

일본어 텍스트 가독성 분석*

—구JLPT 기출문제 기반 일본어 텍스트의 한자 및 한자 어휘 가독성 판단 프로그램의 개발—

김 유 영**

<目 次>

1. 들어가며	3.3 'AJ-JpnRa Tool'의 실행화면 및 사용법
2. 일본어 텍스트의 가독성 분석	
3. 일본어 텍스트 가독성 분석 프로그램, 'AJ-JpnRa Tool v.1.2.0.2'의 개발	4. 'AJ-JpnRa Tool'을 활용한 연구 예
3.1 AJ-JpnRa Tool의 가독성 판단 설계방침	4.1. 일본어 교육과 학습
3.2 'AJ-JpnRa Tool'의 사양 및 설치	4.2. 알기 쉬운 일본어(やさしい日本語)
	5. 나가며

Key Word : 可読性, 漢字教育, 表記, 形態素分析, 日本語能力試験(JLPT)

1. 들어가며

‘가독성(可読性・readability)’의 사전적 의미는 텍스트를 ‘읽고 이해할 수 있는 정도’ 혹은 텍스트를 ‘읽기 쉬운 정도’라고 간단히 정의할 수 있다¹⁾. 읽고 이해하기 쉬운 텍스트 즉, ‘가독성’이 높은 텍스트는 독자들로 하여금 그 내용을

* 이 논문은 2013년도 동덕여자대학교 연구비 지원에 의하여 수행된 것임(연구번호: 201302530).

This study was supported by the Dongduk Women's University grant(No. 201302530).

** 金晴泳(동덕여자대학교 일본어과 교수), 일본어 어휘론.

1) 『日本国語大辞典』第2版(2010), *Oxford Advanced Learner's Dictionary*. 8th Edition (2010).

쉽게 기억할 수 있게 해 주며, 또한 많은 양의 텍스트를 보다 더 빠르고 오래 동안 읽을 수 있게 해 준다. 그런데 이와 같은 텍스트의 가독성에는 다양한 요소가 영향을 미치는데, 이는 크게 다음의 (1)과 같이 ‘리더(Reader·読者, 이하 리더) 영역’과 ‘텍스트(Text·テキスト, 이하 텍스트) 영역’이라는 두 차원으로 구분하여 생각해 볼 수 있다.

(1) [텍스트 가독성에 영향을 미치는 요인]

- a. 리더Reader 영역 : 사전 지식, 독해 능력, 흥미 및 동기
- b. 텍스트Text 영역 : 내용, 언어형식(길이·표기·어휘·문법·문체 등), 구성형식(디자인·구성·폰트 등)

이와 같이 다양한 요인에 의해 결정되는 텍스트 가독성은 응용하기에 따라 여러 분야에 중요한 의미를 갖는다. 예를 들어 (1)a와 같은 ‘리더 영역’의 요인이 고정적이라고 할 때, 정부와 같은 공적 기관의 공지 혹은 안내 등의 경우, 효과적인 정보전달을 위해 (1)b와 같은 ‘텍스트 영역’의 요인을 조절하여 가독성 높은 텍스트를 수요자에게 제공할 필요가 있다(e.g. 외국인 주민을 대상으로 한 고용보험 안내). 또한, 상업 출판에 있어서도, ‘텍스트 영역’의 요인을 조절하여 대상 독자에게 적합한 가독성 높은 텍스트를 제공하는 것을 통해 매출을 높이고 이익을 극대화 할 수 있다(e.g. 대학 신입생을 위한 경제 입문서).

한편, (1)b의 ‘텍스트 영역’의 요인이 고정적이라고 가정한다면, 해당 텍스트의 가독성 즉, 난이도에 적합한 수준의 학습자의 학습에 그 해당 텍스트를 사용하는 것을 통해 학습의 효과를 극대화 시키는 등, 교육 및 학습 분야에도 응용할 수 있다(e.g. JLPT N2 학습자를 위한 일본 동화 선집).

이에 본고에서는 일본어 텍스트 가독성 분석 프로그램(AJ-JpnRa Tool)을 개발하는 것을 통해, 특정 일본어 텍스트를 얼마나 읽고 이해하기 쉬운가를 판별하기 위한 객관적인 기준을 제시하고, 동시에 이를 기반으로 일본어 텍스트의 레벨을 판단할 수 있는 수단을 제공하고자 한다. 이와 같은 연구는 앞으로, 일본어 교육 및 학습 그리고 알기 쉬운 일본어(やさしい日本語)²⁾ 등의 일본어

2) 알기 쉬운 일본어(やさしい日本語): 외국인 주민과의 커뮤니케이션의 어려움을 극복하기 위해, 기존 ‘배리어 프리(Barrier Free)’ 혹은 유니버설 디자인(Universal Design)이라

연구에 활용할 수 있을 것이라고 생각한다.

2. 일본어 텍스트의 가독성 분석

본 연구에서 다루고자 하는 가독성은 앞서 언급한 요인 중에서 (1)b의 ‘텍스트 영역’, 그 중에서도 ‘언어형식’에 관한 내용이라고 할 수 있는데, 이를 구체적으로 일본어 텍스트 가독성에 대입시켜 본다면 ‘일본어 텍스트의 길이(단락, 문장, 형태소 개수)’, ‘일본어 텍스트의 표기(한자, 히라가나, 가타카나 등)’, ‘일본어 텍스트에 사용된 한자’, ‘일본어 텍스트에 사용된 어휘’, ‘일본어 텍스트에 사용된 문법’, ‘일본어 텍스트의 문체’ 등이 일본어 텍스트의 가독성에 영향을 미치는 ‘언어형식’ 요인이 된다. 이 중에서 이번에 개발된 ‘AJ-JpnRa Tool’ 프로그램에서 채택한 가독성 판별요소는 다음의 (2)와 같다. 그러나 본 연구에서는 지금까지 개발된 (2)a·b·c와 같은 요소를 중심으로 일본어 텍스트 가독성을 판단하고자 하며, (2)d·e 요소에 관해서는 금후의 과제로 삼고자 한다.

(2) [일본어 텍스트 가독성 분석 프로그램 ‘AJ-JpnRa Tool’의 가독성 판단 요소]

- a. 일본어 텍스트의 길이: 단락·문장·형태소 개수 (개발완료)
- b. 일본어 텍스트의 표기: 한자·가나 비율 (개발완료)
- c. 일본어 텍스트에 사용된 한자 난이도 (개발완료)
- d. 일본어 텍스트에 사용된 어휘 난이도 (개발 중)
- e. 일본어 텍스트에 사용된 문법 난이도 (개발 중)

그런데, 위와 같은 가독성 판별 요소를 기반으로 텍스트를 분석하는 데에는 그 무엇보다 중요한 점이 있는데, 그것은 바로 ‘기준’이다. 예를 들어, 과연 어느 정도의 (2)a‘일본어 텍스트의 길이’가 초급 일본어 학습자에게 적합한가? 그리

는 개념을 언어에 도입시켜, 언어의 장벽을 없애거나 완화하는 ‘언어 배리어 프리 Language barrier free’ 라는 개념(김유영, 2013b:19)을 기반으로 한 일본어.

참고) 「やさしい日本語」とは、普通の日本語よりも簡単で、外国人もわかりやすい日本語のこと - <http://human.cc.hirosaki-u.ac.jp/kokugo/EJ1a.htm>

고 과연 어느 정도의 (2)c‘한자 난이도’를 가진 텍스트가 외국인 주민에게 정보를 전달하는 데에 적합한가와 같은 문제를 해결하고자 한다면, 우선 이를 판단할 객관적이고 일정한 ‘기준’이 필요하게 된다.

이에 본 연구에서는 일본어 능력시험(이하, JLPT)의 출제기준을 주된 지표로 삼아 기준을 개발했는데, JLPT는 일본어를 교육하는 많은 교육기관에서 일본어의 구사 능력을 테스트하는 기준으로 사용되고 있으며 이는 한국뿐만 아니라 많은 다른 나라들도 그 사정이 크게 다르지 않기 때문이다. 우선, (2)a‘일본어 텍스트의 길이’와 (2)b‘한자·가나 비율’의 경우, 지난 20년간의 JLPT 기출문제의 독해·문법 영역의 독해 지문을 각 레벨별로 데이터베이스화 하여 산출한 통계를 기준으로 삼았다. 그리고 (2)c‘한자 난이도’의 경우, 지난 20년간의 JLPT의 한자·어휘 영역의 기출문제를 지표로 삼았다.

3. 일본어 텍스트 가독성 분석 프로그램, ‘AJ-JpnRa Tool v.1.2.0.2’의 개발

본고에서는 김유영(2013b)에서 사용된 베타버전의 일본어 텍스트 가독성 분석 프로그램인 ‘AJ-JpnRa Tool v.1.0.0.0’을 대폭 개선하여, 더욱 향상된 일본어 텍스트 가독성 분석 프로그램을 개발 및 공개하고자 하는데, 그 구체적인 프로그램에 관한 개요는 다음과 같다.

3.1 AJ-JpnRa Tool의 가독성 판단 설계방침

본 연구를 통해 개발된 ‘AJ-JpnRa Tool v.1.2.0.2’는 2절의 가독성 판단 요소 중, (2)a‘일본어 텍스트의 길이’와 (2)b‘일본어 텍스트의 표기’, 그리고 (2)c‘일본어 텍스트에 사용된 한자 난이도’를 기준으로 일본어 텍스트의 가독성을 종합적으로 판단하는데, 각각의 구체적인 기준에 대한 설정 방침은 다음과 같다.

3.1.1 ‘일본어 텍스트의 길이’와 ‘일본어 텍스트의 표기’

우선 (2)a‘일본어 텍스트의 길이’와 (2)b‘일본어 텍스트의 표기’의 경우, 본고에서는 최근 20년간³⁾의 구舊JLPT(1급~4급) 독해·문법 영역 중 독해 지문만을 선별하여 ‘JLPT 독해 지문 데이터베이스’를 구축하고, 이 데이터베이스를 다음의 (3)과 같은 항목을 기준으로 통계적으로 분석하여 AJ-JpnRa Tool의 텍스트 가독성 분석 기준을 수립했다.

(3) [구JLPT 독해 지문 데이터베이스 분석 항목]

a. ‘일본어 텍스트의 길이’ : (2)a

- 1) 전체 단락 수, 전체 문장 수, 전체 형태소 수
- 2) 텍스트 당 평균 단락 수, 텍스트 당 평균 문장 수, 텍스트 당 평균 형태소 수
- 3) 단락 당 평균 문장 수, 단락 당 평균 형태소 수
- 4) 문장 당 평균 형태소 수

b. ‘일본어 텍스트의 표기’ : (2)b

- 1) 전체 문자 대비 한자 비율
- 2) 가나 VS 한자의 사용 비율

참고로 구JLPT의 독해·문법 영역의 기출문제의 독해 지문은 다음의 (4)와 같은 문항을 선별했는데, 기출문제 중에는 (5)와 같이 그래프 및 그림을 보고 답하는 문항이 20년간 총 11문항이 포함되어 있으며 이와 같은 문항은 지문 선정에서 제외했다. 또한 기출 문제의 문항 중, ‘안내문’과 ‘문항’ 그리고 ‘보기’ 및 ‘인용정보’도 역시 그 대상에서 제외하고 순수한 독해 지문만을 채택했다 (e.g. II. 次の文章を読んで、後の問いに答えなさい。答えは、1・2・3・4から最も適当なものを一つ選びなさい。… 중략 … (佐江衆一『けんかの仕方教えます』による) … 중략 … 問1 ①「中学の三年間を通じて一人の友だちと深くつきあうことはなかった」とあるが、理由はなせか。… 중략 … 1. 野球がきらいだったから。4). 마지막으로 주석의 경우, 본문의 어휘를 주석의 내용으로 교체했다.

3) 최근 20년간(1990년~2009년), 일본어능력시험(JLPT) 기출문제.

4) 1991년 JLPT 2급, 독해·문법 기출문제 중, II-1번 문항.

(4) [구JLPT 독해·문법 영역 기출문제의 독해 지문 선정 문항]

a. 구JLPT 1급

- 1) 1990~1991·1993·1998년 : I·II·III 1~5
- 2) 1992·1994년 : I·II·III 1~7
- 3) 1995~1996·1999년 : I·II·III 1~6
- 4) 1997년 : I·III 1~5 (저작권 문제로 미공개 문항인 II 생략)
- 5) 2000~2001년 : I·II 1~4·III 1~3
- 6) 2002·2004·2008·2009년 1차 : I·II 1~4·III 1~4
- 7) 2003년 : I·II 1~3·III 1~4
- 8) 2005~2007·2009년 2차 : I·II 1~4·III 1~5

b. 구JLPT 2급

- 1) 1990년 : I·II·III 1~4
- 2) 1991·1999년 : I·II·III 1~7
- 3) 1992·1995년 : I·II·III 1~5
- 4) 1993~1994·1996~1998년 : I·II·III 1~6
- 5) 2000년 : I·II 1~3·III 1~6
- 6) 2001년 : I·II·III 1~4
- 7) 2002·2004~2005·2007·2009년 2차 : I·II 1~3·III 1~5
- 8) 2003·2008·2009년 1차 : I·II 1~3·III 1~4
- 9) 2006년 : I·II 1~3·III 1~3

c. 구JLPT 3급

- 1) 1990~1991년 : IV
- 2) 1992년 : IV 1~4
- 3) 1993~1996년 : V 1~4·VI
- 4) 1997~2009년 2차 : V·VI

d. 구JLPT 4급

- 1) 1990년 : V 1~6
- 2) 1991년 : V 1~5
- 3) 1992~1993년 : V 1~3
- 4) 1994~1995년 : VI 1~3
- 5) 1996~1999~2001년 : V·VI 1~3

(5) [구JLPT 2008년 기출문제 중 독해·문법 영역 III-(5)번 문항 예시]

III-(5) 下のグラフは、1971年度と2007年度に行われた調査での「会社を選ぶときどんなことを最も重視したか」という質問に対する新入社員の答えをまとめたものである。

그런데, 이와 같은 통계 자료를 텍스트 가독성 판단에 적용함에 있어서 유의할 점이 있는데, 우선 (3)a'일본어 텍스트의 길이'의 경우, 기출문제 지문의 단락 수·문장 수·형태소 수의 평균값을 단순 비교하여 텍스트의 가독성을 판단할

수 없다는 점이다. 이는 구JLPT 기출문제의 문항 유형에 따라 단락 수 및 문장 수 그리고 형태소 수는 텍스트의 레벨과 무관하게 변화할 수 있어 그 값의 편차가 크기 때문이다. 구JLPT 1급의 독해 지문이라고 하더라도, ‘한자’와 ‘문법 요소’의 난이도는 높지만 텍스트의 절대 길이가 짧은 경우도 있는데, 실제로 1급 기출문제의 독해 지문의 텍스트 당 평균 단락 수가 약 4.27개인 반면, 구JLPT 2급의 텍스트 당 평균 단락 수는 4.51개로, 구JLPT 2급의 평균 단락 수가 1급의 것을 상회하고 있는 것을 알 수 있다. 또한 구JLPT 3급과 4급의 경우, 한 단락이 한 두 문장으로 구성된 경우를 많이 찾아볼 수 있기 때문에, ‘텍스트 당 평균 단락 수’·‘텍스트 당 평균 문장 수’ 통계가 텍스트의 레벨과 무관하게 문항 유형에 의해 영향을 받을 수 있다. 예를 들어, 구JLPT 2급의 ‘텍스트 당 평균 단락 수’와 ‘텍스트 당 평균 문장 수’가 각각 4.51개, 12.18개인 반면, 3급의 경우 5.4개, 14.06개로 2급의 것을 상회하는 결과를 보인다.

따라서 본 연구에서는 위와 같은 문항의 유형에 따른 영향을 최소화하기 위해, (3)a의 분석 항목 중, 다음의 (6)과 같은 분석 항목을 주요 기준으로 삼아 구JLPT 독해 지문 텍스트(4 참조)의 통계를 분석하고 그 결과를 통계적 기법을 사용하여 일본어 텍스트의 가독성 분석 기준을 상정했다. 이와 같이 상정된 상세한 기준을 앞서서 언급한 (3)b‘일본어 텍스트의 표기’의 기준과 함께 AJ-JpnRa Tool에 탑재했는데 이를 구체적으로 정리하면 다음의 표1~4·그림 1과 같다.

(6) [(3)a·b 중에서 채택한 JLPT 독해 지문 데이터베이스 분석 항목 및 목적]

- a. 단락 당 문장 수 : 단락 구조와 단조로움의 정도를 파악
- b. 텍스트·단락·문장 당 형태소 수 : 텍스트·단락·문장의 길이 및 이를 통해 텍스트의 단조로움의 정도를 파악
- c. 한자·히라가나·가타카나 비율 : ‘전제 문자 대비 한자 비율’과 ‘가나 VS 한자의 사용 비율’을 통해 난이도 정도를 파악

표 1. 구JLPT 독해 지문의 ‘일본어 텍스트의 길이’의 평균 μ

평균 μ	n	단락당 문장 수	텍스트당 형태소 수	단락당 형태소 수	문장당 형태소 수
1급	169	3.607810651	318.6982249	101.3119527	29.1904142
2급	169	3.193727811	260.6153846	74.02218935	23.69402367
3급	52	2.98	152.2307692	39.89961538	13.41961538
4급	47	2.559574468	54.4893617	26.35021277	10.8612766

표 2. 구JLPT 독해 지문의 ‘일본어 텍스트의 길이’의 표준편차 σ

표준편차 σ	n	단락당 문장 수	텍스트당 형태소 수	단락당 형태소 수	문장당 형태소 수
1급	169	2.011769077	225.3068227	55.9217492	8.985781914
2급	169	1.542466237	179.4232825	37.65785012	7.652078018
3급	52	1.289769701	87.80473331	24.00636476	5.725971092
4급	47	1.415095568	44.32286305	21.02281602	6.293104445

표 3. 구JLPT 독해 지문의 ‘일본어 텍스트의 표기’의 평균 μ ·표준편차 σ

표준편차 σ	n	평균 μ		표준편차 σ	
		전체대비 한자비율	한자·가나 비율	전체대비 한자비율	한자·가나 비율
1급	169	0.283826627	0.477100592	0.065132484	0.177373014
2급	169	0.238913609	0.381360947	0.058897769	0.151700421
3급	52	0.140842308	0.194615385	0.040609495	0.069690113
4급	47	0.061774468	0.084680851	0.054051169	0.078070511

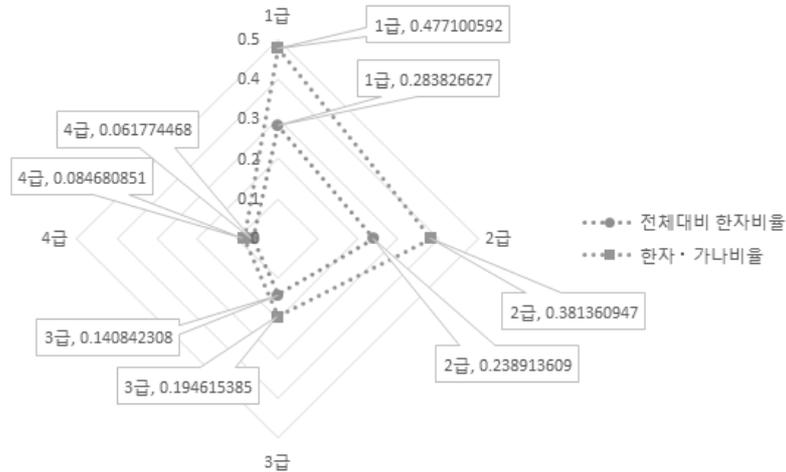


그림 1. 구JLPT 독해 지문의 ‘전체대비 한자비율’ 및 ‘한자·가나 비율’의 평균_μ

그런데, 2010년부터 새롭게 실시되고 있는 新JLPT의 경우, 그 출제기준이 아직 공개되어 있지 않으며, 구JLPT와 달리 N3급(구JLPT 2급과 3급 사이의 난이도⁵⁾)이 추가 되었다. 따라서 본 연구에서는 위와 같은 ‘일본어 텍스트의 길이’에 대한 기준을 통해 분석한 텍스트는 구JLPT의 기준에 따라 텍스트 가독성 레벨을 부여하고 이를 다시 新JLPT 레벨로 환산하여 가독성 결과를 도출하는데, 그 구체적인 환산표는 다음의 (7)과 같다.

(7) [‘일본어 텍스트의 길이’ 텍스트 가독성 레벨 환산 표]

- a. 구JLPT 1급 ⇨ 新JLPT N1급
- b. 구JLPT 2급 ⇨ 新JLPT N2급
- c. 구JLPT 3급 ⇨ 新JLPT N3·4급
- d. 구JLPT 4급 ⇨ 新JLPT N5급

3.1.2 일본어 텍스트에 사용된 한자 난이도

이어, (2)c‘일본어 텍스트에 사용된 한자 난이도’의 경우 다음의 세 가지 데이

5) 일본어능력시험 공식 사이트(<http://jlpt.or.kr>) 참조.

터베이스(표2)를 기반으로 텍스트의 한자 난이도를 판별하는 데, 우선 자체 조사를 통해 구축한 최근 20년간의 구JLPT(1급~4급)의 기출한자⁶⁾ 2195자를 기반으로 한 「일본어능력시험 기출 한자」 데이터베이스를 구축했다. 그리고 일본 초등학교의 학년별 교육한자인 『小学校学習指導要領』의 부록 『学年別漢字配当表』 1006자 및 ‘중학교학년별교육한자’ 1130자 그리고 이를 합한 일본어의 상용한자인 『常用漢字』 2136자를 여기에 더해 한자 난이도 분석용 데이터베이스를 구축했다.

단, 신JLPT의 경우 앞서 언급한 바와 같이 출제기준이 공개되어 있지 않으며, 구JLPT와 달리 N3급이 추가되었기 때문에, ‘AJ-JpnRa Tool’은 구JLPT를 기반으로 하고, 여기에 공개된 신JLPT의 출제 기준을 반영하여 기준을 수정했다. 특히 N3의 경우, 필자가 설정한 임의 기준에 따라 한자 난이도 기준을 설정했는데, 본 ‘JLPT 기출한자 데이터베이스’의 구체적인 난이도 구분 기준은 다음의 (8)과 같다. 그리고 이를 바탕으로 선정한 AJ-JpnRa Tool의 JLPT 한자 난이도 설정 기준 목록의 일부 예시는 (9)와 같다.

표 4. 『学年別漢字配当表』, ‘중학교학년별 교육한자’, 『常用漢字』 및 ‘JLPT 기출한자’ 데이터베이스 디자인(김유영, 2013b:27)

필드 이름	데이터 형식	비고
no	숫자	일련번호
character	짧은 텍스트	한자
character_ver	짧은 텍스트	이체자
radical	짧은 텍스트	부수
strokes	숫자	획수
lv	짧은 텍스트	레벨
add_year	숫자	추가년도
read01	짧은 텍스트	읽기 / 음
read02	짧은 텍스트	
read03	짧은 텍스트	
read04	짧은 텍스트	
read05	짧은 텍스트	
mean01	짧은 텍스트	뜻 / 훈
mean02	짧은 텍스트	
mean03	짧은 텍스트	
mean04	짧은 텍스트	
mean05	짧은 텍스트	
mean06	짧은 텍스트	
mean07	짧은 텍스트	
mean08	짧은 텍스트	
mean09	짧은 텍스트	
mean10	짧은 텍스트	
eg01	짧은 텍스트	예문
eg02	짧은 텍스트	
eg03	짧은 텍스트	

6) 新新 일본어능력시험의 언어지식 중 문자·어휘 영역 기준.

(8) [AJ-JpnRa Tool의 한자 데이터베이스 난이도 설정 기준]

- a. 구JLPT와 신JLPT의 출제 기준과 구JLPT의 기출문제를 기반으로 각 한자의 난이도 설정. 단, N3의 경우 구JLPT의 2급·3급의 기출 문제를 합해, 그 중 가장 빈도가 높은 항목을 기준으로 설정. e.g. 신JLPT N2: 기존시험(구JLPT)의 2급과 거의 같은 난이도⁷⁾
 - b. 기출 문제는 하위 급수를 우선 기준으로 적용하여 난이도 설정. e.g. N4와 N5에 공통된 기출문제의 경우, N5 난이도로 설정
 - c. JLPT의 문자·어휘·문법 영역의 경우, 기출문제의 '지문'과 '정답'에 출현한 한자에만 한정하여 난이도 설정.
- e.g. 1990년 구JLPT 4급 문자·어휘 영역 기출문제의 경우, 1번(持) 정답에 출현한 한자만을 선정.
 : II-3-(2) もって → 1. 持つて(○) / 2. 待つて / 3. 侍って / 4. 持つて

(9) ['AJ-JpnRa Tool'의 JLPT 기출 한자별 난이도 설정 기준 목록-일부 발췌]

- a. 『学年別漢字配当表』의 초등학교 1학년 한자 : 70자
 一右雨円王音下火花貝学氣九休玉金空月犬見五口校左三山子四糸字耳七車手
 十出女小上森人水正生青夕石赤千川先早草足村大男竹中虫町天田土二日入年
 白八百文木本名目立力林六
- b. 중학교 학년별 교육한자 중 일부 : 550자 중, 200자 발췌
 駐芝審槽浦弔寮旋腐舶軟劬搬芋寧馱宛宜圈皆彈惱載還芳杉崎患微較稻硫硬泊
 香弧又邪妨慈猶津荒猫怠繁即碎伐塞騷占塗鍵慮迎慢遡魔脂迅擊菌睡蚊督酬腺
 籍徐撲濕迫淚漏循撤辛燥恣徵濯徹房湧枋埋充尿附衡互堀涼茨涉敏洪御眠双渴
 涯鉛髮乏玩菓敷釜懇懸途簿乾弘淡赴渡訴積懲壞滯壤扇滑壁閑拔滋麻扶裕默滅
 丈称暇援与殿吐及姬忙潜尋戒祈祉墜茂唾旬抄尺歲憩尾潰戾醇培仙崎翫寮粒遣
 袖頼偏鈍賠袋弁介苗瀬旨溶腰欄拓既旦礎麵契被晶奈畜網滂普聽添棄 ……
- c. 상용한자(초등학교 및 중학교 학년별 교육한자) 범위 밖 한자 : 85자
 猪磐嚙銚雫悸痰窺拌癩孟涌喀區溜迂拇庄鯨珂空龍鋒伊攬駿鳴筑旭濡勿罹嶼耶
 牡隴幌竈咬烏柵笹亘咳鞅館嶺槌釘柏洩喘俳菽幡笠幹叱柴徊媿桿只塙阿榭僭梯
 侯梢鏘貴榜頰杜輻霞杖礫栖栗埃汲沫這
- d. JLPT 기출한자 중 N3급 한자 : 301자
 配伝費居申冷迎具全段政公皮側販季鉄米共取受財枝老支反熱将參存孫司流定
 案輸美性録号解普冊葉臣可保非許虫戸募祝神易原役無要府返州玉王割旧選能

7) 일본어능력시험 공식 사이트, 새로운 JLPT 인정기준
 : <http://jlpt.or.kr/jlpt/jlpt4.asp?Mcode=4> 참조.

械觀論第番庁慣活園信面殺寢折実初根数格量節省畑久糸庭祖任例幸忘肌島變
 券身对在敗草荷当妻枚制達利示替育压失再的過減込官宙価約現角精胃忙眠若
 城汚浴守算完占丸軍形科涙宇容路岩石宅宿個組曲温央決箱兩談由並移付式客
 血渡限愛歳歴他互菓経登痛刊願貝器鼻材余期造連繪)香表応良打払福毛仮勞
 苦景令助骨級氷加必麦欲単亡予条谷座内念陸欠次細礼倍部法呼命泉和局善成
 息最大章接徒到型順卵告交駐營線点干平窓緑向幸未未団札波比固相湯列位困
 機用化各関直落吸針系飛因般包件馬泣差様残油治濯酒甘席続柱指際

e. JLPT 기출한자 중 N5급 한자 : 103자

千火母一食七分万東話電山読飲八氣来足校社間週道会左毎南休土月北日出水今
 高言少車書門長小日店木四本安手九外多立女子上白雨先行年国生五新天二馱金
 前耳時父大人右中入語学名男口古百何見花六後買開空友川午半西魚三十下

김유영(2013b:28·31)을 바탕으로 추가 및 수정

따라서 AJ-JpnRa Tool의 (2)c‘한자 난이도’ 판단은, 3절에서 언급한 최근 20년간의 구JLPT(1급-4급) 독해·문법 영역 중 독해 지문을 위 (8)·(9)와 같은 기준을 통해 구축된 한자 난이도 데이터베이스를 이용하여 통계적으로 분석한 결과를 기준으로 삼아 이루어진다. 그 구체적인 분석 결과는 다음의 표 5·6과 같다.

표 5. 구JLPT 독해 지문의 ‘일본어 텍스트에 사용된 한자 난이도’ 평균₁

한자 별 난이도 구JLPT 독해지문	N5	N4	N3	N2	N1	JLPT 초과
1급	0.2873373	0.2649018	0.2391314	0.1434544	0.0630834	0
2급	0.3463166	0.2783047	0.2068018	0.1168757	0.0486101	0
3급	0.6386635	0.3434442	0.0116654	0.00395	0.0022788	0
4급	0.7862191	0.0222915	0	0	0	0

표 6. 구JLPT 독해 지문의 ‘일본어 텍스트에 사용된 한자 난이도’ 표준편차 σ

한자 별 난이도 구JLPT 독해지문	N5	N4	N3	N2	N1	JLPT 초과
1급	0.0104092	0.0066545	0.0060325	0.0034599	0.0013689	0
2급	0.0115388	0.0086308	0.0054418	0.003596	0.0014084	0
3급	0.020431	0.0191962	0.0013897	0.0003437	0.0001975	0
4급	0.1534614	0.0039973	0	0	0	0

3.1.3 일본어 텍스트 가독성 판단 분석 모형

‘AJ-JpnRa Tool’의 가독성 판단 분석은, 3.1.1과 3.1.2절에서 제시한 구JLPT의 독해지문의 난이도 판정 기준, 즉 구JLPT의 각 레벨 별 ‘일본어 텍스트의 길이’, ‘일본어 텍스트의 표기’, ‘일본어 텍스트에 사용된 한자 난이도’ 각각의 평균(μ)과 표준편차(σ)에, 가독성을 분석하고자 하는 텍스트의 각각에 해당하는 값(x)을 다음 (10)의 식에 대응하여 표준값(z)을 구한 후, 이 텍스트의 원수치(x)가 각각의 레벨의 평균과 얼마나 상이한가를 비교하여 가독성을 판단한다. 참고로 표준값(z)이란, 원수치(x)가 평균에서 얼마나 떨어져 있는지를 나타내며, 음수이면 평균이하, 양수이면 평균이상임을 나타낸다.

(10) **표준값(Z-Score, Z-Value)** : 통계학적으로 정규분포 안에서 각각의 경우 혹은 값이 표준편차 상에 어떤 위치를 차지하는지를 계산

$$z = \frac{x - \mu}{\sigma}$$

(11) [1999년 1급 기출문제, 독해·문법 I-1 문항 판독 결과 일부 예시]

1. テキストの長さと言記におけるレベル判定 :
 - 文書/パラグラフ : **N1級(0.19)** > N2級(0.52) > N3~N4級(0.79) > N5級(1.02)
 - 形態素/テキスト : **N1級(1.58)** > N2級(2.30) > N3~N4級(5.94) > N5級(13.98)
 - 形態素/パラグラフ : **N1級(0.20)** > N2級(1.02) > N3~N4級(3.02) > N5級(4.09)
 - 形態素/文書 : **N1級(0.12)** > N2級(0.57) > N3~N4級(2.56) > N5級(2.74)

漢字/テキスト : **N1級(0.32)** > N2級(1.11) > N3-N4級(4.03) > N5級(4.49)

漢字/仮名 : **N1級(0.24)** > N2級(0.91) > N3-N4級(4.67) > N5級(5.58)

2. テキストにおける漢字レベルによるレベル判定 :

- ➔ N1級として : N5級の漢字は平均値より‘-0.4’離れていて, 平均値以下 / N4級の漢字は平均値より‘0.2’離れていて, 平均値以上 / N3級の漢字は平均値より‘-0.1’離れていて, 平均値以下 / N2級の漢字は平均値より‘0.8’離れていて, 平均値以上 / N1級の漢字は平均値より‘-0.5’離れていて, 平均値以下.

3.2 ‘AJ-JpnRa Tool’의 사양 및 설치

‘AJ-JpnRa Tool’은 윈도 기반의 32비트 프로그램으로, Visual Studio .NET 2010과 MS-Access 2010을 기반으로 개발되었으며, 주요 기능 및 자세한 시스템 요구 사양은 다음의 (12)·(13)과 같다.

(12) [‘AJ-JpnRa Tool’의 주요 기능]

- a. ‘일본어 텍스트의 길이’ 및 ‘한자’ 기반 일본어 텍스트 가독성 분석
 - 1) 단일 파일 혹은 폴더 내 복수 파일 일괄 분석
 - 2) 일본어와 한국어 등 복수의 언어를 포함한 다언어 텍스트 분석 가능
- b. 단일 파일 형태소 분석(차센茶筌, 띄어쓰기, 가나읽기 형식 지원)

(13) [‘AJ-JpnRa Tool’의 시스템 요구 사양 및 배포(다운로드)]

- a. 운영체제 : Window XP 이상, Window Vista, Window 7, Window 8, Window 8.1
- b. 메모리 : 최소 1G 이상
- c. 하드 디스크 : 최소 100MB
- d. 필요 프로그램 : 반드시 Microsoft Access 2010 이상 버전이 설치되어 있을 것
- e. 기타 사항 : 형태소 분석기 Mecab 0.996 내장, 최신버전은 ‘AJ-JpnRa Tool 1.2.0.2’
- f. 프로그램 다운로드 및 지원
 - 1) <http://www.japanese.or.kr/JapaneseUtile.aspx>
 - 2) <http://sourceforge.net/projects/ajjpnratool>
- g. 메일링리스트 : <https://groups.google.com/forum/#!forum/aj-jpnra-tool>

상기 (13)f에서 다운로드 받은 AJ-JpnRa Tool구체적인 프로그램의 설치방

법은 다음의 그림2의 순서와 같다. 단, AJ-JpnRa Tool을 인스톨 하고 사용하는 데에 있어서 주의할 점은 다음의 (14)과 같다.

(14) [AJ-JpnRa Tool의 설치 및 사용상 유의점]

- a. Microsoft Access 2010 이상 : AJ-JpnRa Tool은 반드시 Microsoft Access 2010 이상의 버전이 설치되어 있는 환경에서만 사용가능
- b. 인스톨 시, 그림2-c 단계에서 반드시 'MeCab 설치를 체크'하여 선택할 것
- c. 그림2-d와 같은 단계에서 반드시 인코딩을 'UTF-8'로 선택할 것
- d. 그림2-e와 같은 단계에서 반드시 '예(Y)'를 선택할 것
- e. 분석 대상은 'UTF-8(유니코드)'형식의 텍스트 파일일 것 e.g. ○○○.txt

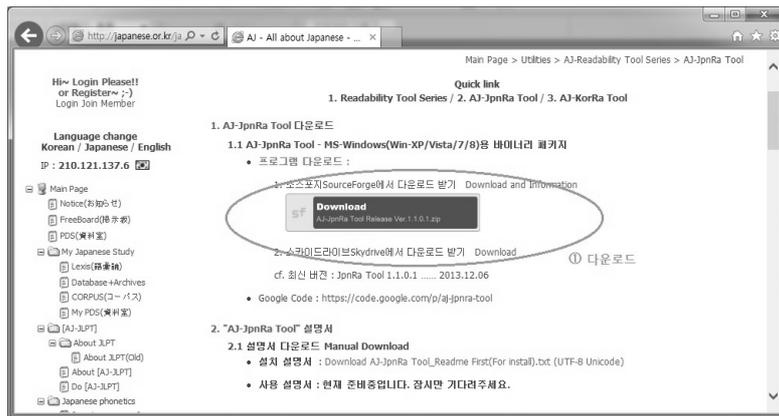


그림 2-a. 프로그램 다운로드

http://japanese.or.kr/japanesutil/Readability Tool Series/AJ-JpnRa_Tool.aspx



그림2-b. 다운로드 파일의 압축해제 및 'Setup.exe'파일을 클릭하여 설치

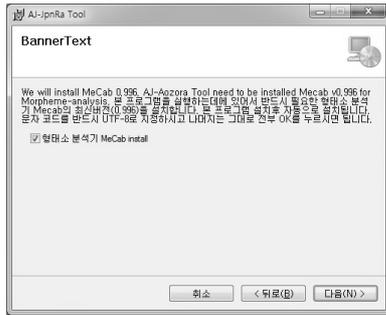


그림2-c. MeCab설치 선택

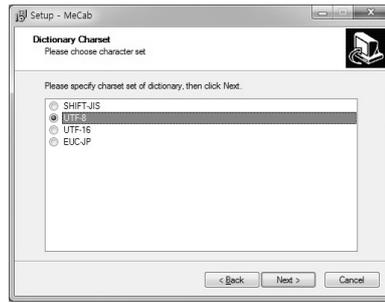


그림2-d. 반드시 UTF-8 선택

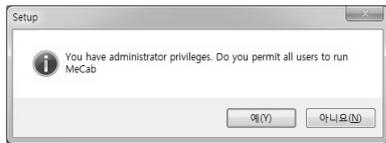


그림2-e. MeCab설정 '예(Y)' 선택

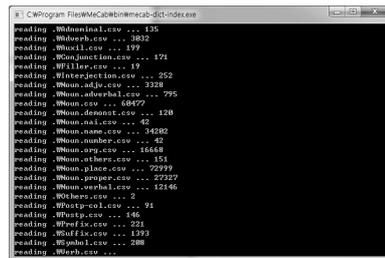


그림2-f. MeCab설정 완료화면



그림2-g. 바탕화면의 아이콘 중 좌측의 'AJ-JpnRa Tool'을 클릭하여 실행

3.3 'AJ-JpnRa Tool'의 실행화면 및 사용법

3.3.1 일본어 텍스트 가독성 분석

AJ-RaTool 프로그램의 메인 기능은 '일본어 텍스트의 가독성 분석'으로, 여기에는 폴더에 들어 있는 텍스트 파일을 일괄 분석하는 기능(이하, 폴더 일괄 분석)과 단일 텍스트 파일을 분석하는 기능(이하, 단일 텍스트 분석)이 탑재되어 있다. 우선, 폴더 일괄 분석의 경우, 가독성 분석을 원하는 일본어 텍스트를 하나의 폴더에 모아 두고, 그림3-a과 같이 「File」-「Open Jpn Text Folder and

Analysis - 漢字レベル」 메뉴를 실행한 후, 원하는 폴더를 지정하면 자동 분석이 진행된다. 그리고 단일 텍스트 분석은 그림3-b와 같이 좌측 하단의 「教育漢字・JLPT漢字レベル判定」을 선택하면 텍스트 파일을 지정하는 팝업 창이 뜨는데, 여기에서 분석하고자 하는 일본어 텍스트 파일을 지정하면 해당 파일의 가독성 분석이 자동으로 진행된다. 단, 텍스트의 레벨을 판단하는 가독성 판단의 경우 단일 파일 분석의 경우에만 가능하며, 폴더 일괄 분석은 언어 통계만을 제공한다.

그리고 텍스트 파일은 공히 UTF-8(유니코드)의 인코딩으로 저장된 텍스트일 경우에만 문제없이 분석을 행할 수 있으므로, 반드시 분석하고자 하는 텍스트 파일을 사전에 UTF-8 형식으로 인코딩을 변환해 둘 필요가 있다.

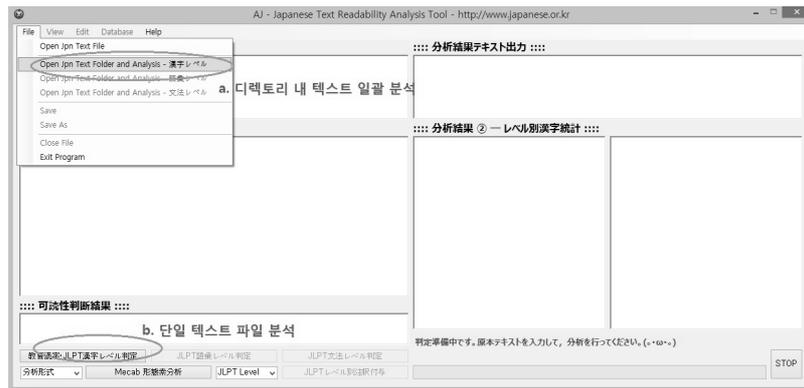


그림 3. 일본어 텍스트 가독성 분석

위와 같이 일본어 텍스트의 가독성 분석을 실시하면, 다음의 그림4와 그림5(팝업)와 같은 분석 결과를 얻을 수 있는데, 결과 화면에 대한 자세한 사항은 이어지는 (14)와 같다.

(14) [일본어 텍스트의 가독성 분석 결과 화면 설명]

- a. 좌측 상단의 「**原本テキスト情報**」 결과 창 : 그림4-a
 - 1) 폴더 일괄분석 : 분석 대상 폴더 경로와 파일 개수 정보를 출력
 - 2) 단일 텍스트 분석 : 분석하고자 하는 텍스트 전문 출력
- b. 우측 상단의 「**分析結果テキスト出力**」 결과 창 : 그림4-b
 - 1) 폴더 일괄분석 : 결과 없음
 - 2) 단일 텍스트 분석 : 해당 텍스트의 형태소 분석 결과(分かち書き) 출력
- c. 좌측 중단의 「**分析結果①-全体統計**」 결과 창 : 그림4-c
 - : 한자·가나 비율 및 문단·문장·형태소 수 통계
- d. 우측 하단의 「**分析結果②-レベル別漢字統計**」 결과 창 : 그림4-d
 - : 교육한자의 학년별·JLPT 레벨별 출현 한자 통계(연어 수)
- e. 좌측 하단의 「**可読性判断結果**」 결과 창 : 그림4-e
 - : 상기 c·d의 결과를 종합하여 일본어 텍스트 가독성 판단 결과 출력
- f. 팝업 된 「**教育課程・常用漢字・JLPTレベル別, 出現漢字一覧表**」 결과 창 : 그림5
 - : d의 통계에 출현한 교육한자의 학년별·JLPT 레벨별 한자 목록 및 통계(개별어 수)

3.3.2 일본어 텍스트 형태소 분석

AJ-JpnRa Tool은 일본어 형태소 분석기 MeCab를 내장하고 있어, 일본어 텍스트 가독성 분석뿐만 아니라, 간단한 일본어 형태소 분석도 가능하다. 일본어 텍스트의 형태소 분석을 수행하기 위해서는 우선, 프로그램을 실행하고, 「File」 메뉴의 「Open Jpn Text File」을 선택하고 텍스트 파일을 열어 「原本テキスト情報」 창에 텍스트를 입력하거나, 직접 「原本テキスト情報」 창에 일본어 텍스트를 입력한다. 그 후에, 그림6-a에서 보이는 좌측 최 하단의 「分析形式」를 선택하여 「茶筌形式分析」「分かち書き分析」「読み分析」중 원하는 분석 형식을 선택한 후, 그림6-b에서 보이는 「MeCab 形態素分析」을 선택하면 우측 상단의 「分析結果テキスト出力」 결과 창(그림6-c)에서 형태소 분석 결과를 확인할 수 있다.

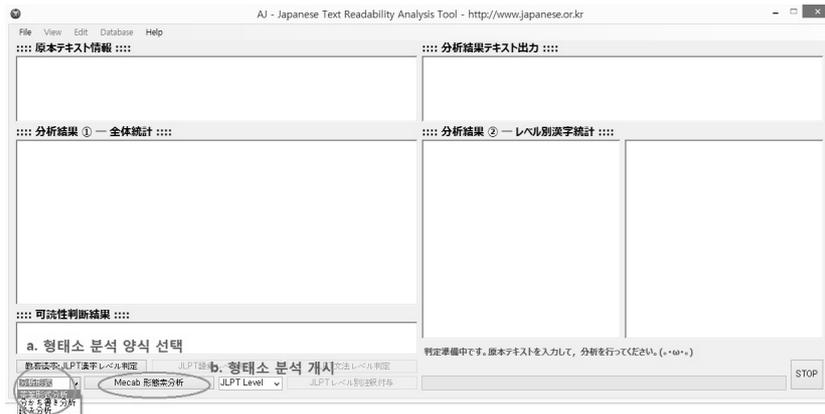


그림 6. 일본어 텍스트 형태소 분석 메뉴

4. 'AJ-JpnRa Tool'을 활용한 연구 예

4.1. 일본어 교육과 학습

지금까지 일본어학습자와 교수자를 위한 일본어 학습 및 교육 시스템은 수 없이 많이 개발되어 왔으며, 최근에 이르러서는 정보통신기술의 발달과 대중화에 의해, ICT(Information & Communication Technology)를 활용한 일본어 학습 및 교육 시스템이 크게 주목받고 있다. 이와 같은 ICT에 의한 교육 및 학습 환경의 변화는 일본어에도 많은 영향을 미쳤는데, 그 중 한 가지를 들자면 우선 일본어 텍스트의 전자화를 들 수 있을 것이다. ICT를 활용한 e-Learning이 확대됨에 따라 PC로부터 무선 모바일 기기에 이르기까지 전자 텍스트가 대량으로 생산되고 있으며 그 필요성 또한 더욱 더 높아지고 있다. 이러한 전자 텍스트의 생산량과 그 수요는 지금까지의 전통적인 매체 즉, 인쇄 출판물을 넘어서기 시작했으며, 단순히 기존의 인쇄 출판물을 전자 텍스트화하는 데에서 그치지 않고, 텍스트의 생성 단계에서부터 대량의 전자 텍스트로 생산되어 소비되고 또한 그대로 소멸되는 시대로 접어들게 되었다. 일본어 학습에 다양

하게 응용할 수 있는 일본어 전자 텍스트는 인터넷 뉴스나 웹 사이트, 그리고 블로그나 SNS 등을 통해 급속도로 성장하고 있으며, 상업적 교재뿐만 아니라 공교육의 교과서에 이르기까지, 기존의 매체들도 속속 전자 텍스트 화에 동참하고 있다. 이와 같은 전자 텍스트가 ICT를 활용한 일본어 학습 및 교육에 있어서 유용할 것이라는 것은 자명한 사실이지만, 이들을 모두를 그대로 학습과 교육에 사용할 수는 없다. 일본어 교육 및 학습에 유효한 텍스트 즉, 학습자의 레벨에 맞으면서도 올바른 일본어 전자 텍스트를 선별하거나 생산하여 이를 일본어 교육과 학습에 응용할 필요가 있기 때문이다. 그리고 이는 두말할 필요도 없이 전자 텍스트뿐만 아니라, 종래의 일반 인쇄 출판물에 사용되는 일본어 텍스트에도 해당되는 내용이라고 할 수 있다.

따라서 본 연구를 통해 개발된 AJ-JpnRa Tool과 같이 일본어 전자 텍스트의 가독성 기준을 정립하고 이를 판단하는 수단을 교수자와 학습자에게 제공하는 것은 수준별 일본어 교육과 학습에 필요한 텍스트를 확보 및 생산 가능할 수 있게 해 주어 더욱 효과적인 일본어 학습과 교육을 실시하는 것을 가능케 한다. 또한 AJ-JpnRa Tool을 이용한 일본어 텍스트 가독성 분석을 통해 텍스트가 그 대상으로 삼고 있는 학습자들의 수준에 적합하게 제공되고 있는가에 대한 사후 검증이 가능하게 된다. 그리고 학습자 스스로도 자신의 수준에 맞는 일본어 텍스트를 찾을 수 있게 되는 등, 일본어 텍스트의 가독성 분석을 누구나 쉽게 간단히 행할 수 있게 됨에 따라, 학습자의 자기 주도형 학습에도 도움이 된다. 예를 들어, 드라마 및 영화 대본 그리고 노래 가사 및 인터넷 신문 기사 등 일본어 학습자가 스스로 얻을 수 있는 일본어 텍스트는 무한하지만, 이들 텍스트가 과연 학습자 스스로에게 적합한 수준의 텍스트인가를 판단하는 것은 쉬운 일이 아니며, 이는 교수자의 입장에서 바라보아도 그 상황은 크게 달라지지 않는다.

4.2. 알기 쉬운 일본어(やさしい日本語)

그리고 일본어 텍스트 가독성 분석은, ‘알기 쉬운 일본어(やさしい日本語)’ 등 외국인을 대상으로 언어적 장벽을 완화 혹은 없애기 위한 일본어 텍스트의 작성 규정 및 기존 텍스트의 분석을 위한 기초 연구 자료로 응용될 수 있다.

김유영(2013b)에서 언급한 바와 같이, 일본 후생노동성의 ‘공적 정보제공’을 위한 텍스트를 AJ-JpnRa Tool을 응용하여 학년별 교육한자와 일본어능력시험의 레벨별 한자와의 대조를 실시한 결과, 현재 일본 후생노동성에 의한 외국인 주민을 대상으로 하는 공적 정보제공을 목적으로 하고 있는 일본어 텍스트에는 외국인 주민이 ‘배리어’를 크게 느낄 수 있을 만큼 높은 비율의 한자가 사용되고 있으며, 사용된 한자의 레벨 또한 높아, 적극적인 ‘알기 쉬운 일본어’에 대한 대응의 필요성이 절실하다는 것을 확인할 수 있었다(그림7 참조). 추가하자면 이와 같은 ‘공적 정보제공’이라는 측면에서 ‘알기 쉬운 일본어’에 관한 논의는 비단 일본뿐만 아니라 급격하게 다문화 사회로 진입하고 있는 한국에 있어서도 시사하는 바가 크다.

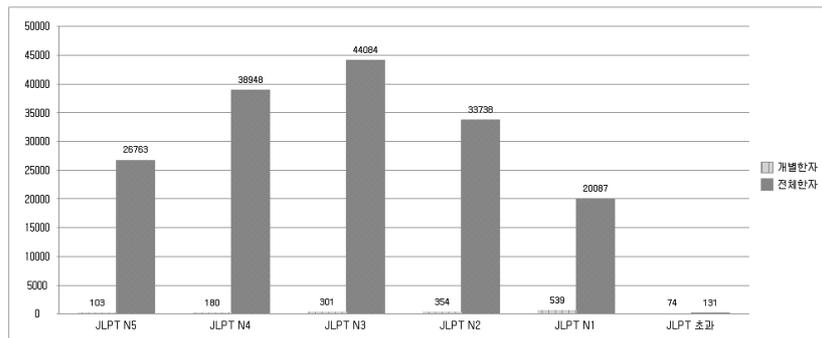


그림 7. AJ-JpnRa Tool을 사용하여 일본 후생노동성이 제공하는 ‘동일본대지진 관련 공적 정보제공’을 위한 일본어 텍스트를 분석한 결과(김유영, 2003b:33)

5. 나가며

본 연구를 통해 개발된 일본어 텍스트 가독성 분석 프로그램 ‘AJ-JpnRa Tool’은 예를 들어 공적 기관이 정보제공을 위해 일본어 텍스트를 작성할 때, 그리고 교수가 일본어 교육을 위해 학습자의 능력에 맞는 텍스트를 작성하

거나 학습자에게 기존에 존재하는 일본어 텍스트 중에서 학습자의 수준에 맞는 가독성을 가진 일본어 텍스트를 선택하여 제공할 때와 같은 경우 등 다양한 분야의 일본어 연구와 교육에 응용할 수 있다.

현재 'AJ-JpnRa Tool'의 기능은 한자와 한자 어휘에 한정된 텍스트 가독성 분석에 한정되어 있으나, 프로그램 상의 메뉴에서 찾아 볼 수 있듯이 앞으로 일본어 어휘·문법 등에 이르기까지 보다 더 다양한 측면에서의 종합적인 분석 및 고찰을 가능하게 할 수 있도록 프로그램을 보완해 나아갈 예정이다. 그리고 앞선 4절에서 언급한 연구 예시를 실제로 수행하여 일본어 텍스트 가독성에 관한 연구와 논의를 계속해서 진행하고자 한다.

【参考文献】

- 김유영(2012) 「일본어 연구를 위한『靑空文庫(아오조라문고)』 데이터베이스의 구축과 활용-전자 텍스트 처리 프로그램 "AJ-Aozora Tool"을 활용한 데이터베이스 구축 모델 개발」『일본어학연구』35, 한국일본어학회, pp.35-53.
- _____ (2013a) 「일본어 연구를 위한『靑空文庫』 데이터베이스 구축 프로그램」『AJ-Aozora Tool』『日本言語文化』24호, 한국일본어문화학회, pp.131-150.
- _____ (2013b) 「동일본대지진과 일본사회의 언어-외국인 주민을 대상으로 하는 ‘공적 정보제공’에 있어서의 ‘알기 쉬운 일본어やさしい日本語-」『일본학보』97집, 한국일본학회, pp.17-38.
- 이도열(2012) 「일본어능력 평가시험간의 비교적도 고찰: JLPT N1 레벨과 EJU 일본어 득점에 대한 비교적도 분석을 중심으로」『日本語文學』Vol.52, 한국일본어문화학회, pp.55-77.
- 大隅敦子(2009) 「新しい「日本語能力試験」--構成概念の構築と新しいレベルの設定(特集 日本語教育の最新トピック(2))」『日本語学』28-11, 明治書院, pp.24-35.
- 加藤扶久美・岩本阿由美・遠藤祥子(2007) 「漢字教材用データベースの活用」『富山大学留学生センター紀要』6, 富山大学留学生センター, pp.1-12.
- 河村よし子(1998) 「読解のためのレベル判定システムの構築: 語彙チェッカーの開発と活用」『日本語教育方法研究会誌』5-2, 日本語教育方法研究会, pp.10-11.
- サワシュボラ・林良彦(2006) 「漢字の自律学習のための読み誤りの自動生成(<特集> 言語の学習・教育)」『電子情報通信学会技術研究報告. TL, 思考と言

- 語』106(363), 一般社団法人電子情報通信学会, pp.25-30.
- 日本国語大辞典編集委員会編(2001)『日本国語大辞典』第2版, 小学館
- 吉岡頼子(2012)「日本語能力試験における「読解」問題」『日本語教育論集』-21, 姫路獨協大学大学院言語教育研究科日本語教育コース, pp.73-80.
- Akaike, H. 1974. "A new look at the statistical model identification." *IEEE Transactions on Automatic Control* 19-6, 716-723.
- Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*, Cambridge University Press.
- Chall, J. S. 1958. *Readability: An appraisal of research and application*. Columbus, OH: Ohio State University Press, pp.1-202
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.), New York: Academic Press.
- Dale, E. & Chall, J. S. 1948. "A formula for predicting readability." *Educational Research Bulletin* 27, pp.37-54.
- Doak, C. C., L. G. Doak, and J. H. Root. 1996. *Teaching patients with low literacy skills*. Philadelphia: J. P. Lippincott Company, pp.1-207.
- DuBay, W. H. 2006. *Smart language: Readers, Readability, and the Grading of Text*. Costa Mesa: Impact Information, pp.1-151.
- Flesch, R. 1946. *The art of plain talk*. New York: Harpers.
- Fry, E. 1977. "Fry's readability graph: Clarifications, validity and extension to level 17." *Journal of Reading* 21-3, pp.242-252.
- Fry, E. 2002. "Readability versus leveling." *Reading teacher* 56, no. 3:286-292.
- Gray, W. S. and B. Leary. 1935. *What makes a book readable*. Chicago: Chicago University Press, pp.1-388
- Gunning, R. 1952. *The Technique of Clear Writing*, New York: McGraw-Hill.
- Klare, G. R. 1963. *The measurement of readability*. T. Ames, IA: Iowa State University Press, pp.1-358
- McLaughlin, H. 1969. "SMOG grading: A new readability formula." *Journal of Reading* 12-8, pp.639-646.
- Oxford University Press. 2010. *Oxford Advanced Learner's Dictionary*. Oxford University Press.
- Si, L. & Callan, J. 2001. "A statistical model for scientific readability." *Proceedings of the Tenth International Conference on Information and Knowledge Management*, pp.574-576.
- Spache, G. 1953. "A new readability formula for primary grade reading materials." *Elementary School Journal* 53-7, pp.410-413.

일본어 텍스트 가독성 분석 김유영 / 235

Zenger, W. E. & Zenger, S. K. 1976. *Handbooks for Evaluating and Selecting Textbooks*, Belmont, California: Fearon Publishers.

◇투 고 일 : 2014. 01. 30

◇심사개시 : 2014. 02. 24

◇계재확정 : 2014. 03. 30

〈Abstract〉

An analytical Study on the Readability of Japanese text

—With a focus on Developing an Japanese text analysis Program Based
on Previous format of JLPT—

In this paper, we conducted a study on the variety factors influence the text readability'. Especially through an analysis of 'the length of text(paragraphs, sentences, and the number of morphemes)', 'Japanese notation(Kanji, Hiragana, Katakana, etc.)', 'the Chinese character in Japanese text', we determined a certain standard and built a tool(AJ-JpnRa Tool) of 'the readability of Japanese text'.

For this purpose, we built the Japanese Language Proficiency Test(JLPT, 1 to 4 degree) database of the last 20 years: 1990 to 2009.

The 'AJ-JpnRa Tool' established through this study can provide objective criteria for determining the text readability, how easy a certain Japanese text is to read and understand for readers. And through this we will be allowed to have wide application: Japanese-language education and learning and the 'Easy Japanese(やさしい日本語)'.