

## 주택디자인에서 건축가들의 어휘 사용행태 및 기본어휘에 관한 연구

### A Study on the Lexicon-Use Behaviour of Architects & the Basic Lexicons in House Design

윤 대 한\*  
Youn, Dae-Han

#### Abstract

This paper analyzed statistically two corpora that were constructed from the texts about house designs written by Korean architects and PA Awards architects. The main results are as follows; (1) The numbers of words in Korean house-design corpus were 9,352 and those of words in PA Awards house design corpus were 2,379. The former were 18.7% and the latter 4.8% of about 50,000 words regarded as the real using scale in actual life. (2) When the architects described their house designs, the lexicon-concentration phenomenon was pervasive in both groups. Therefore, we can infer that the high-frequency lexicons are very important in house design. (3) The architects' behaviour patterns of using the house-design lexicons, went by rules according to the word frequency order. The tendency formulas of them had the R<sup>2</sup> values which were more than 90%. (4) In Korean house design corpus, the high frequency lexicons were '공간', '층', '주택', '집', '대지', '거실', and '실'. In PA awards house design corpus, they were 'house', 'room', 'space', 'living', 'wall', 'level' and 'area'. From these results, We can tell that 'space' is the highest frequency word in house design of the two groups, and that '대지' and 'wall' are the words that reveal well the differences between the two groups.

Keywords : house design, architect, behaviour, lexicons, corpus linguistics, PA Awards

주요어 : 주택디자인, 건축가, 어휘사용행태, 기본어휘, 문치언어학

## 1. 들어가는 말

### 1. 연구의 배경과 목적

최근 들어 건축가들의 말과 글에 대해 건축가와 건축이론가들 뿐만 아니라 일반인들의 관심이 높아졌는데, 건축가가 자신의 작품과 관련하여 직접 쓴 글과 말은 그들의 사고를 반영하는 일차 자료로써 그 가치를 충분히 갖는다. 과거에 비해 건축가들이 다양한 글쓰기 활동을 통해 건축담론을 생산하고 있고, 이것들이 잡지, 서적, 강연회 등을 통해 전달되면서 건축디자인 분야에 많은 영향을 미치고 있다.

말과 글에서 사람의 언어 사용행태는 많은 종류의 어휘를 평균적으로 사용하는 것이 아니라 일정 규모의 특정 어휘를 주로 많이 사용하는 것으로 알려져 있다. 이 같은 어휘 사용행태를 고려한다면, 특정 집단의 사용어휘 규모뿐만 아니라 사용빈도가 높은 어휘의 규모와 그 종류를 찾아내는 일은 매우 가치있는 작업이다. 언어학에서 잘 다루지 못했던 이 분야는 최근의 비약적인 컴퓨터관련 기술발달 덕분에 대규모 언어에 대한 계량적 분석이 가능하게 되었고, 그로부터 얻은 결과들은 특정

개인의 주관성 있는 이론보다 설득력 있게 받아들여지면서 많이 응용되고 있다. 국내 언어학에서 이 같은 접근 방법론은 1980년대 후반기에 문치언어학으로 소개되면서 본격적으로 시작되었지만, 건축분야에서 구체적으로 연구된 사례는 거의 없다<sup>1)</sup>.

건축에서 주택은 인간생활을 담고 반영하는 건축디자인의 가장 기본적인 분야로, 개성있는 건축가의 예술성이 잘 표현될 수 있을 뿐만 아니라 건축디자인의 목적과 동기의 측면에서 가장 단순하고 순수하여 건축의 시작이며 끝으로 생각되기도 하며, 사보아 주택과 같이 건축가의 사상이나 시대정신을 반영할 수도 있는 범주이기도 하다<sup>2)</sup>.

1) 건축분야에 홍성민(2005)의 “교회건축물에 대한 사용자의 인식 특성에 관한 연구”가 있으며, 연구진행을 위해 문치언어학에서 강조하는 자연언어를 어휘수집 방법에 이용하였다. 그러나 전반적인 문치언어학의 방법론을 적용한 것은 아니다.

2) “주택의 문제는 시대의 문제이다”(Le Corbusier), “단독주택은 젊은 건축가들이 성장할 수 있는 토양이며 시대정신은 단독주택에 진정으로 표현될 수 있다.”(Paul Rudolph, 1975), “단독주택만이 건축의 또 다른 문제인 은유, 상징 인습 등의 수용 문제를 해결할 수 있는 유일한 소재이다.”(Peter Eisenman, 1975), “단독주택은 언제나 비평의 출발점이 되어왔다. 건축가들은 이를 통해 많은 것을 배우고 있다.”(Charles Gwathmey, 1977) 등에서 건축가들의 생각을 엿볼 수 있다.

\*정희원(주저자), 광주대학교 건축학과 부교수, 공학박사

본 글은 주택디자인과 관련하여 국내외 건축가들이 서술한 텍스트를 문치언어학의 방법론에 따라 계량적으로 분석함으로써, 주택디자인과 관련한 건축디자인 어휘의 규모 및 사용행태를 밝히고, 이 결과로부터 주택디자인관련 주요 기본어휘를 찾고 이를 비교하는 것을 목적으로 한다.

2. 연구방법과 범위

본 글의 주요 연구방법은 문치언어학의 방법론을 따라 진행되었다. 먼저 주택디자인과 관련하여 자연언어에 해당하는 국내외 건축가들이 쓴 텍스트를 수집하여 말뭉치를 구성한 뒤, 연구에 적합한 기준에 따라 어휘를 분절 및 정리하고 이에 필요한 어휘정보를 연결시킨 뒤 어휘빈도를 측정하여 그 결과에 따라 빈도순위별로 어휘를 정리함으로써 각 말뭉치 안에서의 기본어휘를 찾아내었고 이를 비교하였다.

말뭉치 구성 대상으로 국내의 것은 국내 주요 건축잡지 게재된 작품을, 외국의 것은 영어로 작성되었고 인지도가 높은 PA Awards 주택부문 수상작을 대상으로 하였다. 시간적 범위는 *Progressive Architecture*의 폐간 시점, 국내 잡지의 창간시점, 텍스트양 등을 고려하여 국내는 1981-1994년, PA는 1966-1994년 이다.

II. 문치언어학과 어휘빈도

1. 문치언어학(Corpus Linguistics)

문치언어학은 언어 전문가나 비평가가 행하는 언어의 조작성과 인위성을 피하고, 언어의 자연상태를 포착하려는 목적으로, 일정한 원칙에 따라 실생활에서 사용된 말, 즉 자연언어를 수집하여 만든 ‘말뭉치(Corpus)’<sup>3)</sup>를 컴퓨터로 처리하고 분석하여 다양한 언어정보를 추출해 내는 언어학의 한 분야이다. 역사적으로 말뭉치의 개념은 실증적 자료를 중시하던 1950년대 미국 구조주의 언어학자들이 실제 말이나 글로 발화된 언어 자료를 일차적 설명 대상으로 삼은 것을 그 효시로 볼 수 있다.

말뭉치를 이용한 실질적인 연구는 1959년 더럼(Durham) 대학에서, 1960년 이후에는 런던대학에서 영어의 실제 쓰임을 광범위하게 조사한 것에서 찾을 수 있고, 본격화된 것은 1963년부터 2년 동안 미국 브라운(Brown) 대학에서 구성한 브라운 말뭉치(Brown Corpus)<sup>4)</sup>이며, 널리 유포된 계기는 1980년대 대량의 어휘를 기반으로 한 자료 이용법이 활성화되면서 추진된 영국의 「코빌드(COBUILD)」<sup>5)</sup> 계획이다.

3) 말뭉치는 일반적으로 ‘자연 상태의 언어, 즉 언어의 있는 그대로 모습을 잘 드러내 보여 줄 수 있는 자료를 모아 놓은 것’을 뜻한다.  
 4) 15개 장르 500편의 글에서 2,000 어절씩 추출하여 총 100만 어절 말뭉치를 구축하고 컴퓨터를 이용하여 이를 분석해 어휘에 관한 다양한 통계를 제시하였다. 현대적 의미의 문치언어학 시조가 되는 것이다  
 5) COBUILD는 Collins Birmingham University International Language Database의 약자이다.

다른 언어학 분야와는 달리 말뭉치를 이용하는 문치언어학의 핵심은 연구방법론이다. 문치언어학은 말뭉치를 대상으로 자연언어에 대한 계량적 결과를 분석함으로써 과거의 오류를 수정하고 새로운 것을 찾아내고<sup>6)</sup>, 언어 연구의 객관성, 신뢰성, 대량성 등을 확보할 수 있게 되었다<sup>7)</sup>. 따라서 이제 언어의 확률적 사실에 대한 판단은 개인의 직관이 아니라 대량의 언어자료를 분석한 문치언어학적 방법에 의해서 결정되는 시대가 되었고<sup>8)</sup>, 이 결과들은 사전편찬, 정보검색, 번역기, 분야별 교육의 우선순위 결정, 언어사용 집단의 의식 분석 등에 결정적인 역할을 하고 있다.

국내에서는 1988년에 이상섭이 ‘Corpus’를 ‘말뭉치’<sup>9)</sup>로, ‘Corpus Linguistics’를 ‘문치언어학’으로 번역하면서 처음 소개되었다. 이후 연세대학교의 ‘연세한국어말뭉치’가 1988년부터 구축되고 있고, 한국과학기술원, 고려대학교, 국립국어연구원 등에서도 본격적으로 말뭉치를 구축하고 있으며, 1998년부터 ‘21세기 세계계획: 국어정보화 추진 중장기 사업’의 일환으로 국가적 대규모 말뭉치가 구축되고 있다.

2. 문치언어학의 연구방법론

연구방법론이 핵심인 문치언어학에서 사용하는 일반적인 연구진행 과정은 다음과 같다.

- a. 말뭉치 설계: 표본 선정기준을 정함
- b. 원시말뭉치 구성: 어휘를 수집하고 컴퓨터에 입력
- c. 분절작업: 텍스트를 원하는 단위<sup>10)</sup>로 분절
- d. 태깅(tagging): 분절된 단위에 관련 어휘정보를 연결
- e. 어휘정리 작업
- f. 통계작업: 어휘빈도를 측정하고, 이 자료를 토대로 사용 어휘규모, 어휘 활용태, 기본어휘 등을 밝힘
- e. 어휘의미 사례 분석: 문장 안에서의 어휘의미 분석<sup>11)</sup>

3. 어휘, 어휘소, 형태소

어휘는 어휘소(lexeme)들의 총체로 정의되며, 어휘소를 원소로 하는 다양한 범주의 부분집합들을 가질 수 있다. 그러나 어휘는 단순한 어휘소들의 집합을 가리키는 것보다는, 기초어휘, 생활어휘, 방언어휘, 전문어휘, 누구의 작품어휘 등과 같이 일정한 성격을 지닌 단어들의 집합을 가리키는 경향이 강하다. 따라서 건축가들의 주택디자인

6) 실제로 문치언어학은 개인적 직관의 연구와 자연언어 사이에 많은 차이가 있음을 밝혔고, 기존 사전들이 갖고 있는 오류를 지적하였다.  
 7) 박영환, 「말뭉치에 기반한 형태소 분석기 및 철자 검사기의 구현」, 연세대석사학위논문, 1991, p.2.  
 8) 이상섭의, 「우리말 낱말 빈도 조사 표본의 선정기준」, 「새한국어사전 편찬을 위한 사전편찬학 연구」, 제3집, 탑출판사, 1990, p.176.  
 9) 말뭉치 외에 정광, 이기용, 김홍규, 임혜창, 강범모 등은 ‘말모듬’, ‘언어 데이터베이스’, ‘코퍼스’ 등으로 번역하여 사용하기도 한다.  
 10) 이 때 단위로는 어휘소나 형태소 단위가 선택된다.  
 11) 고빈도 어휘나 가치있는 특정 어휘에 대해 실제 사용된 문장 안에서의 어휘의미를 분석함으로써 어휘의 의미활용태를 알아낸다.

표 1. 국어와 영어에서 문장 분절단위에 따른 분절 예

구분		분절 예
국어	문장	우리나라의 최고령자가 누구인지
	어휘소단위	우리나라, -의, 최고령자, -가, 누구, -이, -지
	형태소단위	우리, 나라, -의, 최-, 고령-, -자, -가, 누구, -이, -나, -지
영어	어휘소단위	understand
	형태소단위	under-, stand

어휘도 국어에서 하나의 부분집합 어휘가 될 수 있다. 언어학에서 문장 분절을 위한 단위에는 음소(phoneme), 형태소(morpheme), 어휘소가 있다. 음소는 의미를 변별시켜 주는 최소 단위이다. 음소들이 모여 단어를 구성하거나 '의미를 가진 최소 단위'를 형성하면 이것이 형태소이다. 어휘소는 단어 수에 상관없이 하나의 의미단위를 말하며 단어가 중심이 되며 형태소가 포함된다. 문치언어학에서 텍스트를 분석 단위로 분절할 때 그 기준을 형태소나 어휘소에 두지만, 그 기준을 어디에 두는가에 따라 약간의 통계적 차이가 발생한다. 따라서 분석하고자 하는 분야와 목적에 적합한 분할 단위를 설정할 필요가 있다.

4. 어휘빈도와 기본어휘

개수 혹은 회수를 의미하는 빈도는 절대영점이 있고 '1'을 단위(unit)로 하며 확률과 직접적이고 불가분의 관련성을 갖는 확률적 현상이다. 따라서 빈도분석은 정상분포 등의 비현실적 가정 없이도 직접적으로 확률적 대상이 될 수 있다<sup>12)</sup>. 빈도자료는 다른 형태의 자료들과 비교하여 일반적으로 단순하고 자연적이라는 속성이 있고, 자료 수집의 간편함과 수집된 자료의 해석적 명확함 때문에 여러 분야에서 중요한 역할을 담당하고 있다.

말뭉치 안에서 어휘사용 회수를 의미하는 '어휘빈도' 그 자체는 단어의 물리적 성질의 측정치로, 해당 어휘가 언어사용 주체들에 의해 얼마나 자주 쓰이는가를 나타내 주는 지표이다. 따라서 어휘빈도는 말뭉치 내에서 사용된 어휘의 상대적 중요성을 가늠하는 준거가 될 수 있다. 그러나 보다 더 중요한 사실은 그와 같은 물리적인 성질을 결정하는 것이 어휘사용자 의식의 어떤 성질이라는 사실이다<sup>13)</sup>. 따라서 수량화된 어휘빈도는 언어사용 주체의 의식을 드러낼 수 있으며, 그 정보는 하나의 객관적 근거가 될 수 있다<sup>14)</sup>.

어휘빈도는 일반적으로 ①사용어휘의 구조, ②언어사용 행태 ③어휘사용 주체의 의식이나 사회의 지배사상과 같은 정보<sup>15)</sup> ④관련 교육의 기본방향 설정을 위한 기준 설정<sup>16)</sup> ④어휘의 다양한 의미 가운데 의미의 상대적 중요성을 가늠하는 정보 제공 등과 같은 기능들을 갖는다<sup>17)</sup>.

'기본어휘'는 말뭉치 안에서 어휘 사용빈도가 상대적으로 매우 높은 어휘들이다<sup>18)</sup>. 사람들이 어휘를 집중해서 사용한다는 일반적인 사실을 고려한다면, 기본어휘는 집중정도에 따라 상대적으로 매우 중요한 가치를 갖는다. 실생활에서 아무리 풍부한 어휘를 구사하는 사람이라도 일반적으로 모국어의 10% 이상을 알고 있는 사람이 드물다는 사실을 고려한다면<sup>19)</sup>, 언어 사용집단의 전체 어휘규모뿐만 아니라 기본어휘를 찾아내는 것도 매우 중요한 일이다.

III. 주택디자인 말뭉치 구성

1. 텍스트 수집

국내 건축가들의 주택디자인 텍스트는 「건축문화」(1981-현재)를 중심으로 1981-1994년 동안 게재된 작품을 대상으로 수집하였다. 연도별로 작품수 및 텍스트 양이 부족할 경우 2차적으로 「건축사」(1963-현재)와 「건축과환경」(1984-현재)를 통하여 보완하여 총 218개의 작품이 수집되었다.

PA Awards<sup>20)</sup> 건축부문은 매년 약 800여개 작품들이 출품되었지만, 수상작은 그 가운데 건물용도와는 관계없이 약 10여개였다. 당선작 비율이 출품작 대비 평균 1.9%이지만, 주택은 당선작 대비 평균 18.8%로 다른 용도의 건축물과 비교하여 상대적으로 PA에서 비중있게 다루어지고 있다고 볼 수 있다. PA Awards로부터의 건축가들의 텍스트 수집은 1996년 PA Awards의 심사기준 변경

12) 박광배, 「빈도분석」, 성원사, 1992, p.5.  
 13) 조평환, 「언어심리학」, 민음사, 1991, p.78.  
 14) 계량적 연구에 대해 비판적이었던 전통문학연구, 평론, 문학사 등에서도 문학현상의 빈도, 정도, 변화 등을 나타내는 수량개념은 낡은 것이 아니다. 자연언어의 수량표현은 '많은', '좀 더' 등과 같은 모호한 주관적 척도에 의존함으로써 소통 및 검증의 장애를 일으키는데, 이런 폐단을 줄이기 위해서라도 계량적 분석의 활용은 절실하다.(김홍규, 「고시조 데이터베이스 색인어 검색 및 분석 방안」, 「한국어 데이터베이스의 설계 및 응용을 위한 기초 연구」, 민음사, 1995, p.136.)

15) Luhn(1967)은 문헌에 출현한 단어들은 문헌 내용분석을 위해 사용될 수 있고, 단어 출현빈도가 주제어로서의 중요성을 측정하는 기준이 된다고 보았다. 심리학에서는 단어 재인 속도(Solomon and Postman, 1952; Morton, 1970; 김민식, 정찬섭, 1989)나 기억망 속에서의 정보 인출(Conrad, 1972; Andersson, 1983) 연구 등에서 볼 수 있는 것처럼, 단어빈도에 관한 지식이 인지구조 및 정보처리 과정을 이해하는데 중요한 설명 변수로서 다루어진다.  
 16) 건축과 관련하여 건축교육 등에서 교육의 우선순위를 결정할 수 있는 실마리를 제공할 수 있다.  
 17) 정찬섭 외, 「우리말 낱말 빈도 조사 표본의 선정 기준」, 「새 한국어 사전편찬을 위한 사전편찬학 연구」, 제3집, 탑출판사, 1990, pp.7-10.  
 18) 말뭉치에서 기본어휘를 정하는 기준으로 여전히 어휘 빈도수를 중심으로 하고 있다. 빈도수 외에 어휘사용 폭과 사용주체의 인식도 고려할 필요성이 제기되기도 한다.  
 19) Jespersen, Growth & Structure of English Language, 1930, p.196.  
 20) PA Awards는 근대건축이 분화되고 상업화되면서 그 전성기를 맞고 있던 1953년 당시 PA의 편집장이었던 Creighton과 그의 동료들이 그 이전부터 매년 실시되어 오던 Design Preview Selection을 공모전 수준으로 끌어올리기로 결정한 것을 계기로 시작되었다.

시점<sup>21)</sup>과 1996년 초 폐간 시점을 고려하고, 국내 텍스트 수집 기간을 포함하면서 적절한 텍스트 양을 확보하기 위해 1966년부터 1994년까지 주택작품 74개를 대상으로 하였다. PA Awards의 텍스트는 Project, Location, Site, Program, Design Solution, Construction and Materials, Jury Comment로 구성되어 있지만, 국내 텍스트의 성격과 유사한 범주인 Program과 Design Solution에 해당하는 텍스트만을 사용하였다.

2. 텍스트의 분절

텍스트 분절은 어휘소 단위를 기준으로 하였다. 영어와 비교하여 다소 복잡한 문법구조를 갖고 있는 국어에서는 연구 목적에 맞게 다음과 같은 기준들이 적용되었다.

명사의 경우, 예를 들어 공간과 관련하여 ‘공간이,’ ‘공간은,’ ‘공간을,’ ‘공간과,’ ‘공간에,’ ‘공간의,’ ‘공간으로,’ ‘공간에서(도)(의)’ 등의 어휘들이 실제 사용되었지만, 이들은 ‘공간’이라는 동일한 어휘소에 다양한 조사가 붙은 것이기 때문에, 본 글에서는 어휘소 단위만 다루고, 기능어에 해당하는 조사는 제외하였다. 형용사와 동사는 어간이 동일한 어휘들을 정리한 다음, 그 가운데 빈도수가 가장 높은 어휘(대표어휘)로 통일시켜 사용하였다. 즉, ‘작은,’ ‘작고,’ ‘작다’와 같은 형용사는 빈도수가 가장 높게 조사된 ‘작은’이란 어휘로, ‘두다,’ ‘두어,’ ‘두고,’ ‘두었다,’ ‘두는,’ ‘둔다,’ ‘두면서’ 등의 동사는 빈도수가 가장 높은 어휘 ‘두다’로 통일시켜 사용하였다.

3. 주택디자인 말뭉치의 구성 및 평가

(1) 주택디자인 말뭉치의 구성

구성된 국내주택디자인 말뭉치의 연도별 작품수, 어휘 규모, 평균 어휘규모는 <표 2>와 같다. 구성된 말뭉치의 사용어휘 전체규모는 조사(助詞)를 제외하고서<sup>22)</sup> 40,454개, 연도별 평균 작품수는 15.6개, 연도별 평균어휘규모는 2,890개, 작품당 평균어휘규모는 약 186개였다. 국내의 경우 80년대 초반에 비해 1985년 이후 작품당 어휘 규모가 확연히 많음을 볼 수 있다. 해당 건축가는 모두 154명<sup>23)</sup>이었고, 일인당 작품수는 평균 1.4개이다.

PA Awards는 매년 수상 작품수가 다르다. 구성된 PA

21) PA Awards의 심사관점은 1966년에 PA 편집장인 J.M. Dixon은 심사평가기준을 그 동안의 근대주의에 근거한 단일 기준이었던 배타적인 방법에서 포용적인 방법으로 전환하는 새로운 평가 방법을 선언하였다. 즉, 당시의 새로운 다원주의적 경향을 포용하는 방법으로 작품을 평가할 것을 결정하였고, 또한 상대적으로 주관적이며 변화하고 있는 건축개념의 가치를 객관적인 평가기준으로는 결정할 수 없으며, 우수성을 결정하는 가치관도 또한 변하며 가치관의 변화야말로 현대 사회의 본질을 공식적으로 선언하였다.

22) 제외된 조사 가운데 대표적인 것들에는 ~의(960개), ~을(839개), ~에(554개), ~이(480개), ~를(417개), ~과(406개), ~으로(394개), ~은(309개), ~가(285개), ~로(265개), ~에서(217회), ~인(200회) 등이 있으며, 이 결과는 정영미(1993)의 연구결과와 거의 일치한다.

23) 공동작품의 경우 건축가 모두를 포함하였다. 218개의 작품 가운데 2명 공동작품이 16개였고, 3명 공동작품이 1개였다.

Awards 주택디자인 말뭉치의 작품별 어휘규모는 <표 3>과 같다. 사용어휘 전체규모는 11,860개이며, 작품당 평

표 2. 연도별 국내주택디자인 말뭉치 구성

연도	작품수	어휘규모	작품당 평균 어휘규모
1981	25	2,430	97.2
1982	20	2,117	105.9
1983	8*	503	62.9
1984	7*	963	137.6
1985	24	4,783	199.3
1986	10	2,003	200.3
1987	14	3,117	222.6
1988	18	2,858	158.8
1989	16	4,050	253.1
1990	13	4,556	350.5
1991	18	3,335	185.3
1992	13	3,448	265.2
1993	18	2,858	158.8
1994	14	3,433	245.2
합	218	40,454	
평균	15.6	2,889.6	

\*수집대상 3개의 잡지에 더 이상의 주택작품이 없음

표 3. PA Awards 주택디자인 말뭉치 구성

번호	어휘 규모	연도	번호	어휘 규모	연도	번호	어휘 규모	연도
1	153	1966	26	141	1980	51	99	1986
2	379	1967	27	208	1980	52	157	1986
3	221	1967	28	115	1980	53	156	1986
4	98	1967	29	81	1980	54	83	1987
5	225	1967	30	343	1980	55	156	1988
6	98	1969	31	101	1980	56	124	1989
7	126	1969	32	34	1980	57	236	1989
8	118	1970	33	56	1980	58	107	1989
9	141	1972	34	201	1980	59	115	1989
10	94	1973	35	103	1980	60	177	1991
11	82	1973	36	90	1981	61	211	1991
12	74	1975	37	244	1982	62	155	1991
13	160	1976	38	97	1982	63	199	1991
14	126	1976	39	212	1982	64	300	1992
15	242	1977	40	106	1982	65	152	1992
16	193	1977	41	218	1982	66	164	1992
17	87	1977	42	123	1983	67	182	1992
18	203	1977	43	225	1984	68	143	1992
19	210	1978	44	126	1984	69	163	1993
20	364	1978	45	63	1984	70	242	1993
21	156	1978	46	126	1984	71	98	1993
22	213	1978	47	239	1985	72	155	1993
23	366	1979	48	64	1985	73	230	1993
24	123	1979	49	121	1985	74	132	1994
25	152	1980	50	83	1985			

\*PA Awards의 전체 어휘규모=11,860개

\*작품당 평균 어휘규모=160.3개

국어어휘규모 어휘규모는 160개이며, 해당 건축가는 65명으로 일인당 작품수는 1.1개 이다. 두 말뭉치에서 작품당 평균어휘규모 차이는 26개, 일인당 작품수 차이는 0.3개 이다.

(2) 말뭉치 평가

특정 개인어휘를 분석하지 않는 경우, 말뭉치의 적절성에 대한 평가요소는 말뭉치의 규모 및 글쓴이의 다양성이다.

말뭉치의 규모는 가능한 클수록 좋지만, 시간과 인력, 그리고 비용의 한계가 있기 때문에 개별연구의 경우 한계상황을 고려한 의미있는 통계치를 얻을 수 있는 적정규모를 설계할 필요가 있다. 국가적 사업을 통한 대규모 말뭉치의 경우 그 규모가 100만 개 이상이며 1억 개가 넘는 것도 있지만, 일반적인 국내의 관련 개별 연구논문의 경우 구성된 말뭉치의 규모 분포는 약 2,000-50,000 개가 주를 이룬다.

구성된 두 말뭉치에서 각각 어휘규모 40,454개 및 11,860개, 작품당 어휘규모 186개 및 160개, 전체 어휘규모에 대한 개별작품 어휘규모의 점유비 0.46% 및 1.35%, 전체 작품수에 대한 개별건축가의 점유비 0.64% 및 1.49% 등의 수치를 고려하면 두 말뭉치의 규모가 적절하고, 말뭉치 내에서의 다양성과 균형성을 확보하고 있는 것으로 볼 수 있다.

IV. 건축가들의 주택디자인 어휘사용 행태 및 기본어휘

1. 건축가들의 주택디자인 어휘사용행태

말뭉치를 통한 언어 사용주체의 어휘사용행태는 사용어휘의 규모, 어휘빈도 순위별 어휘수, 어휘활용비, 기본어휘 등을 통하여 알 수 있다. 본 글에서 구성한 국내 및 PA Awards 주택디자인 말뭉치를 정리한 결과는 <표 4> 및 <표 5>와 같다.

(1) 주택디자인 어휘규모

국내 주택디자인 말뭉치에서의 어휘사용 총규모는 40,454개, 개별 어휘종류는 9,352개로 조사되었다. 국내 최대 규모의 「표준국어대사전」<sup>24)</sup>(1999) 어휘종류 약 50만 개와 비교하면, 국내주택디자인 어휘종류 규모는 약 1.9% 정도이다. 국내 자연언어에 기초한 말뭉치에 근거해 편찬 계획에 있는 「연세한국어사전」의 약 20만개 어휘와 비교하면 약 4.7% 정도이다. 김광해(1993)는 각종 어휘조사 결과를 고려하여 볼 때 세계 여러 나라에서 실제 사용되고 있는 어휘규모는 5만-6만 개 정도로 보고 있는데, 이를 적용하면 국내주택디자인관련 어휘종류는

표 4. 국내주택디자인 말뭉치에서 건축가들의 건축디자인 어휘사용행태

빈도 순위	사용 빈도 회수	개별 어휘 개수	누적 어휘 개수	어휘	어휘 사용 회수	전체 어휘종류에 대한 활용비(%)		전체 어휘 사용량에 대한 활용비(%)		한어휘의 활용비
						비	누적비	비	누적비	
1	670	1	1	공간	670	0.011	0.011	1.656	1.656	1.656
2	479	1	2	수(불완전 명사)	479	0.011	0.021	1.184	2.840	1.184
3	455	1	3	것	455	0.011	0.032	1.125	3.965	1.125
4	438	1	4	층	438	0.011	0.043	1.083	5.048	1.083
5	420	1	5	주택	420	0.011	0.053	1.038	6.086	1.038
6	342	1	6	이다	342	0.011	0.064	0.845	6.931	0.845
7	329	1	7	이(지시대 명사)	329	0.011	0.075	0.813	7.744	0.813
8	324	1	8	집	324	0.011	0.086	0.801	8.545	0.801
9	320	1	9	있는	320	0.011	0.096	0.791	9.336	0.791
10	280	1	10	대지	280	0.011	0.107	0.692	10.028	0.692
11	270	1	11	있다	270	0.011	0.118	0.667	10.696	0.667
12	241	1	12	거실	241	0.011	0.128	0.596	11.292	0.596
13	235	1	13	실	235	0.011	0.139	0.581	11.873	0.581
14	209	1	14	이(2)	209	0.011	0.150	0.517	12.389	0.517
15	205	1	15	그(지시대 명사)	205	0.011	0.160	0.507	12.896	0.507
16	202	1	16	방	202	0.011	0.171	0.499	13.395	0.499
17	195	1	17	건물	195	0.011	0.182	0.482	13.877	0.482
18	184	1	18	형태	184	0.011	0.192	0.455	14.332	0.455
19	169	1	19	정원	169	0.011	0.203	0.418	14.750	0.418
20	156	1	20	건축주	156	0.011	0.214	0.386	15.136	0.386
21	151	1	21	등(etc)	151	0.011	0.225	0.373	15.509	0.373
22	149	1	22	쪽	149	0.011	0.235	0.368	15.877	0.368
23	147	1	23	생활	147	0.011	0.246	0.363	16.240	0.363
24	133	1	24	내부	133	0.011	0.257	0.329	16.569	0.329
25	126	1	25	형	126	0.011	0.267	0.311	16.881	0.311
26	125	1	26	감	125	0.011	0.278	0.309	17.190	0.309
27	124	1	27	면	124	0.011	0.289	0.307	17.496	0.307
28	123	1	28	주거	123	0.011	0.299	0.304	17.800	0.304
29	120	1	29	층	120	0.011	0.310	0.297	18.097	0.297
30	115	1	30	지붕	115	0.011	0.321	0.284	18.381	0.284
31	114	1	31	벽	114	0.011	0.331	0.282	18.663	0.282
32	112	1	32	부분	112	0.011	0.342	0.277	18.940	0.277
33	111	1	33	외부	111	0.011	0.353	0.274	19.214	0.274
34	108	1	34	하였다	108	0.011	0.364	0.267	19.481	0.267
35	106	3	37	계단, 자연 도로	318	0.032	0.396	0.786	20.267	0.262
36	104	1	38	우리	104	0.011	0.406	0.257	20.524	0.257
37	103	1	39	가족	103	0.011	0.417	0.255	20.779	0.255
38	101	1	40	대한	101	0.011	0.428	0.250	21.029	0.250
39	100	1	41	식당	100	0.011	0.438	0.247	21.276	0.247
40	97	1	42	현관	97	0.011	0.449	0.240	21.516	0.240
41	94	1	43	곳	94	0.011	0.460	0.232	21.748	0.232
42	92	1	44	또한	92	0.011	0.470	0.227	21.975	0.227
43	91	2	46	일(1), 그리고	182	0.021	0.492	0.450	22.425	0.225
44	90	2	48	평면, 계획	180	0.021	0.513	0.445	22.870	0.222
45	89	1	49	설계	89	0.011	0.524	0.220	23.090	0.220
46	86	1	50	있도록	86	0.011	0.535	0.213	23.303	0.213
47	85	3	53	중심, 할, 건축	255	0.032	0.567	0.630	23.933	0.210
48	84	1	54	마당	84	0.011	0.577	0.208	24.141	0.208
49	83	2	56	남, 지하	166	0.021	0.599	0.410	24.551	0.205
50	82	3	59	창, 때, 위, 해	246	0.032	0.631	0.608	25.159	0.203
51	80	1	60	속	80	0.011	0.642	0.198	25.357	0.198
52	79	1	61	하는	79	0.011	0.652	0.195	25.552	0.195
53	78	1	62	분위기	78	0.011	0.663	0.193	25.745	0.193
54	75	1	63	하나	75	0.011	0.674	0.185	25.930	0.185
55	74	1	64	느낌	74	0.011	0.684	0.183	26.113	0.183
56	72	1	65	작은	72	0.011	0.695	0.178	26.291	0.178
57	70	2	67	기능, 및	140	0.021	0.716	0.346	26.637	0.173
58	69	2	69	개념, 때문	138	0.021	0.738	0.341	26.979	0.171
59	68	1	70	색	68	0.011	0.749	0.168	27.147	0.168
60	67	1	71	한다	67	0.011	0.759	0.166	27.312	0.166
61	66	3	74	요소, 두어, 따라	198	0.032	0.791	0.489	27.802	0.163

24) 국립국어연구원이 8년 동안 500명의 인력과 예산 112억 원을 들여 3권짜리로 완간한 사전으로 말뭉치를 기반으로 하였다. 중복 단어는 각각 계산해서 전문어 190,000개, 북한어 70,000개, 방언 20,000개, 옛말 12,000개를 수록하고 있다.

표 4. 계속

빈도 순위	사용 빈도 회수	개별 어휘 개수	누적 어휘 개수	어휘 사용 회수	전체 어휘종류에 대한 활용비(%)		전체 어휘 사용량에 대한 활용비(%)		
					비	누적비	비	누적비	한어휘의 활용비
62	65	1	75	65	0.011	0.802	0.161	27.962	0.161
63	64	6	81	384	0.064	0.866	0.949	28.912	0.158
64	62	5	86	310	0.053	0.920	0.766	29.678	0.153
65	61	3	89	183	0.032	0.952	0.452	30.130	0.151
66	60	3	92	180	0.032	0.984	0.445	30.575	0.148
67	59	3	95	177	0.032	1.016	0.438	31.013	0.146
68	58	2	97	116	0.021	1.037	0.287	31.300	0.143
69	56	4	101	224	0.043	1.080	0.554	31.853	0.138
70	55	3	104	165	0.032	1.112	0.408	32.261	0.136
71	54	1	105	54	0.011	1.123	0.133	32.395	0.133
72	53	2	107	106	0.021	1.144	0.262	32.657	0.131
73	52	1	108	52	0.011	1.155	0.129	32.785	0.129
74	51	3	111	153	0.032	1.187	0.378	33.163	0.126
75	50	5	116	250	0.053	1.240	0.618	33.781	0.124
76	49	6	122	294	0.064	1.305	0.727	34.508	0.121
77	48	1	123	48	0.011	1.315	0.119	34.627	0.119
78	47	1	124	47	0.011	1.326	0.116	34.743	0.116
79	46	3	127	138	0.032	1.358	0.341	35.084	0.114
80	45	3	130	135	0.032	1.390	0.334	35.418	0.111
81	44	1	131	44	0.011	1.401	0.109	35.527	0.109
82	43	5	136	215	0.053	1.454	0.531	36.058	0.106
83	40	5	141	200	0.053	1.508	0.494	36.552	0.099
84	39	7	148	273	0.075	1.583	0.675	37.227	0.096
85	38	5	153	190	0.053	1.636	0.470	37.697	0.094
86	37	7	160	259	0.075	1.711	0.640	38.337	0.091
87	36	8	168	288	0.086	1.796	0.712	39.049	0.089
88	35	5	173	175	0.053	1.850	0.433	39.482	0.087
89	34	8	181	272	0.086	1.935	0.672	40.154	0.084
90	33	3	184	99	0.032	1.967	0.245	40.399	0.082
91	32	10	194	320	0.107	2.074	0.791	41.190	0.079
92	31	9	203	279	0.096	2.171	0.690	41.879	0.077
93	30	6	209	180	0.064	2.235	0.445	42.324	0.074
94	29	8	217	232	0.086	2.320	0.573	42.898	0.072
95	28	11	228	308	0.118	2.438	0.761	43.659	0.069
96	27	12	240	324	0.128	2.566	0.801	44.460	0.067
97	26	11	251	286	0.118	2.684	0.707	45.167	0.064
98	25	5	256	125	0.053	2.737	0.309	45.476	0.062
99	24	12	268	288	0.128	2.866	0.712	46.188	0.059
100	23	19	287	437	0.203	3.069	1.080	47.268	0.057
101	22	12	299	264	0.128	3.197	0.653	47.921	0.054
102	21	20	319	420	0.214	3.411	1.038	48.959	0.052
103	20	14	333	280	0.150	3.561	0.692	49.651	0.049
104	19	14	347	266	0.150	3.710	0.658	50.309	0.047
105	18	20	367	360	0.214	3.924	0.890	51.199	0.044
106	17	29	396	493	0.310	4.234	1.219	52.417	0.042
107	16	21	417	336	0.225	4.459	0.831	53.248	0.040
108	15	24	441	360	0.257	4.716	0.890	54.138	0.037
109	14	40	481	560	0.428	5.143	1.384	55.522	0.035
110	13	37	518	481	0.396	5.539	1.189	56.711	0.032
111	12	37	555	444	0.396	5.935	1.098	57.809	0.030
112	11	58	613	638	0.620	6.555	1.577	59.386	0.027
113	10	63	676	630	0.674	7.228	1.557	60.943	0.025
114	9	72	748	648	0.770	7.998	1.602	62.545	0.022
115	8	97	845	776	1.037	9.036	1.918	64.463	0.020
116	7	105	950	735	1.123	10.158	1.817	66.280	0.017
117	6	173	1,123	1,038	1.850	12.008	2.566	68.846	0.015
118	5	237	1,360	1,185	2.534	14.542	2.929	71.775	0.012
119	4	315	1,675	1,260	3.368	17.911	3.115	74.890	0.010
120	3	590	2,265	1,770	6.309	24.219	4.375	79.265	0.007
121	2	1,301	3,566	2,602	13.911	38.131	6.432	85.697	0.005
122	1	5,786	9,352	5,786	61.869	100	14.303	100	0.002
합계		9,352	40,454	100			100		

표 5. PA Awards 말뭉치에서 건축가들의 건축디자인 어휘 사용행태

빈도 순위	사용 빈도 회수	개별 어휘 개수	누적 어휘 개수	어휘	어휘 사용 회수	전체 어휘종류에 대한 활용비(%)		전체 어휘사용량에 대한 활용비(%)		
						비	누적비	비	누적비	한어휘의 활용비
1	1181	1	1	the	1,181	0.042	0.042	9.958	9.958	9.958
2	477	1	2	a	477	0.042	0.084	4.022	13.980	4.022
3	472	1	3	and	472	0.042	0.126	3.980	17.960	3.980
4	432	1	4	of	432	0.042	0.168	3.642	21.602	3.642
5	274	1	5	to	274	0.042	0.210	2.310	23.912	2.310
6	212	1	6	house	212	0.042	0.252	1.788	25.700	1.788
7	192	1	7	is	192	0.042	0.294	1.619	27.319	1.619
8	188	1	8	in	188	0.042	0.336	1.585	28.904	1.585
9	135	1	9	for	135	0.042	0.378	1.138	30.042	1.138
10	120	1	10	with	120	0.042	0.420	1.012	31.054	1.012
11	109	1	11	on	109	0.042	0.462	0.919	31.973	0.919
12	103	1	12	an	103	0.042	0.504	0.868	32.841	0.868
13	101	1	13	room	101	0.042	0.546	0.852	33.693	0.852
14	92	1	14	as	92	0.042	0.588	0.776	34.469	0.776
15	91	1	15	are	91	0.042	0.631	0.767	35.236	0.767
16	84	1	16	that	84	0.042	0.673	0.708	35.944	0.708
17	82	1	17	space	82	0.042	0.715	0.691	36.636	0.691
18	79	1	18	by	79	0.042	0.757	0.666	37.302	0.666
19	78	1	19	from	78	0.042	0.799	0.658	37.960	0.658
20	72	1	20	living	72	0.042	0.841	0.607	38.567	0.607
21	64	1	21	be	64	0.042	0.883	0.540	39.106	0.540
22	62	1	22	wall	62	0.042	0.925	0.523	39.629	0.523
23	59	1	23	level	59	0.042	0.967	0.497	40.126	0.497
24	57	2	25	area, at	114	0.084	1.051	0.961	41.088	0.481
25	53	2	27	this, two	106	0.084	1.135	0.894	41.981	0.447
26	51	1	28	it	51	0.042	1.177	0.430	42.411	0.430
27	49	1	29	bedroom	49	0.042	1.219	0.413	42.825	0.413
28	47	2	31	which, one	94	0.084	1.303	0.793	43.617	0.396
29	44	1	32	building	44	0.042	1.345	0.371	43.988	0.371
30	36	2	34	floor, client	72	0.084	1.429	0.607	44.595	0.304
31	35	1	35	view	35	0.042	1.471	0.295	44.890	0.295
32	33	1	36	its	33	0.042	1.513	0.278	45.169	0.278
33	32	1	37	or	32	0.042	1.555	0.270	45.438	0.270
34	31	4	41	main, three, garden, site	124	0.168	1.723	1.046	46.484	0.261
35	28	4	45	family, facade, guest, through	112	0.168	1.892	0.944	47.428	0.236
36	27	1	46	roof	27	0.042	1.934	0.228	47.656	0.228
37	26	4	50	kitchen, private, design, into	104	0.168	2.102	0.877	48.533	0.219
38	25	1	51	between	25	0.042	2.144	0.211	48.744	0.211
39	24	3	54	architect, element has	72	0.126	2.270	0.607	49.351	0.202
40	23	4	58	frame, studio, was, will	92	0.168	2.438	0.776	50.126	0.194
41	22	2	60	dining, side	44	0.084	2.522	0.371	50.497	0.185
42	21	4	64	entrance 외	84	0.168	2.690	0.708	51.206	0.177
43	19	8	72	study 외	152	0.336	3.026	1.282	52.487	0.160
44	18	5	77	plan 외	90	0.210	3.237	0.759	53.246	0.152
45	17	5	82	window 외	85	0.210	3.447	0.717	53.963	0.143
46	16	7	89	small 외	112	0.294	3.741	0.944	54.907	0.135
47	15	14	103	storage 외	210	0.588	4.330	1.771	56.678	0.126
48	14	13	116	architecture 가	182	0.546	4.876	1.535	58.212	0.118
49	13	13	129	residence 외	169	0.546	5.422	1.425	59.637	0.110
50	12	8	137	high 외	96	0.336	5.759	0.809	60.447	0.101
51	11	16	153	simple 외	176	0.673	6.431	1.484	61.931	0.093
52	10	10	163	center 외	100	0.420	6.852	0.843	62.774	0.084
53	9	25	188	nature 외	225	1.051	7.902	1.897	64.671	0.076
54	8	29	217	single 외	232	1.219	9.121	1.956	66.627	0.067
55	7	34	251	library 외	238	1.429	10.551	2.007	68.634	0.059
56	6	54	305	. right 외	324	2.270	12.821	2.732	71.366	0.051
57	5	58	363	corridor 외	290	2.438	15.259	2.445	73.811	0.042
58	4	116	479	plane 외	464	4.876	20.135	3.912	77.723	0.034
59	3	174	653	basement 외	522	7.314	27.449	4.401	82.125	0.025
60	2	394	1,047	pond 외	788	16.562	44.010	6.644	88.769	0.017
61	1	1,332	2,379	road 외	1,332	55.990	100	11.231	100	0.008
합		2,379			11,860	100		100		

약 15.6-18.7%에 해당한다.

PA Awards 주택디자인 말뭉치에서의 전체 어휘사용규모는 11,860개, 어휘종류는 2,379개로 조사되었다. 이 2,379개의 규모는 영미권에서 가장 큰 약 50만개의 어휘 규모를 가진 「옥스퍼드영어사전」(OED)<sup>25</sup>과 비교하면 약 0.5%, 영미인의 경우 보거나 들어서 이해할 수 있는 ‘인식어휘’(passive vocabulary)의 규모인 약 5만개 대해서는 약 4.8%에 해당하며, 실제로 사용할 수 있는 ‘사용어휘’(active vocabulary)의 범위 약 2천-5천 개의 범위에 들어간다.

영미권의 여러 지역에서 어휘를 수집한 OED의 특성과 국내 말뭉치에서 조사를 제외한 사실을 고려하면, 상대적으로 국내 주택디자인 어휘가 PA Awards의 것과 비교하여 디자인어휘 어휘종류가 보다 더 풍부하다고 볼 수 있다.

(2) 어휘 빈도순위별 어휘규모

<표 4> 및 <표 5>에서와 같이 어휘빈도 순위는 국내 주택디자인 말뭉치에서는 122위까지, PA Awards의 것에서는 61위까지 조사되었다.

어휘빈도수에 따른 어휘수 그래프를 살펴보면, <그림 1>과 같이 국내주택디자인 및 PA Award 두 말뭉치에서 모두 동일하게 직각에 가까운 ‘L’자 모양의 반비례 관계

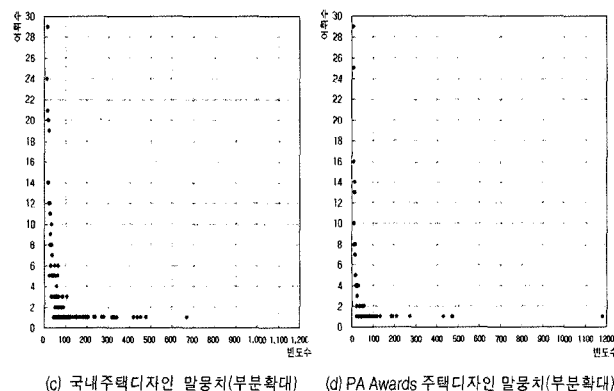
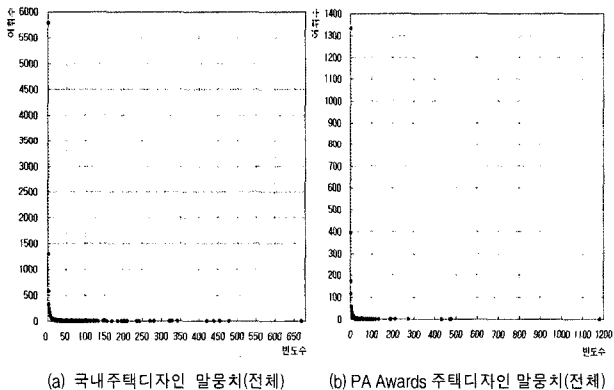


그림 1. 어휘빈도수별 어휘규모

로 나타났다. 부분확대 그림을 보면, 두 말뭉치 모두 어휘사용빈도 1위와 2위 어휘사용빈도 차이는 상대적으로 매우 크고, 어휘빈도수 400-500 사이에서 한 그룹의 어휘가 있으며, 그 이하의 어휘빈도에서는 비슷한 양상으로 어휘수가 증가하고 있다. 국내주택디자인 말뭉치에서는 어휘빈도수 약 100회 이하에서 어휘수가 급격하게 증가하고 있고, PA의 것에서는 약 50회 이하에서 급격하게 증가하는 현상을 보인다.

이 그래프는 주택디자인과 관련한 건축가들의 어휘사용행태는 한정된 어휘규모 안에서 사용빈도가 매우 높은 어휘의 종류는 매우 적고, 사용빈도가 매우 낮은 어휘의 종류는 매우 적은 특성을 보여준다. 즉 주택디자인과 관련한 건축가들의 어휘사용행태는 국어와 영어에서 모두 ‘어휘사용 집중현상’이 매우 심하다는 공통점이 있다. 따라서 이 현상을 고려한다면 고빈도의 어휘들이 저빈도 어휘와 비교하여 상대적으로 그 중요성이 매우 크다고 해석할 수 있다.

‘언어는 언어 사용주체의 의식을 반영한다’<sup>26</sup>는 점을 받아들이고 언어활용에서 어휘집중현상이 각 집단에서 보편적인 현상이라면, 사용빈도가 높은 내용어들은 건축가들의 의식을 잘 반영하는 어휘이며, 집단간 차이가 있을 경우 집단간 특성을 잘 드러내는 중요한 어휘로 볼 수 있다.

(3) 빈도순위에 따른 개별어휘 활용비

빈도순위별 개별어휘의 활용비는 그래프는 <그림 2>에서와 같이, 두 말뭉치 모두 반비례 관계의 비슷한 형태이며, 국내의 경우 빈도순위 약 30위까지, PA의 경우는 빈도순위 약 10위까지 급격하게 감소하다 그 이후 완만하게 감소한다. 이를 통해 두 말뭉치 모두에서 고빈도 어휘의 활용도가 매우 높음을 알 수 있다. 이들 그래프 상에서 나타나는 활용비의 감소 추세식은 그 설명력이 두 집단 모두 98%가 넘는 것으로 나타나, 개별어휘의 활용비 감소는 일정법칙에 따라 감소하고 있는 것으로

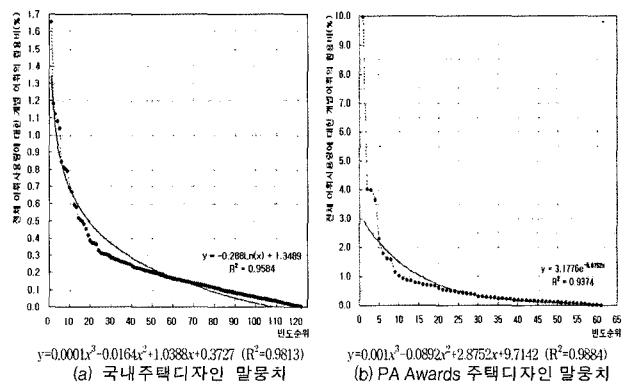


그림 2. 빈도순위별 개별어휘의 활용비 및 추세식

25) OED(TheOxfordEnglish Dictionary)는 영어권, 북미부터 남아프리카까지, 오스트레일리아와 뉴질랜드부터 캐리비안까지의 모든 단어를 포함하며, 옛말까지 포함하고 있다.

26) 사피어(Sapir)와 워프(Whorf)가 세운 ‘언어상대가설(言語相對假說)’에 따르면 언어가 인간의 사고나 사유를 반영할 뿐만 아니라 그 언어를 쓰는 사람들의 사고방식에 영향을 미친다.

해석할 수 있다.

구체적으로 국내 말뭉치에서 1회만 사용된 어휘는 5,786개로 전체 어휘종류의 약 62%이지만, 그 활용비는 전체의 약 14%정도 이다. 그와는 대조적으로 670회 사용된 빈도순위 1위 어휘는 전체 어휘종류의 0.011%를 차지하지만 그 활용비는 전체의 1.7%를 차지하고 있어 빈도수가 1회인 한어휘에 비해 그 활용도가 828배이다.

PA Awards 주택디자인 말뭉치의 경우 1회만 사용된 어휘는 1,332개로 전체 어휘종류의 약 56%이지만, 그 활용비는 전체의 약 11% 밖에 되지 않는다. 1,181회 사용된 빈도순위 1위의 어휘는 전체 어휘종류의 0.042%이지만 그 활용비는 전체의 약 10% 이며 이것은 1회 사용된 한어휘에 비해 그 활용도가 1,245배에 해당한다.

(4) 빈도순위에 따른 어휘활용

빈도순위에 따른 어휘활용 누적비 그래프 <그림 3>은 두 말뭉치 모두 고빈도 어휘에서 어휘활용비 누적값이 급격하게 증가하고 중빈도 어휘에서는 서서히 증가하다가 저빈도 어휘에서 다시 급격하게 증가하는 완만한 'S' 형태이다. 이 곡선도 일정 법칙에 따라 변화하고 있는 것으로 나타났으며 각 추세선의 설명력은 각각 약 96%, 94%로 매우 높다.

국내주택디자인 말뭉치의 경우 빈도순위 상위 10위까지의 고빈도 어휘가 전체 어휘활용비의 약 10%, 20위까지가 약 15%, 50위까지가 약 25%, 100위까지가 약 50% 정도의 활용도를 갖고 있다.

PA Awards 주택디자인 말뭉치에서 빈도순위 상위 5위까지의 어휘가 전체 어휘활용비의 약 25%를, 10위까지가 약 30%, 23위까지가 약 40%, 40위까지가 약 50%, 50위까지가 60%정도의 활용도를 갖고 있다<sup>27)</sup>. 국내와 비교하여 PA의 고빈도 어휘활용도가 높은 것은 고빈도 어휘에 기능어(the, a, an 등)들이 많고 그 활용도가 매우 높기 때문이다.

(5) 누적어휘수에 따른 어휘활용비

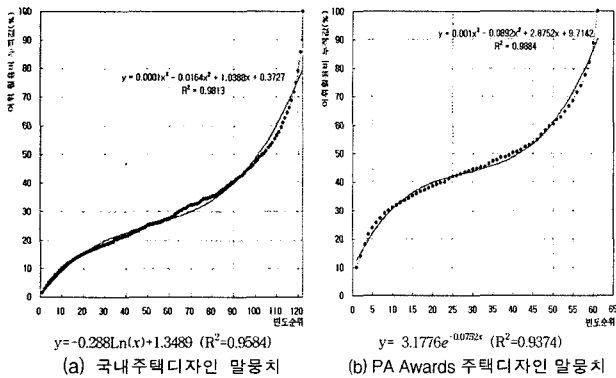


그림 3. 빈도순위에 따른 어휘활용 누적비 및 추세식

27) 영어를 텍스트로 사용한 다른 연구에서 밝힌 빈도순위 50위까지의 단어가 차지하는 비가 60%를 차지한다는 결과(조명환, 앞의 책, p.78.)와 비교하면 동일한 결과이다.

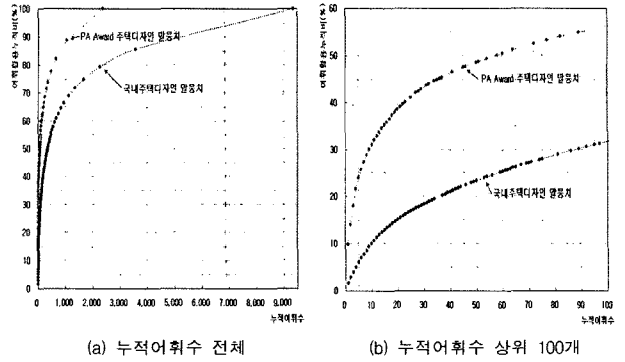


그림 4. 누적어휘수(상위빈도순)에 따른 어휘활용 누적비

국내 및 PA Awards 주택디자인 말뭉치에서 상위빈도 어휘를 기준으로 한 누적어휘수에 따른 어휘활용비 그래프 <그림 4>는 개략적으로 상위 10개 어휘까지 급격하게 증가하고, 20개까지 그 증가세가 둔화되고, 30위부터는 완만하게 증가하는 비슷한 형태를 보인다.

국내의 경우 빈도수가 높은 상위 10개 어휘가 약 10%, 20개가 약 15%, 30개가 약 18%, 50개가 약 23%, 90개가 약 30%의 활용비를, PA의 경우 상위 10개의 어휘가 약 31%, 20개가 약 38%, 30개가 약 43%, 60개가 약 50%의 활용비를 보이고 있다. 두 말뭉치를 비교하면, 동일 누적어휘수일 때 PA의 어휘활용도가 국내의 것보다 높으며, 두 말뭉치에서 상위 약 30개부터 두 말뭉치에서의 활용비 차이는 25% 내외로 거의 일정하다.

2. 주택디자인에서의 기본어휘

지금까지의 유사연구에서 기본어휘 선정은 일반적으로 빈도수를 기준으로 하여 선정되어 왔다. 즉 기본어휘는 말뭉치 내에서 활용도가 매우 높은 즉, 상대적으로 사용빈도가 높은 어휘들이다. 그러나 일반적인 어휘사용행태에서 고빈도와 저빈도 어휘를 제외하고서는 사용빈도 및 활용도가 연속적이기 때문에 기본어휘를 구분하는 정확한 기준을 찾아내기 어렵다.

국내 및 PA Awards 주택디자인 말뭉치로부터 조사된 최상위 빈도의 기본어휘 50개는 <표 6>과 같다. 두 말뭉치에서의 기본어휘 대부분은 명사<sup>28)</sup>와 기능어<sup>29)</sup>들로 나타났다. 이들 기본어휘는 국내 주택말뭉치에서는 약 23%, PA에서는 약 49%의 어휘활용도를 갖고 있는데, 이것은

28) 건축디자인과 관련하여 명사 이외에 관심을 둘 수 있는 것으로 서술어휘에 해당하는 형용사와 동사가 있으며 이들 어휘들에 대해 보다 자세하게 연구될 필요성이 있다. 국내주택디자인 말뭉치에서 빈도수가 높은 형용사로는 '작은', '느낄수있는', '넓은', '좋은', '다른' 등이, 동사로는 '위하다', '주다', '두다', '만들다', '따르다' 등이 조사되었다. PA Awards 주택디자인 말뭉치에서 형용사로는 'main', 'private', 'large', 'low', 'used' 등이, 동사로는 'make', 'lead', 'take', 'allow', 'represent' 등 조사되었다.

29) 기능어(function word)는 말과 말 또는 문장과 문장 사이에서 문법적인 관계를 나타내는 어휘들로, 국어에서는 조사, 어미, 접속사 등이, 영어에서는 관사, 정관사, 전치사, 관계대명사, 조동사, 의문사, 대명사 등이 속하며, 국어가 영어에 비해 기능어가 발달되어 있다.



표 6. 국내 PA Awards 주택디자인에서의 기본어휘 50개

구분	말뭉치	
	국내주택디자인	PA Awards 주택디자인
어휘	공간, 수(불완전명사), 것, 층, 주택, 이다, 이(지시대명사), 집, 있는, 대지, 있다, 거실, 실, 이(2), 그(지시대명사), 방, 건물, 형태, 정원, 건축주, 등(etc), 쪽, 생활, 내부, 형(形), 감(感), 면(面), 주거, 축, 지붕, 벽, 부분, 외부, 하(였)다, 계단, 자연, 도로, 우리, 가족, 대한, 식당, 현관, 곳, 또한, 일(1), 그리고, 평면, 계획, 설계, 있도록	the, a, and, of, to, house, is, in, for, with, on, an, room, as, are, that, space, by, from, living, be, wall, level, area, at, this, two, it, bedroom, one, which, building, floor, client, view, its, or, main, garden, site, three, family, facade, guest, through, roof, kitchen, private, design, into
활용도	23.3%	48.6%(33.8%*)

\*관사(the, a, an)를 제외하였을 경우의 활용도

근본적으로 국어와 영어의 문법적 구조 차이에 기인하는 것으로 보인다. 즉, 관사와 같은 기능어들이 영어 말뭉치에서 고빈도 어휘를 차지하고 있기 때문으로, 이 차이를 고려하면 활용도 차이는 줄어든다.

기본어휘를 단순히 어휘빈도가 높은 어휘로 볼 경우 기능어들이 많이 포함되기 때문에, 주택디자인과 관련하여 건축가들의 의식을 잘 반영하는 의미있는 어휘들을 찾아내기 위해서는 기본어휘에서 기능어를 제외하고 내용어 중심으로 살펴볼 필요가 있다. 본 글에서는 내용어<sup>30)</sup> 중심의 고빈도 기본어휘를 ‘주요기본어휘’로 정의한다.

두 말뭉치에서 빈도순위 약 50위까지의 어휘들 가운데 주요기본어휘에 해당하는 어휘들을 약 10개씩 묶어 3그룹으로 나누어 정리하면 <표 7>과 같다. 이들 주요기본어휘들의 개별 어휘활용비는 주요기본어휘 순위 1, 2위를 제외하고서는 <그림 2>의 결과와는 다르게 국내의 것이 PA보다 높게 나타났고, 전체 활용비에서도 국내의 것이 약 2% 정도 높게 나타났다.

주요기본어휘 순위에 따른 어휘활용비는 <그림 5>와 같이, 두 말뭉치 모두에서 거의 동일한 형태의 감소추세를 보여준다. 이를 설명하는 두 집단의 추세선 설명력(R<sup>2</sup>)이 모두 97% 이상으로 매우 높게 분석되었는데, 이는 주요기본어휘의 순위에 따른 어휘활용도는 언어가 달라도 일정한 계량적 법칙처럼 활용되고 있다고 해석할 수 있다.

주요기본어휘에서 1, 2 순위의 어휘활용비가 3위 이하의 어휘들과 비교하여 어휘활용비가 상대적으로 매우 높은 것으로 나타났으며, 1, 2위의 어휘는 PA가, 나머지 순위의 어휘들은 국내의 것들이 어휘활용도가 상대적으로 높게 나타났다.

<표 7>에서 찾아진 주요기본어휘들을 대상으로, 두 집단 모두에서 공통적인 것, 국내주택디자인 주요기본어휘에만 있는 것, PA Awards 주택디자인 주요기본어휘에만

표 7. 국내 및 PA Awards 주택디자인 말뭉치에서의 주요기본어휘

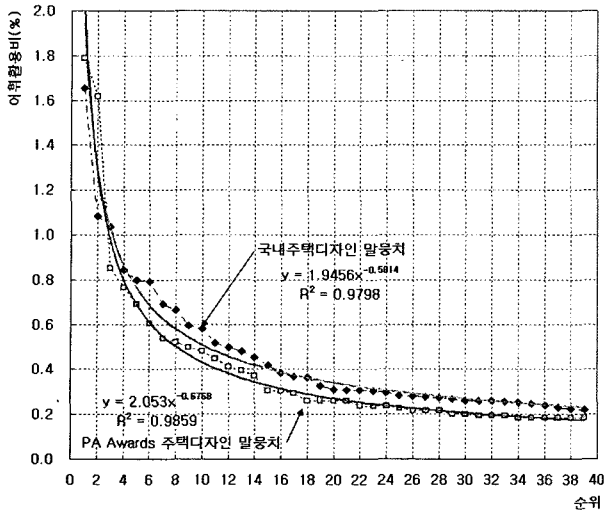
그룹	번호	국내주택디자인 말뭉치			PA Awards 주택디자인 말뭉치		
		빈도 순위*	주요 기본어휘	어휘 활용비**	빈도 순위*	주요 기본어휘	어휘 활용비**
A	1	1	공간	1.656	6	house	1.788
	2	4	층	1.083	7	is	1.619
	3	5	주택	1.038	13	room	0.852
	4	6	이다	0.845	15	are	0.767
	5	8	집	0.801	17	space	0.691
	6	9	있는	0.791	20	living	0.607
	7	10	대지	0.692	21	be	0.540
	8	11	있다	0.667	22	wall	0.523
	9	12	거실	0.596	23	level	0.497
	10	13	실	0.581	24	area	0.481
소계			8.750	소계		8.365	
B	11	14	이(2)	0.517	26	two	0.447
	12	16	방	0.499	29	bedroom	0.413
	13	17	건물	0.482	30	one	0.396
	14	18	형태	0.455	32	building	0.371
	15	19	정원	0.418	33	floor	0.304
	16	20	건축주	0.386		client	0.304
	17	22	쪽	0.368	38	view	0.295
	18	23	생활	0.363		main	0.261
	19	24	내부	0.329		garden	0.261
	20	25	형(形)	0.311		site	0.261
21	26	감(感)	0.309	three	0.261		
소계			4.437	소계		3.574	
C	22	27	면(面)	0.307	42	family	0.236
	23	28	주거	0.304		facade	0.236
	24	29	축	0.297		guest	0.236
	25	30	지붕	0.284	46	roof	0.228
	26	31	벽	0.282	47	kitchen	0.219
	27	32	부분	0.277		private	0.219
	28	33	외부	0.274		design	0.219
	29	34	하(였)다	0.267	52	element	0.202
	30	35	계단	0.262		architect	0.202
	31		자연	0.262	55	frame	0.194
32	도로	0.262	studio	0.194			
33	38	우리	0.257	was		0.194	
소계			3.335	소계		2.579	
합			16.522	합		14.518	

\*말뭉치 내에서의 빈도순위/\*\* 전체 어휘사용량에 대한 개별어휘 활용비

있는 것으로 구분하여 정리하면 <표 8>과 같다.

<표 7>에서 현 단계에서 의미파악이 어려운 동사를 제외하고서, 주택디자인과 관련 최상위 빈도의 주요기본어휘 10개를 살펴보면, 국내에서는 ‘공간’, ‘층’, ‘주택’, ‘집’, ‘대지’, ‘거실’, ‘실’ 등이, PA Awards에서는 ‘house’, ‘room’, ‘space’, ‘living’, ‘wall’, ‘level’, ‘area’ 등의 어휘들로 나타났다. 주택디자인 텍스트 수집대상이 대부분 2층 단독주택이었던 점을 고려하여 ‘주택’, ‘층’, ‘house’

30) 내용어(content word)는 실질적인 의미 내용을 가진 어휘들로써 명사, 형용사, 동사 등이 해당한다.



국내 :  $y = 1.9456x^{-0.5614}$  ( $R^2 = 0.9798$ )  
 P A :  $y = 2.0530x^{-0.6768}$  ( $R^2 = 0.9859$ )

그림 5. 국내 및 PAAwards 말뭉치에서 주요기본어휘의 어휘활용비

등의 어휘를 제외하면, 주택디자인과 관련하여 국내 건축가들이 가장 많이 사용하는 의미있는 어휘는 ‘공간’, ‘대지’, ‘거실’, PA에서는 ‘room’, ‘space’, ‘wall’, ‘level’이다.

이 같은 결과로부터 주요기본어휘 가운데 주택디자인과 관련하여 두 집단의 건축가들이 가장 많이 사용하는 대표 어휘는 ‘공간(space)’이며, 국내건축가들이 상대적으로 중시하는 어휘는 ‘대지’, PA Awards 건축가들이 상대적으로 중시하는 어휘는 ‘wall’과 ‘level’이다.

주요기본어휘수를 확대한 <표 8>을 보면, 주택디자인과 관련하여 국내 건축가들은 ‘대지’, ‘정원’, ‘외부’, ‘자연’, ‘도로’와 관련된 것과 ‘느낌(感)’, ‘계단’에 많은 중점을 두고 있고, 이와는 다르게 PA Awards 건축가들은

표 8. 두 집단간 주택디자인 주요기본어휘

국내에만 있는 것	주택디자인 주요기본어휘			PA에만 있는 것
	국내	PA	활용비 차이*	
쪽, 생활, 내부형, 감(感), 면(面), 축, 부분, 외부, 계단, 자연, 도로, 우리	공간	space	(+)0.965	level, area, one. view, main. three, family, facade, guest, kitchen, private, design, element, architect, frame, studio
	층	floor	(+)0.779	
	주택, 집, 주거	house	(+)0.355	
	대지	site	(+)0.431	
	실, 방	room, bedroom	(-)0.185	
	이(2)	two	(+)0.070	
	건물	building	(+)0.111	
	정원	garden	(+)0.261	
	건축주	client	(+)0.082	
	지붕	roof	(+)0.056	
벽	wall	(-)0.214		

\* (+)는 국내주택디자인 말뭉치의 활용비가 높음  
 (-)는 PA주택디자인 말뭉치의 활용도가 높음을 의미  
 \*\* 거실, living 어휘는 현 단계에서 공통어부 확인이 불가능해 제외

‘wall’, ‘level’, ‘view’, ‘facade’, ‘frame’과 ‘private’ 한 것에 상대적으로 많은 관심을 두고 있는 것을 알 수 있다.

이러한 사실로부터 국내 건축가들은 주택디자인 과정에서 주택과 관련한 외부 요소들과의 관계와 느낌을, PA Awards 건축가들은 실질적인 건축요소를 상대적으로 중시하고 있다고 추론할 수 있다.

건축가들이 주택디자인과 관련하여 자발적으로 사용한 자연언어에서 찾은 주요기본어휘들은 건축가들이 주택디자인 과정에서 매우 중요하게 의식하고 있는 것들을 반영한 것으로, 설문 등을 통해 그 중요도를 측정된 결과와는 다른 보다 중요한 가치를 지닌 것으로 평가할 수 있다.

### V. 맺는 말

본 글은 국내 주택디자인 및 PA Awards 주택디자인과 관련하여 건축가들이 쓴 텍스트를 수집하여 대규모 말뭉치를 구성하고 이를 계량적으로 분석한 것으로, 그에 대한 주요 연구 결과는 다음과 같다.

첫째, 주택디자인과 관련하여 사용한 건축가들의 주택디자인 어휘규모는 국내의 것이 9,352개, PA의 것이 2,379개였다. 두 언어의 특성 차이를 고려하더라도 국내 주택디자인 어휘종류가 PA보다 많다고 볼 수 있다. 조사된 어휘규모는 실제 사용어휘규모 약 5만개에 대해 국내 및 PA 각각 약 18.7% 및 4.8%에 해당하는 규모이며, PA의 조사 결과는 영어의 사용어휘 규모 범위 안에 있어, 본 연구의 결과는 타 연구의 결과를 벗어나지 않았다.

이러한 결과로부터, 국어가 영어보다 건축가들이 사용하는 주택디자인 어휘가 보다 풍부하며, 그 활용 가능성에서도 높은 잠재력을 갖고 있다고 볼 수 있다.

둘째, 건축가들의 주택디자인 어휘사용행태는 국내 및 PA에 모두에서 고빈도 어휘수는 매우 적지만 그 활용비는 매우 높고 그와는 반대로 저빈도 어휘수는 매우 많지만 그 활용비는 매우 낮은, ‘어휘사용 집중현상’이 매우 심한 것으로 나타났다. 인간이 언어를 수단으로 사고하고 인식한다는 점을 받아들인다면, 이 어휘사용 집중현상은 주택디자인과 관련하여 기본어휘가 상대적으로 매우 중요한 가치를 갖고 있음을 의미한다.

셋째, 어휘사용행태를 반영하는 빈도순위별 개별어휘의 활용비, 빈도순위에 따른 어휘활용 누적비, 순위별 주요기본어휘의 어휘활용비 등의 현상을 설명하는 추세식의 설명력(R<sup>2</sup>값)은 모두 90% 이상으로 매우 높게 나타났다. 이 결과는 비록 언어가 다르더라도 건축가들의 집단적인 어휘사용행태는 일정한 계량적 법칙에 따라 이루어지고 있음을 보여준다. 전체 어휘활용비는 PA의 어휘가 국내의 것보다 높지만 기능어를 제외한 내용어 중심의 주요기본어휘 활용비는 국내 주택디자인 어휘가 높았다.

넷째, 거의 대부분이 명사인 최상위 빈도 10위 이내의 주요기본어휘들은 국내에서는 ‘공간’, ‘층’, ‘주택’, ‘집’, ‘대지’, ‘거실’, ‘실’ 순이었고, PA에서는 ‘house’, ‘room’,

‘space’, ‘living’, ‘wall’, ‘level’, ‘area’ 순이었다. 이들 어휘로부터 두 집단 모두 가장 많이 사용하는 어휘는 ‘공간(space)’이며, 두 집단의 차이를 잘 반영하는 어휘는 ‘대지’와 ‘wall’이었다.

다섯째, 주요기본어휘 분석 결과는 국내 건축가들은 주택디자인 과정에서 ‘대지’, ‘자연’, ‘도로’ 등과 같은 주택과 관련한 외부 요소들과의 관계와 정원 그리고 ‘느낌(感)’을, PA Awards 건축가들은 ‘wall’, ‘level’, ‘view’, ‘facade’, ‘frame’과 같은 실질적인 건축요소와 ‘private’한 점을 상대적으로 중시하고 있는 것으로 해석할 수 있다.

본 글은 문치언어학의 방법론을 사용하여 주택디자인에 적용한 것으로, 이 분야에서의 시작단계에 해당하는 기초 연구이다. 따라서 이와 관련한 후속 연구로 언어사용에서 역할이 다른 명사, 형용사, 동사에 대한 분석, 기본어휘에서 드러난 중요성이 높은 고빈도 어휘들의 실제 문장 안에서 활용과 그 의미, 그리고 다른 범주의 말뭉치와 비교분석이 진행되어야 할 필요성이 있다.

## 참 고 문 헌

1. 김광해, 「국어 어휘론 개설」, 집문당, 1993.
2. 박광배, 「빈도분석」, 성원사, 1992.
3. 박영환, 「말뭉치에 기반한 형태소 분석기 및 철자 검사기의 구현」, 연세대석사학위논문, 1991.
4. 이상섭, “날말빈도를 추정하기 위한 말뭉치 자료수집의 실제”, 「인문과학」, 연세대인문과학연구소, 1989.
5. 이상섭외, 「우리말 낱말 빈도 조사 표본의 선정기준」, 새 한국어 사전편찬을 위한 사전편찬학 연구, 제3집, 탑출판사, 1990.
6. 조명환, 「언어심리학」, 민음사, 1991.
7. 정영미, 「국어 어휘의 통계적 특성과 이의 응용」, 인문과학, 연세대인문과학연구소, 1993.
8. 허재혁, 「미국 현대주택 건축의 디자인 주제와 표현 수법에 관한 조사 연구」, 연세대석사학위논문, 1995.
9. 홍성민, 「교회건축물에 대한 사용자의 인식특성에 관한 연구」, 대한건축학회논문집, 2005, 2.

(接受: 2006. 1. 5)