

텍스트마이닝을 위한 패션 속성 분류체계 및 말뭉치 웹사전 구축

장세윤 · 김하연*† · 김송미** · 최우진** · 정진** · 이유리***
서울대학교 생활과학연구소, *군산대학교 의류학과, **서울대학교 의류학과,
***서울대학교 의류학과/서울대학교 생활과학연구소

Development of Online Fashion Thesaurus and Taxonomy for Text Mining

Seyoon Jang · Ha Youn Kim*† · Songmee Kim** ·
Woojin Choi** · Jin Jeong** · Yuri Lee***

Research Institute of Human Ecology, Seoul National University
*Dept. of Clothing & Textiles, Kunsan National University
**Dept. of Textiles, Merchandising, and Fashion Design, Seoul National University
***Dept. of Textiles, Merchandising, and Fashion Design, Seoul National University/
Research Institute of Human Ecology, Seoul National University
Received September 29, 2022; Revised October 26, 2022; Accepted November 8, 2022

Abstract

Text data plays a significant role in understanding and analyzing trends in consumer, business, and social sectors. For text analysis, there must be a corpus that reflects specific domain knowledge. However, in the field of fashion, the professional corpus is insufficient. This study aims to develop a taxonomy and thesaurus that considers the specialty of fashion products. To this end, about 100,000 fashion vocabulary terms were collected by crawling text data from WSGN, Pantone, and online platforms; text subsequently was extracted through pre-processing with Python. The taxonomy was composed of items, silhouettes, details, styles, colors, textiles, and patterns/prints, which are seven attributes of clothes. The corpus was completed through processing synonyms of terms from fashion books such as dictionaries. Finally, 10,294 vocabulary words, including 1,956 standard Korean words, were classified in the taxonomy. All data was then developed into a web dictionary system. Quantitative and qualitative performance tests of the results were conducted through expert reviews. The performance of the thesaurus also was verified by comparing the results of text mining analysis through the previously developed corpus. This study contributes to achieving a text data standard and enables meaningful results of text mining analysis in the fashion field.

Key words: Text mining, Taxonomy, Corpus, Thesaurus, Big data; 텍스트마이닝, 분류체계, 말뭉치, 동의어 사전, 빅데이터

I. 서론

마케팅에서 상품 믹스는 상품을 너비와 길이로 구

†Corresponding author

E-mail: hykim@kunsan.ac.kr

분하여 체계적으로 분류하는 방법이다(Kotler & Armstrong, 2015). 패션 상품의 디자인은 소재, 컬러, 실루엣, 디테일과 같은 의복의 기본 요소들의 창의적인 조합으로 이루어지기 때문에(Lamb & Kallal, 1992; Littrell & Miller 2001; Shin & Kim, 2013), 패션 상품은 디

자인 요소들을 활용하여 체계적 분류가 가능하다. 하지만 이에 대한 명칭은 언어가 가지는 역사성과 사회성, 자의성에 따라 사용되는 시대나 사용 주체에 따라 달라지기 마련이다. 따라서 패션 상품과 관련 어휘를 체계적으로 수집하고 분류하여 분석하는 일련의 과정은 거시적으로는 특정 시대의 사조를 파악하거나 트렌드의 추이를 살펴볼 수 있으며, 미시적으로는 자사의 상품을 관리하거나 인기도를 예측하는 척도로도 활용될 수 있을 것이다.

특히 오늘날 디지털 기술의 발전과 클라우드 서비스의 보급으로 데이터의 수집에서 전송, 저장, 분석 및 폐기에 이르는 데이터 처리 비용이 점점 감소함에 따라 다양한 원천에서 데이터를 수집하고 가공하여 활용하여 가치를 발생시키는 활동은 증가하고 있다. 특히 최근 생성되는 데이터의 80~95%는 텍스트를 비롯한 비정형 데이터로 알려졌다(Gandomi & Haider, 2015). 게다가 소비시장에서 소비자-기업-사회 간 상호작용의 대부분은 텍스트로 이루어지기 때문에 텍스트 데이터는 분석대상으로 중요성을 지닌다. 이러한 텍스트 데이터의 가치가 상승함에 따라 이를 효율적으로 활용하는 빅데이터 중심의 방법론도 주목을 받고 있다. 예를 들어 텍스트마이닝(text mining)은 방대한 비정형 데이터 세트 속에서 의미 있는 정보를 추출하는 것으로 답문을 정량화하여 중요한 어휘를 도출하고 의미를 객관화할 수 있다는 특징이 있다(Kim & Lee, 2018). 따라서 정보의 원천에서 최대한 어휘의 손실 없이 분석하기 위해, 구조화된 텍스트 데이터 세트인 말뭉치를 적용하는 것이 필수적이다.

말뭉치 또는 코퍼스(corpus)는 자연어가 들어 있는 대량의 어휘 집합으로(Bengfort et al., 2018), 이러한 말뭉치는 계층적으로 구조화될 수 있다(Garshol, 2004). 구체적으로 분류의 레벨(level)이 깊어짐에 따라 상위 개념에서부터 보다 구체적인 하위 개념으로 분화될 수 있다. 만약 A의 모든 어휘들이 B의 어휘가 되기도 한다면, A는 B의 하위레벨이라고 표현할 수 있다. 이처럼 말뭉치의 계층적이며 구조적인 형태를 분류체계(taxonomy)라 한다. 말뭉치와 분류체계는 해당 산업에서 합의된 지식이어야 하며 어느 개인에게 국한되는 것이 아니라 그룹 구성원이 모두 동의하는 개념이다. 따라서 이들은 상세하게 기술하여 사람 간 또는 소프트웨어 간에 동일하게 이해될 수 있도록 구축되어야 한다. 특히 온라인 자원이 증가함에 따라 다양한 원천에

서 이종의 데이터를 결합하여 고차원적인 분석을 하기 위해서는 체계적이고 전문적인 분류체계와 말뭉치의 적용이 필수적이다.

현재 의류·패션 분야에서 한국어 기반 텍스트마이닝 연구동향을 살펴보면, 전처리 과정에서 한국어 기반 분석엔진인 ‘MeCab-kr(Kim & Lee, 2022)’이나 ‘espresso K(Song & Lim, 2021)’, ‘KoNLP(Jang & Kim, 2020)’를 주로 적용하고 있다. 하지만 도메인과 어휘들의 구조적 형태를 특정하지 않는 채 일반적으로 사용되는 어휘를 중심으로 정제하는 경우, 같은 개념에 대해 서로 다른 어휘를 사용하는 이음동의어(synonym)(e.g., 스웨트셔츠, 맨투맨, MTM)나 동음이의어(homonym)(e.g., 옷본을 의미하는 패턴과 무늬를 의미하는 패턴)를 고려하지 못하여 어휘의 의미를 정확히 추출하는 데 어려움이 있을 수 있다.

이러한 점을 보완하고자 비교적 최근에는 패션 도메인에서도 패션 상품에 대한 특성과 지식을 반영하여 패션 상품에 전문성을 가진 분류체계가 개발되고 있다. 일례로 Guan et al.(2018, 2019)은 의복의 속성이 시각적으로 명백히 드러나서 데이터로 기록할 수 있는 명시적 요소와 함께 관찰자가 의복에 대한 연상 작용으로 발생하지만, 데이터로 직접 기록하기가 어려운 스타일, 선호도, 바디이미지 등을 포함하는 암시적 요소로 나누어 분류를 제안하였다. 또한 Jang et al.(2022)의 연구에서는 패션 상품의 중요 속성이지만 주관적이라 분류나 표기가 어려운 브랜드 감성에 대한 분류를 위해, 특히 그 감성이 디자이너나 소비자 와도 잘 소통되고 있다고 알려진 20가지 하이엔드 패션 브랜드의 이미지를 딥러닝으로 학습하여 브랜드 감성을 자동으로 분류하기도 하였다. 하지만 이상의 연구들도 다음의 문제를 해결하기에는 여전히 어려움이 있다. 첫째, 패션은 융합 학문이자 산업 내 다양한 이해관계자들이 포함되어 있어 사용하는 주체에 따라 하나의 의미가 여러 가지 어휘로 표현될 수 있다. 따라서 패션의 대표 어휘와 동의어를 포함하는 말뭉치 사전, 즉, 시소러스(thesaurus)가 필요하다. 텍스트마이닝은 텍스트 파싱(parsing), 필터링(filtering), 클러스터링(clustering) 등의 과정에서 주로 텍스트 빈도수에 기반을 두어 동작한다. 그러나 컴퓨터는 맥락 속 어휘의 의미론적 파악, 즉 시맨틱(semantic)을 파악하지 못하므로, 동음이의어와 이음동의어를 서로 다르게 처리하는 문제를 야기할 수 있기 때문이다. 둘째, 패션 어휘는 트렌드에

민감하여 신조어가 자주 등장하고, 외래어 표기도 많으므로, 이러한 신조어를 사전에 반영할 필요가 있다. 반면, 표준어 사전은 표준 단일어를 기반으로 기술됨에 따라 인간을 둘러싼 사회, 문화, 기술 등 환경의 변화에 따른 신조어가 배제되기 쉽다. 신조어를 반영하는 것은 텍스트마이닝의 결과를 질적으로 높여줄 수 있을 것이다(Choi, 2014). 셋째, 패션에 특화된 말뭉치가 컴퓨터 기반으로 구축되지 않아 텍스트마이닝에 바로 적용하기가 어렵다. 텍스트마이닝은 컴퓨터 언어학의 일종으로 대량으로 축적된 인간의 언어를 컴퓨터가 분류하고 유용한 정보를 추출하기 위해 컴퓨팅 자원을 활용한다. 따라서 텍스트를 식별하고 의미 있는 단위로 분해하는 데이터 전처리 과정을 지원하는 한국어 말뭉치 프로그램 형식으로 개발되어 있다. 그러나 패션에 특화된 말뭉치는 웹페이지나 프로그램의 형태로 공개되지 않아서 텍스트마이닝에 바로 사용되기 어렵다.

반면 선형연구(Jang et al., 2022; Kim & Lee, 2022; Park & Choi, 2020a, 2020b, 2021)에서는 패션 산업에서 빅데이터와 인공지능을 활용한 비즈니스를 위해서 데이터 관리를 위한 데이터 표준화가 가장 선행되어야 함을 주장하고 있다. 정부 차원에서도 이러한 표준화된 데이터의 중요성을 높게 인식하고 있다. 이에 따라 2017년부터 지능정보산업 인프라 조성사업에서 국가 주도로 학습용 데이터를 구축하고, 2020년에는 통합 플랫폼인 AI 허브(www.aihub.or.kr)를 통해 수요자에게 데이터 세트를 개방하고 있다. 이 중 한국어 음성 및 자연어 말뭉치는 44개가 포함되어 있으나(Ministry of Science and ICT, 2022), 의류·패션 분야의 전문 말뭉치는 현재까지 보고된 바가 없다. 따라서 본 연구에서는 패션 빅데이터 분석의 정확도를 높이고 패션 데이터의 표준화를 이루기 위해 패션 상품의 특수성을 고려한 분류체계 및 말뭉치를 개발하는데 목적이 있다. 특히 한국어 기반 이음동의어 문제를 최대한 해결함으로써 텍스트마이닝의 성능을 높이고자 한다.

II. 관련 연구

1. 패션 상품의 속성 분류

패션 도메인에서는 온라인 쇼핑과 데이터마이닝 분야에 적용하기 위해 패션 상품의 관점에서 분류체계

를 적용해오고 있다. 인간의 지식을 컴퓨터 시스템이 이해할 수 있도록 가공하는 방법인 지식표현 프레임워크에 따르면, 어휘들의 성질과 특성을 대표하는 특정 속성(attribute)은 이에 대응하는 속성 값(value)으로 표현할 수 있다(Lee, 2018). 따라서 컴퓨터 시스템에서는 특정 객체의 속성을 기술하는 하나의 규칙으로 ‘속성 이름:속성 값’에 대응하는 형식을 취하도록 정하고 있다. 예를 들어 ‘빨간 원피스’는 ‘아이템:원피스’이면서 ‘컬러:빨간색’으로 특성을 기술할 수 있다.

패션 상품의 특성을 규정하는 속성은 아이템뿐 아니라 스타일, 실루엣, 디테일, 컬러, 소재, 패턴과 프린트와 같은 디자인 요소가 거론된다(Littrell & Miller 2001; Shin & Kim, 2013). 이들은 크게 형태적 분류와 의미적 분류로 구분할 수 있다. 형태적 분류는 가시적이며 객관적인 디자인 속성에 따른 분류를 의미한다. 구체적으로 소재, 컬러, 디테일(e.g., 칼라(collar), 여밈방식, 단추의 수) 등과 같은 디자인 요소들이 해당한다. 이러한 속성의 조합은 옷의 형태, 즉, 실루엣을 결정한다. 반면 이러한 패션 속성들의 조합은 착용자와 관찰자로 하여금 특정한 감성이나 스타일, 이미지 등을 연상시킨다. 따라서 이는 의사소통의 수단이 되는 의미적 분류에 해당한다(Guan et al., 2018, 2019). 데이터 표준화를 위해 패션 상품의 분류체계 시스템을 구축한 Park and Choi(2020a, 2020b, 2021)의 일련의 연구들에서는 아이템, 컬러, 소재감, 감성과 같은 속성들이 G(general)레벨과 S(specific)레벨의 계층으로 구성된다. G레벨은 형태, 길이, 여밈 등과 같이 의복을 3차원으로 형성하는 구조적 요소로 보통 하나의 값만 가질 수 있다. 반면 S레벨은 의복 표면의 시각적이고 장식적인 요소들이 해당된다. 예를 들어 아이템의 라운드 넥(neck)은 G레벨이면서, S레벨로 유(U), 보트(boat), 카울(cowl), 크루(crew) 등을 포함한다.

패션 상품의 속성은 소비자 혹은 컴퓨터 비전(computer vision) 분야에서도 의복 디자인을 평가하거나 특징을 인지하기 위한 기준이 될 수 있다. 컴퓨터 공학 분야에서는 컴퓨터 비전의 지도학습을 위해 패션 이미지를 중심으로 데이터 세트를 구축해 오고 있는데(Kiapour et al., 2014; Liu et al., 2016; Wu et al., 2021; Xiao et al., 2017; Zheng et al., 2018; Zou et al., 2019), 분류체계는 이미지 라벨링의 기준이 된다. 예를 들어 DeepFashion(Liu et al., 2016) 데이터 세트는 289,222장의 패션 사진에 50개의 아이템 속성 값 중 하나와 소

재, 패턴과 프린트, 길이, 디테일, 스타일 등의 패션 속성에 해당하는 1,000여 개 속성 값이 함께 라벨링되어 있다. DeepFashion을 포함한 대부분의 패션 데이터 세트들은 주로 아이템을 중심으로 분류되어 있으나, 티셔츠, 바지, 스커트, 신발, 가방 등과 같이 포괄적인 수준의 범주로 이루어져(Mallavarapu et al., 2021), 고차원적인 분석에 한계가 있다. 또한 선행 데이터 세트는 패션 데이터를 주로 온라인 쇼핑몰에서 수집하다 보니, 상품 속성 중 상거래에서 관리의 목적으로 사용되는 객관적인 요소인 아이템, 소재, 컬러, 디테일에 치중되어 분류되고 있다(Sadeh et al., 2019). 이는 형태적 분류에 해당하는 속성들은 비교적 객관적이라 논란의 여지없이 쉽게 수집하고 컴퓨터 비전으로 지도 학습이 가능하기 때문이다. 하지만 패션 상품의 객관적 속성이 같다고 하더라도, 패션 이미지에 담긴 기호나 상징에 따라 해석이 달라질 수 있기 때문에(Jang et al., 2022), 의미적 요소에 대한 학습과 분류도 시도되고 있다. 일례로 Jang et al.(2022)은 패션의 의미적 요소로 패션 브랜드 감성을 11가지로 분류하였고, 이미지 딥러닝을 통해 패션 사진에서 패션 브랜드 감성을 비교적 일관적으로 측정하고 분류했다.

2. 텍스트마이닝과 패션 말뭉치

텍스트마이닝은 잡음(noise) 속에 신호의 유무를 파악하고, 그러한 신호가 어디에 있는지, 어떤 속성을 지녔는지 식별해 내는 것이다. 다시 말해 텍스트마이닝이란 독특한 어휘, 일반적인 구 등을 성분별로 분해한 후에 통계적 원리를 적용하여, 토픽 모델링(topic modelling)이나 문서 군집화(document clustering)를 시행하는 과정이다(Bengfort et al., 2018). 따라서 텍스트마이닝은 말뭉치라 불리는 텍스트 데이터 세트, 즉, 말뭉치 또는 온톨로지의 구축과 활용이 필수적이다.

말뭉치는 어휘집, 의미망(semantic network), 온톨로지(ontology), 시소러스 등으로 불리기도 하지만, 사용자에게 따라 표현의 차이가 있어 명확하게 정의를 내리기 쉽지 않다(Garshol, 2004). 따라서 본 연구에서 말뭉치는 자연어가 들어 있는 대량의 구조화된 어휘 집합으로 정의하였다. 말뭉치는 특정 분야나 영역(domain)에서 훈련된 모델이 있다. 예를 들어 코넬 영화대사 말뭉치(Danescu-Niculescu-Mizil & Lee, 2011)에는 장르, 연도, 인터넷 영화 데이터베이스(internet movie data-

base: IMDb)의 평점, 추천 수, 주인공의 성별 등의 속성 정보가 함께 담겨있다. 이와 같은 특정 분야의 말뭉치는 일반 말뭉치보다 분석력과 모델링이 더 우수하다고 알려져 있는데, 이는 분야마다 서로 다른 언어를 사용하기 때문이다(Bengfort et al., 2018). 예를 들어 의류학에서 ‘스타킹(stocking)’이라는 어휘는 무릎 위로 올라오는 긴 양말류를 지칭하는 어휘로 사용되지만, 생물학에서는 사과나무의 한 종류로 여겨진다. 또한 방송 콘텐츠 분야에서는 스타킹을 2006~2017년 기간 동안 한국에서 방영된 예능 프로그램인 ‘스타킹(Star-king)’으로 연상할 수도 있다.

어휘의 집합인 온톨로지 역시 특정 도메인과 관련된 형식적이고 명시적이며 공유된 개념이다(Gruber, 1995). 하지만 온톨로지는 어휘 간의 관계를 함께 설정(e.g., 상속관계를 나타내는 has-a, 반대 개념을 나타내는 is-contrast-to, 관련성을 나타내는 is-relation-to, 재료를 나타내는 material-of 등)하기 때문에 어휘가 여러 관계에 속할 수 있다는 점에서 말뭉치와 차이가 있다(Park, 2012). 예를 들어 Noy and McGuinness(2001)의 와인 온톨로지에서 포트와인(port wine)은 와인 컬러에 따라 레드 와인이면서 당분이 스위트(sweet) 값인 디저트 와인 값을 가진다. 따라서 포트 레벨의 하위 와인은 레드 와인과 디저트 와인 레벨에 모두 속할 수 있다. 현재 다양한 분야에서 정보를 공유하고, 정보에 대한 주석을 넣기 위해 표준 온톨로지를 개발하고 있다. 예를 들어 의료 분야에서는 국제 표준의료용어체계(systematized nomenclature of medicine clinical terms: SNOMED CT), 기업의 비즈니스용으로 Enterprise Ontology(Uschold et al., 1998) 등이 있다. 패션 분야의 선행 온톨로지로는 SERVIVE Fashion Ontology(Vogiatzis et al., 2012), Indian Garment Ontology(Ajmani et al., 2013), Bollacker et al.(2016) 등이 있다. 그러나 이 역시 패션 도메인의 모든 속성을 포괄하지 못하였고, 한국어를 포함한 동의어가 반영되지 않아 텍스트마이닝에 한계가 있다.

이상에서와 같이 온톨로지는 어휘 간의 관계 추론의 성격이 강하기 때문에 어휘 출현 빈도를 중심으로 하는 텍스트마이닝 분석에는 말뭉치 데이터 세트가 더 유용하게 사용될 수 있을 것이다. 따라서 본 연구는 텍스트로 된 빅데이터에서 깊은 의미를 찾을 수 있는 방법론인 텍스트마이닝을 보다 효과적으로 수행하기 위해 어휘 간의 관계를 찾기보다는 패션 제품의 속성

을 체계적으로 분류하고 대표 어휘를 기준으로 이음동의어를 확장하는 패션 전문 말뭉치를 구축하고자 하였다.

III. 패션 속성의 분류체계와 말뭉치 사전 구축 절차

본 연구는 패션 도메인에서 텍스트를 효과적으로 분석하고 데이터 표준화를 제고하는 패션 분류체계 및 말뭉치를 개발하기 위해 수행되었다. 이를 위해 하이브리드(hybrid) 구축 방법론을 적용하였는데, 이는 상향식(bottom-up)과 하향식(top-down)을 결합하여 구축하는 방법이다. 구체적으로 다양한 정보 원천에서 광범위한 패션 트렌드와 다양한 복종, 신조어 및 동의어의어, 이음동의어를 아우르는 한국어와 영단어를 수집한 뒤 상위 개념을 구축해가는 상향식 분류를 시행하였다. 또한 패션의 주요 속성을 중심으로 개념의 층위를 두어 분류체계를 구성하는 하향식 구축 방법론을 병행하였다. 선행연구에 따르면 말뭉치 사건의 구축을 위한 설계 방법론은 명확히 존재하지 않는다(Noy & McGuinness, 2001). 온톨로지를 설계하는 방법론을 제안한 Noy and McGuinness(2001)는 텍스트 데이터 세트를 구축하는 데 연구자들의 경험이 유용하다고 하였다. 이에 패션 산업에서 경험이 풍부한 패션 머천다이징/리테일 전공 박사 과정 이상 연구자 6인은 구축 과정에서 여러 차례 의견을 수렴하였다. 그뿐만

아니라 분류의 타당성을 검증하기 위해 패션 이미지를 데이터화하는 비즈니스 모델을 가진 기업에서 패션 데이터를 설계하는 핵심 실무자 2인과 지속해서 의견을 교환하는 과정을 거쳤다. 그 결과 1차로 완성된 분류체계와 말뭉치를 웹용 패션 사전 시스템으로 구축하였다. 이후 패션 사전의 효용성을 정량적, 정성적으로 증명하기 위해 전문가 10인을 대상으로 휴리스틱 평가와 함께 일반적인 한국어 분석엔진과의 성능을 비교 분석하는 과정을 수행하였다. 구체적인 구축 절차는 <Fig. 1>과 같다.

1. 1단계: 텍스트 데이터 수집

패션 도메인의 속성을 선정하고, 상위 레벨과 이후 하위 레벨의 분류체계를 규정하기 위해 우선 하향식 접근법을 적용하여 <Table 1>의 패션 전문 사전, 선행 패션 데이터 세트, 전문 서적에서 의복 및 소재와 관련한 분류체계와 어휘를 참고하였다. 또한 실생활에서 주로 사용되거나 최근 사용되는 어휘를 주로 수집하기 위해 상향식 접근법으로 <Table 1>의 웹페이지(HTML)와 같이 글로벌한 패션 트렌드 발표 기관인 WGSN과 컬러 커뮤니케이션에 주로 사용되고 있는 팬톤(Pantone)사에서 사용하고 있는 분류체계와 어휘를 함께 수집하였다. 또한 이음동의어와 영어 어휘의 추가 수집을 위해 2021년 7월 중 국내 대표 쇼핑몰이라고 판단한 더블유컨셉 및 무신사와 해외 쇼핑몰 중

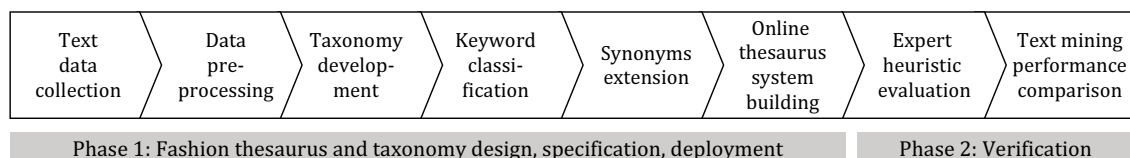


Fig. 1. Fashion thesaurus and taxonomy development process.

Table 1. Sources of bottom-up data collection

Category	Sources
Textbook	Textile Science: Fiber Knowledge Essential to Merchandiser (Ahn, 2016); Lifestyle & Trend (Lee & Park, 2006); Pattern Fashion (Ryu, 2010); Swatch Reference Guide for Fashion Fabrics (Young, 2018)
Dictionary	Fashionpedia (Fashionary, 2021), The Dictionary of Fashion (Tortora & Keiser, 2014)
Literature	Han et al., 2016; Kim & Yoon, 2013; Lee, 2009; Park & Choi, 2020a, 2020b, 2021; Tortora & Keiser, 2014
Dataset	Kiapour et al., 2014; Liu et al., 2016; Wu et al., 2021; Xiao et al., 2017; Zheng et al., 2018; Zou et al., 2019
HTML	Musinsa web store (www.musinsa.com); Pantone website (www.pantone.com); Wconcept web store (www.wconcept.co.kr); WGSN website (www.wgsn.com); Yoox web store (www.yoox.com)

육스닷컴에서 전체 상품의 상품명을 웹스크래퍼로 수집하여 약 10만 개의 어휘를 얻었다.

2. 2단계: 텍스트 데이터 전처리

파이썬의 텍스트 데이터 전처리 코드인 split 함수를 실행하여 불용어와 기호를 제거하고, 의미가 있는 가장 작은 단위로 토큰나이징(tokenizing)하여 유의미한 어휘를 추출하였다. 숫자의 경우 ‘60년대 패션, 3 buttons, 2-way item’처럼 문자와 결합하여 특정한 의미를 나타내기도 하므로 전처리 과정에서 이를 제외하지 않았다. 패션 머천다이징/리테일 전공 박사 2인이 중복된 어휘와 불용어를 추가로 제거하는 과정을 거쳐 한글 13,160개, 영어로 된 어휘 12,575개를 포함하여 총 25,735개의 어휘를 추출하였다.

3. 3단계: 분류체계 구축

이번 단계에서는 선행연구(Park & Choi, 2020a, 2020b)를 참고로 아이템, 실루엣, 디테일, 스타일, 컬러, 소재, 패턴과 프린트의 7대 속성에 대한 개념을 중심으로 분류체계를 구성하였다. 7대 속성은 너비에 배치하고, 각각의 하위 속성은 길이와 깊이를 포함하는 밑단의 레벨로 설정하였다. 레벨의 숫자가 커질수록 하위 개념에 해당하게 되며, 상위의 레벨은 하위 어휘들의 집합으로 볼 수 있다. 첫 번째 레벨은 가장 큰 분류 개념으로 모든 속성 값을 아우를 수 있다.

예를 들어 속성이 아이템인 경우, 1레벨은 아이템이 신체의 어느 부분을 중심으로 걸쳐지는지에 따라 ‘상의, 하의, 풀바디(full body)’로 나뉜다. ‘상의’ 레벨은 다시 신체 위에 직접 입혀지는지, 혹은 옷 위에 겹쳐 입혀지는지에 따라 ‘탑(top)’과 ‘아우터(outer)’의 2레벨로 분기되는 방식이다. 본 연구의 분류체계 방식에 따르면 분류 레벨이 깊어질수록 자세한 정보를 담을 수 있으나, 지나치게 상세하여 어휘의 활용 빈도가 떨어지는 것을 방지하고자 모든 어휘는 5레벨 이내에 위치시켰다. 그런데도 현재 분류체계에서 포괄하지 못하는 어휘가 생길 경우를 대비하여, 분류체계 내에 ‘기타’ 레벨을 포함시켰다. 이러한 규칙을 적용하여 아이템의 최종 1레벨은 ‘상의, 하의, 풀바디, 기타 아이템’으로 구성되었다. 레벨의 구조와 분기는 <Table 1>의 전공 서적과 패션 사전, 선행연구 등을 참고하여 설정하였다.

4. 4단계: 어휘 분류 및 정의

연구자들은 앞의 2단계에서 분류된 어휘를 하나씩 검토하여 같은 의미로 사용되는 이음동의어들을 하나로 묶고, 그중에서 대표되는 한국어/영어 어휘를 선정하였다. 대표 어휘가 선정된 이후에는 이 어휘가 3단계의 어떤 속성과 레벨에 속하는지를 중심으로 분류하였다. 예를 들어 비즈니스 셔츠(business shirts), 화이트 셔츠(white shirts), 와이셔츠, 와이샤쓰, y셔츠, 정장셔츠, 남방셔츠, 남방, 머슬핏셔츠, 흰색와이셔츠 등의 어휘를 포함하는 대표 어휘 ‘드레스 셔츠(영어 대표어: dress shirts)’는 속성이 아이템인 값 중에서 ‘상의(1레벨)-탑(2레벨)-셔츠(3레벨)-드레스 셔츠(4레벨)’로 정의되었다. 하지만 ‘피치’라는 어휘의 경우, 의미에 따라 컬러 속성에서 과일 빛의 관용색명으로 쓰이거나 소재 속성에서 가공방식 중 하나인 기모에도 포함될 수 있다. 온톨로지와 달리 말뭉치는 어휘가 여러 레벨에 중첩될 수 없으므로 배타성을 띠도록 분류하였다. 따라서 연구자들은 문장 내에서 어휘의 시맨틱을 파악하여, 현재 좀 더 많은 의미로 사용되고 있는 속성에 어휘를 최종 포함시켰다. 그 결과 피치는 컬러 속성의 ‘유채색(1레벨)-오렌지(2레벨)-밝은톤(3레벨)-페일오렌지(4레벨)’와 동의어 개념인 피치 컬러를 지칭하는데 사용되었고, 피치 가공은 소재 속성에서 ‘가공(1레벨)-표면표현(2레벨)-외관가공(3레벨)-피치스킨(4레벨)’으로 분류되었다. 어휘를 분류하는 과정에서 ‘사이즈, 시즌, 트렌드 테마, 판매, 지속가능성’과 관련된 어휘들도 도출되었으나, 상품의 외관 및 특성과 관련성이 다소 적다고 판단되어 본 말뭉치 구축 과정에서 제외하였다. 이 과정을 거치며 한글 4,626개, 영어 3,134개, 총 7,760개의 대표 어휘가 구축되었으며, 연구자들은 세 차례 교차검증을 통해 대표 어휘를 7대 속성 내에서 최대 5층위의 레벨 내에 분류시켰다. 또한 외부 전문가와 다섯 차례 이상의 의견 수렴 과정을 거치며, 분류체계의 타당성을 확보하고자 하였다. 이후 연구자들은 분류체계 내 대표 어휘를 기준으로 사전적인 정의와 어휘를 잘 나타내는 이미지를 추가로 수집하였다.

5. 5단계: 이음동의어 문제의 해결

패션 말뭉치의 확장성을 위해 연구자들은 『봉제용어 순화집』(Korean Apparel Industry Association, 2009),

구글 쇼핑 검색 등을 통해 실무에서 사용되고 있는 어휘들을 추가로 수집하였다. 한국어의 경우 학문적 어휘와 실생활 어휘에 괴리가 있거나, 영어를 한국어로 표현하는 데 표기의 차이가 있을 수 있기 때문에 이들을 최대한 포괄할 수 있도록 이음동의어를 수집하였다. 예를 들어 의복의 실루엣을 구성하는 소매의 형태의 일종인 ‘래글런’의 경우, 영문 표기는 ‘raglan’으로 하였지만, 한국어 이음동의어로 ‘라글란, 라그랑, 라그란, 래글런, 나그랑, 나그란’ 등 함께 사용되는 어휘를 추가하였다.

6. 6단계: 온라인 패션 말뭉치 사전 구축

구축된 패션 말뭉치 사전의 공공 성과 활용도를 높이기 위해 연구자들은 웹사전 형태의 시스템으로 구축하고, 검색과 정보 이용이 제한이 없도록 온라인에 공개하였다. 웹사전에서 제공하는 정보의 범위는 속성별 분류체계 정보, 대표 어휘, 한국어 동의어, 영어 대표 어휘, 영어 동의어, 관련 이미지와 정의, 출처이다.

7. 7단계: 전문가 대상 패션 분류체계 및 말뭉치 사전의 효용성 검증

연구자들이 설계한 패션 말뭉치의 효용성을 증명하기 위해, 산-학-연 전문가 10명에게 정확도에 대한 휴리스틱 검증을 요청하였다. 전문가들은 패션 말뭉치 사전의 구축 목적, 구축 과정, 활용 예시 등에 대한 사전 설명을 충분히 듣고 2022년 8월 중 비대면으로 패션 말뭉치 사전의 정확도에 대한 양적·질적 평가를 실시하였다. 평가문항은 7대 속성에 해당하는 분류체계와 말뭉치의 적합성을 5점의 리커트 척도(1점: 전혀 그렇지 않다~5점: 아주 그렇다)와 자유기술로 측정하였으며, Kim(2008)의 웹사이트 신뢰성 평가지표 중 정보에 대한 진실성, 전문성, 접근성, 활용성, 성과 차원의 18문항을 5점 리커트 척도로 측정하였다.

평가에 참여한 전문가는 산업체에서 패션 디자이너, 패션 머천다이징, 해외 영업 및 패션 데이터 관리팀 소속의 5명이 참여하였으며, 학계에서는 패션 디자인, 의류 소재, 의복 구성, 패션 머천다이징 전공의 교수 4명, 마지막으로 연구소에서는 의류 소재를 담당하는 전문가 1명이 설문에 참여하였다. 전문가들의 졸업 후 평균 업체 경력은 14.85년(최소 11년~최대 16.5년), 직위는 과장급 3명, 차장급 3명, 조교수 2명, 부교수 2명

으로 평가에 전문성을 갖추었다고 볼 수 있다.

8. 8단계: 패션 분류체계 및 말뭉치 사전의 성능 검증

패션 말뭉치의 실질적 활용을 증명하기 위해 본 연구에서 개발한 말뭉치와 일반 말뭉치 사전을 중심으로 텍스트마이닝을 실시하였다. 일반 말뭉치 사전의 경우, 데이터 분석코드를 직접 입력하지 않더라도 웹기반으로 쉽게 텍스트마이닝을 지원하는 프로그램인 The IMC사의 텍스트톰(TEXTOM 3.5)을 사용하였다. 텍스트톰은 의류 패션 분야의 선행연구(Choi & Lee, 2020; Han, 2019; Song & Lim, 2021)에서 텍스트마이닝 분석도구로 주로 사용된 SaaS(software-as-a-service)형 제품이다. 해당 제품은 연구자들이 텍스트마이닝용 서버와 코드를 직접 구축하지 않더라도 한국어나 영어로 작성된 웹페이지나 블로그, 카페, 뉴스 등을 수집하여, 정제 및 텍스트마이닝 분석을 수행하고 그 결과를 직관적으로 시각화하는 기능(TEXTOM, n.d.)을 서비스 형태로 이용할 수 있다.

텍스트마이닝의 전처리 과정에서는 문장에서 형태소를 분류한 뒤 연구목적에 따라 분석하고자 하는 품사를 중심으로 어휘를 추출하게 된다. 텍스트톰은 형태소 분석 시 어휘를 추출하는 최소 단위를 띄어쓰기가 없는 복합 어휘로 추출하는 ‘espresso K’ 분석엔진과 기존 한국어 사전에 수록된 최소 의미단위를 기준으로 추출하는 ‘MeCab’ 분석엔진을 선택할 수 있다(Sung, 2020). 예를 들어 ‘가을원피스’라는 어휘에서 명사를 추출할 경우, ‘espresso K’ 분석엔진은 복합 명사 형태 그대로 ‘가을원피스’를 추출할 수 있지만, ‘MeCab’ 분석엔진을 사용할 경우에는 최소 의미단위인 ‘가을’과 ‘원피스’로 분리하여 추출할 수 있다. 본 연구에서는 말뭉치를 상향식으로 구축하는 과정에서 수집한 어휘를 최소 의미단위로 추출하였기 때문에, 동일한 조건에서 검증을 수행하기 위해 ‘MeCab’ 분석엔진을 이용하였다. 검증 과정에서 형태소 분석은 한국어와 영단어, 숫자를 중심으로 명사와 형용사만 선별하여 추출하였다.

한국어와 영어 어휘의 성능을 모두 비교 평가해 보기 위해 2022년 9월 12일자에 무신사와 더블유컨셉 내 상품 정보를 일부 수집하였다. 무신사나 더블유컨셉과 같은 온라인 플랫폼은 입점업체가 상품을 게시할 때 패션 전문 어휘와 실용 어휘, 영어 어휘를 혼재해서

사용하는 경향이 있다. 따라서 이들은 정제된 표현과 문법을 준수하는 기사 글보다 텍스트의 정제 및 분석에 난이도가 높을 것으로 보았고, 이러한 이유로 해당 텍스트를 검증대상으로 삼았다. 구체적으로 무신사는 무신사스탠다드 상품 중에서 주간 베스트 1~5위에 해당하는 상품의 상세 설명 글을 수집하여 한국어 사전의 성능을 검증하였다. 더블유컨셉은 상품명을 주로 영단어로 표기하고 있어 주간 베스트 상품 중 상위 99개의 상품명을 수집하여 영어 사전의 성능을 검증하고자 하였다.

IV. 연구결과

연구자들은 상품 속성의 관리를 용이하게 하고 텍스트마이닝의 정확도를 높이기 위해 한-영 이음동의어와 속성 내 분류체계가 포함된 패션 말뭉치 사전의 버전 1.0을 구축하였다. 말뭉치 사전에는 7대 패션 속성의 깊이가 최대 5레벨로 확장되며 특성이 구체화되는 분류체계와 함께 분류체계 내 대표 어휘 1,956개, 한국어 및 영어 이음동의어를 포함하여 10,294개의 어휘가 포함되었다. 예를 들어 세로 방향에 섬유이 솔로 된 이랑모양의 파일을 낸 코듀로이(corduroy)는 <Fig. 2>에서와 같이 속성이 소재인 분류체계에 포함된다. 또한 분류체계 상 ‘제직방법-우븐-파일직’에 속한다. 코듀로이의 한국어 동의어는 ‘골텐, 코르텐, 골뎡, 고르뎡’이 있으며, 함께 사용되는 영어 동의어로 ‘cordroy, corduroy’를 지정해 두었다.

1. 패션 도메인의 분류체계 및 말뭉치 구축

1) 아이템

아이템이란 옷의 종류로 티셔츠, 재킷, 치마 등과 같은 품목을 의미한다. 이러한 패션 아이템 분류를 통해 패션 상품을 필터링하고, 상품에 대한 설명을 태그하거나 소비자가 관심을 가질 만한 유사한 상품을 추천하기 위한 기초 데이터로 활용할 수 있다(Seo & Shin, 2019). Han et al.(2016)을 포함한 선행연구에서는 아이템은 품목과 목적, 용도, 혜택, 착용자, 분리 형태에 따라 분류할 수 있었으나, 상품 관리 주체에 따라 분류의 상세함이나 분류 방식에 차이가 컸다. 예를 들어 후드티는 티셔츠 또는 아우터의 하위 개념으로 보았지만, 때로는 상위의 하위 레벨로 취급하기도 했다. 다른 관점으로 아이템은 ‘형용사+명사’의 조합으로 된 상품명에서 명사에 해당하는 품목으로 보기도 하였다(Sadeh et al., 2019).

본 연구에서는 어휘에 대한 정의와 사진자료를 참고하여 먼저 신체의 어느 위치에 해당하는 의복인지를 중심으로 아이템의 1레벨을 상의, 하의, 풀바디로 분류하였다. 상의는 다시 내의인지 내의 위에 입혀지는 외의인지에 따라, 하의의 경우 두 다리를 각각 감싸는 형태인지 서로 연결되어 두 다리를 하나로 덮는 형태인지에 따라 바지와 스커트로 분류하였다. 풀바디의 경우는 품목에 따라 점프수트(jumpsuit), 오버롤(overall), 바디수트(bodysuit), 드레스(dress), 셋업(setup)/세트(set)로 구성되었다. 예를 들어 A라인 스커트의 경우

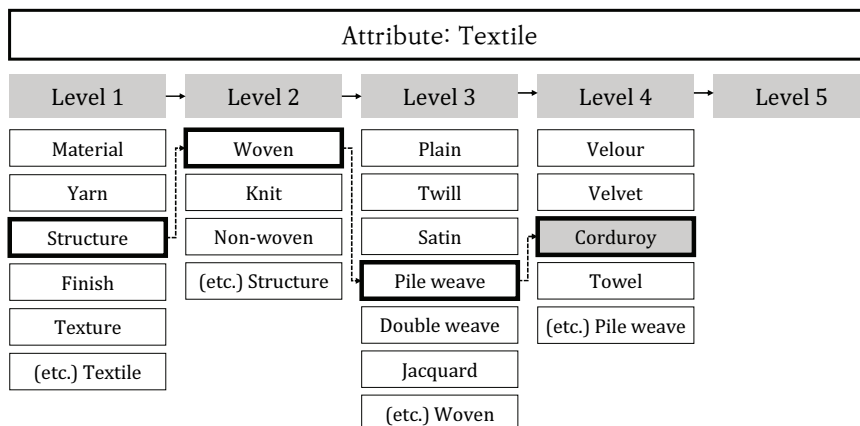


Fig. 2. Illustration of textile taxonomy.

‘하의(1레벨)-스커트(2레벨)-A라인 스커트(3레벨)’로 분류된다. 여기서 2레벨의 스커트는 허리선과 엉덩이 둘레, 밑단의 실루엣에 따른 조합과 총 길이, 그리고 디테일과의 조합에 따라 3레벨에 다양한 종류의 스커트가 포함된다. 아이템 속성의 1레벨은 4개이며, 최대 5레벨까지 형성되어 419개의 대표 어휘를 포함하여 한국어 1,190개, 영어 981개로 총 2,171개의 어휘가 포함되었다.

2) 스타일

스타일은 대상의 형태적 특징이 균질화, 유형화된 것으로 외적으로 조형적 특징과 함께 내적으로 표의를 지닌다. 패션 상품 속성 중 비교적 객관적으로 추출할 수 있어 유형화된 컬러나 소재, 디테일 속성들과 달리 스타일은 소비자나 디자이너의 주관성이 크게 반영되는 패션 속성이며 문화적, 감정적으로 달리 해석되는 속성이기도 하다(Guan et al., 2018). 이에 착용자들은 자신의 정체성에 일관성을 주기 위한 가치관과 취향을 옷의 스타일로 반영하기도 한다(Lee, 2015).

선행연구에서는 스타일을 주로 클래식(classic), 모던(modern)과 같은 이미지로 구분하였다(Hong et al., 2014; Kwon & Park, 2021). 한편 Kwon and Park(2021)은 전통적인 이미지 분류법에 소셜 미디어 데이터 분석을 더하여 패션 스타일의 범위를 확장하였다. 하지만 소셜 미디어 데이터에 치중하다 보니 현재 메가트렌드인 캐주얼(casual)과 스포티(sporty) 스타일에 치중되었다. 또한 이미지 인식을 통한 스타일 분류기 개발 연구(Takagi et al., 2017)에서는 페미닌을 포함하여 총 14개의 서로 다른 여성복 스타일을 학습시켜 패션 사진을 분류하였는데, 하위 14개의 스타일 값이 서로 차별화되지 못하여 분류의 타당성이 부족하다는 한계점이 있다.

본 연구에서는 Lee and Park(2006)을 참고로 스타일을 문화와 시대로 구성하였으며, 다른 한편으로 패션 트렌드와 관련하여 장식 정도나 성별, 시각적 이미지에 따라 스타일을 분류하는 관행에 따라 성별과 패션 이미지에 따른 스타일 분류를 추가로 설정하였다. 문화적 특성에 기반한 스타일은 하나의 문화가 형성되는 범위를 고려하여 2레벨에서 민족/지역 기반 문화 특성(ethnic), 사회문화적 움직임(subculture)을 기반으로 한 특성 및 일상생활과 관련한 문화(lifestyle)를 기준으로 분류하였다. 성별에 따른 스타일 분류는 2레벨에서

성별 특성과 연령 특성을 함께 고려하여 페미닌(feminine), 걸리쉬(girlish), 보이쉬(boyish), 매니쉬(mannish)로 분류하였고, 추가적으로 젠더리스(genderless)와 양성적(androgynous) 특성으로 분류하였다. 시대적 특성은 특히 예술 사조 및 패션 스타일과 관련성을 보이므로 이를 중심으로 하위 어휘를 구성하였다. 특정 예술 사조와는 무관하지만 과거를 연상시키는 스타일인 레트로(retro), 빈티지(vintage)도 추가로 분류하였다. 장식적 특성은 장식의 정도를 고려하여 미니멀(minimal)과 맥시멀(maximal)로 분류하였다. 패션 이미지는 선행연구(Hong et al., 2014; Kwon & Park, 2021)를 참고로 캐주얼, 클래식, 아방가르드(avant-garde)와 같은 대표 이미지 어휘들로 구성하였다. 또한 텍스트 전처리 결과 등 한 제품을 스타일링에 따라 여러 방법으로 표현할 수 있는 ‘양면착용(reversible), 3-way, 체인지웨어, 롤업(roll-up)’ 등의 어휘와 ‘탈부착 가능한(detachable)’ 어휘가 수집되어 이를 1레벨의 ‘스타일링’으로 추가하였다. 결과적으로 스타일 속성에 1레벨은 6개로 분류되었으며 최대 4레벨까지 형성되었다. 스타일의 대표 어휘 203개를 포함하여 한국어 575개, 영어 444개, 총 1,019개의 어휘가 구성되었다.

3) 실루엣

의복에서의 형태 개념은 인체 형태, 의복 실루엣에 의해 만들어지는 외형, 실루엣 내부의 각 부분의 외형 선 등 세 가지 기본적인 형태로 다루어질 수 있다(Horn & Gurel, 1981). DeLong(1987)은 의복 이미지의 형성은 먼저 실루엣에 의해 이루어지고, 관찰 단계를 거쳐 최종적으로 실루엣과 세부적인 부분의 상호작용에 의해 결정된다고 하였다.

선행연구에서 실루엣을 분류한 속성들을 살펴보면 주로 길이, 형태, 소매 길이, 소매 형태, 넥라인 등을 포함한다. Park and Choi(2020a)는 상의와 하의의 실루엣을 형태, 길이, 밑단에 따라 구분하였으며, 상의의 경우 구체적으로 넥라인, 어깨선, 소매 형태, 소매 길이, 소매 밑단으로 분류하였다. 본 연구는 이를 참고하였으나, 상의, 하의, 풀바디의 구조를 결정하는 속성을 보다 효율적이고 체계적으로 포괄하기 위해 길이, 맞음새(fit), 소매 형태, 어깨 형태, 넥라인, 백라인(backline), 허리선, 그리고 밑단(hemline)으로 분류하였다. 길이와 둘레는 형태의 구조적 변화와 관련되는 부분을 기준으로 하위에 상의 길이, 아우터 길이, 소매

길이, 바지 길이, 밑위 길이, 치마 길이로 분류하였다. 둘레는 가슴과 허리, 엉덩이 둘레로 분류하였다. 맞음새는 신체에 얼마나 밀착되는지를 기준으로 하였으며, 소매 형태는 소매의 확장된 부분이 어깨점부터 손목 사이의 하완부 중 어디에 있는지를 기준으로 하위 어휘들을 분류하였다. 어깨 형태의 경우 어깨의 확장 정도를 기준이 되었으며, 넥라인은 파여진 정도와 디자인에 따라 분류되었다. 백라인의 경우 파인 깊이에 따라 하위 어휘를 구성하였고, 허리선의 경우 허리선 높이를 기준으로 분류하였다. 마지막으로 밑단은 대칭 여부에 따라 대칭과 비대칭으로 분류하였다. 실루엣 속성은 1레벨은 11개이며, 최대 4레벨까지 대표 어휘 197개로 분류되었으며, 한국어 455개, 영어 406개로 구성되어 총 861개의 어휘가 포함되었다.

4) 디테일

의복의 디테일은 의복 전체에 대하여 세부적인 부분을 의미한다. 선행연구에서는 디테일을 장식적 디테일과 구조적 디테일로 나누고 있다(Park & Choi, 2020a). 장식적 디테일은 의복에 덧붙이면 새로운 느낌을 주는 디테일 요소로 플리즈(pleats), 셔링(shirring), 턱(tuck) 등이 포함된다. 구조적 디테일은 의복을 구성하고 있는 구조적 측면의 디테일을 의미하는 것으로 넥라인, 칼라, 소매, 포켓(pocket) 등이 해당된다(Kim & Yoon, 2013). 그러나 선행연구에서는 실루엣과 디테일의 구분이 모호한 경우가 많았다. 본 연구에서는 실루엣은 구조적으로 변경할 수 있으며, 탈부착이 불가능한 것으로 정의하였고, 디테일은 탈부착이 가능한 것으로 정의하고 기준에 따라 하위 레벨에 속하는 어휘들을 분류하였다. 예컨대 선행연구(Park & Choi, 2020a)에서는 넥라인을 디테일로 정의하였지만, 넥라인은 의복을 구성함에 있어 탈부착이 불가능하기 때문에 본 연구에서는 실루엣으로 분류하였다.

본 연구에서는 선행연구에서 디테일의 정의에 부합하는 요소들을 분류하여 1레벨을 봉제, 구성, 절개선, 여밈, 트리밍(trimming), 패치(patch), 자수, 주름, 부자재, 부속품으로 설정하였다. 예를 들어 구성의 경우 2레벨에는 칼라, 라펠(lapel), 밑단 디테일, 소매 디테일, 주머니, 후드와 같이 의복 구성에 적용되는 디테일 요소들이 포함되어 있다. 2레벨의 칼라의 경우는 다시 셔츠칼라, 스탠드칼라, 플랫칼라 등으로 세분화되었다. 결과적으로 디테일 속성의 1레벨은 7개이며 최대

4레벨까지 형성되어 346개의 대표 어휘가 생성되었고, 한국어 970개, 영어 623개, 총 1,593개의 어휘들이 포함되었다.

5) 패턴과 프린트

프린트와 패턴은 소재의 표면에 나타난 모양인 문양이나 색채와 관련한 속성이다(Kim et al., 2015). 선행연구에서는 패턴과 프린트를 내용적 특성과 표현적 특성으로 구분하여 하위 차원을 분류하거나(Baek & Bae, 2017), 패턴과 프린트를 스트라이프(stripe), 체크 패턴(check), 자연 무늬, 에스닉 무늬 등으로 패턴의 하위 항목을 구성하여 분류하고 있다(Park & Choi, 2020a). 하지만 패턴과 프린트를 통해 나타낼 수 있는 다양한 시각적 효과까지는 다루지 못했다는 한계점이 존재한다. 패턴은 직조나 프린트에 따라 수없이 많은 변형을 생성하기 때문에 본 연구에서는 패턴과 프린트의 프린팅 기법이나 위치적인 특성까지 고려한 포괄적인 분류체계를 생성하고자 하였다.

가장 먼저 패턴과 프린트가 없는 경우를 분류하기 위해 1레벨에 무지(solid)를 두었으며, 무늬가 있는 경우에는 무늬의 모티프(motif)와 반복 여부를 중심으로 구성하였다. 무늬의 종류인 모티프는 선행연구(Ryu, 2010)를 참고로 체크, 스트라이프, 도트(dot), 플로럴(floral), 애니멀(animal), 오브젝트(object) 등으로 분류하였다. 그러나 모티프의 경우 소재의 직조 방식과 혼용하여 사용되는 어휘가 있어 이를 소재로 분류할 것인지 패턴과 프린트로 분류할 것인지 논의하였다. 그 결과 본 연구에서는 외관상 보이는 형태에 초점을 맞추어 어휘들을 분류하기로 하였다. 예를 들어 하운즈투스(houndstooth) 체크는 ‘프린팅이 아닌 직조 방식으로 패턴을 만드는 것’이지만, 외관상 보이는 무늬에 더 중점을 두어 본 연구에서는 체크 패턴의 하위 항목으로 구성하였다. 또한 1레벨의 무늬 반복은 패턴이나 프린트가 연속적인지에 따라 2레벨에서 원 포인트(one point) 무늬인지 사방연속 무늬인지 구분하였다. 이 밖에도 소재 표면에 다양한 무늬를 생성하는 기법으로 번 아웃(burn out), 타이다이(tie-dye)와 같은 다양한 효과를 분류하기 위해 1레벨에 무늬 기법을 추가하였다. 그 결과 패턴과 프린트 속성은 1레벨이 5개이며, 최대 4레벨까지 형성되어 221개의 대표 어휘가 생성되었고, 한국어 1,038개, 영어 782개, 총 1,820개의 어휘들이 포함되었다.

6) 소재

의류에 사용하는 소재에 따라 용도, 목표 소비자의 연령, 가격대 등이 결정될 수 있다. 소재는 컬러와 함께 브랜드의 아이덴티티를 소비자와 소통할 수 있는 매개체 역할을 하며, 의류의 중요한 속성으로 밝혀졌다(Heo & Nah, 2019). 패션 이미지의 소재 속성 연구(Park & Choi, 2020b)에서는 소재의 특성을 크게 재질, 구조, 시각적 질감, 위치, 무늬의 다섯 가지로 구분하였다. 특히 재질, 구조, 질감에 대한 정보는 소비자가 의류에 대해 지각하는 감성 또는 실루엣에 영향을 줄 수 있으며, 소재의 속성 체계를 이루는 주요 항목이라고 밝혔다(Park & Choi, 2020b).

소재 속성의 1레벨은 선행연구(Ahn, 2016; Young, 2018)를 참고로 소재의 재료, 원사, 직물 구조, 가공, 소재감의 5가지 범주로 분류하였다. 재료는 면, 모, 폴리에스터 등의 소재에 대한 가장 기본적인 정보를 나타내며, 섬유와 혼용률이 관련되어 있다. 구조는 실이나 직조 방법에 따른 분류이다. 소재감은 의류 소재의 표면에서 나타나는 특성으로 촉각과 시각적인 측면으로 구분될 수 있다. 질감의 촉각적 측면은 의류 소재의 원사 종류, 소재의 조직, 가공 등과 같이 물리적인 성질에 의해 느껴지는 촉각적 감성이며, 시각적 측면은 의류의 표면에서 느껴지는 촉각적 경험이 시각적으로 전이된 질감이다(Choo, 2001). 재료는 2레벨에서 소재의 자연적 채취 여부에 따라 천연과 인조로 구분된다. 가공은 2레벨에서 가공 기법에 따라 염색, 워싱(washing), 표면표현, 기능성으로 분류된다. 소재감의 2레벨은 감각 기관에 따라 광택, 비침 등을 포함하는 시각적 소재감과 유연성, 탄성, 계절감 등을 포함하는 촉각적 소재감으로 구성되었다. 쇼핑몰에서 상품의 소개를 위해 빈번히 사용되는 원사명 역시 2레벨에서 종류를 확대하였다. 직물의 구조는 2레벨에서 소재의 짜임에 따라 직물과 편성물, 짜이지 않은 난우븐(non-woven)의 3가지로 분류되었다. 최종적으로 소재는 1레벨이 6개이며, 최대 5레벨까지 분류되어 대표 어휘 396개를 포함하여 한국어 1,038개, 영어 569개로, 총 1,607의 어휘가 포함되었다.

7) 컬러

컬러에 대한 구분과 명확한 의사소통을 위해서는 객관적인 기호나 숫자를 활용한 체계적인 분류가 필요하다. 이는 니즈에 따라(Seo & Kim, 2019), 먼셀(Mun-

sell), NCS, 오스트발트(Ostwald) 등과 같이 다양한 표색계(color system)가 개발되었다. 이 시스템들은 주로 R(빨강), Y(노랑), G(초록), B(파랑), P(보라) 등의 계통으로 분류해서 컬러를 분류한다(Cha, 2018). 컬러를 분류하는 또 다른 방법은 컬러의 연상을 불러일으키는 관용색명을 사용하는 것인데, 이는 소비자에게 친근함을 주는 이유로 상품 컬러명이나 유행 컬러명의 선정에 많이 활용된다(Lee, 2009). 대표적으로 팬톤 컬러의 경우, 관용색명을 사용하여 새로운 컬러의 추가가 용이하다는 장점이 있지만 구조적으로 분류하기는 어려움이 따른다. 또한 색은 색상과 더불어, 농담, 명암, 강약 등의 상태로도 구별될 수 있다. 그 예로 톤(tone)은 색의 농담과 선명함, 탁함 등을 구별하기 위하여 사용되며 명도와 채도를 합한 복합 개념이다(Lee, 2009).

컬러는 그 다양성에도 불구하고 신경망을 통해 의류 이미지의 속성을 분류하는 연구(Chakraborty et al., 2020)에서는 학습의 편의성을 높이기 위해 주로 10여 개의 단순화된 색상 값만이 사용되었다. Park and Choi (2020b)는 기존의 컬러 체계에서 더 확장하여 26개의 컬러와 13개의 톤으로 구성하였으나, 계통색에 대한 분류가 직관적으로 어려울 뿐만 아니라 일반 소비자와 소통이 효율적일 수 있는 관용색명이 부재하다는 한계를 지닌다. 이에 본 연구는 계통색뿐 아니라 패션 상품명과 트렌드 컬러에 자주 사용되는 카멜(camel), 카키(khaki), 민트(mint) 등의 관용색명을 포함한 세분화된 새로운 컬러 분류를 제시하고자 하였다. 또한 다양한 컬러를 포괄하기 위해 색상에 다양한 톤을 조합하여 컬러의 기본 분류체계를 완성하였다.

구체적으로 살펴보면 1레벨은 유채색, 무채색, 특수색으로 구성된다. 유채색의 하위 레벨에는 선행연구(Park & Choi, 2020b)에서 계통색으로 주로 사용되는 레드(red), 오렌지(orange), 옐로우 그린(yellow green) 등이 있다. 여기에 패션 상품명에서 빈번히 사용되는 컬러 10종을 추가하여 총 18가지 컬러를 유채색의 2레벨로 선정하였다. 무채색은 흰색, 회색, 검정색이 2레벨에 포함되었으며, 특수색에는 광택이 있는 메탈(metal) 컬러, 투명색, 멀티 컬러 등이 2레벨에 포함되었다. 유채색 및 무채색 중 회색의 경우, 하위 레벨은 네 가지 톤, 즉, 원색 톤, 어두운 톤, 밝은 톤, 형광 톤이 결합되었다. 그러나 일부 계통색은 형광색으로 구현이 불가능하여 ‘색명+톤’의 규칙을 적용하지 않았다.

컬러의 경우 1레벨은 4개이며, 이들은 최대 4레벨까지 형성되어 대표 어휘 174개를 포함한 한국어 798개, 영어 425개로, 총 1,223개의 어휘가 구성되었다.

이상에서와 같이 한국어를 기준으로 대표 어휘는 1,956개, 총 어휘 수는 10,294개가 말뭉치로 구축되었다. 속성별로 구축된 어휘의 수는 <Table 2>와 같다.

2. 온라인 패션 분류체계 및 말뭉치 사전 시스템 구축 결과

패션 분류체계 및 말뭉치의 쉬운 관리와 이용을 위해 연구자들은 <Fig. 3>과 같이 웹사이트 형태의 온라인 시스템을 구축하였다. 이용자들은 웹사이트에서 찾고자 하는 개념어를 두 가지 방식으로 검색할 수 있다. 첫 번째는 <Fig. 3>의 왼쪽 예시와 같이 드릴다운(drill down) 형식으로 속성별 분류체계에서 상위 레벨(요약된 레벨)에서부터 하위 레벨(상세한 레벨)로 특성을 구체화해가며 정보를 찾는 방식이다. 이는 이용자가 개념

어를 정확히 모르는 상태에서 필요한 개념어를 찾거나 할 때 사용할 수 있다. 두 번째는 검색창을 이용하여 직접 입력하는 방식으로, 개념어는 알고 있지만 정확한 뜻을 알고 싶을 때 사용하면 편리하며, 사용자는 패션 어휘에 대해 기술된 정보를 확인할 수 있다. 웹사이트에서 제공하는 정보는 <Fig. 3>의 우측 예시와 같이 속성 정보, 분류체계 위치, 한국어 동의어, 영어 대표어, 영어 동의어, 관련 이미지와 정의, 출처이다. 추가적으로 현재 수록된 어휘 전체를 엑셀 파일로 다운로드하여 분석 및 품질 관리의 편의성을 추구하였다. 웹사이트의 관리자는 관리자 페이지를 통해 신규 어휘나 신규 속성을 등록하거나, 분류에 대한 정보를 수정할 수 있다.

3. 패션 분류체계 및 말뭉치 사전의 정확도 검증

패션 분류체계 및 말뭉치 사전에 대한 전문가 평가의 결과는 <Table 3>과 같다. 우선 전문가들은 패션 말뭉치 사전의 분류체계(평균 4.38점) 및 말뭉치(4.31점)

Table 2. Number of fashion thesaurus

Attributes	Korean		English		Total
	Standard vocabulary	Synonym	Standard vocabulary	Synonym	
Item	419	771	373	608	2,171
Style	203	372	162	282	1,019
Silhouette	197	258	151	255	861
Detail	346	624	260	363	1,593
Pattern & Print	221	817	188	594	1,820
Textile	396	642	300	269	1,607
Color	174	624	139	286	1,223
Total	1,956	4,108	1,573	2,657	10,294



Fig. 3. Illustration of online fashion thesaurus.

Table 3. Evaluation score of thesaurus

Evaluation items	Average of evaluation score*			
	Overall	Taxonomy	Thesaurus	
Attributes of fashion	Item	4.08	4.08	4.08
	Style	4.35	4.45	4.25
	Silhouette	4.32	4.33	4.30
	Detail	4.13	4.13	4.13
	Pattern & Print	4.53	4.61	4.45
	Textile	4.48	4.53	4.43
	Color	4.53	4.53	4.53
	Total	4.34	4.38	4.31
Web dictionary	Credibility	4.19		
	Usefulness	4.22		
	Accessibility	3.98		
	Usability	4.30		
	Satisfaction	4.19		
	Intention to use	4.25		

*: 5-point likert scale (1: strongly disagree-5: strongly agree)

에 대해 비교적 정확하게 구성되었다고 평가하였다. 전문가들은 분류체계와 말뭉치에 대한 수정 의견을 함께 제시하였는데, 연구자들은 전문가들의 정성적인 평가 의견을 반영하여 패션 말뭉치 사전을 최종 수정하였다. 전문가들은 웹사전에 대한 효용성 역시 5점 만점에 신뢰도(4.19점), 유용성(4.22점), 접근성(3.98점), 업무 활용도(4.30점), 전반적 만족도(4.13점), 향후 이용의도(4.25점)를 평가하였으며 그 결과 산-학-연의 관점에서 효용성을 어느 정도 확보하고 있다고 평가하였다.

4. 패션 말뭉치 사전의 활용

본 연구에서 제안한 패션 말뭉치 사전은 기본적으로 텍스트마이닝의 정제 단계에서 어휘 추출의 기준 값으로 활용이 가능하다. 이에 패션 말뭉치의 실질적 활용을 증명하기 위해 본 연구의 말뭉치와 일반 한국어 분석엔진을 중심으로 텍스트마이닝을 실시하였다. 주요 결과는 <Table 4>-<Table 5>와 같다.

먼저 무신사 데이터의 경우 최소 의미단위를 기준으로 추출하는 텍스트의 ‘McCab’은 ‘셔츠’, ‘팬츠’와 같이 일반적으로 널리 사용되는 패션 어휘를 추출하는데 어려움이 적었으나, 패션 어휘가 결합된 합성어

나 전문 용어는 한 번에 정제시키지 못했다. 예를 들어 ‘테일러드 재킷’의 경우, 패션 말뭉치 사전은 이를 단번에 인식할 수 있었으나, ‘McCab’은 ‘테일러’+‘드’+‘재킷’, ‘테일러드’+‘재킷’ 또는 ‘테일’+‘러드’+‘재킷’으로 분리해서 인식했다. ‘젠더리스’와 같이 비교적 최근에 등장한 전문 용어의 경우에도 일반 사전은 하나의 최소 의미단위가 아닌 ‘젠더’와 ‘리스’로 분리하여 인식하였다. 따라서 ‘McCab’을 이용할 경우, 연구자들은 도출된 어휘의 앞뒤 어휘를 확인하여 의미에 맞게 조합하는 과정을 추가적으로 수행해야 할 것이다.

이밖에도 패션 말뭉치 사전은 일반 사전보다 약 2~4배 많은 패션 전문 어휘를 도출해냈다. <Table 4>를 살펴보면, 패션 말뭉치 사전의 경우 전반적으로 정확도가 높은 것을 알 수 있었는데, 외래어를 한글로 직접 표기하는 경우가 많은 패션 상품의 특성상 대표어와 이음동어어에 해당하는 어휘들이 수집되어 텍스트마이닝에 반영된 결과로 볼 수 있다. 또한 패션 상품과 소재에 대한 전문적인 지식이 더해져서는 텍스트의 정확도가 높아진 것으로 생각할 수 있다. 특히 소재의 경우, ‘원단’, ‘코튼’, ‘폴리우레탄’이라는 큰 범주의 소재 종류를 포함하여 ‘인터록’, ‘프렌치테리’, ‘싱글저지’, ‘스판텍스’와 같은 상세한 수준의 소재 종류를 도출할 수 있었다. 패션 말뭉치 사전은 소재의 종류뿐 아니라

Table 4. Results of Musinsa text mining ('MeCab' vs Fashion thesaurus)

Item	Style	Silhouette	Detail	Pattern & Print	Textile	Color
◎Pants	◎Basic	●Neckline	◎Band	N.A.	◎Cotton	N.A.
◎Shirts	◎Gender	●Oversize	◎String		◎Fabric	
◎Suit	●Genderless	●Regular fit	●Bartack		◎Polyurethane	
◎T-shirt	○Less	●Round neckline	●Belt		●Bio washing	
●Sweatshirts		●Silhouette	●Button		●Finishing	
●Tailored jacket		○Regular	●Gather		●French terry	
○D		○Wide	●Hemline		●Interlock	
○Lored			●Lapel		●Polish	
○Sweat			●Pad		●Single jersey	
○Tail			●Pocket		●Spandex	
○Tailor			●Waist-band		●Stretch	

●Fashion thesaurus only; ○'MeCab' engine only; ◎Both
N.A.: Word not derived

Table 5. Results of Wconcept text mining ('MeCab' vs Fashion thesaurus)

Item	Style	Silhouette	Detail	Pattern & Print	Textile	Color
◎Cardigan	N.A.	◎Crop	◎Button	●Argyle	◎Cotton	◎Black
◎Coat		●Half	◎Collar	●Artwork	◎Leather	◎Blue
◎Dress		●Long	◎Hood	●Herringbone	◎Melange	◎Brown
◎Jacket		●Midi	◎Zip	●Square	◎Mohair	◎Ivory
●Blazer jacket		●Round neckline	●Pocket	●Stripe	◎Tweed	◎Navy
●Bolero		●Short	●Stitch		◎Wool	◎Pink
●Trench coat		●Sleeveless			●Cashmere	◎2color
●Sweatshirt		●Square neckline			●Hard	◎3color
●Varsity jacket		○Neck			●Fur	●Camel
○Blaze					●Knit	●Cream
○Trench					●Lambskin	●Mint
○Var						●Mustard

●Fashion thesaurus only; ○'MeCab' engine only; ◎Both
N.A.: Word not derived

‘바이오 워싱’, ‘광택’ 등 소재 가공 방법 및 ‘신축성’과 같은 촉감과 관련된 전문 어휘도 인식하여 일반 사전 보다 우수한 성능을 보이는 것으로 나타났다.

다음으로 상품명이 주로 영어로 구성되어 있는 더블유컨셉 데이터를 대상으로 텍스트를 정제한 결과 <Table 5>도 무신사 쇼핑몰의 텍스트마이닝 결과와 비슷한 경향을 나타냈다. 일반 사전에서 정제된 ‘jacket’의 경우, 재킷만 단독 어휘로 취급이 되어 어떤 특성의 재킷인지 확인하기 어려웠다. 학교나 소속팀의 로고를 부착한 유니폼 형태의 ‘바시티(varsity) 재킷’의 경우, 일반 사전은 ‘바시’, ‘티’, ‘재킷’의 각각 다른 단어로 정제되었는데, ‘바시’+‘티’ 또는 ‘바’+‘시티’로 일관되지 않게 분리되었다. ‘FW’ 또는 ‘SS’와 같이 시즌을 나타내는 어휘도 개별적인 알파벳으로 분리되었다.

현재 패션 말뭉치 사전에서는 시즌의 속성을 포함

하고 있지 않지만, 추후 소비 맥락과 함께 분석하기 위해서는 계절감이나 이벤트, 시즌 등을 하나의 속성으로 추가해야 할 필요성을 느꼈다. 또한 원단 직조 방법 중 하나인 ‘tweed’는 패션 말뭉치 사전에서는 제대로 분리해 낸 반면, 일반 사전에서는 ‘귀엽다’는 형용사 혹은 ‘숫자 2’와 관련된 명사/형용사인 ‘twee’로 잘못 정제되어 나타났다. 컬러의 경우에도 패션 말뭉치 사전은 패션 상품에서 자주 사용되는 크림(cream), 카멜, 머스타드(mustard), 민트를 포함하여 10개의 관용색명을 정제하며 일반 사전의 어휘보다 풍부하게 컬러명을 인식할 수 있었다.

V. 결론 및 논의

마케팅 분야에서 특정 분야의 표준화된 분류체계

와 말뚱치를 구축하는 것은 패션 상품 속성의 관리를 용이하게 하고 텍스트마이닝의 정확도를 높일 수 있기 때문에 날로 중요해지고 있다. 이에 연구자들은 패션 상품의 특성을 기술하는 데 주로 사용되는 속성인 아이템, 스타일, 실루엣, 디테일, 패턴과 프린트, 컬러, 소재를 중심으로 어휘를 수집하여 말뚱치로 구축하였다. 또한 어휘의 표준화를 위해 7대 속성별로 상위 1레벨에서부터 하위 최대 5레벨에 이르기까지 정보를 상세하게 분화시켜 나가는 하향식 개발법과 다양한 원천에서 추출한 텍스트에서 상위 개념을 생성하는 상향식 접근법을 혼합하여 패션에 특화된 분류체계를 구축하였다. 또한 분류체계 내 대표 어휘를 중심으로 한국어 및 영어로 된 이음동의어를 추가로 수집하여 텍스트마이닝의 정확도를 높이고자 하였다. 결과적으로 7대 패션 속성 내 대표 어휘 1,956개는 이음동의어를 포함하여 10,294개의 말뚱치로 확장되었고, 대표 어휘들은 7대 속성의 분류체계 내에 특정 위치에 속하게 분류되었다.

본 패션 사전은 다음과 같은 의의를 지닌다. 첫째, 패션 상품을 구성하는 7대 패션 속성에 대해 단순히 어휘를 나열하는 것에서 그치지 않고, 최대 5클래스의 분류체계로 체계적으로 구성함과 동시에, 이음동의어를 모두 포함하는 한국어 전용 패션 사전으로 텍스트마이닝 연구에서 학술적 의의가 있다. 같은 의미를 가진 여러 단어를 한 단어로 인식할 수 있게 구축함으로써 텍스트마이닝 연구뿐만 아니라, 텍스트로 이루어진 빅데이터를 분류하고 AI 학습에도 활용할 수 있으며, 학습 성과 또한 높을 것으로 기대할 수 있다. 본 사전의 동의어 처리 시스템은 AI로 학습해야 하는 텍스트 기반 챗봇(chatbot) 서비스와 같은 인터페이스를 통한 고객의 질의 맥락을 이해하고, 즉각적이고 정확하게 응대할 수 있도록 인공지능 학습에 기초 자료로 사용될 수 있을 것이다.

둘째, 패션 사전은 웹사전의 형태로 구축하였기 때문에 실무자나 연구자 모두가 쉽고 편리하게 사용할 수 있을 것이다. 특히 웹 기반으로 쉽게 어휘와 분류체계를 관리할 수 있기 때문에 시간의 흐름에 따라 새로 등장하는 패션 어휘의 편입이 용이하다. 예를 들어 최신 트렌드로 ‘언더보옴(underboob)’ 패션이 유행하기 시작하였는데, 이는 현재 패션 사전에서 정의되지 않은 어휘이다. 수정과 변형이 어려운 전통적인 사전의 형태라면 언더보옴은 속성이 실루엣인 분류체계 내 ‘길이

(1레벨)-탑길이(2레벨)-크롭(3레벨)’ 또는 ‘기타 탑길이(3레벨)’로 분류되고 정의할 수 있다. 그러나 웹사전의 새 어휘 추가 기능을 이용한다면 탑길이(2레벨)의 하위 분류에 언더보옴(3레벨)을 추가하고, 이에 대한 이음동의어로 ‘밑 가슴 옷’, ‘underboob’을 편리하게 추가할 수도 있다. 이는 새롭게 생겨나는 어휘를 쉽게 시스템에 반영함으로써 텍스트 분석을 통한 트렌드 분석에도 무리 없이 활용될 수 있을 것이다.

셋째, 업체에서는 활용 목적에 따라 속성의 분류체계와 어휘 값을 재구성할 수 있을 것이다. 예를 들어 가능성이 중요한 아웃도어 복종의 경우, 디자인보다는 소재의 무게감이나 용도에 따른 분류가 중요하게 사용될 수 있다. 이 경우에는 소재 속성에서 두께(3레벨)와 아이템의 3레벨(e.g., 코트, 티셔츠, 블라우스 등)을 중심으로 원단 데이터의 정보를 구축하거나, 또는 데이터 분석 시 변수로 적용할 수 있을 것이다. 또한 텍스트마이닝 방법론을 수행하는 패션 연구자들이 빈도를 기준으로 상위 핵심어를 선별하고 중요도를 해석하고자 할 때, 특정 형태소로 분석범위를 한정하는 것과 같이 패션의 속성을 중심으로 분류하고 해석하기가 용이해질 것이다.

넷째, 본 패션 사전은 봉제공장을 포함한 산업현장, 잡지, 오픈플랫폼, 패션 전문 서적, 트렌드 정보회사, 기 발간된 패션 사전에서 사용되고 있는 용어를 수집하고 같은 뜻으로 사용되는 다른 단어를 동의어 처리함으로써 패션 산업의 데이터 표준을 이루는데 기여했다고 판단된다. 패션 브랜드가 사용하고 있는 용어 중 같은 의미이지만 상이한 단어가 있고(e.g., 니트를 스웨터로 사용하는 경우가 있고 다이마루를 의미하는 경우가 있음), IT 기업과 한 단어를 다른 단어로 사용하는 경우(e.g., 패턴을 패션 기업에서는 옷본을 의미하기도 하지만, IT 기업에서는 무늬를 의미)도 있어 데이터 표준이 부재하다는 문제점이 거론된다(Kim & Lee, 2022). 이와 같은 문제로 IT 기업과의 협업 시 용어 정리하는데 많은 시간이 소요되고 있다는 문제점 또한 해결할 수 있을 것으로 보이며, 디지털 트랜스포메이션을 시도하는 많은 패션 기업에서 실무에 본 연구의 결과를 활용할 수 있을 것으로 기대된다.

마지막으로 텍스트 분석뿐만 아니라 이미지 데이터의 메타데이터화(meta-data)가 가능하다. 메타데이터는 웹 검색에 유용하도록 데이터에 대한 데이터를 명시화한 것(Mizoguchi, 2005/2012)으로 상품명 등에서

충분히 상품 속성에 대한 텍스트를 유추하기 어려운 경우에 관리자 페이지에서 7대 속성에 해당하는 정보 값을 텍스트로 입력할 수 있을 것이다. 이를 통해 향후 산업에서 개인의 취향에 더 가까운 상품을 검색하거나 추천하기 위해 필요한 원시 데이터의 정확도가 향상될 수 있을 것이다.

그럼에도 불구하고 다음과 같은 한계점과 후속 연구의 필요성을 지닌다. 먼저 본 연구에서 패션 도메인 지식을 바탕으로 포괄적인 패션 특화 말뭉치를 개발하였음에도 불구하고 데이터 확장의 여지가 있다. 첫째, 패션 상품을 구성하는 7대 주요 속성 외에도 소비자 행동이나 감정과 관련된 어휘, 마케팅 믹스와 관련된 어휘, 트렌드의 테마를 표현하는 어휘 등은 본 연구에서 일부 제외되었다. 이러한 어휘들도 연구목적에 따라 패션 사전의 속성으로 확장이 가능할 것이다. 둘째, K-fashion이 글로벌하게 활동하는 현지점에서 현지 빅데이터를 분석하고 상품화하는 것을 지원하기 위해 중국어, 일본어, 베트남어, 인도네시아어 등 다양한 언어로 말뭉치의 확대가 가능할 것이다. 셋째, 어휘는 생성과 변천, 소멸을 거친다. 현재 시점에서 주로 사용되는 어휘를 수집하였기 때문에 시점을 망라하여 이음동의어를 수집할 필요가 있다. 이를 통해 시대별로 주로 사용되었던 어휘, 새로 등장한 어휘에 대한 고찰 및 시대상과 패션 어휘의 관계를 연구할 수 있을 것이다.

1. 사사

해당사항 없음

2. 연구윤리

해당사항 없음

3. 데이터 및 자료 가용성

본 연구에 사용된 데이터는 합당한 요청이 있는 경우 교신 저자가 제공 가능함.

4. 이해관계 상충

저자 YL은 2021년 4월부터 본 학술지의 편집위원장이지만, 편집과정에 전혀 참여하지 않아 이해관계 상충 문제가 없음.

5. 연구비 지원

본 연구는 한국콘텐츠진흥원의 ‘소상공인의 패션디자인 향상을 위한 지능형 패션 수요 예측 및 판로 분석 기술 개발(R2020040102)’ 사업의 연구비를 지원받아 수행되었음.

6. 저자의 기여

SK, WC와 JJ는 데이터 수집 및 분석을 주로 담당하였고, YL, SJ와 HYK는 이론적 틀 고찰 및 원고 작성을 주로 담당하였다. 모든 저자가 최종 원고를 읽고 승인하였음.

7. 저자 정보

장세윤 서울대학교 생활과학연구소, 책임연구원
김하연 군산대학교 의류학과, 조교수
김송미 서울대학교 의류학과, 박사수료
최우진 서울대학교 의류학과, 대학원생
정진 서울대학교 의류학과, 대학원생
이유리 서울대학교 의류학과, 교수/
 서울대학교 생활과학연구소, 겸무연구원

References

Ahn, D. J. (2016). *Textile science: Merchandiser에게 꼭 필요한 섬유지식*(2판) [Textile science: Fiber knowledge essential to merchandiser (2nd ed.)]. Seoul: Hanol.

Ajmani, S., Ghosh, H., Mallik, A., & Chaudhury, S. (2013). An ontology based personalized garment recommendation system. *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), USA, 03*, 17–20. doi:10.1109/WI-IAT.2013.143

Baek, J. H., & Bae, S. J. (2017). A study on the utilization characteristics of art print for strengthening visual identity of fashion brand. *Journal of the Korean Society of Design Culture*, 23(3), 401–419. doi:10.18208/ksdc.2017.23.3.401

Bengfort, B., Bilbro, R., & Ojeda, T. (2018). *Applied text analysis with Python: Enabling language-aware data products with machine learning*. Sebastopol, CA: O'Reilly Media, Inc.

Bollacker, K., Díaz-Rodríguez, N., & Li, X. (2016, January 1). *Beyond clothing ontologies: Modeling fashion with subjective influence networks*. Paper presented at the KDD Workshop on Machine Learning meets Fashion, San Francisco, CA.

Cha, S. (2018). A study on the tone classification system of KS A 0011 in the perspective of chromatics. *Journal of Korea Society of Color Studies*, 32(2), 129–150. doi:10.17289/jksocs.32.2.201805.129

Chakraborty, S., Hoque, S. M. A., & Kabir, S. M. F. (2020). Predicting fashion trend using runway images: application of logistic regression in trend forecasting. *International Journal of Fashion Design, Technology and Education*, 13(3), 376–386. doi:10.1080/17543266.2020.1829096

Choi, J. I. (2014). 불한 웹사전의 신조어 수용 [A reflection on neologisms in the French-Korean dictionary on the web].

- Etudes de Langues et Litterature Francaises*, 100, 813–834.
- Choi, Y.-H., & Lee, K.-H. (2020). Ethical fashion research trend using text mining: Network analysis of the published literature 2009–2019. *Fashion & Textile Research Journal*, 22(2), 181–191. doi:10.5805/SFTI.2020.22.2.181
- Choo, S. (2001). *Images of fashion fabric based on color and texture* (Unpublished doctoral dissertation). Yonsei University, Seoul.
- Danescu-Niculescu-Mizil, C., & Lee, L. (2011). Chameleons in imagined conversations: A new approach to understanding coordination of linguistic style in dialogs. *Proceedings of the 2nd Workshop on Cognitive Modeling and Computational Linguistics, USA*, 76–87. doi:10.48550/arXiv.1106.3077
- DeLong, M. R. (1987). *The way we look: A framework for visual analysis of dress*. Ames, IA: Iowa State University Press.
- Fashionary. (2021). *Fashionpedia: The visual dictionary of fashion design*. Hong Kong: Fashionary International Limited.
- Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007
- Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science*, 30(4), 378–391. doi:10.1177/0165551504045856
- Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5-6), 907–928. doi:10.1006/ijhc.1995.1081
- Guan, C., Qin, S., Ling, W., & Long, Y. (2018). Enhancing apparel data based on fashion theory for developing a novel apparel style recommendation system. In Á. Rocha, H. Adeli, L. P. Reis, & S. Costanzo (Eds.), *Trends and advances in information systems and technologies: Vol. 3* (pp. 31–40). Cham: Springer. doi:10.1007/978-3-319-77700-9_4
- Guan, C., Qin, S., & Long, Y. (2019). Apparel-based deep learning system design for apparel style recommendation. *International Journal of Clothing Science and Technology*, 31(3), 376–389. doi:10.1108/IJCST-02-2018-0019
- Han, K. H. (2019). A study on the consumer's perception of Hi-Seoul fashion show using big data analysis. *Journal of Fashion Business*, 23(5), 81–95. doi:10.12940/jfb.2019.23.5.81
- Han, S.-Y., Cho, Y., & Lee, Y. (2016). The effect of the fashion product classification method in online shopping sites. *Journal of the Korean Society of Clothing and Textiles*, 40(2), 287–304. doi:10.5850/JKSC.2016.40.2.287
- Heo, S. H., & Nah, K. (2019). Brand relationship of material and color identity according to color scheme. *Journal of the Korean Society of Design Culture*, 25(3), 523–534. doi:10.18208/ksdc.2019.25.3.523
- Hong, Y. J., Kim, R. R., Lim, S. E., & Kim, Y. I., (2014). The type and characteristics of the modern men's fashion images. *Journal of the Korean Society of Costume*, 64(5), 18–29. doi:10.7233/jksc.2014.64.5.018
- Horn, M. J., & Gurel, L. M. (1981). *The second skin: An interdisciplinary study of clothing* (3rd ed.). Boston, MA: Houghton Mifflin Company
- Jang, N., & Kim, M.-J. (2020). Material as a key element of fashion trend in 2010–2019 - Text mining analysis -. *Fashion & Textile Research Journal*, 22(5), 551–560. doi:10.5805/SFTI.2020.22.5.551
- Jang, S., Kim, H. Y., Lee, Y., Soel, J., Kim, S., & Lee, S.-g. (2022). Deep learning for classification of high-end fashion brand sensibility. *Journal of the Korean Society of Clothing and Textiles*, 46(1), 165–181. doi:10.5850/JKSC.2022.46.1.165
- Kiapour, M. H., Yamaguchi, K., Berg, A. C., & Berg, T. L. (2014). Hipster wars: Discovering elements of fashion styles. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer vision -- ECCV 2014 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I* (pp. 472–488). Cham: Springer. doi:10.1007/978-3-319-10590-1_31
- Kim, C. E., & Lee, J. H. (2018). Trends of big data and artificial intelligence in the fashion industry. *Journal of the Korean Society of Clothing and Textiles*, 42(1), 148–158. doi:10.5850/JKSC.2018.42.1.148
- Kim, H., & Lee, H. K. (2022). Changes in the perception of second-hand fashion consumption in the post-pandemic era. *Fashion & Textile Research Journal*, 24(1), 66–80. doi:10.5805/SFTI.2022.24.1.66
- Kim, J. Y., Oh, K. W., & Jung, H. J. (2015). Creating the idea of textile print pattern design using the visual expression of popular music. *Fashion & Textile Research Journal*, 17(4), 524–540. doi:10.5805/SFTI.2015.17.4.524
- Kim, S.-Y., & Yoon, S.-W. (2013). Fashion design development and type analysis of redesign using clothing details. *Journal of the Korea Fashion & Costume Design Association*, 15(1), 191–206.
- Kim, Y.-K. (2008). Development of indicators for evaluating the web credibility by goodness-of-fit analysis. *Journal of the Korean Society for Information Management*, 25(4), 185–204. doi:10.3743/KOSIM.2008.25.4.185
- Korean Apparel Industry Association. (2009, October). 봉제용어 순화집 [Refined terminology of sewing] [PDF document]. *The Korean Society of Clothing and Textiles*. Retrieved from [https://www.ksct.or.kr/up_file/dboard_dataroom%5B19%](https://www.ksct.or.kr/up_file/dboard_dataroom%5B19%5D)

- 5D%5B0%5D
- Kotler, P., & Armstrong, G. (2015). *Principles of marketing* (16th ed.). Boston, MA: Pearson.
- Kwon, S., & Park, M. (2021). A study on the design characteristics of athleisure look in image-based SNS. *Journal of the Korean Society of Clothing and Textiles*, 45(1), 17–27. doi:10.5850/JKSCT.2021.45.1.17
- Lamb, J. M., & Kallal, M. J. (1992). A conceptual framework for apparel design. *Clothing and Textiles Research Journal*, 10(2), 42–47. doi:10.1177/0887302X9201000207
- Lee, H. G. (2015). *The types of style* (Unpublished doctoral dissertation). Seoul National University, Seoul.
- Lee, J.-J., & Park, E. K. (2006). *Lifestyle & trend* (Rev. ed.). Seoul: Yekyong Publishing Co.
- Lee, K.-H. (2009). Planning of systematic color coordinate system and manufacture of color paper for establishment of basic color education. *Journal of the Korean Society for Clothing Industry*, 11(5), 709–719.
- Lee, K. M. (2018). *Artificial intelligence*. Paju: Life & Power Press Co., Ltd.
- Littrell, M. A., & Miller, N. J. (2001) Marketing across cultures: Consumers' perceptions of product complexity, familiarity, and compatibility. *Journal of Global Marketing*, 15(1), 67–86. doi:10.1300/J042v15n01_05
- Liu, Z., Luo, P., Qiu, S., Wang, X., & Tang, X. (2016). Deep-Fashion: Powering robust clothes recognition and retrieval with rich annotations. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), USA*, 1096–1104. doi:10.1109/CVPR.2016.124
- Mallavarapu, T., Cranfill, L., Kim, E. H., Parizi, R. M., Morris, J., & Son, J. (2021). A federated approach for fine-grained classification of fashion apparel. *Machine Learning with Applications*, 6:100118. doi:10.1016/j.mlwa.2021.100118
- Ministry of Science and ICT. (2022, July 12). AI 학습용 데이터 190종 추가 개방...총 381종으로 늘어 [Additional 190 types of data for AI learning will be opened...a total of 381 types]. *대한민국 정책브리핑* [Korea Policy Briefing]. Retrieved from <https://www.korea.kr/news/policyNewsView.do?newsId=148903465>
- Mizoguchi, R. (2012). *Ontology engineering* (K.-S. Choi, & D.-S. Hwang, Trans.). Seoul: Dooyangsa. (Original work published 2005)
- Noy, N. F., & McGuinness, D. L. (2001, March). Ontology development 101: A guide to creating your first ontology [PDF document]. *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880*. Retrieved from https://protege.stanford.edu/publications/ontology_development/ontology101.pdf
- Park, J. H. (2012). *한국어 동사 의미망 구축 방법론* [Methodology for constructing the meaning network of Korean verbs]. Seoul: Chung Woon.
- Park, N., & Choi, Y. (2020a). A form attribute classification system for fashion items. *International Journal of Fashion Design, Technology and Education*, 13(1), 11–19. doi:10.1080/17543266.2019.1664641
- Park, N., & Choi, Y. (2020b). Color & texture attribute classification system of fashion item image for standardizing learning data in fashion AI. *Journal of the Korean Society of Clothing and Textiles*, 44(2), 354–368. doi:10.5850/JKSCT.2020.44.2.354
- Park, N., & Choi, Y. (2021). Classification system of fashion emotion for the standardization of data. *Journal of the Korean Society of Clothing and Textiles*, 45(6), 949–964. doi:10.5850/JKSCT.2021.45.6.949
- Ryu, H.-J. (2010). *Pattern fashion*. Seoul: Idambooks.
- Sadeh, G., Fritz, L., Shalev, G., & Oks, E. (2019, June 15). Joint visual-textual embedding for multimodal style search. *Cornell University*. Retrieved from <https://arxiv.org/abs/1906.06620>
- Seo, Y., & Kim, Y. (2019). Characteristics and network analysis of seasonal fashion trend color. *Journal of Korea Society of Color Studies*, 33(2), 17–28. doi:10.17289/jkses.33.2.2019.05.17
- Seo, Y., & Shin, K.-s. (2019). Hierarchical convolutional neural networks for fashion image classification. *Expert Systems with Applications*, 116, 328–339. doi:10.1016/j.eswa.2018.09.022
- Shin, S.-Y., & Kim, Y.-I. (2013). The classification of fashion frame and fashion image of Korean women in their twenties and thirties. *Journal of the Korean Society of Costume*, 63(4), 118–131. doi:10.7233/jksc.2013.63.4.118
- Song, E.-y., & Lim, H.-s. (2021). Perceptions and trends of digital fashion technology - A big data analysis -. *Fashion & Textile Research Journal*, 23(3), 380–389. doi:10.5805/SFTI.2021.23.3.380
- Sung, K.-S. (2020). Social media big data analysis of Z-generation fashion. *Journal of the Korea Fashion & Costume Design Association*, 22(3), 49–61. doi:10.30751/kfeda.2020.2.3.49
- Takagi, M., Simo-Serra, E., Iizuka, S., & Ishikawa, H. (2017). What makes a style: Experimental analysis of fashion prediction. *Proceedings of 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Italy*, 2247–2253. doi:10.1109/ICCVW.2017.263
- TEXTOM. (n.d.) About TEXTOM. *TEXTOM*. Retrieved from <https://www.textom.co.kr/home/sub/overview.php?pnm=1>
- Tortora, P. G., & Keiser, S. J. (2014). *The fairchild books dic-*

- tionary of fashion*. (4th ed.). New York, NY: Bloomsburg Publishing Inc.
- Uschold, M., King, M., Moralee, S., & Zorgios, Y. (1998). *The enterprise ontology*. *The Knowledge Engineering Review*, 13(1), 31–89. doi:10.1017/S0269888998001088
- Vogiatzis, D., Pierrakos, D., Paliouras, G., Jenkyn-Jones, S., & Possen, B. J. H. H. A. (2012). Expert and community based style advice. *Expert Systems with Application*, 39(12), 10647–10655. doi:10.1016/j.eswa.2012.02.178
- Wu, H., Gao, Y., Guo, X., Al-Halah, Z., Rennie, S., Grauman, K., & Feris, R. (2021). Fashion IQ: A new dataset towards retrieving images by natural language feedback. *Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), USA*, 11302–11312. doi:10.1109/CVPR46437.2021.01115
- Xiao, H., Rasul, K., & Vollgraf, R. (2017, August 25). Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. *Cornell University*. Retrieved from <https://arxiv.org/abs/1708.07747>
- Young, D. E. (2018). *Swatch reference guide for fashion fabrics* (4th ed.). New York, NY: Bloomsburg Publishing Inc.
- Zheng, S., Yang, F., Kiapour, M. H., & Piramuthu, R. (2018). ModaNet: A large-scale street fashion dataset with polygon annotations. *Proceedings of the 26th ACM international conference on Multimedia, Korea*, 1670–1678. doi:10.1145/3240508.3240652
- Zou, X., Kong, X., Wong, W., Wang, C., Liu, Y., & Cao, Y. (2019). FashionAI: A hierarchical dataset for fashion understanding. *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), USA*, 296–304. doi:10.1109/CVPRW.2019.00039