

텍스트마이닝을 활용한 온라인 판매 여성 청바지 상품명에 나타난 키워드의 정보 특성 분석

강 여 선[†]

덕성여자대학교 의상디자인 전공

A Study on Keyword Information Characteristics of Product Names for Online Sales of Women's Jeans Using Text Mining

Yeo Sun Kang[†]

Dept. of Fashion Design, Duksung Women's University

Received August 3, 2022; Revised (October 6, 2022; December 1, 2022); Accepted December 5, 2022

Abstract

This study used text mining to extract 2,842 keywords from 7,397 product names and organized them into categories in order to analyze the characteristics of keywords appearing in the product names of jeans after 2020. The item category included denim and Chungbaji [청바지], and Ilja [일자], while the silhouette category included wide and bootcut. In addition, high-waist and banding comprised the making sector, and the materials category consisted of napping, spandex, and soft blue. Denim surpassed the others in frequency, co-occurrence frequency, and centrality, and co-appeared with various other keywords. Also, the co-appearance of item and silhouette was prominent, and there were many keyword combinations that showed characteristics related to (a) high waist; (b) hemline detail; (c) rubber band; and (d) partial tearing. Furthermore, idiom expressions such as 'slim fit' and 'back tearing', which were not highlighted in the co-occurrence frequency, were additionally confirmed through correlation. Therefore, the product name analysis effectively identified the detailed characteristics of the silhouette and the making of jeans preferred by consumers.

Key words: Jeans, Text mining, Online shopping, Bigdata, Product title; 청바지, 텍스트마이닝, 온라인쇼핑, 빅데이터, 상품명

I. 서 론

2021년 12월 온라인쇼핑 연간 거래액은 역대 최고치를 경신하였고 의복 거래액은 전년 대비 12.6% 증가하였으며 상품군별 비중도 4위에 올랐다(Statistics Korea, 2022). 온라인쇼핑은 시간과 장소에 따른 제약이 적고 다양한 상품을 쉽게 비교할 수 있다는 편리함 때

문에 계속 성장하고 있다. 다만, 의류의 경우는 복종과 스타일이 매우 다양할 뿐 아니라 같은 스타일이라도 브랜드마다 출시하는 제품이 각양각색이며 특히, 개인의 신체 치수와 취향에 따라 선호 스타일과 호칭도 달라지므로 온라인쇼핑에서 원하는 제품을 찾는 과정은 다른 제품군에 비해 훨씬 어렵고 불편하다. 때로 하루 6,000종의 신상품이 출시되는(Kim, 2022) 등 수시로 많은 신제품이 출시되는 특성으로 인해 제품 검색에 더 많은 시간과 노력이 필요하다.

[†]Corresponding author

E-mail: yskang@duksung.ac.kr

온라인쇼핑에서 소비자는 아이템, 스타일, 색상, 치수 등 원하는 정보로 상품을 검색한 후 해당 상품의 상세정보를 확인하고 구매를 결정한다. 즉, 상품명에 포함된 정보는 검색과정에서 소비자와 상품을 연결하는 고리로서 온라인쇼핑의 시발점이라 하겠다. 또한, 소비자는 상품명에 표기된 정보를 통해 제품에 대한 긍정적 기대치를 갖게 되면 제품의 상세페이지에서 추가정보를 확인하고 구매를 진행한다. 따라서 상품명에 소개된 제품정보는 제품과 소비자 간의 연결고리뿐 아니라 제품에 대한 기대치 형성에 영향을 미치는 매우 중요한 요소라 하겠다. 이에 패션 브랜드들은 온라인쇼핑 환경에서 제품의 노출 빈도를 높이고 소비자에게 제품 특성을 잘 전달할 수 있는 효율적인 검색어를 상품명에 포함하길 원한다.

상품명에는 일반적으로 스타일, 소재, 사이즈, 색상 등의 정보와 아이템명, 브랜드명 등이 포함된다. 그러나 같은 정보라도 한글이나 영어 등 언어가 다르거나 영어의 한글식 표기가 다를 수 있다. 또한, 전공용어 대신 소비자들이 사용하는 용어를 사용하기도 한다. 심지어 제품 특성과 상관없는 검색어를 포함시켜 제품의 돌발적 노출을 유도하기도 한다. 그러나 상품명에는 외래어가 압도적으로 많으며(Jeong, & Lee, 2016; Kim, 1985; Kuk 2010) 소비자가 정확한 외래어 표기법을 모르거나 판매자가 작성한 표기와 다르게 검색하면 제품은 검색되지 않는다. 또한, 사이버 쇼핑물 의류 제품의 상품명에는 소재, 구성 관련 내용은 명사로, 이미지나 소재 느낌 관련 내용은 주로 형용사로 표현하나 대부분 판매자가 주관적으로 선택한 단어이므로(Shim & Kim, 2008) 같은 정보를 의미해도 단어 형태에 따라 검색 결과는 달라진다. 이렇듯 용어의 다양한 표현방식이나 불명확한 정보는 소비자와 제품의 연결을 방해하거나 부정확한 정보로부터 잘못된 기대치를 형성하게 하므로 온라인쇼핑의 만족도를 떨어뜨리는 중요 원인이 될 수 있다. 특히, 온라인쇼핑 규모가 커질수록, 다양한 제품이 빠르게 소개될수록 제품 검색과 정보의 전달을 담당하는 키워드는 그 중요성이 더욱 커진다. 따라서 소비자에게 다양한 제품을 소개하고 제품정보를 제공하는 상품명을 조사하여, 상품에 포함된 키워드들의 정보 종류, 구성 특성을 구체적으로 분석한다면 이는 판매자와 구매자 모두에게 유용한 정보로 활용될 수 있을 것이다.

II. 이론적 배경

1. 상품명 및 상품 정보

상품명은 구매를 자극하고 소비자를 설득하는 수단으로 제품의 형태, 장식, 실용성 못지않게 중요한 구매기준이며(Kim, 2005) 동시에 상품의 구성과 기능에 관한 내용을 전달하는 통로이기도 하다. 특히, 온라인 쇼핑의 경우, 상품명에 포함된 문구와 설명은 상품이 미지와 함께 상품에 대한 기대치 형성에 직접 작용한다. 따라서 상품에 관한 정보를 정확하게 전달하고 상품에 대한 긍정적 기대치가 형성될 수 있도록 상품명 구성할 필요가 있다.

80년대 중반 이후 상품명의 중요성에 대한 인식이 확대되면서 관련 연구가 활발하게 이루어졌다(Kuk, 2010). 상품명에 관한 연구는 사회적, 언어적, 문화적 접근의 연구들이 많으며 연구 대상도 식품류(Hong & Song, 2015; Kim, 1985; Kuk, 2010)와 화장품류(Jeong & Lee, 2016; Kim, 1985)가 많은 편이다. 그중 의류 분야 연구는 1985년 상품명의 어휘분석(Kim, 1985)에 의류제품을 포함한 것과 사이버 쇼핑물 의류제품(Shim & Kim, 2008) 등으로 매우 제한적이다.

Kim(1985)은 의류, 제화, 제과, 가전제품 및 화장품의 상품명을 분석하여 외래어 비중이 크고 특히, 의류, 제화, 화장품이 더 그러하며 이는 외국 상품 선호 및 모방에서 기인한다고 하였다. 또한, 전문적인 문법규칙이 유명무실해 지는 경우도 많다고 하였다. Shim and Kim(2008)은 사이버 쇼핑물 여성 의류 BEST 100의 상품명 분석을 통해, 상품명에는 주로 소재, 구성, 이미지 관련 내용이 많으며 쇼핑물 유형에 따라 사용하는 단어 수가 다르고 소재와 구성에 관한 내용은 명사로, 이미지나 소재의 느낌에 관한 내용은 형용사를 주로 사용하나 대부분 판매자가 주관적으로 선택한 단어이기 때문에 소비자 불만요인이 될 수 있으므로 객관적 설명 체계가 필요하다고 하였다. 이처럼 의류 상품명은 외래어가 특히 많고 판매자의 주관적 표현에 따라 구성되는 특징을 보이나 이에 관한 연구는 거의 없는 상황이다. 특히 온라인쇼핑에서 검색 결과와 상품에 대한 기대치를 결정짓는 중요한 요소이므로 상품명 구성하는 용어의 특징을 파악할 필요가 있다.

상품명과 더불어 제품정보에 관한 선행연구를 살펴보면, 스포츠 브랜드 관련 빅데이터에서 제품 관련

용어의 비중이 50%로 가장 많아(Lee et al., 2021) 제품 정보에 대한 높은 관심을 알 수 있었고 반팔 티셔츠 구매리뷰를 통해 소비자들이 제품 구매과정에서 사이즈, 심미성, 소재의 품질 및 세탁 정보에 관심이 높음(Han, 2021)을 알 수 있었다. 또한, 의류 쇼핑몰사이트에서 제공하는 정보, 특히 여성 테일러드 재킷의 사이즈 정보의 부족과 부정확성(Lee & Kim, 2020), 티셔츠 제품의 치수 정보 부족 및 치수 부정확으로 인한 맞춤새 예측의 어려움(Lee et al., 2009), 그 외 화면상의 제품과 실제 상품과의 차이 문제와 상품정보 부족(Kim et al., 2013) 등의 문제가 보고되었다. 이처럼 기사, 제품의 상세정보 또는 구매리뷰를 대상으로 제품정보의 중요성과 문제점 및 소비자들의 관심을 분석하고 있으나 온라인쇼핑의 시발점으로서, 소비자에게 다양한 제품을 소개하고 제품 정보를 제공하는 상품명에 관한 연구는 찾아보기 힘든 상황이다.

2. 패션 분야의 텍스트데이터 분석법

온라인쇼핑의 확대에 따라 브랜드들은 소비자들이 남긴 검색, 구매 등의 빅데이터로부터 소비자들의 관심과 반응을 분석하고자 노력하고 있다. 특히, 최근 빅데이터 분석이 활성화되면서 온라인쇼핑 환경에 존재하는 텍스트데이터들은 실제 활용 가능한 가치정보로 변화하였다. 이런 데이터들은 수시로 생성되고 형태가 다양하며 규모도 방대할 뿐 아니라 실제 소비자들의 생각을 내포하고 있어 그 의의가 크다.

패션 분야에서도 최근 텍스트데이터를 활용한 연구가 빠르게 증가하고 있으며 필요한 정보를 추출하기 위해 다음의 방법을 활용하고 있다. 국내외 스포츠브랜드 관련 텍스트데이터에서 중요 키워드를 추출하고 키워드 간 관계를 파악하기 위해 TF(Term Frequency) 기반의 빈도와 N-gram 네트워크를 분석하거나(Kim & Lee, 2018), 문장에서 단어를 추출하는 TF-IDF(Term Frequency-Inverse Document Frequency) 기반의 빈도를 분석하거나(Kim, 2020) 빈도분석과 변동계수(Coefficient of variation)를 활용해 연관어를 분석함으로써 소재와 연관된 단어를 추출하였다(Jang & Kim, 2020). 또한, 소비자의 생각을 파악하기 위해 오피니언마이닝으로 국내외 스포츠용품 브랜드에 대한 소비자들의 인식을 파악하거나(Kim & Lee, 2018) 제품 리뷰 데이터의 TF-IDF 빈도, 연결중심성 및 감성분석을 통해 티

셔츠 제품에 대한 소비자 관심과 감정을 분석하였으며(Han, 2021) 골프웨어에 대한 소비자 인식 연구에서는 빈도와 상위 30개 키워드들의 중심성을 비교하였다(Lee & Lee, 2018). 방대한 텍스트데이터로부터 수개의 주제를 파악하는 토픽모델링으로는, 아노락 관련 기사에서 LDA 알고리즘으로 3가지 토픽을 추출한 경우(Kim, 2020)와 수영복 관련 텍스트데이터로부터 빈도, 공출현빈도 및 연결중심성으로 추출한 단어들의 군집을 분석하는 CONCOR(Convergence of iterated correlations)분석을 활용하였다(Lee et al., 2017). 블로그 텍스트데이터로부터 남성 스트라이프 셔츠의 디자인 속성과 감성 어휘를 파악한 연구(An & Park, 2018)도 빈도분석, 상관관계에 의해 매트릭스 도출 그리고 군집을 분류하기 위한 CONCOR분석을 하였다.

본 연구의 목적은 여자 청바지 상품명을 구성하는 중요 키워드를 추출하고 키워드들의 정보 특징을 분석함으로써 상품명의 정보전달 및 검색의 효율성을 높이는 방안 모색에 도움이 되는 자료를 마련하는 것이다. 따라서 TF 기반의 빈도분석으로 중요 키워드를 추출한 후 키워드의 정보를 분류하고 그 특징을 파악한 후 공출현빈도와 중심성 분석을 통해 다빈도 키워드들의 연결 관계를 파악하며 나아가 상관관계 분석을 통해 특정 조합을 이루는 키워드 쌍을 파악하고자 한다.

III. 연구방법 및 절차

1. 데이터 수집 및 정제

1) 데이터 수집(Data crawling)

청바지 상품명은 국내 1위 포털사이트인 네이버 쇼핑목록에서 수집하였다. 네이버는 국내 1위 포털사이트이며(Internet Trend, 2022) 특정 브랜드나 특정 집단(성별, 연령별)을 대상으로 하는 사이트보다 사용자 폭이 넓다고 생각했기 때문이다. 또한, 청바지는 네이버 쇼핑에서 하의류 중 팬츠와 별도로 분류되어 있어 상품간 유사성이 높으므로 상품명의 정보를 파악하기에 유리한 상황이다. 더불어 본 연구는 텍스트마이닝을 활용해 상품명의 정보 특징을 분석하는 초기 단계 연구이므로 명확한 결과 도출을 통해 상품명 분석 방법의 타당성을 검증하고자 청바지를 연구대상으로 정했다.

청바지 상품명의 나열 조건은 ‘네이버랭킹순’, ‘낮

은(높은)가격순’, ‘등록일순’, ‘리뷰많은순’, ‘리뷰좋은순’이나 본 연구는 다음과 같은 이유로 ‘리뷰많은순’으로 수집했다. 첫째, 랭킹순으로 수집한 경우, 10,000개의 상품명 중 7,127개가(2022년 1월 11일 기준) 중복데이터로 판명되어 데이터 수가 크게 줄었다. 둘째, 실제 구매가 진행된 상품명의 특성을 파악함으로써 결과의 활용도를 높이고자 하였다. 때로 인위적인 리뷰 수 조작으로 인해 순위가 달라지나 수집 데이터의 수를 늘리고 순위를 배제하고 분석함으로써 이로 인한 영향을 최소화할 수 있다고 판단했다. 한편, 등록된 5백만 개 이상의 상품을 ‘리뷰많은순’으로 정렬하고 상품 등록일을 확인한 결과 최근 2년간의 제품 수는 약 10,000 개 정도였다. 따라서 2022년 1월 13일, 네이버쇼핑에서 여자 청바지 상품명을 ‘리뷰많은순’으로 정렬하고 2020년 이후 상품명을 리스트리(listly)로 수집했다.

2) 데이터 정제(Data refinement)

수집한 데이터에서 중복데이터(2,391개)와 다른 아이템(44개)을 삭제하고 총 7,397개를 분석하였다. 다른 아이템은 남자 데님(14개), 임부용 청바지(5개), 티셔츠(6개), 셔츠(5개), 재킷(3개), 후디(3개), 원피스(3개), 후드집업(2개), 그 외(3개)이었으며 연도별 데이터 수는 <Table 1>과 같다.

2. 분석 방법

1) 텍스트마이닝(Text mining)

텍스트마이닝은 자연어 처리 기술을 바탕으로 비정형 텍스트데이터로부터 원하는 정보를 검색, 추출, 분석하는 빅데이터 분석 방법이다. 따라서 7,397개의 상품명의 텍스트데이터에서 유의미한 키워드를 추출하기 위해 패션 분야 연구에 많이 사용하는 Espresso K(Han, 2021)와 MeCab(Choi & Lee, 2020)을 활용했다. 1차 추출된 단어 중복수 표현이나 띄어쓰기가 다른 경우는 상위빈도의 단어 형태로 통일하였고 무의미한 조사나 감탄사 같은 불용어는 삭제하는 등의 정제과정을 거쳤다. 그러나 의류 상품명에는 표준어와

별도로 외래어 및 전공용어가 매우 많은 특징이 있으며 특히, 계절, 의류 호칭 등 영문과 숫자가 혼용되고 스키니, 찢청 등 사전적 표현이 아닌 용어도 있어 Espresso K와 MeCab으로 추출한 최종 키워드에는 하나의 용어가 2개로 분리되거나 일부가 삭제되는 문제가 다수 발생했다. 따라서 수동으로 형태소분석 및 정제과정을 재진행하여 최종 2,842개의 키워드를 추출했다.

다만, 청바지 상품명에 실제로 사용된 다양한 키워드를 파악하기 위해 유사어 및 같은 의미의 다른 표현은 통일하지 않았다. 즉, 아이템 용어인 데님과 진, 다양한 색상 표현인 진청, D톤 등, 유사한 길이이나 같은 길이로 통일하기 어려운 8부, 9부, Ankle 등은 그대로 분석하였다. 한편 브랜드명은 붙여 써서 고유명사로 인식하도록 하였고 투 버튼, 투 포켓, 투 컬러 등도 붙여 써서 고유명사화했다. 수동정제는 빈도 0.01% 이상인 단어들에서 오류가 발견되지 않을 때까지 반복하였다.

2) 분석 절차

먼저, 청바지 상품명의 중요 키워드와 그 특징을 파악하기 위해 TF(Term Frequency) 기반의 빈도분석에서 출현율 0.01% 이상을 기준으로 다빈도 키워드를 추출하고 키워드의 정보에 따라 분류했다. 정보 종류는 ‘리뷰많은순’ 상위 100개 상품명에 나타난 정보에 따라, 성별, 연령, 브랜드, 계절, 아이템, 실루엣(핏), 제작 방법(의복구성 방법), 바지길이, 소재, 색상, 스타일, 호칭, 판매정보(판매점, 가격, 할인정보 등), 기타의 14개로 구분하였다.

더불어 상품명 구성의 맥락을 파악하기 위해 N-gram 분석으로 공출현빈도를 분석했다. 본 연구의 목적은 상품명에 나타난 정보들의 비중과 관계를 파악하는 것이므로 단어들의 출현 순서(데님-부츠컷, 부츠컷-데님)보다는 공출현 여부에 의미를 두었다. 다음으로 중요 키워드들의 네트워크 구조와 연결 관계를 객관적으로 파악하기 위해 중심성을 분석하였다. 중심성이란 한 키워드가 전체 네트워크에서 중심에 위치하는 정도를 나타내는 수치로 연결중심성(degree centrality), 근접중심성(closeness centrality), 매개중심성(between centrality), 위세중심성(eigenvector centrality)을 분석할 수 있다. 중심성 분석의 대상 노드 수는 대체로 100개 내외를 활용하나 연구에 따라 다양하며 의류 분야는 30~100개(Choi & Lee, 2020; Jang & Kim, 2020;

Table 1. Distribution of product titles by year

	2020	2021	2022	Total
N	1,794	5,391	212	7,397
%	24.3	72.9	2.8	100.0

Kim, 2020; Lee & Lee, 2018; Lee et al., 2021)를 활용하였다. 이에 본 연구는 선행연구와 더불어 정량적 기준인 빈도분석 결과를 반영하여 노드 수를 정했다.

한편, 빈도분석, 공출현빈도 분석, 중심성 분석이 모두 다빈도 키워드의 특징과 관계를 파악하기 위한 분석이므로 빈도와 상관없이 밀접한 관계를 보이는 키워드 조합을 파악하기 위해 단어 간 유사도를 측정하는 상관계수를 활용했다. 단어 간 유사도를 측정하는 방법은 여러 가지이나 본 연구에서는 같이 변화하는 키워드 쌍을 파악함으로써 청바지 상품명의 관용적 표현을 확인하기 위해 상관계수를 분석하였다. 결과분석의 기준은 상품명의 병렬식 단어 표현 특성을 고려하여 상관계수 0.20 이상, 즉, 관계있으나 낮은 관계(TEXTOM, 2017)를 기준으로 삼았다.

IV. 결 과

1. 텍스트 정제 방법별 키워드 추출 결과 비교

텍스트마이닝은 단어 정제 방법에 따라 추출되는 단어가 달라지므로 분석 결과에 미치는 영향이 크다. 특히, 패션 관련 단어들은 일반적 단어 추출 과정과 달리 영어, 숫자가 혼용되거나 영어의 한글 표현이 사전적 표현이 아닌 경우도 많다. 본 연구는 먼저 일반적으로 활용하는 Espresso K와 MeCab을 활용하여 단어를 추출했으나 <Table 2>와 같이 고유명사를 그대로 추출하는 Espresso K는 고유명사 특성이 강한 부츠컷, 하이웨스트 등은 그대로 추출하였으나 ‘스키니’는 보다 보편적 단어인 ‘스키’로, ‘배기’는 ‘배’로, ‘생지’는 ‘생’으로 추출하였다. 일본어 사전을 토대로 한국어에 맞춰 개발되었고 명사 중심으로 추출하는 MeCab은 ‘부츠컷’을 ‘부츠’와 ‘컷’으로 분리하였고 ‘하이웨스트’와 ‘하이웨스트’는 원문에 따라 ‘하’와 ‘이스트’ 또는 ‘웨스트’로 분리하였으며 ‘롤업’은 ‘롤’과 ‘업’으로 ‘리바이스’는 ‘리’와 ‘바이스’로, ‘청반바지’는 ‘청반’과 ‘바지’로 분리하였다. 또한, 호칭(S, M, L, 55, 4 etc.), 색상, 계절(21FW etc.)처럼 영어와 숫자가 혼합된 경우는 둘 다 모두 S-XL를 ‘S’와 ‘XL’로, ‘4col’을 ‘4’와 ‘col’로 분리하였다.

이런 문제들은 텍스트마이닝을 위한 패션 전문 사전을 활용하면 해결될 수 있다. 그러나 정제용 패션 전문 사전을 구축하는 것은 또 다른 연구 영역으로 많은

시간과 노력이 필요하다. 또한 현재 효율적이고 정확한 오픈소스 사전도 거의 없는 상황이다. 이에 본 연구는 수동정제로 최종 키워드를 추출했다. 수동정제로 추출한 상위 50개 단어와 Espresso K 및 MeCab의 단어를 비교한 결과 각각 28, 31개가 포함되었다. 특히, 진, 히든밴딩, 찢청, 2버튼, 크롭 등 10개는 수동정제 과정에서 새롭게 상위 빈도에 포함되었다.

2. 텍스트 마이닝 분석 결과

1) 빈도분석(Frequency of keywords)

최종 추출된 2,842 키워드의 빈도 및 분포 결과는 <Table 3>과 같다. 가장 많이 출현한 키워드는 데님으로 9.91% 출현했다. 그 외에 1% 이상 출현한 키워드는 청바지, 여성, 일자, 기모, 부츠컷, 와이드, 팬츠, 진, 하이웨스트, 슬림, 스키니, 스판, 밴딩, 워싱, 배기, 연청 순의 17개였다. 0.5% 이상 출현한 키워드는 32개, 0.1% 이상은 132개였으며 0.05% 이상은 219개였다. 분포율이 낮아질수록 키워드 수가 급증해 0.01% 이상은 0.02%보다 약 170% 늘어나 761개였다.

누적분포율은 상위 32개가 58.9%, 132개(0.1% 이상)는 약 80%, 761개(0.01% 이상)는 약 95%에 달했다. 나머지 2,081개(73%) 키워드의 총 비중은 3% 미만이었다. 이처럼 전체 키워드 수는 매우 많으나 다빈도 키워드는 약 30~100개이므로 효율적 검색을 위해 다빈도 키워드를 활용할 필요가 있겠다.

2) 정보 종류에 따른 키워드 분류(Keyword classification according to information type)

키워드 특성을 구체적으로 살펴보기 위해 분포율 0.01% 이상인 761개 키워드를 정보에 따라 분류하였다(Table 4). 정보 종류는 아이템, 실루엣(핏), 의복구성(제작 방법), 길이, 소재, 색상, 스타일, 호칭, 성별, 타깃연령, 브랜드, 계절, 판매정보(판매점, 가격)로 구분했다. 또한, 0.02% 이상인 구간(집단 1)과 0.02% 미만~0.01% 이상 구간(집단 2)을 나누어 정보 종류별 분포를 비교하였으며 0.1% 이상(132개) 키워드를 카테고리별로 정리하였다. <Fig. 1>은 가독성을 위해 56개(0.3% 이상) 키워드를 분류한 결과이다.

가장 비중이 큰 카테고리는 ‘아이템(21.64%)’이었고 다음으로 ‘실루엣(18.59%)’, ‘의복구성(13.26%)’, ‘소재(10.42%)’, ‘브랜드명(9.13%)’, ‘색상(5.96%)’이

Table 2. The comparison of 50 keywords extracted by Espresso K, Mecab, and manual refining

Rank	Espresso K		MeCab		Manual refinement	
	Word	N	Word	N	Word	N
1	Denimpants	2,100	Denim	4,343	Denim	4,326
2	Chungbaji [청바지]	1,521	Chungbaji [청바지]	1,934	Chungbaji [청바지]	2,031
3	Denim	1,482	Woman	1,580	Woman	1,655
4	Woman	1,377	Ilja [일자]	1,363	Ilja [일자]	1,364
5	Ilja [일자]	991	Napping	1,332	Napping	1,339
6	Bootscut	953	Cut	1,280	Bootscut	1,264
7	Highwest [하이웨스트]	806	Boots	1,266	Wide	1,129
8	Slim	786	Ding [딩]	1,246	Pants	1,065
9	Wide	752	Pants	1,147	Jeans	1,038
10	Pants	618	Wide	1,126	Highwaist	990
11	Spandex	578	Slim	822	Slim	843
12	Ski	537	Skinny	807	Skinny	809
13	Banding	527	East	800	Spandex	785
14	Ba [배]	426	Spandex	790	Banding	723
15	Light blue	383	Ha [하]	736	Washing	608
16	Straight	357	Washing	610	Baggy	554
17	Washing	348	Baggy	549	Light blue	473
18	1	327	Light blue	473	Semi	419
19	2	308	Semi	417	Cutting	414
20	Female	300	Baji[바지]	414	Dark blue	394
21	3	278	Cutting	412	Straight	388
22	Semi	273	Dark blue	388	Hidden banding	372
23	Dark blue	264	Straight	387	Fit	357
24	4	245	Hidden	372	Long	339
25	DP	241	Fit	342	Female	332
26	S	232	Female	332	Summer	266
27	Cutting	218	Button	310	Baji [바지]	251
28	Levis	179	2	275	Black	228
29	Summer	178	Summer	267	Boyfriendfit	224
30	5	177	Waist	254	Hemline	222
31	Skinnyjean	176	Black	241	Midblue	220
32	Midblue	173	Le [리]	231	Slit	296
33	Banbaji [반바지]	171	Hemline	223	Blueblack	208
34	Blueblack	168	Midblue	218	Winter	204
35	B	167	Boyfriend	215	Cheongbanbaji [청반바지]	195
36	M	167	Up	203	Rollup	194

Bold letters indicates the keywords not included in the final list.

Shaded cell indicates newly appeared keywords in the final list.

Table 1. Continued I

Rank	Espresso K		MeCab		Manual refinement	
	Word	N	Word	N	Word	N
37	D	157	S	202	Raw denim	191
38	21	154	Winter	202	Banbaji [반바지]	189
39	Winter	148	Blueblack	201	2021	184
40	W	145	Roll	194	Levis	183
41	NC	144	Raw denim	194	Ripped jeans	165
42	Boyfit	142	Chengban [청반]	194	Daily	158
43	Daily	140	Banbaji [반바지]	193	2button	157
44	Sang [생]	140	Color	192	Cotton	150
45	A	139	2021	183	Crop	145
46	Rollup	136	Vise [바이스]	178	Short pants	145
47	Baji [바지]	133	Daily	167	Pocket	142
48	Jambangee	129	Cheong [청]	162	Store	141
49	Vintage	119	Ban [밴]	158	Vintage	140
50	PT	118	Pocket	152	Side	140

Bold letters indicates the keywords not included in the final list.
Shaded cell indicates newly appeared keywords in the final list.

Table 3. Frequency and cumulative distribution rate

Rate	5%	1%	0.5%	0.4%	0.3%	0.2%	0.1%	0.05%	0.04%	0.03%	0.02%	0.01%	Total
Number of words	1	17	32	40	56	76	132	219	256	327	447	761	2,582
Increment	-	16	15	8	16	20	55	87	37	71	120	314	1,821
CPE%	9.9	48.1	58.9	62.5	67.7	72.5	80.4	85.8	87.36	89.3	91.6	94.5	100.0

며 이들의 누적분포율은 79.04%다. 그 외 성별(4.67%), 계절(2.90%), 옷길이(2.61%), 스타일(1.91%), 호칭(1.51%) 순이었다. 따라서 청바지 제품명을 구성하는 메인 키워드는 아이템, 실루엣, 의복구성, 소재, 브랜드, 색상 키워드라 하겠다. 한편 761개 중 41.26%의 키워드가 포함된 집단 2에서는 브랜드명(1.86%)이 가장 많았고 다음으로 의복구성(0.23%), 소재(0.17%), 호칭(0.12%) 순이었다. 브랜드와 호칭 정보 분포가 커졌음을 알 수 있다.

카테고리별 대표 키워드는 다음과 같다. 먼저, 아이템 키워드로는 데님(9.91%), 청바지(4.65%), 진(2.33%)의 비중이 컸고 그 외 일반 바지를 뜻하는 팬츠(2.44%), 바지(0.57%)의 비중이 컸다. 다음으로 실루엣 키워드는 일자(3.12%), 부츠컷(2.89%), 와이드(2.59%), 슬림(1.93%), 스키니(1.85%), 배기(1.27%), 스트레이트(0.89

% 순이었고 그 외 핏, 보이프렌드핏, 루즈, 통, 테이퍼드, 조거, 맘핏 등의 키워드가 나타났다. 일자, 부츠컷, 와이드의 분포율이 2% 이상으로 높게 나타난 결과를 통해 기본 실루엣인 일자에 대한 소비자들의 상시적 관심 및 부츠컷 및 와이드에 대한 현시점에서의 높은 관심을 엿볼 수 있겠다.

의복구성 키워드들은 청바지 구성 특성을 상세히 나타냈으며 크게 1) 허리의 높이와 구성법 관련 키워드, 2) 밑단 디테일 관련 키워드, 3) 청바지의 부분적 찢어짐에 관한 키워드로 분류되었다. 먼저 허리높이 관련 다빈도 키워드는 하이웨스트(2.27%), 하이라이즈(0.26%), 반하이웨스트(0.41%) 등이었고 허리 구성법은 밴딩(1.66%), 히든밴딩(0.85%), 뒷밴딩(0.22%) 등이었다. 다음으로 밑단의 디테일과 관련된 키워드는 컷팅(0.95%), 트임(0.68%), 밑단(0.51%), 룰업(0.44%),

Table 4. Distribution of 761 keywords by information type

Category	Group 1 (0.02% or more)	Group 2 (0.01-0.02%)	Sub total	132 words with a frequency of 0.1% and more
Item	21.62	0.02	21.64	Denim (9.91), Cheongbaji [청바지] (4.65), Pants (2.44), Jeans (2.38), Baji [바지] (0.57), Short jeans (0.45), Banbaji [반바지] (0.43), Short pants (0.33), Hot pants (0.14), Suspenders (0.10)
Silhouette	18.54	0.05	18.59	Ilja [일자] (3.12), Bootscut (2.89), Wide (2.59), Slim (1.93), Skinny (1.85), Baggy (1.27), Semi (0.96), Straight (0.89), Fit (0.82), Boyfriendfit (0.51), Loose (0.30) Tong [통] (0.28), Tapered (0.15), Jogger (0.14), Momfit (0.13), Flare (0.12), Napal [나팔] (0.12)
Making	13.03	0.23	13.26	Highwaist (2.27), Banding (1.66), Cutting (0.95), Hidden banding (0.85), Hemline (0.51), Slit (0.58), Rollup (0.44), Ripped jeans (0.38), 2 button (0.36), Pocket (0.33), Side (0.32), Incision (0.30), Loose-Lesstight (0.30), Unbalance (0.30), Highrise (0.26), Waist (0.25), Button (0.24), Point (0.23), Pintuck (0.22), Backbanding (0.22), Stitch (0.18), Half highwaist (0.14), Diagonal (0.14), Back (0.12), Belt (0.11)
Fabric	10.25	0.17	10.42	Napping (3.07), Spandex (1.80), Washing (1.39), Raw denim (0.44), Cotton (0.34), Embroidery (0.23), Cotton flannel (0.22), Demage (0.22), Warm (0.21), Bonding (0.19), Cool (0.19), Mink (0.15), Lining (0.14), Brush (0.11), Blending (0.11), Stone (0.11)
Brand	7.27	1.86	9.13	Levis (0.42), Jambangee (0.30), Guess (0.24), Plac jeans (0.17), CalvinKlein (0.17), Destroyed (0.16), Siero (0.14), Bangbang (0.14), Buckaroo (0.14), 8Seconds (0.14), NIX (0.12), Jeep (0.11), Benetton (0.11), CLRIDE.N (0.10), Indigo Bank (0.10)
Color	5.91	0.05	5.96	Light blue (1.08), Dark blue (0.90), Black (0.52), Midblue (0.50), Blueblack (0.48), Indigo (0.25), Grey (0.21), 2 color (0.18), Blue (0.17), Light color (0.15), Color (0.15), D tone (0.15), M tone (0.14), Ice (0.14), Color block (0.10)
Sex	4.67	-	4.67	Woman (3.79), Female (0.76), Unisex (0.10)
Season	2.87	0.03	2.90	Summer (0.61), Winter (0.47), 2021 (0.42), New item (0.30), Fal (0.29), FW (0.28), Spring (0.27), 2020 (0.12)
Length	2.59	0.02	2.61	Long (0.78), Crop (0.33), Caprants (0.31), 5bu [5부] (0.20), 4bu [4부] (0.16), 3bu [3부] (0.16), Ankle (0.16), Short (0.14), 8bu [8부] (0.13)
Style	1.91	0.06	1.97	Daily (0.36), Vintage (0.32), Secondhand (0.27), Basic (0.19), Comport (0.19), Casual (0.16), Basic (0.13)
Size designation	1.41	0.12	1.53	S (0.26), S-L (0.17), Short girl (0.15), M (0.14)
Sale information	0.98	-	0.98	Store (0.32), Price (0.19), 3 types (0.18), 2 types (0.11)
Age & Others	0.52	0.28	0.80	-
Total	91.57	2.89	94.46	

언발(0.30%), 사선(0.14%) 등이었다. 세 번째로 청바지의 부분적 찢어짐 관련 키워드는 찢침(0.38%), 옆(0.32%) 절개(0.30%) 등이었다. 따라서 의복구성 키워드 분석을 통해 최근 유행하는 허리높이, 착용 편리성과 호칭의 유연성을 위한 허리의 밴딩 구성, 그리고 밑단과 찢어진 부위 등에 관한 소비자 선호를 구체적으로

파악할 수 있었다. 길이 키워드에서도 롱(0.78%)이 크롭(0.33%)이나 9부(0.31%)의 2배 이상으로 나타났으며 이 또한 현재 소비자들의 선호를 보여주는 결과이다.

소재와 색상 키워드는 다음과 같다. 데이터의 수집 시기로 인해 기모(3.07%)가 가장 많았고 용(0.22%),



Fig. 1 Distribution of 56 keywords by information.

웍(0.21%), 본딩(0.19%), 밍크(0.15%)도 다빈도 키워드였다. 그 외 스판(1.80%), 생지(0.44%), 면(0.34%), 혼방(0.11%) 등 소재 특성 키워드와 워싱(1.39%), 자수(0.23%), 데미지(0.22%) 등 가공 특성 키워드가 있었다. 워싱 방법은 브러쉬(0.11%)와 스톤(0.11%), 색상은 연청(1.08%) > 진청(0.90%) > 블랙(0.52%) > 중청(0.50%) > 흑청(0.48%) 순이었다.

3) 호칭 키워드 분석(Size designation keywords)

카테고리별 분류에서 호칭 키워드의 비중은 1.53%로 아이템, 실루엣, 의복구성 등에 비해 작았으나 판매 장소의 3배, 가격의 8배였다. 호칭 정보는 주로 상품의 상세정보에 표시되므로 당연한 결과이다. 그러나 호칭은 의류 상품에서 중요한 키워드(Lee & Lee, 2018)로서 중요한 제품 구매요인이므로(Lee et al., 2021) 상품명에 표시된 호칭의 특징을 상세히 살펴보았다. 전체 2,582개에서 호칭 키워드의 분포는 <Table 5> 1.61%였다.

호칭 키워드의 종류는 1) 호칭, 2) 체형의 2가지로 나뉘었으며 호칭은 숫자 또는 문자로 표시되었고 숫자는 허리둘레치수나 의복 치수, 문자는 S~XL가 사용

되었다. KS K 0051:2019(Korean Agency for Technology and Standards [KATS], 2019)에서 캐주얼웨어로 분류되는 청바지는 5 cm 간격의 허리둘레치수 호칭이 권장되나 실제 상품명에는 문자 호칭이 5배 많았다. 문자 호칭(0.80%) 다음으로 체형 키워드(0.42%), 숫자 호칭(0.23%) 순이었다. 이 결과는 실루엣과 의복구성 카테고리의 키워드 분석 결과와 연계하여 해석할 수 있다. 1) 와이드와 일자 등 여유로운 핏, 2) 밴딩으로 구성된 허리, 3) 스판 소재의 비중이 컸기 때문에 허리둘레치수 호칭보다 착용 가능 범위가 더 넓은 문자 호칭이 많았다고 하겠다.

한편, 호칭의 크기를 비교하면, 문자 호칭과 허리둘레치수 호칭 모두 작은 호칭 > 큰 호칭 > 중간 호칭 순이었다. 문자 호칭은 S(0.26%) > M(0.14%) > 빅사이즈(0.09%) > L(0.08%) 순이나 큰 호칭인 빅사이즈와 L을 합하면 M보다 많다. 허리둘레 호칭도 27 inch와 28 inch가 가장 많고 다음으로 36 > 32 > 38 > 34 > 30 순이었다. 참고로 제8차 사이즈코리아(Size Korea, 2022)의 20대 여성 배꼽수준허리둘레(바지 착용 위치) 평균은 79.2 cm(31.18 inch)이다.

Table 5. Size information included in 2,582 keywords

Character designation	Number designation									Body characteristic			Others		
	Waist size			Clothing size											
Word	N	%	Word	N	%	Word	N	%	Word	N	%	Word	N	%	
S	112	0.26	27	12	0.03	55	10	0.02	Short girl	64	0.15	Size	29	0.07	
S-L	73	0.17	28	7	0.02	77	7	0.02	Regular	32	0.07	3 Size	26	0.06	
M	62	0.14	36	7	0.02	88	7	0.02	Tall girl	26	0.06	Various size	12	0.03	
Big size	41	0.09	38	6	0.01	66	5	0.01	Slim	26	0.06	6 Size	1	0.00	
L	36	0.08	32	6	0.01	55-77	3	0.01	Long leg	14	0.03	1 Size	1	0.00	
Up to 3XL	4	0.01	34	5	0.01	44	2	0.00	Normal stature	12	0.03	2 Size	1	0.00	
S-3XL	3	0.01	30	5	0.01	44-66	2	0.00	Skinny girl	3	0.01				
XL	3	0.01	29	3	0.01	Up to 88	2	0.00	155	2	0.00				
Up tp XL	3	0.01	35	2	0.00	99	1	0.00	Chubby	1	0.00				
S-XL	3	0.01	33	2	0.00	55-88	1	0.00	150	1	0.00				
3XL	2	0.00	24	1	0.00				160	1	0.00				
Up to 2XL	2	0.00	23	1	0.00										
2XL	2	0.00	67	1	0.00										
S-2XL	2	0.00	25-38	1	0.00										
Plus size	1	0.00	26-35	1	0.00										
XS	1	0.00													
Sub total	350	0.80	Sub total	60	0.14	Sub total	40	0.09	Sub total	182	0.42	Sub total	70	0.16	

체형 키워드로는 키와 관련된 키워드가 압도적으로 많았고 특히, 작은 키를 나타내는 키작녀(0.15%)가 큰키(0.06%)나 보통키(0.03%)에 비해 2~4배 많았다. 키를 숫자로 나타낸 경우도 150, 155, 160으로 모두 작은 키였다. 키는 바지길이 맞춤새를 가늠하는 기준이므로 키가 작은 소비자를 타겟으로 길이 수선의 불편함 및 밑단 수선으로 인한 실루엣 및 디테일 변화문제 가 없음을 상품명에 미리 알리기 위함으로 볼 수 있다.

3. 네트워크 분석(Network analysis)

1) N-gram 분석

빈도분석 및 카테고리 분류를 통해 다빈도 키워드의 특징은 파악하였으나 키워드 간 관계를 파악하기 위해 공출현빈도를 분석하였다. 공출현빈도 상위 50쌍을 카테고리별로 분류하였고<Table 6> 상위 키워드

간 관계도를 <Fig. 2>에 시각화하였다. 가장 비중이 큰 아이템 카테고리의 경우, 데님. 청바지, 팬츠, 진의 공출현이 많았으며 대체로 실루엣 키워드와 공출현하였다. 특히, 다빈도 순 1위인 데님은 데님-부츠컷(441), 데님-와이드(409), 데님-일자(389) 순으로 공출현빈도 1~3위에 오르는 동시에 소재, 의복구성, 색상 등 가장 다양한 종류의 키워드들과 공출현하였다. 따라서 데님은 청바지 상품명의 대표 키워드로서 실루엣, 소재, 의복구성, 색상 등 다양한 키워드들과 연계하여 상품 특성을 설명한다고 하겠다. 그 외 진은 스키니(218)와 공출현이 많았다.

실루엣 키워드들은 아이템 키워드 외에 소재, 의복구성 및 핏 키워드와도 공출현했다. 부츠컷과 일자는 기모, 스판텍스와, 와이드와 일자는 하이웨이스트와의 공출현이 많았다. 한편, 다빈도 키워드였던 슬림과 세미의 공출현도 나타났다. 슬림은 일자(215) 및 부츠

Table 6. Co-occurrence keywords by information type (N-gram analysis)

Keywords by category		Co-occurrence word (N) by category					
		Item	Silhouette and fit	Making	Fabric	Brand	Color
Denim	Short Pants (76)	Bootscut (441)	Ilja [일자] (389)	Banding (135)	Napping (281)	Dark blue (87)	
		Wide (409)					
		Highwaist (75)					
	Banbaji [반바지] (72)	Baggy (182)	Skinny (139)	Spandex (100)			
Item	Chungbaji	Ilja [일자] (261)	Bootscut (219)				
		Wide (120)	Baggy (105)				
		Skinny (85)					
		Pants	Bootscut (122)	Wide (85)	Baggy (81)		
	Jeans	Skinny (218)					
Sex	Woman	Chungbaji (141)	Skinny (71)		Napping (89)	Levis (127)	Jambangee (92)
		Denim (95)					
Fabc	Napping	Denim (281)	Bootscut (102)		Cotton flannel (88)		
		Chungbaji (69)	Ilja [일자] (91)				
	Spandex	Denim (100)	Bootscut (73)	Banding (90)	Napping (71)		
	Chungbaji (72)	Ilja [일자] (71)					
	Washing	Denim (119)					
Making	Highwaist	Denim (75)	Wide (72)				
			Ilja [일자] (71)				
	Banding	Denim (135)		Napping (82)			
	Cutting	Denim (72)		Hemline (86)			
Fit	Slim	Ilja [일자] (216)	Bootscut (105)				
		Bootscut (186)	Wide (90)				
		Straight (76)					
	Semi		Wide (88)				
	Fit						
Color	Dark blue	Denim (73)	Wide (96)				
		Denim (87)	Bootscut (70)				
		Black	Jeans (106)				
Year	2021						FW (84)

컷(105)과, 세미는 부츠컷(186) 및 와이드(90)와의 공출현이 많아 핏-실루엣 간 특정 키워드 조합을 확인할 수 있었다. 그 외 커팅-밑단(82)처럼 디테일 특징을 나

타내는 공출현도 있었으며 색상의 경우, 연청과 진청은 데님과, 블랙은 진과 공출현하며 뚜렷한 차이를 보였다. 이처럼 공출현빈도 분석을 통해아이템, 실루엣

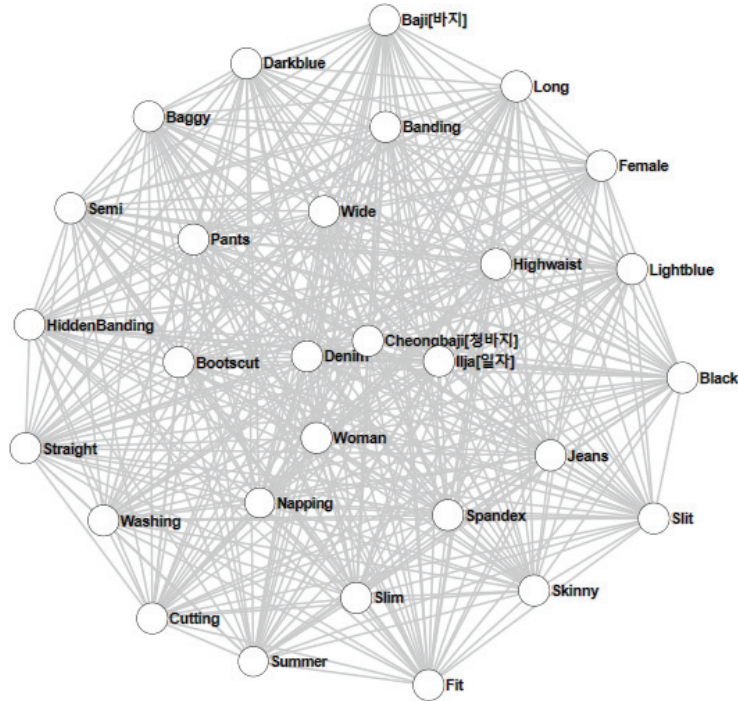


Fig. 2. Co-occurrence relationship between 29 keywords (N-gram analysis).

및 디테일 간의 특정 조합을 파악할 수 있었으며 이 또한 현재 소비자가 선호하는 청바지의 구체적 구성 특징을 나타낸다고 하겠다.

2) 중심성분석(Centrality analysis)

공출현빈도 분석을 통해 다빈도 키워드 조합을 파악하였으나 공출현 키워드들의 중요성과 영향력을 분석하고 중심성을 정량화하기 위해 중심성을 분석하였다. 선행연구의 노드 개수와 빈도분석 결과에 따라 132개(0.05% 이상) 키워드를 분석하였고 상위 32개(0.5% 이상)를 <Table 7>에 제시하였다. 순위는 실제 값에 따라 표시하였다.

연결중심성은 연결된 노드의 개수가 많을수록 값이 크고 중심 단어임을 의미하고 근접중심성은 연결 경로 즉 거리를 나타내므로 값이 크면 직접 연결관계를 의미한다. 매개중심성은 네트워크 구축의 매개자 역할을 의미하며 값이 1에 가까울수록 네트워크에 미치는 영향이 크다. 위세중심성은 연결중심성 값에 가중치를 부여하여 영향력을 더한 값이다. 본 연구에서는 연결중심성과 위세중심성의 순위가 유사하므로 연

결빈도에 따른 중심성 이외에 가중치에 따른 차이는 적다고 하겠다.

키워드별로 살펴보면, 데님은 빈도 순위 1위이며 공출현빈도에서 대표 키워드로 분석된 결과와 마찬가지로 모든 중심성도 1위였다. 데님이 청바지의 대표 키워드임을 다시 입증하는 결과이다. 데님과 더불어 여성과 청바지는 4가지 중심성에서 모두 5위 안에 들었고 일자와 와이드는 모두 10위 안에 들었다. 따라서 데님, 청바지, 여성, 일자, 와이드가 영향력 있는 중요 키워드라 하겠다. 그 외 3가지 중심성에서 10위 안에 든 키워드는 팬츠와 진이었고 2가지 중심성에서 10위 안에 든 키워드는 기모, 부츠컷, 하이웨이스트, 슬림, 스판, 밴딩, 워싱의 7개였다.

중심성 순위가 모두 높은 데님, 여성, 청바지를 제외한 나머지 9개 키워드는 연결중심성이나 위세중심성은 높으나 근접중심성과 매개중심성이 낮거나 또는 그 반대인 경우로 나뉘었다. 먼저 연결중심성과 위세중심성이 더 높은 키워드는 일자, 기모, 부츠컷, 하이웨이스트, 스판덱스였고 반대로 양적 연결보다 근접중심성과 매개중심성이 높은 키워드는 와이드, 슬림,

Table 7. Centrality analysis of the top 32 keywords

Frequency		Degree centrality		Closeness centrality		Between centrality		Eigenvector centrality	
Keyword	Rank	Value	Rank	Value	Rank	Value	Rank	Value	Rank
Denim	1	60.806	1	0.664	1	0.009	1	0.469	1
Cheongbaji [청바지]	2	39.544	2	0.657	4	0.008	4	0.359	2
Woman	3	32.251	3	0.657	5	0.008	5	0.293	3
Ilja [일자]	4	24.449	4	0.645	9	0.008	9	0.265	4
Napping	5	24.384	5	0.635	11	0.007	11	0.255	5
Bootscut	6	19.449	6	0.629	13	0.007	13	0.210	7
Wide	7	18.513	8	0.654	6	0.008	6	0.211	6
Pants	8	16.989	11	0.654	7	0.008	7	0.178	9
Jeans	9	18.433	9	0.664	2	0.009	2	0.161	12
Highwaist	10	19.445	7	0.623	17	0.006	17	0.201	8
Slim	11	16.300	12	0.664	3	0.009	3	0.172	11
Skinny	12	13.156	14	0.626	15	0.007	15	0.128	14
Spandex	13	17.395	10	0.632	12	0.007	12	0.178	10
Banding	14	13.574	13	0.645	10	0.008	10	0.142	13
Washing	15	10.924	15	0.654	8	0.008	8	0.110	15
Baggy	16	9.251	17	0.617	21	0.006	21	0.100	16
Lightblue	17	9.373	16	0.617	20	0.006	20	0.098	17
Semi	18	8.152	21	0.612	22	0.006	22	0.093	18
Cutting	19	8.262	19	0.620	19	0.006	19	0.083	21
Darkblue	20	8.160	20	0.626	16	0.007	16	0.091	19
Straight	21	5.833	26	0.612	23	0.006	23	0.064	25
HiddenBanding	22	6.947	22	0.590	29	0.005	29	0.077	23
Fit	23	6.943	23	0.623	18	0.006	18	0.078	22
Long	24	6.563	24	0.567	36	0.004	36	0.077	24
Female	25	8.471	18	0.629	14	0.007	14	0.083	20
Slit	26	4.909	30	0.595	28	0.005	28	0.052	31
Summer	27	6.251	25	0.590	30	0.005	30	0.059	27
Baji [바지]	28	5.551	27	0.603	26	0.005	27	0.055	28
Black	29	5.084	29	0.609	25	0.006	25	0.052	30
BoyfriendFit	30	4.118	34	0.587	31	0.005	31	0.043	34
Hemline	31	4.513	33	0.600	27	0.006	26	0.044	33
MidBlue	32	4.673	31	0.612	24	0.006	24	0.052	29

밴딩, 워싱이었다. 일자, 기모, 부츠컷의 연결중심성 순위는 빈도 순위와 같이 4~6위로 매우 높았고 하이웨이스트는 빈도 순위보다 높은 7위였다. 그러나 근접중심성과 매개중심성은 일자 9위, 기모 11위, 부츠컷 13위, 하이웨이스트 17위로 연결중심성보다 5~7단계 낮

았다. 따라서 이 키워드들은 다른 키워드와의 양적 연결은 많으나 영향력은 상대적으로 크지 않다고 하겠다. 스판덱스도 같은 경향이었으며 그 외 10위권 밖의 키워드 중 스키니, 배기, 히든밴딩, 룡도 같은 경향을 보였다. 반대로 양적 연결보다 근접중심성과 매개중심

성이 높은 키워드 중 와이드는 일자(4위)나 부츠컷(6위)보다 연결중심성은 7위로 낮으나 근접중심성과 매개중심성은 6위로 일자(9위), 부츠컷(13위)보다 높았다. 따라서 일자나 부츠컷보다 양적 연결은 다소 약하나 다른 키워드와 근거리로 연결되어 매개 키워드 역할을 함으로써 영향력은 더 크다고 하겠다. 슬림 또한 빈도 순위는 11위이나 근접중심성과 매개중심성은 3위로 매우 높았고 워싱과 밴딩도 연결중심성은 10위권 밖이나 근접중심성과 매개중심성은 8위, 10위로 높았다. 따라서 양적 연결보다 근거리 연결과 매개 키워드로서의 역할이 더 큰 키워드라 하겠다.

3) 상관관계 분석(Correlation analysis)

출현빈도 및 중심성분석을 통해 다빈도 키워드 간 관계를 파악할 수 있었다. 그러나 출현빈도와 별도로 특정 관계를 보이는 키워드를 분석하기 위해 761개 키

워드의 상관계수를 분석하였다. 상품명의 나열식 구조 특성을 고려하여 유사도 0.200(낮으나 상관관계 있음)을 기준으로 판단하였다(Table 8).

2021-FW(0.575)의 상관관이 가장 높았고 2~5위는 밉크-용(0.356), 하이라이즈-리바이스(0.322), 커팅-밑단(0.313), 트임-옆선(0.300)이었다. 빈도와 공출현빈도 순위가 매우 낮았던 키워드들이었다. 상관관계를 보인 키워드들의 특성을 보면, 먼저 블랙-진(0.249), 스키니-진(0.223), 통-바지(0.243), 슬림-핏(0.237) 등 흔히 사용되는 관용구적 표현의 상관관이 높았다. 다음으로 의복구성이나 소재의 특성을 구체적으로 설명하는 키워드의 상관관이 높았다. 즉, 커팅-밑단(0.313), 트임-옆선(0.300), 뒤-찢어짐(0.226)처럼 의복 구성 특징을 구체적으로 설명하거나 워싱-스톤(0.231), 워싱-브러쉬(0.206)처럼 방법을 구체화하거나 밉크-용(0.356), 기모-용(0.258)처럼 소재 특성을 강조하는 경우들이

Table 8. Correlation of 132 keywords

Correlation of 32 keywords of 0.5% or more (value)	Category	Other highly correlated keywords (value)
Denim-Cheongbaji [청바지] (0.267) Denim-Jeans (0.256) Baji[바지]-Tong[통] (0.243) Cheongbaji [청바지]-Spandex (0.212) Jeans-Black (0.249) Jeans-Skinny (0.223)	Item	Hotpants-Shortjeans (0.226) Hotpants-Banbaji [반바지] (0.228) ShortJeans-3bu [3부] (0.241)
Slim-Fit (0.237) Bootscut-Semi (0.216)	Silhouette	Tapered-CLRIDE.n (0.200)
Cutting-Hemline (0.313) Slit-Sideseam (0.300)	Making	Highrise-Calvinklein (0.286) Highrise-Levis (0.322) Back-Incision (0.226)
-	Fabric	Mink-CottonFlannel (0.356) Napping-Cotton flannel (0.287) Napping-Winter (0.265) Washing-Stone (0.231) Washing-Jambangee (0.224) Washing-Brush (0.206)
-	Color	Blue-8second (0.231) Mtone-Nix (0.238) Mtone-Guess (0.217) Mtone-Buckaroo (0.208) Dtone-Nix (0.215) Dtone-Guess (0.211)
-	Season	2021-FW (0.575) FW-3types (0.228) Spring-Fall (0.234)
-	Length	Ankle-CalvinKlein (0.241)
-	Size designation	S-M (0.239)

었다. 마지막으로 브랜드별로 사용되는 키워드가 존재했다. 하이라이즈는 리바이스(0.322), 캘빈클라인(0.286)과 높은 상관을 보였고 색상 키워드인 M톤은 게스(0.217), 버커루(0.208), 닉스(0.238), D톤은 닉스(0.215), 게스(0.211), 청은 8세컨즈(0.231)와 높은 상관을 보였다. 그 외 데님-청바지(0.267), 호칭 중 S-M(0.239)의 상관이 높았다.

이처럼 상관분석을 통해 빈도나 공출현빈도는 다소 낮으나 청바지 특성 표현에 흔히 사용하는 관용적 표현, 구체적인 의복 구성법이나 소재 특성을 전달하는 키워드 조합을 파악할 수 있었으며 상품명 구성 시 이런 특징을 활용할 필요가 있겠다.

V. 결론 및 제언

본 연구는 온라인쇼핑에서 구매로 연결된 여성 청바지 상품명에 텍스트마이닝으로 중요 키워드를 추출하고 키워드의 정보와 관계를 분석함으로써 상품명 구성 방안을 위한 기초자료를 제공하고자 하였다. 이를 위해 네이버쇼핑의 여성 청바지 카테고리에서 '리뷰 많은순' 10,000개의 상품명에 수집한 후 데이터클리닝 및 수동정제 과정을 거쳐 2,842개 키워드를 추출하고, 빈도, 공출현빈도, 중심성 및 상관관계를 분석하여 중요 키워드와 다빈도 키워드 간 관계를 분석하였다.

빈도분석결과 1위 키워드는 '데님'이었고 청바지, 여성, 일자, 기모, 부츠컷, 와이드, 팬츠, 진, 하이웨스트, 슬림, 스키니, 스판, 밴딩, 워싱, 배기, 연청 등이 상위 키워드였다. 키워드 정보와 카테고리별 대표 키워드를 파악하기 위해 0.01% 이상인 키워드(761개)를 14개 카테고리로 분류한 결과, 가장 비중이 큰 카테고리는 아이টে이였고 그 외 실루엣, 의복구성, 소재, 브랜드명, 색상의 6개 카테고리가 전체의 80%를 차지했다. 따라서 청바지 상품명은 주로 6개 카테고리의 키워드로 구성되었음을 알 수 있다. 아이টে이 카테고리의 대표 키워드는 데님, 청바지, 진이었고 실루엣의 경우 일자, 와이드, 부츠컷이었다. '일자(기본 실루엣)'에 대한 소비자들의 상시적 관심과 부츠컷 및 와이드에 대한 높은 관심이 나타난 결과라 하겠다. 길이 키워드에서는 롱이 크롭이나 9부보다 2배 이상 많았으며 색상 키워드는 연청 > 진청 > 블랙 순이었다. 또한, 의복구성 키워드에서는 1) 허리높이(하이웨스트, 하이라이즈), 2) 허리구성(밴딩, 히든밴딩, 뒷밴딩), 3) 밑단 디테일(컷팅, 트임, 롤업, 언발), 4) 부분적 찢어짐(찢

청, 옆, 절개) 등 최근 유행하는 구성적 특성을 상세히 파악할 수 있었다.

이 결과를 바탕으로 구매가 진행된 상품명에는 아이টে이 대표 키워드와 동시에 현지점에서 소비자가 선호하는 실루엣, 의복구성, 길이, 색상 등의 특성을 구체적으로 전달하는 키워드가 포함되었음을 알 수 있었으며 이에 상품명 연구의 의의가 있다고 하겠다. 특히 이런 정보들은 검색 수단으로 사용되므로 다빈도 키워드를 사용하는 것이 더 좋은 검색 결과를 가져올 수 있을 것이다. 예로, 스트레이트 대신 일자, 루즈나 통바지 대신 와이드, 밴드 대신 밴딩, 하이라이즈 대신 하이웨스트를 사용한다면 청바지 상품의 노출 빈도는 증가할 수 있을 것이다. 따라서 다양한 아이টে이를 대상으로 상품명 분석 연구를 진행하여 상품명 구성 시 소비자와 판매자의 연결을 도모할 수 있는 키워드를 활용할 수 있도록 자료가 구축될 필요가 있다.

한편, 호칭 카테고리의 비중은 크지 않았으며 호칭 정보가 주로 상품의 상세페이지에 표시되므로 당연한 결과이다. 다만, 의류 호칭은 구매 과정에서 소비자에게 중요한 정보이므로 특징을 살펴본 결과, 체형 키워드와 호칭(문자, 숫자) 키워드로 구분되었으며 키를 나타내는 키워드에서 작은 키를 나타내는 '키작네'가 '큰키', '보통키'보다 월등히 많았다. 키는 바지길이를 가늠하는 기준이므로 키가 작은 소비자들에게 길이 수선 시 발생하는 실루엣(부츠컷) 및 밑단 디테일의 변화가 없음을 알리려는 의도로 볼 수 있다. 한편, 착용대상 범위가 넓은 문자 호칭이 허리둘레치수 호칭보다 월등히 많았으며 이는 와이드, 일자, 배기 등 여유로운 실루엣, 허리의 밴딩 그리고 스판 소재가 많았던 결과가 반영된 것으로 해석할 수 있다.

다음으로 다빈도 키워드 간 관계 분석을 위해 상위 132개 키워드의 공출현빈도를 살펴본 결과, 아이টে이 키워드와 실루엣 키워드 간 공출현이 두드러졌다. 특히, 빈도순 1위인 데님은 부츠컷, 와이드, 일자와 공출현 빈도 1~3위에 오르는 동시에 소재, 의복구성, 색상 등 가장 다양한 종류의 키워드와 공출현하였다. 이는 데님이 청바지 대표 키워드로서 제품 특성을 설명하는 다양한 특성의 키워드와 함께 출현했음을 뜻한다. 또한, 하이웨스트-와이드/일자, 배기-밴딩 등 실루엣과 의복구성 특성을 보여주는 키워드 조합이 두드러졌고 슬림-일자, 세미-부츠컷, 절개-밑단 등과 같이 의복구성 특성을 구체화하는 키워드 조합도 많았다. 따라서 아이টে이, 실루엣 및 의복구성 특징을 전달하는 대

표 키워드는 필수적으로 포함하여 제품에 대한 소비자의 관심을 유도할 필요가 있겠다.

공출현빈도에서 나타난 키워드 조합들의 중요도와 영향력을 정량화하기 위해 연결중심성, 위세중심성, 근접중심성, 매개중심성을 분석하였다. 데님은 빈도와 공출현빈도의 결과와 마찬가지로 중심성도 모두 1위로 나타나 청바지의 대표 키워드임이 재입증되었다. 그 외 청바지, 여성, 일자, 와이드의 중심성이 매우 높아 가장 영향력 있는 키워드들이라 하겠다. 다만, 와이드는 일자에 비해 연결중심성은 다소 낮으나 근접중심성과 매개중심성은 더 높았다. 즉, 양적 연결은 일자보다 다소 약하나 다른 키워드와 근거리로 연결되며 매개 키워드 역할을 담당함으로써 네트워크에서의 영향력은 더 컸다. 슬림, 워싱, 밴딩도 양적 연결보다 근거리 연결과 매개 키워드로서의 역할이 더 컸다. 따라서 데님과 더불어 와이드처럼 현시점에서 소비자가 관심 갖는 키워드가 청바지 상품명의 중심 키워드로서 의복구성 및 소재 키워드들과 근접하게 배치되는 특징을 알 수 있다. 한편, 기모, 부츠컷, 하이웨이스트는 연결중심성이 높으나 근접중심성과 매개중심성은 그보다 5-7단계 낮았다. 양적 연결은 많으나 영향력은 상대적으로 크지 않다고 하겠다.

다빈도 키워드 이외에 밀접한 관계가 나타난 키워드 조합을 상관계수로 분석한 결과, 빈도는 적으나 밉크-용, 커팅-밑단, 트임-옆선처럼 현재 소비자들이 선호하는 소재 및 의복구성을 설명하는 키워드 쌍, 그리고 2021-FW, 블랙-진, 스키니-진, 바지-통, 슬림-핏 등 관용적 표현들의 상관성이 높았다.

종합하면, 온라인 판매 여성 청바지 상품 중 구매로 연결된 상품명에 나타난 정보는 다음과 같은 특징을 갖는다. 대표 키워드인 데님을 중심으로 실루엣, 의복구성, 소재, 브랜드, 색상 정보를 주로 포함하며 특히, 현재 소비자가 선호하는 실루엣과 의복구성 특징에 관한 키워드들이 공출현하는 경향이 뚜렷했으며 아이템(데님) 키워드와 실루엣 키워드가 상호 근거리로 위치하여 소비자가 연달아 정보를 인식할 수 있도록 구성되어 있었다. 더불어 트임-옆선, 밉크-용과 같이 현시점에서 소비자가 관심 있는 의복구성 디테일이나 소재 특성을 구체적으로 나타내는 키워드 쌍도 나타났다. 한편, 상품의 상세페이지에서 확인하는 호칭 키워드는 그 비중이 크진 않았으나 키작녀의 경우 상품명에 체형 정보를 포함함으로써 바지길이 맞춤새를

예상할 수 있도록 편의를 제공하기도 하였다. 그러나 유사한 의미라도 주로 사용하는 키워드가 존재하는 점, 키작녀와 같이 비전문 키워드의 사용, 밴딩처럼 외래어의 특정한 표현방식 등을 활용하여 정보를 전달해야 하는 점은 유념할 부분이다.

본 연구는 상품명 빅데이터를 활용함으로써 기존 설문조사보다 쉽고 빠르게 조사대상을 확보하는 동시에 최근 2년 이내 구매된 상품들의 제품 정보를 구체적으로 파악할 수 있었다. 이를 통해 예측된 트렌드와 별도로 현시점에서 소비자들이 실제로 선호하는 실루엣, 의복구성 및 소재 특성을 파악할 수 있었으며 동시에 청바지 상품명에 포함되어야 할 정보의 종류와 사용되는 용어의 특징을 구체적으로 파악할 수 있어 이에 연구의 의의가 있다고 하겠다. 더불어 단어나 열형 빅데이터를 활용한 의류 제품 정보 분석의 사례를 제시함으로써 추후 다양한 상품명과 제품 정보 분석 연구의 참고 자료로 활용될 수 있을 것이다. 다만, 네이버 쇼핑목록에 국한한 점, 데이터 수집 시기 및 추출 방법에 따른 한계점이 있으므로 연구결과를 일반화하기에는 무리가 있다. 이후 상품명 관련 연구가 활성화되어 상품명에 포함된 정보 종류, 사용 용어 및 구성 방법의 기준이 마련되어 소비자와 판매자가 쉽고 효율적으로 소통할 수 있는 환경이 조성되기를 희망한다.

1. 사사

해당사항 없음

2. 연구윤리

해당사항 없음

3. 데이터 및 자료 가용성

본 연구에 사용된 데이터는 연구기간에 수집하여 연구자가 수동정제하였으므로 공개가 불가함.

4. 이해관계 상충

저자 YSK는 2015년 4월부터 본 학술지의 편집위원이지만, 편집과정에 전혀 참여하지 않아 이해관계 상충 문제가 없음.

5. 연구비 지원

해당사항 없음

6. 저자의 기여

본인이 연구 전체를 담당하였음.

7. 저자 정보

강여선 덕성여자대학교 의상디자인전공, 교수

References

- An, H., & Park, M. (2018). A study on the evaluation of fashion design based on big data text analysis-Focus on semantic network analysis of design elements and emotional terms-. *Journal of the Korean Society of Clothing and Textiles*, 42(3), 428–437. doi:10.5850/JKSC.2018.42.3.428
- 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
- Han, K. h. (2021). A convergence analysis study on online review of short-sleeved T-shirt. *The Korean Society of Science & Art*, 39(4), 541–555. doi:10.17548/ksaf.2021.09.30.541
- Hong, M., & Song, H. (2015) A study linguistic features and sociocultural meaning of food product names. *The Journal of Linguistic Science*, 72, 417–438.
- Internet Trend. (2022, September 22). 검색엔진 [Search engine]. *Internet Trend*. Retrieved from <http://www.internettrend.co.kr/trendForward.tsp>
- 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
- Jeong, J.-W., & Lee, J.-S. (2016) An analysis on the product name for the domestic cosmetics brands. *Journal of the Korea Fashion & Costume Design Association*, 18(1), 91–103.
- Kim, G.-A., Kim, C.-K., & Kim M. J. (2013) An analysis of the problem for providing product information in internet shopping mall. *Journal of Korea Design Forum*, 38, 387–397. doi:10.21326/ksdt.2013..38.035
- Kim, J.-H., & Lee, J.-M. (2018). Comparison and analysis of domestic and foreign sports brands using test mining and opinion mining analysis. *The Journal of the Korea Contents Association*, 18(6), 217–234.
- Kim, J.-S. (2005). 독일 광고에 나타난 상품명명의 분석과 네이밍 전략 [Analysis of product names and naming strategy in German advertisements]. *Koreanische Zeitschrift fuer Deutschunterricht*, 32, 89–110.
- Kim, K. O. (1985). *Linguistic analysis of commodity name* (Unpublished master's thesis). Yonsei University, Seoul. Yonsei Education Graduate School, Seoul, Korea.
- Kim, M. H. (2022, July 5). ‘날마다 ‘6천종 신제품’ Z세대 홀린 중국 패스트패션의 비밀은? [6,000 new products' everyday, What is the secret of China's Fast-fashion that attracted generation Z?]. *The Hankyoreh*. Retrieved from https://www.hani.co.kr/arti/international/international_general/1049587.htm
- Kim, S. R. (2020). A convergence study on the perception of anorak fashion using big data analysis. *The Korean Society of Science & Art*, 38(4), 43–55. doi:10.17548/ksaf.2020.09.30.43
- Korean Agency for Technology and Standards. (2019, July 26) KS K 0051:2009 Sizing systems for female adult's garments. *Korean Standards & Certifications*. Retrieved from <https://standard.go.kr/KSCI/standardIntro/getStandardSearchList.do>
- Korean Agency for Technology and Standards. (2022, January 20). 인체 항목 검색 > 측정 항목 검색 [Body item search > Measurement item search]. *Size Korea*. Retrieved from <https://sizekorea.kr/human-meas-search/human-data-search/meas-item>
- Kuk, S. Y. (2010) *A study on the classification of confectionary brand names* (Unpublished master's thesis). Korea National University of Education, Cheongju.
- Lee, A., Lee, J. H. (2018). A study of perception of golfwear using big data analysis. *Fashion & Textile Research Journal*, 20(5), 533–547. doi:10.5805/SFTI.2018.20.5.533
- Lee, A. L., & Kim, H. E. (2020). The sizing communication in online apparel retail websites - Focusing on ready-to-wear women's tailored jacket -. *Fashion & Textile Research Journal*, 22(5), 617–627. doi:10.5805/SFTI.2020.22.6.617
- Lee, J.-H., Lee, J.-M., Kim, W.-K., & Kim, H.-G. (2017). A study on perception of swimsuit using big data text-mining analysis. *Korean Journal of Sport Science*, 28(1), 104–116.
- Lee, H., Nam, Y. J., & Choi, Y. L. (2009). Questionnaire study for commercialization of souvenir T-shirts. *Journal of the Korean Society for Clothing Industry*, 11(4), 614–620.
- Lee, M., Chung, K., & Park, J. (2021) Big data analysis of global sportswear brands. *Journal of the Korean Society of Costume*, 71(5), 96–116. doi:10.7233/jksc.2021.71.5.096
- Shim, J.-Y., & Kim, S. Y. (2008) A study for fashion product name of BEST 100 in cyber shopping mall. *The Research Journal of the Costume Culture*, 16(5), 955–962. doi:10.29049/rjcc.2008.16.5.955
- Statistics Korea. (2022, February 3). 2021년 12월 및 연간 온라인쇼핑동향(2021년 4/4분기 온라인 해외 직접 판매 및 구매 통계 포함 [December 2021 and annual online shopping trends (including online overseas direct sales and purchase statistics for Q4 2021)]. *Statistics Korea*. Retrieved from https://kostat.go.kr/portal/korea/kor_nw/1/1/index.board?bmode=read&aSeq=416587
- TEXTOM. (2017, May 2). 텍스트 이용전, 알아두면 좋은 데이터 분석 용어 [Before using TEXTOM, useful data analysis terminology to know]. *Big Data Assistant TEXTOM Blog*. Retrieved from <https://blog.naver.com/textom/220997038623>