

(09_22)조합형과 완성형

조합형은 한글의 창제 원리인 초성, 중성, 종성의 개념을 살려서 한글이 사용할 수 있는 비트를 세 개로 쪼개서 초성, 중성, 종성에게 각각 할당하는 방법이다. 완성형은 한글 한 글자를 각각 독립된 개념으로 생각하고 각 글자에 코드를 부여하는 방법이다. 완성형과 조합형은 각자의 장단점을 가지고 치열했던 논쟁을 일으켰다. 조합형은 한글의 창제 원리를 고스란히 담고 있고, 조합이라는 방식 덕에 모든 한글을 표현할 수 있다는 장점이 있다. 하지만 코드를 글자로 변환하기 위해서는 2바이트를 비트 단위로 쪼개서 해석을 해야하기 때문에 처리상의 부담이 있었다. 반면 완성형은 숫자가 글자를 그대로 의미하기 때문에 비트 단위 해석의 부담은 없었지만 한글의 창제 원리를 무시하고 마치 한글을 한자처럼 취급한다는 단점이 있었다. 오히려 완성형을 자소 단위로 쪼개기 위해서는 추가적인 해석 부담이 있다. 결론적으로 보았을 때, 당시에만 해도 PC의 메모리가 그다지 넉넉한 편이 아니었기 때문에 완성형이 국가표준으로 되어 있었고, 조합형이 태생적으로 각 컴퓨터 제조사들의 자체 한글표기 포맷으로 발달해 왔기 때문에 완전히 호환되지 않는 독자 규격들이 난립하다가 결국 세를 통합하지 못하고 조합형은 비주류로 남게 되었다. 표준 완성형 인코딩 방식의 다른 이름은 EUCKR이다. EUCKR은 KS X 1003과 KS X 1001으로 구성된다. KS X 1001에는 한글 외에도 그림 문자와 한자 등의 글자가 많기 때문에 그렇지 않아도 작은 이 영역에 들어가지 못한 한글이 생겼다. 바로 “똥”이나 “뿔” 등이다. 자주 사용되지 않는 몇몇 글자들을 EUCKR에서 사용할 수 없다는 점이 어떤 사람들에게는 별 문제가 아니었지만 어떤 사람들에게는 큰 문제였다. 예를 들어 고어를 많이 사용하는 국어학자라거나, 채팅이나 소설에서 독특한 표현(?)을 사용하는 사람들에게 EUCKR은 족쇄나 마찬가지였다. 이 때 등장한 것이 마이크로소프트의 CP949 즉, 코드 페이지 949이다. CP949는 마이크로소프트가 KS X 1001에 없는 한글 8822글자를 추가해서 EUCKR을 확장한 완성형 인코딩 방식이다. 그래서 CP949는 확장 완성형 또는 통합형 한글 코드라고도 불리운다. CP949를 도입해서 어느 정도 논쟁이 정리되었으나, CP949에 대해서도 논란이 있었다. MS에서는 완성형의 문제를 이전부터 인식하고 있었고, 표준 완성형의 한계를 넘기 위해 글자를 새로 정리하려 했으나 역시나 기존 인코딩과의 호환을 넘을 수 없어 추가되는 문자를 기존 완성형 코드가 사용하지 않는 영역에 추가했다. 이 분쟁이 완전히 사라진 것은 유니코드 2.0이 등장한 이후다. 유니코드 1.x대 버전에도 조합형 한글은 들어가 있었으나, 완성형 한글은 모든 글자를 수록하지 않는 등 엉망이었다. 2.0으로 버전이 올라가면서 한글 영역이 대이동, 완성형 형태의 조합가능한 11172자의 모든 한글 음절이 모두 들어갔다. 즉 유니코드를 쓰면 완성형이든 조합형이든 모두 호환이 가능하다는 것. 이 때문에 조합형 완성형 논쟁은 더이상 무의미하게 되었다. 그리고 간단한 수식을 통해 초중종성 분리가 가능한 유니코드의 특성상, 유니코드는 조합형의 특성에 더욱 가깝다고 볼 수 있다. 그리고 실제로 미리 지정된 11172개 코드 말고도 초성, 중성, 종성을 분리해 코드로 나열한 실제 과거 조합형과 매우 유사한 형태도 정식 유니코드 활용 법으로 인정되기도 한다.