

# 문서 자동 요약 기술의 발전을 위한 학제적 접근의 필요성과 과제\*

서 혁\*\*

## <차 례>

- I. 머리말
- II. 요약하기 연구와 문서 자동 요약
- III. 학습자와 컴퓨터의 요약 능력 비교  
—MS Word 2003을 중심으로
- IV. 학제적 측면에서의 시사점

## I. 머리말

국어교육이 학문의 한 분야로 자리 잡은 지도 짧게는 20년, 길게는 반 세기를 넘어서고 있다. 2011년에 만 20년으로 성년을 맞이하는 국어교육 학회도 새로운 도약을 위해, 미래지향적 측면에서 연구 영역의 확장과 새로운 연구 방법의 도입이 절실한 때가 아닌가 생각한다.

국어교육은 인문학적이면서도 사회학적 성격을 함께 지니고 있다. 교육 연구 자체가 사회학적, 인문학적, 과학적 성격을 가진 응용적 학문이다. 국어는 도구이자 문화이며, 그 교육 및 연구는 인문학적, 사회학적 성

---

\* 이 논문은 제46회 국어교육학회 학술발표대회(2010. 8. 28)에서 발표한 것을 수정·보완한 것이다.

\*\* 이화여대 국어교육과 교수(shcore@hanmail.net)

찰과 탐구를 지향하되, 기본적으로 과학적, 공학적 접근에도 관심을 둘 필요가 있다. 학문의 발전은 연구 내용, 연구 분야, 연구 방법의 개발·확장·축적에 따른 인적·사회적·역사적 기여에 달려 있다. 국어교육은 지금까지 교육을 통하여 수없이 많은 인적, 사회적 기여를 해 온 것이 사실이지만, 좀 더 과학적이고 체계적인 접근 노력이 없이는 현재의 위상을 유지하기도 쉽지 않을 수 있다. 인접 학문과의 학제적 연구는 새로운 학문의 개발은 물론 기존 학문 영역의 발전을 위해서도 필수적인 과제가 되고 있다.

모든 학문 영역이 그러하듯이 국어교육 역시 궁극적으로 인간의 이해, 기술, 문화와 문명, 존재 모두에 답하고 새로운 의미를 부여할 수 있어야 할 것이다. 학습자, 국어 이해와 표현, 언어문화예술 감상에 초점을 두고 있는 국어 교육 연구도 단순히 텍스트와 학습자라는 나무와 숲에서 더 나아가 인간과 삶의 세계와의 유기적 관계와 발전에 대해서도 관심을 가져야 할 것이다. 2010년 8월에 개최된 국어교육학회 학술대회 주제인 “국어교육과 디지털 리터러시” 역시 이러한 맥락에서 설정된 한 시도로 이해된다. 특히 현재도 그러하지만 미래 사회는 IT와 디지털 기술 및 연구가 인간의 삶에 더욱 중요한 영향을 미칠 것임이 분명하며, 향후 더욱 다양한 접근과 시도가 있어야 할 것으로 생각된다.

이러한 맥락에서 1990년대 이후 비약적으로 발전하고 있는 컴퓨터 정보 과학 분야의 문서 자동 요약(또는 자동 문서 요약, Automatic Document Summarization)과 관련하여 국어교육 전공자의 시각에서 각 영역의 연구 성과들을 간단히 살펴보고, 외람되지만 학제적 상호 참조가 필요하다고 생각되는 몇 가지 의견을 제시하고자 한다.<sup>1)</sup> 특히 기존에 국어교육 분야에서 주로 학습자를 대상으로 한 요약 관련 연구 성과들을 간략히 살펴본 후, 컴퓨터 정보 과학 분야에서 연구되어 온 정보 검색과 문서 자동 요약

1) 참고로 필자는 컴퓨터 공학 분야에는 문외한이지만 컴퓨터를 활용한 정보 검색과 처리, 문서 자동 요약을 중심으로 한 몇몇 연구 성과와 기술들을 대략적으로나마 살펴보면서, 기존에 이루어져 온 국어학 및 국어교육 관련 연구 성과들이 상호 학제적 관련성이 생겼던 것보다도 훨씬 깊다는 것을 확인할 수 있었다.

과 관련한 연구들의 주요 내용들을 비교 검토한 후 상호 참조가 필요한 향후 과제들에 대해서 간단히 언급하고자 한다.

“요약”에 대한 연구는 텍스트언어학에서 1970년대 이후 Kintsch & van Dijk(1978), Brown & Day(1980), Han & Goldman(1983) 등을 중심으로 비교적 활발하게 연구되어 왔고, 국어교육 분야에서도 현재까지 비교적 꾸준하게 진행되어 왔다.

인간과 컴퓨터의 요약 능력의 비교 대조를 위해, 다소 시의성은 떨어지지만 상용화 되어 널리 보급된 바 있는 MS Word 2003 버전에 적용되었던 문서 자동 요약 기능을 활용했다. MS Word는 주지하다시피 세계적으로 일반화 된 프로그램으로서, 국내에서도 아래아 한글 프로그램과 함께 점차 그 사용자가 늘고 있는 추세이다. 흥미로운 것은 MS Word 2003 버전에 포함되어 있던 문서 자동 요약 기능이, 2007년 버전에서는 영어 버전에서만 사용자가 선택 설치하도록 되어 있고, 2010년 버전에서는 영어와 한글 버전 모두에서 제거되었다는 점이다. MS Word의 문서 자동 요약 기능의 원리나 후속 버전에서의 기능 삭제 이유 등에 대해서는 잘 알려져 있지 않으나, 아직은 완전하지 못한 기술 수준의 한계를 인정한 것이 아닌가 생각된다.

MS Word의 문서 자동 요약 기능이 비록 관련 분야에서 최신의 연구 성과들을 오히려 반영하고 있는 것은 아니며, 최근에서는 더욱 다양하고도 효율적인 성과물들이 개발되어 웹 검색 등에서 활발하게 적용되고 있는 것으로 알려져 있다. 더구나 MS Word의 문서 자동 요약 기능의 기술적 과정이나 절차, 원리 등이 공개되어 있는 것도 아니고 그것에 접근하기도 쉽지 않다. 실제로 본 연구와 관련하여 마이크로소프트사의 국제지사 등을 통해 관련 연구의 협조 요청을 수차례 시도했으나 해당 내용들은 대외비로서 공개할 수 없다는 입장이었다. 그런데 관련 분야의 전문가인 강인수 교수의 조언이나 추교남(2007) 등의 연구 성과에 따르면 MS Word의 문서 자동 요약 기능보다도 훨씬 우수한 연구 성과들이 보고되어 오고 있다. 후술하겠지만 추교남(2007)의 자동 요약과 관련한 주요 방법들의 비교에서, MS Word의 방식은 43%의 수행률을 보임으로써 60~90%에 이르

는 다른 방법들<sup>2)</sup>에 비해서 가장 낮게 나타났다.

이러한 상황에도 불구하고 본고에서 MS Word 2003 버전을 중심으로 인간의 요약 능력과 컴퓨터의 문서 자동 요약 기능을 비교하게 된 것은, 연구자 개인의 오래된 관심과 호기심의 결과이기도 하지만, 문서 자동 요약과 관련한 컴퓨터 공학의 성과들이 실제 사용자들을 위해 제공되었다는 점에서 가장 직접적이고도 구체적으로 확인할 수 있는 방법이기 때문이다.

## II. 요약하기 연구와 문서 자동 요약

### 1. 요약하기 연구의 흐름과 제한점

국어교육의 핵심 목표는 국어 이해·표현을 중심으로 하는 문식성(文飾性, literacy), 사고력, 언어문화예술 감상 능력 신장을 핵심으로 한다. 문식성의 가장 기본적인 내용은 요약하기이다.<sup>3)</sup> 국외에서의 요약하기 연구 성과를 요약하면 다음과 같다.<sup>4)</sup>

1980년대를 전후하여 요약규칙과 학습자에 대한 적용 연구가 구체화되기 시작했다. Kintsch & van Dijk(1978)에서는 텍스트의 중요 내용을 거시 규칙 전략을 통하여 요약하는 이론적 근거를 제공하였으며, 요약하기 규칙으로 “삭제, 선택, 일반화, 재구성”을 제시하고 있다. 한편 Brown & Day(1980)에서는 “복사-삭제, 상위어로 일반화(superordination), 중심문장(topic sentence) 선택, 중심문장 생성”을, Han & Goldman(1983)에서는 “복사, 결합,

2) 최근까지 이루어진 주요 자동 요약 방법으로는 “문맥 분석 기반 방법”, “육하원칙 활성화 기반 방법”, “두문(頭文, 신문기사의 lead) 기반 방법”, “제목 기반 방법”, “제목+중요어휘 위치 기반 방법” 등이 있다.

3) 최근에는 멀티미디어 등 매체의 발달과 다양화에 따라, 문자 중심의 읽고 쓰기 능력을 가리키던 문식성의 개념이 소리, 문자, (동)영상 등 복합매체에 힘입어 ‘다중 문식성(multiliteracies)’이 강조되고 있다.

4) 국내외 요약하기 연구와 관련한 주요 성과는 장철순(2008)에 잘 정리되어 있다.

선택, 생성”의 네 가지를 제시하고 있다(서혁, 1991 : 34). 그런데 이들이 제시하고 있는 삭제와 선택은 거의 동시에 이루어지며, 선택에는 삭제의 개념이 부수적으로 뒤따르는 것으로 볼 수도 있다. 따라서 요약규칙은 “중심문장 선택, 상위어로 일반화, 중심문장의 생성(재구성)”의 셋으로 압축될 수 있다.

Baumann(1986)에서는 직접교수법을 활용하여 중심내용(main idea)의 이해 능력을 향상시키기 위한 수업의 단계를 상세하게 제시하고 있다. Brown & Day(1983)에서는 초·중·고등학생과 대학생들의 전략 양상을 비교하여 학년에 따른 요약 능력의 차이를 밝혀내고, 요약하기에 관한 지식과 전략을 가르쳐야 할 필요성을 제시하고 있다. Winograd(1984)에서는 능숙한 독자와 미숙한 독자를 비교하여 글을 요약할 때의 전략적인 어려움을 실험 결과를 통해 보여주고 있다. 한편 James Flood & Diane Lapp(1986)에서는, 서혁(1991)에서도 강조하고 있듯이, 중심내용 파악을 읽기와 쓰기의 관련성 차원에서 접근하고 있다. David Cooper(2006)에서도 텍스트 의미 구성 전략으로서의 요약하기의 개념 및 전략을 소개하고 있다.

국내의 요약하기 연구로는 주로 학습자의 독해 처리와 글의 이해, 요약 능력과 관련하여 이루어져 왔는데, 천한신(1987), 서혁(1991, 1994), 이삼형(1994), 임두순(1989), 김재봉(1996, 1999), 김봉순(1996, 1999), 이수희(2005), 장철순(2008)을 비롯하여 상당수의 연구 결과들이 최근까지 이루어져 오고 있다.

천한신(1987)에서는 요약 능력과 인지 발달과의 상관성, 중요도 파악 능력과 요약 능력의 관계를 다루고 있다. 서혁(1991)에서는 문장과 단락의 중요도 파악 능력이 학습자의 요약 능력과 밀접한 관련이 있다는 점을 학습자의 수준에 따라 구체적으로 분석, 제시하고 있다. 임두순(1989)에서는 독해 전략 훈련의 학습 효과와 관련하여 중심 내용 파악 전략과 정교화 전략을 비교·대조한 후, “수정, 삭제, 첨가, 재구조화”를 중심으로 하는 정교화 전략이 효과적임을 제시하고 있다.

한편 이삼형(1994)에서는 요약이 독자들의 정보처리 능력을 평가할 수 있는 주요 지침이라고 보고, 텍스트 내용 구조 분석을 통해 정보들 간의 구체적인 관계 유형을 제시함으로써 단락의 구분과 중심내용 파악의 이론적 기초를 제공하고 있다. 김봉순(1996)에서는 담화 표지의 상세화를 통한

텍스트의 의미구조를 밝히고자 했으며, 김봉순(1999)에서는 설명적 텍스트의 의미구조를 포괄성, 선호성, 상호성이라는 세 원리에 의해 분석하고 중심내용이 결정되는 원리를 규명하고자 했다. 또한 김재봉(1999)에서는 요약 전략을 분절 구조, 주제 구조, 관계 구조의 활용 전략으로 제시하고 요약의 방법과 기능에 대해 구체화하고 있다. 이수희(2005)에서는 독자의 요약하기 목적과 내용, 형식에 따라 다양한 요약이 가능함을 보여주고 있다.

그런데 요약하기의 결과는 몇 가지 변인에 따라 요약 결과에 차이가 있을 수 있다. 즉, 요약하는 사람의 언어능력과 지식, 목적과 관심의 차이에 따라 달라질 수 있다. 여기에서 ‘지식’이라 함은 어휘, 문법, 텍스트에 대한 지식과 세상사적 지식을 포함한다. 텍스트에 대한 지식은 텍스트의 유형, 특성, 구조, 소통의 맥락<sup>5)</sup>을 포함한다. ‘목적’이라 함은 글을 읽는 독자의 읽기 목적을 말하는 것으로 전체적인 텍스트의 주요 정보나 필자의 메시지 파악, 특징의 세부 정보 파악 등을 가리킨다. ‘관심’이라 함은 글을 읽는 독자의 주요 관심사나 해당 텍스트 내용과 관련한 세부 관심 사항을 가리킨다. 이때의 관심은 때때로 독자의 세상사적 지식과 비례할 수도 있고 반비례 할 수도 있다. 즉, 잘 알고 있는 내용이기 때문에 관심이 높을 수도 있고, 반대로 잘 모르는 내용이기 때문에 더욱 관심을 가지고 읽을 수도 있다는 점이다. 따라서 ‘관심’은 대체로 ‘목적’에 포함될 수 있다.

또한 요약하기는 전체 텍스트 내용을 얼마나 충실하게 반영하는지 여부에 따라 객관적 요약과 주관적 요약으로 나눌 수도 있다. 객관적 요약은 전체 텍스트 내용을 중심으로 객관적으로 충실하게 정보를 선별, 압축, 재생산하는 것을 가리킨다. 이는 기존에 일반적으로 요약하기 교육이나 연구에서 취했던 관점에 해당한다. 반면에 주관적 요약은 독자의 특수한 목적에 따라 의도적으로 특정 보의 취득이나 과제해결을 목적으로 텍스트 정보를 선별, 압축, 재생산하는 것을 가리킨다. 여기에는 독자의 비판적 해석이나 새로운 통찰적 정보가 추가될 수 있다. 객관적 요약과 주관적 요약은 각각 일반 목적의 요약, 특수 목적의 요약이라고 부를 수도 있다.

5) 이때의 맥락은 텍스트 소통의 상황 맥락과 사회문화적 맥락을 모두 포함한다.

일반적으로 요약하기 능력을 평가하고자 할 때는 텍스트 내용을 얼마나 잘 이해하고 정보를 효과적으로 선별, 압축, 재생산하는가를 살펴보고자 하기 때문에 객관적 요약의 관점을 취한다. 초기의 요약 연구나 요약 교육도 대체로 객관적 요약에 더 비중이 놓여져 온 것이 사실이다. 그런데 읽기 교육은 텍스트에 대한 객관적 이해와 함께 궁극적으로는 독자의 주관적이고 비판적인 텍스트의 생산을 지향해야 한다는 점에서 객관적 요약하기(일반 목적의 요약하기)와 주관적 요약하기(특수 목적의 요약하기)가 모두 포함되어야 할 것이다.

그러나 본 연구에서는 객관적인 문서 요약 능력을 중심으로 하는 문서 자동 요약이 논의의 핵심이기 때문에 객관적 요약에 초점을 두고자 한다. 그래야만 요약 능력과 수준의 객관적 비교 대조가 가능하기 때문이기도 하다.

참고로 요약하는 능력의 차이를 Marton과 Säljö(1976)는 4가지 이해 수준으로 다음 <표 1>과 같이 제시한 바 있다.

<표 1> 요약하기의 인지적 수준, Marton과 Säljö(1976)

4수준	필자의 논의점의 요약, 지지되는 증거사항, 결론에 도달하는 개인적 사고 내용을 모두 포함하는 경우 <ul style="list-style-type: none"> <li>• 요약 규칙이 효과적으로 사용됨</li> <li>• 주제문의 작성이 자신의 스키마로 재구성되어 있음</li> </ul>
3수준	결론적인 내용을 요약적으로 제시하기는 하지만, 증거제시의 불충분과 개인적 사고 내용의 불명확성을 보여주는 경우 <ul style="list-style-type: none"> <li>• 주제문에 해당하는 내용이 포함되어 있기는 하지만 그것을 지지하는 세부 정보가 불충분함</li> <li>• 일부 핵심 정보만을 복사, 삭제 규칙에 따라 옮겨 놓음</li> </ul>
2수준	결론을 도출하지 못하고 세부 사항만 상술한 경우 <ul style="list-style-type: none"> <li>• 주제문에 해당되는 내용이 빠짐</li> <li>• 정보의 일부분만을 옮겨 나열해 놓음</li> </ul>
1수준	결론이나 세부사항이 모두 불충분하고 합당치 못한 정보 제시, 개인적 사고의 취약성을 보여주는 부분 <ul style="list-style-type: none"> <li>• 주제문에 해당되는 내용, 세부 사항에 대한 내용 모두 불충분함</li> <li>• 텍스트와 관련 없는 내용을 기술</li> </ul>

전술한 국내외 연구 성과들을 살펴보면 인간의 독해 과정에 초점을 둔 요약하기 연구는 주로 담화·텍스트의 처리 과정, 담화·텍스트의 구조 및 정보의미의 관계, 담화표지, 중심내용 파악과 요약 능력 지도에 초점이 있는 것을 알 수 있다. 특히 요약의 결과물이 단순히 문장의 주요 내용을 간추리는 방식이 아니라 핵심정보들을 바탕으로 (학습)독자의 사고 과정을 거쳐 새롭게 재구성하는 데 초점을 두고 있음을 알 수 있다. 이는 문서 자동 요약이, 자연언어 처리의 어려움과 한계 때문에, 중심 문장의 추출에 초점을 둔 것과 차이가 있음을 말해 준다.

## 2. 문서 자동 요약 연구 현황과 제한점

문서 자동 요약에 대한 연구는 1960년대에 시작된 것으로 알려져 있다(김금영 외, 2002). 국내에서는 1990년대 후반에 21세기 세종계획 프로젝트가 가동되면서 국어 말뭉치(코퍼스) 구축 작업이 본격화되면서 더욱 탄력을 받게 되었다고 할 수 있다. 특히 1990년대 들어 전 세계적으로 컴퓨터 보급이 일반화 되면서 인터넷 등을 통한 효과적인 정보 검색 기술의 필요성에 따라 인간의 언어를 컴퓨터 등으로 처리하는 자연언어처리 연구가 핵심적인 과제로 대두되었다. 여기에는 정보검색, 문서 자동 분류, 문서 자동 요약, 맞춤법 검사기, 기계번역, 음성합성, 음성인식 등이 속한다. 다양한 연구 노력의 결과 최근의 자연언어 처리 기술은 매우 빠른 속도로 발전하고 있는 것으로 보인다.

그런데 컴퓨터가 인간의 의사소통 능력과 유사한 자연언어 처리 능력을 갖추기 위해서는, 음운, 형태, 어휘, 통사, 담화·텍스트, 화용과 관련한 언어 규칙은 물론 세상사적 지식을 포괄하는 알고리즘으로 재구성되어야 하는 새로운 엄청난 작업이기 때문에 지금보다도 훨씬 완벽한 수준의 인공지능이 만들어지기까지는 아직도 상당한 시일이 걸릴 것으로 예상된다. 자연언어 처리가 어려운 가장 큰 요인 중의 하나는 자연언어의 중의성(重義性)에 기인한다는 데 큰 이의는 없는 것으로 보인다.



자연언어 처리 분야에서도 가장 일찍이 연구되기 시작한 분야가 바로 문서 자동 요약이 아닌가 생각된다. 문서 자동 요약은 단일 혹은 다중의 문서에서 중심 내용을 추출하는 정보과학 기술 또는 그 결과물을 가리킨다. 문서 자동 요약은 인터넷 등의 웹 검색 엔진을 통해 사용자에게 효율적이고 편리한 정보 검색 및 처리를 위한 것이다.

김금영 외(2002)에 따르면 자동 문서 요약은 접근 방법에 따라 크게 문장추출 기반 요약 시스템과 정보추출 기반 요약 시스템으로 나눌 수 있다. 문장추출 기반 요약 시스템은 통계적인 방법, 위치, 단서 단어를 이용하여 중심 문장을 선택하여 요약문을 생성하는 것을 말한다. 구체적으로는 통계적 정보를 이용한 요약,<sup>6)</sup> 문장 특성을 이용한 요약,<sup>7)</sup> 의미 구조를 이용한 요약,<sup>8)</sup> 수사적 구조를 이용한 요약으로 나눌 수 있다.

정보 추출기반 문서 요약 시스템은 정해진 개념들을 문서 내에서 추출하여 주어진 템플릿을 채우는 방식의 요약문을 말한다. 이와 유사한 질의기반 자동문서 요약은 주어진 1개 이상의 명사로 이루어진 질의에 대해서 관련 있는 문장 단위를 추출하는 방식을 말한다. 이는 인터넷 검색 사이트에서 찾고자 하는 관련 어휘를 입력하면 웹상에 존재하는 다양한 문서 중에서 관련되는 내용을 요약적으로 제시해 주는 데 유용하다.

초기의 영어 문서 자동 요약 프로그램인 ECON(English Text CONdensing)에서는 각 문장에서 중요한 단어만을 남겨두고 지우는 삭제규칙을 중심으로 이루어진 것으로 알려져 있다.<sup>9)</sup> 그런데 문장을 구성하는 단어들의 취

6) 단어의 출현 빈도와 말뭉치 내의 단어 사용 빈도를 이용하여 문장을 추출하는 방식을 말한다. 클러스터 구조를 이용한 문서 자동 요약 방법이 대표적이라고 할 수 있는데, 이는 텍스트 내에서 단어의 중요도를 고려하여 색인어를 추출한 후, 문장 간의 관련도, 유사도를 계산하여 문장 클러스터를 작성한 후 유사도를 통해 관련성이 가장 높은 문장들을 중심으로 요약문을 산출하는 것이다. 이때 요약문의 분량에 따라 대형 클러스터에서 소형 클러스터로 점차 요약 텍스트에 포함되는 내용이 확장되어 간다.

7) 중심 문장을 추출하여 요약문에 나타낼 수 있는 어휘 특성과 확률 정보를 이용하여 요약문을 생성하는 방식을 말한다.

8) 텍스트 내의 중심문장과 개념들의 연관 관계를 바탕으로 한 요약.

9) 김태희·박혁로·신중호(1999), 검색/요약/필터링을 위한 텍스트 이해 모형 및 처리 기술 개발, 과학기술부, p.12.

사선택은 자칫 너무 일반적인 진술에 그쳐 실질적 의미가 없는 요약문을 구성하거나 원문의 의도와 다른 요약문을 생성할 우려가 있다. 실제로 요약 규칙에서의 삭제규칙은 선택규칙에 부차적으로 적용되는 규칙으로 이해하는 것이 타당하다.

문서 자동 요약 결과물 평가 기준의 핵심은 재현율(Recall Ratio)과 정확률(Precision Ratio)이다. 재현율은 전체 어휘 중 중심 내용으로 압축하는 비율을 가리키고(요약어휘량/ 전체어휘량), 정확률은 요약 결과물의 문법적 정확도를 가리킨다. 그런데 재현율과 정확률이 떨어지는 것을 방지하기 위해 현재로서는 대체로 중요 문장 추출 방식을 선택하고 있는 것으로 보인다. 이는 다음과 같은 진술에서도 확인된다.

“요약문의 질을 높이기 위한 좀 더 적극적인 시도으로써 절 단위의 요약을 하거나, 문장의 절을 삭제하는 식의 요약을 고려할 수 있으나, 정보전달 오류의 가능성을 높이고, 정보의 내용이 약화되는 가능성이 매우 높게 존재한다. 물론, 인공지능의 정교하고 광범위한 지식을 동원한다면 어느 정도 성과가 있을 수 있지만 현재로서는 그러한 연구는 제대로 시도된 바가 없다. 결국 현재 시점에서 상용화를 위한 요약은 가장 적절한 문장을 선택하는 문제라고 볼 수 있다.” (김태희 외, 1999 : 67)

아울러 김태희 외(1999 : 67~68)에서는 복합 방법에 의한 문서 자동 요약 방법이 그나마 가장 효과적이라고 제시하고 있다. 즉, 단어 출현 빈도를 중심으로 한 통계적 방법과 텍스트 구조 지식 등을 중심으로 하는 지식기반적 방법의 총체적 시각(Wholism)이 필요함을 언급하고 있다.

추교남(2007)에서는 하위범주화(의미)사전을 통한 ‘문맥 분석 기반 방법’이 기존에 이루어진 다양한 자동 요약 방법<sup>10)</sup> 들 중 가장 효과적임을 제

10) 자동요약의 주요 방법은 연구자에 따라 다양하게 명명되고 있기도 하다. 통계적 빈도수, 어휘의미장, 제목기반, 두문기반, 구조중심, 의미기반 개념망, 온톨로지 등. 여기에서 개념망이란 ‘상하위, 동의, 유의, 반의, 포함(전체-부분), 오용, 관련어’ 관계를 통해 상호 의미 관계를 구축하는 것이다. 개념망 구축은 국어사전 혹은 전문용어사전에 등재된 표제어에 대해서 개념어 간의 의미적인 관계를 연결시키는 것으로서, 아직까지는 국어에 대한 전문 지식을 가진 사람에 의해 하나하나 수작업으로 구축되는 것이 일

시하고 있다.<sup>11)</sup>

〈표 2〉 주요 자동 요약 방법의 정확도 비교(추교남, 2007 : 109)

구분	문맥 분석 기반 방법	육하원칙 활성화 기반 방법	두문 기반 방법	제목 기반 방법	제목+중요 어휘 위치 기반 방법	MS-Word
정확도	91%	74%	70%	68%	55%	43%

문맥 분석 기반 방법이 일종의 개념망을 통한 의미적 접근이 강한 반면에, 육하원칙 활성화 기반 방법은 수사구조를, 나머지 ‘두문, 제목, 제목 + 주요 어휘 위치’ 기반의 방법들은 통계적 빈도수에 더 의존하고 있는 것으로 판단된다. 물론 컴퓨터로 이루어지는 자동 요약 기능의 속성상 모든 방법에서 통계적 연산 과정은 기본적으로 활용되는 방법이라고 봐야 할 것이다.

자동 요약의 정확도가 대체로 70~90%까지 이르고 있다는 것은 상당히 놀라운 사실이다. 물론 아직은 실험적 수준인 경우가 많고, 완성되어 상용화되기까지는 자연언어와 세상사 지식의 개념망을 확충하는 등의 온톨로지<sup>12)</sup> 개발 구축 노력이 뒤따라야 할 것이다.

반적이다.

- 11) 기존의 다양한 자동 요약 분석 방법과 원리에 대한 소개는 강인수(2010)를 참고할 수 있다. 2010년 8월 국어교육학회 학술발표 대회에서 강인수 교수의 발표는 문서 자동 요약과 관련한 많은 새로운 정보와 시사점을 제공해 준 바 있으며, 본 연구에도 많은 도움이 되었다. 발표자께 감사드린다.
- 12) 전술한 개념망을 더욱 체계적으로 컴퓨터 데이터 베이스화 하고자 하는 노력이 온톨로지의 구축이라 할 수 있다. 즉, 온톨로지는 공유된 개념화(shared conceptualization)에 대한 정형화되고 명시적인 명세서(formal and explicit specification)로서, 단어와 관계들로 구성된 일종의 사전으로서 생각할 수 있다. 여기에는 특정 영역(분야)과 관련된 단어들이 계층적으로 표현되어 있고, 추가적으로 이를 확장할 수 있는 추론 규칙이 포함되어 있어, 웹 기반의 지식 처리나 응용 프로그램 사이의 지식 공유, 재사용 등이 가능토록 되어 있다. 온톨로지는 시맨틱 웹 응용의 가장 중심적 개념으로서, 이를 표현하기 위해 스키마와 구문 구조 등을 정의한 언어가 온톨로지 언어(ontology language)이다 (네이버 용어 사전 참조). 온톨로지의 구축은 컴퓨터를 통해 정보를 체계적이고 효율적으로 검색하고 재구성할 수 있도록 하기 위한 것으로 인공지능 완성의 기본 조건이 된다고 할 수 있다.

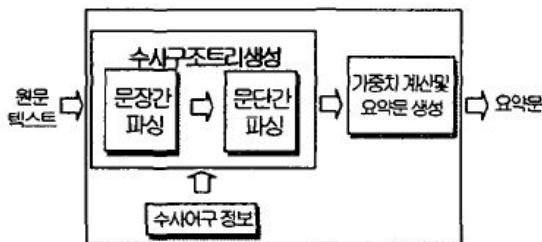
여기에서 MS-Word의 방법은 2003 버전을 가리키는 것으로 보이는데 정확률이 가장 뒤떨어지는 것으로 나타난다. 김계성 외(2003)의 ‘단락 자동 구분을 이용한 문서 요약 시스템’ 연구에서도 재현율과 정확률이 각각 .64와 .49로서, MS Word의 .33, .39에 비해 훨씬 뛰어남을 보여주고 있다. 2007버전부터 해당 기능을 제거하기 시작하는 이유를 확인할 수 있다. 본고에서 MS Word의 자동 요약과 학습자의 요약 수행 결과를 비교해 본 실험 결과는 별도의 절에서 후술된다. 또한 본고에서 관심을 가지고 있는 수사구조를 이용한 자동 요약 방법과 관련하여 김태희 외(1999 : 68)에서는 일찍이 다음과 같이 기술한 바 있다.

“수사어구 데이터베이스와 서술어의 수사관계 패턴 데이터베이스를 강화하여 수사어구 추출단계의 기능보강 및 서술어구 추출단계의 기능을 보강하여 수사어구의 위치에 관계없이 입력문장의 수사관계를 규명할 수 있도록 향후 연구가 계속되어야 할 것이다.”

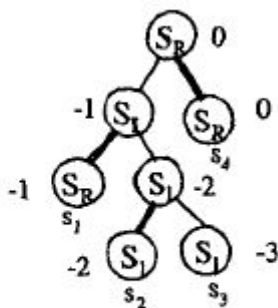
그런데 이 지적에서 “수사어구의 위치에 관계없이”라는 진술은 본 연구자의 입장에서는 재고의 여지가 있다고 판단된다. 이 연구물이 1999년도의 것으로 상당한 기간이 흘렀다는 점을 감안해야 하겠으나, 수사관계에서 수사어구의 위치는 문맥에 따라 차이는 있겠으나 요약과 관련하여 중요한 의미를 갖는 정보에 해당한다. 다시 말해서 텍스트는 특정 맥락에서 발신자가 자신의 전달하고자 하는 정보를 글의 전개 과정에 따라 제시하는 것으로 비유하자면 건축물의 평면적 전개도와 같다. 일반적으로 문자 텍스트의 경우 입체적 사고를 문자와 종이라는 평면에 구조화하여 제시하는 방식이다. 이때 비록 평면적 문자의 나열이라고 할지라도 텍스트 자체는 일정한 방향글의 전개 방향과 운동량(정보의 중요도)을 갖는 벡터(Vector)에 해당한다고 볼 수 있다. 이런 맥락에서 문서 자동 요약은 기본적으로 텍스트의 벡터 값을 구하는 것에서 출발한다고 볼 수 있다. 따라서 수사어구를 비롯한 텍스트 요소들의 위치 값은 인간의 텍스트 처리 과정에서는 물론 컴퓨터를 활용한 문서 자동 요약에서도 중요한 의미를 갖는다.

실제로 김건오 외(2002)의 “어휘 클러스터링을 이용한 자동 문서 요약”을 비롯하여 최근의 자동 요약 관련 연구물에서는 텍스트 내용어휘 분석 시 벡터로 변환시켜 컴퓨터 저장 공간에 벡터화의 개념을 도입하여 접근하고 있다. 다만 이때의 벡터의 개념이 정확히 텍스트 전개의 흐름을 반영하고 있지는 못하다는 생각이 든다. 텍스트 전개의 논리적 흐름보다는 기본적으로 중요도 판정을 위한 가중치 계산에 초점을 두고 있는 것으로 보이기 때문이다.

이유리·최기선(1999)에서는 수사구조를 이용한 텍스트 자동 요약을 시도하고 있다. 위 논문에서는 텍스트 내 단어들의 의미 연결구조를 이용하는 요약 시스템의 경우 WordNet과 같은 시소러스에 저장된 의미 관계 정보만을 이용하므로 특정 텍스트에서의 단어들의 관계를 파악하는 데는 한계가 있다는 문제의식에서 출발한다. 이를 보완해 줄 수 있는 단서로 텍스트의 수사 구조에 주목하는 것이다. 이때의 수사 구조는 대체로 담화 표지를 활용한 문장 간의 연결 관계에 초점을 두고 제한적 실험을 시도하고 있다. 해당 논문에서는 담화표지 대신에 ‘수사어구’라는 표현을 사용하고 있으며, “그러니까, 그래서, 왜냐하면, 즉, 뿐만 아니라, 따라서, 예를 들어, 하지만, 따라서” 등의 접속어구, 지시어, 숙어표현에 제한하여 분석하고 있다. 이 경우 한 단락 정도의 짧은 텍스트에서는 효과적으로 작동할 수 있으나 여러 단락으로 이루어진 긴 텍스트의 경우에는 적용이 쉽지 않다는 단점이 있다.



〈그림 5〉 수사구조를 활용한 요약시스템의 구성  
(이유리·최기선, 1999 : 98)



〈그림 6〉가중치가 부가된 수사구조 트리  
(이유리 · 최기선, 1999 : 101)

수사구조를 활용한 문서 자동 요약은 기존에 어휘소 반복을 중심으로 하는 빈도수 조사 방법에 비해 진일보한 것이라 할 수 있으나, 텍스트의 전반적 구조 분석의 방법으로는 아직 미흡하다고 판단된다. 즉, 텍스트의 거시구조 파악 노력이 좀 더 필요하고, 어휘소의 반복성과 의미관계, 상하위어 등을 포괄하는 통합적 온톨로지 체계의 도입이 요구된다.

아울러 텍스트의 수사구조와 수사어구의 정밀화 작업이 필수적으로 체계화 되어야 할 것이다. 또한 수사구조 역시 문장 간, 문단 간 수사 구조의 개념에 더하여 텍스트 전체의 수사구조로 확장되어야 할 것이다. 이와 관련하여서는 Kintsch & van Dijk(1978), 서혁(1996), 김봉순(1996), 이해운(2007) 등에서 논의된 바 있는 텍스트 구조 분석, 거시구조, 초구조(Superstructure) 등의 개념을 참조할 필요가 있다. 특히 이해운(2007)에서는 담화표지에 따른 텍스트 구조분석을 통해, 동화/ 신문기사/ 학술논문이 각각 “연쇄(sequene)/연접, 상세화/연접, 상세화, 수단”으로 상이한 텍스트 구조를 갖는다고 보고, 궁극적으로 텍스트에 따른 초구조(Superstructure)의 차이를 밝히고자 시도하고 있다. 이 역시 수사구조를 활용한 문서 자동 요약에 시사하는 바가 크며, 텍스트 연구자들의 과제가 되리라고 본다.

서혁(1996)에서는 담화의 구조와 관련한 논의에서 수사적 구조를 텍스트 전체에 흐르는 특징적인 표현 장치와 전체적인 구조에 가까운 개념으로 보고자 했다. 예컨대 특정 정보나 주장을 나열하는 경우, “……는 다음

과 같이 세 가지 측면에서 살펴볼 수 있다. 첫째, ……(또는 먼저, 우선, 무엇보다도 등), 둘째, ……, 셋째, ……”와 같은 특징적 구조를 갖는 경우가 많다. 이때 ‘첫째’, ‘둘째’, ‘셋째’ 등의 서수사는 예컨대 “우리 첫째 아이는……”에서의 ‘첫째’와 혼란의 여지가 있으나, 담화표지로서 작동할 때의 “첫째, 둘째, 셋째”는 바로 다음에 반점(.)으로 이어지고, 또한 대체로 담화상의 등간격을 유지하려는 속성을 지닌다는 점에서 비교적 분명하게 확인될 수 있다. 이는 주술호응의 규칙에 비교하여 텍스트의 수사적 호응 구조의 하나로 볼 수 있다.

수사구조를 활용한 문서 자동 요약 연구에서는 주로 문장 간 또는 문단 간 수사 구조를 중심으로 접근하고 있는데, 담화·텍스트 전체의 수사(호응)구조나 초구조(Superstructure)에 대한 관심과 연구 도입이 필요하다고 판단된다. 물론 이는 국어 텍스트 연구자들의 과제가 될 것이다. 그밖에 문서 자동 요약과 관련한 몇 가지 문제점과 과제를 지면 관계상 간략히 제시하면 다음과 같다.

첫째, 문서 자동 요약과 관련한 종합적 접근법의 문제로서 온톨로지의 구축 작업과 관련된다. 즉, 온톨로지의 구축작업과 함께 수사구조, 통계적 빈도수, 제목 기반 등의 방법이 종합적으로 활용되는 방안이 필요하다고 본다. 그 기술적 가능성에 대해서는 문외한인 필자로서는 알 수 없으나, 현재의 연구와 기술 발달 속도로 그리 멀지 않은 기간 내에 충분히 가능하리라고 판단된다. 특히 이는 단순히 컴퓨터 공학자나 언어학자들에 의해 온톨로지의 기계적 구축을 기다릴 문제가 아니라, 인간인 학습자들의 문서 요약에서 보여주는 다양한 특징과 동원되는 능력, 지식, 전략 등의 문제를 정밀하게 기술하고 분석함으로써 훨씬 더 효과적이고도 체계적으로 접근될 수 있고, 연구 기간도 단축시킬 수 있을 것이다.

둘째, 맥락적 정보에 대한 고려가 부족하다는 점이다. 즉, 기존의 자동 요약 방법들에서 문맥 기반의 방법들은 많이 연구되고 있으나, 발신자와 수신자의 상황맥락, 사회문화적 맥락 정보에 기반한 ‘맥락적 정보 기반’의 방법이 시도되지 않는 것이 좀 의아하게 생각된다. 이는 물론 필자가 과묵한 탓일 수 있으나, 아직까지는 주어진 텍스트 내의 문맥 기반에 제한

된 방법이 주를 이루고 있는 듯하다. 물론 종합적인 맥락 기반의 방법 역시 보다 완전한 온톨로지의 구축과 밀접한 관련이 있을 것으로 판단된다. 맥락적 정보 속에는 소통의 상황맥락과 관련되는 발신자의 의도, 목적, 태도(양태 등) 등은 물론, 관련 사회문화적 지식과 가치관 등의 사회문화적 맥락도 포함될 수 있다.

셋째, 요약 시 텍스트의 생산(표현)적 측면과 관련되는 기술의 미흡이다. 즉, 현재의 문서 자동 요약은 주로 텍스트 정보의 중요도 판단에 정보의 선별에 머물러 있고, 이마저 완전한 수준에 이르지 못하고 있다. 학습자에게 요구되는 요약 능력과 같은 수준이 되기 위해서는 컴퓨터의 자연 언어 의미 처리 능력과 문법적 표현이 능력이 앞으로 훨씬 더 보강되어야 하는 것은 물론, 인간의 요약 능력에 대한 보다 정밀한 연구가 뒤따라야 할 것이다.

### III. 학습자와 컴퓨터의 요약 능력 비교 —MS Word 2003을 중심으로<sup>13)</sup>

본 연구에서는 현행 교육과정에서 학습을 진행하고 있는 초등학교 6학년 한 학급의 어린이들과 MS word 2003프로그램에 <세대 문화의 갈등 (부록 1)>이라는 글을 요약하게 하였다.<sup>14)</sup>

13) 이 글에서 MS Word 2003의 자동 요약 실험 비교 및 기타 자료 정리 부분은 필자가 담당했던 2010-1학기 이화여대 대학원 석·박사 과정 수업에서 오은정, 곽진영, 김규연, 김선경, 고경은, 김보람, 문영은, 이소라, 박윤진 선생이 본 연구에 많은 도움을 주었다. 특히 오은정 선생은 학습자 실험 등을 중심으로 주도적으로 수고해 주었다. 크게 감사드린다.

14) 학습자들은 서울시 동작구에 소재한 S초등학교 6학년 한 학급 어린이 35명으로 구성되었다. 35명 중에 19명이 남학생, 16명이 여학생이며 이들은 2010년 4월, 해당 학교 교육과정과 평가 지침에 의해 읽기 분야 성취도 평가(지필고사, 25문항 100점 만점)를 보았다. 성취도 평가 결과 읽기 수준의 ‘상(上)’에 해당하는 아동이 24명으로 69%, ‘중(中)’에 해당하는 아동은 9명으로 26%, ‘하(下)’에 해당하는 아동은 2명으로 5%로 나



MS Word 프로그램 2003에서는 25%, 50%, 75%의 선택사항을 차례로 실행한 것이다. 이 자동 요약 기능은 리포트, 정부 보고서, 기술 문서, 논문 등과 같은 구조화된 문서에서 아주 잘 작동하며, 비교적 구조화되지 않는 문서에서는 잘 작동하지 않는다고 소개되고 있다(MS Word 2003 매뉴얼).<sup>15)</sup> 자동 요약 기능은 문서에서 문장을 분석하고 점수를 매겨 주요 문장을 결정하는 방식으로, 만약 문서에서 자주 사용되는 단어를 포함하는 문장은 높은 점수를 얻게 되는 형식이다.

## 1. 학습자의 요약 능력 평가

본 연구에서는 서혁(1994)과 7차 교육과정에서 제시하는 원리를 조합하여 아래 <표 3>의 요약규준을 구성하여 학습자와 워드 프로그램의 요약문을 평정하였다. 평가는 현장 교사를 포함하여 대학원 석·박사 과정에 재학 중인 3명의 전문가들에 의해 실시되었다.

---

타났다. 평가 점수를 읽기 수준으로 가름하는 기준은 성취도 평가의 평균이 89.20으로 평균보다 높은 점수를 얻은 아동과 평균보다 20점의 차이로 낮은 아동, 20점 이상의 차이로 낮은 아동을 나눈 것이다. 참고로, 연구대상들이 소속된 학교는 사회경제적 지위가 중상급에 해당하는 전형적인 서울 아파트촌으로 학습자들의 수준이 문식성을 비롯한 여러 가지 면에서 높게 나타나는 학교이다.

- 15) MS 워드 2003에서는 문서 자동 요약 방법을 총 일곱 가지로 소개하고 있다. 다음 진술은 그 중 하나이다. “요약 내용을 작성한 후에는 문서의 주요 내용이 포함되어 있는지 확인하십시오. 내용 문장은 제대로 다듬어진 깔끔한 문장이 아니므로 문장을 잘 다듬어서 사용하십시오.” 이 진술을 통해 자동요약 기능은 처음부터 ‘적절한 요약’을 해낼 수 없다는 사실을 전제하고 있다는 것을 알 수 있다. 자동요약을 시행한 후 주요 내용이 포함되어 있는지 확인해야 하고, 문장을 다듬어야 하는데, 이러한 결과물을 요약하기로 볼 수는 없다. 중심 내용 파악이라는 요약하기의 가장 기본 요소를 자동요약은 적절하게 수행하고 있지 못하기 때문이다. 그리고 자동요약은 문서에 텍스트 상자, 틀, 표 등 기호, 도형 등은 자동요약하지 않는다는 것을 밝히고 있다. 자동요약이 텍스트의 의미가 도식적으로 표현되는 것을 아직 인지하지 못한다는 것을 의미할 것이다.

〈표 3〉 요약하기 평정 기준

평정 기준	내용적 요소			형식적 요소	
	맥락성(20)	완결성(20)	독창성(20)	조건 준수(20)	세부 사항(20)
평정 질문	중심 내용과 세부 내용을 정확하게 구분하고, 글의 흐름에 따라 왜곡하지 않고 논지를 기술했는가?	요약문이 하나의 완결된 글의 형태로 통일성(coherence)을 갖추고 있는가?	이해한 내용을 자신의 표현으로 재구성하였는가?	제시된 조건에 따라 글을 작성하였는가?	맞춤법, 단락 나누기 등 국어지식의 사용이 올바르게 반영되었는가?

연구대상 35명의 학습자에게 <부록>의 ‘세대 문화의 갈등’<sup>16)</sup>이라는 글을 읽게 하고 800자 내외로 요약하게 한 결과를 위 <표 3>의 기준에 의해 평정하였다.

〈표 4〉 학습자들의 요약하기 평정 결과

연번	이동명	요약하기 평정 기준					총점
		맥락성	완결성	독창성	조건 준수	세부 사항	
평균		14.71	15.86	16.14	17.31	16.00	80.02

평정 결과는 대체로 우수한 편으로 드러났으며 점수 합산과 관련하여 형식적 요소 중 ‘조건 준수’ 항목이 가장 높은 점수 획득을 나타내었다. 이는 800자 내외라는 제한된 조건에 600~800자, 800~1000자를 작성한 학생에게 2점의 감점을 주었기 때문에(다른 기준의 점수 폭은 5점) 생긴 현상일 수도 있지만 대체로 명시적이고 기계적인 부분에 대해서는 대체로 어려움 없이 과제를 수행하기 때문으로 보인다. 한편, 맥락성과 관련된 부분이 가장 낮게 나왔는데, 이를 통해 글의 중심 내용과 세부 내용을 구분하여 텍스트 문맥의 흐름에 따라서 핵심 내용을 도출해 내는 과제를 학생들이 상대적으로 어려워한다는 것을 알 수 있다. 전문가들의 평가 결과에

16) 초등학교 6학년 교사용지도서에 수록된 글로서, 전체 내용은 뒤의 부록을 참조할 수 있다.

따라 만점(100점)을 받은 학생의 요약문은 다음과 같다.

요즘처럼 사회가 빠른 속도로 변화하는 시대에는 자연히 세대 차이가 커지고 그에 따른 문화의 갈등도 한층 더 심해지고 있다. 세대차이로 생기는 문화적 갈등은 어린이의 옷차림과 언어 등을 통해 드러난다. 이런 세대문화의 갈등이 생기는 까닭은 무엇일까? 첫째, 대화의 단절이다. 각자 자기 일에 시간을 투자하기 바쁜 현대사회에서 대화하는 시간이 무척 줄어들고 있다. 둘째, 서로 상대방의 입장을 생각하지 않는 이기주의도 하나의 요인이 된다. 그 외의 요인 중에는 대중매체의 영향도 크다고 할 수 있다.

이러한 세대문화의 갈등을 해결하기 위한 방안에는 세 가지가 있다. 첫째, 서로 이해의 폭을 넓히기 위해 대화를 자주 나누어야 한다. 둘째, 어린이와 어른이 서로의 장점을 받아들여 조화시키는 노력을 해야 한다. 셋째, 어린이와 어른이 함께 도덕적인 사회를 만들기 위해 마음을 모아야 한다. 그리하여 진실과 정직이 바탕을 이루는 사회를 함께 만들어 나갈 수 있다면 세대문화는 조화롭게 발전할 것이다.(34번 최\*\*)

위 요약문을 쓴 학생의 경우 원문 텍스트에서 구조적 읽기의 흔적을 찾아 볼 수 있다. 이를테면, ‘빠른 속도로 변화’, ‘옷차림’, ‘언어’, ‘까닭’, ‘어떻게’ 등의 단락을 이끌만한 단어에는 □표시를 하였다. 또한 ‘하나’, ‘다음’, ‘첫째’, ‘둘째’ 등속의 서수 구실을 하는 단어에는 ①, ② 등의 원문자 표시를 해둔 것을 볼 수 있었다. 텍스트를 읽는 동안 이 학생의 머리에서 중심 내용과 세부 내용이 효과적으로 판단되고, 예시와 근거의 적절한 구분과 변별이 이뤄지고 있다는 것을 알 수 있다. 그리고 추출된 문장과 문장 사이에 적절한 어말어미나 접속사 등을 사용하여 요약문 자체의 흐름이 원문 텍스트의 그것처럼 자연스럽게 이어지고 있는 것을 볼 수 있었다.

## 2. 컴퓨터의 자동 요약 능력 평가

본 연구에서 적용한 컴퓨터에 의한 문서 자동 요약은 MS Word 2003

버전을 윈도우 비스타 운영체제를 사용하는 Sense R70 컴퓨터에서 구동시킨 것이다. 요약 결과는 <부록>을 참고할 수 있다. 이 <부록>에는 대상 텍스트 원문과 13%, 25%의 요약 내용을 함께 제시한 것이다. MS Word 2003의 요약 결과를 앞에 제시된 요약 평정 기준에 따라 평가한 결과는 다음 <표 5>와 같다. 평정 방법은 앞서 학습자 요약문 평정과 같다.

〈표 5〉 MS Word 2003의 요약문 평정 결과

요약 비율	요약하기 평정 기준					총점
	맥락성 (20)	완결성 (20)	독창성 (20)	조건 준수 (20)	세부 사항 (20)	
13 %	5	15	5	20	15	60
25 %	5	15	5	20	10	55
27 %	5	15	5	20	10	55
75 %	5	20	5	20	15	65
평균	5.00	16.25	5.00	20.00	12.50	58.75

위 <표 5>에서와 같이 학습자들의 총점과 비교했을 때 20점 이상의 점수 차이를 확인할 수 있다. 워드 프로그램의 경우, 선택상황에서 결정된 조건은 착오 없이 진행되기 때문에 조건 준수 면에서 만점인 20점을 받을 수 있었다. 그리고 원문 텍스트와 독립적으로 글을 읽어 보았을 때 인상적으로 완결되어 보이는 경향이 있기도 하였다. 하지만 제시된 원문 텍스트와 요약문의 비교대조를 통해서도 알 수 있듯이, 빈도수가 높은 어휘소가 포함된 문장을 중심으로 단순히 특정 문장들을 선택 나열하고 있음을 볼 수 있다. 따라서 통일성(coherence)이 현저하게 떨어진다. 또한 원문 텍스트에서 한 글자도 틀리지 않게 옮겨 놓았기 때문에 원문 텍스트와 깊은 관련을 맺으면서도 또 하나의 독창적인 텍스트를 생성하는 데에는 실패했다고 볼 수 있다. 요컨대 MS Word 2003의 자동요약 기능은 다음과 같은 특징과 문제점을 보여준다.

첫째, 단어 빈도수가 높은 것을 핵심 단어로 인식한다는 점이다. 즉, 반복되는 단어를 핵심 단어로 인식한다. 이 때 핵심 단어는 하나만 선정

되는 것이 아니라 문단에서 여러 번 반복되는 단어는 모두 핵심 단어로 인식하는 것으로 보인다. 실제로 형태소 분석기 ‘말샘41’을 통해서 실험 텍스트를 분석한 결과 단어빈도수 결과를 반영하는 것으로 확인되었다.

둘째, 문장에 대한 정보중요도의 판단, 즉 중심 문장의 선택 방법은 핵심 단어가 많이 들어가는 문장을 중심으로 한다. 또한 반복되는 단어(핵심단어)가 복수일 경우, 이 단어들이 모두 들어간 문장을 핵심 문장으로 선정하는 경향이 강하다. 이는 단어 빈도수에 의한 가중치에 따라 중심문장을 추출하는 것을 의미한다.

셋째, 선행문장 우선 선택의 원리이다. 즉, 동일 또는 유사한 정보 중 요도를 지니는 문장들의 경우 선행 문장을 우선적으로 선택한다. 핵심 단어가 모두 들어간 문장이 여러 개라면 문단에서 가장 앞에 있는 문장을 핵심 문장으로 선정하는 것을 보여준다. 두드러지게 반복되는 단어가 없는 경우에는 두괄식 구성을 기준으로 가장 앞 문장을 핵심 문장으로 선정하는 것으로 보인다. 별도의 실험문을 통해서 확인한 결과에 의하면, 동일한 위계의 문장 구조가 병렬적으로 나열될 때에도 최선행 문장을 요약문으로 반영했다.<sup>17)</sup>

넷째, 상위어 판단이 불가하다는 점이다. 전술한 별도의 실험문 분석 결과, 상위어로 대체 능력이 없음을 보여 주었다. 이는 MS Word 2003의 자동 요약 기능이 단어 빈도수 중심의 통계적 방법에 따른 문장 추출 방식을 중심으로 제작되었고, 문맥정보 기반의 개념망은 활용되지 않는 것으로 판단된다.

다섯째, 텍스트 구조 파악 능력을 갖추고 있지 못한 것으로 보인다. 별도의 실험문 분석 결과에 의하면 두괄식, 미괄식, 중괄식 역시 요약에 영향을 주지 못하는 것으로 판단된다.

이상의 실험 결과를 정리하자면, MS Word 2003의 자동 요약 기능은

17) 이는 박진영 선생의 실험을 통해서도 확인되었다. 즉, “나는 ( )을 좋아한다.”는 동일 구조 문장으로 ( )안에 다양한 과일의 종류를 나열한 병렬 텍스트 구조 실험에서 첫째 문장을 중심문장으로 선택하였다. 이는 선행문장 우선 선택의 원리와 함께, 후술되는 바와 같이, 상위어 인식이 불가함을 보여준다.

기계적인 조건의 준수, 세부사항의 이행에 있어서는 인간 학습자보다 오류가 적을 수 있겠으나 글의 맥락성을 살려 독창적으로 요약문을 생성하는 데에는 효과적이지 않다는 것을 알 수 있다. 우수한 학습자와 워드 프로그램의 요약 능력을 비교해 보았을 때 워드프로그램의 요약기능이 제 역할을 하기 위해서는 몇 가지의 추가적인 기능과 활동이 필요해 보인다.<sup>18)</sup>

첫째, 자동 요약 기능을 비율이나 문서 안에서의 배치 정도의 한정된 선택 상황 속에서 이뤄지게 하지 말고 문서작성자의 인지전략과 궤를 같이 하는 안내-선택의 옵션을 두는 것이 좋을 듯하다. 즉, 요약문을 덩어리째 제시하는 현 방식보다는 문장단위나 단락단위의 표식에서 ‘이 문장을 살릴까요?’ 등의 안내 팝업창이 뜨는 방식도 가능할 것이다. 인공지능적 요소가 보다 인간의 두뇌활동과 비슷해지기 전에는 문서 작성자의 인지전략을 적절히 안내하면서 이뤄내는 요약기능이 더 유효할 듯하다.

둘째, ‘우선’, ‘다음으로는’ 등의 개조식 서술 문두를 파악하는 것이 필요하다. 위에서 예시한 워드프로그램의 요약문에서는 ‘첫째’, ‘둘째’ 등으로 시작하는 두괄식 어구에서조차 핵심문장이 변별되지 못하고 있다. 따라서 원문 텍스트의 논리 구조를 이끄는 상용구 등이 프로그램 내에서 해독이 되도록 조치하는 것이 좋겠다.

셋째, 우수한 학습자들이 자신만의 시각적 이미지를 원문 텍스트에 표시한 후 요약한 것을 참고하여 문서 작성자가 핵심 단어나 핵심 어구를 음영 등의 기법으로 표시하게 한 후 요약기능을 실행하면 좋을 것이다. 작성자가 표시한 부분이 추가 선택사항으로 요약 기능에 추가된다면 핵심 문장이 사장되는 오류를 피할 수 있을 것이기 때문이다.

워드 프로그램의 요약문 평정 결과는 초등학생 하위 학습자의 점수와 비슷한 수준으로 낮게 나왔다. 워드 프로그램의 요약 점수 낮은 주요 원인은 맥락성과 독창성, 문장 조어 능력 등의 인지-초인지 전략이 인간

18) 이는 오은정, 김규연, 김선경 선생의 논의와 제안이기도 하다. 검색 결과 확인 질문 방법 등은 현재 인터넷 검색에서 활용되고 있기도 하나, 전체적으로 시사해 주는 바가 많은 참신한 아이디어라고 판단된다. 앞으로 기술 보완 및 개발이 과제이다.

두뇌에서는 일정한 훈련 후에 자연스럽게 재현되는 데 비하여 워드 프로그램에서는 단순한 선택 상황에서 제한된 전략 사용만이 이뤄지고 있기 때문에 보인다.

#### IV. 학제적 측면에서의 시사점

컴퓨터를 통한 자연언어 처리, 특히 문서 자동 요약 연구의 성과들은 국어교육에 시사하는 바도 많다. 즉, 국어교육에서는 학습자들의 정보 검색 능력과 활용 능력, 정보의 이해와 판단, 활용과 표현 능력 신장 등 디지털 리터러시 교육에 시사점을 얻을 수 있다. 실제로 학습자들이 자신이 원하는 정보를 효과적으로 검색하고 활용하는 정보 검색 및 활용 능력은 기본적인 교육 내용의 한 부분으로 고려될 필요가 있다. 지금까지는 인터넷 상에서 필요한 정보 검색은 단순히 한두 단어의 키워드를 입력하거나 주어진 틀에 질문을 입력하면 되는 정도로 매우 가볍게 처리되어 오지는 않았는지 반성이 필요하다. 무작정 상세화하여 다양한 관련 어휘를 입력하는 것도 정보 검색에 실패할 가능성이 높다. 따라서 효율적인 정보 검색을 위해서는 자신이 원하는 정보를 가장 효과적으로 검색하고 판단하기 위해 상하위어 활용, 관련어휘와 지식의 활용, 상세화 정도에 대해 상당한 인지적 노력이 필요하다. 더 나아가서는 문서 자동 요약의 과정과 절차, 원리를 이해하게 되면 정확하고 효과적인 언어 표현 및 이해 능력의 필요성과 중요성과 그 가치의 내면화 교육과도 밀접한 관련을 갖는다.

지금까지 논의한 결과를 바탕으로 학제적 관점에서 문서 자동 요약이 주는 시사점을 간단히 제시하면 다음과 같다.

첫째, 맥락정보 기반 접근법 활용의 필요성과 가능성이다. 제목이나 문맥 단서뿐만 아니라 담화·텍스트의 상황맥락, 사회문화적 맥락 등의 맥락 정보 활용 가능성을 고려해야 할 것이다. 특히 발신자, 수신자, 더 나아가서는 의도, 목적, 메시지, 관련지식(스키마)을 종합적으로 고려할 수

있는 방안이 온톨로지의 구축과 함께 고려되어야 할 것이다.

둘째, 담화 표지와 텍스트의 구조를 좀 더 적극적으로 활용하는 방안이 모색되어야 할 것이다. 기존의 담화 표지에 대한 연구 성과의 반영도 아직은 미흡한 수준으로 보이며, 특히 텍스트의 구조(초구조, 거시구조, 담화·텍스트 차원의 수사적 호응 관계) 등은 아직도 거의 반영되지 못하고 있는 것으로 보인다. 특히 텍스트 전개 구조를 참조한 입체적 벡터의 개념(텍스트 전개 방향, 텍스트 부분의 중요도)이나 등간격 지향의 원리 등(첫째, 둘째, 셋째 등)은 구체적으로 공동 연구가 필요한 부분으로 생각된다.<sup>19)</sup>

문서 자동 요약 연구는 주로 기사문이나 설명적 텍스트 등 정보 전달적 텍스트를 중심으로 연구되어 왔으며, 주로 정보의 검색을 목적으로 이루어져 왔다. 이는 컴퓨터에 의한 정보의 창출이나 생산, 특히 비유적 표현의 이해 및 처리와 관련하여서는 아직도 어려운 과제임을 말해 준다. 그러나 스스로 그림을 그리는 컴퓨터 프로그램이 나오고 있듯이, 언젠가 스스로 시와 소설을 짓는 프로그램이 일반화 되는 날도 멀지 않을 것이다.

중요한 것은 이러한 컴퓨터 공학의 발달의 기본 바탕에는 인간의 정보처리 과정에 대한 정밀한 이해가 핵심적인 역할을 하게 된다. 즉, 인간의 언어 학습과 이해, 표현의 인지적·정의적 사고 과정에 대한 규명이 핵심이라는 점이다. 컴퓨터의 정보 처리 프로그램에는 학습과정이라고 불리는 전처리 과정이 필수적인데, 컴퓨터 프로그래머나 전문가뿐만 아니라 컴퓨터 역시 하나의 학습자라고 볼 수 있다. 따라서 국어교육에서 학습자의 언어처리 과정에 대한 과학적 탐구는 기여할 수 있는 바가 크다. 컴퓨터의 언어처리 모형의 기본 원리 역시 인간의 언어 처리 과정을 기반으로 하고 있는 부분이 크기 때문이다. 따라서 학습자의 국어 학습 과정, 이해

19) 부연하자면, 인간의 독해 처리 과정에서 설명적 텍스트의 기본 구조는 “처음(대상의 제시, 논의의 필요성과 전개 계획, 맥락 등)－가운데(대상에 대한 설명, 주요 내용 전개)－끝(주요 내용에 대한 요약 정리, 부가, 전망 등)”을 갖는 것이 일반적이다. 즉, “……에 대해 살펴본다. 무엇에 대한 구체적인 주요 내용(먼저, 무엇보다도, 첫째, 우선, 다음으로, 둘째, 또 다른 등), 마무리(정리, 확인, 평가, 전망 등. 지금까지……에 대해 살펴보았다. 향후……할 것이다. 여기에는 김계성 외(2003) 등에서 적용된 단락 자동 구분을 이용한 문서 요약 시스템의 방식을 참조할 필요가 있다고 판단된다.



와 표현 과정에서의 실수와 오류, 사실적·추론적·비판적 이해와 표현의 인지적 사고 발달 과정을 핵심으로 하는 국어교육의 연구 내용은 컴퓨터 언어 처리에 기여할 수 있는 바가 크다. 지금까지 언어와 관련하여 이루어져온 음운, 어휘, 형태, 통사, 담화·텍스트 구조, 담화·텍스트의 이해와 처리 과정, 인지 언어학 등 모든 연구들이 문서 자동 요약 연구에만도 관련되지 않는 부분이 없을 정도이기 때문이다. 그리고 생략, 지시, 대응 이론, 스키마와 개념망 이론 등이 완성될 때 문서 자동 요약을 비롯한 컴퓨터 언어 처리에 대한 과제가 좀 더 완전하게 해결될 수 있다는 시각이 일반적이다. 이는 인간의 언어이해와 표현의 정밀한 과정이 해결되어야만 문서 자동 요약은 물론 컴퓨터 언어 처리가 좀 더 완벽해질 수 있다는 뜻이다.

따라서 학습자의 언어 이해와 표현 과정에 대한 보다 정밀한 탐색과 연구는 국어교육은 물론 컴퓨터 정보 처리 연구에도 중요한 자산이 되며, 여기에 상호 학제적 연구의 필요성과 중요성이 존재한다. 비전문가의 관점에서 접근한 부족한 글을 향후 수정 보완할 것을 기약하며 독자의 양해를 구한다.\*

---

\* 본 논문은 2010. 10. 30. 투고되었으며, 2010. 11. 5. 심사가 시작되어 2010. 11. 9. 심사가 종료되었음.

▣ 참고문헌

- 강인수(2010), “문서 자동 요약의 현황과 과제”, 『국어교육과 디지털 리터러시』, 국어교육학회 제46회 전국학술대회발표대회 자료집, 국어교육학회.
- 고영중 · 박진우 · 서정연(2002), “문장 중요도를 이용한 자동 문서 범주화”, 『소프트웨어 및 응용』 제29권 제6, 한국정보과학회, 417~424.
- 김건오(2002), “어휘 클러스터링을 이용한 주제어 판별 기반의 자동 문서 요약”, 서강대 대학원 석사학위논문.
- 김건오 · 고영중 · 서정연(2002), “어휘 클러스터링을 이용한 자동 문서 요약”, 『한국정보과학회 학술발표논문집』 제29권 제1호(B), 463~465.
- 김계성 · 이현주 · 이상조(2003), “단락 자동 구분을 이용한 문서 요약 시스템”, 『소프트웨어 및 응용』 제30권 제7·8호, 한국정보과학회, 681~686.
- 김금영 · 강인호 · 안동연 · 정성중 · 박순철(2002), “질의기반 자동문서 요약”, 『춘계학술발표논문집』, 제9권 제1호, 한국정보처리학회.
- 김봉순(1999), “설명적 텍스트의 중심내용 분석 원리”, 『국어교육학연구』 9집, 국어교육학회.
- 김재봉(1995), “문 주제 중심의 텍스트 요약과 거시규칙”, 『텍스트언어학』 3, 박이정, 31~82.
- 김재봉(1999), 『텍스트 요약 전략에 대한 국어교육학적 연구』, 집문당.
- 김태희 · 박혁로 · 신중호(1999), “검색/요약/필터링을 위한 텍스트 이해 모형 및 처리 기술 개발”, 과학기술부 연구개발정보센터.
- 남기종(2003), “신문기사 명사-동사 공기패턴 정보를 이용한 문서 자동 요약”, 전남대 대학원 석사학위논문.
- 류동원(2001), “2단계 문서분할을 통한 자동 문서요약 시스템”, 포항공과대학교 대학원 석사학위논문.
- 박건숙(2006), “자연언어처리를 위한 구문, 의미 정보 구축”, 『한국어의미학』 제19집 2006. 4, 한국어의미학회, 1~35.
- 박진호(2006), “자연언어처리와 국어 연구”, 『국어국문학』 제143권, 국어국문학회, 97~122.
- 반 다이크, 정시호 역(1995), 『텍스트학』, (주)민음사.
- 배희숙 · 백혜승 · 최기선(2003), “문서 자동요약을 위한 말뭉치 기반 언어정보 추출 : 백과사전 인물 관련 항목의 계량언어학적 연구”, 『계량언어학』 제2집, 39~52.
- 배희숙 · 최기선(1997), “정보검색 시스템의 성능 향상을 위한 복합어 전자사전”, 『프

- 랑스문화예술연구』 제6집, 프랑스문화예술학회, 69~82.
- 서 혁(1991), “단락, 문장의 중요도 파악과 단락의 주제문 작성 능력이 요약에 미치는 효과”, 서울대학교 대학원 석사학위논문.
- 서 혁(1994), “요약능력과 요약 규칙”, 『국어교육학연구』 4집, 국어교육학회, 113~142.
- 윤재민 외(2004), “육하원칙 활성화도를 이용한 신문기사 자동추출요약”, 『소프트웨어 및 응용』 제31권 제4호, 한국정보과학회.
- 윤재민 · 강인수 · 권오욱 · 배재학 · 이종혁(2002), “확장된 6하원칙을 이용한 신문기사 자동요약”, 학술발표논문집, 29(1B), 한국정보과학회, 466~468.
- 이삼형(1994), “설명적 텍스트의 내용 구조 분석 방법과 교육적 적용 연구”, 서울대학교 대학원 박사학위 논문.
- 이수희(2005), “독자의 목적에 따른 요약하기 체계화 연구”, 경인교육대학교 교육대학원 석사학위논문.
- 이유리 · 최기선(1999), “수사구조를 이용한 텍스트 자동요약”, 학술발표 논문집, 한국정보과학회 언어공학연구회, 97~102. <http://uci.or.kr/G300-cX1072144.vn0p97>.
- 이해운(2007), “담화표지에 의한 텍스트 구조 분석”, 『독어학』 제15집, 한국독어학회.
- 장철순(2008), “요약하기 실태 분석과 교수·학습 방안”, 이화여자대학교 교육대학원 석사학위논문.
- 천한신(1987), “인지 발달에 따른 글 내용의 중요도 파악 능력의 발달에 관한 연구”, 서울대학교 대학원 석사학위 논문.
- 추교남(2007), “멀티미디어 XML 문서에 대한 의미 분석 기반의 지능적 자동 요약”, 인천대학교 대학원 박사학위논문.
- 한경수(2000), “질의분해를 이용한 적합성 피드백 기반 자동 문서요약”, 고려대학교 대학원 석사학위논문.
- 황금하 · 신지애 · 최기선(2008), “개념 및 관계 분류를 통한 분야 온톨로지 구축”, 『소프트웨어 및 응용』 제35권 제9호, 한국정보과학회.
- Kintsch, W. and van Dijk, T.A.(1978), “Towards a model of text comprehension and production”, *Psychological Review* 85, 363~394.
- Marton, F. & Säljö, F.(1976), “On qualitative differences in learning : I -Outcome and process”, *British Journal of Educational Psychology*, Vol.46, 4~11.
- New London Group(1996). “A Pedagogy of Multiliteracies : Designing social futures”, *Harvard Educational Review* 66, 60~91.
- Wittrock, M. C. (1990). “Generative processes of comprehension”, *Educational Psychologist*, 24 (4), 345~376.

## &lt;초록&gt;

## 문서 자동 요약 기술의 발전을 위한 학제적 접근의 필요성과 과제

서 혁

국어교육에서 학습자의 언어처리 과정에 대한 과학적 탐구는 기여할 수 있는 바가 크다. 컴퓨터 언어처리 모형의 기본 원리 역시 인간의 언어 처리 과정을 기반으로 하고 있는 측면이 많기 때문이다. 따라서 학습자의 국어 학습 과정, 이해와 표현 과정에서의 실수와 오류, 사실적·추론적·비판적 이해와 표현의 인지적 사고 발달 과정을 핵심으로 하는 국어교육의 연구 내용은 컴퓨터 언어 처리에 기여할 수 있는 바가 크다. 지금까지 언어와 관련하여 이루어져 온 음운, 어휘, 형태, 통사, 담화·텍스트 구조, 담화·텍스트의 이해와 처리 과정, 인지 언어학 등 모든 연구들이 문서 자동 요약 연구에만도 관련되지 않는 부분이 없을 정도이다. 그리고 생략, 지시, 대응 이론, 스키마와 개념망 이론 등이 완성될 때 문서 자동 요약을 비롯한 컴퓨터 언어 처리에 대한 과제가 좀 더 완전하게 해결될 수 있다는 시각이 일반적이다. 이는 인간의 언어이해와 표현의 정밀한 과정이 해결되어야만 문서 자동 요약은 물론 컴퓨터 언어 처리가 좀 더 완벽해질 수 있다는 뜻이다. 따라서 학습자의 언어 이해와 표현 과정에 대한 보다 정밀한 탐색과 연구는 국어교육은 물론 컴퓨터 정보 처리 연구에도 중요한 자산이 되며, 여기에 상호 학제적 연구의 필요성과 중요성이 존재한다. 지금까지 논의한 결과를 바탕으로 학제적 관점에서 문서 자동 요약이 주는 시사점을 간단히 제시하면 다음과 같다.

첫째, 맥락정보 기반 접근법 활용의 필요성과 가능성이다. 제목이나 문맥 단서뿐만 아니라 담화·텍스트의 상황맥락, 사회문화적 맥락 등의 맥락 정보 활용 가능성을 고려해야 할 것이다. 특히 발신자, 수신자, 더 나아가서는 의도, 목적, 메시지, 관련지식(스키마)을 종합적으로 고려할 수 있는 방안이 온톨로지의 구축과 함께 고려되어야 할 것이다.

둘째, 담화 표지와 텍스트의 구조를 좀 더 적극적으로 활용하는 방안이 모색되어야 할 것이다. 기존의 담화 표지에 대한 연구 성과의 반영도 아직은 미흡한 수준으로 보이며, 특히 텍스트의 구조(초구조, 거시구조, 담화·텍스트 차원의 수사적

호응 관계) 등은 아직도 거의 반영되지 못하고 있는 것으로 보인다. 특히 텍스트 전개 구조를 참조한 입체적 벡터의 개념(텍스트 전개 방향, 텍스트 부분의 중요도)이나 등간격 지향성의 원리(첫째, 둘째, 셋째 등)은 구체적으로 공동 연구가 필요한 부분으로 생각된다.

【핵심어】 국어교육, 문서 자동 요약, 요약하기, 학제적 연구, 담화표지, 텍스트 구조

<Abstract>

## Some Prospects on the Research of Automatic Document Summarization for Interdisciplinary Approach

Suh, Hyuk

This thesis suggests some ideas on the interdisciplinary approach for the research of Auto Document Summarization (hereafter ADS) between Korean Education(hereafter KE) and Computer Engineering. The principles of language processing in computer are fundamentally based on the cognitive process of human. Therefore, the contents and results of KE which is interested in reading comprehension and expression including summarization contribute a great deal to ADS. The results of researches accomplished in KE are accumulated so much so far such as text comprehension, discourse and text structure, word acquisition, etc. The research on ADS will be more concrete when it refer to the process of text comprehension and expression of human being including learners. Therefore, interdisciplinary approach between KE and CE cannot be emphasized too much.

In this point some suggestions and prospects on ADS research are presented as follows. Firstly, the need and possibility of context-based approach on ADS. When we research on ADS, we need to consider on sociocultural context as well as context of discourse and text, furthermore addressor, addressee, intention, purpose, message and schemata when we construct ontologies. Secondly, we need to consider the discourse marker and text structure more deeply and precisely, including micro-structure, macro-structure, super-structure and sequences of rhetorical expressions.

**【Key words】** Korean Education, Automatic Document Summarization, summarization, Discourse Marker, text structure

## 【부록】

## 요약 대상 텍스트 원문 및 MS Word 2003의 요약 결과\*

신세대 어린이 문화. 이는 우리가 익숙하게 듣는 말로서, 어른 문화와는 다른 어린이만의 독특한 생각이나 행동방식을 일컫는 말이다. 그런데 요즘 어린이 문화에 대해 우려하는 어른이 적지 않다. 그러나 과연 이것이 우리 어린이만의 문제인가? 어린이 문화는 우리 사회 문화의 일부일 뿐이다. 지금의 어른도 한 때는 어린이였으며, 우리 어린이도 자라 나서 언젠가는 미래의 어린이를 걱정하는 어른이 될 것이다. 따라서, 어린이 문화는 어린이만의 문제가 될 수 없다. 요즘처럼 사회가 빠른 속도로 변화하는 시대에는 당연히 세대 차이가 더욱 커지고 그에 따른 문화의 갈등도 한층 더 심해지게 마련이다. 그렇다면 이러한 세대 문화 갈등의 실태는 어떠한지, 그 원인은 무엇인지, 그리고 세대문화의 갈등을 어떻게 극복해야 하는지에 대해서 살펴보도록 하자.

세대차이로 생기는 문화적 갈등은 어린이의 옷차림, 언어 사용 등에서 그 실상을 엿볼 수 있다. 어린이의 옷차림은 옛날과는 많이 다르다. 예전에는 부모님이 사 주신 옷을 단정하게 입었지만, 요즘 어린이는 유행에 따르면서도 자신의 독특한 개성을 살려 입는다. 길이가 길고 헐렁헐렁한 바지, 이해할 수 없는 외국어로 가득한 옷, 아주 짧은 반바지나 치마를 입은 어린이를 쉽게 볼 수 있다. 그래서 어른들은 이런 어린이의 옷차림에 대해 안타까워하고 걱정하기도 한다. 언어 사용에서도 세대문화의 갈등은 일어난다. 물론, 언어의 세대 갈등은 어제오늘의 일은 아니다. 언어는 사용하는 사람의 얼굴이면서 그 사람이 살고 있는 사회문화를 반영하는 것이다. 그러므로 어린이의 말이나 말투, 글이나 글투를 보면 그들의 생각을 엿볼 수 있다. 예를 들어, ‘썰렁하다’, ‘웬 내숭’ 등과 같은 말을 때나

\* 13% 요약(372자, 밑줄과 함께 진한 글씨체)과 25%(밑줄만 친 부분, 764자) 요약의 결과

장소를 가리지 않고 사용한다. 어른이 어린이에게 그들이 사용하는 언어에 대해 설명을 요구하면, “왜요? 좋잖아요?”라는 말로 되받아서 도리어 어른들을 이해성이 부족한 사람으로 만들곤 한다. 어른들은 이러한 언어 사용이 계속되면, 고운 우리말이 사라질 것이며, 바른 심성이 파괴될 수 있을 것이라고 우려한다.

그러면 이와 같은 세대 문화의 갈등이 생기는 까닭은 무엇일까? 그 하나는 대화의 단절이다. 사람들은 복잡한 사회에서 살아남기 위해 각자 자기 일에 대부분의 시간을 투자한다. 그렇기 때문에 가족이 함께 대화를 나누는 시간이 줄어들었다. 그래서 부모는 자녀가 무엇을 고민하고 어떤 일을 하고 있는지 잘 알지 못하고 자녀 또한 부모의 생각이나 관심사를 이해하지 못한다. 결국, 서로 점점 멀어지게 되어 세대문화의 갈등이 생기는 것이다. 다른 하나는 이기주의이다. 어린이와 어른은 모두 자기 입장에 서만 생각하고 상대방을 고려하지 않는다. 어른은 어린이의 개방적인 사고를 그저 억누르려고만 한다. 반면에, 어린이는 지금까지 지켜 온 것들을 무시하거나 거부하려고 한다. 이 밖에도 세대 문화의 갈등을 부추기는 요인이 또 있다. 텔레비전, 라디오, 영화, 신문, 잡지 등 대중 매체의 영향이 바로 그것이다. 대중 매체의 내용을 살펴보면 어린이와 어른의 관심사가 분명하게 구분된다. 텔레비전의 오락 프로그램만 하더라도 대부분이 신세대만을 위한 오락 프로그램과 기성 세대만을 위한 오락 프로그램으로 나뉘어져 있다. 이렇게 텔레비전 프로그램을 나누어 버리면 시청률은 상승할지 몰라도 가족 사이의 세대 갈등은 점점 심해질 것이다.

그러면 이러한 세대 문화의 갈등을 극복하기 위해서는 어떻게 하여야 할까? 첫째, 서로 이해의 폭을 넓히기 위해 대화를 자주 나누어야 한다. 상대방의 문화를 무조건 나쁘다고만 할 수 없다. 어린이 문화는 미래의 어른 문화로 피어나게 될 씨앗이라 할 수 있다. 또, 어른 문화는 이러한 어린이 문화가 성장할 수 있도록 도와주는 밑거름이라고 할 수 있다. 그러므로 어린이와 어른은 서로 이해하여 세대 문화의 갈등을 대화로 풀어 나가야 할 것이다. 마음 속에 있는 생각을 터놓으면 상대방과의 거리를 좁힐 수 있다. 둘째, 어린이와 어른 서로의 장점을 받아들여 조화시키



려는 노력을 해야 한다. 어린이는 창의적이고 기발한 상상력을 가지고 있다. 그리고 사물을 새롭게 볼 수 있는 능력도 있다. 예를 들어, 어른이 시계를 보고 시간을 나타내는 기계로만 이해할 때, 어린이는 리듬 측정기로도 이해한다는 것이다. 이처럼, 어린이는 툭툭 튀는 새로운 생각을 가지고 있다. 반면에 어른들은 어린이들이 겪어 보지 못한 다양한 경험과 풍부한 지혜를 가지고 있다. 그러므로 어른은 어린이의 독창성과 상상력을 어린이는 어른의 경험과 지혜를 받아들여 발전시켜야 한다. 셋째, 어린이와 어른은 함께 도덕적인 사회를 만들기 위해 노력해야 한다. 사회가 도덕적으로 흔들리면 어린이는 어른을 불신하게 될 것이다. 도덕적으로 타락한 어른은 어린이들의 고민을 진정으로 이해할 수 없을 것이며 이해하려 들지도 않을 것이다. 그러므로 우리는 바른 사회를 만들어 세대 문화의 갈등을 극복해야 한다.

지금까지 우리는 세대 문화의 갈등의 실상과 원인, 그리고 그 극복 방안에 대하여 살펴보았다. 세대 문화의 갈등은 어린이만의 문제가 결코 아니다. 이것은 어른에게도 책임이 있다. 그러므로 세대 문화의 갈등을 해소하기 위해서는 서로를 이해하려는 마음으로 대화의 기회를 많이 가져야 한다. 그리고 어린이는 어른들의 경험과 지혜를 배우고, 어른들은 어린이의 처지에서 이해하려는 노력이 필요하다. 그리하여 진실과 정직이 바탕을 이루는 사회를 함께 만들어 나가야 한다. (총 2,660자. 지문 출처 : 교육과학기술부(2007). 초등학교 국어과 교사용 지도서, 대한교과서주식회사.)