

초등학생들이 사용하는 채팅 언어의 분석

정희택* · 정영순*

여수대학교 교육대학원 컴퓨터교육전공

요 약

본 연구에서는 초등학생들이 사용하는 채팅 언어를 분석하고 이를 통해 발견된 특성을 제시하며, 나아가 이를 해결할 수 있는 방안을 제안한다. 채팅언어의 분석은 문장들 간의 유기적 관계를 고려하여 형식상, 내용상, 통계상의 특성으로 구분하여 분석한다. 분석 결과로 유아발음으로 변형되어 쓰기, 음절의 앞 뒤 바꾸어 쓰기, 상대방이 불분명한 대화와 같은 새로운 특성이 발견되었다. 분석 결과를 통하여 언어의 오용을 최소화하고, 아동의 정서발달에 긍정적인 채팅 문화를 정착시키기 위해 채팅 관련 컴퓨터 시스템적 방안을 제안한다.

Analysis on the Chatting Language of Elementary School Students

Hee-taek Ceong* · Yeong-sun Jeong*

Yosu National University, Dept of Computer Education

ABSTRACT

In this research, we analyze chatting language of elementary school students, present the properties which are a result of this analysis and propose the method to solve it. Analysis on the chatting language considering organic relation of sentences is classified into syntactic, semantic and statistical characteristics. As a result of analysis, new characteristics is discovered writing by the way children pronounce their words, writing by switching the order of syllable's frontal and rear ends, inarticulate conversation of the other man. To minimize language misuse and to establish a chatting culture that is favorable to the development of children's emotional development, we propose the methods of computer system.

Keyword : Chatting language, Data mining, Syntactic analysis, Semantic analysis, Statistical analysis

1. 서 론

컴퓨터 통신 인구가 급증하면서 급기야는 컴퓨터 통신에 쓰이는 ‘통신 언어’가 따로 생겨나기에 이르렀다. 통신 언어는 크게 채팅 언어(chatting language), 도배언어(emoticon language), 네트워크 게임상의 게임언어(game language), 전자게시판언어(bulletin board system language), 핸드폰언어(cellular phone language)로 분류한다[1,2]. 통신 언어는 정보의 비공개성이나 비속도성 그리고 현장성 등 여러 가지 문제를 해결하였다[3]. 그러나 다음과 같은 문제점들을 가지고 있다. 첫째, 통신 언어는 의사소통에 장애를 주기도 한다. 짧은 시간에 좀 더 많은 정보를 주고받기 위해 언어를 축약하거나 기호 등을 이용하여 새로운 언어들 만들어 통신상에서 사용하는 것이다. 하지만, 인터넷이라는 가상공간의 특성을 감안할 때 왜곡된 통신 언어는 네티즌 상호간의 의사소통 장애를 끊임없이 일으키는 결과를 초래한다. 둘째, 부정적으로 말하기, 냉소적으로 말하기 등의 언어 형태는 개인의 정서를 황폐하게 한다[4]. 인터넷상에서 사용자는 다른 사람의 간섭을 받지 않는 가상공간 속에서 남들과 다른 존재성을 부각시키기 위해 자신들만의 공간을 확보하여 현실 공간으로부터 일탈하려는 욕구를 표출한다. 따라서 일탈 현상은 언어의 변형이나 비속어, 냉소적인 표현 등으로 나타나게 되며 결국 언어의 품위가 떨어지게 되고 아울러 개인의 정서는 메탈라가고 인간을 경시하는 풍조로 이어지게 되는 것이다. 마지막으로, 통신 언어는 국어의 언어 규범에 혼란을 준다[3]. 가상공간에서 사용되는 통신 언어는 일상 언어에 영향을 미치게 된다.

본 연구에서는 통신 언어의 특성 연구를 위해 채팅 언어를 선택하여 분석한다. 이는 채팅 언어가 온라인상에서 실시간으로 대화가 이루어지며, 입말체에 가까운 문체를 구

사하고, 시간과 타자의 효율성을 중요시한다는 특성을 지니고 있는 언어이기 때문이다. 본 연구에서 분석대상은 초등학생들이 사용하는 채팅 언어를 대상을 수행한다. 이는 초등학생들은 아직 언어 발달 단계에 머무르고 있다는 점을 감안할 때 그들이 사용하는 채팅 언어는 자칫 국어에 대한 오류개념을 습득하게 할 뿐 아니라, 정서적인 면에서도 부정적인 영향을 미치게 될 것이라 판단되었기 때문이다. 아울러 연구자가 초등학교 교사로서 학생들의 언어 사용과 인터넷에서의 언어 사용에 극단적인 몇몇 경우를 경험하면서, 이러한 연구를 통해 문제를 분석하고 나아가 완화할 수 있는 방안을 모색하고자 본 연구를 수행하였다.



<그림 1> 채팅언어 분석 과정

본 연구에서는 그림 1과 같이 분석 방법으로써 크게 언어적 관점과 통계적 관점에서 수행한다. 언어적 관점은 채팅 언어도 하나의 언어라고 하는 관점에서 초등학생들의 채팅 언어가 갖는 특성들(primitives)을 분석하였다. 이를 위해 채팅 언어가 갖는 외형적인 형식상의 특성과 의미적 특성을 다룬 내용상의 특성을 분석한다. 결과로 생성된 특성들은 일반적인 채팅 언어의 특성 이외에도 초등학생들이 갖는 특성을 분석할 수 있었다. 다음으로 앞서 분석된 특성들의 통계적 특성을 분석한다. 각 특성들의 빈도와 자판기 배열에 따른 빈도 특성을 분석한다. 이

는 각 특성들의 유효성을 확인하기 위한 과정이다. 마지막으로 각 특성들이 채팅 자료 상에 나타나는 연관성을 분석한다. 하나의 채팅 문장에 하나이상의 특성들이 존재하고 그들 간에 연관성을 분석하여 특성들 간의 관계를 분석하고자 한다. 이러한 분석을 통해 언어의 오용을 최소화하고 긍정적인 채팅 문화 정착을 위한 컴퓨터 시스템적 방안을 제안한다. 아울러 기존 연구에서는 단순 언어적 관점에서만이 분석이 이루어졌고 초·중·고등학생을 대상으로 한 분석이 이루어지지 않았다.

논문의 구성은 다음과 같다. 2장에서는 통신 언어 특성을 분석한 기존 연구 및 데이터 탐사 기법을 제시한다. 3장에서는 초·중·고등학생들이 사용하는 채팅 언어의 특성을 형식적, 내용적 특성에 근거하여 언어적 분석을 수행한다. 4장에서는 분석된 특성들을 기반으로 단순 통계적 특성 및 연관 규칙 탐사 기법을 이용한 분석을 수행한다. 마지막으로 5장에서는 결론 및 분석결과로써 시스템적 제언을 기술한다.

2. 관련 연구

2.1 언어학적 분석에 대한 기존 연구

통신언어에 대한 언어학적 관점에서 분석한 기존 연구들은, 발견된 현상에 대한 국어학적 관점에서 분석만을 수행하거나 각 현상을 나열하는 형태의 분석만이 이루어졌다. 먼저 [5]의 연구에서는 통신 언어가 국어생활에 미치는 특성을 분석한 것으로 분석기준을 크게 음운과 표기, 어휘와 변형, 통사와 문법으로 구분하여 분석하였다. [6]에서는 전자게시판에서 사용하는 통신언어의 특성을 분석하였다. 이 연구에서는 분석 기준을 모음의 변화, 자음의 변화, 문장 종결어미의 형태 변이, 신조어와 의미의 변화, 기호의 사용 양상으로 구분하여 수행하였다. 고교

생들의 사이버 어휘특성을 분석한 [7]의 연구에서는 사용된 어휘들에서 발견된 특성을 기반으로 분석 기준을 제시하였다. 제시한 분석 기준으로는 연음 표기 어휘, 자·모음 동화 어휘, 음운과 음절 첨가 어휘, 음운과 음절 탈락 및 축약 어휘, 준말 어휘, 문자나 기호 조합 어휘, 약어 어휘, 외국어 어휘, 은어, 유행어, 신조어 어휘, 비속어, 욕설 어휘, 방언 어휘로 구분하였다. 이러한 분석은 각 단위 어휘 수준에서 분석을 수행하였을 뿐 문장단위의 분석은 포함하고 있지 않다. [4]에서는 통신언어에서 발견되는 현상들에 기인하여 분석 기준을 크게 9가지로 구분하였다. 소리 나는 대로 쓰는 현상, 음절을 줄여 쓰는 현상, 문말에 불필요한 자음을 붙여 쓰는 현상, 모음을 바꾸어 쓰는 현상, 은어·신조어·비속어 남용 현상, 기호를 이용하여 쓰는 현상, 존댓말을 사용하지 않는 현상, 부정적·냉소적으로 응답하는 현상으로 구분하였다. 마지막으로 [3]에서는 통신언어의 분석 기준으로 크게 음운적 특성과 형태적 특성, 어휘적 특성으로 구분하여 수행하였다. 이러한 기존 연구들은 통신 언어의 현상적 특성을 현상적으로 구분하여나 나열하였으며, 언어의 특성이 갖는 의미전달 관점 및 통계상의 특성을 분석하고 있지 않다. 본 연구에서는 기존 연구들의 이런 분석 기준들을 포함한 정형화된(formal) 분석 틀을 제안하고자 한다. 기존 연구들에서 제안한 분석 기준들의 내용은 본론에서 기술하기 때문에 여기서 지면관계상 기술하지 않는다.

2.2 데이터 탐사 기법

본 연구에서는 통신 언어의 통계적 특성을 분석하기 위한, 하나의 방법으로 데이터 탐사방법을 이용한다. 이를 위해 데이터 탐사기법에 대한 특성을 간략히 기술한다. 데이터 탐사(mining)는 대용량의 데이터로부터

더 유용하게 활용될 수 있는 지식을 효과적으로 찾아내는 지식 탐사의 한 연구 분야이다. 데이터 탐사방안은 일반적으로 순차패턴(sequential patterns), 분류(classification), 클러스터링(clustering), 연관규칙(association rules)으로 구분한다[8,9,10,11,12,13].

본 연구는 채팅 언어의 사용에서 어휘간에 또는 문장간에 존재하는 연관성을 알아보기 위해 연관규칙 탐사를 이용한다. 연관규칙은 항목 집합으로 표현된 묶음에서 각 항목간의 연관성을 반영하는 규칙이다. 연관규칙을 탐사하는 방법은 기본적으로 미리 결정된 최소 지지도 이상을 만족하는 항목 집합에 대해 연관 규칙을 생성하는 단계로 이루어진다. 데이터 사이의 연관성을 찾는 연관 규칙 탐사에 관한 분야는 한 데이터 항목들의 그룹과 다른 데이터 항목들의 그룹 사이에 존재하는 연관성 탐사에 응용된다.

데이터 사이에 존재하는 연관 규칙 탐사 기법에는 AIS, Apriori 등이 있다[9,10,11,12,13]. 연관 규칙 탐사의 기본적인 골격을 다음의 2단계로 구성한다. 1단계는 사용자가 정의한 최소 지지도를 만족하는 데이터 항목의 집합을 탐사한다. 각각의 데이터 항목에 대한 지지도를 계산하여 미리 정의된 최소 지지도를 만족하는 데이터 항목들을 추출해 낸다. 2단계는 연관 규칙을 생성한다. 1단계에서 탐사된 빈발 데이터 집합을 이용하며, 데이터의 부분 집합에서 생성된 규칙 중 사용자가 정의한 최소 신뢰도를 만족하는 규칙들을 최종 규칙으로 생성한다.

3. 초등학생들이 사용하는 채팅언어의 언어적 특성 분석

채팅언어의 언어적 특성은 크게 두 가지로 구분하였다. 첫째, 채팅 언어가 갖는 외형적, 구조적으로 나타나는 ‘형식상의 특성’과 둘째, 채팅 언어의 의미적 특성을 다룬 ‘내용상의 특성’이 있다.

3.1 형식상의 특성

가. 된소리적기와 소리 나는 대로 적기
된소리적기나 소리 나는 대로 적기는 채팅언어의 전형적인 입말 특징에 기인한다. 특히 된소리적기는 타자의 비경제성에도 불구하고 소리에 변화를 주는 것과 강한 어조를 남기려는 의도에서 사용한 것이다. 소리나는 대로 적는 것은 대체로 음소를 줄이고, 쉽게 적을 수 있다는 표기상의 이로운 점이 있다.

- (1)된소리 적기:쌀앙(사랑),그렇게(그렇게)
- (2)소리나는 대로 적기:싫어(시러),월래(원래)

나. 유아발음으로 변형되어 쓰기

소리나는 대로 쓰기의 규칙을 벗어난 특성을 새롭게 발견할 수 있었는데, ‘유아발음으로 변형되어 쓰기’이다. 문자 대화를 하다 보면 가장 취약한 부분이 바로 화자의 감정을 표현하지 못한다는 것인데, 채팅에서는 평소의 어투보다 더 강조된 유아발음을 사용하게 된다 : 아라쳐(알았어), 무서버(무서워), 씨러(싫어), 업떠요(없어요).

다. 음운 탈락과 첨가

음운변동은 일어나는 가장 큰 이유는 ‘언어의 경제성’때문이다. 그 외에도 표현 효과를 높이고, 청각 인상을 보다 뚜렷하게 하며, 유추현상의 이유도 있고, 일상에서 말을 잘못 사용한 것이 그대로 굳어지며 변동이 일어나는 경우도 있다.

- (1) 음운 탈락

음운 탈락이란 어떤 음운이 완전히 떨어져 나가는 것을 뜻한다. 음운탈락은 다시 닿소리탈락과 홀소리탈락이 있다. 음운을 없애거나 줄이는 것은 채팅언어의 가장 일반적인 특징이다. 채팅언어는 입말의 특징이 있기 때문에 입말에서 소리 줄임이나 소리 없앰과 비슷한 현상으로 나타난다. 그런데 채

팅언어가 입말보다 소리 없앰이나 줄임이 더 심한 것은 속도와 시각성이라는 두 가지 환경 때문이다.

(가) 당소리 탈락:괜찬은데(괜찮은데),귀차나서(귀찮아서)

(나) 홀소리 탈락:외로운가바(외로운가봐),조야(취야)

(2) 음운 첨가

음절 끝소리에 당소리를 덧붙이는 현상이 있다. 이것은 상대방에게 인상을 강하게 심으려는 의도와 어형에 변화를 주어 딱딱함을 없애려는 의도이다.

(가) 당소리 첨가

- 1) ㅂ: 나은뎀(나은데),넵(네)
- 2) ㅅ: 해욏(해요),안녕하시욏(안녕하시오)
- 3) ㅇ: 갈넝(갈네), ~거등(~거든)
- 4) ㅁ: 뵙(네), 누구신깁(누구신가)

(나) 홀소리 첨가: 알쥬(알지)

라. 음절 생략과 첨가

(1) 음절 생략

소리의 줄임이나 없앰에서 나타나는 특징은 주로 홀소리를 없애거나 줄이는 것이다. 홀소리는 입말에서도 쉽게 줄여지는 특징이 있다. 그러나 당소리는 울림소리 당소리가 쉽게 없어지지 않는 특징이 있기 때문에 채팅언어에서도 쉽게 없애지 않는다.

(가)사이 홀소리 없애고 줄임:멤(마음),앤(애인),갈켜(가르켜),넴(너무)

(나)끝 홀소리 없애고 줄임:울(우리),웃지말(웃지 마라),해주셈(해주세요)

(다) 음절을 하나만 없애기:감사요(감사해요),알썰요(알았어요),쩍다(전부다)

(라)하나 이상의 음절을 없애고 줄임:비번(비밀번호),훤배(훤씬 배로)

(2) 음절 첨가

음절을 첨가하여 상대방에게 표현의 변화를 주어 강한 인상을 주려고 하거나, 단순히 어

형에 변화를 주어 딱딱함을 없애려는 의도이다.

(가) 음절 늘임: 애들아아(애들아)

(나) -여: 사절여(사절이에요)

마. 홀소리 바꿈

채팅 언어에서 홀소리를 바꾸어 사용하는 경우가 많다. 홀소리의 바꿈은 입말체의 특성이며, 변화를 주어 규범에서 이탈하려는 표현 의도에서 나타난 것으로 정리된다: 읍는거(없는거), 그란다(그런다), 승질(성질).

바. 띄어쓰기 오류

채팅 언어에서는 사용자의 편의에 따라 띄어쓰기가 무시된다. 강조하기 위해서나, 타자의 효율성을 위한 것이다: 재미있는 분이시네요, 딸꾹질 멈춰요, 트랜스 젠더

사. 홀소리만 적기와 당소리만 적기

일종의 음운 탈락 현상으로 홀소리나 당소리만 쓰이는 경우이다. 주로 의성어나 의태어 또는 간단한 인사말 등을 표기할 때 쓰인다.

(1) 홀소리만 적기:ㄱ(예), ㅠ(유), ㄷㅠㄷㅠ(우유우유), ㅏㅑㅓㅑ(아야아야)

(2) 당소리만 적기:ㅋㅋㅋㅋ,ㅎㅎㅎ

아. 음절의 앞 뒤 바꾸어 쓰기

음절을 생략하는 것과 또 다른 현상의 하나로 새롭게 발견한 것이 음절의 앞과 뒤를 바꾸는 현상이다. 일종의 은유적인 표현으로 상대방에게 직접적으로 말하기에 거북한 낱말의 경우에 쓰인다: 랄지(지랄)

자. 문장의 미완결성

채팅 언어에서는 문장을 완결하지 않고 쓰는 경우가 있다. 한 단어나 한 어절로 문장을 대신하는 경우가 있는가하면, 후부요소가 생략된 경우와, 명사형으로 서술어가 끝

나는 경우로서, 이는 채팅어가 갖는 입말체의 특성을 나타내는 대표적인 현상이다.

- (1) 한 단어로 대화를 주고받는 경우 : 또여, 그냥, 네에
- (2) 한 어절로 대화를 주고받는 경우 : 안녕하세요, 어쨌세요
- (3) 후부요소가 생략된 경우 : 그럼 이만
- (4) 명사형으로 서술어가 끝나는 경우 : 빼집

차. 조사의 생략과 오용

조사의 생략과 오용 현상은 입말에서 자연스럽게 일어나는 현상이다. 문장 성분에서 조사가 생략되는 일이 빈번한 것은 한 컴퓨터 자판기에서 한 타수라고 줄이려는 노력이다.

- (1) 조사의 생략 : 나 전화(가) 생겨서...,
- (2) 조사의 오용 : 바로 여러분에(게) 달려 있습니다.

카. 어순의 혼란

국어의 어순은 정치법에서 주어+목적어+서술어의 구조인데, 수식어가 어느 곳에 놓이느냐에 따라서 부자연스러운 경우가 많다. 채팅에서 국어의 어순에 맞지 않는 문장을 구사하는 것은 국어의 어순을 미처 잘 알지 못해서 표기하는 경우도 있지만 입말의 특성을 갖기 때문이다 : 념 여기 복잡하다.(여기 너무 복잡하다.).

3.2 내용상의 특성

가. 한 문장에서 나타나는 내용상의 특성

(1) 상징 기호를 이용한 표현

상징기호가 채팅에서 많이 쓰이는 이유는 채팅 언어가 글말과 거의 비슷한 대화 환경을 가지고 있기 때문이다. 화자는 자기의 표정이나 감정 등을 컴퓨터 문자판을 최대한 이용하여 상징화한다.

(가) 감정 표현:ㅠ_ㅠ(슬프다), - -(느낌 없다), >_<(부끄럽다) _-(화남),

(나) 특정 사람, 사물 표현:凸(상대방을 화나게 하는 표시), 乃(엄지손가락, 추켜세움)

(다) 문장의 기호화:sㅏ이(사이), ㅁ2(빨리), 하2(하이·안녕)

(라) 자음과 모음을 따로 분리하는 경우:오 ㅈ 불러주세요(왜불렀어요)

(2) 은어와 신어

(가) 은어(cant) 사용

통신언어에서 나타나는 뚜렷한 특징 가운데 하나가 은어의 사용이다. 이것은 채팅언어가 익명성이라는 대화 환경과 대화 대상이 비슷한 집단이라는 특수성 때문이다. 은어 사용은 또래집단의 확인으로 이어지고 나아가 그들 사이에 배타적 공감대를 만드는 기재로 작용하며, 이것은 계층간의 대화 단절에 매우 중요한 요인으로 작용한다 : 대빵(대장), 쪽팔리스(창피해서), 허덤(바보).

(나) 신어의 생성

신어에는 전혀 새로운 단어형을 만들어내는 경우도 있으나 이는 극히 드물며, 기존어휘를 이용하거나, 외래어나 방언에서 차용하는 경우가 많다. 이러한 현상들은 채팅언어의 생략성에서 볼 때 자연스러운 현상이다.

1) 새로이 창조하는 방법:리하이(다시 안녕하세요), 초딩(초등학생채팅)

2) 기존의 어휘를 이용하여 만드는 방법 : 주번(주민등록번호)

(다) 통신 메타언어(meta-language)생성

통신 메타언어(meta-language)의 발달은 통신언어가 하나의 정보 전달 매체로 작용하기 때문에 통신언어의 운용상 반드시 필요하다. 그런데 통신 메타언어가 은어와 비슷하게 생성된다는 것이 문제다. 메타언어는 보편성과 합리성을 가지고 사회적으로 공인되는 과정을 거쳐야 한다 :방장(채팅방을 개

설한 사람),방청소(채팅방을 정리하는 것)

(3) 비속어와 욕설

채팅 언어에서 비속어나 욕설의 사용은 통신언어 특성상 어쩌면 당연하다고 하겠다. 통신언어의 익명성이나 폐쇄성으로 상대를 직접 대면하고 말할 때보다 말하는 사람은 훨씬 자유스러운 분위기에서 대화를 한다.

(가) 비속어(slang) 사용 : 갈구지마, 쪽팔리다,

(나) 욕설(abusive language)의 사용 : 싸가지, 대가리

(4) 방언

방언이 전국적으로 확대되는 것은 채팅을 하는 사람들이 전국적인 분포를 보이고, 전자매체가 공간적 제약을 받지 않기 때문이다. 그리고 채팅 언어 사용자들이 규범성이나 일상성을 벗어나려는 표현 욕구에 의해 방언의 사용도 자연스러워 지고 있는 것 같다.

(가) 방언이 쓰이는 경우:니(네), 무신(무슨), 하노(하니), 마라요(마세요)

(나) 여러 지방의 방언을 섞어서 사용하는 경우

대화명을 기준으로 한 명의 대화자가 사용하는 방언 내용을 비교한 결과 경상도와 전라도, 충청도의 방언을 골고루 사용하고 있는 것을 알 수 있다: 꺼져버려 : 걱정마 니 걱정이나 실컷혀(충청도), 꺼져버려:나는 잠시 쉬어야긔다(전라도), 꺼져버려:싫은디(전라도), 꺼져버려:니는 쓰레기가?(경상도).

(5) 외국어의 사용

초등학생들이 사용하는 채팅 언어에서 외국어의 사용은 많지 않다. 아무래도 성인에 비해 외국어 어휘력이 떨어지기 때문에 사용상 한계가 있는 것으로 보인다: 미투(나도 역시 그래), 굿잡(잘 했어요), 마미(엄마)

(6) 부름말과 존댓말

(가) 부름말 ‘님’

‘님’은 존경의 뜻을 나타내는 접미사였다. 그러나 통신언어에서는 의존명사나 대명사의 용법으로 일반화하였다. 채팅 언어에서 ‘님’은 상대방의 지위나 나이 등의 요인을 고려하지 않은 채 ‘두루 높임’의 형식으로 쓰이고 있다 : 가면님, 꿈님, 냄새님, 대성님

(나) 존댓말을 사용하지 않는 경우

채팅 언어에서는 문장의 끝에 ‘여’를 붙여 표현하는 두루 높임인 ‘해요체’를 사용하고 있거나 ‘해체’를 많이 사용하는 것을 볼 수 있다. 이는 대화방이 또래끼리 방을 분류해 놓고 있기 때문에 연령층이 비슷할 것이라는 생각을 할 수 있다.

나. 문장과 문장 간에 나타나는 내용상의 특성

(1) 부정적, 냉소적으로 응답하는 현상

아동들은 대화방에서 아무 이유 없이 냉소적으로 말하거나 부정적인 표현을 사용하여 상대방을 언짢게 하는 경우가 많다. 이는 아동들의 정서에도 좋지 않은 영향을 미치게 되며, 통신 예절에도 어긋나는 예라고 할 수 있다.

(2) 상대방이 불분명한 대화

대화방에서 상대방의 대화를 끝까지 읽고 나서 자신의 의견을 말하는 이가 있는가 하면 상대의 의견은 전혀 개의치 않고 자신의 의견만 말하는 경우가 있다. 그 외에도 화자의 타자 속도가 차이가 나기 때문에 나타나는 특성이자.

(3) 같은 내용을 반복하기

자신의 의견을 강조하기 위해서, 또는 장난삼아 같은 내용을 계속 반복하는 것을 볼 수 있다. 컴퓨터가 복사 기능을 가지고 있기 때문에 더 자유롭게 사용된다고 여겨진다.

4. 초등학생들이 사용하는 채팅언어의 통계적 특성 분석

채팅 언어가 어떤 경향을 띠며 사용되고 있는가를 알아보기 위한 ‘통계상의 특성’으로 구분하여 분석한다.

언어의 분석을 위해서는 어떤 특성을 가진 언어들이 많이 사용되고 있으며 또한 어떠한 이유 때문이지를 알아볼 필요가 있다. 이를 밝히기 위한 분석 자료의 수집 방법은 초등학생들이 사용하는 채팅방(<http://buddybuddy.co.kr>과 <http://chat.sayclub.com>)에 접속한 후 무작위로 대화의 내용을 복사하였다. 초등학생들이 사용하는 채팅방의 내용을 수집하였다고 해서 모두 내용이 초등학생의 것이라는 것은 보장할 수 없다. 그러나 순도 높은 자료의 수집을 위해 연구자가 각 채팅 내용을 출력하여 확인하였고, 의미 있는 자료 수집을 위해 방학이 포함된 2003년 12월 15일부터 2월 18일까지 2개월간 자료 수집을 수행하였다. 분석 대상자는 대화명을 기준으로 모두 248명이며 수집한 총 대화 수는 6499 개이고, 어휘 수는 14,496 개다.

본 연구에서는 통계적 특성 분석하기 위해 단순 통계적 특성 분석과 분석된 특성들 간의 연관성을 분석하기 위한 연관 규칙 탐사를 이용한 분석을 수행한다.

4.1 단순 통계적 특성 분석

단순 통계적 특성 분석은 앞서 분석된 특성들의 존재를 확인함으로써 분석된 요소들의 유효성을 제시한다. 이를 위해 세 가지로 구분 수행한다. 첫째, 어휘 단위별 빈도를 분석한다. 채팅 언어의 가장 기본이 되는 요소인 어휘별 빈도를 분석한다. 둘째, 문장 단위별 빈도를 분석한다. 여러 어휘들이 하나의 문장을 이룰 때 갖는 특성을 기반으로 빈도를 분석한다. 셋째, 채팅에 사용된 전체 어휘

의 글쇠 빈도를 비교하여 분석한다.

가. 어휘 단위별 사용 빈도 분석

전체 어휘단위별 사용 빈도를 분석한 결과는 표 1과 같다. 한 대화에 사용된 어휘의 수는 평균 2.23개인 것으로 나타났다. 이것은 채팅 대화에서는 간결한 대화가 이루어지고 있음을 의미한다. ‘정상적인 어휘’가 39.8%로 가장 높게 나타났다. 문장의 완성도가 낮았던 표 2의 결과와 비교하면 정상적인 어휘가 많음을 알 수 있다. ‘상징 기호를 이용한 어휘’가 18.7%로 가장 높은 비율을 차지하고 있다. 상징 기호의 대부분이 특수 문자나 기호를 사용한다. ‘띄어쓰기 오류 어휘’가 8.8%로 나타났다. 채팅에 관심을 갖고 참가하는 학령이 대부분 고학년 아동들인데 이 결과를 보면 띄어쓰기 오류가 많은 편이다.

<표 1> 어휘단위별 사용 빈도

구분	기준	빈도	%
기타	정상적인 어휘	5,775	39.8
내용상 특성	상징기호를 이용한 어휘	2,709	18.7
	띄어쓰기 오류 어휘	1,274	8.8
형식상 특성	홀소리와 닿소리만 적은 어휘	1,159	8.0
	음운이 탈락되거나 첨가된 어휘	680	4.7
	음절이 생략되거나 첨가된 어휘	593	4.1
내용상 특성	은어와 신어 어휘	569	3.9
	비속어와 욕설 어휘	417	2.9
형식상 특성	바꿈 한 어휘	392	2.7
내용상 특성	변형된 부름말과 존댓말 어휘	259	1.8
	외국어 어휘	247	0.0
형식상 특성	된소리 적기와 소리 나는 대로 적은 어휘	238	0.0
내용상 특성	방언 어휘	131	0.0
형식상 특성	유아발음으로 변형된 어휘	45	0.0
	음절의 앞뒤가 바뀐 어휘	8	0.0
	계	14,496	100

나. 문장 단위별 사용 빈도 분석

문장 단위별 빈도 분석 결과는 표 2와 같다. 인터넷 채팅의 대화 내용에서 정상적인 문장은 6.1%만 구사되고 있다. 결국 전체 대화 내용의 93.9%는 긍정적이지 못한 문장을 구사하는 것으로 나타났다. 그 중 ‘미완결된 문장’이 37.2%로 가장 많은 것으로 나타났는데 이는 평소 입말에서도 나타나는 특성이므로 글말의 자연스러운 현상이라고 여겨진다. ‘상징기호를 이용한 문장’과 ‘흠소리나 닿소리만 적은 문장’은 각각 18.8%와 11.2%로 나타났다. 이 두 가지 현상은 통신 언어가 갖는 가장 대표적인 특성이다. ‘상대방이 불분명한 대화’가 11.3%를 차지하고 있다. 부정적·냉소적으로 응답한 문장’이 9.9%로 나타났다. 아동들이 부정적인 대화 분위기에 너무 오랫동안 노출되면 결국 정서 발달에 부정적인 영향을 미치게 되리라는 우려를 갖게 하는 결과이다.

<표 2> 문장단위별 사용 빈도

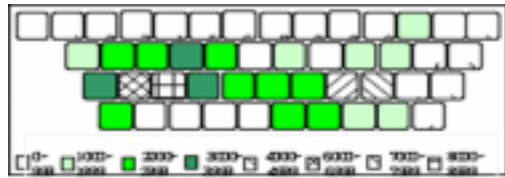
구분	기준	빈도	%
형식상 특성	미완결된 문장	3,803	37.2
내용상 특성	상징기호를 이용한 문장	1,927	18.8
	상대방이 불분명한 대화	1,157	11.3
형식상 특성	흠소리나 닿소리만 적은 문장	1,141	11.2
내용상 특성	부정적·냉소적으로 응답한 문장	1,011	9.9
기타	정상적인 문장	627	6.1
형식상 특성	조사를 생략하거나 오용한 문장	403	3.9
내용상 특성	같은 내용을 반복한 문장	120	1.2
형식상 특성	어순이 틀린 문장	38	0.4
	계	10,227	100

다. 자판기 배열에 따른 특성

<표 3> 글쇠 사용 빈도

글쇠	빈도	글쇠	빈도	글쇠	빈도	글쇠	빈도	글쇠	빈도
~	640	7	22	ㅉ	215	{	0	“	2
·	0	*	62	ㅊ	2378	[14	‘	297
!	1111	8	21	ㅌ	292	}	0	ㅋ	2602
1	258	(14	ㅍ	2069]	17	ㅌ	328
@	1	9	15	ㅍ	410	ㅍ	3193	ㅌ	831
2	84)	32	ㅍ	3652	ㅍ	6134	ㅌ	273
#	0	0	297	ㅉ	589	o	8570	ㅌ	500
3	61	_	603	ㅉ	2951	ㅌ	3905	ㅌ	2223
\$	0	-	1570	ㅌ	796	ㅎ	2700	-	2785
4	87	+	81	ㅋ	1236	ㅌ	2966	<	155
%	3	=	174	ㅌ	687	ㅌ	2985	,	1185
5	45		1	ㅎ	16	ㅌ	7501	>	152
^	213	\	3	ㅎ	1612		4932	.	1560
6	27	ㅉ	189	ㅋ	70	:	6	?	1026
&	6	ㅌ	1450	ㅋ	1486	;	742	/	227

채팅 언어에 사용된 글쇠의 빈도수를 2벌식 컴퓨터 자판기를 기준으로 조사하였으며, 기능키의 빈도수는 조사하지 않았다. 그 외에도 이모티콘에 사용되는 특수 기호나 숫자의 빈도수는 제외하였다. 채팅 언어를 자판을 기반으로 분석한 결과는 표 3, 표 4와 그림 1이다.



(그림 2) Shift 키를 사용하지 않은 글쇠의 사용빈도

그림 1에서는 높은 빈도를 차지하는 글쇠들이 대부분 세 개의 줄에 집중되어 있는 것을 볼 수 있다. 또한 자음의 빈도가 더 높다. 이는 한글의 구조상 당연한 결과라고 보인다. 자음 중에서는 ‘ㄴ’과 ‘ㅇ’ 글쇠가 가장 많이 사용하였고, 모음 중에서는 ‘ㅏ’와 ‘ㅣ’

글쇠를 가장 많이 사용하고 있는 것으로 나타났다.

<표 4> 손가락별 사용 빈도

손가락 (글쇠배당)	빈도	%	손가락 (글쇠배당)	빈도	%		
왼손	소지(8)	9444	11.3	오른손	소지(20)	6648	8.0
	약지(6)	9139	11.0		약지(7)	8301	10.0
	중지(6)	11823	14.2		중지(6)	9611	11.5
	검지(12)	15115	18.1		검지(10)	13259	15.9
	계	45,521	54.6		계	37,819	45.4

표 4에서 손가락을 비교해 보았을 때는 왼쪽손가락을 9.2% 더 많이 사용하고 있는 것으로 나타났다. 이것은 한글이 한 글자에 자음이 더 많이 포함되어 있기 때문이라고 여겨진다. 또, 손가락 중에는 검지를 가장 많이 사용하고 있다. 하지만 오른쪽 소지에는 왼쪽 소지에 비해 12개의 글쇠가 더 배당되어 있는데도 불구하고, 사용 빈도가 더 낮은 것으로 나타났다.

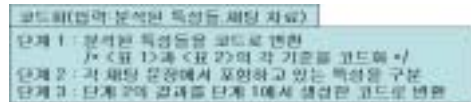
4.2 연관 규칙 탐사를 이용한 채팅 언어 분석

채팅 언어의 내용과 어휘들 간에 어떤 규칙들이 있는지 알기 위해 연관 규칙을 탐사를 하였다. 먼저 수집된 한글 채팅 자료를 데이터 처리를 했다. 그 결과 채팅 전체 데이터는 모두 6,499줄이고, 데이터의 크기는 783KB이다. 전체 데이터에서 아이디별로 채팅에 사용된 어휘와 문장을 분류한 아이디별 어휘와 문장의 집합 데이터는 248줄이었다.

분석 대상이 되는 데이터들은 세 가지로 구분하였다. 먼저, 채팅에 사용된 어휘들의 집합에 대한 연관 규칙탐사를 수행하였다. 이는 사용된 어휘들 간에 특성을 분석하기 위함이다. 다음으로 채팅에 사용된 문장들의 집합들에 대한 연관 규칙 탐사를 수행하였

다. 이는 채팅이란 상호간에 문장을 주고받음으로써 이루어지기 때문에 문장들 간에 특성을 분석하기 위함이다. 마지막으로 채팅 아이디별로 구분하여 특성을 분석하였다. 이는 특정 사용자의 어휘와 문장의 특성을 분석하기 위함이다.

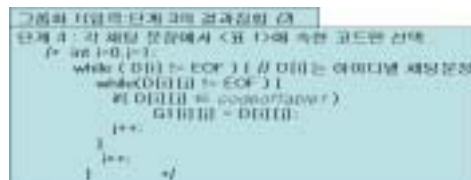
연관 규칙 탐사를 이용한 분석을 위해 다음과 같이 먼저 코드화 단계를 수행한다. 이는 자료 집합에 대한 연관 규칙을 생성하기 전에, 각 채팅 문장에 포함된 형식상의 특성과 내용상의 특성으로 구분하고, 이를 <표 1>과 <표 2>의 기준에 근거하여 코드화한다.



코드화된 결과의해 생성된 채팅 문장의 자료구조는 다음과 같다고 가정한다.

아이디1	특성 코드	...	특성 코드
아이디2	특성 코드	...	특성 코드
...			
아이디n	특성 코드	...	특성 코드

이러한 결과에 대해, 어휘들에 대한, 문장들에 대한, 그리고 아이디별 연관 규칙 탐사를 수행하기 위해 자료를 그룹화 한다.



부정적 또는 냉소적 문장이기 쉽고 이는 미완결 문장을 생성한다.

<표 6> 문장 집합에 대한 규칙

순	규칙	신뢰도(%)	지지도(%)
1	6 <- 9	29.6	70.6
2	6 <- 4	17.8	76.4
3	6 <- 5	17.6	35.6
4	6 <- 2	15.6	63.5

다. 아이디별 어휘 및 문장에 대한 연관 규칙 탐사

각 아이디별 어휘 및 문장에 대한 연관 규칙 탐사 결과는 다음 표와 같다. 다음 각 규칙은 앞선 연관 규칙 결과와 유사한 결과를 생성하였다. 당연한 규칙으로 홀소리 닿소리만 적은 문장(5)인 경우 홀소리 닿소리만 적은 어휘(24)만을 사용한다. 그러나, 이러한 규칙들의 공통적 특성은 정상적인 어휘, 띄어쓰기 오류, 또는 상대방이 불분명한 문장을 사용하더라도 대다수 채팅언어는 홀소리 닿소리 어휘(24)를 사용한 미완결 문장(6)임을 알 수 있다.

<표 7> 아이디별 어휘와 문장에 대한 규칙

순	규칙	신뢰도(%)	지지도(%)
1	24 <- 5	66.1	95.7
2	24 <- 6, 5	54.4	95.6
3	6 <- 21, 4	48.4	95.8
4	6 <- 21, 2	42.3	96.2
5	6 <- 4, 11	41.1	97.1
6	24 <- 10, 5	49.6	95.1
7	24 <- 21, 5	44.8	95.5
8	24 <- 5, 11	40.7	95.0
9	24 <- 6, 10, 5	45.6	95.6
10	6 <- 10, 21, 4	40.7	96.0

5. 결론 및 개선방안

5.1 결론

본 연구는 통신상에서 급속도로 퍼져가고 있는 통신 언어의 특성을 분석하였다. 통신 언어 중 채팅 언어를 선택하여 분석하였는데 그 이유는 채팅 언어가 통신 언어의 특성을 가장 대표할 만하다고 여겨졌기 때문이다. 연구 대상은 아직까지 연구가 활발하지 않다고 여겨지는 초등학생을 대상으로 분석 하였다. 분석 내용은 기존 연구의 분석 기준을 비교하여 형식상 특성과 내용상 특성, 그리고 통계상의 특성으로 구분하였다. 첫째, 형식상 특성은 채팅 언어의 외형적, 구조적으로 나타나는 특성이다. 대표적인 특성으로는 소리 나는 대로 적기, 음운 탈락과 첨가, 음절 생략과 첨가, 바꿈, 띄어쓰기 오류, 홀소리만 적기와 닿소리만 적기, 유야무야음으로 변형되어 쓰기, 음절의 앞 뒤 바꾸어 쓰기, 문장의 미완결성, 조사의 생략과 오용, 어순의 혼란 등으로 분석하였다. 둘째, 내용상 특성은 채팅 언어의 의미적 특성이다. 대표적인 특성으로는 부호를 이용한 표현, 은어와 신어, 비속어와 욕설, 방언, 외국어, 부정적·냉소적으로 응답하는 현상, 부름말과 존댓말, 같은 내용을 반복하기, 상대방이 불분명한 대화 등으로 분석하였다. 셋째, 통계상 특성은 크게 대화의 문장 빈도와 어휘의 빈도, 그리고 자판기 배열에 따라 나타나는 특성, 연관규칙 탐사를 통해 나타나는 특성으로 구분하여 분석하였다. 문장 빈도 분석 결과 ‘미완결된 문장’을 가장 많이 사용하였다. 어휘의 빈도는 ‘상징 기호를 이용한 어휘’를 가장 많이 사용하였다. 자판기 배열에 따라 나타나는 특성은 ‘ㄴ’과 ‘ㅇ’ 그리고 ‘ㅏ’와 ‘ㅣ’글쇠를 가장 많이 사용하고 있는 것으로 나타났다. 손가락을 비교해 보았을 때는 왼쪽손가락을 더 많이 사용하고 있는 것으로 나타났다. 연관 규칙 탐사를 통해 나타난 특성은, 글말이나 입말과 구별되는 특성으로, ‘상징 기호를 이용한 문장’과 ‘상대방이 불분명한 대화’, ‘홀소리나 닿소리만 적은 내

용’, ‘부정적·냉소적으로 응답한 내용’의 경우는 대부분이 ‘미완결 문장’을 구사하고 있는 것으로 나타났다. 또 ‘상징 기호를 이용한 어휘’는 대부분 띄어쓰기가 무시되고 있다.

본 연구의 특성과 기존 연구의 특성을 비교하면 표 8과 같다. 표 8과 더불어 본 연구에서는 기존 연구에서 분석하지 못한 새로운 특성을 발견하였다. 즉, 분석을 통해 유아 발음으로 변형되어 쓰기, 음절의 앞 뒤 바꾸어 쓰기, 상대방이 불분명한 대화의 특성을 발견할 수 있었다. 이는 기존 연구의 대상이 성인이거나 청소년들이었기 때문에 발견할 수 없었지만, 본 분석의 대상이 초등학생이기 때문에 발견한 특성이다.

<표 8> 연구별 특성 비교

	형식상 특성	내용상의 특성		통계상의 특성
		한 문장에서 특성	문장간의 특성	
임규홍[3]	○	○	×	×
백경녀[4]	○	○	○	×
도호근[5]	○	○	×	×
한세규[6]	○	○	×	×
정정덕[7]	○	○	×	×
본 연구	○	○	○	○

5.2 개선방안

본 연구에서 수행한 채팅 언어의 분석 결과를 통해 언어의 오용을 최소화하고, 아동의 정서발달에 긍정적인 채팅 문화를 정착시키기 위한 다음과 같은 컴퓨터 시스템적 방안을 제안하고자 한다.

첫째, 문장의 종결어미를 자동으로 완성시키는 기능이 필요하다. 즉, ‘엔터’ 글쇠를 누르면 문장의 종결어미를 자동으로 완성시키는 프로그램이 필요하다. 통계적 특성에서 ‘부정적·냉소적으로 응답한 문장’은 대부분 ‘미완결 문장’을 구사하고 있는 것으로 나타났다. 예를 들면 채팅 대화창에서 ‘안녕’이라

고 타이핑한 후 엔터 글쇠를 누르게 되면 ‘안녕하세요?’라고 완결된 문장이 대화글에 올라가게 한다. 결국 통계적 특성의 결과에 비추어 볼 때 미완결 문장 줄이기는 부정적·냉소적으로 응답한 문장과 상대방이 불분명한 대화를 줄이는 효과를 가져 올 수 있게 될 것이다. 둘째, 띄어쓰기 오류를 자동 수정 및 보완할 수 있는 기능이 필요하다. 즉, ‘맞춤법 오류 자동 수정 프로그램’이 필요하다. 현재 개발된 한글프로그램들의 경우에는 띄어쓰기가 틀렸거나 오타가 쓰였을 경우에는 틀렸다는 표시가 나타나게 할 수 있다. 여기서 좀 더 발전시켜 채팅 대화창에서 최소한 어휘의 오류를 수정해 주거나 정확한 표현을 실시간으로 제시해줄 수 있는 프로그램이 필요하다. 어렸을 때의 언어 습관은 쉽게 고쳐지지 않는다고 한다. 올바른 용어의 용례를 제시함으로써 올바른 언어 사용을 지향해야 한다. 정상적인 어휘가 39.8%이고, 정상적인 문장이 6.1%라는 통계상의 특성을 비추어 볼 때 적극적인 노력이 필요한 부분이다. 셋째, 철자의 오류를 줄이기 위해 자판기에서 ‘shift’ 글쇠 위치를 이동할 필요가 있다. 영문에 비해 한글은 구조상 ‘shift’ 글쇠를 많이 사용할 수밖에 없게 되어있다. 본 분석의 결과에서 왼손의 소지가 많이 사용되고 있는 결과를 볼 수 있었다. 하지만 소지는 다른 손가락에 비해 손가락의 길이도 짧고 ‘shift’글쇠 주변에 기능키들이 많아 자칫 오타를 많이 경험했을 것이다. 본 분석에서도 □□왔네’의 경우는 □□왔네’를 쓰는 과정에서 사용자가 번거로움을 피하기 위해서 아니면 오타로 인해서 발생한 경우라고 여겨진다. 그래서 ‘shift’ 글쇠를 ‘Caps Lock’과 자리를 바꿀 것을 제안한다. 그 이유는 소지를 구부리지 않고 타이핑 할 수 있는 위치이기 때문이다. 또는 맨 아래줄의 기능키 중 자주 쓰이지 않는 글쇠를 빼고 ‘Enter’ 글쇠처럼 크기를 늘려준다. 소지 사

용의 편리성을 좀 더 증대시키기 위해서이다. 넷째, 익명성의 최소화를 위해 대화자를 간편하게 지정할 수 있는 기능이 필요하다. 즉, 대화명 리스트를 제공하여 기능키로 간단히 사용할 수 있게 한다. 이것은 복잡하게 보이는 대화명을 간단하게 기능키로 대신하여 사용하게 한다. 그 결과 익명성의 최소화를 통해 상대방이 불분명한 대화를 줄이게 되고 연관규칙에서 발견한 결과로 미완결된 문장을 줄이게 된다. 이는 또 문장의 완성도를 높이게 되며 부정적·냉소적인 대화를 줄이는 효과를 가져 올 수 있다.

참고문헌

- [1] 서미정,(2002), 인터넷을 통한 청소년의 채팅 실태 조사 연구, 강원대학교대학원 석사학위논문
- [2] 권연진,(2000). 컴퓨터 통신 언어의 유형별 실태 및 바람직한 방안, 언어과학학술지, pp.2-16.
- [3] 임규홍,(2000). 컴퓨터 통신 언어에 대하여, 배달말학회, pp. 33-56.
- [4] 백경녀,(2001). 청소년의 언어사용 실태와 개선방안 연구, 카톨릭 대학교 교육대학원 석학위논문.
- [5] 도효근,(2001). 통신 언어가 국어 생활에 미치는 역기능 연구, 어문연구학회논문, pp. 31-53.
- [6] 한세규,(2002), 전자게시판의 언어 연구, 경기대학교 교육대학원 석사학위논문.
- [7] 정정덕,(2000), 요즘 고교생들의 사이버 어휘 특징, 창원대학교 인문과학연구소학술지 7권 1호, pp. 1-8.
- [8] 장남석·홍성완·장재호, (1999), 데이터마이닝, 대청출판사.
- [9] 이정원 외 6인,(2001). 데이터 마이닝 알고리즘의 분류 및 분석, 한국정보학회 논문지, pp. 279-300.
- [10] 강현철 외 4인,(2001), 데이터 마이닝(방법론 및 활용), 자유아카데미.
- [11] W. Lee, S.J. Stolfo and K.W. Mok,(2002). Algorithms for Mining system

audit data, Data Mining, Rough Sets and Granular Computing, T.Y. Lin, Y.Y. Yao and L.A. Zed도(eds), Physica-Verlag.

- [12] B. Lets, A. Swami and J. Widom, (1997), Clustering Association Rules. In Proc. of the 13th International Conference on Data Engineering.
- [13] N. Pasquier, Y. Bastide, R. Taouil and L. Lakhal, (1999). Closed Set based Discovery of samll covers for Association Rules, In Proc. 15emes Journees Bases de Donnees Avancees, pp.361-381.

저자 소개

정희택



1992년: 전남대학교 전산통계학과
1995년: 전남대학교 전산통계학과 석사
1999년: 전남대학교 전산통계학과 박사
1999년 - 현재 : 여수대학교 조교수

관심분야 : 분산처리시스템, 컴퓨터교육시스템, 워크플루우시스템, 데이터탐사, 바이오정보학
E-mail : htceong@yosu.ac.kr

정영순



1988년: 광주교육대학교 졸업
2004년: 여수대학교 컴퓨터교육 석사
1988년 - 현재 : 여수부영초등학교 교사
관심분야 : 컴퓨터교육, 데이터탐사, 교육소

소프트웨어 개발

E-mail : haha56@hanmail.net