

한국어 작성에 의한 일본어 작문지원시스템 설계

장원, 여상화, 심철민, 김태완

Japanese composition system design based on writing Korean

Won Chang, SangHwa Yhu, Chulmin Sim, Taewan Kim

본 논문은 일본어 작문지원시스템에 관한 것으로서, 사용자가 일본어에 대한 언어지식이 전혀 없더라도 한국어 작성 능력만으로 일본어를 작문할 수 있도록 지원하는 일본어 작문지원 시스템 개발 방법에 관한 것이다.

본 논문에서 제안하는 일본어 작문지원 시스템은, 사용자가 일어로 작문하고자 하는 문장을 한국어의 어절단위 또는 연속된 문자열로 작성해 나가면, 시스템이 자동개입하여 한국어 분석기술과 일본어로의 변환/생성 기술로 일본어의 대응 문자열을 생성해 준다. 따라서 한국어 작성능력만으로 일본어를 작문할 수 있도록 지원한다. 특히, 한국어 분석의 중의성과 일본어 변환시 대응어 선택의 중의성 문제가 있을 경우에는, 사용자가 시스템이 사용자 대화창에 제공하는 한국어 도움말 정보를 참조하여 선택함으로써 문제점을 해소하여 정확한 작문을 유도한다.

1. 서론

근래 국가간의 정보교류가 급속히 증가함으로써, 언어장벽을 해소하기 위한 노력이 꾸준히 진행되어 왔다. 그 노력의 결과로 국내에서도 영어,일본어,중국어,러시아를 우리말로 번역하거나 우리말을 영어,일어로 번역하는 시스템들이 상용화 단계에 이르렀다. 또한 각종 전자 사전과 작문지원, 문서 요약, 다언어 정보검색 시스템들이 상품화되어 있다. 이들은 모두 국가간의 언어 장벽을 해소하기 위한 훌륭한 도구가 될 것이다.

그러나, 아직까지는 각 시스템들의 기능과 품질면에서 사용자의 요구를 충족하지 못하고 있다.

작문지원시스템의 종래 기술은, 사용자가 매체에 기록된 인사말, 답변서, 사과문 등과 같은 분류에 따른 상황별 예문, 예문을 주제어로 분류한 주제별 인덱스, 구절단위의 예제, 특정 양식에 따른 정형표현의 자료, 또는 단어, 숙어, 관용구 등의 사전류를 참조해서 직접 외국어를 작문하여 입력한다[1][2]. 이러한 지원 방법은 사용자에게 기술적인 외국어의 언어지식을 컴퓨터상에 제시해 주는 것에 불과하다.

본 논문은 일본어를 전혀 모르는 사용자라고 하더라도 한국어를 이해하고 한국어 문장을 작성할 수 있다면 시스템의 개입에 의해 한국어와 일본어간의 언어정보처리기술로 일본어를 반자동으로 생성해 주는 시스템의 개발 방법에 대해 기술한다.

이하에서 문자열이라 함은 하나의 형태소 또는 복수개의 형태소로 이루어진 복합형태소를 지칭하는 것이다. 따라서, 문자열은 복수개의 어절을 포함하기도 한다.

2. 종래의 작문지원기 또는 번역시스템과의 차이점

종래의 작문지원기[1][2]들과 본 논문에서 제안하는 방법의 작문지원기를 비교하면 다음과 같은 차이가 있다.

1. 종래의 방법은 한국어의 작성 능력 뿐만 아니라 외국어에 대한 일반적인 지식이 있어야 한다.
2. 종래의 방법에서는 사용자가 필요한 지식을 스스로 찾아 참조하고 입력하기 까다로운 일본어를 직접 입력해야 한다. 이러한 방법은 작성시간의 지연과 사용자가 많은 요소들을 결정하여 조합하여야 하는 번거로움이 있다.
3. 사용자의 능력에 따라 작문 오류의 여지가 많고, 스스로 검증하기 어렵다는 점이다.

그리고, 한일번역시스템[3]으로 번역함으로써 간접적으로 작문효과를 얻을 수 있는데, 번역시스템과의 차별성을 기술하면,

1. 번역시스템의 번역 품질이 의사교환에 쓰일 정도로 정확하고 실용적이지 못하다.
2. 기계번역시스템은 대량의 문서 내용을 빠른 시간에 번역하기 위한 도구이며, 본 작문지원시스템은 상대적으로 정확하고 실용적인 일본어 문장 작성과 함께 일본어에 대한 언어지식을 습득할 수 있다는 차이가 있다.

3. 시스템의 구성

본 시스템의 구성은 일본어에 미숙한 사용자가 한국어 문장 작성 능력만으로 실용적인 일본어를 작문할 수 있도록 지원하며, 그 과정의 반복으로 자연스럽게 일본어에 대한 언어적 지식을 습득해 가는 학습효

과를 갖도록 시스템을 구성한다.

특히, 한국어의 중의성이나 일본어로 변환시 대응어 선정에 중의성 문제가 있는 것들에 대해서는 지원 시스템이 분석/변환 기술을 이용하여 사용자에게 문제점을 한국어 도움말 정보와 함께 제시하고 사용자가 이를 참조하여 선택하므로써, 양 국어간의 언어적 대응 문제점들을 해결한다. 또한 일본어의 한자, 히라가나, 가다가나, 특수기호 등의 자종 또는 코드의 문제점을 사용자가 의식하지 않고 작문할 수 있도록 구성함으로써 완성도가 높은 일본어의 실용문장 작성을 반자동으로 지원할 수 있도록한다. 아래 시스템 구성도를 소개한다.

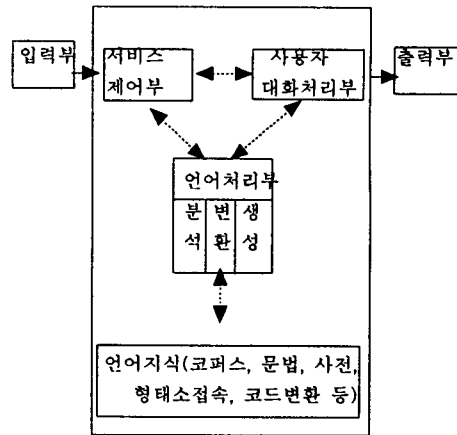


그림 3.1 시스템 구성도

3.1 서비스 제어부

서비스제어부는 서비스 요구사항에 따른 수행을 위한 환경설정과 초기화 등의 준비를 하는 제1과정과, 사용자가 한국어 작성을 시작하면 제1과정에서 기억된 서비스 요구사항 조건에 따라 입력 문자열에 대한 형태(예: 어절단위, 다른 자종의 시작 또는 문장 부호 등)를 검사하여 서비스 개입여부를 판단하는 제2과정과, 대상 문자열을 언어처리부에 전달하고 언어처리의 단계별 결과를 보존/관리하며, 언어처리부와 사용자 대화처리부간의 중의성해소와 미등록어 처리, 대응어 선택 문제에 대한 데이터의 결정 및 변경사항 처리를 제어하는 제3과정의 업무를 수행한다. 또한 사용자가 작문 과정에서 일부를 취소하거나 변경했을 경

우의 처리와 제어를 하는 제 4과정으로 분류할 수 있다. 제3과정의 업무는 언어처리부와 사용자대화 처리부의 기술에서 상세히 설명한다.

3.2 언어처리부

언어처리부는 한국어 분석기와 한국어로부터 일본어로의 변환기, 일본어의 형태소 생성기로 구성된다.

3.2.1 한국어 분석기

한국어 분석기에 의해 분석된 문자열의 분석정보는 형태소, 의미, 자종 정보를 포함하며 이들 정보는 분석 사전으로부터 얻어진다.

표3.1 한국어 분석정보

1. 형태정보	강+을+건너+다+
2. 의미정보	공간(장소)명사+ 조사'을' + 이동성자동사
3. 자종정보	한글+한글+

한국어 분석기는 한국어 좌우접속정보테이블과 코퍼스로부터 추출된 통계치에 의해 부여된 빈도, 형태소의 길이를 고려한 Left-to-Light 온라인 분석방법이다.

본 시스템에 필요한 분석기의 기능은 다음과 같다.

1. 입력은 문장단위가 아닌 사용자의 서비스 개입시점에 대한 요구사항에 따라 어절 또는 임의 크기의 문자열이 된다. 예를 들면 '그는', '할 수 있는', '-삽입문-' 또는 문장기호와 문장기호의 사이의 문자열 형태이다.
2. 시스템은 Left-to-Right On-line 분석 방법으로 개입시점에서 전회에 분석되어 저장된 분석결과와 속성을 기반으로 우측에 올 수 있는 한국어 형태소의 분석을 진행한다.

3. 또한, 대화에 의한 중의성해소 방법으로, 분석 결과에 다품사, 다의어 등의 중의성이 있는 단어는 사용자대화 처리부에 의뢰해서 사용자로부터 결과를 선택할 수 있도록 하므로써 분석기의 오분석 또는 복수개의 분석결과를 배제한다. 예를 들면, '나는 나는 새를'이라는 문자열에 대해 대명사 '나'와 동사 '날다'의 선택을 사용자 대화처리부에서 사용자로부터 결정을 유도한다.
4. 사용자가 입력하고 있는 문자열로부터 이후 조합될 수 있는 문자열(후보단어)이 사전에 존재하면 분석결과와 함께 출력하여 사용자가 입력 예정 문자열을 선택할 수 있도록 한다. 만약 사용자가 후보단어를 선택하였다면 사용자입력창에는 선택한 문자열을 부가하고, 변환과 생성을 거쳐 일본어출력창에 이에 대응하는 일본어 문자열이 부가될 것이다.

step1:

사용자입력창	
나는 *갈	*금회입력분

일본어출력창	
私は	

사용자대화창	* 사용자가
① 가+리	②를 선택.
*② 가+리 수밖에 없	*②,③,④는
③ 가+리 필요가 있	후보단어임
④ 가+리 수 없	

step2:

사용자입력창	
나는 갈 수밖에 없	

일본어출력창	
私はいくしかな	

그림 3.2 입출력창과 사용자 대화창

상기의 그림 3.2를 참조하여 설명하면, 사용자가 서비스개입시점을 어절단위로 선택하였고 현재, 시스템의 한국어 입력창에 '나는'이라는 한국어 문자열과 이에 대한 작문으로 일본어 출력창에 '와다 구시와'와 같이 출력되어 있다고 하자. 이후 사용

자가 '갈'이라는 어절을 입력하였다면 서비스 제어부는 어절단위가 입력되었으므로 언어처리부에 처리를 의뢰한다. 언어처리부의 한국어 분석기는 동사어간 '가'와 관형형어미 'ㄹ'과 함께 'ㄹ 수밖에 없' 또는 'ㄹ 필요가 있' 또는 'ㄹ 수 없'과 같은 후보단어를 빈도 순으로 출력하여 사용자 대화처리부에 사용자 선택을 의뢰한다.

만약 사용자가 'ㄹ 수밖에 없'를 선택했다면 시스템의 한국어 입력창은 '나는 갈 수밖에 없'으로 변경되고 일본어출력창에도 '갈 수밖에 없'에 해당하는 일본어 문자열이 추가된다. 따라서, 사용자는 한국어 작성중에 시스템으로부터 사용빈도가 높은 입력 후보단어를 제공받을 수 있다. 또한 사용자의 선택에 따라 후보단어들의 빈도를 시스템이 자동으로 변경한다.

- 미등록어의 처리에 있어서는 사용자 대화처리부에 의뢰하여 고유명사 처리를 할 것인가 또는 대응하는 문자열을 재입력받을 것인가를 결정한다. 예를 들면, '한빛초등학교'를 고유명사 로 처리할 것을 시스템에 요구할 수 있다. 또 '고르다'라는 동사가 미등록어일 경우 시스템은 재입력을 요구할 수 있으며, 사용자는 '선택하다'와 같이 재입력하므로써 시스템의 미등록어 문제를 해결할 수 있다.

3.2.2 한-일 변환

변환기는 분석기의 출력인 분석정보를 입력하여 변환문법을 수행한다. 변환문법은 한국어와 일본어간의 대조 분석 자료를 코드화한 것이다.

분석정보에서 자종정보는 주로 영숫자와 기호 등의 무변환 대상 문자열 정보로 이용된다. 변환시 복수개의 대응어 문제점이 있을 경우 변환기는 각각의 대응어에 대한 한국어 도움말 정보를 부가하여 사용자대화처리부에 사용자의 선택을 의뢰함으로써 문제점을 해결한다.

표 3.2 한일어간 대조분석

	한국어	일본어
1. 일반동사	공간(장소)명사+ 조사(을) 이동성자동사(건너다)	장소명사+ 은+ 渡る
2. 일본어 특정어법의 동사	탈것(명사)+ 조사(을)+ 타다	탈것+ に+ 乗る
3. 형용동사와 '명사+이다(하다)	이였(하였)겠지 이(하)다면 이였(하였)다면	だったろう ならば だったら

3.2.3 일본어 형태소 생성기

한-일변환과정에서 출력되는 일본어 문자열과 생성코드의 쌍을 받아 자연스러운 일본어 활용언의 시제,태 등을 고려한 어간과 어미의 적절한 결합을 생성해 낸다.

한국어와 달리 일본어는 띄어쓰기하지 않으며 한국어와 어순이 거의 같다. 그러나 한국어와 다른 어순을 갖는 형태도 있으므로 이에 대한 처리와 활용언의 어간과 어미의 적절한 생성이 주된 처리이다. 일본어 형태소생성기는 문자열 간의 일본어 좌우접속정보테이블의 정보에 따라 처리한다.

표 3.3 일본어 형태소생성을 위한 좌우접속정보테이블 형태

우측 형태소 \ 좌측 형태소	동사가 정형	...	형용사 어미 ?형	관정사 ?형	사행 변격 ?형
오단동사(라행)		...			
오단동사(나행)		...			
....
모음동사		...			
사변명사		...			
명사		...			

일본어 좌우접속정보테이블은 한국어에 대응하는 일본어 문자열의 좌우 접속정보를 분류하여 특정 좌접속정보와 우접속정보가 조합되는 규칙을 정보화한 테이블이다.

3.3 사용자대화처리부

한국어 분석결과에 중의성과 대응어 선정의 문제점이 있는 경우, 이 문제점을 해결하기 위하여, 선택 후보 리스트(통계치에 의한 우선순위와 후보리스트에 없음을 나타내는 선택사항포함)를 작성하고 사용자와 대화를 통하여 사용자가 결정한 리스트 항목을 언어처리부에 전달하는 업무를 담당한다.

또한 미등록어일 경우, 사용자에게 그 단어에 상응하는 다른 단어의 입력을 제공받거나 고유명사처리를 의뢰받아 이에 필요한 데이터와 제어를 해당처리부에 넘겨주는 역할을 한다.

4. 시스템의 처리 흐름

이하에서는 첨부된 그림 4.1을 중심으로 본 시스템의 처리흐름을 상세히 설명한다.

그림 4.1에서 P1,P2,P3,P12,P13은 서비스제어부의 역할이며, P6,P10은 사용자대화처리부의 역할이며, 나머지는 언어처리부의 역할이다.

P1. 사용자는 작문을 시작하기 전에 출력할 일본어의 코드(Shift-JIS, EUC 등)명을 지정하거나, 또는 서비스 개입 시점(상황)은 어떤 단위로 지정할 것인가를 선택한다. 한국어는 어절단위의 띄어쓰기를 하므로 어절단위로 또는 문장부호를 경계로 또는 한자어의 출현등의 상황에 따라 서비스받는 시점

을 적절히 정할 수 있다. 시스템은 서비스에 관련된 환경을 설정한다.

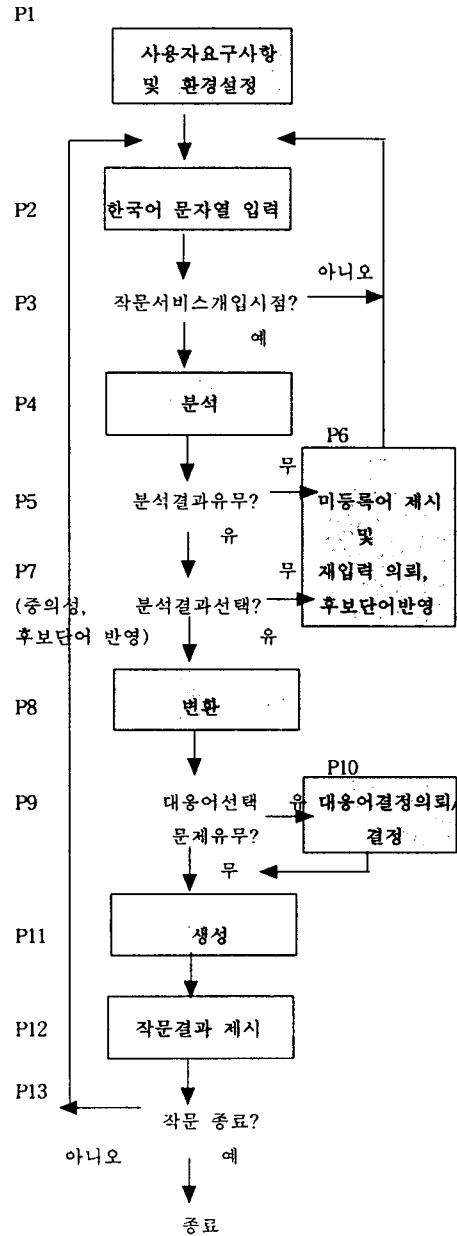


그림 4.1 시스템 흐름도.

P2. 이제 사용자가 한국어 작성을 시작한다.

- P3. P1에서 기억된 서비스 요구사항에 따라서, 서비스 개입시점 파악을 위해 A2의 문자열에 대한 형태를 검사하여 서비스 개입여부를 판단한다. 서비스 개입조건을 충족하면 대상문자열을 언어처리부에 보낸다. 개입시점이 아닐 경우에는 사용자의 입력이 계속된다.
- P4. 언어처리부는 대상문자열을 분석부에 의뢰한다. 분석부는 문자열을 분석하여 결과를 출력한다. 분석결과는 접속가능한 것들과 접속가능한 후보단어도 포함한다.
- P5. 분석결과가 하나일 경우에는 확정하여 다음 단계인 변환으로 넘겨준다. 분석결과가 없을 경우 사용자대화처리부(P10)에 미등록어처리 또는 사용자 재입력을 의뢰한다.
- P6. 분석결과가 없거나 예비후보도 존재하지 않을 경우가 있을 수 있는데, 이것은 명백한 미등록어이므로 미등록어로 제시하고 사용자에게 고유명사로 처리하거나 이에 상응하는 다른 단어를 요구하여 미등록어 문제를 해소한다.
- P7. 분석결과에 중의성이 있을 경우 또는 후보단어가 있을 경우, 사용자는 하나의 분석결과 또는 후보단어를 선택한다. 이때 사용자 대화창에는 분석결과에 통계치에 의한 우선순위와 적절한 분석결과가 없다는 선택사항을 포함한다. 또, 분석결과는 없지만 후보단어가 존재한다는 것은, 문자열의 형태가 서비스요구 범위에 들어 있지만 긴 단어 또는 복수개의 어절로 이루어진 숙어/관용구일 가능성이 있다.
- 예를 들어, 전단계에서 동사‘하’에 대해서 처리하였고 이번에 ‘르’를입력하는 중에 시스템이 개입하였다. 만약 분석사전에 ‘르 수 있’이나 ‘~르 수 없’과 같은 형태가 등록되어 있다면 분석기는 이 후보단어들을 사용자대화처리부에 주어 사용자로부터 선택할 수 있도록 한다.
- P8. 중의성이 해소된 분석결과를 받아 변환을 한다.
- P9. 대응어 선택의 문제점이 있을 경우는 사용자 대화처리부(A10)에 의뢰하여 사용자로부터 선택받는다. 문제점이 없을 경우 바로 생성으로 넘긴다.
- P11. 전번회에 생성된 결과와 현재 생성해야할 문자열에 대해서 생성을 위한 좌우접속정보 테이블의 생성규칙에 따라 일본어 문자열을 생성한다.
- P12. 현재까지 작문된 결과를 출력한다. 이때 현재까지 처리된 한국어에 대한 분석결과도 볼 수 있다. 사용자가 문장분석에 관여하였기 때문에 스스로 점검할 수 있는 방편으로 오작문을 방지하기 위해 필요할 것이다.
- A13. 이 시스템은 반드시 문장단위의 작문에 구애받지 않으므로 사용자가 언제든지 종료할 수 있다.
- 종료후의 사용자 요구사항 처리는 생략한다.

5. 결론

본 논문에서는 일본어에 미숙한 사용자에게 한국어와 일본어간의 자연언어처리 기술을 통하여 일본어의 작문을 지원하고, 특히, 한국어 분석의 중의성이나 일본어 대응어 선택의 중의성, 미등록어의 문제를 사용자의 한국어 지식으로 모두 처리할 수 있게 하므로써, 일본어의 언어지식에 관계없이 완성도가 높은 대화형 일본어 작문지원서비스를 할 수 있는 시스템 구현방법에 대해서 기술하였다.

본 시스템과 같은 구현방법의 역으로 '일본어 작성에 의한 한국어 작문시스템'을 설계할 수 있으며, 이 시스템의 도메인으로 특정 전문분야 또는 비즈니스 관련 작문을 목표로 했을 때 더욱 실용가치가 있으리라고 생각한다. 또한 이 방법을 이용하여 한일간 채팅시스템의 개발이 가능하다.

[참고 문헌]

1. 깊은말 정보사, やわら일본어(일본어 작문지원S/W)사용자 매뉴얼, 1997년
2. 3A Software, AUTO-Pro Business Writer (영작문지원S/W)사용자 매뉴얼, 1997년
3. 드림 씨앤씨, 드림KJ SeoulTokyo (한일번역S/W)사용자 매뉴얼, 1997년
4. 시스템공학연구소, 한일기계번역시스템개발에 관한 연구(3) 연구 보고서, 1991년
5. 시스템공학연구소, 일한기계번역시스템 개발 연구(2) 연구보고서, 1996년
6. 시스템공학연구소, 한일어 대조비교분석(1) 연구보고서, 1988년
7. 이주행, 현대국어문법론, 대한교과서 주식회사, 1993