

다국어 정보처리를 위한 불어 전처리에 관한 연구

서 래 원[†] · 박 세 영^{††} · 유 성 준^{†††}

요 약

본 논문에서는 다국어 정보 처리의 일환으로 불어 정보처리의 기반이 되는 형태소 분석기의 효율성을 높이는 방법에 대해 고찰하였다. 형태소 결합 현상으로 생성된 어절을 분리해 줌으로써 불필요한 단어 유형의 추가를 막아주고, 형태소 분석 알고리즘의 부하를 줄여 주고자 하는 목적으로 문의 전처리의 필요성을 강조하였다. 문의 전처리 중 어절의 형태 분리에 관한 형태 분리 규칙을 개시하였으며, 이 분리 규칙은 복문을 단문으로 분리 시켜주는 단문화 분리 규칙 포함하고 있다.

Preprocessing of the French Sentence for Multilingual Information Processing

Lai-Won Seo[†] · Se-Young Park^{††} · Seong-Joo Yoo^{†††}

ABSTRACT

The purpose of this paper was to present the method of progress efficiency of morphosyntactical analyzer for French information processing in the view of multilingual information processing. This study indicated that the method prohibit the additional useless word type by decomposing words which were created by morphological amalgamation. Findings also suggested the need of preprocessing in order to decrease the overload of morphosyntactical analysis algorithm. In addition, general rules were proposed to divide word form and to decompose complex sentence.

1. 서 론

정보 창출 및 소비를 위한 정보기기 보급의 가속화와 더불어 인터넷과 같은 정보 기반시설에 대한 지속적인 투자가 이루어지고 있으며 그 결과 오늘날의 사용자들은 국경과 지리적인 한계를 극복하고 다양한 정보들을 신속히 획득할 수 있게 되었다. 그러나 그와 같은 풍성한 물적 조건에도 불구하고, 유통되고 있는 정보들이 표현된 언어의 이질성으로 인해 사용자들은 원하는 정보에 대한 검색에서 뿐만 아니라 획득한 정보의 이용에서도 여전히 많은 어려움을 겪고 있다. 따라서 보다 효율적인 정보 검색과 활용을 위해서는 틀

적 기반뿐만 아니라 다국어 정보처리 기술 개발에 대한 많은 관심과 투자가 시급한 형편이다. 이와 같은 맥락에서 이루어지고 있는 다국어 정보처리 기술 개발의 일환으로, 본 연구에서는 불어 형태소 분석기의 효율성을 높이는 방법에 대해 고찰하고자 한다.

자연언어 처리에 있어서 가장 어려운 것이 되고 선행되어야 할 것은, 분석대상이 될 문(文)의 전처리 작업이다. [3] 문의 전처리는 형태소 분석기의 효율성을 높이기 위한 방법으로, 형태소 분석에 들어가기 앞서 행해지는 전처리 과정에 해당된다. 문의 전처리는 크게 텍스트를 분석의 기본 단위인 문장으로 절단하는 것과 절단된 각 문장을 개별언어의 특성에 맞게 다시 단어형태로 절단하는 것을 목표로 한다.

불어 문장에서는 /au/의 경우에서와 같이 두 개의 이상의 형태소들이 형태 결합 및 변화 과정을 거친 후

† 중신회원 배재대학교 컴퓨터공학부 교수
†† 정 회원 한국전자통신연구원 지식정보연구부 부장
††† 정 회원 한국전자통신연구원 지식정보연구부 지식정보집계 연구팀 팀장, 책임연구원
논문집수 1999년 10월 11일, 심사완료 : 2000년 3월 11일

단일 어절을 구성하는 현상을 자주 관찰할 수 있다. 따라서 불어 문장의 전처리 작업에는 이러한 현상에 대한 처리가 필수적이라 하겠다. 본 연구에서는 불어 문장의 전처리 작업에 필수적인 불어의 형태소 결합 현상을 연구하고 그의 해결 방법을 제시하고자 한다

2. 불어 형태소 결합 특성

불어에서 나타나는 형태소 결합현상 중 가장 대표적인 예로서 전치사 /à [-에]/와 정관사 /le/가 결합하여 구성된 축약관사 /au/를 들 수 있다. 형태소 결합 현상으로는,

- 연결 형태(concatenation form)
정관사 /le/와 관계사 /quel[어떠한]/이 결합되어 복합 관계사 /lequel/이 된과 같이, 단순히 두 단어가 연결되어 구성된 형태
예) /lequel/, /auquel/, /duquelles/, ...
- 축약 형태(elision form)
두 단어 사이의 모음이 탈락되면서 apostrophe가 추가되어 결합되는 형태.
예) /s'il/, /d'une/, /c'est/, /s'aime/, /l'on/, ...
- 결합형태(hyphenated form)
자음이나 모음의 탈락이 없이 단순히 hyphen으로 연결되는 형태.
예) /là-dedans/, /ci-contre/, /aime-je/, ...
- 아말감 형태(amalgamation form)
앞에서 본 축약관사의 예에서와 같이 두 단어가 아말감 현상에 의해 결합된 후 새로운 형태의 어절로 변환된 형태.
예) /au/, /du/, /aux/, /des/, /aucun/, ...
등 4 가지로 분류할 수 있다.

축약관사 /au/와 같은 경우의 예는 단일 기능에 기반한 기존의 어떠한 문법범주에도 속하지 않는다. 축약관사를 위한 새로운 문법범주를 만들지 않으려면 축약관사 /au/는 /à+/le/로, /du/는 /de+/le/로와 같이 분리되어야 한다.

이와 같은 어절 형태 변형의 이점은, 외형적으로 한 어절을 이루면서 두 가지 통사적 기능을 갖는 어절들을 각각 한가지의 통사적 기능을 갖는 단어로 분리하는 데에 있으며, 이렇게 분리해 줌으로써 기존의 문법범주만으로는 처리가 곤란한 듯이 보이던 문자열(이

절)을 문법범주로 범주화시킬 수 있는 문자열(단어)로 분리시켜준다. 이와 같은 처리 방법은 언어학적으로도 그 처리 근거의 타당성이 이미 입증되었으며 이질적인 기능을 갖는 형태소들이 결합된 형태들을 위한 새로운 범주를 만들지 않아도 되므로 언어 처리과정의 효율성을 증대시키는 결과를 가져온다

전처리과정의 일환인 어절의 형태분리단계는 문자열의 아말감(amalgamation)에 의해 생성된 형태들을 결합 이전의 형태소들로 분리하는 것을 그 일차적인 목적으로 한다. 그러나 형태분리단계에서 아말감된 형태를 단순히 원래의 형태열(regular form)로 바꾸어만 준다면 분리된 결과는 분리 이전의 단계보다도 입력문의 증의성을 배가시키는 바람직하지 못한 결과를 초래하게 된다. 예를 들자면, 분리 단계에서 아말감 형태 /au/를 단순히 /à/와 /le/로 나누어만 놓는다면, context free 상 구문분석단계에서, /le/가 정관사 혹은 약세형 대명사가 될 수 있음으로 인해 /le/의 증의성 제거를 위한 추가작업을 해야만 한다. 그러나, 분리의 단계에서 /au/에서 분리된 /le/에 "정관사"라는 정보를 제공한다면, 그와 같은 추가 작업은 피할 수가 있을 것이다[10] 그러므로 형태 분리 작업은 가능한, 다음의 예에서와 같이, 분리된 결과 형태의 문법범주에 대한 정보제공과 병행되어야 한다

au	→	à le, D ⁺
au	→	à, P, LON le, D, DEF, MAS, SIN

위에서 제시한 4가지 불어 형태소 결합 현상 중 아말감 형태 결합에 관한 언어학적 연구는 Jacqueline PINCHON의 "Morphosyntaxe du français : Étude de cas"에서 찾아 볼 수 있으며, LACL, CERIL, APAST, ENSSAT 등의 형태소 분석기 그리고 최근 RALI(Le Laboratoire de Recherche Appliqué en Linguistique Informatique)에서 제공하는 웹사이트 상의 불어 형태소 분석기를 살펴보면 단순 형태 분리 방법을 사용하고 있을 뿐, 절의 경계에서 나타나는 관계대명사나 종속 접속사 등을 그대로 태그하여 사용하고 있다[11]

3. 어절 형태 분리 기법

축약, 결합, 연결, 아말감 된 형태들은 형태소 분석의 단계에 들어가기 전 단계인 문의 전처리 과정에서 형태분리 및 도치가 필요로써 형태소분석의 효율성을 증가시킬 수 있다.

3.1 Elision Form의 분리

단순 축약형태는 축약되는 두 형태소가 모음탈락 현상을 거친 후 apostrophe로 연결되어 한 형태로 묶인 것을 말한다. 불어에서는 '모음충돌 회피현상'이 두드러져, 첫째 단어가 모음으로 끝나고 다음 단어가 모음으로 시작될 경우 축약현상이 나타나게 된다. 첫째 단어의 마지막 모음이 탈락하고 그 자리에 apostrophe가 삽입되며 두 단어의 축약이 이루어진다. 탈락되는 모음 중 대부분은 /e/이지만, 때로 /i/가 탈락되는 경우도 있다. 그러나 /i/가 탈락될 수 있는 경우는 매우 제한되어 있어서 /s'il/ 혹은 /s'ils/의 경우에만 가능하다.

/on se aime à la mesure que on se connaît mieux/
 ↓ 축약현상
 /on s'aime à la mesure qu'on se connaît mieux/
 [서로 잘 알개 될 수록 사랑하게 된다]

단순 축약형태는 불어 문서에서 매우 많이 나타나는 현상으로, 약 60만 어절을 갖고있는 SYDO²⁰ corpus에 나타나는 빈도수의 조사 결과 약 6%의 어절이 축약형태 어절에 해당하였다.

축약형태는 일반적으로 apostrophe자리에 생략된 모음을 복원해주고, apostrophe가 있던 위치를 기점으로 두 단어로 절단함으로써 그 복원이 가능하다. 생략되었던 모음은 일반적으로 /s'il/ 과 /s'ils/ 두 축약형태를 제외하고는 모두 /e/ 로 복원된다. 그러나 /aujourd'hui [오늘]/나 /grand'mère [할머니]/ 에서와 같이 apostrophe가 형태소의 일부를 구성하는 경우도 있다. 이와 같은 예는 그 수가 매우 적어서 특수 단어 목록을 만들어 처리하는 것이 타당하다.

3.2 Hyphenated Form의 분리

대부분의 경우 hyphen으로 결합된 어절은 의미상 구성 요소들이 분리될 수 없는 관계 즉 복합어로서의 기능을 갖는다. 그러나 때로는 hyphen이 단순히 단어 사이의 경계선을 나타내는 경우도 있다. 후자의 경우에 해당되는 결합형태는 다음 세가지 경우에 나타난다.

- /ce chien-ci [이 개]/, /celui-ci [이 사람]/, /cette voiture-là [저 자동차]/, /jusque-là [지금까지]/, /par-là [저기애]/ 등에서와 같이 /ce/를 수반한 명사구나 전치사구가 particle /ci [여기]/ 혹은 /là [저기]/를 만날 경우, /ci/ 와 /là/는 hyphen으로 결합된다.
- /aime-t-il. [그는.. 좋아하는가]/, /est-elle... [그녀는..

인가]/, /partez-vous. [당신은...떠나시오 ; 당신은...떠나는가]/ 등에서와 같이 대명사와 동사가 도치될 때 대명사와 동사 사이에 hyphen이 삽입된다. 그런데 동사가 모음으로 끝나고 대명사가 모음으로 시작할 경우에는 발음 순화를 위해 추가로 음운 /t/가 삽입되어야 한다.

- /moi-même [나 자신]/, /lui-même [그 자신]/, /nous-même [우리 자신]/ 등에서와 같이 인칭대명사에 /même [자신]/ 형태가 붙은 경우도 있다 이 경우 /même/는 그 문법 기능으로서 형용사의 기능을 갖는 것으로 처리된다. 그러므로 차후 형태소 분석에서 갖게 될 중의성을 줄이기 위하여, 형태 분리 단계에서 미리 | même. F, ADJ, GRN, SIN | 란 문법정보를 제공해야 한다

3.3 Orthographic Amalgamation Form(concatenation form)의 분리

/auquel/, /pourquoi/에서와 같이 결합형태가 여러 단순형태소의 단순한 concatenation에 의한 경우를 orthographic amalgamation form라 한다. 이 경우 문의 전처리작업은 다음의 예에서와 같이 문자열을 단순히 끊어 주는 것에 불과하다.

| auquel | → | au | quel |
 | pourquoi | → | pour | quoi |,

3.3.1 복합관계사의 형태분리

복합관계사는 관사(혹은 축약관사) 부분과 /que/의 변형형태로 된 부분 등으로 분리될 수 있다.

lequel	→	le, D, DEF, MAS, SIN	quel
laquelle	→	la, D, DEF, FEM, SIN	quelle
lesquels	→	les, D, DEF, MAS, PLU	quels
lesquelles	→	les, D, DEF, FEM, PLU	quelles

auquel	→	au	quel
auxquels	→	au	quels
auxquelles	→	au	quelles
duquel	→	du	quel
desquels	→	des	quels
desquelles	→	des	quelles

3.3.2 지시사의 형태 분리

/celui [그 사람 {남성 단수}], /celle [그 사람(여성 단수)], /ceux [그 사람 {남성 복수}], /celles [그 사람 (여성 복수)], /ceci [이것], /cela [저것], /ça [이

기에/ 등과 같은 지시대명사들은 언뜻 보기에 어절의 표면형태를 분리해야할 필요성을 나타내지 않는다. 그러나 이러한 대명사의 형태분리의 타당성은 문자열 /celui-ci/와 명사구 /ce chien-ci/를 비교해 봄으로써 혹은 문자열 /celui qui... [..한 그사람]/와 명사구 /le chien qui... [..한 그 개]/를 비교해 봄으로써 나타나는 isomorphie에서 찾을 수 있다. 따라서, 지시대명사들은 /celui/, /ceux/, /celles/은 정관사에 해당되는 부분 (/ce/, /cette/, /ces/)과 /lui [그]/, /elle [그녀]/, /eux [그들]/, /elles [그녀들]/와 같은 3인칭 대명사 부분으로 분리될 수 있다

celui →	ce, D, DEF, MAS, SIN lui, F, NOM-PRO, MAS, SIN, PE3
celle →	cette, D, DEF, FEM, SIN elle, F, NOM-PRO, FEM, SIN, PE3
ceux →	ces, D, DEF, MAS, PLU eux, F, NOM-PRO, MAS, PLU, PE3
celles →	ces, D, DEF, FEM, PLU elles, F, NOM-PRO, FEM, PLU, PE3

일반적으로 대명사는 정관사와 함께 사용될 수 없다. 따라서 위에서 제시된 바와 같은 지시사의 형태분리는 일종의 예외적 규칙에 해당한다고 볼 수도 있다. 이를 보완하기 위해서는 형태분리에서 얻어진 /lui/, /elle/, /eux/, /elles/에는 NOM-COM(보통명사)의 문법 속성을 제공함으로써 진짜 대명사 /lui/, /elle/, /eux/, /elles/과 차별화 해야한다.

celui →	ce, D, DEF, MAS, SIN lui, F, NOM-COM, MAS, SIN, PE3
celle →	cette, D, DEF, FEM, SIN elle, F, NOM-COM, FEM, SIN, PE3
ceux →	ces, D, DEF, MAS, PLU eux, F, NOM-COM, MAS, PLU, PE3
celles →	ces, D, DEF, FEM, PLU elles, F, NOM-COM, FEM, PLU, PE3

3.4 Morphological Amalgamation Form의 분리

축약관사와 같은 morphological amalgamation form은 형태분리에 의해 나온 구성 요소들의 형태가 분리전의 문자열 형태의 일부가 아니라 전혀 다른 형태를 취한다는 점에서 orthographic amalgamation form과 구분된다. 따라서 morphological amalgamation form은 orthographic amalgamation form에서와 같은 단순한 형태절단에 의한 형태분리를 할 수가 없다. /auquel/,

/duquel/과 같은 복합관계사에서 볼 수 있는 것처럼, orthographic amalgamation form는 경우에 따라 morphological amalgamation form를 구성 요소로 가질 수 있음을 주의해야 한다. 따라서 형태분리 규칙의 수를 줄이기 위하여 그리고 각각의 규칙의 재사용을 가능하게 하기 위하여 형태분리는 다음과 같이 2차에 걸쳐서 처리되어야 한다.

- ① orthographic decomposition
- ② morphological decomposition

위와 같은 절차에 의한다면, 복합관계사 /auquel/의 형태는 | auquel | → | au | quel | → | à, P | le, D | quel | 순으로 규칙이 연속적으로 적용되어 분리된다.

종속절을 이끄는 관계대명사와 종속접속사에 관계된 형태들이 /qu-/형태에 속한다. 그와 같은 형태들은 /que/와 대명사 혹은 대용지 부사(anaphoric adverb)의 아말감 된 형태로 분석된다 이러한 분석의 이점은 (Q' /que/, /de/, /si/)와 같은 절(proposition)의 경계를 나타내는 표지들의 목록을 줄이는데 있다.

예) | qui | → | que, Q | nil, F, NOM-PRO |

때로는 /qu-/ 형태의 처리는 형태분리뿐만 아니라 분리된 요소들의 출현 순서의 조정도 요구한다. 특히 전치사가 선행된 관계사의 경우가 그에 해당된다. 예를 들자면, /à qui/, /pourquoi/ 등과 같은 경우 전치사는 /que/에 지배되는 것이 아니라 잠정적인 대명사에 지배됨을 고려하여야 한다 따라서 전치사와 /que/의 위치는 아래와 같이 도치된다.

à qui →	que, Q à, P x, F, NOM-PRO
auquel →	à, P le, D quel → que, Q à, P le, D x, F, NOM-COM

이러한 처리는 전치사에 대한 인식이 이미 이루어졌음을 전제로 한다. 그리고 때에 따라서는 아래의 같이 orthographic decomposition와 morphological decomposition 처리가 연속되어야 함도 의미한다.

4. 복문의 단문화

4.1 어절 분리처리의 목적

/qu-/ 형태로의 형태분리는 문장의 표층 형태의 수를 줄이는데 그 목적을 두는 것은 물론, 명사구 인식

에도 큰 의미를 두고 있다. 잘 알려진 바와 같이 /que/ 형태로의 분리는 관계대명사를 종속접속사 /que/와 대명사(혹은 대명사구)로 분리 구분하는 것이므로 대명사가 대신하는 종속절의 명사구 인식을 가능하게 한다. 또 다른 형태 분리 목적으로는 절(propositions)의 인식을 들 수 있다. 이는 /que/가 문장 표층에서는 직접 나타나 있지 않았다 하더라도 절의 경계를 의미하는 /que/ 형태를 분리 유추해냄으로써 절의 경계를 확정짓게 된다.

/que/ 형태분리 처리의 근본적 목적은, 관계사의 형태를 절의 경계를 나타내는 /que/와 전치사 그리고 조응성을 갖는 대명사 혹은 부사의 구조로 변형시키는데 있다. 이때 나타나는 전치사는 때로 관계사에 선행되어 나타날 수 있는 선행 전치사를 의미하는 것이다. 다음의 예가 이 경우들을 잘 보여주고 있다.

/le docteur qui court/	-> /le docteur que lu court/
/le docteur à qui je pensais/	-> /le docteur que à lui je pensais/
/le docteur dont je parle/	-> /le docteur que de lu je parle/
/le pays où je vais/	-> /le pays que là je vais/
/ce pour quoi je suis venu/	-> /ce que pour ce je suis venu/

/de/ 혹은 /à/로 시작되는 원형 종속절 또는 /si/로 시작되는 의문형 종속절 이외에 모든 종속절이 /que/의 형태로 분리 분석되게 하고자 모든 종속접속사 형태 즉 /quand/, /comment/, /comme/, /combien/ 까지 /que/ 형태분리 분석범위를 넓혔다.

이러한 '복문의 단문 전처리' 작업은 3가지의 이득을 준다. 우선 종속접속사의 수를 현저히 줄여준다. 분리 분석된 모든 종속절은 /que/로 시작되므로 절의 경계가 확실해 지고 복문의 단문화 작업이 순조로워 진다. 그리고 종속절을 하나의 독립된 절(독립 단문)처럼 분석되게 하는 문장 구성요소를 삽입 완성시키게 된다.

(le docteur) que [lui]SN [court]SV (뛰어가는 의사)
[le docteur]SN [court]SV (의사가 뛰어간다)

절로의 분석으로 유도하는 것을 목적으로 삼고 있는 /que/ 형태분리 분석방법은 때로 그 목적을 제대로 달성하지 못하는 경우가 있다. /que/ 형태분리 분석에 의

해 독립된 절로 형태 변형이 이루어졌을 때, 종속절의 선두에 위치하는 명사구, 전치사구, 부사구는 독립된 문장구성으로서의 제 위치를 완벽하게 찾아내지 못하고 있다.

/le docteur à qui je pensais/ (내가 생각해보던 의사)
-> /le docteur que à lui je pensais/

이와 같이 변형된 종속절에서 완전히 독립된 절의 문장구조를 알아내는 것은 아직 힘들다.

이러한 형태 변형은 많은 조응소(anaphoriques)형태를 줄여줌은 물론 "절 사이의 조응관계(anaphore intraphrastique)"를 "문장 사이의 조응관계(anaphore interphrastique)"로 바꾸어줌으로써 문장 자동 처리를 가능하게 해 준다.

/le chien qui est noir aboie/ (검은 개가 짖는다)
-> /le chien [(que) lu est noir] aboie/ (개 [김은] 짖는다)
-> /∅ [(que) le chien est noir] aboie/
/le chien est noir il aboie/
-> /le chien est noir lui, il aboie/ (개가 짖다, 그가 짖는다)
-> /[(le chien est noir) aboie]/*

첫 예문의 /le chien/과 두 번째 예문의 /le chien est noir/가 각각 조응소 /lui/와 연관되는데, 이는 각각 /est noir/와 /aboie/ 술어와 결합되는 문장 구조를 이루게 된다. 이러한 분석은 물론 두 발화체가 완전히 동일한 것이리는 전제하여 가능하다.

관계접속사 형태분리 분석에 있어서 주어 대명사 /il/을 분리 형태로 사용하지 않고 강세형 대명사 /lui/를 사용하는 이유는, 다음 예문에서 보는 바와 같이 종속절의 조응소가 주어가 아닌 보어일 가능성이 있기 때문이다.

/le chien avec qui Paul se promène aboie/ (풀이 데리고 산보하던 개가 짖는다)
/Paul se promène avec le chien. Celui-ci aboie/ (풀이 개와 산보하다/그 개가 짖는다)

따라서 주어로써 /qui/의 형태분리 분석과 보어로써 /qui/의 형태분리 분석에 있어서 일관성을 유지하기 위해 /qui/의 심층구조 표현에 있어서 두 경우 모두 강세형 대명사를 사용하는 것이 타당하다.

4.2 단순 관계대명사

우선 단순 관계대명사로 연결된 복문의 분리를 살펴 보면, 단순관계대명사는 /qui/, /que/, /quoi/, /dont/, /où/ 등의 형태를 갖는 관계대명사들의 형태 분리에 관계된 것이다.

단순 관계대명사는 두 가지 사용법으로 나눌 수 있다. 즉 독립적인 사용과 전치사절로의 사용이다. 이 중 두 관계대명사(/que/와 /dont/)는 전치사가 선행되어 사용될 수 없다. 반면 나머지 세 개는 다음과 같이 전치사가 선행되어 사용되거나 혹은 독자적으로 사용될 수도 있다

```
/le docteur qui court/, /le docteur à qui je pensais/
/le pays où je vais/, /le pays par où je passe/
/cc pour quoi je suis venu/, /quoi que tu fasses/
```

4.2.1 전치사 없이 사용

단순 관계대명사가 전치사 없이 단독으로 사용되는 경우는 다음과 같이 분석될 수 있다.

```
| qui | → | que, Q | Ø, F, NOM-PRO, GRN, NBN |
| dont | → | que, Q | de, P | Ø, F, NOM-PRO, GRN,
           NBN |
| où | → | que, Q | là, W, PRO |
| que | → | que, Q |
| quoi | → | que, Q | Ø, F, NOM-PRO, GRN, NBN |
예) /le chien dont je parle/
    → | le | chien | que, Q | de, P | Ø, F,
       NOM-PRO, GRN, NBN | je | parle |
```

전치사 없이 사용되는 단순 관계대명사의 형태분리에 있어서 몇 가지 고려해 봐야 할 사항을 살펴보면 다음과 같다.

- 단순 관계대명사의 형태분리 분석결과 중, /dont/의 분리 형태에 있어서 전치사 /de/의 등장과 /où/의 분리 형태에 있어서 조응소 /là/의 등장은 /dont/과 /où/가 각각 /de/ 전치사구 문장구성격 장소보어(complément locatif)의 자리를 차지함을 보이고 있다
- 장소보어가 여러 가지 전치사에 의해 선행되거나 특히 /où/ 그 스스로가 전치사에 의해 선행될 수 있으므로, /où/를 | que, Q | Ø, P, LOC | Ø, F, NOM-PRO, GRN, NBN | 분리하는 것 보다 /là/로 처리하는 것이 더 합당하다고 본다.

- /que/ 형태도 중의성 문제를 안고 있다. 즉 /que/의 사용에 있어서 부지의 형태로서의 사용을 제외하고는 관계대명사로 쓰이거나 절의 경계를 나타내는 단순한 접속사의 역할을 가질 수 있다. 그리고 이 두 가지의 사용을 문맥상에서 분석해 내는 것이 쉽지 않을 것 같다

```
/le sentiment que tu ressens./
/le sentiment que tu ressens de la fierté/
```

- /tu fait quoi?/와 같이 불어에 아직 완전히 젖어 들지 않은 구어체, 논문이나 기사 등 정보처리의 대상이 되는 문헌들에서 매우 드물게 나타나는 구어체 형태의 문장을 제외한다면, 전치사가 없는 /quoi/는 양보절의 앞머리에 나타나는 부정대명사로서만 나타나게 된다.

```
/quoi que tu fasses, .../
```

4.2.2 전치사절로 사용

관계대명사가 전치사에 의해 선행되었을 때, /que/와 강세형 대명사가 융합된 것이라 간주된다 전치사구의 전형적인 형태에 상응하기 위하여 전치사는 /que/와 대명사 사이에 삽입되어야 한다 따라서 분리된 표변 형태의 도치가 필요하게 된다.

```
| xxx, P | qui | → | que, Q | xxx P | Ø, F, NOM-
                  PRO, GRN, NBN |
| xxx, P | où | → | que, Q | xxx P | là, W, PRO |
| xxx, P | quoi | → | que, Q | xxx P | Ø, F,
                   NOM-PRO, GRN, NBN |
예) /le chien à qui je parle/
    → | le | chien | que, Q | à P | Ø, F,
       NOM-PRO | je | parle |
```

4.2.3 ce+관계대명사

관계대명사 앞에 /ce/형태가 붙은 형태의 분리를 가능하게 하는 규칙에 관한 것이다. 이러한 형태로는 /celui/, /celle/ 혹은 /ci/나 /là/로 시작되는 /ce/형태들을 말한다. 이 경우 /ce/형태가 갖는 역할은 대명사의 역할일 뿐 한정사의 역할은 아니다. 이 대명사를 다시 조응적 한정사 /ce/와 명사 Ø로 나눌 수 있다.

ce que →	ce, D, DEF, GRN, SIN Ø, F, NOM-PRO, GRN, SIN que, Q
ce qui →	ce, D, DEF, GRN, SIN Ø, F, NOM-PRO, GRN, SIN que, Q Ø, F, NOM-PRO, GRN, SIN
ce dont →	ce, D, DEF, GRN, SIN Ø, F, NOM-PRO, GRN, SIN que, Q de, P Ø, F, NOM-PRO, GRN, SIN
ce xxx, P quoi →	ce, D, DEF, GRN, SIN Ø, F, NOM-PRO, GRN, SIN que, Q xxx, P Ø, F, NOM-PRO, GRN, SIN
예) /ce pour quoi je suis venu/	→ ce, D, DEF Ø, F, NOM-PRO que, Q pour, P Ø, F, NOM-PRO je suis venu

4.3 관계대명사가 아닌 종속접속사

분해될 수 없는 /que/를 제외하면, 관계대명사가 아닌 종속접속사는 다음과 같다.

/quand/, /comment/, /comme/, /combien/
--

/où/에서 같이 관계대명사가 아닌 종속접속사들을 /que/와 조응 부사의 융합 형태로 이해할 수 있다.

/quand/	= /que alors/
/comment/	= /comme/ = /que ainsi/
/combien/	= /que tant/

/où/와 마찬가지로, 전치시에 의해 선행되는 것도 있다.

quand →	que, Q alors, W, PRO
comment →	que, Q ainsi, W, PRO
comme →	que, Q ainsi, W, PRO
combien →	que, Q tant, W, PRO
xxx, P quand →	que, Q xxx, P alors, W, PRO
xxx, P combien →	que, Q xxx, P tant, W, PRO
예) /je sais comment tu écris/	→ je sais que, Q ainsi, W, PRO tu écris

4.4 /quel/ 형태

절차 융합 형태 처리에서, 복합 관계사의 형태 분리는 한정사가 선행되는 /quel/ 형태로 분리되었으며, 때로는 전치사가 선행되기도 하였다. 복합 관계사의 완전한 분석을 위해 /quel/, /quelle/, /quels/ 그리고 /quelles/ 들을 /que/와 대명사(혹은 대명사구)가 융합

된 형태로 간주하여야 한다. 이는 단순 관계대명사의 구성과 일관성을 유지하기 위한 구조를 갖게 함이다.

/quel/ 형태는 /que/와 /tel/이 융합된 형태로 볼 수 있다. 따라서 /quel/을 다음과 같이 그 분리 분석 규칙을 정하였다.

quel →	que, Q tel, F, NOM-PRO, MAS, SIN
quelle →	que, Q telle, F, NOM-PRO, FEM, SIN
quels →	que, Q tels, F, NOM-PRO, MAS, PLU
quelles →	que, Q telles, F, NOM-PRO, FEM, PLU
예) /je sais quel est son nom/	→ je sais que, Q tel, F, NOM-PRO est son nom

/quel/ 형태가 한정 지시사에 의해 선행되는 경우, 관계절의 선두에 /que/를 두고 다음 규칙을 적용하여 형태분리를 하게 된다

le, D, DEF, MAS, SIN quel	→ que, Q le, D, DEF, MAS, SIN tel, F, NOM-PRO, MAS, SIN
la, D, DEF, FEM, SIN quelle	→ que, Q la, D, DEF, FEM, SIN telle, F, NOM-PRO, FEM, SIN
les, D, DEF, MAS, PLU quels	→ que, Q les, D, DEF, MAS, PLU tels, F, NOM-PRO, MAS, PLU
les, D, DEF, MAS, PLU quelles	→ que, Q les, D, DEF, MAS, PLU telles, F, NOM-PRO, FEM, PLU

마찬가지로 /quel/ 형태가 전치사와 한정 지시사 한 쌍에 의해 선행되는 경우, 관계절의 선두에 /que/를 두고 다음 규칙을 적용하여 형태분리를 하게 된다.

xxx, P le, D, DEF, MAS, SIN quel	→ que, Q xxx, P le, D, DEF, MAS, SIN tel, F, NOM-PRO, MAS, SIN
xxx, P la, D, DEF, FEM, SIN quelle	→ que, Q xxx, P la, D, DEF, FEM, SIN telle, F, NOM-PRO, FEM, SIN
xxx, P les, D, DEF, MAS, PLU quels	→ que, Q les, D, DEF, MAS, PLU tels, F, NOM-PRO, MAS, PLU
xxx, P les, D, DEF, MAS, PLU quelles	→ que, Q les, D, DEF, MAS, PLU telles, F, NOM-PRO, FEM, PLU
예) /le chien auquel je parle/	→ le chien que, Q à, P le, D, DEF tel, F, NOM-COM je parle

4. 맺음말

본 논문에서는, 불어 문서에서 나타나는 어절 형태의 특성 중 형태소 결합 현상에 관하여 살펴보았다. 형태소 분석기의 기본 입력 단위는 분장이다. 입력 문장이 얼마나 전처리 되었는가는 형태소 분석기에 걸리는 부하와 밀접한 관계를 나타낸다.

본문에서 제시한 축약형태, 결합형태, 연결형태, 아말감형태 등 네 가지 형태소 결합 현상에 대한 분류는 불어에서 나타날 수 있는 모든 결합 현상을 분류한 것이다. 이러한 어절 형태 결합 현상으로 이루어진 어절의 출현 빈도는 불어 문서에서 평균 10% 이상에 이른다. 본 논문의 기본 데이터로 사용된 60만 어절의 SYDO corpus에서 나타난 빈도를 살펴보면 축약형태가 가장 많은 빈도를 보이며 6%였으며, 네 가지 모든 형태의 출현 빈도는 11%에 이르렀다. 형태소 분석기에 있어서 분석 대상의 11%의 어절을 전처리 시킬 수 있다는 것은 그만큼의 부하가 줄어드는 것이므로 형태소 분석기의 효율성 증진을 의미한다.

따라서 본 논문에서는, 형태소 결합 현상으로 생성된 어절을 분리해 줌으로써 불필요한 단어 유형의 추가를 막아주고, 형태소 분석 알고리즘의 부하를 줄여주고자 하는 목적으로 문의 전처리의 필요성을 강조하였으며, 문의 전처리 중 어절의 형태 분리에 관한 총 96개의 형태 분리 규칙을 제시하였으며, 그중 문의 전처리 중 복문을 단문으로 분리 시켜주는 단문화 분리 규칙 43개를 포함하고 있다.

참 고 문 헌

[1] Allen, J. Natural Language Understanding, Second Ed. The Benjamin/Cummings Publishing Company, INC, 1994

[2] Berrendonner, A. *Cours critique de grammaire générative*. Lyon : PUL, 1983.

[3] Berrendonner, A. "Grammaire pour un analyseur aspects morphologiques." Lyon : document de travail du groupe SYDO, 1984.

[4] Bouché, R. "Le syntagme nominal, une nouvelle approche des base de données textuelles," in *Actes*

du colloque terminologie et industries de la langue, META Journal des traducteurs Vol.34 / No.3, Montréal, 1989.

[5] Chaumier, J. *Le traitement linguistique de l'information documentaire - L'analyse documentaire* Paris: Entreprise Moderne d'Édition, 1977.

[6] Harris, Z-S "La structure distributionnelle," in *Langages*, No.20, Paris : Larousse, 1970.

[7] Jones, K. S. "How much has information technology contributed to linguistics?," In J. T Coppock (ed), *Information technology and scholarly disciplines*, 1997.

[8] LeGuern, M. "Un analyseur morpho-syntaxique pour l'indexation automatique," *Le français moderne - Tome LIX-No.1*, Paris : CILF (Conseil International de la Langue Française), 1991.

[9] Pinchon, J. *Morphosyntaxe du français : Étude de cas*. Paris : Hachette, 1986.

[10] Seo, L. Contribution au traitement automatique du verbe français lexique, morphologie et syntaxe. thèse en doctorat. Lyon université-lumière, 1995.

[11] <http://www-rah.iro.umontreal.ca>



서 래 원

e-mail : seo@mail.pachan.ac.kr

1985년 Université Lyon 진산언어학과 졸업(학사)

1987년 Université Lyon 진산언어학과(전산언어학 석사)

1994년 Université Lyon 정보통신학과(정보통신학 박사)

1988년~1994년 프랑스 S.I.R., 선임연구원

1992년~1994년 프랑스 LyonIII 대학교 정보통신과, 겸임교수

1995년~1997년 한국전자통신연구원 자연언어처리연구실, 선임연구원

1998년~현재 배재대학교 정보통신학부 조교수

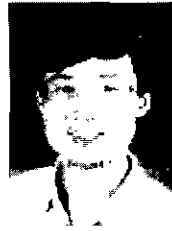
관심분야 자연언어처리, HCI, 멀티미디어 정보 검색, 전문가 시스템



박 세 영

e-mail sypark@computer.etrn.re.kr
1980년 경북대학교 전자공학과
졸업(학사)
1982년 KAIST 전산학과 졸업
(석사)
1989년 프랑스 파리 7 대학(박사)
1982년 한국전자통신연구원 입소

1992년 지식정보연구실장
1995년 미국 미주리 주립대학 초빙연구원
1997년 자연어처리연구실장
1998년 자연어처리연구부장
1999년 지식정보연구부장
관심분야 : 자연어처리, 한국어 정보처리, 정보검색, 자동번역



유 성 준

e-mail : sdotyoo@lycos.co.kr
1982년 고려대학교 전자공학과
졸업(학사)
1990년 고려대학교 대학원 전자공
학과(정보처리전공 석사)
1996년 시라큐스대학교 전기 및 전
산학과 졸업(전산학 박사)

1982년~현재 한국전자통신연구원 지식정보검색연구팀장,
책임연구원
관심분야 : 멀티미디어 정보 검색, 구조화 멀티미디어
문서 정보 검색, 멀티미디어 데이터베이스