

경영전문석사학위 논문

사직서 빅데이터 분석을 통한
이직 및 퇴사 원인 연구
- 콜센터 근무자를 대상으로 -

2022년 2월

서울과학종합대학원대학교

김형백

경영전문석사학위 논문

사직서 빅데이터 분석을 통한
이직 및 퇴사 원인 연구
- 콜센터 근무자를 대상으로 -

2022년 2월

서울과학종합대학원대학교

김 형 백

사직서 빅데이터 분석을 통한 이직 및 퇴사 원인 연구
- 콜센터 근무자를 대상으로 -

지도교수 장 중 호

이 논문을 경영학 석사 학위논문으로 제출함
2022년 2월
서울과학종합대학원대학교
김 형 백

김형백의 석사 학위논문을 인준함
2022년 1월

위 원 장 _____ 김 보 영 _____ (인)

위 원 _____ 문 달 주 _____ (인)

위 원 _____ 장 중 호 _____ (인)

초 록

본 연구의 목적은 콜센터 근로자들의 사직서(辭職書)를 빅데이터 분석 기법을 통해 분석하여 이직 및 퇴직의 원인을 살펴보는 것이다. 이에 본 연구에서는 국내 대표 콜센터 업체의 8년간 52,212건의 방대한 사직원 데이터를 대상으로 연구 분석하였다. 사직서의 '구제적인 퇴사 사유' 필드를 대상으로 한 본 연구에서 분석 방법은 빅데이터 텍스트 마이닝 기법인 '단어 빈도 분석', '네트워크 분석', 'LDA 토픽 모델링 분석'을 수행하였고, 자료의 특성에 따라 표본을 전체, 성별, 코로나 전후별, 근속기간별, 연령별로 분류하여 진행하였다. 본 연구의 의의는 각 분류별로 3가지 텍스트 마이닝 기법을 모두 적용해 분석함으로써 오해석을 최대한 방지하고 각 분석 기법의 단점을 보완하려고 노력하였다는 점이다. 본 연구의 결과를 요약하면 다음과 같다.

첫째, 남녀 성비는 남자 30%, 여자 70%의 비율을 보였다. 또한 근속 1년 미만인 경우가 76%에 달해 기존 연구 결과보다 훨씬 높은 비율을 보였으며 연령대는 20대, 30대 40대가 고른 분포를 보인 것으로 나타났다.

둘째, 분석 대상 전체를 대상으로 추출한 퇴직 사유의 단어 빈도의 상위 10개 단어를 살펴보면, '개인', '업무', '이직', '계약', '종료', '건강', '무단', '결근', '적용', '만료'이다.

셋째, 성별 분석 결과에서는 남자는 이직을 위한 퇴사 비율이 높았고 여자는 업무 적응을 이유로 퇴사하는 경우가 높았다. 여자의 경우 업무 적응 외에도 '건강'상의 이유로 이직하는 경우가 많이 매핑되고 있으며 특히 '간호', '집안', '자녀', '육아' 관련 키워드가 각 토픽에 많이 등장하는 특징을 보였다.

넷째, 코로나 사태 이후(2020년도 이후)에는 퇴사 사유에 '코로나'가 직접 언급되기 시작했으며, 코로나 이전과 비교했을 때 이직을 위한 퇴사의 비율이 큰 폭으로 줄어 코로나 사태 이후 실업률 증가 등과 맞물려 자연스럽게 이직에 대한 의도가 줄어들었음을 확인할 수 있었다. 특히 주목할 점은 코로나 사태 이후에는 대부분의 토픽에서 '스트레스'와 '치료' 키워드가 포함된다는 것이다.

다섯째, 1년 이상 근속자의 경우 이직을 위한 퇴사의 비율이 급격히 높아지는

것을 확인할 수 있었고, 건강상의 이유로 퇴사하는 경우가 많아지는 경향을 파악하였다.

여섯째, 연령별 분석에서는 각 연령대에서 확연하게 차이를 보였다. 20대의 경우 안정적 취업을 위한 퇴사의 비율이 높고, 30대의 경우 다른 연령대에 비해 ‘이직’을 위한 퇴사 비율이 가장 높았으며 ‘집안’, ‘부모님’, ‘육아’ 문제가 매우 중요한 주제로 파악됨으로써, 여성 근로자가 30대에 접어들어 결혼, 출산 등을 통해 육아 문제로 이어져 퇴사로 이어지는 패턴으로 파악되었다. 40대도 ‘적성’에 따른 ‘업종’과 ‘근무’의 어려움을 여전히 호소하고 있으며, ‘자녀’와 ‘육아’의 부담에 따른 퇴사가 여전히 높은 비중을 차지하고 있음을 알 수 있었다.

분석 결과 기존 이직 관련 연구의 선행요인들과 일치하는 변수들을 파악하였으며, 각 분류별 뚜렷한 퇴사 사유와 흐름을 파악할 수 있었다. 따라서, 콜센터 상담원들의 고용 안정과 기업의 인재유지를 위해 본 연구에서 분류한 인구통계학적 분류별로 유효한 정책들을 수립하고 시행할 필요가 있다. 본 연구에서 도출된 결론을 바탕으로 기업의 인사 및 교육 정책을 수립한다면 콜센터 근무자들의 잦은 이직을 줄이고 근속 기간을 연장하는데 긍정적인 영향을 미칠 것으로 예상된다.

- 주제어 : 사직원 빅데이터 분석, 빈도 분석, 네트워크 분석, LDA 토픽 모델링 분석, 콜센터, 퇴사 사유, 이직 사유

목 차

제 I 장 서 론	1
제 1절 연구 배경 및 목적	1
제 2절 논문의 구성	3
제 II 장 이론적 배경	4
제 1절 이직 및 퇴사관련 질적연구	4
제 2절 이직 및 퇴사관련 빅데이터 연구	9
2.1 빅데이터 분석	9
2.2 이직 및 퇴사관련 텍스트 마이닝 연구	11
제 3절 콜센터 근무자의 이직에 관한 연구	13
제 4절 텍스트마이닝(빈도,네트워크분석,토픽모델링)	14
4.1 빈도 분석	14
4.2 네트워크 분석	14
4.3 LDA 토픽 모델링 분석	16
제 III 장 연구 절차 및 방법	18
제 1절 연구 대상 및 절차	18
제 2절 연구 방법	19
2.1 데이터 수집	19
2.2 데이터 전처리	19
2.3 빈도 분석	20
2.4 네트워크 분석	20
2.5 LDA 토픽 모델링 분석	20

제 IV장 연구 결과	21
제 1절 표본 전체 분석 결과	21
제 2절 성별 분석 결과	25
2.1 남자	25
2.2 여자	29
제 3절 코로나 전후 분석 결과	33
3.1 코로나 이전	33
3.2 코로나 이후	36
제 4절 근속년수별 분석 결과	40
4.1 1년 미만 근속자	40
4.2 1년 이상 근속 근로자	44
제 5절 연령대별 분석 결과	48
5.1 20대	48
5.2 30대	52
5.3 40대 이상	55
제 V장 결론 및 시사점	60
제 1절 연구 결과의 요약	60
제 2절 연구의 논의 및 시사점	62
제 3절 연구의 한계점 및 연구 방향	62

표 목 차

<표 1. 연구대상의 통계학적 특성>	18
<표 2. 전체 표본의 퇴직 사유 빈도 분석 결과>	21
<표 3. 전체 표본을 대상으로 토픽 모델링 결과>	23
<표 4. 남자 직원의 빈도 분석 결과>	26
<표 5. 남자 직원의 토픽 모델링 결과>	27
<표 6. 여자 직원의 빈도 분석 결과>	30
<표 7. 여자 직원의 토픽 모델링 결과>	31
<표 8. 코로나 이전의 빈도 분석 결과>	33
<표 9. 코로나 이전 퇴사자의 토픽 모델링 결과>	35
<표 10. 코로나 이후 빈도 분석 결과 >	37
<표 11. 코로나 이후 퇴사자의 토픽 모델링 결과>	38
<표 12. 1년 미만 근속자의 단어 빈도 분석 결과>	41
<표 13. 1년 미만 근속자의 토픽 모델링 결과>	42
<표 14. 1년 이상 근속자의 단어 빈도 분석 결과>	44
<표 15. 1년 이상 근속자의 토픽 모델링 결과>	46
<표 16. 20대의 빈도 분석 결과>	49
<표 17. 20대의 토픽 모델링 결과>	50
<표 18. 30대의 단어 빈도 분석 결과>	52
<표 19. 30대의 토픽 모델링 결과>	53
<표 20. 40대의 빈도 분석 결과>	56
<표 21. 40대의 토픽 모델링 결과>	57

그림 목 차

<그림 1. LDA 모델의 데이터 처리 과정>	17
<그림 2. 연구 진행 절차>	19
<그림 3. 전체 표본에 대한 네트워크 분석 결과 >	23
<그림 4. 전체 표본 대상의 토픽간 유사도 거리 맵>	25
<그림 5. 남자 직원에 대한 네트워크 분석 결과>	27
<그림 6. 남자 직원의 토픽간 유사도 맵>	29
<그림 7. 여자 직원에 대한 네트워크 분석 결과>	31
<그림 8. 여자 직원의 토픽간 유사도 맵>	32
<그림 9. 코로나 이전의 네트워크 분석 결과>	34
<그림 10. 코로나 이전 퇴사자의 토픽간 유사도 맵>	36
<그림 11. 코로나 이후 네트워크 분석 결과>	38
<그림 12. 코로나 이후 퇴사자의 토픽간 유사도 맵>	40
<그림 13. 1년 미만 근속자의 네트워크 분석 결과>	42
<그림 14. 1년 미만 근속자의 토픽간 유사도 맵>	44
<그림 15. 1년 이상 근속자의 네트워크 분석 결과>	46
<그림 16. 1년 이상 근속자의 토픽간 유사도 맵>	48
<그림 17. 20대의 네트워크 분석 결과>	50
<그림 18. 20대의 토픽간 유사도 맵>	51
<그림 19. 30대의 네트워크 분석 결과 >	53
<그림 20. 30대의 토픽간 유사도 맵>	55
<그림 21. 40대의 네트워크 분석 결과>	57
<그림 22. 40대의 토픽간 유사도 맵>	59

제 I 장 서 론

제1절 연구배경 및 목적

전 세계적으로 코로나19(COVID-19) 바이러스의 유행이 급격한 속도로 확산되고 장기화되면서 기존의 일상과 경제활동에 전면적인 변화에 직면하게 되었다. 이러한 현상에 따라 국내에서도, ‘사회적 거리두기’를 도입하여 사람 간의 접촉을 줄이는 대책을 강화하였고, 생활 속 거리두기로 전환한 후 접촉을 통한 감염병의 확산을 막기 위해 외출자제, 재택근무 등을 시행하며 전국적으로 참여하고 있다. 이에 따라, 외출을 자제하는 분위기가 형성되면서 소비자의 소비 패턴 또한 변화하고 있다. 기존의 오프라인 중심의 소비는 빠르게 디지털화로 전환되고 있으며, 최근 소비자의 소비 패턴은 비대면 소비로 확산되는 추세이다. 이에 따라 비대면 채널을 활용한 상담 수요가 늘어나면서 콜센터 산업을 지속적으로 성장하고 있으며, 최근에는 AI 기술을 접목한 인공지능 콜센터(AICC¹⁾) 시장도 경쟁이 가속화되고 있다.

국내 콜센터는 1990년대부터 소비자 중심의 시장변화, 소비자의 다양한 욕구 충족 및 고객관리 차원의 마케팅 필요성이 부각되면서 등장하기 시작하였다. 최근에는 정보통신 기술의 발전과 함께 효율적인 고객관리에 중요한 채널관리 차원의 극대화로 인해 비용 효율성을 추구하는 대표적 산업으로 성장하고 있다(Lee & Cho, 2019). 조사기관마다 차이가 있지만, 국내 콜센터 및 텔레마케팅 서비스업체는 2015년 기준 약 4만여 개, 관련 업종 근로자도 약 100만 명, 이중 콜센터와 텔레마케팅 종사자는 약 40만 명 수준으로 추정되고 있다(The Ministry of Employment and Labor, 2017). 최근 4차 산업혁명 시대의 도래로 향후 사라질 일자리로 전망됨에도 불구하고 관련 종사자 수는 2.2% 수준으로 꾸준히 증가할 것으로 전망되고 있다.

콜센터 종사자는 상대적으로 고용 불안정성이 높은 계약직이나 외주 용역 형식으로 운영이 이루어지고 있으며, 콜센터 업무는 대표적인 감정노동이 높은 직군임

1) Artificial Intellectual Call Center : 인공지능 기술이 적용된 콜센터

과 동시에 열악한 노동환경으로 인해 양질의 서비스로 전환되지 못하고 있는 실정이다. 즉 콜센터 종사자들은 주당 약 40시간제의 근무 형태를 취하지만, 불안정한 휴식시간이나 전자감시 등으로 장시간 노동에 상응하는 고강도 수준의 업무를 하고 있으며, 여성, 비정규직, 저임금 등으로 대표된다(The Ministry of Employment and Labor, 2017). 또한 감정노동이 수반되며, 컴퓨터를 이용한 전자감시 기반 성과 모니터링제 등 콜센터 종사자들은 불안정한 고용구조, 열악한 고용환경, 장기간 대고객 서비스로 인한 스트레스 증가 등으로 인해 이직율도 약 16%로 높게 나타나고 있다.

특히, 근속기간이 3년 미만에서 이직률은 약 45% 수준인 것으로 조사되고 있다(The Ministry of Employment and Labor, 2017). 이로 인해 콜센터 산업의 인력 수급의 불균형과 함께, 이로 인한 채용 소요비용의 증대, 양질의 서비스 품질 유지의 어려움, 나아가 기업의 이미지에도 타격을 입힐 수 있다. 따라서, 콜센터 운영의 핵심 요소 중 하나는 우수한 상담원을 확보하고 유지하는 것이다. 즉 콜센터의 인적자원인 상담원의 모집 및 선발에서부터 교육 훈련을 통한 인적자원의 개발, 평가 및 경력 관리에 이어 인적자원의 유지를 위한 퇴사 및 이직 관리 등 인적자원관리가 콜센터 운영의 성공 여부가 달려있다고 해도 과언이 아닐 것이다.

국외에서 시작된 이직연구는 초기 Mark와 Simon(1958)의 조직균형이론(theory of organizational equilibrium)을 시작으로 기대충족모형(Met-Expectation model; Porter, Steers, 1973), 매개 연쇄모형(Intermediate linkages model; Mobley, 1977) 등 기존 모형에 다양한 변수를 추가하는 방식으로 발전되었으며, 이후 모형의 통합을 시도하거나 직무배태성(job embeddedness) 및 촉발사건(shok) 등을 검증하여 기업의 인재손실에 대한 개입방안을 모색하는 궁극적 목표를 지닌 채 약 60년간 지속되고 있다.(Koh, Goh, 1995; Lee, Mitchell, 1994)

국내 이직 연구들 대부분은 직무 스트레스나 업무요구(job demands) 변수로 인해 직무 만족(job satisfaction) 및 조직몰입(organizational commitment) 등과 같은 직무·조직에 대한 긍정적 태도 수준이 낮아져 이직 의도가 증가하고, 실제 이직 행동으로 이어지는 논리를 기본으로 하고 있다(Eo, 2015; Jo, Kim, 2018; Kim, Lee, Lee, 2019; Oh et al., 2007). 이러한 태도 변수와 이직 의도 간 관련성을 검증한 단

일연구들 외에도 개별적 연구들을 통합하여 검증한 메타 연구가 다양한 업종을 대상으로 수행되었고, 기관에서 해마다 반복 측정되어 누적된 대규모 패널을 이용한 여러 종단연구가 수행되었다.(Kim, Jyung, 2010; Lee, 2008; Lee, Kang, 2018; Lee et al., 2014; Lee et al., 2016; Moon, 2015)

최근 주목할 만한 연구 동향은 국외에서 가져온 변수에 한정된 이직요인들에서 벗어나기 위한 다양한 시도를 하고 있다는 점이다. 즉, 이직이란 개인의 주관적 경험과 환경적 요인이 결합된 복합적인 의사결정과정이기 때문에 이러한 이직 행동의 원인에 직장이나 업무와는 상대적으로 거리가 멀거나 무관한, 가족이나 대인관계, 거주지 등 다양한 결정 요인이 존재하며, 이러한 변수들은 국내의 노동시장과 사회적 특성이 반영되어 있다. 따라서 이러한 변수들을 추가하여 이직모형을 확장하기 위한 다양한 탐색연구가 수행되었으며, 이러한 연구들은 주로 질적 연구방법론을 취하고 있고, 그 외 웹상의 플랫폼을 대상으로 텍스트 마이닝 기법을 이용한 몇몇 빅데이터 연구도 시행되었다(Eun, Koo, Jung, 2018; Jung, Suh, 2019; Kang, 2016; Kim, Kim, Kim, 2020; Lee, Kim, Kang, 2017; Lee, Kim, Seung, Yoo, 2019)

이에 본 연구에서는 국내 대표 콜센터 업체의 8년간의 방대한 사직원 데이터를 다양한 빅데이터 텍스트 마이닝 기법을 적용해 분석해 이직 및 퇴사의 원인을 연구해보고자 한다. 기존의 이직 연구와는 달리 퇴사 당사자 및 관리자가 작성한 ‘구체적 퇴사 사유’를 분석함으로써 정형화된 설문 분석의 한계에서 벗어나 실질적인 의도를 파악하고자 노력하였다. 이를 통해 콜센터 산업의 인력 수급의 불균형, 채용 소요비용에 대한 부담을 줄이고 나아가 고용 안정을 통해 양질의 서비스 품질을 유지하기 위한 인적자원 방향을 제시하고자 하였다.

제2절 논문의 구성

본 연구는 총 5장으로 구성되며, 각 장의 내용은 다음과 같다.

제 1장 서론에서는 연구의 배경 및 목적, 논문의 구성 등에 대해 서술하였다.

제 2장 이론적 배경에서는 이직 및 퇴사 관련 선행 연구를 질적 연구와 빅데이터 연구로 구분해서 살펴보고, 특히 본 연구의 주제에 해당하는 콜센터 근무자의

이직 연구에 대해서 살펴보았다. 또한 본 연구의 주된 방법론에 해당하는 빅데이터 분석에 대한 이론적 배경을 살펴보고 빈도 분석, 네트워크 분석, 토픽 모델링 등 본연구에 활용된 텍스트 마이닝 이론을 정의하였다.

제 3장 연구 절차 및 방법에서는 본 연구에서 사용할 데이터의 수집과 전처리 및 텍스트 마이닝 분석 방법을 기술하였다. 전체 수집 데이터에서 의미있는 사직원 52,212건을 선별하고 수동 정제 작업을 수행했으며, 기본적인 형태소 분석으로 토큰을 생성하는 방법을 서술하였다. 본 연구에서 정한 6가지 분류 체계(‘전체’, ‘성별’, ‘코로나 전후’, ‘근속기간(1년)별’, ‘연령별’)에 따라 각각 빈도 분석, 네트워크 분석, LDA 토픽 모델링 분석에 대해 기술하였다.

제 4장 연구 결과에서는 6가지 분류별로 시행한 텍스트 마이닝 분석 결과를 각각 확인하여 기술하였다. 첫째, 빈도 분석 결과는 상위 30개의 단어를 확인하고 특징을 분석하였다. 둘째, 네트워크 분석은 데이터 전처리 과정에서의 형태소 분석, 명사 추출 등 토큰화 과정에서 파편화될 수 있는 단어의 연결 고리를 찾고 정확한 해석을 위해 수행했으며 이를 그래프로 시각화해서 단어 연관성을 확인했다. 셋째, LDA 토픽 모델링은 퇴직 사유 문장 전체를 대상으로 토픽 모델링함으로써 주제를 유추 및 파악함으로써 숨겨진 의미를 찾아 그래프 및 표로 시각화하였다. 이와 같은 다양한 텍스트 마이닝 방법들을 결합해서 결과를 해석함으로써 데이터 분석의 정확도를 높이고 숨겨진 의미까지 파악하는 노력을 기울였다.

제 5장 결론 및 제언에서는 연구의 분석 결과에 대한 요약과 시사점을 도출하였고, 연구가 가지는 한계점과 앞으로의 연구 방향에 대하여 서술하였다.

제 II장 이론적 배경

제1절 이직 및 퇴사관련 질적연구

질적 분석방법을 사용하여 직장인들의 이직요인을 탐색한 연구는 매년 약 2편씩 출판되고 있다. 양적 연구보다 상대적으로 편수는 적은 편이며, 주로 연구대상

과의 면담을 시행 후 자료에 대한 주제와 영역을 분류한 현상학적 연구(phenomenological research), 내러티브 연구(narrative research), 합의적 질적 연구(consensual qualitative research)이다. 또한, 특정 성별이나 연령, 직종을 대상으로 한 연구들에서부터 다양한 직종의 이직요인을 통합적으로 도출한 연구에 이르기까지 자료의 형태가 다양하며, 이 중 간호사 직종을 대상으로 한 연구가 우세한 편이다(Kim, Kim, Kim, 2020; Hur, Paek, 2012; Jang, 2008).

면담 과정에서 사용하는 주된 질문은 ‘이직을 고려하고 있는 이유’ 및 ‘과거 다니던 직장을 그만둔 이유’, 그리고 ‘이직을 하지 않고 직장에 남아있는 이유’ 세 가지로 구분할 수 있으며, 이 중 잔류하는 이유의 경우 직무배태성의 개념을 적용한 것이다. 직무 배태성(job embeddedness)은 직장인들이 직장을 떠나는 이유에 초점을 맞추기보다 직장에 남아있는 이유를 포괄적으로 측정하는 변수이며, 최근 국내에서 질적 연구를 통해 이직 사유와의 차별성 및 이직에 대한 영향력이 확인되었고, 지속적으로 양적 연구들을 통해서 효과가 검증되고 있다(Eun, Koo, Jung, 2018; Lee et al., 2019; Lee et al., 2014).

위와 같이 기존 질적 연구에서 이직의사를 지닌 이유, 이직사유, 잔류사유로 구분되는 면담을 통해 도출된 변수들을 살펴보면 다음과 같다. 우선 첫째로, 심층적 탐색연구라 할지라도 기존 양적 연구에서 검증되던 변수들이 유사하게 도출된다. 예를 들어, 직무 관련 변수로는 직무만족 및 만족의 원인이 되는 변수들이 가장 빈번하게 도출되었는데, 다양성(variety) 및 통제감(autonomy)과 같은 직무특성이론(job characteristic theory; Hackman, Oldham, 1975) 내 변수들, 그리고 적성, 흥미 등을 들 수 있다. 또한, 직무스트레스의 원인이 되는 양적·질적 과부하(work overload)와 함께 개인이 맡은 역할과는 무관한 잡무를 발생케 하는 비합리적 업무체계나 역할분담 등도 빈번하게 도출되었다(Eun et al., 2018; Jeon, Wang, 2014; Lee et al., 2019).

다음으로 조직과 관련된 변수로는 주로 상사 및 동료와의 갈등과 같은 사내 대인관계가 대부분을 차지하며, 개인의 업무부진, 보상 및 승진을 결정하는 인사제도, 업무현장의 분위기 등과 연관됨과 동시에 부정적으로 인식되는 조직문화와 2078 이종현, 이재은 조직정치(organizational politics) 등이 있다(Eun et al., 2018; Jang,

Chang, 2008; Kim et al., 2020; Lee et al., 2019). 또한, 업무수행 및 재직경험을 통해 얻어지는 개인의 경력(career)도 상당한 비율을 차지한다. 연구에서는 경력설계의 가능성이나 한계, 혹은 업무에서 얻어지는 전문성 등으로 이직과 잔류가 결정되며, 국내에서 검증된 양적 연구들에서는 빈번하게 나타나지만, 국외 최신 메타연구 등과 비교해봤을 때 국내 노동시장의 특성 및 미래에 대한 고용불안이 반영된 차별화된 변수라 볼 수 있다(Rubenstein et al., 2018; Lee et al., 2019). 그 외 재직 중인 직장의 인지도가 있다. 이는 잠재고객이 평가하는 기업 이미지나 재직 중인 종업원들에 의해 지각된 기업의 외적 명성(perceived external prestige; Mael, Ashforth, 1992)으로 볼 수 있으며, 해당 변수는 현재까지 국외 연구에서 이직 의도나 이직의 선행변수로 빈번히 연구되고 있으나, 국내에서 이직과의 관련성을 검증한 사례는 적은 편이다(Herrbach, Mignonac, Gatignon, 2007; Kamasak, 2019).

두 번째로는 인구통계변수와 관련된 영역이다. 우선 연봉 및 인센티브 등 금전적 변수가 가장 많은 빈도를 차지하며, 그 외 앞서 언급한 조직문화 중 성차별적 조직문화나 상사의 성희롱과 같은 특정 성별과 관련된 변수도 존재한다(Lee et al., 2019). 과거 직장인들의 이직사유가 직무스트레스나 대인관계, 기업 비전이었던 것에 비해 급여나 연봉은 약 10년 전부터 매년 직장인 설문조사에서 이직사유로 최상위권을 차지하고 있다(Job Korea, 2020.04.27.; Lee et al., 2014). 따라서 직무·조직 태도변수가 많은 연구를 통해 이직의 선행요인으로 검증되었다 할지라도 급여와 관련된 인구 통계학적 변수는 이에 견줄 만큼 영향력이 높아져 있어 연구에 제외되어서는 안 될 중요한 변수이다. 마지막으로 주목할만한 내용은 기존 양적 연구에서 드물게 확인되거나 이전에 검증되지 않았던 새로운 변수들이다.

첫째, 직무 관련 변수는 양적 연구에서 검증되었던 직무특성 및 적성이나 흥미 외에도 원치 않은 업무(직무배치), 업무교육의 부재, 직무 효능감이 나타났다(Kim et al., 2020; Eun et al., 2018; Jeon, Wang, 2014; Lee et al., 2019; Shin, Kim, 2016). 직무배치는 개인이 지닌 기술 및 전공 적합성과 관련된 변수이다(Cho, Lee, 2020; Robst, 2007). 이러한 직무배치와 더불어 업무교육은 이전까지는 주로 조직의 투자비용 및 성과, 직무만족과의 관련성이 검증되던 변수인 반면, 질적 연구에서는 개인의 전문성 및 경력과의 연결성에 의해 이직과 잔류가 결정된다. 또한, 직무효

능감의 경우 Bandura(1986, 1997)의 이론에 따르면, 사회인지이론을 근거로 일반적 자기효능감과 과제 특수적 자기효능감(task specific occupational self-efficacy)으로 구분되며, 특정 과업이나 활동들을 성공적으로 수행할 수 있다는 능력에 대한 신념인 직무 효능감(vocational self-efficacy) 개념도 존재한다(Gardner, Pierce, 1998). 그러나 기존의 직무효능감이 업무요구에 대응할 수 있는 스스로 평가한 능력에 초점이 맞춰진 반면, 질적 연구방법론을 통해 도출된 직무 효능감은 개인의 업무 정체성(task identity) 및 수행의 가치를 주변에서 인정하는지 여부 및 사내평판 등에 의해 결정되는 조직기반 자긍심(organizational based self-esteem; Pierce et al., 1989)과 유사하다는 점이 큰 차이점이다(Lee et al., 2019; Shin et al., 2016).

둘째, 조직 관련 변수의 사례는 구조조정이나 고용형태(비정규직) 등에 의한 직장 내 고용불안과 같은 국내 노동시장의 상황을 반영한 변수들이 도출되었으며, 보상 및 승진과 연관된 불합리한 인사 고과시스템, 복리후생에 대한 불만을 들 수 있다. 고용불안의 경우 기존 연구에서 확인되었던 재직 중인 직장에서의 해고 불안 외에도 낮은 취업률 및 높은 실업률 등 외부 요인이 포함된다(Eun et al., 2018). 그리고 인사고과시스템의 경우 기존의 연구에서는 주로 조직 공정성(Greenberg, 국내 이직 관련 질적 연구 및 빅데이터 연구에서 나타난 이직 선행요인에 대한 종합적 검증 2079 Cropanzano, 2001) 중 상호작용공정성(interactional justice)과 정보공정성(informational justice)으로 대체되고 있어 인사고과시스템 자체에 대한 평가 및 만족이라기보다 사내 정보에 대한 공개 여부나 상사로부터의 대우 등과 연관된 변수이며, Roberts, Reed(1996)의 모형과 같은 인사고과 시스템 특성에 대한 이해, 수용 정도를 검증한 연구는 드물다. 그리고 복리후생의 경우 조직에 대한 맹목적 충성정보다는 일과 생활의 균형(work-life balance; WLB)을 중시하는, 2000년대 후반 이후부터 노동시장에 나타난 워라벨세대의 특성을 반영한다고 볼 수 있다(Park, Chang, 2015; SERI, 2009). 최근 연봉보다 혁신적인 근무제도나 다양한 복지제도에 대한 관심을 더 지닌 직장인들이 증가했고, COVID-19로 인해 일과 삶이 통합(work-life integration, WLI)되고 있는 상황을 고려할 때 기업의 비용지출 및 인재 확보 전략에 있어 매우 중요한 변수이다(Job Korea, 2017; Lee, 2019).

셋째, 새로운 변수 중 개인적 영역에 속한 변수의 사례를 살펴보면, 가장 많이

도출된 것은 가정과 관련된 변수들이다. 직장인들은 가정과 직장이라는 두 영역에서 필연적으로 역할을 지니게 되며, 이러한 역할을 수행하기 위한 시간과 비용, 가족의 지지 등과 같은 자원을 필요로 한다. 그러나 많은 요구가 주어질 때 사용 가능한 자원이 한정적이라면 일-가정 갈등(work-family conflict)이 발생한다. 기존 연구들에서는 전이이론(spillover theory)이나 보상이론(compensation theory)을 근거로 일과 가정이 영향력을 주고받거나 역의 관계를 지니고 각 영역의 손실을 보상받는 등의 작용을 통해 직무만족이 낮아지거나 직무스트레스가 상승하여 이직의도에 영향을 끼치게 된다(Champoux, 1980; Near, Rice, Hunt, 1980; Staines, 1980). 가정영역 내 선행변수로는 배우자의 지지(spouse support)나 결혼만족(marital satisfaction), 가정 중심성(family centrality) 등이 있으나(Carr, Boyar, Gregory, 2007; Thomson et al., 2015), 질적 연구에서 나타난 가정영역의 변수들은 주로 결혼과 육아 등 사건 자체라는 점이다(Jeon et al., 2014; Lee et al., 2019). 물론 해당 사건들로 인해 금전적 지출 증가 및 배우자의 지지에 대한 요구 등 개인마다 서로 다른 영향력을 지닌 지각 변수가 나타나 이직이 발생하는 매개과정(mediation process)은 유사할 수 있으나, 최우선적으로 선행되는 촉발사건들 자체에 대한 영향력 검증도 필요하다. 추가로 가정영역에 있어 부모의 영향이나 사회적 영역이라 할 수 있는 또래의 영향은 국내에서만 나타나는 변수이다. 이는 가족 중심사회인 우리나라에서 개인이 사회적 위치를 직업이나 재직 중인 기업에 대한 부모의 인정 및 주변 또래들과의 비교를 통해 확인하려는 경향성을 지니고 있기 때문이다(Lee et al., 2019; Yi, Sim, 2016). 또한, 통근거리(시간)나 직장의 위치, 직장으로 인해 거주하게 된 지역의 편의성, 학업병행의 어려움 및 진학 등의 변수가 존재한다(Kang et al., 2017; Lee et al., 2019). 기존의 양적 변수들에서 검증되지 않아 그 예측력에 대한 타당성은 가늠할 수 없으나, 탐색적 연구들에서 하나의 주제나 영역으로 분류된 변수이기 때문에 이 역시 다른 이직·잔류사유 등과 함께 비교 검증할 필요성이 있다. 그 외 우울과 불안 등 부정적 정서 수준의 증가, 현실에 안주하는 태도, 자신에 대한 불신, 생계유지만을 위한 목표 등 개인 정서적이고, 신념 및 가치관에 관련된 변수들도 살펴볼 수 있다(Eun et al., 2018; Kim et al., 2020) 그러나 이러한 지극히 개인적인 변수들은 삶이나 직장에서 겪은 사건들 혹은 개인의 성격에 영향을 받는

것으로, 각 질적 연구들의 면담절차 내 상세 질문이나 자료분류 기준에 의해 영역도
출 여부가 결정되거나, 다른 변수로 대체될 가능성도 있다(Zimmerman, 2008).

요약하자면, 이직 관련 질적 연구들은 기존 국내·외 양적 연구에서 검증된 변
수들이 유사하게 나타나고 있으나, 연구의 목적과 부합되는 심층적 탐색결과에 따
른 변수들도 추가로 나타나고 있다. 즉, 국내 노동시장의 상황과 현재 직장인들의
특성을 반영하는 기준에 관찰되지 않았던 변수들 (이종현, 이재은) 이 이직의 요인
으로서 직무, 조직, 개인의 각 영역에 도출하고 있어 해당 변수들에 대한 신뢰도
및 타당도 확보가 요구된다

제2절 이직 및 퇴사관련 빅데이터(텍스트마이닝)연구

2.1 빅데이터 분석

IT기술의 발전을 토대로 ‘제4차 산업혁명’의 혁신적인 변화는 사회 전방면에 영
향을 미치고 있다. 이미 사물인터넷(IoT), 인공지능(AI), 머신러닝(Machine learn-
ing) 등의 핵심기술은 보편화되고 있으며 이러한 새로운 기술 혁신의 바탕에는 디
지탈 환경에서 방대한 규모로 생성되는 빅 데이터의 정보의 축적이 규모를 더욱
크게 만들고 있다(박선우, 2018).

‘빅데이터’(Big data)란 2011년 맥킨지의 보고서에서 처음 사용한 용어로서 기존
의 데이터 처리기술 방식으로는 수집, 저장, 관리, 분석이 어려운 방대한 규모의 정
보 및 데이터를 의미한다(Manyika et al., 2011). 삼성경제연구소 (2011)는 빅데이터
(Big data) 분석에 대하여 데이터의 수집과 분석을 통해 가치 있는 정보를 추출하
여 고객의 요구에 능동적으로 대응하고 시장의 변화를 예측하기 위한 정보화 기술
이라 정의하였다. 빅데이터의 대표적 특징으로 데이터의 양(Volume), 다양성
(Variety), 속도(Velocity) 등이 있다(McAfee & Brynjolfsson, 2012). 이서구(2015)는
빅데이터에 대하여 3가지로 나누어 설명하였다.

첫째 데이터의 양(Volume)은 데이터의 크기를 의미한다. 현재 매 초당 연결되는 데이터의 양은 20년 전의 인터넷상에서 교환되는 모든 데이터의 양을 합친 것보다 많다. 오늘날의 다양한 기업에서는 굳이 인터넷이 아니더라도 매우 큰 용량의 데이터를 손쉽게 다룬다.

둘째, 빅데이터의 특징은 속도(Velocity)이다. 빠른 속도의 분석은 그 자체로 경쟁력의 원천이 될 수 있기 때문에 현대의 비즈니스에서도 무한한 속도의 경쟁을 통해 우위를 논하고 있다. 빅데이터는 이와 같은 빠른 속도의 분석을 가능하게 함으로서 소비자들의 니즈를 파악하는 데에 중요한 도구로서 활용된다. 예를 들어, 고객의 인터넷 쇼핑 정보를 통해 소비자에 관한 특성과 행동정보가 실시간으로 분석되어 해당 고객의 니즈를 파악하여 최적의 대안이 제시되는 경우가 있다. 이러한 빠른 속도의 실시간 빅데이터 분석은 해당 기업의 고객 가치를 증대시켜 결과적으로 경쟁력 제고에 도움이 된다.

셋째, 다양성(Variety)이다. 빅데이터의 다양성이란, 정형화 및 비정형화 자료까지 폭넓게 포함하는 것을 의미한다. 그 자료의 유형은 텍스트 메시지일 수도 있고, 이미지일 수도 있다. 또는 센서가 읽어낸 정보일 수도 있고, 스마트 폰에서 발생된 GPS 정보일 수도 있다. 이러한 다양한 유형의 자료를 포함한다는 점에서 기존의 데이터와 차이를 보인다.

빅데이터의 특징에 대하여 양, 속도, 그리고 다양성 이외에 진실성(Veracity)이 추가되었다(Schroeck et al., 2012). 데이터의 진실성이란, 품질과 신뢰성을 의미한다. 예를 들어, 친환경 에너지 생산에 연관성이 높은 날씨의 불확실성, 에너지 수요량 예측 및 재생자원에 대한 에너지 생산은 서로 관계가 있어 자료의 불확실성으로 상당한 영향을 받으므로 진실성(Veracity)의 특징의 중요성이 주목된다.

빅데이터의 분석 기술 중 하나인 텍스트 마이닝(Text Mining)은 비정형 혹은 반정형 텍스트 데이터에서 자연어처리 기술을 기반으로 정보와 지식을 추출하는 기술로 정의된다(Feldman & Dagan, 1995). 텍스트 마이닝 기술을 통해 방대한 텍스트에서 의미있는 유용한 정보를 추출하고, 다른 정보와의 연계성을 파악을 통해, 텍스트가 가진 카테고리를 찾아 키워드의 방향성을 확인할 수 있다. 기존의 단순한 정보 검색과는 달리, 그 이상의 결과를 얻어낼 수 있으며 컴퓨터가 인간이 사용해

는 언어(자연어)를 분석하고 대용량 언어 자원과 통계적, 규칙적 알고리즘을 사용하여 그 안에 숨겨진 정보를 발굴해낸다. 주요 응용분야로 문서 분류, 문서 군집, 정보 추출, 문서 요약 등이 있다(김정숙, 2012). 또한, 텍스트 마이닝 분석은 기계학습, 정보 검색, 자연어처리, 데이터 마이닝 등 여러 학문 분야의 연구 성과에 기반을 둔 응용 분야라 할 수 있다(Kao & Poteet, 2007). 그렇기 때문에 이를 기반으로, 빅데이터는 개인 행동, 소비자 선택, 탐색 행동, 교통 흐름, 질병관리 등 다양한 분야에서 패턴을 분석할 수 있을 뿐만 아니라 정보의 예측력까지 높은 앞으로 전망이 매우 좋은 기술이다(George et al., 2014).

빅데이터는 다양한 종류의 규모 데이터 분석을 통해 새로운 가치를 창출할 수 있다는 점에서 산업적, 정책적 활용에 높은 관심을 받고 있다. 특히, 4차 산업혁명 논의의 계기로 데이터에 대한 관심과 함께, 정부 차원에서 추진해온 공공데이터 및 데이터 관련 산업 활성화 정책을 통해 기존의 규제에 비해 다양하고 방대한 데이터들이 실제 활용될 수 있는 환경이 구축되었다(이규범 외, 2017). 각 학문별로도 분야별 특성을 고려한 빅데이터 활용이 이루어지고 있으며, 실제 사회과학 분야의 연구에서도 다양한 방식으로 시도되고 있다(심원섭 외, 2018; 손정렬, 2020)

2.2 이직 및 퇴사 관련 텍스트 마이닝 연구

정보통신기술의 발전으로 인터넷, 모바일, SNS 분야에서 매우 많은 정보가 지속적으로 생성되고 있으며, 특히 SNS에서 나타나는 텍스트 정보 등의 비정형 데이터를 포함한 빅데이터 연구도 기하급수적으로 증가하고 있다(Han, Jin, 2014). 2016년부터 약 1~2편씩 출판되고 있는 이직 관련 국내 빅데이터 연구는 연구 경향성을 분석한 문헌 및 이직과의 직접적 관련성이 적은 연구들을 제외하면 매우 드물다(Choi, Kim, 2019). 해당 연구들은 주로 국내 온라인 사이트의 기업 리뷰 정보를 이용하여 텍스트 마이닝 기법을 통해 이직을 예측하는 토픽을 탐색하는 방식이다. 일부 연구를 제외하면 한 연구당 데이터는 1.6만~6만 건이며, 방대한 양의 데이터에 근거한 결과이나, 대부분의 연구가 연구자 개인의 해석과 요약에 근거한 정성적 방법이 아닌, 사전에 이론을 통해 토픽 집합을 구성하여 텍스트를 코드화하는 정량적

내용분석법을 사용하였다. 즉, 연구자의 주관에 의한 단어의 특성 및 의미 파악 과정에서 정보가 누락되거나 오해석의 가능성을 낮출 수는 있으나, 검증되지 않은 추가적인 잠재변수를 확인하기는 상대적으로 어렵다(Cheung, Jak, 2016). 게다가 대부분의 연구에서 이러한 토픽 집합을 구성하는 틀이 자료를 수집한 온라인사이트의 직무만족 평가요소에 초점이 맞춰져 있어 이러한 제약점은 더욱 부각된다(Kim, Kang, Lim, 2016).

그럼에도 불구하고 일부 연구는 텍스트 정보에서 이직자와 재직자를 구분 후 만족요인과 불만족요인을 비교 분석하였고, 직무만족의 하위요인에는 기존 이론적 틀에 적합한 직무내용을 비롯한 승진기회, 급여, 복지, 조직문화 등 다양한 직무·조직 요소를 포함하고 있다. 게다가 다양한 업종 및 집단 간 특성, 그리고 시기별로 가능한 비교분석을 통해 결과를 도출하였다(Jung, Suh, 2019; Lee, Kim, Kang, 2017). 따라서 이러한 장점으로 인해 후속 조직 연구들에서 온라인 리뷰 데이터의 활용 가능성 및 다양한 경영전략 등의 시사점을 제공해 준다.

국내 이직 관련 텍스트 마이닝 연구에서 온라인 기업리뷰 사이트의 평가틀과 직무만족 이론을 통해 도출한 요인에 대한 내용을 종합적으로 살펴보면 다음과 같다(Jung, Suh, 2019; Kim, Kang, Lim, 2016; Lee, Kim, Kang, 2017). 우선은 직무의 내용이다. 이는 질적 연구에서의 결과와 마찬가지로 업무의 강도나 효율성, 야근 등과 관련되며, 직무 스트레스를 발생시켜 이직에 이르는 과정을 검증한 기존 연구에서 빈번하게 나타나는 요구변수이다. 이 외에도 일과 가정의 균형 및 가정에 관련된 단어들을 본 요인에 포함하고 있거나 따로 분류하고 있어 앞서 언급한 워라벨 세대의 특성이 반영된 중요한 결과라 볼 수 있다(Bae, Kim, 2009; Park, Chang, 2015; Kim, Kang, Lim, 2016).

다음은 조직문화를 들 수 있다. 이는 ‘보수적, 군대, 사내 정치’ 등의 단어를 포함하며, 업무현장의 분위기와 인사고과시스템에 대한 불만 등 조직 자체의 특성을 반영하는 요인이라 볼 수 있다. 특히, 조직정치의 경우 직무 불만족이나 철회행동(withdrawal behavior) 등에 관련된 변수로, 업무환경에 대한 불확실성을 증가시켜 개인의 스트레스를 증가시키고, 조직의 성과를 저해하는 변수이며, 국내에서도 마찬가지로 이직과의 관련성이 연구되고 있다(Ferris, Kacmar, 1992; Ferris et al.,

1995).

세 번째는 급여와 복지이다. 이는 ‘연봉, 복리후생, 연차’ 등의 단어가 포함되며, 급여와 복지요인이 구분되기도 한다(Jung, Suh, 2019). 연구에서 우세분석(dominance analysis)을 시행한 경우, IT 직군이나 화학/제조, 서비스 등 다양한 직군에서 대체로 다른 요인들보다 상대적으로 중요한 요인으로 국내 이직 관련 질적 연구 및 빅데이터 연구에서 나타난 이직 선행요인에 대한 종합적 검증 2081로 나타나고 있다(Jung, Suh, 2019; Kim, Kang, Lim, 2016). 네 번째는 승진기회이다. 승진은 급여와 같은 금전적 보상이 아닌, 인사고과시스템에 의해 부여되는 구조적 보상이며, 연구에 따라 인사(human resource)나 자기개발(self development)과 연관된 단어로 포함되기도 하고, 특정 직군(IT)에서만 중요성이 높게 나타나기도 한다(Jung, Suh, 2019). 이외에도 경영진에 대한 만족도와 근무시간, 휴가, 고용형태, 기업 간 관계, 프로젝트, 평판, 근무지 등 앞서 언급한 요인들과 공통성을 지니거나 질적 연구결과에서 관찰 가능했던 요인들이 존재한다(Jung, Suh, 2019).

요약하자면, 이직 관련 텍스트 마이닝 연구는 방대한 데이터를 근거로 하지만 상대적으로 연구 수가 부족하기 때문에 토픽이나 관련 변수들에 대해 확신하기는 아직 이르다. 또한, 이직의 선행요인인 직무 만족에 대한 이론을 토픽모형으로 설정하고 있어 이직과의 직접적 관련성을 속단하기에도 이르다고 할 수 있다. 그러나 소수의 연구라 할지라도 기존 양적 연구에 따른 결과들이 중요한 토픽으로 도출되고 있고, 아직 검증되지 않은 잠재적 변수들도 질적 연구와 유사하게 파악 가능하여 마찬가지로 양적 자료에 의한 이직에의 영향력 검증이 요구된다.

제3절 콜센터 근무자의 이직에 관한 연구

콜센터 근무자의 이직관련 국내 연구는 주로 직무특성과 직무만족, 조직몰입을 주요 변수로 보고 이를 이직의도와 연관시켜 분석하는 형태를 보인다. 특히 직무만족 요인으로는 콜센터 근로자의 감정노동, 스트레스 반응 등 감정노동 측면에서의 접근이 많은 편이다.

이석인(2007)은 콜센터 근무자의 이직의도 영향요인을 파악하기 위해 설문 분석을 통해 연구하였고, 분석결과 직무만족과 조직몰입은 이직의도를 감소시키는 중요한 요인들로 분석이 되었고, 직무특성은 이들 두 변수를 매개로 하여 이직의도를 간접적으로 감소시킴을 확인하였다.

김정임(2015)은 콜센터 근로자의 감정노동, 스트레반응이 이직의도에 미치는 영향을 파악하기 위해 설문 분석을 통해 연구하였고, 분석결과 감정노동, 스트레스반응 및 이직의도의 상관관계에서는 모두 양의 상관관계가 있는 것으로 분석되어 대상자의 감정노동 수준이 높을수록 신체화, 우울 및 분노의 스트레스반응은 높아지는 것으로 나타났다.

기존의 연구를 요약하자면, 콜센터 근로자들의 경우 감정노동 및 스트레스 반응이 직무 만족과 조직 몰입에 영향을 미치며, 이는 이직 의도와 직접적인 상관관계를 가진다는 것이다.

제4절 텍스트마이닝(빈도,네트워크분석,토픽모델링)

4.1 빈도 분석

본 연구에서는 텍스트마이닝 기법에서 가장 쉽고 보편적으로 활용되는 단어 빈도 분석을 첫 단계로 수행한다. 빈도 분석은 문서별 단어 출현 빈도를 보여줌으로써 본격적인 텍스트 분석 전 데이터에 대한 이해와 흐름을 살펴보기 위한 기초 분석을 할 수 있다. 다만 빈도 분석만으로는 토큰화된 단어들 간의 정확한 관계와 숨겨진 의미 파악이 불가능하므로, 본 연구에서는 후술할 네트워크 분석 및 토픽 모델링 기법을 동시에 적용해 데이터 분석의 정확도를 높이려고 한다.

4.2 네트워크 분석

본 연구는 네트워크 분석을 통해서 사직서의 구체적 퇴사 사유에서 발췌된 각

토큰(형태소 분석 후 추출한 명사 단어)들 간에 어떤 유기적 관계를 형성하고 있는지를 분석하고자 한다. 네트워크 분석을 하지 않을 경우 토큰화되고 파편화된 단어의 유기적 관계를 파악할 수 없어 데이터 해석의 오류를 일으킬 가능성이 높아진다. 일반적으로 이러한 유기적 관계 속성 간에 일어나는 상호작용은 연결망 구조로 묘사될 수 있는데 이를 네트워크라고 한다. 네트워크 분석에서 연결망의 구조적 변수는 노드(nodes)라고 불리며, 본 연구에서 네트워크 노드란 퇴사 사유를 나타낸다. 네트워크 안에서 퇴사 사유 간 연계는 연결 관계(link 또는 tie)로 나타내며, 퇴사 사유들의 지속적인 연결 관계를 통해서 경로(path)가 구축된다. 그리고 네트워크 안에서 퇴사 사유들이 연결되는데 필요한 최소한의 연결 관계를 경로거리(path length 또는 path distance)라고 할 수 있다. 일반적으로 네트워크는 특정 연구 분야들 간의 연계관계 빈도에 의해서 측정되며, 상대적으로 연결 관계가 많은 연구협력 분야들을 모아 군집, 집단 또는 클러스터(clusters)라고 한다. 따라서 군집은 다중적 연결 관계를 통해서 연계가 중복적으로 일어나는 융합 네트워크 안의 연구 분야 집단을 의미한다(허정은 외, 2013).

이러한 연계의 중복성(redundancy)은 네트워크를 탄력적으로 만들어 몇몇 분야가 사라지거나 연결 관계가 부분적으로 단절되어도 전체적인 연구 군집의 연계성은 위축되지 않게 된다. 네트워크 안에서 연구 분야들의 상대적 위치는 지도화(mapping)되고 분석될 수 있는데, 이러한 관계적 속성을 나타내는 중요한 개념으로 중심성(centrality)을 들 수 있다. 일반적으로 중심성 측정이란, 네트워크 노드가 네트워크 안의 다른 노드들과 연계되는 빈도와 노드의 거리에 초점을 맞춘다(Scott, 1991). 따라서 중심성 지수를 측정하는 것은 연구 내에서 형성되고 있는 연구 분야들의 유기적 연계 구조를 나타내 줄 수 있다. 중심성 진단은 다른 분야들과 연결 관계가 있는 특정 연구 분야를 파악하며, 그 특정 분야는 다른 분야들과 연계되는 최단 경로의 일부가 된다.

Knocke와 Kuklinski(1982)의 개념에 따르면, 중심성은 네트워크 안에서 형성되는 모든 연결 관계 중에서 특정 연구 분야가 포함되어 있는 연계 관계의 총 수를 전체에서 차지하는 비율로 나타낼 수 있다. 이러한 중심성 연구 분야는 융합연구 네트워크를 형성하도록 하는 중요 역할을 담당하게 되며, 연구 분야들간에 상호 연계

가 이루어지도록 영향력을 미치게 된다. 연결 중심성(degree centrality)은 네트워크 안에서 특정 연구 분야에 인접하여 연계되어 있는 상이한 분야의 수를 나타내는 것으로 다른 연구 분야들로부터 연계된 빈도가 높을수록 연결 중심성이 높다고 볼 수 있다. 매개 중심성(betweenness centrality)은 네트워크 안에서 특정 연구 분야가 상이한 두 연구 분야의 “사이”에 위치하는 정도를 의미한다(Scott, 1991). 이는 자체적으로 연계될 수 없는 상이한 두 분야를 매개시켜주는 다리(bridge) 역할을 수행하는 연구 분야의 역량을 나타낸다. 연구 분야의 매개 중심성은 특정 연구 분야에서 다른 분야로의 경로 함수로 나타낼 수 있다. 따라서 매개 중심성은 특정 연구 분야가 네트워크 안에서 연구의 흐름을 제어하거나 조정할 수 있는 정도를 측정한다고 볼 수 있다.

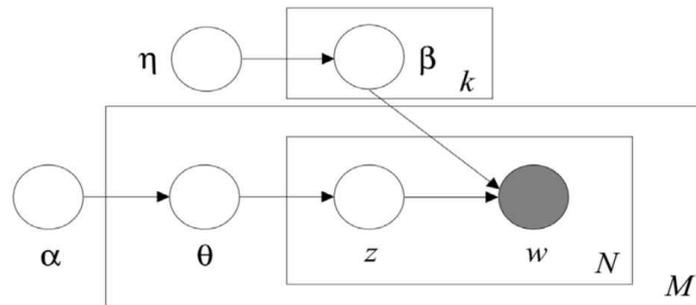
근접 중심성(closeness centrality)은 네트워크 안에서 연구 분야의 잠재적 독립성을 나타낸다. Scott(1991)의 개념에 따르면, 근접 중심성은 융합연구가 일어날 수 있는 두 연구 분야 사이에 존재하는 최단 경로의 거리(geodesic distance)를 의미하는 것으로, 거리가 짧을수록 연구 분야 간 근접성이 높아져 연구의 흐름이나 확산이 밀접히 일어나게 된다고 본다.

4.3 LDA 토픽 모델링 분석

본 연구에서는 토픽 모델링 방법 중 대표적으로 활용되는 LDA(Latent Dirichlet Allocation) 모델을 사용하였다. 토픽모델링이란 문서 집합의 추상적인 ‘토픽’을 발견하기 위한 텍스트마이닝 기법으로 하나의 문서가 특정 단어들의 집합이라고 가정하고 문서에 나오는 단어를 확률적으로 계산하여 본문의 숨겨진 의미구조를 발견하는 통계적 모델 중 하나이다. 1990년 Deerwester et al.을 통해 등장한 최초의 토픽 모델인 LSI(Latent Semantic Indexing)는 문서-단어 행렬을 문서-토픽 행렬과 토픽-단어 행렬로 분해하는 과정을 기술하였고, 2001년 Hofmann et al.은 LSI를 기반으로 단어의 출현 빈도를 확률로 대체하는 모형인 pLSI(probabilistic Latent Semantic Indexing) 모델을 기술하였다. pLSI 모델은 확률을 적용하지만 생성된 모델에 새로운 문서가 입력될 경우 기존 모델을 적용할 수 없는 오버피팅

(Overfitting) 문제가 존재하였고, 이후 2003년 Blei et al.은 pLSI의 불완전한 확률 모델을 보완해 LDA 모델을 발표하면서 현재의 토픽모델링 분야를 정착시켰다.

LDA 모델은 문서나 단어 등 관찰된 변수를 통해 문서 구조처럼 보이지 않는 변수를 추론하는 것을 목적으로 하며 이를 통해 전체 문서 집합의 토픽, 각 문서별 토픽 비율, 각 토픽에 포함된 단어의 분포 등을 도출할 수 있다. LDA 모델의 데이터 처리 과정은 <그림 1>과 같다. 사전에 문서-토픽별 분포값(θ)과 토픽의 단어 분포값(β) 및 토픽의 개수(k)를 하이퍼 파라미터(Hyper parameter)로 미리 입력하여야 하며 입력된 값을 통해 단어(w)를 관측해 단어마다 적절한 토픽번호(z)를 정해준다. 그리고 모델을 반복하며 θ , β 값을 갱신, 모든 z 값 중 가장 높은 z 값을 찾아 문서에 있는 각각의 단어들이 어디에 속해야 하는지 추론한다.



<그림 1. LDA 모델의 데이터 처리 과정>

- M : 문서의 개수
- N : 문서에 속한 단어의 개수
- W : 단어
- Z : 해당 단어가 속한 토픽번호
- k : 토픽의 개수 (Hyper parameter)
- α : 문서-토픽별 θ 분포값 (Hyper parameter)
- η : 토픽-단어별 β 분포값 (Hyper parameter)
- θ : 문서별 토픽의 분포
- β : 토픽의 단어 분포

III장 연구 절차 및 방법

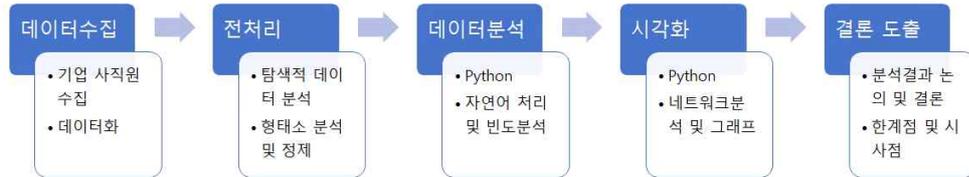
제1절 연구대상 및 절차

본 연구의 목적은 콜센터 근무자들의 사직원을 빅데이터 기법을 활용해 분석하여 실질적인 퇴사 및 이직의 원인을 알아보는 것이다. 이를 위해 국내 최대의 아웃소싱 콜센터 운영업체의 8년간 사직원 데이터를 수집하고 정제해서 연구에 활용하였다. 연구 대상의 통계학적 특성은 <표 1>과 같다.

구분		빈도	비율
전체		52,212	100%
성별	남자	15,916	30%
	여자	36,296	70%
연령	20대	16,986	33%
	30대	19,561	37%
	40대 이상	15,665	30%
근속기간	1년 미만	39,857	76%
	1년 이상	12,355	24%
코로나 전후 퇴사	코로나 이전	39,529	76%
	코로나 이후	12,683	24%

<표 1. 연구대상의 통계학적 특성>

전체적인 연구 진행 절차는 <그림 2>와 같다. 데이터 수집 및 정제는 MS-Excel을 활용했고, 전처리, 데이터 분석 및 시각화는 모두 Python 3.7.7 버전을 활용하였다.



<그림 2. 연구 진행 절차>

제2절 연구방법

2.1 데이터 수집

분석을 위해 국내 최대의 콜센터 아웃소싱 업체인 T사의 8년간(2014년~2021년) 사직원 데이터를 수집하였고, 해당 데이터의 구조를 파악하여 최종 연구 대상 데이터를 확정하였다. 총 54,377건의 사직서를 수집하였고, 데이터 분석에 필수적인 필드(입사일, 퇴사일, 생년월일, 성별, 구체적 퇴사 사유, 면담 기록)가 두 개 이상 없는 데이터를 제거하였고, 퇴사 사유가 없는 경우에는 면담 기록으로 대체하였다.

1차 데이터 탐색을 통해 연구에 활용할 수 있는 사직원 총 52,212건을 최종 확정했으며, MS-Excel을 활용해 데이터를 정제하고 분류(데이터화)하였다. 해당 데이터는 <표 1>의 특성을 반영해 분류했으며, 각 분류별로 CSV 파일로 내보내 빅데이터 분석의 기반 자료로 활용하였다.

2.2 데이터 전처리

데이터 전처리는 입력 데이터로 사용될 “구체적 퇴사 사유”를 대상으로 진행하였다. 첫째, 정규표현식을 이용해 한글과 공백을 제외한 모든 문자를 제거하였다. 둘째, KoNLPy 패키지를 활용해 형태소를 분석하고 명사 단위로 토큰화하였다. 셋째, 일반적인 접미어와 분석에 영향을 미치지 않는 ‘위해’, ‘사정’, ‘사유’, ‘상의’, ‘

불가', '이상', '회사', '처리', '예정', '사직', '퇴사', '이유', '요청', '퇴직', '문제', '이전', '다른', '이후', '바운드', '때문' 등은 불용어로 처리하였다.

2.3 빈도 분석

빈도 분석은 전처리된 퇴사 사유 데이터에서 어떤 단어가 연관되어 있는 지를 확인하기 위해서 시행되었고, 빈도 수가 높은 상위 30개 단어들을 추출하였다. 빈도 분석을 위해 파이썬의 Counter 패키지를 활용하였으며, Pandas 패키지를 통해 데이터 프레임을 관리하였다.

2.4 네트워크 분석

네트워크 분석은 파편화된 토큰(단어)들의 연관 관계를 파악하기 위해 시행되었거, 그래프 시각화를 통해 분석 결과를 쉽게 이해할 수 있도록 하였다. 전처리된 토큰들의 연관 규칙을 분석해 연관성 높은 토큰들로 데이터 프레임을 구성하기 위해 Apriori 알고리즘을 사용하였다. 파이썬 apriori 패키지를 활용하였으며, min_support는 0.01로 설정한 후 연관 토큰들을 분류하였다. 네트워크 분석 후 그래프를 그리는 것은 파이썬 networkx 패키지를 활용하였으며, kamada_kawai 레이아웃으로 결과물을 시각화 하였다.

2.5 LDA 토픽 모델링 분석

앞서 전처리가 완료된 토큰 데이터를 각각 분류별로 LDA 모델에 적용하여 3,000번의 학습을 통해 주요 토픽 20개를 도출하였다. 파이썬 gensim 패키지에 포함된 LdaModel, CoherenceMetric, corpora, PerplexityMetric, CoherenceModel을 임포트하여 프로그래밍하고 실험하였으며, LDA 모델에 대한 하이퍼 파라미터 값은 $k=20$, $\alpha=0.1$, $n=0.01$ 로 진행하였다(<그림 1> 참고). 분석의 용이성을 위해 추가로 시행한 토픽과 키워드의 시각화는 파이썬 pyLDAvis 패키지를 활용하였다.

제 IV장 연구 결과

제1절 표본전체 분석 결과

(1) 단어 빈도 분석 - 상위 30

분석 대상 전체를 대상으로 추출한 퇴직 사유의 단어 빈도는 <표 2> 와 같다. 상위 10개 단어를 살펴보면, ‘개인’, ‘업무’, ‘이직’, ‘계약’, ‘종료’, ‘건강’, ‘무단’, ‘결근’, ‘적용’, ‘만료’이다. 의미 파악이 어려운 ‘개인 (사유)’를 제외하면, 업무 적용 문제, 프로젝트 종료 및 기간 만료로 인한 계약 만료, 이직을 위한 퇴사 순으로 빈도가 높은 것을 알 수 있다.

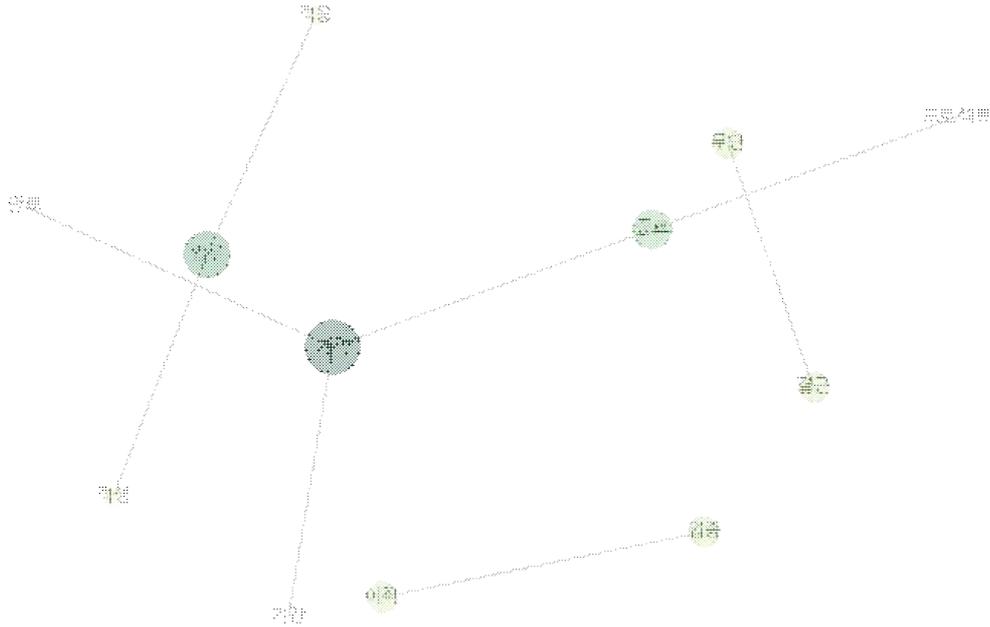
순위	단어	빈도수	순위	단어	빈도수
1	개인	12297	16	어려움	1016
2	업무	7979	17	기간	857
3	이직	6831	18	연락	850
4	계약	6187	19	준비	846
5	종료	4359	20	이사	828
6	건강	4098	21	프로젝트	763
7	무단	3872	22	지속	671
8	결근	3866	23	육아	650
9	적용	3347	24	고용	650
10	만료	2481	25	집안	641
11	근무	1934	26	결정	588
12	적성	1517	27	희망	582
13	전직	1507	28	스트레스	581

14	학업	1436		29	변경	556
15	업종	1248		30	치료	546

<표 2. 전체 표본의 퇴직 사유 빈도 분석 결과>

(2) 네트워크 분석

분석 대상 전체를 대상으로 Apriori 알고리즘으로 연관 토큰을 만들어 네트워크 그래프로 시각화하였으며 그 결과는 <그림 3>이다. 첫째, ‘업무’, ‘계약’, ‘종료’는 중요 단어로 나타났다. 둘째, ‘무단’, ‘결근’이 이직으로 이어지는 경우가 상당한 비중을 차지함을 알 수 있었다. 이는 빈도 분석에서도 파악된 내용으로 특정 사유 없이 무단 결근으로 퇴사 처리된 경우, 관리자에 의해 면담 기록 형태로 퇴사 사유로 기록된 것이 그 원인으로 추정된다. 셋째, 계약 종료의 경우 프로젝트 종료와 기간 만료가 퇴사로 이어진 것임을 알 수 있었다. 넷째, ‘업무’로 인한 퇴사의 경우 업무 적용의 문제와 업무가 적성에 맞지 않는 경우가 많음을 알 수 있었다. 다섯째, ‘이직’으로 인한 퇴사의 경우 다른 ‘업종’으로 이직을 희망하는 경우가 많음을 알 수 있었다.



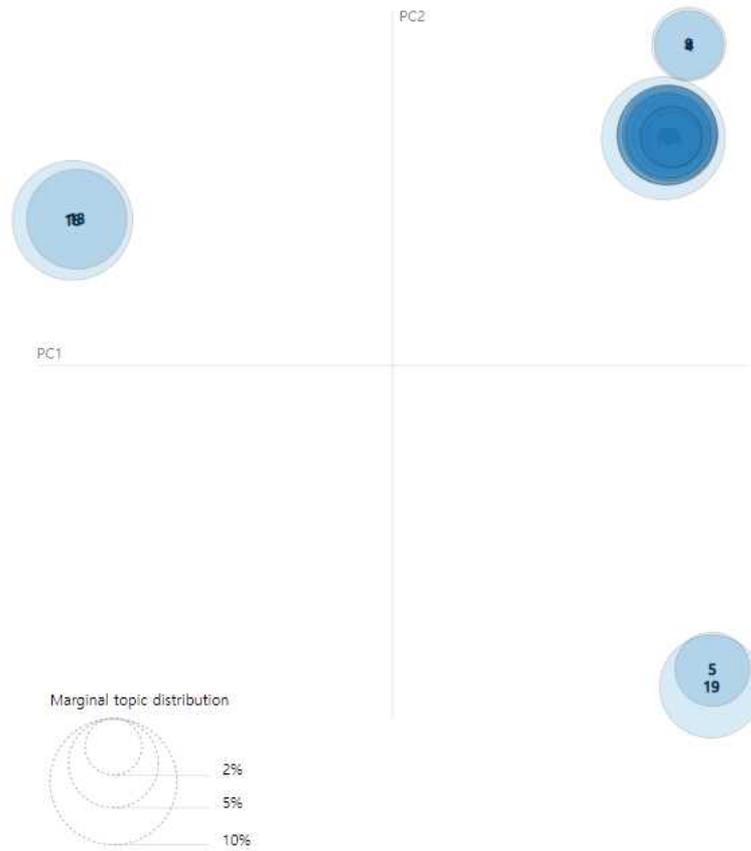
<그림 3. 전체 표본에 대한 네트워크 분석 결과 >

(3) 토픽 모델링 분석

전체 표본을 대상으로 한 LDA 토픽 모델링 분석은 131개의 고유한 토픽과 52,210개의 문장을 대상으로 수행했으며 그 결과는 <표 3>과 같다. 설정된 하이퍼 파라미터 값에 따라 총 20개의 토픽이 선정되고 각 토픽에 포함된 주요 키워드를 확인할 수 있다. 각 키워드 앞의 숫자들은 해당 키워드가 그 토픽에 기여한 점수를 나타낸다. 또한 토픽들간의 유사성을 파악할 수 있도록 시각화 한 결과가 <그림 4>인데, 이를 통해 토픽 #4와 #9, #5와 #19, #13과 #18이 유사한 토픽임을 알 수 있다. 이를 기반으로 분석한 전체 표본을 대상으로 한 실질적인 토픽 모델링의 결과는 ‘학업’과 ‘병간호’, ‘급여’와 ‘병 치료’가 매우 중요한 퇴사 및 이직의 원인으로 등장하였다. 이는 앞서 수행한 빈도 분석과 네트워크 분석에서는 발견하기 어려웠던 실질적인 사유가 추가적으로 파악된 것으로 볼 수 있다.

토픽	키워드(상위 5)
#1	'0.280*"연락" + 0.219*"진행" + 0.211*"부모님" + 0.169*"두절" + 0.067*"운영"'
#2	'0.331*"준비" + 0.208*"휴식" + 0.141*"공부" + 0.094*"판단" + 0.089*"자격증"'
#3	'0.286*"악화" + 0.196*"결정" + 0.176*"지속" + 0.169*"자녀" + 0.073*"상태"'
#4	'0.398*"간호" + 0.170*"원거리" + 0.134*"발령" + 0.117*"부족" + 0.082*"관리"'
#5	'0.358*"급여" + 0.265*"수술" + 0.099*"허리" + 0.093*"부분" + 0.080*"디스크"'
#6	'0.309*"이사" + 0.279*"육아" + 0.168*"희망" + 0.073*"지방" + 0.065*"지역"'
#7	'0.242*"교육" + 0.236*"이동" + 0.170*"상담" + 0.121*"센터" + 0.084*"조건"'
#8	'0.645*"학업" + 0.117*"병행" + 0.100*"부득이" + 0.077*"복귀" + 0.056*"관계"'
#9	'0.285*"기타" + 0.209*"복학" + 0.141*"생각" + 0.139*"학교" + 0.083*"확인"'
#10	'0.366*"어려움" + 0.197*"가족" + 0.106*"사업" + 0.105*"간병" + 0.092*"승계"'
#11	'0.306*"직종" + 0.302*"취업" + 0.149*"전공" + 0.100*"집중" + 0.089*"대학"'
#12	'0.206*"고객" + 0.179*"직장" + 0.129*"관련" + 0.125*"전환" + 0.109*"출근"'
#13	'0.386*"업종" + 0.228*"부담" + 0.105*"전직" + 0.095*"본인" + 0.093*"대환"'
#14	'0.322*"통보" + 0.254*"문자" + 0.227*"의사" + 0.185*"계속" + 0.000*"코로나"'
#15	'0.300*"스트레스" + 0.289*"변경" + 0.118*"장기" + 0.104*"부서" + 0.101*"단기"'
#16	'0.615*"만료" + 0.309*"적성" + 0.041*"응대" + 0.034*"입사" + 0.000*"고용"'
#17	'0.392*"병간호" + 0.264*"어머니" + 0.237*"건강악화" + 0.099*"유학" + 0.000*"입신"'
#18	'0.488*"근무" + 0.287*"전직" + 0.171*"프로젝트" + 0.052*"환경" + 0.000*"코로나"'
#19	'0.364*"기간" + 0.159*"집안" + 0.146*"치료" + 0.111*"병원" + 0.076*"입원"'
#20	'0.254*"해지" + 0.175*"난이도" + 0.164*"출퇴근" + 0.129*"시간" + 0.097*"거리"'

<표 3. 전체 표본을 대상으로 토픽 모델링 결과>



<그림 4. 전체 표본 대상의 토픽간 유사도 거리 맵>

제2절 성별분석결과

2.1 남자

(1) 단어 빈도 분석 - 상위 30

남자 직원 전체를 대상으로 추출한 퇴직 사유의 단어 빈도는 <표 4> 와 같다. 빈도 분석 결과 남자 직원들의 퇴직 사유는 ‘이직’, ‘업무’, ‘계약’, ‘무단결근’, ‘건강’

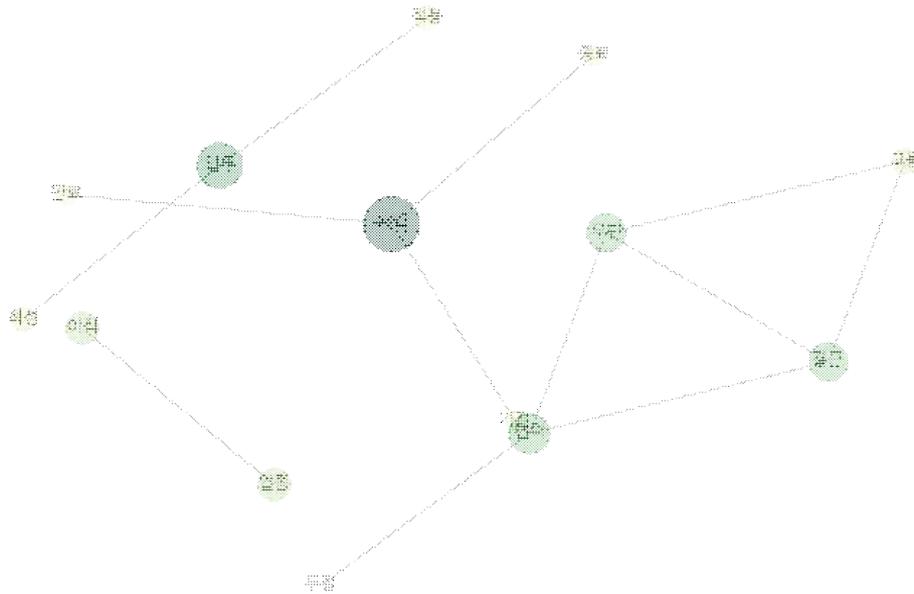
순이었다. 전체 표본을 대상으로 한 경우에 비해 ‘업무 적응’ 보다는 ‘이직’이 더 높은 특징을 보여준다. 특히 개인 ‘사업’을 위해 퇴사하는 경우도 비교적 높은 빈도를 보이는데 오직 남자 직원의 경우에서만 파악되는 특징이었다.

순위	단어	빈도수	순위	단어	빈도수
1	개인	3358	16	준비	358
2	이직	2578	17	연락	323
3	업무	2106	18	복합	300
4	계약	1985	19	어려움	273
5	결근	1357	20	기간	273
6	무단	1349	21	집안	222
7	종료	1338	22	고용	213
8	건강	874	23	결정	209
9	적응	836	24	사업	208
10	만료	777	25	지속	199
11	학업	608	26	이사	191
12	근무	572	27	두절	191
13	전직	561	28	진행	188
14	업종	559	29	희망	187
15	적성	462	30	급여	187

<표 4. 남자 직원의 빈도 분석 결과>

(2) 네트워크 분석

남자 직원 전체에 대한 네트워크 분석 결과는 <그림 5>이다. 첫째, 전체 표본 분석 결과에서 알 수 있었듯이 ‘계약’, ‘업무’, ‘결근’이 중요한 단어로 나타났다. 둘째, ‘연락’ 이 중요한 단어로 등장하고 ‘두절’, ‘무단’, ‘결근’과 연결이 되는데, 이는 실제 퇴직 사유를 확인한 결과 무단 결근 후 연락이 두절되어 퇴사 처리하는 경우로 파악이 되었다. 이 부분은 남자 직원일 경우 나타나는 특이한 형태로 남자 직원의 경우, 무단 결근 후 연락 두절로 이어지는 형태가 많은 것을 알 수 있었다.



<그림 5. 남자 직원에 대한 네트워크 분석 결과>

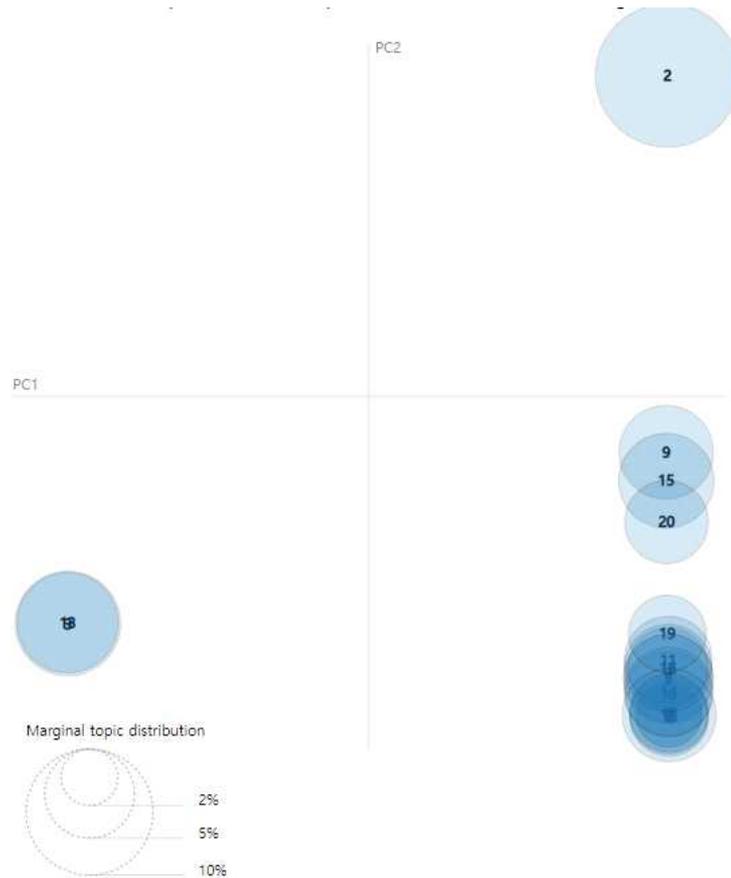
(3) 토픽모델링 분석

남자 직원을 대상으로 한 LDA 토픽 모델링 분석은 49개의 고유한 토픽과 15,916개의 문장을 대상으로 수행되었으며 그 결과는 <표 5>와 같다. 첫째, 토픽간 유사도 맵 <그림 6>을 참고하면 토픽 #2가 높은 비중을 가지면서 다른 토픽과 유사성이 없는 특성을 가짐을 볼 수 있다. 둘째, 토픽 #2의 경우 기여도가 0인 ‘고용’을 제거하고 해석을 해보면 남자 직원의 경우 계약 기간이 만료된 후 개인 사업을 이유로 퇴직을 하는 경우가 높다는 것을 알 수 있었다.

토픽	키워드(상위 5)
#1	'0.985* "지속" + 0.000* "악화" + 0.000* "수술" + 0.000* "사업" + 0.000* "스트레스"
#2	'0.596* "만료" + 0.194* "기간" + 0.122* "사업" + 0.087* "변경" + 0.000* "고용"
#3	'0.521* "업종" + 0.244* "이사" + 0.232* "악화" + 0.000* "고용" + 0.000* "계약해지"

#4	'0.575*직종" + 0.415*취업" + 0.000*악화" + 0.000*사업" + 0.000*스트레스'
#5	'0.428*집안" + 0.365*직장" + 0.202*아버지" + 0.000*스트레스" + 0.000*수술'
#6	'0.403*프로젝트" + 0.307*간호" + 0.282*수술" + 0.000*고용" + 0.000*악화'
#7	'0.530*복합" + 0.245*학교" + 0.219*대환" + 0.000*악화" + 0.000*사업'
#8	'0.388*치료" + 0.344*고객" + 0.261*장기" + 0.000*고용" + 0.000*수술'
#9	'0.996*학업" + 0.000*고용" + 0.000*사업" + 0.000*악화" + 0.000*학교'
#10	'0.386*희망" + 0.341*가족" + 0.267*전환" + 0.000*스트레스" + 0.000*수술'
#11	'0.369*해지" + 0.343*결정" + 0.285*군입대" + 0.000*고용" + 0.000*스트레스'
#12	'0.716*어려움" + 0.278*부도남" + 0.000*사업" + 0.000*스트레스" + 0.000*학교'
#13	'0.779*진직" + 0.218*업종" + 0.000*사업" + 0.000*스트레스" + 0.000*학교'
#14	'0.378*진행" + 0.337*교육" + 0.279*휴식" + 0.000*스트레스" + 0.000*사업'
#15	'0.995*근무" + 0.000*스트레스" + 0.000*만료" + 0.000*수술" + 0.000*사업'
#16	'0.627*연락" + 0.367*두절" + 0.000*만료" + 0.000*사업" + 0.000*스트레스'
#17	'0.338*공부" + 0.337*시간" + 0.319*출퇴근" + 0.000*이사" + 0.000*사업'
#18	'0.326*급여" + 0.307*스트레스" + 0.215*관련" + 0.148*본인" + 0.000*고용'
#19	'0.993*준비" + 0.000*수술" + 0.000*사업" + 0.000*악화" + 0.000*학교'
#20	'0.995*적성" + 0.000*학교" + 0.000*사업" + 0.000*스트레스" + 0.000*악화'

<표 5. 남자 직원의 토픽 모델링 결과>



<그림 6. 남자 직원의 토픽간 유사도 맵>

2.2 여자

(1) 단어 빈도 분석 - 상위 30

여자 직원 전체를 대상으로 추출한 퇴직 사유의 단어 빈도는 <표 6>과 같다. 남자 직원들에 비해 ‘업무 적응’의 어려움과 ‘건강’상의 이유로 퇴사하는 경우가 매우 높음을 알 수 있었다. ‘육아’와 ‘자녀’ 등의 가정 관련 키워드가 처음으로 등장하면서 남성보다는 여성이 가정과 자녀 문제로 퇴사하는 경우가 높음을 알 수 있었다.

순위	단어	빈도수	순위	단어	빈도수
1	개인	8940	16	업종	689
2	업무	5870	17	육아	639
3	이직	4252	18	이사	637
4	계약	4203	19	프로젝트	595
5	건강	3223	20	기간	584
6	종료	3022	21	연락	527
7	무단	2522	22	준비	488
8	적응	2511	23	지속	472
9	결근	2508	24	고용	437
10	만료	1704	25	간호	429
11	근무	1361	26	스트레스	422
12	적성	1055	27	집안	419
13	전직	946	28	치료	401
14	학업	829	29	자녀	401
15	어려움	742	30	변경	400

<표 6. 여자 직원의 빈도 분석 결과>

(2) 네트워크 분석

여자 직원 전체에 대한 네트워크 분석 결과는 <그림 7> 이다. 모든 분류에서 공통적으로 중요한 단어로 나타나는 ‘계약 종료’, ‘업무 적응’ 외에 특이한 사항으로 ‘건강’ 상의 이유로 ‘이직’하는 경우가 많이 매핑(Mapping)됨을 알 수 있었다. 남자 직원의 경우 ‘업종 전환’을 이유로 이직을 하는 것이 많은 것에 비해 특이한 현상임을 파악할 수 있었다.

#9	'0.442* "자녀" + 0.286* "교육" + 0.145* "통보" + 0.120* "문자" + 0.000* "포기" '
#10	'0.482* "어려움" + 0.225* "기타" + 0.200* "고객" + 0.090* "입사" + 0.000* "포기" '
#11	'0.286* "업종" + 0.268* "스트레스" + 0.163* "수술" + 0.095* "대환" + 0.073* "신청" '
#12	'0.246* "변경" + 0.225* "부담" + 0.203* "급여" + 0.121* "근무" + 0.070* "환경" '
#13	'0.341* "직장" + 0.309* "병원" + 0.234* "입원" + 0.110* "의사" + 0.000* "코로나" '
#14	'0.306* "근무" + 0.300* "고용" + 0.216* "학업" + 0.061* "발생" + 0.059* "승계" '
#15	'0.254* "관련" + 0.236* "장기" + 0.158* "자격증" + 0.122* "시험" + 0.116* "부분" '
#16	'0.728* "만료" + 0.194* "이사" + 0.052* "원거리" + 0.024* "지역" + 0.000* "단기" '
#17	'0.532* "육아" + 0.374* "지속" + 0.089* "근무" + 0.000* "코로나" + 0.000* "해지" '
#18	'0.349* "치료" + 0.241* "휴식" + 0.146* "상담" + 0.145* "상태" + 0.115* "필요" '
#19	'0.614* "전직" + 0.143* "병간호" + 0.141* "업종" + 0.098* "어머니" + 0.000* "포기" '
#20	'0.288* "악화" + 0.200* "난이도" + 0.174* "출퇴근" + 0.143* "시간" + 0.110* "거리" '

<표 7. 여자 직원의 토픽 모델링 결과>



<그림 8. 여자 직원의 토픽간 유사도 맵>

제3절 코로나 전후 분석 결과

3.1 코로나 이전

(1) 단어 빈도 분석

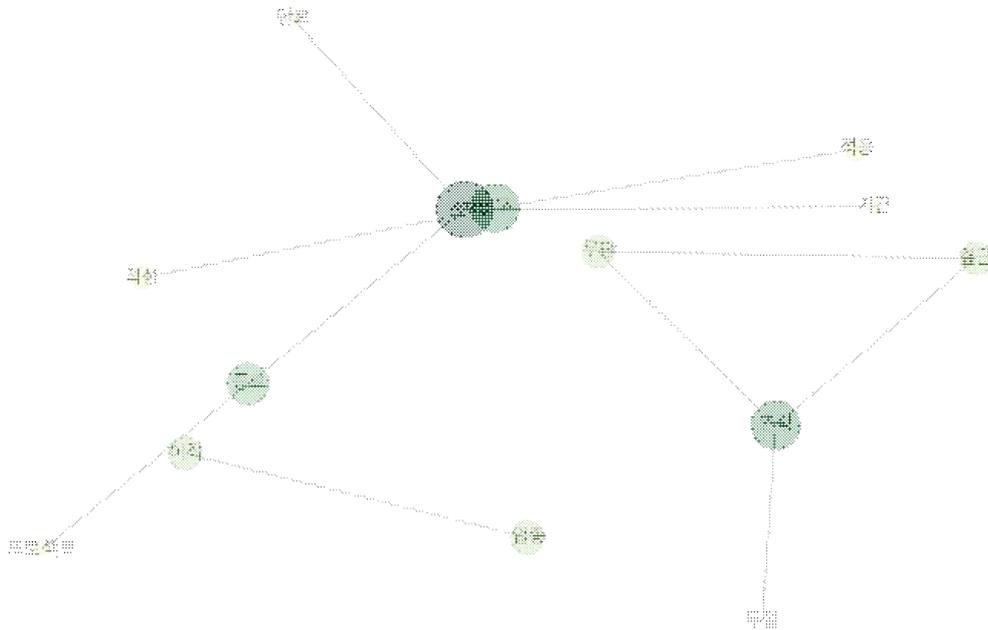
코로나 발생 이전(2019년까지) 퇴사자들을 대상으로 추출한 퇴직 사유의 단어 빈도는 <표 8>과 같다. ‘개인’, ‘업무’, ‘이직’, ‘계약’ 등 전체 표본을 대상으로 높은 빈도인 일반적인 사유들로 구성된 것을 알 수 있었다.

순위	단어	빈도수	순위	단어	빈도수
1	개인	8930	16	어려움	795
2	업무	6173	17	연락	794
3	이직	5194	18	프로젝트	710
4	계약	3966	19	준비	691
5	무단	3285	20	이사	636
6	결근	3269	21	지속	581
7	건강	3044	22	기간	575
8	종료	2993	23	육아	538
9	적응	2644	24	집안	515
10	만료	1668	25	희망	479
11	근무	1552	26	결정	479
12	학업	1222	27	두절	474
13	전직	1218	28	변경	457
14	적성	1118	29	복학	442
15	업종	958	30	치료	438

<표 8. 코로나 이전의 빈도 분석 결과>

(2) 네트워크 분석

코로나 발생 이전 퇴사자들의 퇴직 사유에 대한 네트워크 분석 <그림 9 참조> 역시 전체 표본을 대상으로 분석한 결과와 일치하며 큰 특이사항은 발견되지 않았다. ‘계약 만료’ 및 ‘업무 적응’과 ‘이직’을 위한 ‘업종’ 변경이 중요한 묶음으로 파악되었다.



<그림 9. 코로나 이전의 네트워크 분석 결과>

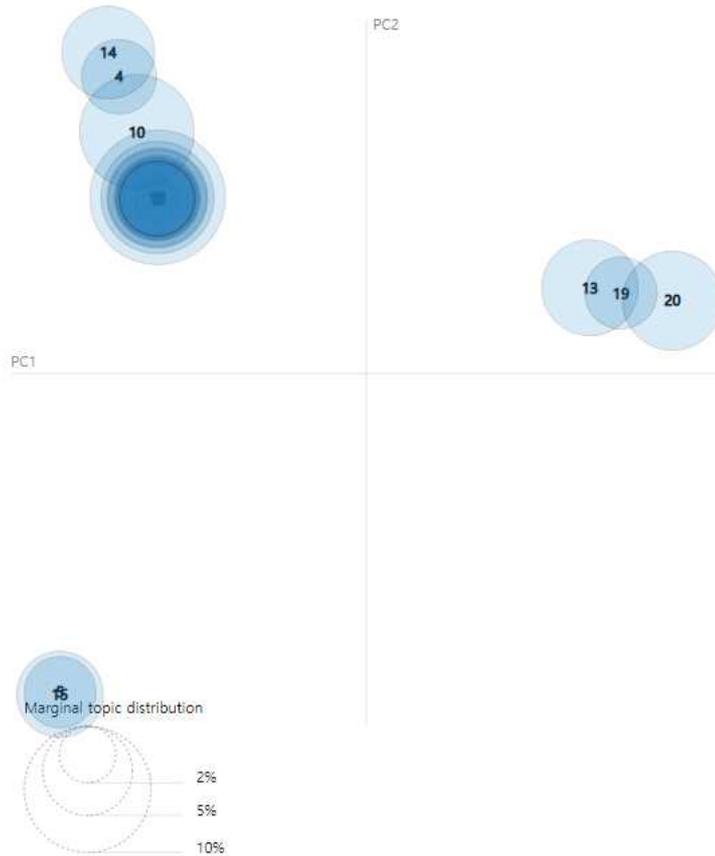
(3) 토픽 모델링 분석

코로나 이전 퇴사자를 대상으로 한 LDA 토픽 모델링 분석은 104개의 고유한 토픽과 39,527개의 문장을 대상으로 수행되었으며 그 결과는 <표 9>과 같다. 표본 전체 결과와 유사한 패턴을 보이고 있는데, 특이할 점은 토픽간 유사도 맵 (<그림 10> 참고)에서 볼 수 있듯이 토픽 #5와 #15의 내용이다. 두 토픽은 매우 유사함을 알 수 있는데, 토픽 #5에는 ‘스트레스’, ‘준비’, ‘시험’, ‘근속’, ‘자격증’이, 토픽 #15에는 ‘준비’, ‘직장’, ‘통보’, ‘문자’, ‘결혼’ 키워드가 도출되었다. 이를 통해 코로나 이전

에는 ‘자격증’ ‘시험’이나 ‘결혼’ ‘준비’ 등을 위해 이직 및 퇴사를 한 것도 큰 흐름이 있음을 파악할 수 있었다.

토픽	키워드(상위 5)
#1	'0.318*연락" + 0.184*두절" + 0.177*사업" + 0.112*학교" + 0.093*아버지"
#2	'0.404*기간" + 0.312*결정" + 0.216*전환" + 0.065*정규직" + 0.000*소속"
#3	'0.716*본인" + 0.270*기타" + 0.000*군입대" + 0.000*포기" + 0.000*아버지"
#4	'0.228*고객" + 0.199*치료" + 0.165*건강악화" + 0.155*필요" + 0.131*장기간"
#5	'0.345*스트레스" + 0.163*준비" + 0.152*시험" + 0.140*근속" + 0.110*자격증"
#6	'0.306*학업" + 0.297*진직" + 0.279*업종" + 0.116*집안" + 0.000*포기"
#7	'0.224*악화" + 0.204*관련" + 0.195*공부" + 0.154*해외" + 0.116*유학"
#8	'0.413*이사" + 0.274*휴식" + 0.142*지방" + 0.085*부족" + 0.082*직함"
#9	'0.339*급여" + 0.202*취업" + 0.167*출근" + 0.116*상태" + 0.112*거주지"
#10	'0.465*적성" + 0.167*변경" + 0.110*가족" + 0.097*간호" + 0.087*입사"
#11	'0.380*진행" + 0.221*부모님" + 0.136*간병" + 0.123*의사" + 0.098*가계"
#12	'0.260*부담" + 0.248*직종" + 0.212*대환" + 0.171*신청" + 0.106*운영"
#13	'0.383*어려움" + 0.238*교육" + 0.122*근무" + 0.090*부득이" + 0.085*생각"
#14	'0.215*병간호" + 0.168*장기" + 0.125*입원" + 0.122*병원" + 0.096*치료"
#15	'0.418*준비" + 0.251*직장" + 0.172*통보" + 0.104*문자" + 0.052*결혼"
#16	'0.305*육아" + 0.255*발생" + 0.228*복귀" + 0.205*자녀" + 0.000*포기"
#17	'0.303*복합" + 0.162*출퇴근" + 0.158*시간" + 0.143*난이도" + 0.125*거리"
#18	'0.612*만료" + 0.161*고용" + 0.142*희망" + 0.062*승계" + 0.022*원거리"
#19	'0.529*지속" + 0.304*수술" + 0.162*근무" + 0.000*포기" + 0.000*발생"
#20	'0.560*근무" + 0.312*프로젝트" + 0.064*부분" + 0.060*환경" + 0.000*군입대"

<표 9. 코로나 이전 퇴사자의 토픽 모델링 결과>



<그림 10. 코로나 이전 퇴사자의 토픽간 유사도 맵>

3.2코로나 이후

(1) 단어 빈도 분석 - 상위 30

코로나 이후(2020년 이후) 퇴사자들의 퇴직 사유에 대한 단어 빈도는 <표 10>과 같다. 퇴사 사유에 ‘코로나’가 직접 언급되기 시작했으며, 상위 10개 단어를 살펴보면 ‘개인’, ‘계약’, ‘업무’, ‘이직’, ‘종료’, ‘건강’, ‘만료’, ‘적용’, ‘결근’, ‘무단’으로 전체 표본에서의 빈도와 크게 다르지는 않음을 알 수 있었다. 코로나 이전과 비교했을 때 이직을 위한 퇴사의 비율이 큰 폭으로 줄어 코로나 사태 이후 실업률 증가 등

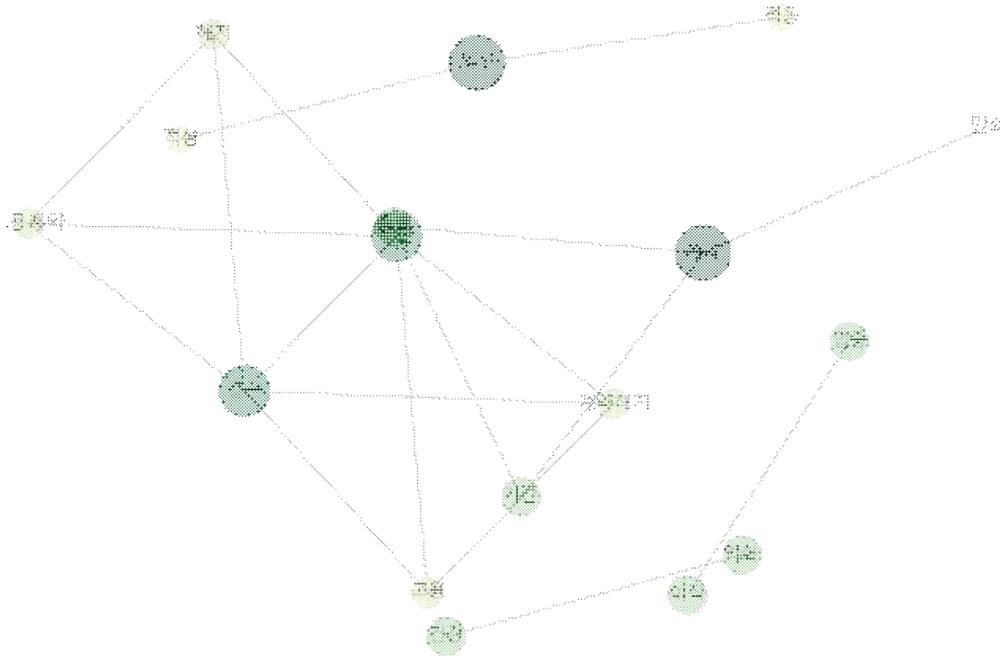
과 맞물려 자연스럽게 이직에 대한 의도가 줄어들었음을 확인할 수 있었다.

순위	단어	빈도수	순위	단어	빈도수
1	개인	3368	16	전직	289
2	계약	2222	17	기간	282
3	업무	1803	18	어려움	220
4	이직	1636	19	학업	215
5	종료	1367	20	악화	201
6	건강	1053	21	고용계약	200
7	만료	813	22	이사	192
8	적응	703	23	계약해지	181
9	결근	597	24	준비	155
10	무단	587	25	스트레스	155
11	적성	399	26	코로나	147
12	근무	381	27	집안	126
13	고용	358	28	난이도	122
14	해지	320	29	출퇴근	122
15	업종	290	30	급여	120

<표 10. 코로나 이후 빈도 분석 결과 >

(2) 네트워크 분석

코로나 이후 퇴사자들의 퇴직 사유에 대한 네트워크 분석 결과는 <그림 11>과 같다. 코로나 이전 퇴사자들의 네트워크 분석 결과 (<그림 9> 참고)와 비교해보면 네트워크 노드가 많아지고 복잡해진 것을 알 수 있다. 코로나 이전과 비교해 가장 큰 특징은 '건강 악화'가 직접적으로 네트워크 노드를 구성하며 등장했다는 것이다. 빈도 분석 결과와 종합해볼 때 '코로나'로 인한 '건강 악화'가 퇴사 사유의 직접적 원인의 하나로 확인되었다.



<그림 11. 코로나 이후 네트워크 분석 결과>

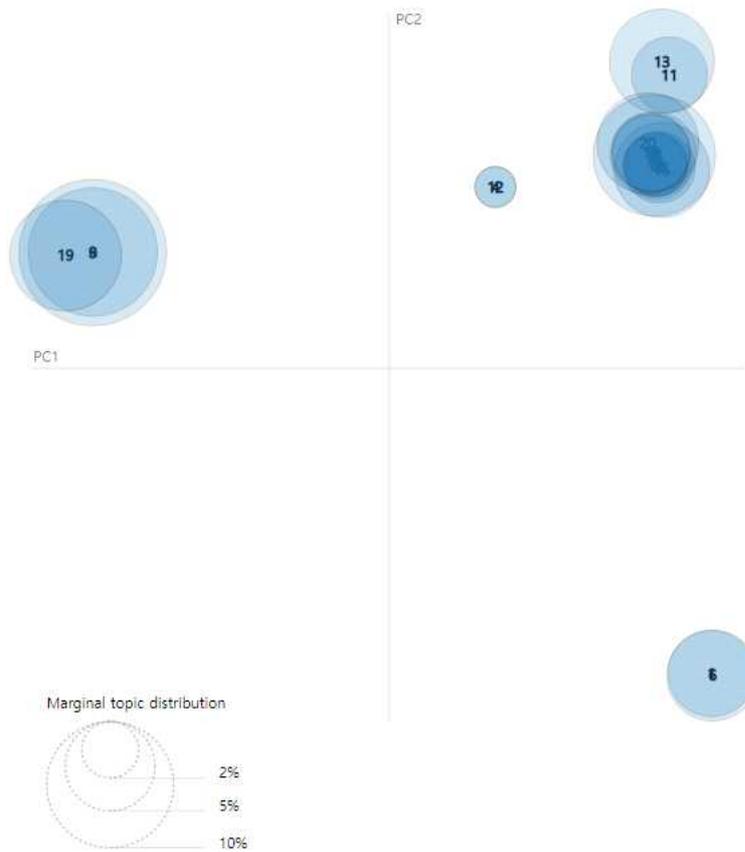
(3) 토픽 모델링 분석

코로나 이후 퇴사자를 대상으로 한 LDA 토픽 모델링 분석은 31개의 고유한 토픽과 12,683개의 문장을 대상으로 수행되었으며 그 결과는 <표 11>과 같다. 코로나 이전과 비교해서 토픽 모델링에서 파악할 수 있는 가장 큰 특징은 대부분의 토픽에서 ‘스트레스’와 ‘치료’ 키워드가 포함된다는 것이다. <표 11>에서 알 수 있듯이 토픽에 기여하는 수준이 아주 낮거나 0으로 표시된다고 하더라도 그 문장의 숨은 의미로 파악할 수 있다는 점에서, 코로나 사태 이후 건강 관련 스트레스 지수가 올라가고 퇴사 사유로 이어진다는 점을 확인할 수 있었다.

토픽	키워드(상위 5)
#1	'0.717*"업종" + 0.279*"치료" + 0.000*"집안" + 0.000*"스트레스" + 0.000*"이사"'
#2	'0.576*"악화" + 0.420*"단기" + 0.000*"치료" + 0.000*"집안" + 0.000*"스트레스"'
#3	'0.990*"스트레스" + 0.000*"업종" + 0.000*"집안" + 0.000*"치료" + 0.000*"이사"'

#4	'0.032*"업종" + 0.032*"치료" + 0.032*"집안" + 0.032*"스트레스" + 0.032*"이사"'
#5	'0.987*"난이도" + 0.000*"업종" + 0.000*"집안" + 0.000*"치료" + 0.000*"이사"'
#6	'0.790*"전직" + 0.206*"업종" + 0.000*"집안" + 0.000*"스트레스" + 0.000*"치료"'
#7	'0.989*"기타" + 0.000*"업종" + 0.000*"이사" + 0.000*"출퇴근" + 0.000*"치료"'
#8	'0.251*"고용" + 0.251*"결근" + 0.249*"무단" + 0.247*"계약해지" + 0.000*"출퇴근"'
#9	'0.301*"결근" + 0.296*"무단" + 0.201*"해지" + 0.200*"고용계약" + 0.000*"업종"'
#10	'0.986*"출퇴근" + 0.000*"업종" + 0.000*"집안" + 0.000*"스트레스" + 0.000*"이사"'
#11	'0.687*"준비" + 0.307*"가족" + 0.000*"업종" + 0.000*"출퇴근" + 0.000*"치료"'
#12	'0.032*"업종" + 0.032*"치료" + 0.032*"집안" + 0.032*"스트레스" + 0.032*"이사"'
#13	'0.474*"어려움" + 0.239*"간호" + 0.206*"희망" + 0.079*"가족" + 0.000*"업종"'
#14	'0.556*"적성" + 0.442*"기간" + 0.000*"출퇴근" + 0.000*"집안" + 0.000*"스트레스"'
#15	'0.995*"근무" + 0.000*"집안" + 0.000*"업종" + 0.000*"스트레스" + 0.000*"치료"'
#16	'0.992*"이사" + 0.000*"집안" + 0.000*"치료" + 0.000*"스트레스" + 0.000*"업종"'
#17	'0.530*"코로나" + 0.463*"부담" + 0.000*"집안" + 0.000*"출퇴근" + 0.000*"업종"'
#18	'0.631*"집안" + 0.362*"육아" + 0.000*"치료" + 0.000*"스트레스" + 0.000*"업종"'
#19	'0.295*"고용" + 0.236*"해지" + 0.233*"결근" + 0.233*"무단" + 0.000*"집안"'
#20	'0.545*"학업" + 0.245*"급여" + 0.207*"결정" + 0.000*"출퇴근" + 0.000*"업종"'

<표 11. 코로나 이후 퇴사자의 토픽 모델링 결과>



<그림 12. 코로나 이후 퇴사자의 토픽간 유사도 맵>

제4절 근속년수별분석결과

4.1 1년미만근속근로자

(1) 단어 빈도 분석 - 상위 30

1년 미만 근속자를 대상으로 추출한 퇴직 사유의 단어 빈도는 <표 12>와 같다. 본 연구를 진행하면서 1년 미만 근속 근로자의 퇴사율이 76%에 달한다는 것을 알

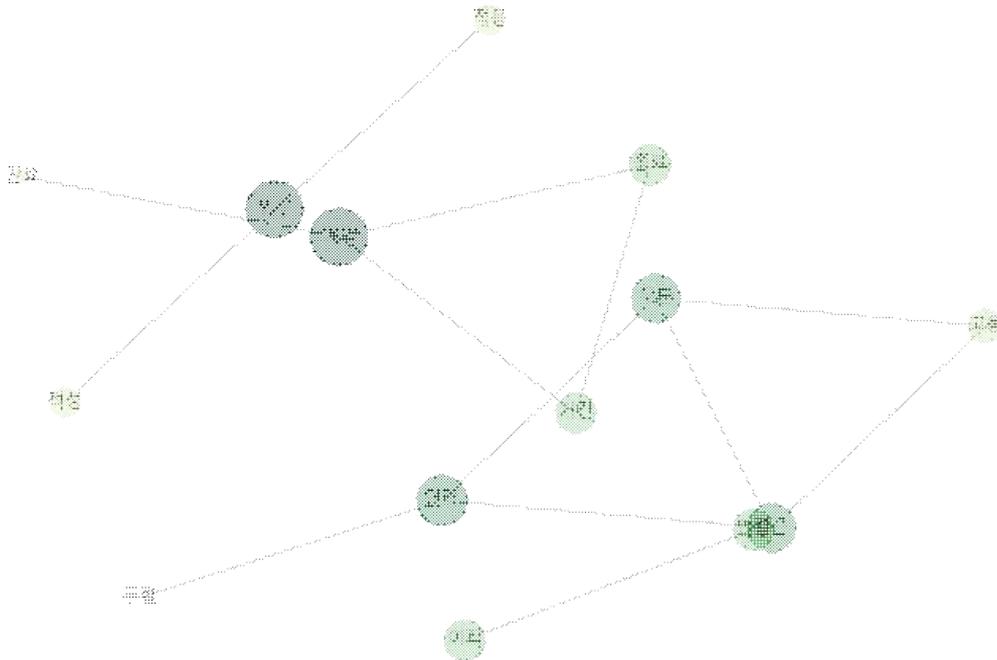
수 있었다. 자사 소속 콜센터가 아닌 아웃소싱 전문 콜센터의 특성이 반영된 것일 수 있으므로 데이터 해석에 참고가 되어야 할 것이다. 1년 미만 근속 근로자의 경우 ‘업무 적응’, ‘계약 종료’, ‘이직’의 순으로 퇴사 빈도가 높았으며, ‘무단 결근’으로 퇴사 처리 되는 경우가 많은 것으로 파악되었다.

순위	단어	빈도수	순위	단어	빈도수
1	개인	9279	15	업종	816
2	업무	7027	16	연락	815
3	계약	5025	17	기간	746
4	이직	4360	18	준비	606
5	무단	3697	19	지속	535
6	결근	3683	20	집안	531
7	종료	3679	21	고용	520
8	적응	3192	22	이사	518
9	건강	2939	23	두절	489
10	근무	1546	24	복학	449
11	만료	1534	25	간호	444
12	적성	1376	26	스트레스	434
13	학업	1120	27	결정	431
14	전직	1008	28	희망	429
15	어려움	821	29	부담	405

<표 12. 1년 미만 근속자의 단어 빈도 분석 결과>

(2) 네트워크 분석

1년 미만 근속자의 네트워크 분석 결과(<그림 13> 참고) 각각의 단어 중요도가 유사한 많은 노드의 네트워크가 구성되었다. ‘업무 적성 및 적응’, ‘계약 종료’, ‘다른 업종 이직’ 등이 비슷한 중요도를 가지는 것으로 파악되었다.



<그림 13. 1년 미만 근속자의 네트워크 분석 결과>

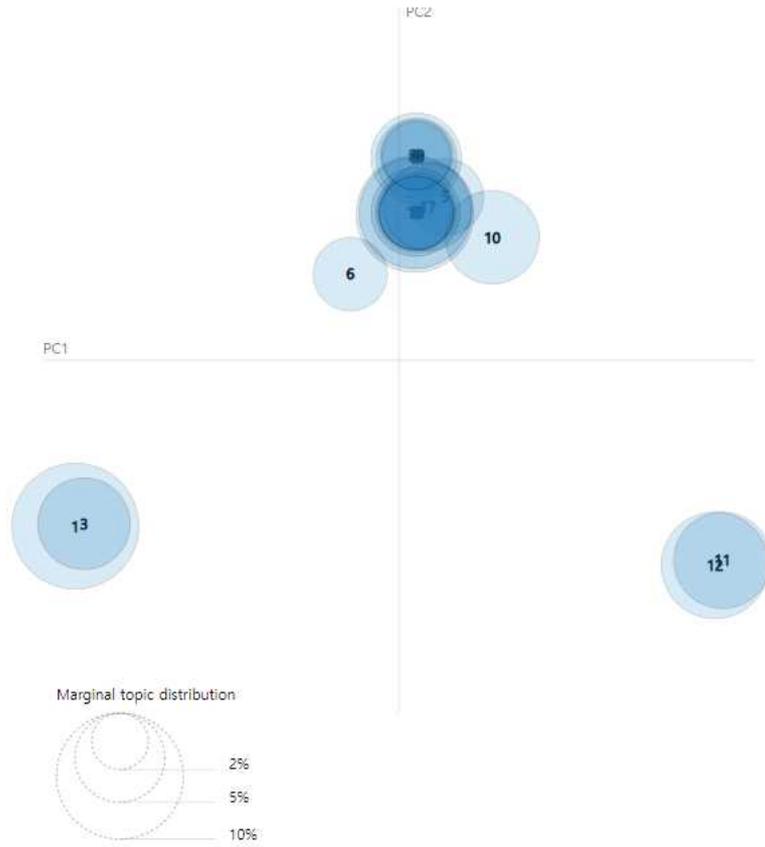
(3) 토픽 모델링 분석

1년 미만 근속자를 대상으로 한 LDA 토픽 모델링 분석은 102개의 고유한 토픽과 39,857개의 문장을 대상으로 수행되었으며 그 결과는 <표 13>과 같다. 앞서 네트워크 분석에서도 알 수 있었듯이 모든 토픽 구성 키워드들이 파편화된 양상을 나타낸다. 이는 퇴사 사유에 사용한 단어가 각각 개별적인 중요도를 가진다는 것으로 해석할 수 있는데, 이런 경우 숨겨진(연결된) 의미를 찾기가 상당히 어렵고 형식적인 사유를 입력했을 가능성이 높음을 암시한다.

토픽	키워드(상위 5)
#1	'0.441*"만료" + 0.260*"업종" + 0.126*"이사" + 0.069*"단기" + 0.058*"직종"'
#2	'0.278*"시간" + 0.246*"출퇴근" + 0.225*"출근" + 0.138*"부분" + 0.107*"유학"'

#3	'0.490* "기간" + 0.177* "부담" + 0.136* "만료" + 0.126* "대환" + 0.069* "허리" '
#4	'0.525* "적성" + 0.268* "연락" + 0.160* "두절" + 0.045* "부족" + 0.000* "해지" '
#5	'0.410* "진행" + 0.334* "난이도" + 0.142* "장기" + 0.109* "장기간" + 0.000* "해지" '
#6	'0.357* "급여" + 0.243* "변경" + 0.214* "전환" + 0.149* "부득이" + 0.032* "직종" '
#7	'0.311* "사업" + 0.226* "관련" + 0.209* "부모님" + 0.142* "공부" + 0.107* "병행" '
#8	'0.383* "결정" + 0.235* "병간호" + 0.209* "취업" + 0.123* "건강악화" + 0.046* "가족" '
#9	'0.174* "입사" + 0.150* "복귀" + 0.149* "발생" + 0.123* "이동" + 0.108* "센터" '
#10	'0.265* "스트레스" + 0.232* "치료" + 0.131* "병원" + 0.130* "입원" + 0.085* "해외" '
#11	'0.426* "어려움" + 0.162* "고객" + 0.119* "근무" + 0.118* "포기" + 0.081* "의사" '
#12	'0.563* "근무" + 0.186* "희망" + 0.108* "기타" + 0.061* "환경" + 0.054* "주말" '
#13	'0.380* "교육" + 0.208* "등보" + 0.157* "대학교" + 0.149* "문자" + 0.102* "자녀" '
#14	'0.247* "직장" + 0.199* "간호" + 0.198* "가족" + 0.144* "아버지" + 0.115* "휴식" '
#15	'0.415* "전직" + 0.302* "고용" + 0.094* "본인" + 0.078* "거리" + 0.068* "생각" '
#16	'0.545* "집안" + 0.180* "간병" + 0.161* "어머니" + 0.109* "일신상" + 0.000* "군입대" '
#17	'0.756* "학업" + 0.174* "수술" + 0.067* "필요" + 0.000* "군입대" + 0.000* "고용계약" '
#18	'0.313* "복학" + 0.229* "악화" + 0.156* "학교" + 0.101* "신청" + 0.073* "상태" '
#19	'0.576* "지속" + 0.296* "프로젝트" + 0.121* "육아" + 0.000* "해지" + 0.000* "유학" '
#20	'0.650* "준비" + 0.168* "시험" + 0.119* "자격증" + 0.059* "공부" + 0.000* "군입대" '

<표 13. 1년 미만 근속자의 토픽 모델링 결과>



<그림 14. 1년 미만 근속자의 토픽간 유사도 맵>

4.2 1년 이상 근속 근로자

(1) 단어 빈도 분석 - 상위 30

1년 이상 근속자를 대상으로 추출한 단어 빈도 분석 결과는 <표 14>와 같다. 1년 미만 근속자의 결과와 비교해볼 때 다음과 같은 특징을 파악할 수 있었다. 첫째, 이직을 위한 퇴사의 비율이 월등히 높다는 것이다. 둘째, 건강상의 이유로 퇴사하는 경우가 많아졌다. 셋째, 무단 결근의 비율이 상당히 낮아진다는 점이다.

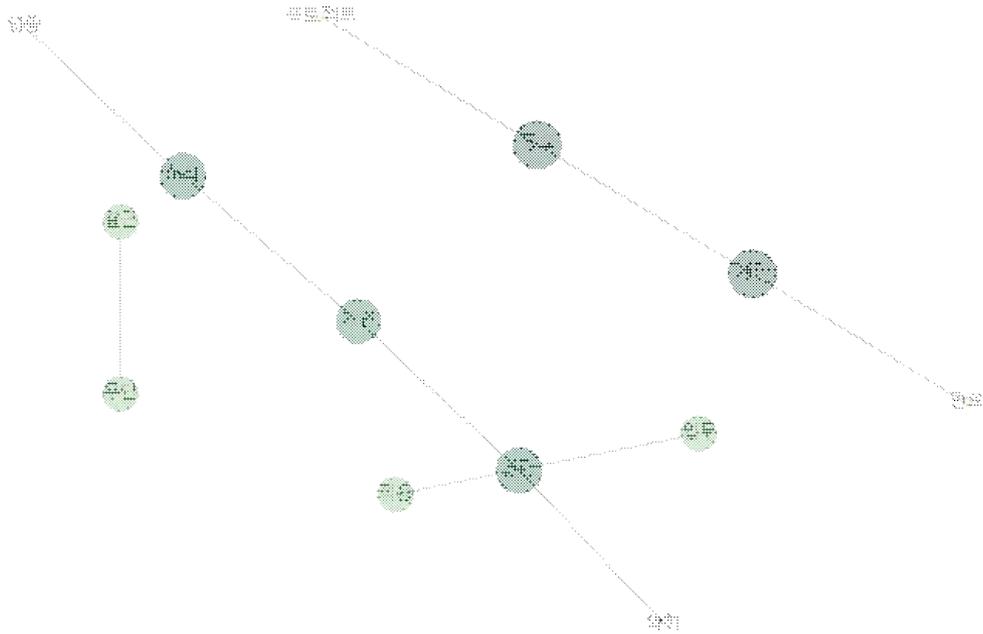
순위	단어	빈도수	순위	단어	빈도수
----	----	-----	----	----	-----

1	개인	3018		16	변경	268
2	이직	2470		17	준비	240
3	계약	1163		18	출퇴근	231
4	건강	1157		19	어려움	195
5	업무	952		20	급여	188
6	만료	947		21	결근	181
7	종료	681		22	전환	179
8	전직	498		23	무단	173
9	프로젝트	486		24	자녀	161
10	업종	431		25	결정	157
11	육아	400		26	악화	157
12	근무	387		27	적용	154
13	학업	317		28	희망	153
14	이사	310		29	이동	148
15	휴식	272		30	스트레스	147

<표 14. 1년 이상 근속자의 단어 빈도 분석 결과>

(2) 네트워크 분석

1년 이상 근속자의 네트워크 분석 결과(<그림 15> 참고) 주목할 점은 건강 상의 이유로 인한 퇴사가 급격한 비율로 높아진다는 것이다. 이는 단어 빈도 분석 결과에서도 알 수 있었는데, ‘건강 악화’가 ‘업종’ ‘이직’으로 이어진다는 점을 파악할 수 있었다.



<그림 15. 1년 이상 근속자의 네트워크 분석 결과>

(3) 토픽 모델링 분석

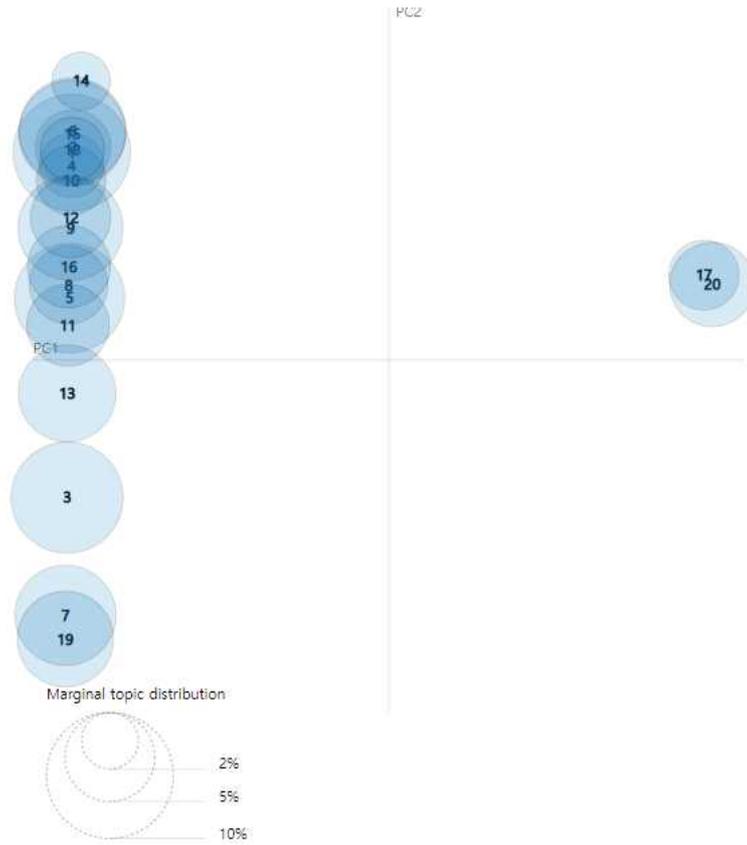
1년 이상 근속자를 대상으로 한 LDA 토픽 모델링 분석은 36개의 고유한 토픽과 12,351개의 문장을 대상으로 수행되었으며 그 결과는 <표 15>와 같다. ‘치료’, ‘전직’, ‘휴식’, ‘결혼’, ‘자녀’, ‘육아’ 등이 주요 키워드로 토픽 모델을 형성함을 알 수 있다. 또한 키워드 구성에 ‘치료’, ‘집안’ 등이 공통적으로 반복됨을 통해 개인의 건강이나 가정 관련 사유로 인한 실질적 퇴사가 증가함을 파악할 수 있다.

특히 <그림 16>의 토픽간 유사도 맵에서 토픽 #17과 #20은 다른 토픽들과는 완전히 다른 토픽 그룹을 만들고 있는 것을 확인할 수 있다. 이 토픽들의 키워드를 확인하면 ‘자녀’, ‘육아’와 관계된 것으로 역시 가정적인 사유로 퇴사하는 경우로 파악되었다.

토픽	키워드(상위 5)
----	-----------

#1	'0.307* "악화" + 0.292* "출퇴근" + 0.245* "시간" + 0.154* "거리" + 0.000* "결정" '
#2	'0.516* "치료" + 0.476* "집안" + 0.000* "지속" + 0.000* "변경" + 0.000* "결근" '
#3	'0.675* "근무" + 0.322* "어려움" + 0.000* "지속" + 0.000* "결정" + 0.000* "치료" '
#4	'0.990* "적응" + 0.000* "결정" + 0.000* "지속" + 0.000* "무단" + 0.000* "변경" '
#5	'0.403* "준비" + 0.302* "결근" + 0.292* "무단" + 0.000* "지속" + 0.000* "변경" '
#6	'0.337* "적성" + 0.257* "전환" + 0.231* "가족" + 0.172* "거주지" + 0.000* "변경" '
#7	'0.996* "전직" + 0.000* "결정" + 0.000* "지속" + 0.000* "무단" + 0.000* "변경" '
#8	'0.994* "휴식" + 0.000* "결정" + 0.000* "지속" + 0.000* "무단" + 0.000* "변경" '
#9	'0.554* "변경" + 0.234* "이동" + 0.208* "기간" + 0.000* "지속" + 0.000* "결근" '
#10	'0.991* "급여" + 0.000* "결정" + 0.000* "지속" + 0.000* "집안" + 0.000* "변경" '
#11	'0.994* "학업" + 0.000* "변경" + 0.000* "지속" + 0.000* "결정" + 0.000* "치료" '
#12	'0.502* "사업" + 0.493* "희망" + 0.000* "집안" + 0.000* "지속" + 0.000* "변경" '
#13	'0.995* "프로젝트" + 0.000* "결정" + 0.000* "무단" + 0.000* "지속" + 0.000* "치료" '
#14	'0.972* "결혼" + 0.001* "무단" + 0.001* "결정" + 0.001* "변경" + 0.001* "집안" '
#15	'0.327* "결정" + 0.250* "고용" + 0.221* "지속" + 0.199* "승계" + 0.000* "집안" '
#16	'0.993* "이사" + 0.000* "변경" + 0.000* "결근" + 0.000* "결정" + 0.000* "집안" '
#17	'0.774* "자녀" + 0.212* "육아" + 0.000* "집안" + 0.000* "지속" + 0.000* "변경" '
#18	'0.989* "스트레스" + 0.000* "무단" + 0.000* "결정" + 0.000* "변경" + 0.000* "집안" '
#19	'0.996* "업종" + 0.000* "결정" + 0.000* "무단" + 0.000* "집안" + 0.000* "변경" '
#20	'0.992* "육아" + 0.000* "결정" + 0.000* "집안" + 0.000* "지속" + 0.000* "변경" '

<표 15. 1년 이상 근속자의 토픽 모델링 결과>



<그림 16. 1년 이상 근속자의 토픽간 유사도 맵>

제5절 연령대별 분석 결과

5.1 20대

(1) 단어 빈도 분석 - 상위 30

20대를 대상으로 추출한 단어 빈도 분석 결과는 <표 16>과 같다. 상위 10개 단어를 살펴보면, ‘개인’, ‘계약’, ‘업무’, ‘이직’, ‘종료’, ‘결근’, ‘무단’, ‘적용’, ‘건강’, ‘만료’

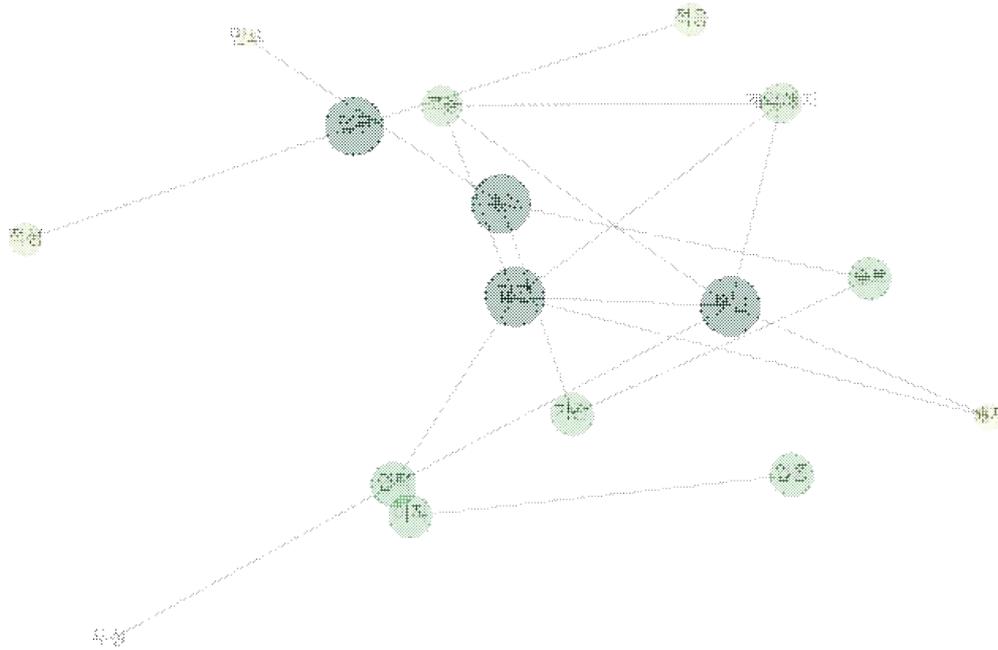
로 전체 표본을 대상으로 한 단어 빈도 결과와 큰 차이를 보이지 않았다.

순위	단어	빈도수		순위	단어	빈도수
1	개인	3355		16	업종	410
2	계약	2859		17	준비	400
3	업무	2764		18	기간	386
4	이직	2025		19	전직	382
5	종료	1969		20	어려움	347
6	결근	1929		21	고용	327
7	무단	1912		22	이사	311
8	적응	1210		23	지속	263
9	건강	1093		24	두절	257
10	만료	953		25	스트레스	239
11	학업	939		26	학교	228
12	근무	667		27	집안	228
13	적성	608		28	해지	210
14	연락	431		29	희망	206
15	복합	424		30	계약해지	204

<표 16. 20대의 빈도 분석 결과>

(2) 네트워크 분석

20대의 네트워크 분석 결과는 <그림 17>과 같다. ‘계약 종료’를 제외할 경우 ‘적성’에 의한 ‘업무 적응’ 문제가 가장 높은 중요도를 보였고, ‘무단 결근’이 퇴사로 이어지는 비중이 높은 것으로 파악되었다. 그래프가 후술할 30대, 40대와는 달리 노드 수가 많아지고 복잡해지는 것은 노드를 구성하는 키워드의 연관성이 떨어지는 것을 의미한다.



<그림 17. 20대의 네트워크 분석 결과>

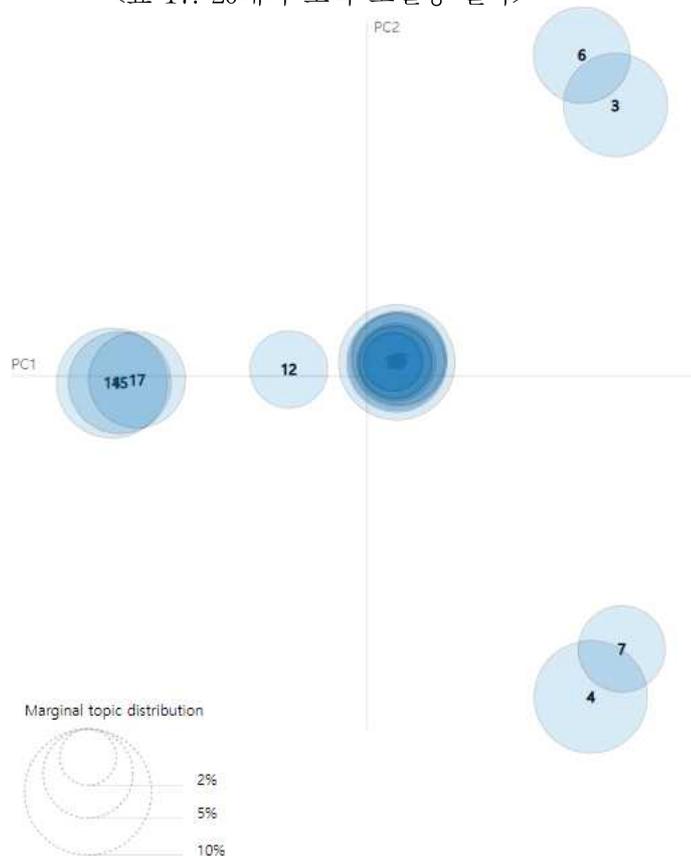
(3) 토픽 모델링 분석

20대를 대상으로 한 LDA 토픽 모델링 분석은 51개의 고유한 토픽과 18,270개의 문장을 대상으로 수행되었으며 그 결과는 <표 17>과 같다. ‘적성’, ‘어려움’, ‘부담’ 등이 주요 토픽을 구성했으며, ‘군입대’와 ‘복학’을 제외할 경우 대부분의 경우 ‘급여’, ‘장기간’이 고려된 취업을 추구하기 위한 퇴사임을 파악할 수 있었다.

토픽	키워드(상위 5)
#1	'0.524*"이사" + 0.234*"가족" + 0.170*"부모님" + 0.069*"간호" + 0.000*"고용"'
#2	'0.565*"난이도" + 0.425*"수술" + 0.000*"적성" + 0.000*"취업" + 0.000*"업종"'
#3	'0.481*"업종" + 0.207*"시간" + 0.201*"고객" + 0.108*"출퇴근" + 0.000*"고용"'
#4	'0.680*"근무" + 0.180*"직장" + 0.137*"관련" + 0.000*"고용" + 0.000*"해지"'
#5	'0.996*"적성" + 0.000*"고용" + 0.000*"계약해지" + 0.000*"해지" + 0.000*"이사"'
#6	'0.610*"전직" + 0.198*"업종" + 0.189*"급여" + 0.000*"해지" + 0.000*"적성"'
#7	'0.757*"어려움" + 0.166*"근무" + 0.071*"출퇴근" + 0.000*"취업" + 0.000*"직종"'

#8	'0.981*"부담" + 0.000*"취업" + 0.000*"직종" + 0.000*"적성" + 0.000*"이사"'
#9	'0.315*"집안" + 0.300*"군입대" + 0.266*"공부" + 0.115*"병간호" + 0.000*"고용"'
#10	'0.431*"연락" + 0.300*"스트레스" + 0.266*"두절" + 0.000*"해지" + 0.000*"고용"'
#11	'0.978*"장기" + 0.000*"취업" + 0.000*"업종" + 0.000*"적성" + 0.000*"집안"'
#12	'0.402*"휴식" + 0.325*"교육" + 0.265*"취업" + 0.000*"고용" + 0.000*"적성"'
#13	'0.971*"통보" + 0.001*"직종" + 0.001*"업종" + 0.001*"적성" + 0.001*"집안"'
#14	'0.343*"악화" + 0.268*"복학" + 0.249*"지속" + 0.137*"대학교" + 0.000*"해지"'
#15	'0.337*"학교" + 0.290*"희망" + 0.198*"복학" + 0.171*"프로젝트" + 0.000*"고용"'
#16	'0.679*"결정" + 0.310*"병원" + 0.000*"취업" + 0.000*"적성" + 0.000*"집안"'
#17	'0.666*"준비" + 0.160*"대학" + 0.088*"취업" + 0.082*"복학" + 0.000*"적성"'
#18	'0.305*"진행" + 0.290*"단기" + 0.278*"직종" + 0.124*"간호" + 0.000*"해지"'
#19	'0.994*"기간" + 0.000*"집안" + 0.000*"적성" + 0.000*"취업" + 0.000*"업종"'
#20	'0.497*"변경" + 0.494*"치료" + 0.000*"적성" + 0.000*"취업" + 0.000*"직종"'

<표 17. 20대의 토픽 모델링 결과>



<그림 18. 20대의 토픽간 유사도 맵>

5.2 30대

30대를 대상으로 추출한 퇴직 사유의 단어 빈도 분석 결과는 <표 18>과 같다. 주목할만한 특징은 첫째, 30대에서 ‘이직’을 위한 퇴사 비율이 가장 높다는 것이다. 둘째, ‘육아’라는 키워드가 처음 등장하면서 출산 후 퇴사로 이어지는 형태가 명확하게 파악되었다. 셋째, ‘학업’에 대한 비중이 상당 부분 낮아졌다.

(1) 단어 빈도 분석 - 상위 30

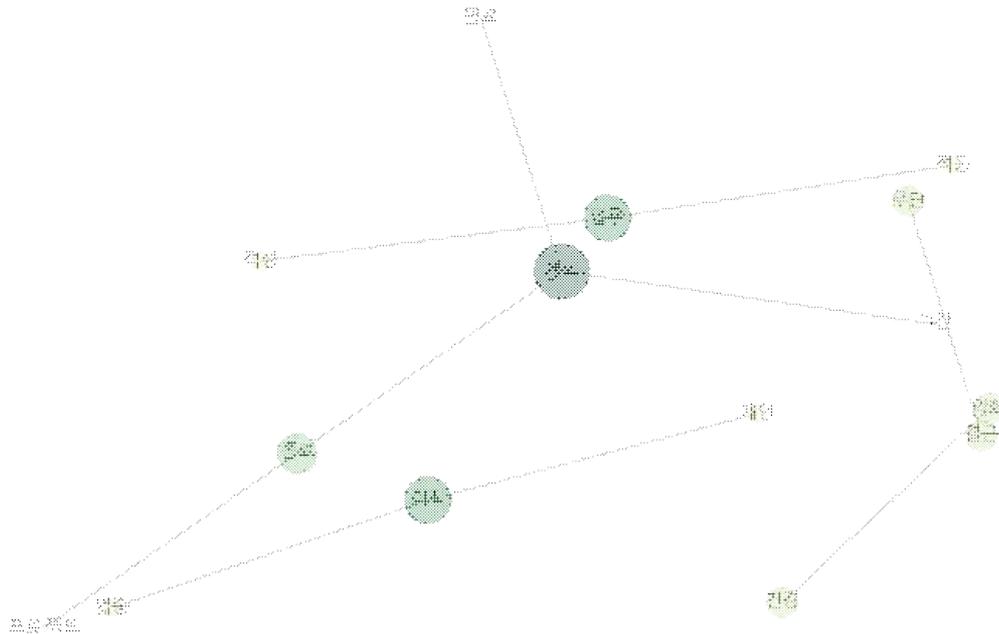
순위	단어	빈도수	순위	단어	빈도수
1	개인	4533	16	어려움	390
2	이직	3068	17	준비	362
3	업무	3027	18	이사	328
4	계약	1880	19	프로젝트	309
5	건강	1606	20	육아	306
6	무단	1458	21	연락	302
7	결근	1450	22	기간	278
8	종료	1399	23	급여	272
9	적용	1238	24	집안	268
10	만료	778	25	결정	263
11	근무	746	26	지속	255
12	전직	701	27	희망	252
13	업종	612	28	출퇴근	237
14	적성	524	29	악화	234
15	학업	448	30	스트레스	233

<표 18. 30대의 단어 빈도 분석 결과>

(2) 네트워크 분석

30대의 퇴사 사유 네트워크 분석 결과(<그림 19> 참고) 특징은 다음과 같다.

첫째, ‘계약 종료’ 외에 다른 ‘업종’으로의 ‘이직’의 중요도가 명확하게 나타난다. 둘째, 20대 분석 결과에는 보이지 않았던 ‘건강’ ‘악화’가 나타나면서 건강 문제로 인한 퇴사가 시작되는 시점임을 알 수 있다.



<그림 19. 30대의 네트워크 분석 결과 >

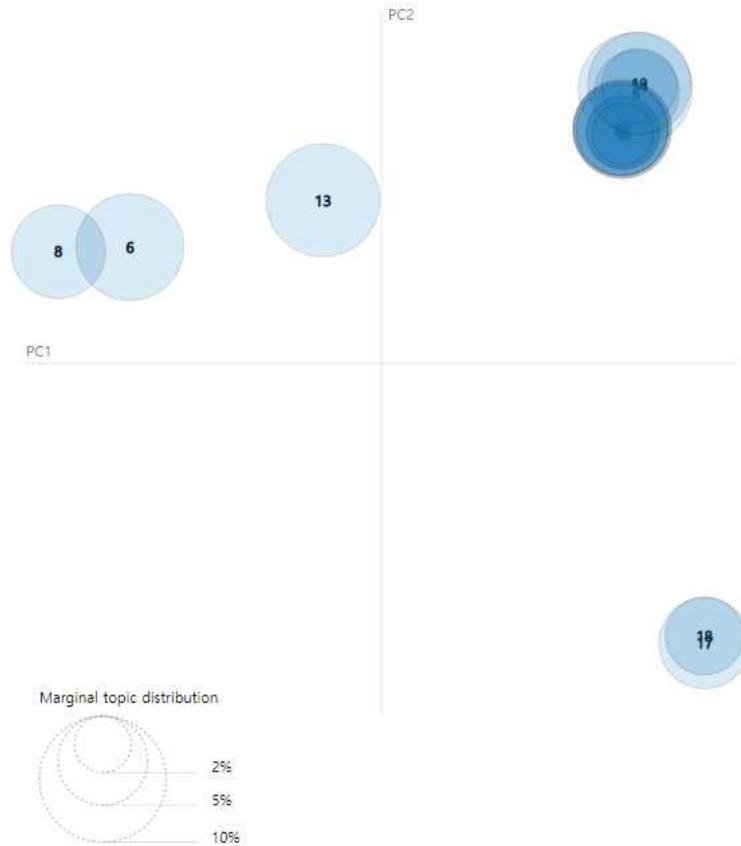
(3) 토픽 모델링 분석

30대를 대상으로 한 LDA 토픽 모델링 분석은 59개의 고유한 토픽과 19,561개의 문장을 대상으로 수행되었으며 그 결과는 <표 19>와 같다. ‘업종’과 ‘전직’은 빈도 분석과 네트워크 분석에서도 파악된 특징이었는데, 토픽 모델링을 통해 ‘집안’, ‘부모님’, ‘육아’ 문제가 매우 중요한 주제로 도출됨을 파악할 수 있었다. 이는 여성 근로자가 30대에 접어들어 결혼, 출산 등을 통해 육아 문제로 이어져 퇴사로 이어지는 패턴으로 추정할 수 있을 것이다.

토픽	키워드(상위 5)
----	-----------

#1	'0.388*이사" + 0.328*사업" + 0.280*가족" + 0.000*부모님" + 0.000*거주지"
#2	'0.713*만료" + 0.166*치료" + 0.084*병원" + 0.035*입원" + 0.000*고용"
#3	'0.395*악화" + 0.279*난이도" + 0.217*장기" + 0.105*입원" + 0.000*집안"
#4	'0.578*학업" + 0.215*부담" + 0.203*대한" + 0.000*고용" + 0.000*부모님"
#5	'0.398*직중" + 0.315*병간호" + 0.280*통보" + 0.000*거주지" + 0.000*시간"
#6	'0.617*업종" + 0.170*전직" + 0.118*이동" + 0.093*공부" + 0.000*고용"
#7	'0.990*집안" + 0.000*시간" + 0.000*입원" + 0.000*거주지" + 0.000*근무"
#8	'0.671*전직" + 0.325*어려움" + 0.000*고용" + 0.000*시간" + 0.000*부모님"
#9	'0.440*간호" + 0.311*부모님" + 0.241*어머니" + 0.000*고용" + 0.000*시간"
#10	'0.686*적성" + 0.311*희망" + 0.000*거주지" + 0.000*부모님" + 0.000*입원"
#11	'0.425*기간" + 0.268*육아" + 0.200*본인" + 0.104*자녀" + 0.000*고용"
#12	'0.530*전환" + 0.463*시간" + 0.000*고용" + 0.000*부모님" + 0.000*거주지"
#13	'0.671*근무" + 0.222*지속" + 0.104*어려움" + 0.000*고용" + 0.000*부모님"
#14	'0.532*직장" + 0.420*휴식" + 0.042*입사" + 0.000*거주지" + 0.000*두절"
#15	'0.306*스트레스" + 0.295*교육" + 0.198*고객" + 0.197*수술" + 0.000*고용"
#16	'0.990*결정" + 0.000*시간" + 0.000*입원" + 0.000*집안" + 0.000*거주지"
#17	'0.534*준비" + 0.200*거리" + 0.148*취업" + 0.114*출퇴근" + 0.000*고용"
#18	'0.307*출퇴근" + 0.255*지방" + 0.245*거주지" + 0.186*결혼" + 0.000*입원"
#19	'0.306*급여" + 0.226*변경" + 0.217*진행" + 0.138*관련" + 0.111*입사"
#20	'0.341*프로젝트" + 0.332*연락" + 0.190*두절" + 0.133*출근" + 0.000*고용"

<표 19. 30대의 토픽 모델링 결과>



<그림 20. 30대의 토픽간 유사도 맵>

5.3 40대 이상

(1) 단어 빈도 분석 - 상위 30

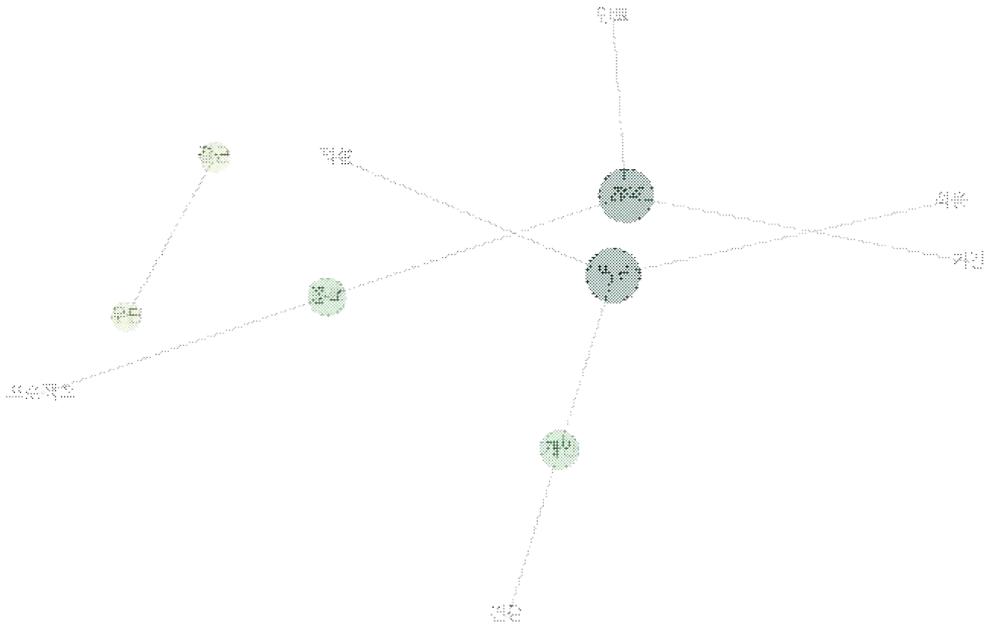
40대 이상을 대상으로 추출한 퇴직 사유의 단어 빈도 분석 결과는 <표 20>과 같다. 그 특징을 살펴보면 첫째, ‘개인’, ‘업무’, ‘이직’, ‘계약’, ‘건강’, ‘종료’, ‘적응’의 단어는 20대와 30대와 유사한 패턴을 보였으나, ‘무단 결근’의 비중은 훨씬 낮아진 것을 파악할 수 있다. 둘째, ‘자녀’와 ‘간호’가 처음으로 등장해 가정적 원인으로 인한 퇴사 사유가 ‘육아’에서 ‘자녀’와 ‘병간호’로 옮겨가는 형태를 파악할 수 있었다.

순위	단어	빈도수		순위	단어	빈도수
1	개인	4381		16	자녀	289
2	업무	2171		17	어려움	278
3	이직	1716		18	간호	227
4	계약	1442		19	업종	225
5	건강	1390		20	변경	201
6	종료	986		21	기간	192
7	적용	893		22	이사	188
8	만료	747		23	치료	171
9	근무	515		24	병간호	170
10	무단	498		25	고용	163
11	결근	484		26	휴식	155
12	전직	420		27	출퇴근	154
13	적성	382		28	지속	151
14	육아	327		29	부담	147
15	프로젝트	300		30	집안	145

<표 20. 40대의 빈도 분석 결과>

(2) 네트워크 분석

40대 이상을 대상으로 한 네트워크 분석 결과는 <그림 21>과 같다. ‘계약’과 ‘업무’가 중요한 단어로 노드를 형성하고 있다. 네트워크 구성이 20대와 30대에 비해 간결해지는 것은 각 노드를 구성하는 키워드의 연관성이 높다는 것을 의미한다. 이는 연령대가 높아질수록 더욱 명확하게 나타날 것으로 예측할 수 있다. 특히 ‘이직’에 해당하는 그래프가 나타나지 않는다는 점도 유의해서 볼만 하다. 이는 퇴직 사유에는 ‘이직’이 포함된다 하더라도 업종 전환이나 기타 구체적인 사유와 연결되지 않았다는 것으로 해석할 수 있다.



<그림 21. 40대의 네트워크 분석 결과>

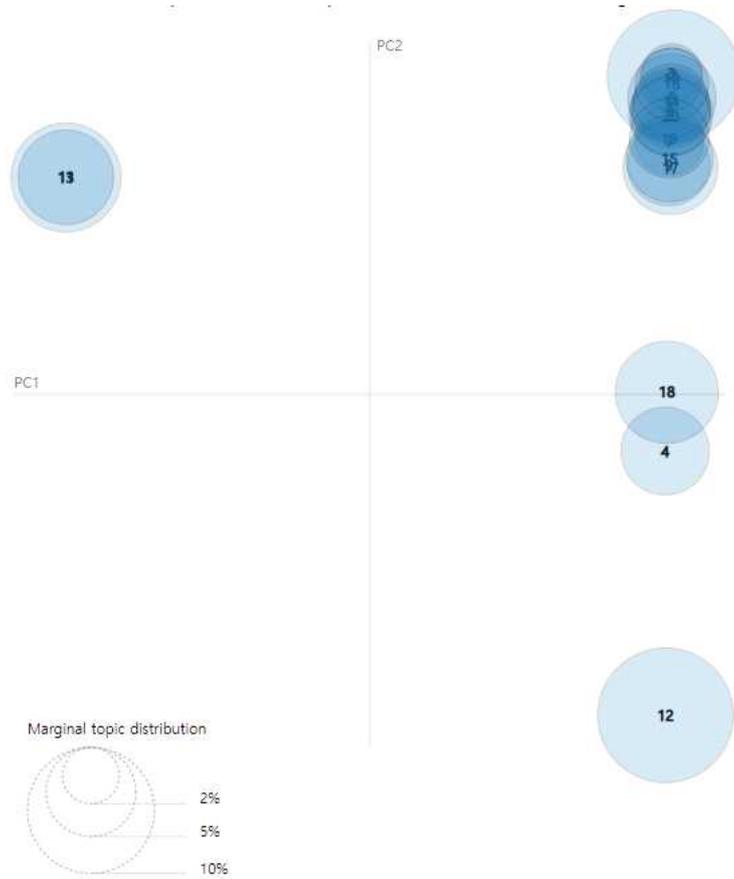
(3) 토픽 모델링 분석

40대 이상을 대상으로 한 LDA 토픽 모델링 분석은 38개의 고유한 토픽과 14,266개의 문장을 대상으로 수행되었으며 그 결과는 <표 21>과 같다. 그 특징을 살펴보면, 첫째, 40대 이상에서도 ‘적성’에 따른 ‘업종’과 ‘근무’의 어려움을 여전히 호소하고 있다는 점이다. 둘째, ‘자녀’와 ‘육아’의 부담에 따른 퇴사가 여전히 높은 비중을 차지하고 있음을 알 수 있다. 셋째, ‘치료’, ‘병간호’, ‘간병’ 등 의료적 행위에 따른 사유가 높은 비중으로 토픽을 이루고 있음을 알 수 있다.

토픽	키워드(상위 5)
#1	'0.635*"이사" + 0.357*"직장" + 0.000*"전환" + 0.000*"난이도" + 0.000*"변경"'
#2	'0.252*"프로젝트" + 0.231*"기간" + 0.210*"고용" + 0.197*"급여" + 0.107*"승계"'
#3	'0.984*"악화" + 0.000*"전환" + 0.000*"난이도" + 0.000*"변경" + 0.000*"연락"'
#4	'0.996*"적성" + 0.000*"전환" + 0.000*"업종" + 0.000*"변경" + 0.000*"시간"'
#5	'0.501*"지속" + 0.491*"병간호" + 0.000*"변경" + 0.000*"업종" + 0.000*"연락"'

#6	'0.553*치료" + 0.439*수술" + 0.000*전환" + 0.000*난이도" + 0.000*변경"'
#7	'0.991*변경" + 0.000*전환" + 0.000*난이도" + 0.000*연락" + 0.000*시간"'
#8	'0.383*부담" + 0.323*부모님" + 0.288*전환" + 0.000*변경" + 0.000*시간"'
#9	'0.992*육아" + 0.000*시간" + 0.000*업종" + 0.000*변경" + 0.000*적성"'
#10	'0.985*휴식" + 0.000*전환" + 0.000*업종" + 0.000*변경" + 0.000*연락"'
#11	'0.861*근무" + 0.135*어려움" + 0.000*전환" + 0.000*업종" + 0.000*변경"'
#12	'0.504*무단" + 0.494*결근" + 0.000*전환" + 0.000*변경" + 0.000*연락"'
#13	'0.583*업종" + 0.413*어려움" + 0.000*전환" + 0.000*변경" + 0.000*연락"'
#14	'0.560*간병" + 0.432*연락" + 0.000*전환" + 0.000*변경" + 0.000*난이도"'
#15	'0.555*간호" + 0.439*가족" + 0.000*전환" + 0.000*변경" + 0.000*난이도"'
#16	'0.519*희망" + 0.471*본인" + 0.000*전환" + 0.000*업종" + 0.000*변경"'
#17	'0.369*결정" + 0.344*난이도" + 0.283*집안" + 0.000*업종" + 0.000*전환"'
#18	'0.825*전직" + 0.172*스트레스" + 0.000*전환" + 0.000*변경" + 0.000*시간"'
#19	'0.991*자녀" + 0.000*난이도" + 0.000*변경" + 0.000*전환" + 0.000*연락"'
#20	'0.579*출퇴근" + 0.413*시간" + 0.000*전환" + 0.000*변경" + 0.000*연락"'

<표 21. 40대의 토픽 모델링 결과>



<그림 22. 40대의 토픽간 유사도 맵>

제 V장 결론 및 시사점

제1절 연구결과의 요약

본 연구는 전 세계적으로 코로나19(COVID-19) 바이러스의 유행이 급격한 속도로 확산되고 장기화되는 시점에서 급격한 성장세를 보이고 있는 콜센터에 근무하는 상담원들의 이직 및 퇴사 요인을 분석하고자 하였다. 기존 연구들이 대부분 면담 및 설문 등을 통한 질적 연구 방법을 채택했지만 표본의 수가 절대적으로 작아서 그 신뢰도를 확보하기가 쉽지 않았다. 따라서 본 연구에서는 국내 최대의 콜센터 아웃소싱 업체의 8년간 52,212건의 사직서의 '구체적 퇴사 사유'를 빅데이터 분석 방법을 활용하여 실질적인 퇴사 사유를 파악하고자 하였다.

본 연구에서는 자료의 특성에 따라 표본을 전체, 성별, 코로나 전후별, 근속기간별, 연령별로 분류하고 각 분류에 대해 빅데이터 분석의 텍스트 마이닝 기법인 빈도 분석, 네트워크 분석, LDA 토픽 모델링 분석을 각각 실행해 오해석을 최대한 방지하고 각 분석 기법의 단점을 보완하려고 노력하였다는 점에서 의의가 있다.

본 연구의 결과를 요약하면 다음과 같다.

첫째, 연구 대상의 인구 통계학적 분석에서 흥미로운 결과가 도출되었다. 이 부분은 연구 대상이 특정 기업 소속 콜센터가 아니라 다양한 기업과 기관들의 콜센터를 대행하는 아웃소싱 전문 BPO²⁾ 업체임을 고려할 필요가 있다. 8년간 52,212명을 대상으로 분석한 결과 남자가 30%, 여자가 70%의 비율을 보였다. 또한 근속 1년 미만인 경우가 76%에 달해 기존 연구 결과보다 훨씬 높은 비율을 보였다. 연령대는 20대, 30대 40대가 고른 분포를 보인 것으로 나타났다.

둘째, 분석 대상 전체를 대상으로 추출한 퇴직 사유의 단어 빈도의 상위 10개 단어를 살펴보면, '개인', '업무', '이직', '계약', '종료', '건강', '무단', '결근', '적용', '만료'이다. 의미 파악이 어려운 '개인 (사유)'를 제외하면, 업무 적용 문제, 프로젝트 종료 및 기간 만료로 인한 계약 만료, 이직을 위한 퇴사 순으로 빈도가 높은 것을

2) Business Process Outsourcing : 업무 처리의 아웃소싱을 담당하는 기업

알 수 있었다. 추가적으로 전체 표본을 대상으로 한 토픽 모델링의 결과를 보면 ‘학업’과 ‘병간호’, ‘급여’와 ‘병 치료’가 매우 중요한 퇴사 및 이직의 원인인 것으로 파악되었다.

셋째, 성별 분석 결과에서는 남자는 이직을 위한 퇴사 비율이 높았고 여자는 업무 적응을 이유로 퇴사하는 경우가 높았다. 남자의 경우 개인 ‘사업’을 위해 퇴사하는 경우가 높고, 무단 결근을 한 이후에는 연락이 두절되는 경우가 높은 것으로 파악되었다. 여자의 경우 업무 적응 외에도 ‘건강’상의 이유로 이직하는 경우가 많이 매핑되고 있으며 특히 ‘간호’, ‘집안’, ‘자녀’, ‘육아’ 관련 키워드가 각 토픽에 많이 등장하는 특징을 보였다.

넷째, 코로나 사태 이전 퇴사자들의 경우에는 전체 표본을 대상으로 분석한 결과와 일치하며 큰 특이사항은 발견되지 않았지만, 네트워크 분석 결과 코로나 이전에는 ‘자격증’ ‘시험’이나 ‘결혼’ ‘준비’ 등을 위해 이직 및 퇴사를 한 것도 큰 흐름이 있음을 파악할 수 있었다. 코로나 사태 이후(2020년도 이후)에는 퇴사 사유에 ‘코로나’가 직접 언급되기 시작했으며, 코로나 이전과 비교했을 때 이직을 위한 퇴사의 비율이 큰 폭으로 줄어 코로나 사태 이후 실업률 증가 등과 맞물려 자연스럽게 이직에 대한 의도가 줄어들었음을 확인할 수 있었다. 특히 주목할 점은 코로나 사태 이후에는 대부분의 토픽에서 ‘스트레스’와 ‘치료’ 키워드가 포함된다는 것이다. <표 11>에서 알 수 있듯이 토픽에 기여하는 수준이 아주 낮거나 0으로 표시된다고 하더라도 그 문장의 숨은 의미로 파악할 수 있다는 점에서, 코로나 사태 이후 건강 관련 스트레스 지수가 올라가고 퇴사 사유로 이어진다는 점을 확인할 수 있었다.

다섯째, 1년 이상 근속자의 경우 이직을 위한 퇴사의 비율이 급격히 높아지는 것을 확인할 수 있었고, 건강상의 이유로 퇴사하는 경우가 많아지는 경향을 파악하였다. 특히 1년 미만 단기 근속자에 비해 무단 결근의 비율이 상당히 낮아지는 특징을 보였다.

여섯째, 연령별 분석에서는 각 연령대에서 확연하게 차이를 보였다. 20대의 경우, ‘군입대’와 ‘복학’을 제외할 경우 대부분의 경우 ‘급여’, ‘장기간’이 고려된 안정적 취업을 위한 퇴사임을 파악할 수 있었다. 30대의 경우 다른 연령대에 비해 ‘이직’을 위한 퇴사 비율이 가장 높았다. 특히 30대에 ‘육아’라는 키워드가 처음으로 높은 빈

도수를 보이는데, 토픽 모델링을 통해서도 ‘집안’, ‘부모님’, ‘육아’ 문제가 매우 중요한 주제로 파악됨으로써, 여성 근로자가 30대에 접어들어 결혼, 출산 등을 통해 육아 문제로 이어져 퇴사로 이어지는 패턴으로 추정할 수 있다. 40대도 ‘적성’에 따른 ‘업종’과 ‘근무’의 어려움을 여전히 호소하고 있으며, , ‘자녀’와 ‘육아’의 부담에 따른 퇴사가 여전히 높은 비중을 차지하고 있음을 알 수 있었다. 특히 ‘치료’, ‘병간호’, ‘간병’ 등이 높은 비중으로 토픽을 이루고 있어 의료행위에 따른 추가적인 부담감이 퇴사로 이어진다는 것을 유추할 수 있었다.

제2절 연구의 논의 및 시사점

본 연구의 논의 및 시사점은 다음과 같다.

첫째, 이직 및 퇴사 사유를 연구하기 위해 기업의 사직원을 대상으로 빅데이터 분석을 수행했고, 해석의 정확도를 높이기 위해 빈도 분석, 네트워크 분석, 토픽 모델링 기법을 동시에 적용해서 각 분류별 인사이트를 찾았다는 점에서 그 의의를 찾을 수 있다.

둘째, 콜센터 상담원들의 고용 안정과 기업의 인재유지를 위해 본 연구에서 분류한 인구통계학적 분류별로 유효한 정책들을 수립하고 시행할 필요가 있다. 본 연구에서 도출된 결론을 바탕으로 기업의 인사 및 교육 정책을 수립한다면 콜센터 근무자들의 잦은 이직율을 줄일 수 있고 근속 기간도 늘릴 수 있을 것으로 예상된다.

셋째, 코로나 사태가 장기화되는 시점에서 건강 관련 스트레스 지수가 높아짐을 알 수 있었다. 실제 코로나 감염 여부와는 상관없이 직원들의 건강 관련 걱정을 해소할 수 있는 방안을 마련하는 것이 직원들의 조직 만족도를 높이는데 기여할 것으로 보인다.

제3절 연구한계점 및 향후 연구방향

첫째, 본 연구에서 사용한 데이터는 국내 최대 콜센터 아웃소싱 업체 한 곳에 만

국한되어 있다는 점이다. 데이터의 양은 충분하게 확보했지만, 특수한 조직 문화 및 고용 계약 조건 등이 개별 근로자에게 미친 영향까지는 파악하지 못했다. 특정 산업군에 매몰되지 않는다는 장점이 있지만, 서비스를 제공하는 기업의 조직 몰입 및 만족도에는 제한이 있을 수 있다.

둘째, 사직원의 '구체적 사유' 상당수가 '개인 사정' 또는 '일신상의 사유'로 되어 있어 실질적인 분석에 제한이 있었다는 점이다.

셋째, 사직원 외의 인사 정보 및 근무 행태를 파악할 수 있는 정보가 데이터화 되지 못했다는 점이다. 퇴사 및 이직을 결심하는데 있어 고용의 형태와 급여 등이 반드시 영향을 미치므로, 개인 신상 정보 및 근무 행태에 대한 자료를 복합적으로 연구할 필요가 있다.

이러한 한계점을 보완하기 위해 향후 연구에서는 사직원의 퇴사 사유와 기업의 인사 관련 데이터, 사내 게시판, 면담 일지 등을 다각적으로 연결해 연구할 필요가 있다. 추가적인 선행요인을 밝히기 위한 이직 관련 탐색 연구가 지속적으로 수행되면 더욱 통찰력 있는 연구와 기업의 고용안전 전략 도출이 가능해질 것이다.

참고문헌

- 김정숙(2012). 빅데이터 활용과 관련기술 고찰, 한국콘텐츠학회지, 10(1), 34-40.
- 김정임, 최보람(2015) 콜센터 근로자의 감정노동, 스트레스반응 및 이직의도에 관한 융합적인 연구
- 박선우(2018). 빅데이터 시대와 데이터 융합, 정보통신방송정책, 30(1), 1-24.
- 손정렬(2020). 빅데이터 연구에서의 빅마켓, 빅네임 그리고 빅네트워크:
국제학술지를 통해 본 도시 빅데이터 연구동향, 대한지리학회지, 5 (2),
161-179
- 심원섭, 최승묵, 심창섭(2018). 관광 빅데이터 분석의 주요 쟁점, 관광연구논총,
30(3), 3-2 .
- 이서구(2015). 빅데이터 분석에 관한 마케팅적 접근, 대한경영학회지, 28(1), 21-35
- 이석인, 이수화(2008) 중업원의 이직의도 영향요인에 관한 실증연구,
한국콘텐츠학회논문지 '08 Vol. 8 No. 2
- 허정은, 양창훈(2013). 네트워크 분석을 통한 융합연구 구조 분석, 기술혁신학회지,
16(4), 883-912.
- Bae, S. H., Kim, M. S. (2009). Mediatinf effect of the perception of organiza-
tional politics on the relationship between organizational justice and
supervisory trust, job strain and turnover intention, Journal of Human
Resource Management Research, 16(3), 123-148
- Carr, J. C., Boyar, S. L., Gregory, B. T. (2007). The moderating effect of
work-family centrality on work--family conflict, organizational atti-
tudes, and turnover behavior, Journal of management, 34(2), 244-262
- Cheung, M. W. L., Jak, S. (2016). Analyzing big data in psychology: A split/an-
alyze/meta-analyze approach, Frontiers in Psychology, 7, 738. (in
Korean)
- Choi J. S., Kim, Y. (2019). A study on research trend for nurses' workplace
bullying in Korea: Focusing on semantic network analysis and topic

- modeling, *Korean Journal of Occupational Health Nursing*, 28(4), 221-229. (in Korean)
- David M. Blei, Andrew Y. Ng & Michael I. Jordan. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Eun, H., Koo, J., Jung, T. (2018). Qualitative research of job turnover and stay, based on job embeddedness, *Korean Journal of Culture and Social Issues*, 24(2), 221-250. (in Korean).
- Feldman R, Dagan I. (1995). Knowledge Discovery in Textual Databases (KDT), *INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, 1, 112-117
- Ferris, G. R., Kacmar, K. M. (1992). Perceptions of organizational politics, *Journal of Management*, 18, 93-116.
- Ferris, G. R., Frink, D. D., Beehr, T., Gilmore, D. (1995). Political fairness and fair politics: The conceptual integration of divergent constructs.” in *Organizational Politics, Justice, and Support*, edited by Cropanzano, R. S., Kacmar, K. M. Quorum Books: Westport, CT
- George, G., M. R. Haas, & A. Pentland(2014), “Big data and management,” *Academy of Management Journal*, 57(2), 321-326.
- Han, G., & Jin, S. (2014). Introduction to big data and the case study of its applications, *Journal of the Korean Data Analysis Society*, 16(3), 1337-1351. (in Korean)
- Herrbach, O., Mignonac, K., Gatignon, A. L. (2007). Exploring the role of perceived external prestige in managers' turnover intentions, *The International Journal of Human Resource Management*, 15(8), 1390-1407
- Hur, S. J., Paek, J. (2012). A Qualitative study on psychological changes in secretaries who change jobs, *Korean Association Of Secretarial Studies*, 21(3), 85-108. (in Korean)
- Jang, S. (2008). The qualitative research on the reentry process of highly-edu-

- cated housewives to their job market, *The Women's Studies*, 74(1), 79-104. (in Korean)
- Jeon, H. S., Wang, E. J. (2014). A study on an exit interview process, influencing the withdrawal of a turnover decision: Semiconductor manufacturing plant cases, *Korean Journal of Industrial and Organizational Psychology*, 27(4), 805-830. (in Korean)
- Job Korea (2017). 76% of office workers 'If the welfare system is good, I will leave even if the salary is low'.
http://www.jobkorea.co.kr/GoodJob/Tip/View?News_No=13900&schCtgr=0&schTxt=%EC%9D%B4%EC%A7%81&Page=7
- Jung, Y., Suh, Y. (2019). Mining the voice of employees: A text mining approach to identifying and analyzing job satisfaction factors from online employee reviews, *Decision Support Systems*, 123, 113074.
- Kamasak, R. (2019). The influence of perceived external prestige on job satisfaction and turnover intentions, *Journal of Business and Economics*, 2(3), 209-221.
- Kang, J. (2016). Comparative analysis of job satisfaction Factors, using LDA topic modeling by industries: The case study of job planet reviews, *Journal of Information Technology Services*, 15(3), 157-17
- Kao, A., & Poteet, S. (2007). *Natural language processing and text mining*. London: Springer Science & Business Media.
- Kim, E. S., Jyung, C. Y. (2010). The role of general characteristics and job satisfaction in voluntary turnover process, *The Journal of Career Education Research*, 23(3), 107-122. (in Korean)
- Kim, D. Kang, J., Lim, J. I. (2016). Comprehensive analysis of job satisfaction factors, using LDA topic modeling by industries: The case study of job planet reviews, *Journal of Information Technology Services*, 15(3), 157-171.

- Kim, K. K., Kim, G. Y., Kim, B. K. (2020). New nurses' experience of turnover: A qualitative meta-synthesis, *Journal of Korean Academy of Nursing Administration*, 26(2), 84-99. (in Korean)
- Koh, H. C., & Goh, C. T. (1995) An analysis of factors affecting turnover intentions of non-managerial clerical staff: A singapore study, *International Journal of Human Resource Management*, 6, 103-25
- Lee, J. H., Kim, M. J., Seung, H. N., Yoo, H. J. (2019). A consensual qualitative study on Korean employees' determinants of turnover, *Korean Journal of Psychology: General*, 38(2), 219-245. (in Korean)
- Lee, J. H., Shin, K. H., Baeck, S. G., Heo, C. G. (2014). The effectiveness of job embeddedness in turnover studies: A meta-analysis, *Korean Journal of Industrial and Organizational Psychology*, 27(4), 743-779. (in Korean)
- Lee, J., Kim, S., Kang, J. (2017). A study on job satisfaction/retention factors and job dissatisfaction/turnover factors by industries using job reviews, *Journal of Information Technology Services*, 16(1), 1-26. (in Korean)
- Lee, S. (2008). The role of job satisfaction and person-job fit in voluntary turnover process: An application of event history analysis, *Korean Journal of Industrial and Organizational Psychology*, 21(4), 705-728. (in Korean)
- Lee T. W. Mitchell T. R. (1994). An Alternative Approach: The Unfolding Model of Voluntary Employee Turnover, *Academy of Management Review*, 19(1), 51-89.
- Lee, Y. Kang, J. (2018). Related factors of turnover intention among Korean hospital nurses: A systematic review and meta-analysis, *Korean Journal of Adult Nursing*, 30(1), 1-17. (in Korean)
- McAfee, A. and E. Brynjolfsson(2012), "Big Data: The management revolution," *Harvard Business Review*, 90(10), 60-66
- Park, S., Chang, Y. C. (2015). The influence of work-life Balance(WLB) support

- system on employee engagement: Focused on moderating effect of perceived organizational support, *Management & Information Systems Review*, 34(5), 49-69. (in Korean)
- Rubenstein, A. L., Eberly, M. B., Lee T. W., Mitchell, T. R. (2018). Surveying the forest: A meta analysis, moderator investigation, and future oriented discussion of the antecedents of voluntary employee turnover, *Personnel psychology*, 71(1), 23-65.
- Schroeck, M., R. Shockley, J. Smart, D. RomeroMorales, & P. Tufano(2012), *Analytics: The Real-world Use of Big Data*, IBM Institute for Business Value
- Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer & Richard Harshman. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6), 391-407.
- Thomas Hofmann. (2001). Unsupervised learning by probabilistic latent semantic analysis. *Machine Learning*, 42(1-2), 177-196.
- The Ministry of Employment and Labor (2017). MOEL publishes health protection handbook for workers engaged in emotional labor.
- Zimmerman, R. D. (2008). Understanding the impact of personality traits on individuals' turnover decisions: A meta-analytic path model, *Personnel Psychology*, 61(2), 309-348