) 114. 싹다(952) 115. 청년(945) 116. 기술(937) 117. 조국(936) 118. 나무(927) 119. 맞다(917) 120. 세계(911) 121. 지나다(907) 122. 나서다(905) 123. 먹다(898) 124. 뽀(893) 125. 수령(890) 126. 전투(890) 127. 벌리다(879) 128. 내리다(871) 129. 시작(871) 130. 평화(865) 131. 경제(864) 132. 함께(864) 133. 사업(859) 134. 진행(859) 135. 입다(855) 136. 전사(戰士)(850) 137. 밀(831) 138. 주제(826) 139. 강화(819) 140. 보다(816) 141. 새롭다(812) 142. 이야기(810) 143. 자리(808) 144. 옛날(803) 145. 하나(799) 146. 방법(798) 147. 생산(798) 148. 일군(795) 149. 공장(794) 150. 앓다(792) 151. 여기(792) 152. 물(791) 153. 쓰다(用)(791) 154. 필요하다(788) 155. 동안(781) 156. 들어가다(778) 157. 얼굴(776) 158. 혁명적(774) 159. 번(771) 160. 등(等)(767) 161. 오르다(763) 162. 정치(762) 163. 쪽(760) 164. 몇(755) 165. 어머니(754) 166. 비롯하다(751) 167. 온(751) 168. 몸(747) 169. 땅(747) 170. 계속(746) 171. 가슴(740) 172. 또한(738) 173. 그러하다(734) 174. 철저하(721) 175. 들리다(718) 176. 빛(718) 177. 깊다(716) 178. 죽다(712) 179. 전쟁(705) 180. 공산주의(702) 181. 승리(702) 182. 떠나다(702) 183. 실현(695) 184. 조적(684) 185. 발전(679) 186. 한참(665) 187. 앓다(660) 189. 살다(658) 190. 처음(655) 191. 노동자(643) 192. 싸우다(643) 193. 높다(638) 194. 모 두(638) 195. 녀자(631) 196. 풀다(629) 197. 특히(626) 198. 안(~가다)(622) 199. 두다(621) 200. 어디(618) 201. 짓다(616) 202. 지도(615) 203. 필요(615) 204. 가지(여러~)(613) 205. 대로(명령~)(612) 206. 사회(607) 207. 시간(606) 208. 흔흔히(606) 209. 바람(602) 210. 느끼다(601) 211. 이리하다(601) 212. 이루다(601) 213. 줄(울~민다)(600) 214. 개(個)(599) 215. 들

4. 어휘 앞의 숫자는 빈도순, 괄호 안의 숫자는 빈도수이다.

다(596) 216. 만나다(594) 217. 전체(594) 218. 조건(594) 219. 보내다(593) 220. 부르다(593)
 221. 자신(593) 222. 관계(590) 223. 얼마(589) 224. 모양(584) 225. 표시(583) 226. 줌(578)
 227. 창조(578) 228. 빨리(578) 229. 학생(571) 230. 그리다(569) 231. 대중(569) 232. 진실
 (569) 233. 준비(564) 234. 부문(563) 235. 역할(563) 236. 성과(562) 237. 공업(561) 238. 생
 기다(560) 239. 국가(560) 240. 년(559) 241. 읊다(559) 242. 칫(559) 243. 통하다(558) 244.
 적극(550) 245. 인간(549) 246. 기본(545) 247. 아직(545) 248. 활동(543) 249. 아버지(541)
 250. 보이다(540) 251. 작다(539) 252. 바라보다(537) 253. 끊임없다(534) 254. 기계(533) 255.
 정신(532) 256. 목소리(529) 257. 수(528) 258. 기초(521) 259. 수행(521) 260. 옷다(521)
 261. 아이(520) 262. 올리다(519) 263. 사랑(518) 264. 시기(時)(518) 265. 어쩌하다(518) 266.
 당원(514) 267. 이것(513) 268. 걸음(512) 269. 이제(512) 270. 키우다(510) 271. 이룩하다
 (510) 272. 늙다(509) 273. 이러하다(508) 274. 교육(499) 275. 과정(499) 276. 꽃(499) 277.
 노래(497) 278. 밤(497) 279. 많이(494) 280. 공화국(492) 281. 작업(492) 282. 다니다(490)
 283. 커다랗다(489) 284. 들다(487) 285. 해결(487) 286. 가장(484) 287. 너(483) 288. 과학
 (482) 289. 봄(482) 290. 정책(480) 291. 사회주의(477) 292. 연구(475) 293. 중(中)(474) 294.
 계획(472) 295. 마을(470) 296. 묻다(468) 297. 주체사상(468) 298. 일으키다(467) 299. 들리
 다(465) 300. 참가(465) 301. 언제나(465) 302. 언다(464) 303. 끝다(460) 304. 마련(459) 305.
 잠(457) 306. 올리다(453) 307. 설비(452) 308. 불(451) 309. 방(448) 310. 사실(448) 311. 서
 로(447) 312. 책등(447) 313. 새(446) 314. 뛰(445) 315. 아침(445) 316. 그러하다(444) 317.
 이름(443) 318. 나타나다(442) 319. 들다(442) 320. 중심(442) 321. 과업(441) 322. 주인(436)
 323. 낡다(435) 324. 네(433) 325. 대다(431) 326. 농민(430) 327. 달리다(430) 328. 지적(428)
 329. 충실하다(428) 330. 안다(427) 331. 쓰다(425) 332. 지키다(424) 333. 혁신(424) 334. 만
 일(424) 335. 쥐다(421) 336. 먼저(420) 337. 자주(自主)(419) 338. 들어오다(417) 339. 편지
 (416) 340. 학교(415) 341. 하여(413) 342. 제대(411) 343. 빛나다(407) 344. 어렵다(406)
 345. 제기(402) 346. 가운데(399) 347. 그레(399) 348. 뜻(399) 349. 해방(398) 350. 직장(396)
 351. 강(394) 352. 꾸리다(394) 353. 떨어지다(394) 354. 구체적(393) 355. 녀인(393) 356. 로
 동(392) 357. 읊다(392) 358. 저녁(391) 359. 그러므로(390) 360. 현실(390) 361. 속도(389)
 362. 책임(384) 363. 제(384) 364. 이르다(至)(384) 365. 보다(一률통한)(383) 366. 바로(380)
 367. 잠깐(375) 368. 군중(374) 369. 예술(374) 370. 순간(372) 371. 퍼(372) 372. 회(372) 373.
 말(馬)(371) 374. 자연(371) 375. 경우(370) 376. 남다(368) 377. 직결(367) 378. 대답(366)
 379. 열(365) 380. 버리다(364) 381. 인민대중(364) 382. 세우다(363) 383. 자주적(362) 384.
 바치다(360) 385. 결정(359) 386. 들어서다(359) 387. 침략(359) 388. 어느(359) 389. 기다리다
 (358) 390. 당조직(358) 391. 당하다(358) 392. 회의(會議)(357) 393. 이미(357) 394. 모습(356)
 395. 아들(356) 396. 향하다(355) 397. 오래(355) 398. 충실성(354) 399. 혁명투쟁(354) 400. 메
 우(353) 401. 중앙(353) 402. 거기(352) 403. 놀라다(352) 404. 내용(352) 405. 낡다(351) 406.
 단결(349) 407. 형식(349) 408. 남(348) 409. 아무(348) 410. 그대르(345) 411. 타다(乘)(345)
 412. 해(344) 413. 인민경제(341) 414. 아무리하다(340) 415. 안례(339) 416. 하루(338) 417.
 흐르다(337) 418. 읊기다(336) 419. 위원회(336) 420. 군사(335) 421. 기간(335) 422. 영웅
 (335) 423. 왜(335) 424. 아름답다(334) 425. 일(기간)(333) 426. 기관(332) 427. 깊이(332)
 428. 꼭(332) 429. 관리(331) 430. 돌아오다(330) 431. 데(330) 432. 원칙(330) 433. 중앙위원
 회(329) 434. 잠(327) 435. 잠시(327) 436. 어리다(327) 437. 농촌(326) 438. 창건(324) 439.
 딸(323) 440. 아래(323) 441. 법(322) 442. 지역(322) 443. 반대(320) 443. 자(者)(320) 444.
 전하다(320) 445. 형상(320) 446. 일어나다(320) 447. 넘다(319) 448. 문화(319) 449. 사건

(319) 450. 당신(318) 451. 상태(318) 452. 신문(318) 453. 끊임없이(318) 454. 돌아가다(317)
 455. 평화적(317) 456. 동원(317) 457. 기업소(315) 458. 민주주의(315) 459. 작용(315) 460.
 임무(315) 461. 전선(314) 342. 현상(314) 463. 후(314) 464. 전기(313) 465. 꺼내다(312)
 466. 길다(311) 467. 밋(310) 468. 공사(309) 469. 정말(309) 470. 뜨겁다(309) 471. 발표(308)
 472. 밝히다(308) 473. 중오(307) 474. 율예(306) 475. 관철(305) 476. 높이(305) 477. 뛰다
 (304) 478. 원수(304) 479. 시대(303) 480. 없이(303) 481. 옷(303) 482. 습(302) 483. 거리
 (300) 484. 결과(300) 485. 불다(299) 486. 쓰다(299) 487. 놀(298) 488. 특성(298) 489. 예
 (298) 490. 움직이다(297) 491. 내놓다(296) 492. 보장(296) 493. 제일(296) 494. 차(296)
 495. 열다(296) 496. 곧다(296) 497. 멀다(295) 498. 기술자(294) 499. 자다(294) 500. 자체
 (294)

3.1. 어종별(語種別) 분류

북한의 고빈도어 500개를 어종별(語種別)로 분류하면 고유어 314개, 한자어 175개이며 나머지 11개는 ‘한자어+고유어’(필요+하다, 충실+하다, 향+하다 등)의 결합 구조를 지닌 단어들이다. 따라서 어종별 분포도는 고유어 62.8%, 한자어 35%, 기타(‘한자어+고유어’)2.2%이다. 들어 있지 않았다.

북한에서는 ‘어휘정리’와 ‘단어만들기’ 사업⁵에 의해 신조어(新造語)인 고유어를 많이 만들어 사용하게 했지만 상위 고빈도어 500개 내에는 하나도 포함되어 있지 않는 것으로 나타났다. 외래어도 마찬가지다.

북한의 ‘어휘정리’는 어휘 다듬기의 형태로 전개되었는데 한자어를 ‘바꾸고치기’(하복>여름옷), ‘살려 고치기’(부들기>버금도드리), ‘만들어 고치기’(늑지>꽃잔디밭) 등의 조어법을 통해 고유어로 다듬었다(이성연 1996). ‘단어만들기’는 고유어의 말뿌리를 실마리로 하여, 우리말의 구조 특성을 살려 의미의 명확성과 대중들의 사상, 감정, 정서, 미감 등을 반영할 수 있는 방향으로 새로운 말을 만들었던 것이다.⁶

그런데 이들 신조어들이 상위 고빈도어에 들어 있지 않은 것을 보면 사용 어휘⁷의 지위를 획득하지 못한 것으로 판단된다.⁸ 이것은 언중들의 자연

5. 김일성의 “조선어의 민족적 특성을 옮겨 살려 나갈 데에 즈음하여”라는 제목의 담화문(1966년)에 의해 ‘단어만들기’ 사업이 전개되었다. 단어의 특성에 따라 ‘일반어 만들기’와 ‘학술용어 만들기’로 나누어 시행되었는데, 한자어를 고유어로 바꾸고 새로 온 고유어를 만들어 냈던 것이다(리익선 1974).

6. “우등불모임, 씩떨구다, 날개바퀴, 비비송곳, 물주머니, 나무라함, 왓새마을, 노루묵, 되살이, 헛웃음증, 엇가기”등의 조어들을 예로 들 수 있다.

7. 기본 어휘는 두 가지 개념으로 쓰인다. 첫째는 포괄적인 개념으로서 작품이나 한정된 언어 사용 국면에서 몇 개인가의 총을 형성하고 있는 것을 전제로 하여, 여러 층에 걸쳐서 공통적으로

스런 언어 사용의 흐름에 의하지 않고 정책 접근법에 의한 언어정책상 만들어진 신조어이기 때문에 해석된다.

3.2. 품사별 분류

상위 고빈도어 500개를 품사별로 분류하면 명사 258개, 동사 118개, 형용사 41개, 부사 42개, 대명사 16개, 관형사 11개, 의존명사 7개, 수사 3개, 조사 2개, 감탄사 1개의 순서로 나타난다. 품사통용의 단어 7개('지금, 이제, 서로, 만일, 먼저, 깊이, 높이')는 예문이 없기 때문에 명사인지 부사인지 식별하기가 어렵지만 '깊이'와 '높이'는 명사로 나머지는 부사로 분류했다.⁹

따라서 명사 51.6%, 동사 23.6%, 형용사 8%, 부사 8.4%, 수사 0.6%, 조사 0.4%, 감탄사 0.2%의 분포를 보이고 있다. 이는 언어 생활에서 명사와 동사가 차지하는 비중이 높음을 보여주는 예라고 하겠다.

접미사가 1개('적') 있는데 접미사를 독립 단어로 인정하여 고빈도어 500개 안에 포함시킨 것은 매우 특이한 빈도 조사 방법이라고 하겠다. 접속부사나 의성부사, 의태부사 등은 포함하지 않았다.

4. 고빈도어의 어휘 양상

4.1. 남한 기초 어휘 목록과의 비교

기초 어휘란 '우리가 생활해 나가는 데 있어서 이것만은 최소한도로 필요하고 그 이상은 여분의 것이라고 생각해도 좋은 비교적 소수의 낱말'(신익성 1990), 혹은 '언어생활에서 빈도수가 높고 분포가 넓으며, 파생이나 합성 등 2차 조어의 근간이 되는 최소한의 필수어'(임지룡 1991)를 말한다. C. K. Ogden 등이 국제적으로 사용되기 위한 제2언어(국제 보조어)로서 도움을 주기 위한 목적으로 선정한 기초 영어(The System of Basic

출현하는 어휘소의 집합을 가리킨다. 잡지를 예로 들면, 실용기사, 문예작품, 취미 등 계되는 내용별로 층을 형성하는 것이다. 다음에, 종합적 개념으로서 언어 사용 영역의 전개를 위해 가장 기본이 되는 어휘의 집합을 가리킨다. 따라서 특정한 목적, 특정한 분야를 위한 '생활 기본 어휘', '학습 기본 어휘'처럼 분야별 기본 어휘 개념으로 사용되는 것이다(김광해 1993).

8. 고유어 신조어는 상위 고빈도어 1000개 안에 들어 있지 않다.

9. '깊이'와 '높이'가 남한의 기초 어휘에서는 명사로 분류되었기 때문이다.

English)가 대표적인 예에 해당한다(김광해 1993).¹⁰

북한의 기초 어휘와 남한의 기초 어휘 비교는 양쪽 언어 사회의 모습과 특징을 파악하는 데 도움을 준다. 그런데 남한에서는 기초 어휘 목록에 관한 연구가 있지만 북한에서 작성된 기초 어휘의 목록을 찾을 수가 없기 때문에 북한의 고빈도어와 남한의 기초 어휘 목록을 통해 살펴 보는 간접적인 방법을 택하기로 한다.

남한의 고빈도어와 기초 어휘의 관계에 대한 연구 또한 아직 없기 때문에 북한의 고빈도어와 남한의 기초 어휘에 대한 직접 비교는 그 상관 관계에 상대적인 한계가 있다. 그러나 기초 어휘는 빈도수가 높다는 점에 의거하여 북한의 고빈도어와 남한의 기초 어휘와의 관계를 살펴 보는 것이다. 이를 통해 북한의 기초 어휘와 언어 사회의 특징을 유추해 볼 수 있으며 나아가서 남한 언어 사회와의 비교도 가능할 것이다.

남한에서는 임지룡(1991)이 객관적 방법에 의해 선정한 고빈도어와 주관적 방법에 의한 의미분야를 절충하여 내용어 1,500개의 기초 어휘 목록을 작성했는데, 북한의 고빈도어 500개 중에서 이 목록과 일치하는 어휘를 비교하기로 한다. 간접적인 방법이지만 북한의 고빈도어와 우리 나라 기초 어휘와의 관계를 알아 볼 수 있는 지표가 될 수 있을 것이다. 그리고 이 지표는 통일 후의 어휘 통일에 대한 국어정책에도 필요한 자료로서 활용될 수 있을 것이다.

남한의 기초 어휘 목록과 일치하는 북한 고빈도어 현황¹¹

1. 사람에 관한 어휘(201)

1. 인체(68)

① 기관(42) : 눈, 머리, 몸, 가슴, 얼굴<5> ② 생리, 질병(26) : 잠, 숨<2>

2. 정신(25) : 사상, 마음, 정신<3>

3. 부류(101)

① 사람 일반(32) : 사람, 청년, 여자, 아이<4>

10. 기초 어휘 목록으로는 영어를 배우는 외국인을 위하여 L. Thorndike 등이 마련한 2,000어휘 표가 있다. 그리고 일본 국제문화진흥회에서 외국인의 일본어 교육에 도움을 주기 위해 선정한 일본어 기본 어휘 2,000어가 있다.

11. ()안의 숫자는 남한의 해당 어휘수이고, < >안의 숫자는 북한의 해당 어휘수이다.

② 친척(34) : 어머니, 아버지, 아들, 딸<4>

③ 직업(34) : 학생, 기술자<2>

4. 기타(7) : 이름<1>

II. 의식주에 관한 것(159)

1. 의생활(21) : 옷<1>

2. 식생활(57) : 없음

3. 주생활(20) : 집, 방<2>

4. 생필품(61) : 없음

III. 사회생활에 관한 어휘(160)

1. 사회조직(30) : 나라, 민족, 세계, 사회, 국가, 마을<5>

2. 제도, 관습(28) : 정치, 전쟁, 법<3>

3. 교통, 통신(30)

① 교통(17) : 길, 차<2>

② 통신(13) : 신문<1>

4. 공공시설(21) : 없음

5. 경제분야(51)

① 경제일반(12) : 일, 사업, 생산<3> ② 상업(17) : 경제<1>

③ 어업(9) : 없음 ④ 농업(6) : 없음 ⑥ 공업(6) : 기술, 공장, 기계<3>

IV. 교육 및 예체능에 관한 어휘(150)

1. 교육일반(39) : 문제, 방법, 과학, 연구, 학교<5>

2. 언어(36) : 말, 이야기<2> 3. 문학(9) : 없음

4. 체육, 오락(37) : 운동<1> 5. 음악(21) : 노래<1>

6. 미술(8) : 없음

V. 자연계에 관한 어휘(165)

1. 천체(6) : 없음 2. 지리, 지형(28) : 땅, 강<2>

3. 자연현상(34) : 물, 빛, 바람, 불, 전기<5> 4. 동물(59) : 말<1>

5. 식물(28) : 나무, 꽃<2> 6. 광물(6) : 없음

VI. 감각 및 인식에 관한 어휘(165)

1. 일반부류(17) : 관계, 시간, 높이<3>

2. 공간(36) : 우, 앞, 밖, 뒤, 안, 사이, 밑, 모양, 가운데, 옆, 아래<11>

3. 시간(34) : 때, 시간, 오늘, 시간, 옛날, 밤, 아침, 저녁, 해, 하루<10>

4. 수량(57)

① 수량(37) : 하나<1> ② 수량단위(20) : 개<1>

5. 추상(21) : 평화, 사랑, 문화, 꿈<4>

VII. 동작에 관한 어휘(250)

하다, 되다, 주다, 가다, 위하다, 대하다, 지다, 가지다, 오다, 나다, 내다, 놓

다, 들다(入), 서다, 시키다, 따르다, 치다, 잡다, 받다, 찾다, 만들다, 맞다, 먹다, 내리다, 입다, 보다, 앉다, 쓰다(用), 오르다, 죽다, 살다, 싸우다, 풀다, 두다, 짓다, 느끼다, 이루다, 듣다, 만나다, 보내다, 부르다, 그리다, 웃다, 다니다, 돌다, 묻다(問), 얻다, 끝다, 그리다, 나타나다, 들다, 놓다, 달리다, 안다, 지키다, 쥐다, 울다, 남다, 버리다, 기다리다, 향하다, 놀라다, 타다, 흐르다, 옮기다, 전하다, 넘다, 뛰다, 붙다, 움직이다, 열다, 자다<72>

VIII. 상태에 관한 어휘(150)

있다, 갈다, 아니다, 알다, 크다, 중요하다, 많다, 좋다, 높다, 새롭다, 필요하다, 깊다, 작다, 어렵다, 젊다, 아름답다, 어리다, 길다, 뜨겁다, 굳다, 멀다<21>

IX. 기타(100)

1. 대명사(15) : 우리, 나, 누구, 무엇, 어디, 자신, 너, 당신<8>
2. 의존명사(10) : 것, 수, 곳, 뿐, 줄, 데<5>
3. 부사(49) : 잘, 다, 더, 못, 또, 다시, 함께, 모두, 안, 좀, 가장, 서로, 먼저, 바로, 이미, 매우, 왜, 꼭, 늘<19>
4. 보조동사, 형용사(9) : 말다, 싶다, 앓다<3>
5. 관형사(17) : 이, 그, 저, 여러, 모든, 몇, 온, 새, 어느<9>

이상의 일치 어휘 현황을 보면, 북한의 고빈도어 500개 중 46%인 230개가 남한의 기초 어휘 1,500개 내에 포함된다. 대분류 항목의 어휘별로는 '사람'은 23개로서 약 1.4%, '의식주'는 3개로서 약 2%, '사회생활'은 18개로서 약 11.2%, '교육 및 예체능'은 9개로서 6%, '자연계'는 10개로서 약 6%, '감각 및 인식'은 30개로서 약 18%, '동작'은 72개로서 28.8%, '상태'는 21개로서 약 14%, '기타'는 44개로서 약 44%로 나타난다.

남북 분단의 역사가 반세기를 넘어서고 그 동안 언어 이질화의 골이 깊어졌다고는 하지만 기초 어휘와의 일치율이 위험 수위는 아닌 것으로 해석된다. 그러나 특정 항목의 어휘는 일치율이 매우 낮게 나타나는 데 이것은 정치·사회체제의 차이에 의한 것이다. 통일 후 언중의 언어 사용 흐름이 자연스럽게 조화를 이룰 것이므로 극복 방안 마련이 크게 어렵지는 않을 것이다.

남한의 기초 어휘 목록이 1,500개이므로 북한의 고빈도어를 1,500개로 확대한다면 일치율은 어떻게 나타날 것인가¹²⁾를 조사해 보았다. 그 결과 약

12. 나머지 고빈도어 1,000개의 예시는 생략했음.

31%인 466개가 남한의 기초 어휘와 일치율을 보였다.

분류 항목별로는 '사람' 56개(27.8%), '의식주' 8개(5%), '사회생활' 28개(17.5%), '교육 및 예체능' 19개(12.6%), '자연계' 29개(17.5%), '감각 및 인식' 50개(30.3%), '동작'은 163개(65.2%), '상태'는 49개(32.6%), '기타'는 64개(64%)로 나타난다.

이에 대한 좀더 구체적인 분석은 고빈도어 500개 단위로 간극을 설정하여 다음 4.2.에서 하겠다.

4.2. 분포 간극의 특징

북한 고빈도어 1,500개와 남한 기초 어휘 목록과의 일치율을 500개 단위의 분포 간극으로 나타내면 <표 2>와 같다.

<표 2> 북한 고빈도어 1,500개의 분포현황13

대분류	소분류	1-500	500-1,000	1,000-1,500	계	%	대분류	소분류	1-500	500-1,000	1001-1,500	계	%
사 랑	기관(42)	5	8	4	17	40.0	예 체 능	문학(9)	0	1	0	1	11
	생리, 질병(26)	2	1	0	3	11.5		체육, 오락(37)	1	0	0	1	2.7
	정신(25)	3	5	1	9	36		음악(21)	1	0	0	1	4.7
	사람일반(32)	4	1	1	6	18.7	자 연 계	미술(8)	0	1	0	1	12.5
	친척(34)	4	5	2	11	32.3		천체(6)	0	1	0	1	16.6
	직업(34)	2	2	1	5	14.7		지리, 지형(28)	2	1	5	8	28.5
	기타(7)	1	4	0	5	71.4		자연현상(34)	5	7	2	14	41.1
의 식 주	의생활(21)	1	0	0	1	4.7		동물(59)	1	0	1	2	3.3
	식생활(57)	0	1	0	1	1.7		식물(28)	2	1	1	4	14.2
	주생활(20)	2	0	0	2	10		광물(6)	0	0	0	0	0
	생활품(61)	0	3	1	4	6.5	감 각 및 인 식	일반부류(17)	3	2	2	7	41.1
사 회 생 활	사회조직(30)	5	2	3	10	33.3		공간(36)	1	3	3	6	16.6
	제도, 관습(28)	3	0	1	4	14.2		시간(34)	10	4	2	16	47.0
	교통(17)	2	0	1	3	17.6		수량(37)	1	2	3	6	16.2
	통신(13)	1	0	1	2	15.3		수량단위(20)	1	1	1	3	15.0
	공공시설(21)	0	1	0	1	4.7		추상(21)	4	7	1	12	57.1
	경제일반(12)	3	0	0	3	25	동 작 상 태	동작(250)	72	42	29	163	65.2
농업(6)	0	0	1	1	16.6	상태(150)		21	15	13	49	32.6	
사 회 생 활	어업(9)	0	0	0	0	0	기 타	대명사(15)	8	0	0	8	53.3
	상업(17)	1	0	0	1	5.8		의존명사(10)	5	0	0	5	50
	공업(6)	3	0	0	3	50		부사(49)	19	13	3	35	71.4
	교 육	교육일반(39)	5	3	2	10		25.6	관형사(17)	9	4	0	13
언어(36)		2	3	0	5	13.8	보조용언(9)	3	0	0	3	33.3	

먼저 분포 간극의 주요 특징을 보면 남한의 기초 어휘를 기준으로 한 북한 고빈도어의 어휘의 분포는 500개, 1,000개, 1,500개 단위의 간극에서 어느 정도 비례 관계가 있음을 알 수 있다. 1~500의 간극에서 분포 합계가

13. ()안의 숫자는 남한의 기초 어휘수이다.

0으로 나타나는 기초 어휘 등 극히 소수를 제외하고는 1,000개와 1,500개의 간극에서도 비슷한 비례 관계가 있다는 것이다.

이는 두 가지의 해석을 가능하게 한다. 첫째는 비록 남한의 기초 어휘를 기준으로 한 것이지만 북한 고어빈도어는 남한의 기초 어휘와 일정한 함수 관계를 가지고 있으며, 통일 후의 국어생활에서 큰 혼란은 피할 수 있다는 점을 시사해 준다는 점이다. 둘째는 북한 고어빈도어의 추출 방법이 의도적으로 사회체제에 맞춘 것이 아니라 나름대로 객관성을 지니고 이루어졌다는 점이다.

다음에 간극 분포의 세부적인 특징을 살펴 보면 다음과 같다.

첫째, 대분류 어휘를 기준으로 '기타'와 '동작'과 관련된 어휘의 일치율이 60%가 넘어서 매우 높게 나타나고, '감각 및 인식'과 '상태'와 관련된 어휘가 30%를 넘어 보통으로 나타난다. 반면에 '의식주'와 관련된 어휘의 일치율은 10% 미만으로서 매우 낮게 나타난다.

둘째, 소분류 항목별 어휘의 경우 일치율 70% 이상은 '관형사', '부사', '기타(사람)'이며, 60% 이상은 '동작', 50% 이상은 '공업', '추상', '대명사', '의존명사'이다. 일치율 50%가 넘는 어휘들은 대부분 일상생활과 관련된 기초 어휘들인 것으로 나타났다. 또한 일치율 30%~49%인 '기관', '정신', '친척', '사회조직', '일반부류(감각 및 인식)', '시간', '상태', '보조용언' 등도 마찬가지이다. 그리고 이들 어휘의 분포 간극을 보더라도 남한과 정치·사회체제가 다르지만 남한 기초 어휘와의 일치율이 비교적 고르게 나타난다. 이는 통일 후의 국어 문제 해결의 어려움이 상대적으로 낮을 수 있음을 시사해 주는 지표라고 하겠다.

셋째, 일치율 20% 미만의 경우는 주된 이유가 남한과 사회 체제와 정치 이데올로기의 차이에 의한 것으로 해석된다. 그리고 상대적으로 분포 간극의 기복이 심하게 나타난다. 특히 외식주는 소항목 어휘 전부가 10% 이하이며, '공공시설', '어업', '상업', '체육, 오락', '음악', '동물'들도 마찬가지인데 통일 후에 국어 정책상의 과제임을 시사해 준다고 하겠다.

4.2. 북한 언어 사회의 특징

북한 고어빈도어와 남한 기초 어휘의 비교를 통해 북한 고어빈도어에는 정치·사회의 이데올로기를 반영하는 어휘들이 많이 있다는 점이다. 특히 '혁명' 관련 합성어나 파생어들이 매우 많이 나타난다. 다음은 고어빈도어 간극

별로 나타난 북한 언어 사회의 특징을 보여주는 관련 어휘들이다.

(1-500) : 인민, 건설, 혁명, 투쟁, 사상, 통일, 위대하다, 적, 수령, 전사, 혁명적, 공산주의, 주체사상, 침략, 혁명투쟁, 중앙위원회, 원쑤<17개>

(501-1,000) : 간부, 군대, 충성심, 혁명위업, 지도자, 탄압, 근로인민, 혁명가, 싸움, 계급적, 혁명사상, 원수님, 혁명화, 령도, 혁명전통, 유일사상교육, 로동계급, 파쇼독재, 혁명활동, 침략자, 혁명성, 당위원회, 항일유격대, 항일무장투쟁, 지도일군, 유일사상체계, 당정책, 당중앙위원회, 총, 항일혁명투쟁, 인민군, 혁명과업, 유일사상<33개>

(1,001-1,500) : 경애하다, 제국주의, 주체시대, 빨찌산, 제국주의자, 생산량, 계급투쟁, 공산주의적, 전투적, 공산주의건설, 당중앙, 혁명운동, 인민군대, 혁명동지, 인민생활, 인민회의, 근로인테리, 인민성, 위원장, 공산주의교양, 미제, 주체사상화, 공산당, 공산주의사상, 유일적지도체계<25개>

5. 결론

북한 고빈도어 상위순 500개를 택하여 어휘 체계와 양상을 고찰하면서 남한의 기초 어휘와의 비교 및 분포 간극의 특징 그리고 북한 언어 사회의 특징도 살펴 보았다. 결과를 요약함으로써 결론을 대신한다.

첫째, 빈도수 조사 본문으로 선택된 문헌과 작품은 북한의 사회체제상 혁명 이데올로기와 관련된 것들의 비중이 높은 편이었다. 따라서 고빈도어로서의 출현 어휘와도 비례하며, 북한 언어 사회의 특징과도 함수 관계를 보여 주었다.

둘째, 어종별(語種別) 분포도는 고유어는 314개로서 62.8%, 한자어는 1755개로서 35%, 기타 11개로서 2.2%이다. 서구 및 동구 외래어도 상위 고빈도어 500개 안에 한 개도 들어 있지 않았다. 그리고 김일성 교시에 따라, '어휘정리'와 '단어만들기' 사업의 일환으로 고유어를 바탕말로 해서 만들어진 신조어(新造語)도 마찬가지다. 이들 어휘가 아직 '사용 어휘'의 지위를 얻지 못한 것으로 생각된다.

셋째, 품사별로 분류하면 명사 258개(51.6%), 동사 118개(23.5%), 형용사(41개(8%), 부사 42개(8.4%), 대명사 16개(3%), 관형사 11개(2%), 수사 3개(0.6%), 조사 2개(0.4%), 감탄사 1개(0.2%)의 순서를 보인다.

넷째, 남한의 기초 어휘를 기준으로 한 북한 고어빈도어 간극상의 분포는 어휘 목록 분야별로 차이를 보인다. 일치도가 높은 것과 낮은 것이 있는데, 전자로는 '사람에 관한 어휘', '감각 및 인식에 관한 어휘', '동작에 관한 어휘', '상태에 관한 어휘' '기타 어휘' 등이다. 후자로는 '의식주에 관한 어휘', '사회생활에 관한 어휘', '교육 및 예체능에 관한 어휘', '자연계에 관한 어휘' 등이다. 이는 사회 체계 및 경제 생활의 차이에서 기인한 것으로 분석된다.

다섯째, 남한 기초 어휘와의 일치도가 높은 어휘는 통일 후의 국어 문제 해결에 긍정적인 면을 보여 주며, 일치도가 거의 없는 분야의 어휘는 국어 문제 해결을 위한 과제가 무엇인가를 보여 준다고 하겠다.

참고 문헌

- 고영근. 1989. 북한의 말과 글. 을유문화사.
 김광해. 1993. 국어 어휘론 개설. 집문당.
 김미영외. 1990. "북한 정기 간행물의 사회언어학적 고찰," 북한·통일연구 논문집. 국토통일원.
 김민수. 1985. 북한의 국어 연구. 고려대학 출판부.
 김종택. 1992. 국어 어휘론. 탑출판사.
 리익선. 1974. 단어만들기 연구. 사회과학출판사.
 문영호외. 1994. 조선어 빈도수 사전. 한국문화사(영인).
 박상훈외. 1985. 우리나라에서의 어휘정리. 사회과학출판사.
 박영순. 1985. "관용어에 대하여," 이울관교수 화갑기념논문집.
 백종학. 1994. "한국인의 어휘사용과 정치성의 지향," 언어 19권.
 심재기. 1982. 국어 어휘론. 집문당.
 유목상. 1976. "통어론적 구성에 의한 어형성에 관한 연구," 성곡논총 5집.
 이상익. 1990. "국어 어휘부의 계량학적 연구," 강신헌교수 화갑기념논문집.
 이성연. 1991. "북한 언어정책의 특징," 성곡논총 22집.
 _____. 1992. "'북한어 조어법에 관한 연구,' 한국언어문학 30집.
 이용주. 1969. "한국어 어휘체계의 특징," 국어교육 15집.
 장태진. 1981. 국어사회학 연구. 과학사.
 _____. 1995. 국어사회언어학 연구. 삼영사.
 조재수. 1986. 북한의 말과 글. 한글학회.
 Coper, R.L. 1989. *Language Planning and Social Change*. Cambridge Univ. Press.

Fishman, J.A. 1972. *Sociolinguistics*. Newbury House Pub..

Fishman, J.A.(ed.). 1974. *Advances in the Language Planning*. Mouton Pub..

Fishman, J.A.(ed.). 1976. *Language and Politics*. Mouton Pub..

Wardhaugh, R. 1992. *An introduction to Sociolinguistics*. Blackwell Pub..

(5011-759) 광주광역시 동구 서석동 조선대학교 사범대학 국어교육과

E-mail : syblee@mail.chosun.ac.kr