

어휘부 구성의 기본 문제

홍재성

1. 머리말

현금의 언어학 논의에서, 또 언어현상에 대한 학제적 논의에서, 어휘연구의 중요성은 점점더 커져가고 있다. 특히 사전학의 성립이나, 전산과학의 응용 분야인 자연언어처리 영역의 발전은 어휘연구의 필요성을 새로이 부각시켰다. 순수한 언어학연구의 테두리 내에서 본다면, M. Gross나 I. Mel'čuk은 어휘부의 실질적인 구성을 자연언어의 언어학적 기술의 중심으로 삼는 중요한 업적을 쌓아 왔다. 통사론 우위의, 통사론 중심의 언어이론이기에 상대적으로 어휘부에 대한 성찰이 소홀했던 생성언어학에서도 이른바 지배 결속이론의 대두 이후 어휘부에 대한 적극적인 언급을 볼 수 있고, 생성언어학의 변종인 일반구조문법이나 어휘·기능문법등에서도 어휘기술이 상당한 중요성을 띄게 되었다.

한편, 구조의미론 영역에서는 어휘의미의 구조적 분석이 시도되고, 이에 근거한 다양한 어휘관계의 설정문제에 대해 꾸준한 연구가 진행되어 왔는데, 그 성과는 현대적 어휘연구의 중요한 기반이 될 것이다.

이 자리에서는 어휘부 구성이 제기하는 몇몇 기본 문제를 일련의 질문 형식으로 열거하여 토론의 단초를 제공하고자 한다.

2. 어휘부 구성의 목표

어휘부 기술을 둘러싼 이론적, 실제적 문제와 그 해결의 논의는, 어휘부의 유형이나 그 구성목표와 긴밀히 상관되어 있음은 자명하다.

어휘부를, 한 언어의 총체적 표상모형의 일부로서, 그 언어에 속하는 가능한 모든 어휘요소와 그와 상관되는 언어적 속성을 망라해서 표상하는 부분으로 규정할 때 어휘부의 유형은 여러 가지로 구분될 수 있을 것이다. 몇 경우만 예를 들어 본다면 다음과 같다.

- ㄱ. 일반언어사전
- ㄴ. 어휘데이터베이스
- ㄷ. 순수한 언어학적 기술로서의 어휘부

근. 자연언어처리 체계의 일부로서의 전산사전

코. 어휘에 대한 지식의 심적 표상으로서의 어휘부

어떠한 형태의 어휘부를 구상하든지 간에 공통적으로 제기되는 일련의 기본적인 문제들이 있겠는데, 우리는 논의 중심을 순수한 언어학적 기술로서의 어휘부구축에 두고 이론적, 방법론적 성찰 또는 선택이 요구되는 몇몇 문제를 제시해 본다.

3. 어휘부 구성의 기본문제

3-1. 언어기술 모형 내에서의 어휘부의 위상

우선, 한 언어에 속하는 것으로 판단될 수 있는 어휘의 총체를 기술하려고 할 때 제기되는 근본적인 문제는 언어의 전반적인 표상모형내에서 어휘부의 위상을 어떻게 설정하는가 하는 것이 되겠다.

여기에 대해서는 크게 두 가지 입장이 있을 것이다. 첫째는, 대부분의 언어학자들이 수용하는 입장으로, 기본적으로 어휘부를 통사부/의미부/형태부/음운부등과 구분하여 그 자율성을 가정하면서, 총체적 모형 내에서 각 부위를 상관시키고 그 상호관련성을 논의하는 것이다. 이러한 가정 내에서도 어휘부의 중요성을 어느 정도 확대시킬 것인가 하는 문제는 이론에 따라 달라지는 것이다. I. Mel'čuk의 의미·텍스트 대응모형은 언어기술의 각 부분을 성층적으로 상관시키면서 어휘부의 중요성을 극대화한 독특한 이론의 예로 볼 수 있다.

둘째는, 어휘부 중심으로, 더 정확히 말한다면, 어휘부와 통사부를 융합시킨 부분을 중심으로, 언어 표상의 모든 국면을 통합하려는 입장이 되겠는데, M. Gross의 어휘 문법은, 형태부/어휘부/통사부의 구별을 두지 않고, 전산어휘부 내에 모든 것을 통괄하려는 시도의 예가 된다. 이러한 시도의 타당성 여부는 흥미있는 논의의 대상이 될 수 있을 것이다.

3-2. 어휘부의 내적구성

3-2-1. 총체적 어휘부 구성의 요구

일반언어사전의 편찬작업을 논의로 한다면, 그간의 언어학적 어휘연구는, 한 언어에 속하는 어휘의 총체를 망라해서 기술을 시도한 경우는 지극히 드물고, 대체로 어휘부의 구성을 위한 이론내적인 원칙이나 원리에 대한 논의에 집중되거나, 이론적 논의의 정당화를 위한 증거 또는 예시으로써 지극히 제한된 어휘부의 단편을 기술했을 뿐이라 해도 과언이 아닐 것이다. 이러한 관점에서 보았을 때, 우선 하나의 개별언어를 대상으로 총체적 어휘부의 구축을 목표로 하는 외연주의적이고 총망라주의적인 어휘연구의 필요성을 제

기할 수 있을 것이다. 이러한 연구는 그 순수한 언어학적 연구로서의 가치도 근거있게 따져질 수 있을 뿐 아니라 언어학적으로 바탕이 든든한 새로운 언어사전의 편찬이나, 자연언어처리 체계에 긴요한 전산사전의 구성을 위해 필수적인 작업으로 여겨진다.

이렇게 총체적 어휘부의 실질적인 구축을 시도한다면, 언어학자들은, 재래의 사전편찬자들이 직면했던 여러 곤혹스러운 문제들을 나름대로 가능한 언어학적으로 근거있게 해결하여야 할 처지에 놓이게 될 것이다.

이제 이와 같은 관점에서 문제를 열거해 보기로 한다.

3-2-2. 어휘목록

어휘부 구축의 가장 기초적 작업중의 하나는 어휘부를 구성하는 항목으로서의 어휘요소를 목록화하는 것이다. 목록화는 특히 어휘요소, 일반사전 편찬작업에 유추해서 표현한다면, 표제어의 배열문제, 표제어의 동형어/다의어 처리의 선택의 문제를 포함한다. 이와 관련된 더욱 근본적인 문제는 일정 시기의 한 언어에 속하는 어휘요소의 총체를 현 단계의 언어학적 논의에 바탕을 둔 기준으로 낱말이 열거하여 엄격히 한정하는 작업이 가능한 것인가 또 그것이 필요한 것인가 하는데 대한 논의일 것이다. 이러한 논의와 병행해서 대강 다음과 같은 질문이 제기된다.

1. 단일어/복합어(합성어와 파생어)를 모두 열거하여 목록화 할 것인가? 또는 단일어만 목록화하고 복합어(특히 파생어)는 조어규칙으로 생성되는 것으로 처리하여 규칙만 명시하는 것으로 그칠 것인가?

2. 단일어/복합어를 모두 목록화하되, 관례적인 일반언어사전에서와 같이 단일목록화하여 자모순서에 의한 배열방식(또는 재배열 regroupement 등 그것에 조금 변경을 가한 배열방식)에 따라 순서화하고 단층적으로 표상할 것인가? 공통핵집어휘/전문어휘/고유명사 등으로 구분될 수 있는 하위어휘부류에 대해서도 똑같은 질문이 제기된다. 전통적인 인쇄언어사전의 경우는 이 모든 하위부류를 단일목록화하고 단층적으로 표상한 대표적 예가 된다. 그러나, 대부분의 언어학자들의 이론적 논의에서 주장되듯이(김영석 교수의 앞의 글 참조) 순수 언어학적 어휘부의 구축을 위해서는 이를 모두 분할하고 어휘부 전체가 적절한 다층적/성층적 내적 구조를 갖도록 하는 선택이 고려되어야 할 듯하다.

우리가 여기서 개인적 견해로 덧붙여 지적하고 싶은 점은 어휘부의 구성을 위해서는 합성어와 속어표현을 낱말이 총망라해서 목록화하고, 또 형식적인 기준으로 유형화하는 것이 필요하리라는 것이다.

3. 어휘요소의 목록화작업은 주어질 어휘요소의 동형어/다의어 처리의

선택문제와 직결되어 있다. 부분적으로 어휘목록의 규모와 내용이 이 선택에 따라 달라지는 것은 두말할 여지가 없다.

3-2-3. 어휘항목 기술의 내용

어휘부 구성을 시도할 때 제기되는 가장 실질적인 문제는 바로 어휘항목 기술의 내용이다. 음운/발음표시, 의미표시, 품사구분이나 하위범주화 속성 등을 기록할 필요성에 대해서는 대부분의 언어학자들이 합의의 보고 있는 내용이지만 각각의 사항의 구체적인 점에 대해서는 논란이 많은 것이 사실이다. 예컨대 의미표시를 형식적 논리언어로 할 것인가, 자연언어를 그대로 메타언어로 사용할 것인가, 또는 의미원소를 가정하고 자질의 집합으로 표상할 것인가 하는 등의 문제에 대하여 심각한 이견이 공존하고 있음이 확인되는 것이다. 하위 범주화 속성의 표시에 있어서도 주어에 대한 정보의 표상 여부를 놓고, Chomsky 계열의 생성언어학과 그 외의 언어모형을 택하는 학파사이에 큰 대조가 있는 것이다. 이와 같은 문제들은 하나 하나 깊이 있고 세밀히 검토되어야 함은 물론이다. 우리는 여기서 간단히 한 가지 점만 덧붙여 지적하는데 그치고자 한다. 그것은 어휘항목 기술의 내용중에 포함되어야 할 어휘·통사 정보는, 품사구분과 같은 어휘범주나 하위범주화 자질, 또는 어휘인여규칙 등에 의거하여 표시하고자 했던 어휘요소 사이의 규칙적인 관계에 관련된 속성에 그칠 것이 아니라, 그것들을 넘어서서 훨씬 더 정밀하고 다양하여야 하리라는 점이다. 조금만 더 구체적으로 언급하면, 일반적인 원리 또는 규칙으로 표상할 수 없는 (그 동안 언어학 논의에서, 제한된 자료에 의거하여 성급하게 제시되었던 일반화된 규칙들이, 사실은 제한된 어휘요소에 묶인 국부적 속성임이 드러난 경우가 많았음을 상기하기 바란다), 따라서 예측할 수 없는 어휘적, 통사적 속성은, 언어학적으로 근거 있고 일관성있는 방법으로 낱낱이 표상되어야 할 것이다. 그런데, 사실은 이와 같은 속성들이 상당한 양에 이르고, 또 그것들이 언어사용자가 내재화하고 있는 어휘지식의 꽤 중요한 부분을 이루고 있어, 어휘부 표상에서 빼놓을 수 없는 것이다. (이와 관련된 구체적 속성들의 예는 M. Gross나 I. Mel'čuk의 어휘 기술에서 볼 수 있다.)

3-3. 어휘부 기술의 형식

마지막으로 지적할 문제는 어휘항목의 기술대상이 되는 언어적 속성을 어떠한 형식으로 표상하는가 하는 것이다. 초기 변형생성문법 이후 최근의 생성언어학에 이르기까지 공통적으로 채택되고 있는 자질체계에 의한 표상방법이 그 주요한 한 선택인데, 또 다른 방식, 예컨대 행렬식 도표, 또는

network와 같은 그래프에 의거한 표상방식의 타당성과 잇점도 충분히 검토해 보아야 할 것이다.

4. 맺는 말

이상에서 우리는, 한 언어의 구체적인 어휘부 구성을 시도할 때, 어떤 이론적/방법론적 틀을 택하건 간에 기본적으로 만나게 될 여러 문제들을 아주 거칠게 나열해 보았다.

결론을 대신하여 한국어 어휘연구를 위한 한 가지 제언을 해 본다면, 그것은 대규모의 어휘목록을—예컨대, 동사어휘 총체, 형용사어휘 총체, 또는 보문명사나 시간명사의 총체를—대상으로 각 개별 어휘요소의 언어적 속성을 총망라해서 낱낱이 체계적으로 연구하는 작업이 이제라도 시작되어야 하지 않을까 하는 것이다. 우리는 이러한 연구를 방법론상으로 총망라주의적, 외연주의적 연구라 특징짓겠는데, 한국어 어휘부의 실질적인 구성시도가, 언어학적 관점에서나, 사전학 또는 전산과학의 응용적 관점에서나 중요하고 시급하다고 판단되는 이 시점에서, 우리는 바로 이러한 유형의 어휘연구의 필요성을 강조하고 싶은 것이다.

참 고 문 헌

- 조항범, 1989. 국어 어휘론 연구사, 국어학 19. 국어학회.
- Andrews, A.D. 1988. Lexical Structure, F. Newmeyer (ed.), *Linguistics: The Cambridge Survey II. Linguistic Theory: Extensions and Implications*, Cambridge University Press.
- Corbin, D. 1987. *Morphologie dérivationnelle et Structuration du lexique*, Max Niemeyer Verlag.
- DRLAV 38, 1988. *Lexique-Nouveaux Modèles*.
- Evens, M.W. (ed.), 1988. *Relational Models of the Lexicon*, Cambridge University Press.
- Gross, M. 1975. *Méthodes en syntaxe*, Paris: Hermann.
- _____. 1988. Sur la structure des articles d'un lexique-grammaire (manuscrit).
- _____. 1990. Lexiques électroniques pour l'analyse automatique des langues naturelles (manuscrit).
- Hudson, R. 1984. *Word Grammar*, Basil Blackwell.
- _____. 1988. *The Linguistic Foundations for Lexical Research and Dictio-*

- nary Design, *International Journal of Lexicography* 1:4, Oxford University Press.
- Mel'čuk, I. 1988a. *Dependancy Syntax*, New York University Press.
- _____. 1988a. Paraphrase et lexique dans la théorie linguistique Sens-Texte, *Lexique* 6.
- _____. et al. 1984. *Dictionnaire explicatif et combinatoire du français contemporain*, Les Presses Universitaires de l'Université Montréal.
- Le Français dans le Monde. 1988. *Lexiques*, numéro thématique-Recherches et Applications.
- Langue Française 69. 1986. *Syntaxe du nom*, Larousse.
- Langue Française 87. 1990. *Dictionnaires électroniques du français*, Larousse.
- Lexicographica*, 1988:4. Max Niemeyer Verlag.
- Lexique 7. 1988. *Lexique et syntaxe en grammaire générative*, Presses de Université de Lille.
- Sabah, G. 1988-89. *L'intelligence artificielle et le langage*, 2 vols, Paris: Hermès.

151-742

서울시 관악구 신림동

서울대학교 불어불문학과