

SPACING PROBLEMS FOR A KOREAN SPELLING CHECKER¹⁾

Byung-Soo Park

It is generally felt among native and foreign speakers of Korean alike that one of the most difficult parts of the Korean writing system is spacing (띄어쓰기). Actually, it is not exaggerating to say that newspapers and magazine are in a state of chaos when it comes to the problem of spacing.

As I see it, the main source of the present confusing practice in spacing comes from nominal compounds. For example, what is the correct way of writing the following compound?

- (1) a. 대학 수학 능력 시험
- b. 대학수학능력시험
- c. 대학수학 능력 시험
- d. 대학수학 능력시험
- e. 대학 수학능력 시험
- etc.

If any of these is acceptable for some reason or other, can anyone say that there is a rule of spacing? The longer the compound, the more serious the problem becomes, and the problem gets deteriorated to the extent that the solution seems dubious:

- (2) a. 대학 수학 능력 시험 출제 위원
- b. 대학수학능력시험출제위원
- c. 대학수학능력시험 출제위원
- etc.

- (3) a. 대학 수학 능력 시험 출제 교수 선정 작업
- b. 대학수학능력시험출제교수선정작업
- c. 대학수학능력시험 출제교수 선정작업
- etc.

1) 경희대학교 문리대 영어영문학과, 언어교육연구원 원장.

What does the Unified Korean Writing System (한글 맞춤법 통일안) say about this problem? There IS a rule of spacing there, but unfortunately, it gives us no definite answer. Let me quote the rule:

(4) Article 61

Words should be written with a space between them, except for particles which should be put together with their previous words without a space. However, words are permitted to be put together in a suitable manner if contextually necessary.

The rule consists of two parts: the main rule and its qualification. The main rule is simple and impeccable. The qualification is most vague on two accounts: it is not clear at all what the two phrases "IN a suitable manner" and "if contextually necessary" mean. Unless "the suitable manner" and the "contextual necessity" are defined unambiguously, the rule say nothing definite. This inevitably leads Korean writers to the chaotic situation about spacing in the Korean writing system. For the rule of spacing as it is now allows one to write the above compounds in any one way shown in (1), (2) and (3).

Now, this serious problem would be a disaster for those who attempt to design a Korean spelling checker for the computer. In designing the spelling checker, what we must have first of all should be a Korean dictionary, that is, a list of Korean words. It is generally agreed that the number of words in a language is finite at a given time (although the list should tend to become longer and longer throughout the time). So it is theoretically possible to have a list of all words in a language at our disposal. In contrast, the number of compounds is infinite, because making compounds with two or more nouns is an infinitely productive process. Observe the following process:

- (5) 수학능력
- 수학능력평가
- 수학능력평가준비
- 수학능력평가준비작업
- 수학능력평가준비작업준비
- 수학능력평가준비작업준비작업
-etc.

For this reason, it would be theoretically impossible to make a complete list of

compounds in a language.

In a writing system in which compounds are allowed to be written without spaces between their component words as in (b), (c), or (d) above, each individual compound is to be taken as a word. (Let us call such a practice of writing compounds "spaceless writing" (붙여쓰기) as opposed to writing with spaces" (띄어쓰기).) Then, in order for our spelling checker to be able to check the spellings of the words, its dictionary should contain all possible compounds in addition to all possible words in Korean. But this would never be possible because the list of Korean words would be infinitely long just as the list of all Korean compounds is infinitely long. This leads us to conclude that it is unrealistic to design a computer spelling checker which will be able to check the spelling of compounds which are written in the spaceless writing system.

Now, let us remind ourselves of where this insurmountable difficulty comes from. It all starts from the particular way in which compounds are written; that is to say, the difficulty is a result from the fact that spaceless writing is permitted in writing compounds. Thus the solution to the difficulty seems to me to be almost obvious. Disallow spaceless writing in writing compounds, and all the difficulty will disappear. If we always keep the rule of "writing with spaces" whether we write words or compounds, the spelling checker equipped with the list of all Korean words will work quite well. Then we should rewrite (5) as follows:

- (6) 수학 능력
 수학 능력 평가
 수학 능력 평가 준비
 수학 능력 평가 준비 작업
etc..

Now in designing a computer spelling checker, we will not have to worry about compounds; we will only take of words. For example, in order to check the spelling of the words in (5), the spelling checker would need a dictionary containing all the compounds as independent lexical entries in addition to the list of the five words 수학, 능력, 평가, 준비, and 작업. Contrastively, in checking a document written in a manner as shown in (6), the spelling checker will work well enough with a dictionary containing only the five words. The difference between the two systems will indeed be striking.

Yet, this obvious solution does not work in the present writing system, in which spaceless writing is permitted "if necessary". Therefore, I propose that the present writing system be modified to the effect that spaceless writing is not

permitted in any case. My proposal means to discard the qualification of the Article 61 in (4). My proposal will make the rule of "writing with spaces" stronger and as a result, we will have a more consistent writing system as far as the practice of "writing with space" is concerned. The ambivalence or sloppiness, which is a characteristic of the present practice of "writing with spaces," should disappear once and for all. Thus, my proposal does not only make it possible to design a realistic Korean spelling checker, but also will it lead to a consistent and unified practice of "writing with spaces" to which everyone will agree.

REFERENCES

- 한글 학회 「한글 맞춤법 통일안」 1991
이 익섭 「국어 표기법 연구」 1992