우효맨 1 - 한국어 특성(현 논문에서 사용하는 내용만 추가)

한국어에는 여러가지 특성이 있는데 그 중 대표적으로 자모 단위 구성과 형태소, 높임표현이 있다. 먼저 한글은 음소 문자로, 자음과 모음을 나타내는 글자가 따로 존재한다. 이 자음과 모음을 합쳐서 음절 단위로 사용하기에, 음절 문자적 성격을 지니고 있으며, 이는 한글은 글자 단위로 나누고, 나뉜 상태에서 자모 단위로도 분리가 가능함을 의미한다.

* 한국어의 자모(자모에 대한 예시는 안넣어도 될 듯..? 내용만 많아짐 일단 추가)

자음:ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ, ㄲ, ㄸ, ㅃ, ㅆ, ㅉ

모음:ㅏ, ㅑ, ㅓ, ㅕ, ㅗ, ㅛ, ㅜ, ㅠ, ㅡ, ㅣ, ㅐ, ㅒ, ㅔ, ㅖ, ㅘ, ㅙ, ㅚ, ㅝ, ㅞ, ㅟ, ㅢ

한국어는 의미를 가진 최소 단위인 형태소로 분리가 가능하다. 형태소는 크게 의존 형태소와 자립 형태소로 구분되며, 의존형태소 중 종결어미, 자립 형태소 중 감탄사 등에서는 높임 표현이 존재하며, 특히 종결어미의 경우, 격식체 비격식체로 분리할 수 있다. 본 논문에서는 높임말 반말 변환 시 문장을 형태소 단위로 분석 후 특정 높임표현을 내포한 형태소를 분석 후 어체를 변환하는 방식을 따른다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **격식체** | | | | **비격식체** | |
| **하십시오** | **하오** | **하게** | **해라** | **해요** | **해** |
| **평서형** | -(ㅂ니)다 | -오 | -네 | -(는/ㄴ)다 | -어요 | -어 |
| **감탄형** |  | -(는)구려 | -(는)구먼 | -(는)구나 | -어요 | -어 |
| **의문형** | -(ㅂ니)까 | -오 | -(느)ㄴ가 | -(느)냐 | -어요 | -어 |
| **명령형** | -(ㅂ)시오 | -오 | -게 | -어라/아라 | -어요 | -어 |
| **청유형** | -(ㅂ)시다 |  | -세 | -자 | -어요 | -어 |

(optional)

* 용언의 활용

용언(동사와 형용사)의 어간이 어미와 결합하는 것을 ‘활용’이라고 하며 이떄 어간이나 어미의 모습이 변하거나 변하지 않는 경우가 있다. 이때 특정한 규칙에 의해 변하는 용언을 규칙 용언, 형태 변화 양상이 예측되지 않는 경우, 불규칙 용언이라고 한다. 본 논문에서는 어체 간 변환 시 용언의 활용을 고려하여 문법적으로 적절하도록 설계하였다.