한국수어사전 표제어 피복 분석과 어휘 정비 후보 도출 -NCS 정보보호·개인정보보호 능력단위명을 대상으로-

한승재*, 박민수*, 김창훈**

*대구대학교 (학부생, **교신저자)

Headword Coverage Analysis of the Korean Sign Language Dictionary and Derivation of Lexical Development Candidates : For NCS Information Security and Personal Data Protection Competency Unit Names

Seung-Jae Han*, Min-Su Park*, Chang-Hoon Kim**

*Daegu University(Undergraduate Student, **Corresponding Author)

요 약

국가직무능력표준(NCS)의 정보보호·개인정보보호 능력단위명 128개를 대상으로 한국수어사전의 표제어 피복 실태를 측정하고 어휘 정비 후보를 도출하였다. 정보·통신 분야 표제어 573개와 대조한 뒤, 사전 전체를 재조회하고 뜻풀이를 검토하여 과소추정과 과대추정을 배제하였다. 표제어와 완전일치하는 능력단위명은 없었다. 고유 어절 144개 중 18개가 표제어로 확인되었고, 구성 어절이 전혀 확인되지 않은 능력단위는 74개였다. 세분류 16개 중 5개는 소속 능력단위 36개의 피복률이 모두 0이었다. 미확인 어절 126개 중 30개를 확보할 때 완전 피복되는 능력단위 수를 최대화하는 정수계획의 최적값은 51개였다. 본 결과는 전달 가능성의 판정이 아니라 현장 사용 조사와 어휘 정비 대상 선정을 위한 기초 자료이다.

I. 서론

정보보호 인력양성은 국가적 과제로 추진되고 있으나, 접근 가능한 형태의 강의를 제공받지 못하는 학습자는 교육 사각지대에 놓인다. 농인 학습자가 대표적인 사례이다. 한국수어는 한국어와 문법 체계가 다른 독립 언어이므로, 한국어 용어가 존재한다는 사실이 대응하는 수어 표현의 존재를 보장하지는 않는다.

정보보호 교육은 특정 직군에 한정되지 않는다. 개인정보 보호법 제28조제2항은 개인정보처리자가 개인정보취급자에게 정기적으로 필요한 교육을 실시하도록 규정한다[1]. 한국수화언어법 제2조제4항은 농인등이 한국수어로 교육받을 권리를 가진다고 규정하며, 같은 조 제3항은 농인과 한국수어사용자를 농인등으로 통칭한다[2]. 따라서 분석 대상의 범위는 연구자의 임의 선정이 아니라 이 두 법령이 규정한 교육 의무와 교육권에서 도출된다.

교육이 의무라면 그 교육에 필요한 어휘가 한국수어사전에 수록되어 있어야 한다. 선행 연구는 국제표준 KS X ISO/IEC 27000:2018 제3절의 표준용어 78개 가운데 한국수어사전 표제어로 확인된 것이 1개임을 보고하였다[3]. 그러나 국제표준은 국내 인력양성의 기준이 아니며, 국내에서 정보보호 직무의 범위를 규정하는 문서는 국가직무능력표준(이하 NCS)이다.

본 논문은 대조 목록을 국제표준 용어에서 NCS 능력단위명으로 전환하고, 측정 결과로부터 어휘 정비 후보 집합을 도출한다. 다만 조어를 제안하지 않으며, 특정 개념이 한국수어로 전달 가능한지 여부도 판정하지 않는다. 본 논문의 측정 대상은 한국수어사전 표제어가 능력단위명을 포괄하는 정도이다.

II. 관련연구

전문 분야 수어의 부재는 2000년에 이미 지적되었으며[4], 15개 기관을 대상으로 한 조사에

서는 정보화 교육 용어의 수어 표기 부족이 교육자의 가장 큰 애로 요인으로 나타나 용어사전 제작이 제언되었다[5]. 그러나 두 연구 모두 결손의 규모를 산출하지는 않았다. 사전에 등재된 어휘를 외부 자료와 전수 대조한 연구도 있으나[6], 이미 등재된 외래어를 대상으로 하였으므로 등재 여부 자체는 측정하지 않았다. 등재 여부를 측정한 선행 연구[3]는 국제표준 용어의 등재율을 산출하였으나, 어떤 어휘를 우선 조사하고 정비할 것인지는 제시하지 않았다.

한편 농인의 참여 없이 만들어진 조어가 사전에 등재되었다가 정비 과정에서 배제된 사례가 보고되었다[7]. 어휘 정비의 방향은 농인 공동체가 결정할 사안이므로, 본 논문은 조어를 제안하지 않고 정비 작업의 대상 후보만을 제시한다. 본 논문은 선행 연구[3]와 동일한 한국수어사전 정보·통신 분야 분류를 대조 대상으로 하되, 대조 목록은 NCS 능력단위명, 판정 지표는 어절 단위 피복률, 산출물은 정비 후보 집합으로 달리한다. 사전 데이터는 선행 연구와 별개로 재수집하였으며, 두 시점의 수집 결과가 일치함을 확인하였다.

III. 측정방법

대조 목록은 NCS 대분류 20 정보통신, 중분류 01 정보기술의 하위 소분류 06 정보보호와 소분류 11 개인정보보호에 속한 능력단위 전체이다[8]. 정보보호는 세분류 13개에 능력단위 108개, 개인정보보호는 세분류 3개에 20개로, 합계 128개이다. 같은 중분류의 소분류 12 산업기술보호는 등록된 능력단위가 없다. 목록은 NCS 포털의 분류별 내보내기 기능으로 수집하였다. 정보보호 직무의 범위는 연구자가 임의로 정하지 않고 국가 표준인 NCS의 분류를 그대로 따랐다.

대조 대상은 한국수어사전 정보·통신 분야 분류 전체이다[9]. 검색어를 입력하면 그 선택에 따라 수집 범위가 달라지므로, 검색어 없이 목록 페이지를 1부터 58까지 차례로 요청하여 574건을 전수 수집하였다. 표제어 기준 고유 항목은 573개이다(수집 시점 2026년 9월 8일).

정규화에는 선행 연구와 같은 방식으로 유니코드 NFC 통일, 공백·가운뎃점·붙임표 제거, 소문자화를 적용하였다. 정규화한 표기가 서로 같으면 해당 표기가 표제어로 확인된 것으로 처리하였다.

능력단위명의 어절 수는 한 어절 4개, 두 어절 41개, 세 어절 70개, 네 어절 12개, 다섯 어절 1개이다. 128개 중 124개가 두 어절 이상이므로, 명칭 전체를 표제어와 완전일치로 대조하면 일치 여부의 2개 값만 얻을 수 있다. 따라서 능력단위명을 어절의 열로 간주하고, 표제어로 포괄되는 어절의 비율을 피복률(coverage)로 정의하였다. 능력단위명의 연속하는 부분열을 공백 없이 연결한 표기가 사전 표제어와 일치하면 그 구간을 확보된 것으로 처리하였다. 왼쪽에서 오른쪽으로 가장 긴 확보 구간을 우선 선택하며 명칭 전체를 순차적으로 탐색하고, 확보된 어절 수를 전체 어절 수로 나눈 값을 피복률로 산출하였다. 이 분할은 동적계획법으로 구한 최적 분할과 동일한 결과를 산출하였다.

대조 범위를 정보·통신 분야 분류로 한정하면 같은 표기가 다른 분류에 등재된 경우를 포착하지 못한다. 이 과소추정을 배제하기 위하여 사전의 대분류 3개 전체를 대상으로 능력단위명 128개, 고유 어절 144개, 두 어절 이상이면서 능력단위명보다 짧은 연속 부분열 174개를 모두 재조회하였다. 그 결과 능력단위명 128개와 부분열 174개는 다른 분류에서도 확인되지 않았고, 고유 어절 13개만 추가로 확인되었다. 이 13개는 일상생활 수어 6개와 전문용어 수어 7개에서 확인되었으며, 정보·통신 분야에서는 확인되지 않았다.

분류 범위를 확대하여 조회하면 반대 방향의 오류가 발생한다. 표기는 같지만 개념이 다른 표제어까지 확인된 것으로 인정하면 등재가 과대추정된다. 따라서 추가로 확인된 13개 표기의 뜻풀이를 검토하였다. 뜻풀이가 특정 분야의 개념에 한정되고 그 분야가 정보보호가 아니면 제외하였다. 뜻풀이가 분야에 한정되지 않거나, 일상생활 수어 분류의 표제어여서 뜻풀이가 없으면 인정하였다. 제외한 5개는 관제(灌祭, 기독

교), 조사(助詞, 국어 교과), 체계(국어 교과), 안전조치(핵물질관리, 정치), 진단(의학)으로, 모두 표기만 같고 뜻이 다른 용어이다. 인정한 8개는 뜻풀이가 분야에 한정되지 않는 시험(기독교)과 점검(의학), 그리고 뜻풀이 없이 등재된 일상생활 수어 계획·연구·운영·정책·환경·설계이다. '시험'과 '점검'은 각 분야의 뜻풀이가 해당 분야에 한정되지 않아 인정하였고, 제외한 '체계'(국어 교과)와 '진단'(의학)은 뜻풀이가 특정 분야의 개념에 한정되어 제외하였다. 설계는 불교 분야(設戒)에도 있으나 일상생활 수어에 별도로 등재되어 있어 인정하였다.

이에 따라 결과를 2개 기준으로 함께 제시한다. 기준 A는 정보·통신 분류의 표제어만 인정한다. 기준 B는 사전 전체에서 표기가 확인되고 뜻풀이 검토를 통과한 표제어까지 인정한다. 본문의 수치는 결손을 과대추정하지 않는 기준 B를 따른다.

IV. 결과

대조 결과는 표 1과 같다.

표 1. NCS 능력단위 128개의 표제어 피복 실태

구분	기준 A	기준 B
능력단위명 완전일치	0 / 128	0 / 128
고유 어절 확인	10 / 144	18 / 144
어절 피복률	12.9%	21.2%
피복률 0.00	88	74
피복률 0.01~0.33	32	31
피복률 0.34~0.50	3	8
피복률 0.51~0.99	5	13
피복률 1.00	0	2

기준 A에서 기준 B로 확장하면 확인된 어절은 10개에서 18개로 늘지만, 명칭 완전일치는 두 기준 모두 0이고 피복률 1.00도 2개에 불과하다. 이는 추가로 확인된 어절이 모두 다른 분야에 등재된 일반어이기 때문이다.

기준 B에서 구성 어절이 모두 확인된 능력단위는 '네트워크 보안 운영'과 '시스템 보안 운영' 2개이다. 두 명칭은 '보안 운영'을 공유하고 대상어만 다르므로, 완전 피복은 새로운 어휘가 확보된 결과가 아니라 기존 어휘 조합이 반복된 결과이다.

피복률 평균이 상대적으로 높은 세분류는 정보보호관리·운영(0.468), 모빌리티보안(0.417), 클라우드 보안 관리·운영(0.315)이었다. 반면 정보보호진단·분석, 보안사고분석대응, 디지털포렌식, 개인정보가명익명처리, 개인정보인증평가의 5개 세분류는 소속 능력단위 36개 모두 피복률이 0이었다. 이들 세분류에는 모의해킹, 보안감사, 위협탐지, 증거 수집, 가명·익명처리, 개인정보 인증심사 준비 등의 능력단위가 속한다.

기준 B에서 미확인 어절은 126개이다. 이 가운데 어떤 어절을 우선 조사하고 정비할 것인지를 다음과 같이 정식화하였다. 능력단위 u의 미확인 어절 집합을 S(u)라 할 때, 어절 k개를 선택하여 S(u)가 선택 집합에 모두 포함되는 능력단위의 수를 최대화한다. 이는 예산 제약 하에서 원소집합이 모두 선택되어야 계수되는 최대 피복 문제의 변형이다. 능력단위 집합 U(|U|=128)와 미확인 어절 집합 W(|W|=126)에 대해 능력단위 u의 미확인 어절 집합을 S(u), 어절 w의 출현 횟수를 n_w, 어절 확보 여부를 x_w∈{0,1}, 완전 피복 여부를 y_u∈{0,1}로 둔다. 1단계는 제약 Σx_w=k 및 y_u≤x_w(∀u, ∀w∈S(u)) 아래 Σy_u를 최대화하고 그 최적값을 z*라 한다. 2단계는 동일 제약과 Σy_u≥z* 아래 확보되는 어절 출현 횟수 Σ n_w·x_w를 최대화한다. 이때 어절 피복률은 고유 어절 수가 아니라 능력단위명 전체의 어절 출현 349개 중 확보된 출현의 비율이며, 이미 구성 어절이 모두 확인된 2개 능력단위는 S(u)=∅이므로 제약 없이 y_u=1로 계수된다. 해는 상기 정수계획의 최적해이며 동일 환경에서 재현됨을 확인하였다. 결과는 표 2와 같다.

표 2. 어절 k개 확보 시의 최적 피복

k	완전 피복 능력단위	어절 피복률
0 (현재)	2 / 128	21.2%
10	21 / 128	35.8%
20	38 / 128	47.6%
30	51 / 128	59.0%
50	71 / 128	73.1%
80	100 / 128	86.0%
126 (전체)	128 / 128	100%

k=30 최적 후보 집합에는 클라우드, 모빌리티, OT보안, SW공급망, 생체인식, 영상정보처

리와 같이 세분류가 대상으로 삼는 기술 명칭이 포함된다. 한편 관리, 분석, 기획, 수립, 구현은 여러 세분류에 반복하여 출현하는 어절이다. 미확인 어절 126개 중 75개는 능력단위 1개에만 출현하고 51개는 2개 이상에 출현하며, 관리는 10개, 기획·클라우드·개인정보는 각각 9개, 생체인식은 8개 능력단위에 출현한다. 소수의 어절이 다수의 능력단위에 공통으로 출현하므로 k가 작을 때 피복률의 상승 폭이 크다. 본 데이터에서는 각 k의 최적해가 더 작은 k의 최적해를 포함하였으나, 이 성질은 일반적으로 보장되지 않으므로 표 2는 순위가 아니라 각 예산에서의 후보 집합으로 해석하여야 한다.

피복률 0인 5개 세분류의 고유 어절 52개 중 39개는 해당 세분류에만 출현하여, 다른 세분류를 모두 정비해도 이들의 어절 피복률은 28.6%에 그친다. 따라서 이들은 독립적인 정비 대상으로 다루어야 한다.

V. 결론

본 논문은 NCS 정보보호·개인정보보호 능력단위명 128개에 대한 한국수어사전의 표제어 피복률을 측정하고 어휘 정비 후보를 도출하였다. 표제어와 완전일치하는 능력단위명은 없었고, 능력단위 74개는 구성 어절이 전혀 확인되지 않았으며, 피복률 0은 정보보호진단·분석, 보안사고분석대응, 디지털포렌식, 개인정보가명익명처리, 개인정보인증평가의 5개 세분류에 집중되어 있었다.

사전에 표제어가 없다는 사실이 해당 개념을 한국수어로 전달할 수 없음을 의미하지는 않는다. 본 결과는 표준화된 사전 표제어에 기반한 어휘 지원이 제한적임을 나타내며, 이러한 제한은 교재·자막·통역 자료와 같이 표준 표기에 의존하는 자료에서 문제가 된다. 현장에서 이미 사용되는 표현의 존재 여부를 조사하고, 없을 경우 농인 공동체의 참여를 통해 조어하며, 등재 전에 농인 이용자의 검증을 거치는 순서가 필요하다. 본 논문의 후보 집합은 첫 단계인 조사의 대상을 좁히는 데 활용될 수 있다.

본 논문의 한계는 다음과 같다. 첫째, 분석 단위가 한국어 어절이므로 한국수어의 언어 단위와 일대일로 대응하지 않는다. 예컨대 k=30 후보 집합에 '영상정보', '영상정보처리', '영상정보처리시스템'이 각각 별도 어절로 포함되는 것이 이러한 불일치를 보여준다. 둘째, 뜻풀이 검토에는 연구자의 해석이 개입한다. 셋째, 후보 집합은 능력단위명의 피복 효율만을 기준으로 하며, 사용 빈도와 개념의 중요도, 농인 학습자의 요구는 반영하지 않는다. 후속 과제는 후보 집합에 대한 현장 사용 조사와 능력단위 요소 층위로의 확장이다.

Acknowledgements

본 논문은 과학기술정보통신부 및 정보통신기획평가원의 'SW중심대학사업'의 지원을 받아 제작되었습니다. (2025-0-00077)

[참고문헌]

[1] 개인정보 보호법, 법률 제20897호, 2025. 4. 1. 일부개정, 2025. 10. 2. 시행.

[2] 한국수화언어법, 법률 제21096호, 2025. 11. 11. 일부개정, 2026. 5. 12. 시행.

[3] 한승재, 김창훈, "정보보호 표준 용어의 한국수어 등재 실태 - KS X ISO/IEC 27000:2018을 기준으로," 한국정보보호학회 충청지부 2026년 정보보호학술논문발표대회, 2026.09.22. pp. 62-64

[4] 김상화, "바른 지식 전달을 위한 전문 수어(手語) 개발의 필요성," 한국수화학회 회보, 제5호, pp. 3-14, 2000.

[5] 강윤주, 원성옥, 윤점룡, 나운환, "청각·언어장애인 정보화교육의 실태와 문제점 분석," 언어치료연구, 제12권 제1호, pp. 217-228, 2003.

[6] 장진석, "영어에서 비롯된 외래어에 대한 한국수화 표현과 미국수화 표현과의 동일성 실태 조사," 특수교육저널: 이론과 실천, 제11권 제4호, pp. 439-463, 2010.

[7] 이현화, 홍성은, "<한국수어-한국어 사전>의 표제어 선정," 특수교육저널: 이론과 실천, 제25권 제2호, pp. 1-20, 2024.

[8] 한국산업인력공단, 국가직무능력표준 20. 정보통신 > 01. 정보기술 > 06. 정보보호, 11. 개인정보보호, https://www.ncs.go.kr (접속일: 2026-09-08).

[9] 국립국어원, 한국수어사전, https://sldict.korean.go.kr (접속일: 2026-09-08). CC BY-NC-ND 2.0 KR