

평가 척도를 이용한 사법통역 평가 사례연구

이지은 이유진 최효은
(이화여대 한국외대 이화여대)

1. 서론

양질의 사법통역 제공은 외국인이 언어 장벽으로 인한 소통의 불편 없이 한국인과 동등한 입장에서 재판을 포함한 사법절차에 참여할 수 있도록 하는 중요한 문제이다. 2018년 사법연감에 의하면 2017년 1심법원에 접수된 외국인 형사사건은 4,469건으로 2012년 대비 38%나 증가하였다. 이와 같이 사법통역 수요가 증가하는 가운데 사법통역 능력과 통역인의 자질을 검증할 수 있는 장치가 마땅히 없었고 체계적인 교육 기회마저 제한적이었다(이지은 2012, 2013). 그간 국내 법원의 통번역인 후보자 선정은 통번역인 후보자들의 이력서에 나타난 경력과 유학 및 거주 경험 등을 토대로 이루어졌고, 통역인의 핵심역량인 통역능력에 대한 검증이 부재했다. 이 때문에 통역 품질을 담보할 제도적 장치가 미비했고 그 결과 사법통역 품질 문제가 빈번히 제기되었다. 이러한 상황에서 지난 해 수원지방법원이 법정통역인 인증제도를 시범실시한 것은 사법통역 인증제 도입의 청신호다.

사법절차에서 제공되는 통역의 정확성 및 중립성 확보를 위해서는 통역인 후보자에 대한 철저한 능력검정을 거쳐서 통역인으로서 적절한 사법통역능력 수준과 윤리 자질 등을 겸비한 후보자를 선발하여 인증하고, 통역인들에 대한 내실 있는 교육을 통한 효율적인 관리 방안이 필요하다(이지은 장원경 2016). 또한 통역 품질 모니터링을 통해 사법통역이 적절하게 이루어지고 있는지 상시 관리 감독하는 사법통역 운영체계가 서비스 품질 관리 차원에서 필요하다(이지은 2012).

그간 국내외 사법통역 연구는 통역인들의 규범 이탈 현상을 포함하여 통역인이 사법절차 및 커뮤니케이션에 미치는 영향을 다양한 관점에서 조명하고 통역품질의 중요성과 통역인 전문교육의 필요성을 강조했다. 하지만 세계적으로 사법통역 평가에 초점을 맞춘 선행연구는 극히 적다. 미국 사법통역사시험에 대한 두 편의 연구논문(Clifford 2005; Stansfield and Hewitt 2005)과 국내 법원 등록 통번역인 후보자의 순차통역 평가를 다룬 이지은과 장원경(2015) 정도다. 이지은과 장원경(2015)은 국내 법원에 등록된 영어와 중국어 통역인 13명의 순차통역을 3점 척도를 이용하여 문장별로 정확성을 평가하였을 때 개인 편차가 있었지만 전반적으로 정확성이 부족한 것을 확인했다.¹⁾ 이 선행연구가 시험과 같은 환경에서 순차통역을 녹음하고, 이를 평가하였다면 본 연구는 실제 공판 기일에 형사법정에서 피고인신문과 증인신문 동안 제공된 대화통역의 품질 평가를 다룬다. 또한 그간 사법통역 연구가 주로 담화분석연구에 기초한 정성평가 성격이 강했다면 본 연구는 사법통역 모니터링 및 인증시험 평가 목적에 활용할 수 있는 평가도구를 개발, 적용하는 평가에 중점을 둔 연구이며, 여러 언어를 복수의 평가자가 동일한 방식으로 평가하기 위해 개발한 평가 척도를 사용하여 평가를 수행했다는 점이 특기할 만하다.

2. 통역 평가와 평가 척도

본 연구에서 다룰 통역 평가는 총괄평가이자 분석적 평가에 해당한다. 학습

1) 판결 선고 순차통역의 정확성은 영어 통역인들의 경우 평균 73.3점(최하 50.1, 최고 79.6), 중국어 통역인들의 경우에는 평균 67점(최하 41.9, 최고 85.3)이었다.

과정을 지지해주는 형성평가와 달리 총괄평가는 학업성취도에 관한 정보를 주거나 교육과정 이수 자격이나 전문자격인증의 결정과 관련된다(Hatim and Mason 1997: 199). 총체적 평가는 개별 품질 요소를 별도로 분석 평가하지 않고 전체적인 수행능력 평가를 하는 것으로 한 개 점수만 도출되는 데에 비해, 분석적 평가는 품질을 구성하는 개별 요소별로 평가하고 각각의 점수를 도출하여 총점을 낼 수 있다(Paloma and Banta 1999: 126).

평가시험을 설계하고 개발할 때 평가에서 나타나는 중요한 요소를 어떻게 평가 및 채점할 것인지 고려해야 하며 이를 위해 평가 루브릭(평가자들이 공유하는 가이드라인과 평정 척도)을 사용하는 것이 일반적이다. 통역 평가의 경우 평가 대상인 통역능력의 구성 요소를 고려한 평가 기준과 척도를 개발하여 사용해야 하는데 이러한 평가 기준에 따른 평가가 규준지향평가(criterion-referenced assessment)이다. 규준은 평가의 타당성에도 직결되며 본 연구도 규준지향평가를 지향한다.

어떤 평가이든 타당성과 신뢰성을 갖추는 것이 필수적이다. 타당성은 평가 점수가 우리가 측정하고자 하는 능력의 지표와 일치하는 정도(Bachman and Palmer 1996: 21)이다. 신뢰성은 평가의 일관성에 관한 문제로 타당성의 필요조건이기도 하다(Moore and Young 1997). 언제, 누가 평가하든지 관계없이 평가의 일관성을 확보하기 위해 동일한 평가 기준과 척도를 사용하는 것이 매우 중요하며, 평가의 신뢰성을 위해서는 평가자 훈련과 사전 조율도 필수적이다.

통역 평가 기준과 평가 척도를 다룬 선행연구를 간단히 살펴보자면 총괄평가를 다룬 연구는 많지 않다. 동시통역 품질 평가 기준으로 청해력/전달, 언어 품질, 논리와 충실성을 제시한 슐다거(Schjoldager 1996)와 누락, 추가, 손동작 및 자세까지 포함한 17가지 순차통역 평가용 세부기준을 제시한 리카도(Riccardo 2002)는 모두 형성평가를 위한 통역 평가 기준을 다루었다. 연구자마다 통역 품질 기준의 분류 방법과 명칭의 차이는 있으나 핵심적인 통역 평가 기준은 정확성, 목표언어 품질, 전달력으로 압축된다. 리(Lee 2008)는 이 세 가지 품질 구성요소를 각 수준별 통역 품질로 구분하고 순차통역 총괄평가용 분석적 평가 척도를 개발했다. 이 세 가지는 뢰히하커(Pöschhacker 2001)의 통역 품질 평가 모델을 고려한 것으로 칼리나(Kalina 2002)가 제시한 내용, 언어적 퍼포먼스, 프레젠테이션의 세 가지 주요 통역 품질 기준과도 일치한다. 정확성

은 내용의 정확성으로 원문에 대한 충실성을 평가하여 누락, 추가, 왜곡 등이 없이 의도된 효과를 잘 전달하는 것이고, 목표언어 품질은 목표텍스트로서 적절함을 가늠하는 언어담화적 측면이며, 전달력은 통역으로서 발화 속도나 전달력, 음성 크기 등 프레젠테이션 요소와 관계된다(Lee 2008: 169-170). 이 세 가지 거시적 평가 기준은 영국 공공서비스 통역평가 척도(Giambruno 2014)에서 그대로 채택된다.²⁾

전문통역 교육기관에서의 통역 평가 기준을 다룬 국내 연구(편혜원 2004, 김혜림 2011, 정철자 2012)는 구체적인 평가 방식을 다루지는 않았으나 평가방식에 대한 교수자 및 평가자로서의 고민을 다루었고, 평가 기준에 대한 동일한 이해와 적용이 필요하다는 점과 복수의 평가자간 사전 조율과 훈련이 평가의 적정성을 구현하는 데에 중요함을 시사했다.

사법통역은 이중언어 능력과 통역능력 이외에도 통역과정에 대한 이해와 해당 절차에서의 담화전략 및 통역인으로서 직무윤리 및 역할에 대한 이해를 필요로 한다. 헤일(Hale 2019)은 사법통역능력을 1)언어 및 담화, 2)맥락, 3)통역, 4)상호작용 등 네 가지 구성 요소 측면에서 설명한다. 언어 및 담화는 이중언어능력과 함께 법률담화를 다룰 수 있는 능력을 의미하며, 통역능력은 이론에 대한 이해와 메타언어능력, 통역방식 다루는 기술을 포함한 전문성을 가리킨다(Hale 2019: 54). 맥락능력은 통역사가 특정 업무를 위해 준비하는 능력이며, 상호작용능력은 참가자 간의 상호작용을 잘 관리할 수 있는 능력을 뜻한다(Hale 2019: 54). 다만 헤일(2019)이 말한 맥락능력이 총괄평가에서 별도 요소로 평가하기 쉽지 않고 불필요하다고 판단하여 본 연구에서는 기존 연구에서 핵심적인 통역 평가 기준으로 드러난 정확성, 목표언어능력, 전달력을 토대로 실용적인 평가 도구를 도출하기로 했다.

최근 들어 통역평가 시험에서는 오류 감점 방식의 채점보다 평가 척도를 이용한 평가방식이 신뢰할 수 있는 평가 루브릭으로 확산되고 있다(Han 2018). 구어의 특성상 통역평가에서는 즉각적으로 평가 기준 적용이 이루어지고 번역평가와 같은 치밀한 분석 시간은 주어지지 않기 때문에 실질적인 이유에 의해

2) 리(2008)과 지암브루노(Giambruno, 2014) 모두 호주와 영국에서 사법통역을 포함한 전반적인 통역 평가에 활용 가능한 기준으로 개발한 것이므로 본 연구에서도 이 기준을 채택했다.

서도 평가 척도 사용이 실용적이고 유용하다(Lee 2008: 170). 통역 비전문가도 평가 척도 사용에 무리가 없었고, 통역 전문가 평가자들과 유사한 평가 결과를 보인 티셀리우스(Tiselius 2009)의 연구결과도 이를 뒷받침한다.

여기에서는 본 연구에서 참고한 리(2008)의 분석적 평가 척도와 이와 동일한 평가기준을 채택한 영국의 공공서비스 통역사 인증시험, 그리고 최근 공개된 호주의 인증통역사시험의 평가 척도를 간략히 소개한다. 리(2008)의 평가 척도는 정확성과 목표언어 품질은 6등급(시험을 미완료한 최하등급 0, 최우수 등급 6), 전달력은 0등급에서 3등급까지 판정하기 위한 품질 기준을 제시하고, 세 가지 평가기준별 배점에 있어서는 통역 교수자들의 의견을 반영하여 전달력의 배점을 상대적으로 낮게 하여 각기 비중을 40%:40%:20%로 제시했다(Lee 2008: 180). 지암브루노(Giambruno 2014: 207)가 소개한 영국 공공서비스 통역 시험 평가 기준표 역시 정확성, 언어사용, 전달력이라는 세 가지 기준에 맞추어 배점 차이를 두지 않고 각 기준별 구간을 동일하게 4등급으로 두고, 3점씩 점수 차이를 두었다. 아래 <표 1>에 정확성 등급 기준을 제시하였는데, 이와 같은 방식으로 전달력과 언어사용도 각기 등급별 세부 기준이 있다.

<표 1> 영국 공공서비스 통역 인증시험 평가 척도 일부

평가 기준	등급(밴드)	점수	세부기준
정확성	A 매우 우수	10~12	<ul style="list-style-type: none"> 원 메시지의 의미를 완전하고 정확하게 전달. 모든 정보를 가감 왜곡 없이 전달. 구어 표현을 완벽히 전달할 수 있으며 해당 주제에 익숙함.
	B 우수	7~9	<ul style="list-style-type: none"> 원 메시지를 정확하게 전달. 한두 개 정도의 사소한 누락/왜곡 있으나 정보의 정확한 전달 및 원 메시지의 완벽한 이해에는 영향을 미치지 않음.
	C 보통(합격)	4~6	<ul style="list-style-type: none"> 원 메시지의 의미 적절하게 전달. 원 메시지의 이해 또는 정보 전달에 영향을 미칠 만큼 심각한 오역, 누락 또는 왜곡 없음.
	D 불합격	1~3	<ul style="list-style-type: none"> 원 메시지의 의미를 전달하지 못하거나 일부만 전달. 원 메시지의 이해 또는 정보 전달에 영향을 미칠 만큼 심각한 오역, 누락, 왜곡.

호주의 통번역인증제도에서도 최근 감점 방식 대신 평가 척도를 이용한 평가방식이 채택 되었다. 인증통역사 인증시험의 순차통역능력 평가 척도에는 크게 두 가지 능력, 즉 전달능력과 언어능력이 있다.³⁾ 전달능력은 의미전달능력, 통역모드 적용능력, 상호작용관리능력, 수사능력 등 네 가지의 하위 구성요소를 중심으로, 언어능력은 의미전달에 필요한 언어능력을 언어방향별 두 가지 구성요소로 구분하여 얼마나 적절하고 유창하게 말하는지, 화용, 어휘, 문법, 통사, 스타일, 어역 등이 어떤 수준인지 평가하는 것이다. 평가 척도는 각 평가 대상이 되는 능력별로 최상등급에서 최하등급까지 1~5로 구분하고 세부 기준에 대한 등급별 설명을 제시한다. 전달능력과 언어능력 대부분의 세부 항목에서 2등급 수준에 도달해야 인증통역사시험 합격이 가능하다. 다만 통역방식 적용과 상호작용관리 능력 둘 중 하나가 3등급 이상이면 인증이 가능하다.

이상 사법통역을 포함하는 통역 평가 관련 문헌연구를 통해 연구동향을 파악하고 여러 언어의 평가자가 참여하며 이 중에는 통역전문가가 아닌 평가자도 포함되는 소수언어의 현실적 필요를 고려하여 어느 정도 직관적으로 사용이 가능하다고 생각되는 10점 척도를 활용한 아래와 같은 평가 척도를 구성했다(<표 2> 참조).

<표 2> 사법통역 평가 척도

평가 기준	등급(점수)	세부기준
정확성	최상(10)	· 원 메시지의 의미를 완전히 정확하게 전달함. · 모든 정보를 누락, 왜곡, 추가 없이 전달함.
	상(8~9)	· 원 메시지를 정확하게 전달함. · 몇 개의 사소한 누락, 왜곡, 추가가 있으나 정보의 정확한 전달 및 원 메시지의 완벽한 이해에는 영향을 미치지 않음.
	중(6~7)	· 원 메시지의 의미를 전반적으로 전달함. · 몇 군데 누락, 왜곡, 추가 등으로 인해 의미와 효과에 영향이 있음.
	하(5 이하)	· 원 메시지의 의미를 제대로 전달하지 못함. · 원 메시지의 이해 및 정보의 전달에 영향을 미칠 만큼

3) <https://www.naati.com.au/certification/certification-testing/>

		<p>심각한 누락, 왜곡, 추가가 있음.</p> <ul style="list-style-type: none"> · 문법, 문장구조, 어휘, 전문용어 구사가 완벽함. · 발음과 억양이 분명하고 정확함.
목표언어	최상(10)	<ul style="list-style-type: none"> · 문법, 문장구조, 어휘, 전문용어를 적절하게 구사함. · 몇 개 정도의 부적절한 언어표현을 구사하나 전체적인 정보의 전달에는 문제가 없음. · 대체로 발음과 억양이 분명하고 정확하지만 발음과 억양이 일부 내용을 이해하는 데 미미한 영향을 미치는 정도임.
	상(8~9)	<ul style="list-style-type: none"> · 문법, 문장구조, 어휘, 전문용어를 적절하게 구사함. · 종종 부적절한 언어표현을 구사하나 전체적인 정보의 전달에는 문제가 없음. · 종종 발음이나 강세가 틀리나 메시지의 전달에는 문제가 없음.
	중(6~7)	<ul style="list-style-type: none"> · 문법, 문장구조, 어휘, 전문용어를 부적절하게 구사함. · 발음이 안 좋고 억양, 어조, 강세가 지나치게 강해서 메시지의 내용을 이해하기 어려움.
	하(5 이하)	<ul style="list-style-type: none"> · 적절한 속도로 통역이 분명하고 부드럽게 이루어짐. · 상황에 맞는 어조와 말투를 사용함.
전달력	최상(10)	<ul style="list-style-type: none"> · 적절한 속도로 통역이 대부분 분명하고 부드럽게 이루어짐. · 상황에 맞는 어조와 말투를 사용함. · 약간의 말실수가 있으나 다시 바로 말하거나 사소한 실수로 의사소통에 영향 미치지 않음.
	상(8~9)	<ul style="list-style-type: none"> · 약간 느리거나 일정하지 않은 속도로 통역이 이루어짐. · 말실수로 메시지 전달력이 다소 약화됨.
	중(6~7)	<ul style="list-style-type: none"> · 전반적으로 느린 속도로 통역이 이루어짐. · 자주 망설임이 있고 말실수가 반복되어 메시지 전달력이 부족함.
	하(5 이하)	

한 등급 내에 3점 이상의 점수 차이를 두면 평가점수가 달라질 수 있다는 점을 감안하여 완벽에 가까운 ‘최상’ 등급을 제외하고는 2점씩 차이가 있게 했다. 적어도 ‘상’ 등급 (8~9점)에 해당해야 적절한 사법통역이 가능하다는 전제로 이에 해당하는 정확성, 목표언어, 전달력의 설명을 적절히 제시하고, ‘중’ 이하 등급은 사법통역 품질에 문제가 있다는 점을 정확성, 목표언어, 전달력의 설명에 반영하여, 평가자들이 평가 기준 적용 과정에서 통역 품질과 능력 차이를 인식할 수 있도록 했다. 사법통역 능력을 판가름하는 인증시험에 이 척도를 활용한다면 세 가지 기준에서 모두 ‘상’ 등급을 충족해야 합격 가능하다. 이 평가

척도는 리(2008)의 세 가지 평가기준을 따른 지암브루노(2014)가 소개한 평가 척도와 같이 평가기준별 가중치를 두지 않고 1:1:1 배점을 하였는데, 이는 내용의 정확성 뿐 아니라 발화스타일을 고려한 충실한 전달과 언어구사의 중요성을 고려한 것으로, 사용이 용이한 단순한 평가 척도를 추구한 때문이다.

3. 사례연구

앞에서 설명한 평가 척도를 이용하여 법정신문에서 이루어진 통역 품질에 대한 평가를 수행한 연구방법과 연구결과를 차례로 소개한다.

3.1. 연구방법

연구자료를 수집한 지방법원은 통역 품질 실태조사 차원에서 최대한 많은 자료를 제공해주기로 계획했으나 공판녹음 상태가 좋지 못하여 평가 대상이 된 자료는 아래 <표 3>과 같이 최종적으로 확정되었다.⁴⁾ 총 14개의 녹음파일은 2017년부터 2018년 상반기 사이 있었던 형사재판 외국인 신문 통역에 해당하며, 총 10시간 34분 분량의 자료에는 7개 언어 통역인 11명의 통역이 담겨 있다. 사건 유형은 교통사고에서부터 살인까지 다양하다. 4번과 5번 사건의 중국어 통역인과 12, 13, 14번 사건의 태국어 통역인은 동일 통역인이다. 이하 재판 관계자와 관련된 정보는 익명화를 위해 모두 바꾸었다. 통역인들의 당시 법원 통역 경력을 햇수로 괄호 안에 표기하였다. 1번 영어 통역인 A가 17년으로 가장 경력이 많았으며, 1년 미만의 통역인도 2명이었다. 나머지 통역인 대부분은 경력 2년으로 확인되었다. 다만 태국어 통역인 K의 경력은 확인 불가하였다. 일어 통역인 C만 통번역대학원 졸업자이며, 해외 대학원 출신인 A를 제외한 나머지 통역인은 국내외에서 전문대 또는 4년제 대학을 졸업하였다.

4) 평가연구 목적으로 별도로 녹음을 한 것이 아니라 실제 공판 녹음자료를 연구용으로 제공받은 것이다. 1차적으로 법원이, 2차적으로 연구진이, 3차적으로 평가자가 평가 가능하다고 확인한 자료만 정리한 것이다. 일부 녹음파일 음질이 좋지 않은 경우 해당 부분을 배제하고 평가를 진행했다.

〈표 3〉 연구자료

번호	언어	통역인 (당시 법원통역 경력)	사건 유형	신문종류	녹음자료 길이
1	영어	A(17년)	마약	피고인신문	70분
2	영어	B(1년 미만)	성폭력	증인신문	76분
3	일본어	C(1년)	살인	피고인신문	130분
4	중국어	D(2년)	성폭력	증인신문	41분
5	중국어	D(3년)	마약	증인신문	28분
6	중국어	E(2년)	성매매알선	증인신문	51분
7	러시아어	F(2년)	교통사고	증인신문	31분
8	러시아어	G(1년 미만)	강제추행	증인신문	78분
9	몽골어	H(2년)	살인	증인신문	13분
10	우즈베크어	I(2년)	도주치상	피고인신문	52분
11	태국어	J(2년)	강간치상	증인신문	38분
12	태국어	K(확인불가)	특수절도	피고인신문	6분
13	태국어	K	특수절도	피고인신문	9분
14	태국어	K	특수절도	피고인신문	11분

평가자료는 석사과정에 해당하는 통역교육을 이수하고, 박사학위 과정에 있거나 교육 경험이 있는 통역 전문가를 우선적으로 섭외했고 언어별로 2명의 평가자를 두었다. 영어, 중국어, 러시아어, 일본어는 통역 전문가 중 적어도 한 명은 사법통역 실무 또는 연구 경험이 있다. 태국어와 몽골어 평가자들은 통역 전문교육을 받지 않았으나 대학에서 해당 언어를 가르치는 교수자들과, 국내 대학에 전공이 개설되어 있지 않는 우즈베크어의 경우 우수한 이중언어 구사자를 평가자로 섭외했다.

평가자들에게는 사법통역에서 요구되는 품질과 정확한 전달의 중요성을 설명하고 <표 2>의 평가 기준과 각 등급에 해당하는 오류 예를 제시하면서 아래 <표 5>의 채점표를 이용하여 평가해줄 것을 요청했다. 위 평가 척도대로 정확성, 목표언어, 전달력 점수를 평가하고, 종합점수를 제시하도록 했다. 평가를 위해 법정신문 녹음자료와 평가목적으로 편집한 증인신문조서 및 피고인 신문조서, 평가 기준 및 채점표를 평가자들에게 이메일로 제공한 후 열흘 내외의 충분한 시간을 갖고 각자 편리한 시간에 평가할 수 있도록 했다. 평가결과를 제출할

때에는 평가자들이 평가점수와 함께 제공한 한국어 신문조서에 주요 오류를 표시하고 간단한 설명을 제공하여 해당 언어를 알지 못하는 연구자가 평가 내용과 문제를 이해할 수 있도록 요청했다.⁵⁾ 12-14번 파일은 동일 사건의 공동피고인에 대한 신문으로 시간도 짧고, 동일 통역인의 통역이라 하나로 묶어 평가했다. 각 언어별 평가자들은 서로 평가 결과를 알지 못하였다.

〈표 5〉 채점표

구분	한→외 통역			외→한 통역		
	정확성	목표 언어	전달력	정확성	목표 언어	전달력
점수 (10점 기준)						
종합점수 (10점 기준)						

3.2 연구결과

녹음파일별로 언어별 각 2인의 평가자들이 제공한 평가점수는 아래 <표 6>과 같다. 전체 한국어→외국어(한외)통역의 종합점수 평균은 6.5, 외국어→한국어(외한)통역 종합점수 평균은 6.8로 언어 양방향 모두 평균 ‘중’ 수준을 보였다. 통역 사건이 다르고 통역 난이도 차이가 있어 통역인 간의 단순 비교가 용이하지 않다. 11명의 통역인 가운데 유일하게 전문통역 교육을 받은 것으로 알려진 3번 일본어 통역은 한외통역은 위스퍼링 방식으로 진행하여 평가가 불가능한 관계로 외한 순차통역 일방향만 평가하였는데 평가 대상 통역인 중 가장 수

5) 본 연구에 참여한 평가자들은 동일한 평가기준표를 사용하여 법정통역인 인증시험 평가를 한 유경험자인 관계로 본 연구를 위한 평가자 교육은 추가로 열지 않았다. 다만 본 연구를 위해 추가로 섭외된 태국어, 몽골어, 우즈베크어, 러시아어 평가자 4인에게는 면담, 전화통화, 이메일 등을 통해 평가 기준을 상세히 설명했다. 추가 섭외된 평가자들에게는 해당 언어를 모르는 연구진도 평가 근거를 알 수 있도록 평가지에 오류를 빠뜨림 없이 기록하도록 당부했다. 평가자가 제시한 오류에 비해 점수가 높다고 판단되는 경우에 연구진이 재검토를 요청했고, 이러한 조율 과정에서 일부 평가자들은 점수를 정정했다. 평가 결과를 통해 일부 평가자들이 실제 법정통역 녹음이라는 점을 의식하여 법원 혹은 통역인에 불이익이 갈 것을 염려하여 후하게 평가한 것과 평가자의 모국어에 따른 평가 점수 차이점이 확인되었다.

준이 높은 통역능력을 보여 ‘상’으로 평가받았으며, 5번 중국어 통역도 ‘상’에 해당하는 등급 평가를 받았다. 7번 러시아어 통역과 12~14번 태국어 통역은 ‘하’ 등급 평가를 받았다. 그 외 1번과 2번 영어 통역, 6번 중국어 통역, 8번 러시아어 통역, 9번 몽골어 통역과 10번 우즈베크어 통역, 그리고 11번 태국어 통역은 ‘중’에 해당하는 등급 평가를 받았다.

〈표 6〉 종합점수, 정확성, 목표언어, 전달력 평가 결과

번호	언어	통역인	종합점수		정확성		목표언어		전달력	
			한외	외한	한외	외한	한외	외한	한외	외한
1	영어	A	6	7	6	6	7	8	7	7
			6	7	6	7	6	8	7	7
2	영어	B	7	8	7	8	7	8	8	8
			7	7	6	7	7	8	8	8
3	일본어	C	·	9	·	9	·	10	·	10
			·	9	·	9	·	10	·	10
4	중국어	D	8	7	8	7	8	7	8	8
			8	7	9	7	8	7	8	8
5	중국어	D	8	8	9	8	8	8	8	7
			8	8	9	8	8	8	8	8
6	중국어	E	6	6	6	6	6	6	6	6
			6	6	6	6	6	6	6	7
7	러시아어	F	5	5	5	5	5	5	5	5
			5	5	5	5	5	6	5	5
8	러시아어	G	7	6	7	6	8	6	7	6
			7	6	7	6	8	6	7	6
9	몽골어	H	7	7	7	7	7	7	8	8
			7	7	7	7	8	7	7	6
10	우즈베크어	I	6	5	5	5	6	6	6	5
			6	5	5	5	7	6	6	5
11	태국어	J	7	7	6	7	7	8	7	7
			6	7	7	6	6	7	6	7
12-14	태국어	K	5	6	4	4	6	8	5	6
			4	6	3	4	3	5	6	7

본 연구에서 개발한 10점 평가 척도를 이용하여 평가하는 방식이 평가자들이 이용하기에 어려움이 없는 것으로 확인되었고, 평가결과도 대체로 신뢰할 만 한 수준이었다. 구체적으로 언어별 평가자간 신뢰도⁶⁾를 살펴보면, 전체 종합점수 중 두 평가자가 같은 등급의 점수를 부여한 비중을 계산한 등급별 퍼센트 일치도(percent agreement)에서 7개 언어 모두 75% 이상으로 신뢰할 만하다는 결과를 얻었다. 점수의 차이가 있어도 종합 평가에서 평가자들이 동일한 등급에 해당하는 평가 결과를 도출한 것은 통역 품질 수준 및 통역능력에 대한 의견일치를 가리키는 의미 있는 결과로 볼 수 있다. 다시 말해 인증시험 평가에서 당락에 대한 의견이 일치할 가능성이 높은 것이다. 종합점수 및 종합점수를 구성하는 정확성, 목표언어, 전달력 점수에 대한 구체적인 등급별 퍼센트 일치도는 아래 <표 7>과 같다.

<표 7> 종합점수, 정확성, 목표언어, 전달력 점수의 등급별 퍼센트 일치도(단위: %)

	종합점수			정확성			목표언어			전달력		
	한외	외한	합	한외	외한	합	한외	외한	합	한외	외한	합
영어	100	50	75	100	50	75	100	100	100	100	100	100
일본어		100	100		100	100		100	100		100	100
중국어	100	100	100	100	100	100	100	100	100	100	66.7	83.3
러시아어	100	100	100	100	100	100	100	50	75	100	100	100
몽골어	100	100	100	100	100	100	0	100	50	0	0	0
우즈베크어	100	100	100	100	100	100	100	100	100	100	100	100
태국어	100	100	100	100	100	100	50	0	25	50	100	75

<표 7>에서 확인할 수 있는 것처럼 종합점수로 본 등급 평가에 대한 평가자 일치도는 영어만 75%이고 모두 100% 일치하였다. 한편 종합점수를 구성하는 정확성, 목표언어, 전달력의 세부 점수 등급별 일치도는 다소 복잡한 양상을

6) 본 연구에서는 동일 자료를 평가한 평가자 숫자가 적어서 ICC(급내상관계수)를 구하는 대신 범주형 변수인 등급을 측정값으로 하여 두 명의 평가자가 같은 등급으로 평가한 자료의 백분율, 즉 퍼센트 일치도(percent agreement)로 평가자간 신뢰도를 분석했다. 퍼센트 일치도는 일반적으로 75% 이상일 때 신뢰 가능한 것으로 여겨진다(Lee 2008: 177; Graham et al. 2012: 9).

보였다. 정확성 평가결과를 살펴보면, 종합점수와 마찬가지로 영어를 제외하면 평가자 일치도가 100%였지만 목표언어와 전달력에서는 평가자들의 의견이 엇갈렸다. 목표언어에 대한 평가에서 영어, 일본어, 중국어, 러시아어, 우즈베크어는 75% 이상 높은 일치도를 보였으나 몽골어와 태국어는 50% 이하의 일치도를 보였다. 전달력에서도 다른 언어는 75% 이상으로 평가등급이 일치하였으나 몽골어의 경우 등급평가가 일치하지 않아 신뢰 수준인 75%를 밑돌았다.

작은 표본으로 단정하기는 어렵지만 이와 같은 언어별 평가자 일치도의 차이는 평가자의 자질 차이에 기인한다고 볼 수 있다. 전술한 바와 같이 영어, 일본어, 중국어, 러시아어의 평가자들은 전문 통역 교육을 이수한 자들로 양 언어의 구사력이나 평가 능력이 고루 우수한데 비해, 우즈베크어와 몽골어, 태국어의 평가자들은 전문적인 통역 교육을 받지 않은 평가자들이었다. 그중에서도 언어별 평가자들의 모국어가 서로 달랐던 몽골어와 태국어를 보면 평가자의 모국어를 따라 한국어와 해당 언어에 대한 평가 결과 차이가 있었다(<표 6> 11번 평가점수 참조). 반면 우즈베크어는 두 평가자의 모국어가 같았고, 세부점수는 다르지만 등급으로 본 평가자 일치도가 높았다. 이러한 평가 결과는 특히 전문 통역 교육과 평가 경험이 없는 평가자에 대한 훈련과 조율이 필수적임을 시사한다.

아래에서는 정확성, 목표언어, 전달력의 세 가지 평가기준별로 구체적인 분석 결과를 제시한다.

3.2.1 정확성

법정신문의 주요한 목적이 실체적 진실 확인임을 고려할 때 통역의 정확성은 가장 중요한 평가 기준이라 할 수 있다. 본 연구에서 평가한 14개 녹음파일의 한외 정확성 평균은 6.4, 외한 정확성 평균은 6.5로 양방향 모두 평균적으로 ‘중’ 수준에 해당한다. 각 녹음파일별로 살펴보면, 정확성 평가에서 높은 점수를 받은 3번(일본어)과 5번(중국어)을 제외하면 평가 대상 대부분이 메시지를 전반적으로 전달하는 수준으로 간주되는 ‘중’에 해당하는 평가 점수를 받았다.

언어별로 살펴보면, 1번과 2번(영어)은 모두 양방향에서 원 메시지의 의미를 전반적으로는 전달하나 부분적으로 누락 또는 왜곡으로 인해 의미와 효과에 영향이 있었다고 평가되어 평가자별 세부적인 점수 차이는 있으나 둘 다 ‘중’에 해당하는 점수를 받았다. 3번(일본어) 통역인 C는 전문 통역사로 한외통역은

동시통역과 유사한 위스퍼링 방식으로 진행했기 때문에 평가가 가능한 외한통역에 한해 평가가 실시됐다. 평가자가 지적한 오류는 모두 언어 치환 가운데 발생하게 되는 변환이거나 사소한 누락, 추가, 왜곡으로 ‘상’ 수준의 점수가 부여됐다.

4번과 5번(중국어)은 모두 통역인 D의 통역이다. 5번은 양방향 모두 정확성에서 ‘상’에 해당하는 점수를 받았으나 그보다 1년 전 통역 녹음파일인 4번은 한외통역은 ‘상’, 외한통역은 ‘중’ 수준의 점수가 부여됐다. 6번(중국어) 통역인 E는 개입을 많이 하며 녹음 후반부에서는 증인의 발화를 정리하여 3인칭으로 전달을 하는 등 원 메시지의 의미와 효과에 영향을 준 것이 ‘중’에 해당하는 평가를 받는 것으로 분석된다.

7번(러시아어)은 교통사고가 일어난 직후의 정황을 증언하는 증인신문임에도 불구하고 양방향 통역에서 모두 심각한 누락이 다수 발생하여 주요한 정황에 대한 정보의 전달에 영향을 미쳤으므로 양방향 모두 ‘하’ 수준의 점수를 받았다.

8번(러시아어)은 한외 방향에서 의미를 전반적으로 정확하게 전달하는 편이나 통역인이 지나치게 개입하며 원 메시지를 왜곡하여 원 메시지의 이해 및 정보의 전달에 영향을 미칠 만큼 심각한 왜곡이 몇 군데 있어서 ‘중’ 수준의 점수로 평가됐다. 외한 방향은 몇 군데 누락으로 인해 의미와 효과에 영향이 있다고 판단하여 ‘중’ 수준인 6점을 부여했다.

9번(몽골어)에 대해 평가자 1과 2는 대체로 적절하게 통역했으나 몇 군데 원 메시지의 의미와 효과의 정확한 전달에 영향을 미치는 누락과 왜곡이 있어 ‘중’으로 평가했다.

10번(우즈베크어)은 정확성에서 ‘하’에 해당하는 낮은 점수를 받았는데 피고인에게 유리한 정황과 관련된 정보를 왜곡하거나 많이 누락하여 전반적으로 정보가 정확하게 전달되지 않았다.

11번(태국어)은 의미와 효과에 영향을 미치는 왜곡이 많이 있었고 신문 과정에서 명확한 대답을 듣기 위해 계속해서 개입하여 보충 질문을 하거나 질문을 끊어서 통역했다. 평가자에 따라 언어방향별 점수는 차이가 있었으나 모두 ‘중’에 해당한다고 평가했다. 정확성에서 매우 낮은 점수를 받은 12-14번(태국어)에 대해 평가자 1과 2 모두 통역 전체적으로 원발화의 명제적 의미마저도

변형될 정도로 부정확한 통역이 대부분이라고 판단했다. 질문은 모두 왜곡되었고 증인의 답변은 통역인이 이해한 정보에 말을 덧붙이는 방식으로 통역됐다.

언어별 평가 결과 및 평가 의견을 종합해 보면, 언어에 상관없이 대체로 동일한 평가 대상에 대해 평가자들이 지적한 문제가 일치하는 경향을 보였다. 대체로 평가자들은 본 연구의 평가 척도에서 제시한 정확성의 기준을 정확히 인지하고 세부 기준에 명시된 누락, 왜곡, 추가의 문제를 적절하게 가려내었으며 그 결과를 바탕으로 적절한 등급의 점수를 부여한 것으로 분석되었다.

등급의 차이를 결정하는 오류의 경중을 보여주기 위해 지면관계상 몇 가지 오류만 예로 제시한다. 해당하는 오류 부분에 밑줄을 그어 표시했다. 먼저 몇 군데 누락, 왜곡, 추가로 인해 의미와 효과에 영향이 있는 ‘중’에 해당하는 오역 사례로 1번(영어)과 4번(중국어) 통역 일부를 제시하고, ‘하’에 해당하는 심각한 오역사례로 7번(러시아어)과 12번(태국어) 통역 일부를 제시한다.

<오류 사례 1>

번호인: 피고인은 2016. 11. 12. 클럽 Bee에서 G, 김성현 일행을 만났고, 그 다음날인 11. 13. 다시 평택부대 앞 커피숍에서 그들을 만났지요.

통역인: At the Club Bee on November 12, 2016, you have met the group of G and the next day on November 13, again you met them uh in front of the coffee shop in Pyeongtaek uh Base.

(피고인은 클럽 B에서 2016년 11월 12일 G 일행을 만났고, 그 다음 날인 11월 13일 다시 그들을 평택 어 부대의 커피숍 어 앞에서 만났지요.)

위의 <사례 1>은 왜곡에 해당하는 예로, 해당 파일을 평가한 영어 평가자들 다 원 메시지가 왜곡되어 원 메시지의 의미에 영향을 미친 경우로 판단한 사례다. 피고인과 G, 김성현 일행이 만난 장소로 ‘평택 부대 앞에 있는 커피숍’을 ‘평택 부대에 있는 커피숍 앞’으로 통역하여 원 메시지를 왜곡했다.

<오류 사례 2>

증인: 我不想去, 他非这样去, 我不去的话就生气。

(가고 싶지 않았는데 그가 그렇게 가야만 한다고 했고, 안 가면 화를 냈습니다.)

통역인: 저는 들어갈려 들어가지 않으려고 했습니다. 하지만 피고인은 저를 강제로 데리고 들어갔습니다.

<사례 2>는 피고인이 왜 두려워했는지에 대한 증인의 진술이 일부 누락되었음을 보여준다. 이 통역인은 여기에 제시하지 않지만 이어지는 진술에서도 가지 않으면 피고인이 화를 냈다는 정보를 빠뜨리는 동일한 오류를 범했다. 결과적으로 증인이 어쩔 수 없이 들어간 이유와 관련된 정보를 전달하지 못해 원 메시지의 의미 전달에 영향을 미쳤다.

<오류 사례 3>7)

증인: Я подошел, когда к Грандеру, Вот это, подъехала от бас терминала в сторону Осана поворачивал эвакуатор, и остановился. И оттуда вышел водитель. Ну (xx) которого ... И подошел сюда. И я поэтому место оставил, ну как бы думаю что, он вызовет (xx) скорую помощь, и пошел вторую машину обратно, к Галине.

(제가 이 그랜저로 갔을 때, 견인차가 버스터미널에서 오산방향으로 나와 멈춰 섰습니다. 그리고 거기서 운전자가 내렸습니다. 그 견인차 (xx)··· 그리고 이쪽으로 왔습니다. 그래서 저는 그가 (xx) 구급차를 부르겠지 라고 생각해 그곳을 떠나왔고, 두 번째 차량 갈리나에게로 갔습니다.)

(중략)

증인: Она не забрала. Просто проезжала мимо, и остановилась.

(그 차는 견인하지 않았습니니다. 그냥 옆을 지나가다 멈춰 섰습니다.)

통역인: 그 견인차가 이렇게 와서 요 옆에 서는 것을 서는 것을 봤습니다.

<사례 3>은 7번(러시아어) 교통사고 사건 증인신문 중 일부로 이 통역인은 전반적으로 누락이 많아 정확성이 낮다. 중략된 부분에서 통역인은 증언을 바로 통역하지 않고 원하는 대답을 듣기 위해 두 차례 질문을 한 후에 위와 같이 통역을 했다. 교통사고 직후의 상황을 설명하는 과정에서 상당히 많은 정보가 누락되고 증인의 발화에서 나타난 구체적인 정황이 전혀 드러나지 않은 것을 알 수 있다.

7) 발췌문의 (xx)는 잘 들리지 않는 부분을 표시한 것이다.

<오류 사례 4>

검사: 택시 뒷좌석이 넓은 것도 아니고 좁은데, 피고인 기준으로 왼쪽에 택시기사의 옷이 걸려있는 것을 보지 못했나요?

통역인: จริง ๆ แล้ว...คือ...ข้างในรถ...ก็...แคบอยู่... แล้วมองเห็นทุกอย่างไ
ด้ แต่...ทำไม...ไม่เห็น...เพื่อนสักเสื้อเอามา ?

(사실...어... 차 안은...어... 좁습니다... 그리고 모든 것을 볼 수 있습니다.
그런데... 어째서 ...친구가 옷을 훔치는 것을 보지 못했습니까?)

<사례 4>의 12번 태국어 통역인 K는 질문을 부분적으로 왜곡하는 정도가 아니라 아예 다른 질문으로 오역했다. 통역 기술 부족으로 원발화에 대한 이해와 기억에 문제가 있는데다 목표언어 표현력의 한계까지 더해져 다른 의미, 다른 형태의 질문을 하게 된 것으로 판단된다. 본문에는 자료를 제시하지 않았지만 피고인이 답한 ‘아닙니다.’를 통역 시 누락하거나 대답 서두에 ‘맞습니다.’를 추가하며 인위적으로 순조로운 신문 진행을 도모하려고 한 것도 평가자가 지적했다.

3.2.2 목표언어

목표언어의 문법, 문장구조, 어휘 또는 전문용어나 발음과 강세가 적절한지를 평가한 목표언어와 관련해서 14개 녹음파일의 한외 목표언어 평균은 6.7, 외한 목표언어 평균은 7.1로 양방향 모두 평균적으로 ‘중’ 수준에 해당한다.

한편 목표언어의 평가는 상당 부분 통역인의 모국어와 연관이 있는 경향을 보였다. 한외 방향과 외한 방향의 양방향에서 고루 높은 목표언어 점수를 받은 5번(중국어) 통역인, 양방향 모두 ‘중’ 수준의 점수를 받은 9번(몽골어) 통역인, 언어 간섭 문제가 있는 6번(중국어)과 사투리를 사용하는 10번(우즈베크어) 통역인을 제외한 나머지 통역인들은 모국어 방향으로의 통역에서 목표언어 점수가 외국어 방향으로의 통역보다 높은 경향을 보였다.

1번과 2번(영어) 통역인은 둘 다 한국어 모국어 화자다. 영어 평가자 1과 2는 1번의 한외통역에 대해 전체적인 내용 전달은 가능하나 종종 부적절한 언어 표현을 구사하며 일부 시제나 태 사용이 미숙한 문제가 있다고 평가했다. 두 평가자는 2번에 대해서도 전반적으로 메시지는 전달하나 어색한 언어표현과 시제

가 종종 나타난다고 지적했다. 한편 모국어인 한국어 방향으로의 통역에 대해 평가자들은 문법, 어휘, 발음, 억양 측면에서 완벽한 것은 아니나 특별히 지적할 만 한 점이 없다고 보았다.

모국어 방향으로의 통역만 평가 가능했던 3번(일본어)은 목표언어 문법, 문장구조, 어휘, 전문용어, 발음, 억양이 완벽하다고 평가 받았다.

4, 5번(중국어) 통역인과 6번(중국어) 통역인은 모두 중국조선어 억양이 있는 한국어를 구사하고 있지만, 두 통역인의 한외통역 목표언어 점수에는 차이가 있었고, 6번(중국어)은 문장구조가 문법에 맞지 않고 어휘가 적절하지 않다는 지적 사항이 한외통역에서 많았다.

한국어가 모국어인 7번(러시아어) 통역인 F의 통역에 대해 평가자 1은 러시아어 문법 실수, 어색한 표현이 많다고 지적했다, 평가자 2도 한외통역은 ‘하’ 수준의 점수를 부여했다. 외한통역은 평가자 2인의 의견이 엇갈려 평가자 1은 ‘하’, 평가자 2는 ‘중’ 수준의 점수로 평가했다.

8번(러시아어) 통역인 G는 러시아어가 모국어다. 러시아어 평가자들은 공통적으로 통역인 G의 외한통역에서 전체적인 정보의 전달에는 문제가 없으나 발음이 틀리거나 언어표현이 적절하지 않은 부분이 군데군데 나타난다고 지적했다.

10번(우즈베크어)에 대해 평가자 1은 표준어가 아닌 사투리를 사용해 그 지역의 사람들이 알아듣지 못하는 단어들을 사용하는 경우가 종종 있다고 지적했다, 평가자 2도 같은 이유로 한외통역에서 ‘중’ 수준의 점수를 부여했다.

11번(태국어)에 대해서는 태국어 발음과 성조가 부정확하고 화자의 어감이 전달되지 않았다는 평가 의견이 있었고, 12번-14번(태국어)에 대해서도 평가자 2는 발음, 성조 문제와 함께 비문이 많고 문법 문제가 많다는 의견과 함께 매우 낮은 점수를 부여했다. 한외통역에서 문장을 짧게 끊어서 통역한 사례가 다수 확인되었는데 이를 통해서도 태국어 표현력이 부족함을 알 수 있었다.

언어별 평가 결과와 평가 의견을 종합해 보면, 평가자들은 일부 예외를 제외하고 모국어 수준의 목표언어 구사 능력에 대해 문법, 문장구조, 어휘, 전문용어를 적절하게 구사하거나 완벽한 수준인 ‘상’에서 ‘최상’ 수준에 해당하는 점수를 부여했다. 한편 평가자들은 전체적으로 정보는 전달되나 부적절한 언어표현이 종종 나타나고 발음이나 강세에도 문제가 있는 경우에 ‘중’ 수준의 6~7 점을, 이러한 문제의 빈도가 잦거나 심각성이 높아 메시지의 전달을 방해하는

수준일 경우 ‘하’에 해당하는 5점을 부여했다.

아래에서는 목표언어의 각 등급 중 평가자들이 지적한 ‘중’, ‘하’에 해당하는 사례를 통해 평가 결과를 좀 더 구체적으로 살펴해보도록 하겠다. 먼저 종종 부적절한 언어표현을 구사하나 전체적인 정보의 전달에는 문제가 없는 ‘중’에 해당하는 6번(중국어)과 8번(러시아어)의 오류 사례를 제시한다.

<오류 사례 5>

검사: 증인은 한국어를 어느 정도 할 수 있나요.
 통역인: 你能讲韩国语多少?
 (한국어를 말할 수 있나요 얼마큼?)

6번(중국어) 통역인은 한외통역에서 의미는 전달되지만 어휘, 통사 구조 등 면에 문제가 있는 비문을 사용하며 통역한 사례가 여러 차례 지적됐다. 위의 사례에서 검사의 질문에 해당하는 중국어 등가 표현은 ‘证人, 你的韩语表达能力如何? (증인의 한국어 표현력은 어떤가요?), 证人, 你的韩语水平如何? (증인의 한국어 수준은 어떤가요?)’이지만 통역인은 한국어 언어 간섭을 받아 어색하게 통역했다. 통역인이 전반적인 의미는 전달했지만 문법에 맞지 않고 부적절한 중국어를 사용한 것이다.

<오류 사례 6>

증인: Вначале водой поливал, по щекам бил, потом просто оставлял в покое.
 (처음에 물을 뿌리고, 뺨을 때리고, 그 다음에 가만히 두었습니다.)
 통역인: 물 때리고, 뺨을 좀 이렇게 살짝 때리고. 뭐, 그냥.
 검사: 뺨 때린다고요?
 통역인: 뺨.
 검사: 물 때린다고요?
 통역인: 물 뿌리고, 뿌리거나 예.

<사례 6>에서 8번(러시아어) 통역인은 부적절한 한국어 표현을 사용해 검사가 알아듣지 못하자 이전 통역을 수정한다. 자가 수정에서 ‘물 때리고’라는 부적절한 표현은 정정됐지만 여전히 문장을 끝맺지 못해 목표언어 구사력에 문

제가 있음을 드러냈다.

이에 비해 문법이나 문장구조, 어휘, 전문용어를 부적절하게 구사하고 발음, 억양, 어조, 강세의 문제로 메시지의 내용을 이해하기 어려운 경우 ‘하’에 해당하는 것으로 평가되는데, 해당 사례로 1번(영어)과 7번(러시아어)의 한외통역 중 일부를 제시한다.

<오류 사례 7>

번호인: (생략) 피고인은 이전에 김송과 통화를 하면서 김송이 톰슨을 바꿔주어 알게 된 것이지요.

통역인: While you were talking to Kim Song, uh Kim Song had changed to Thompson.

(피고인이 김송과 이야기를 하면서 어 김송이 톰슨으로 바꿨습니다.)

‘전화상에서 바꿔주다’의 의미가 통역인의 목표언어 구사력 부족으로 인해 ‘A가 B를 바꾸다’의 의미로 정보가 잘못 전달되었다. 통역인이 ‘put A on the phone’과 같은 적절한 표현 대신 문자적 직역이지만 전혀 의미가 다른 표현 (change)을 사용하여 정확한 의미를 전달하는 데 실패한 것이다.

<오류 사례 8>

번호인: 견인차가 올 때까지 그런저 주변에 다른 사람이 없었지요?

통역인: До... до этой 견인차 до, которая забирает машину, да?
до этой 견인차 никого не было именно возле это у черного машина, правда?

(까지... 견인차까지, 차량을 견인하는 차까지? 견인차까지 즉 검정색 차량 근처에는 아무도 없었죠, 그렇죠?)

<사례 8>은 7번(러시아어) 녹음파일 한외통역의 일부로, 대응어가 있음에도 우리말 음을 그대로 차용한 사례이다. 위 사건에서 ‘견인차’는 반복되어 나오는 중요한 단어로, 이를 적절한 러시아어 대응어로 옮기지 않고 계속 음차 하는 것은 적절한 통역이라 할 수 없다. 통역인의 목표언어 구사력의 문제를 넘어 정확성에도 영향을 미칠 수 있는 문제라 할 수 있다. 또한 목표언어 구사력 부족으로 ‘견인차가 올 때까지’를 ‘견인차까지’로 통역하여 해당 부분의 의미가 전달

되지 않았다.

3.2.3 전달력

적절한 속도 유지, 상황에 맞는 어조 및 말투 사용, 망설임 또는 말실수의 정도를 위주로 살펴본 전달력을 평가한 결과, 14개 녹음파일의 한외 전달력 평균은 6.8, 외한 전달력 평균은 7로 양방향 모두 평균적으로 ‘중’ 수준에 해당한다.

전달력의 평가 결과는 본 연구의 평가 척도에서 제시한 세부 기준대로 발화 속도와 말실수의 빈도 등에 의해 좌우되는 경향을 보였다. 영어의 경우, 전달력 평가에서 1번은 양방향 모두 ‘중’에 해당하는 7점을 받은 반면, 2번은 양방향 모두 ‘상’에 해당하는 8점을 받았다. 영어 평가자 2는 1번에 대해 ‘전체적으로 성량이 작고 자신감이 없다’고 평가한 반면, 2번에 대해서는 ‘적절한 성량과 속도를 유지하며 통역했으며 약간의 말실수가 있으나 바로 수정했다’고 평가했다. 이는 본 연구의 평가 척도에서 제시한 세부 기준에 부합하는 평가 의견으로 적절한 성량과 속도 및 말실수 빈도 여부가 ‘상’과 ‘중’을 나누는 기준이 되었음을 확인할 수 있다.

중국어 평가에서는 평가자 1과 2는 모두 머뭇거림, 발화오류, 간투사 사용이 나타난 부분마다 평가 의견을 제시하며 말실수를 전달력 평가의 주요 기준으로 삼았다.

러시아어 평가에서도 말실수 및 이로 인한 말고침은 전달력을 평가하는 주요한 기준으로 작용했다. 7번에 대해서 러시아어 평가자 1은 외한 방향과 한외 방향 모두에서 ‘잡은 말고침의 통역습관으로 인하여 전달력이 약하거나 떨어진 다’고 지적했다. 아래에서는 전달력의 각 등급 중 평가자들이 지적한 ‘중’, ‘하’에 해당하는 사례를 통해 평가 결과를 좀 더 구체적으로 살펴보도록 하겠다.

말실수로 인해 전달력이 다소 약화된 ‘중’에 해당하는 경우로 1번(영어)과 6번(중국어) 통역 사례를 제시한다. 위 <사례 1>에서 볼 수 있는 것처럼 1번 영어 통역인은 영어로 통역할 때 주저하면서 ‘어’를 빈번히 사용하는 등 발화속도가 처지는 경우가 있었다.

<오류 사례 9>

번호인: 여기 내용을 보면 ‘아는 언니의 소개로 시안마사지에서 일하게 되

었다'는 내용이 적혀 있는데, 이는 사실과 다르지요?

통역인: 那么说, 这儿有通过认识的姐姐的介绍到, 那个, 那个西安按摩店, 所, 那个, 干活儿的, 那与事实是不同吧?

(그렇게 말한다면, 여기에 아는 언니의 소개된 것을 통해 그것, 그 시안마 사지에 가서 그, 일을 했다는 것이 있는데, 그것이 사실과 다르죠?)

증인: 因特网里边, 就是QQ里边儿, 因特网里边儿的。
(인터넷 안에, QQ 안에, 바로 인터넷 안에서요.)

<사례 9>의 중국어 통역인은 발성이 좋고 발화 속도도 빠른 편이었지만 거의 모든 질문을 ‘那么说(그렇게 말한다면)’라는 간투사로 시작했고 ‘那个(그것, 그)’와 같은 간투사를 빈번하게 사용하여 질문이 효과적으로 전달되지 않아 잘 이해하지 못한 증인이 동문서답을 하였다.

자주 망설임이 있고 말실수가 반복되어 메시지 전달력이 부족한 ‘하’(5 이하)에 해당하는 예로 6번(우즈벡어)와 12번(태국어)를 제시한다.

<오류 사례 10>

증인: Koreyscha ismini bilmayman ularning.
(그분들의 한국어 이름을 모릅니다.)

통역인: 한국... 여기 보니까 김주은이라는 이름 자체에 대한... 한국...어로 된 이름을 잘 모르겠습니다. 김주은이 누군지.

<사례 10>에서 6번(우즈벡어) 통역인은 전체적으로 성량이 작고 목소리에 자신감이 없었으며 말끝을 흐린 사례가 많았다. 증인은 말을 더듬거리거나 머뭇거리지 않고 또렷한 목소리에 빠른 속도로 말을 마쳤지만 통역인은 ‘한국어 이름’이라는 표현을 사용하지 못해 ‘한국’에서 말을 멈추고 머뭇거리거나 말을 고치고 얼버무리며 통역을 마쳤다. 통역인의 전달력 부족으로 발화 스타일이 증인과 확연히 달라졌다는 것을 알 수 있다.

위 <사례 4>에서 볼 수 있듯이 12번 태국어 통역인은 목표 언어 구사력과 표현력 부족으로 태국어로 문장을 완성하는 데 어려움을 겪으며 잦은 망설임과 휴지가 있었다.

4. 결론

본 연구는 한 지방법원의 형사재판에서 이뤄진 외국인 신문 총 14건의 녹음파일에 담긴 7개 언어 통역인 11명의 통역 품질을 평가 척도를 이용하여 평가하고 결과를 분석했다. 평가를 위해 본 연구는 선행연구를 토대로 원 발화 메시지에 대한 충실성을 보는 ‘정확성’, 목표텍스트로서 적절한지를 가늠하는 ‘목표언어’, 그리고 통역으로서 발화 속도나 음성과 어조 등이 적절한지를 보는 ‘전달력’, 이 세 가지를 평가 기준으로 삼고, 기준마다 ‘최상(10점)’, ‘상(9-8점)’, ‘중(7-6점)’, ‘하(5점 이하)’의 4개 등급으로 이루어진 10점 척도를 이용하여 평가하여 각 기준별 점수와 종합점수를 도출하도록 했다. 언어별로 각 2인의 평가자들이 평가 방식에 대해 안내 받고 평가를 실시한 결과, 통역인 개별 점수의 차이는 있었지만 종합 평균은 ‘중’에 해당하는 것으로 드러났다. 대부분이 ‘중’에 해당하는 평가를 받았지만 3번 일본어 통역과 5번 중국어 통역이 종합평가에서 ‘상’에 해당하는 등급으로 평가받았고, 7번 러시아어와 12-14번 태국어 통역인은 ‘하’에 해당하는 평가를 받았다.

평가 기준별로 살펴보면 정확성은 메시지를 전반적으로 전달하는 수준으로 간주되는 ‘중’에 해당했다. 평가자들은 정확성에서 평가 척도에 제시된 대로 원 메시지에 제시된 정보에 대해 추가, 누락, 왜곡이 있는지 여부 및 해당 추가, 누락, 왜곡이 정보의 전달에 영향을 미치는 정도를 바탕으로 각 녹음파일의 정확성을 평가했다. 목표언어 역시 양방향 모두 평균 ‘중’ 수준으로 평가자들은 목표언어의 문법, 문장구조, 어휘, 전문용어, 발음 및 강세를 위주로 목표언어 수준을 평가했다. 한편 목표언어의 평가는 상당 부분 통역인의 모국어와 연관이 있는 경향을 보여, 통역인 상당수가 모국어로의 통역에서 상대적으로 높은 점수를 받았다. 전달력 평가 결과 또한 양방향 모두 평균 ‘중’ 수준으로 평가자들은 발화의 속도, 어조 및 말투, 망설임 또는 말실수의 정도를 위주로 전달력을 평가했다. 전달력의 평가 결과 역시 본 연구의 평가 척도 기준대로 통역인의 성량과 발화 속도, 말실수의 빈도에 의해 좌우되는 경향을 보였다.

전반적으로 평가자들은 본 연구의 평가 척도에서 제시한 정확성, 목표언어, 전달력의 기준을 비교적 정확하게 인지하고 세부 기준에 명시된 대로 통역 품질 수준을 가려내고 적절한 등급의 점수를 부여했다. 종합점수의 등급을 기준

으로 살펴 본 평가자간 신뢰도에서 7개 언어 모두 신뢰 가능한 수준인 것으로 확인된 점에 비추어 볼 때, 통역을 전문으로 하지 않는 언어전문가도 평가자에 포함되어 있음에도 불구하고 통역 비전문가가 평가 척도를 사용하는 데 큰 문제가 없으며, 어느 정도 평가자간 유사한 평가 결과를 도출할 수 있음을 보여준 결과라 할 수 있다. 이는 평가 척도의 유용성을 논한 기존의 리(2008), 티셀리우스(2009) 등의 연구 결과를 뒷받침하는 중요한 연구결과다.

다만 본 연구 대상인 사건의 종류, 녹음 길이, 통역의 난이도가 다르다는 변수가 있고, 사법통역 평가 및 교육 경험이 전무한 소수언어 평가자들을 위해서는 적절한 평가를 위한 연구자의 세심한 조율이 불가피했다. 몽골어와 태국어의 세부 점수 퍼센트 일치도만 영어, 일본어, 중국어, 러시아어, 우즈베크어에 비해 상대적으로 낮게 나온 분석 결과를 보면 통일된 평가자 훈련과 적절한 조율이 뒷받침된다면 사법통역 품질평가 및 인증시험 평가 목적에도 평가 척도를 충분히 활용이 가능하다는 것을 확인할 수 있었다.

무엇보다 본 연구의 의의는 사법통역 모니터링 및 인증시험 평가 목적에 비교적 용이하게 활용할 수 있는 평가 도구를 개발하여 실제 사법통역 품질 평가에 적용해 보았다는 데 있다. 특히 소수언어를 포함한 여러 언어를 동시에 동일한 방식으로 평가하기 위해 개발한 평가 척도를 사용하여 평가를 수행했다는 점에서 기존에 담화분석 위주로 이루어진 국내 사법통역 연구의 지평을 넓혔다고 여겨진다.

참고문헌

- 김혜림 (2011) 「교육에서의 통역평가 기준에 대한 고찰: 조사연구를 중심으로」, 『통역과 번역』 13(2): 71-91.
- 이지은 (2012) 『다문화사회의 사법통역』. 서울: 집문당.
- 이지은 (2013) 「사법통역 교육에 관한 실행연구: 사법통역 전문가 기초과정 운영 사례」, 『번역학연구』 14(5): 195-223.
- 이지은, 장원경 (2015) 「사법통역인의 순차통역 평가 사례 연구」, 『언어학연구』 20(1): 99-118.

- 이지은, 장원경 (2016) 「사법통역 제도개선을 위한 소고」, 『이화여대 법학논집』 20(1): 251-277.
- 이혜승 (2008) 「동시통역 품질평가에 대한 경험적 연구 고찰」, 『통번역학연구』 12(1): 91-107.
- 정철자 (2012) 「통역 평가 비교가능성: 교육상황에서의 사례연구」, 『통번역학연구』 16(2): 219-239.
- 편혜원 (2004) 「교사 프로파일에 따른 AB 통역평가에 관한 사례연구」, 『통역과 번역』 6(2): 207-224.
- Angelelli, Claudia V. (2009) 'Using a Rubric to Assess Translation Ability: Defining the Construct', in Claudia V. Angelelli and Holly E. Jacobson (eds) *Testing and Assessment in Translation and Interpreting Studies: A Call for Dialogue between Research and Practice*. Amsterdam: John Benjamins, 13-99.
- Bachman, Lyle F. and Adrian S. Palmer (1996) *Language Testing in Practice: Designing and Developing Useful Language Tests*, Oxford: Oxford University Press.
- Clifford, Andrew (2005) 'Putting the Exam to the Test: Psychometric Validation and Interpreter Certification', *Interpreting* 7(1): 97 - 131.
- Giambruno, Cynthia (ed) (2014) *Assessing Legal Interpreter Quality through Testing and Certification: The Qualitas Project*, University of Alcant/Alicante Publications.
- Graham, Mathew, Anthony Milanowski, and Jackson Miller (2012) 'Measuring and Promoting Inter-Rater Agreement of Teacher and Principal Performance Ratings' The U.S. Department of Education, February. Available at <https://files.eric.ed.gov/fulltext/ED532068.pdf>.
- Hale, Sandra (2019) 'Specialist Legal Interpreters for a Fairer Justice System', in Said Faiq (ed) *Discourse in Translation*, Oxon, NY: Routledge, 47-66.
- Han, Cao (2018) 'Using Rating Scales to Assess Interpretation: Practices, Problems and Prospects', *Interpreting* 20(1): 59-95.
- Hatim, Basil and Ian Mason (1997) *The Translator as Communicator*, London

& New York: Routledge.

- Kalina, Sylvia (2002) 'Quality in Interpreting and its Prerequisites: A Framework for a Comprehensive View', in Garzone, Giuliana and Maurizio Viezzi (ed) *Interpreting in the 21st Century: Challenges and Opportunities*, Amsterdam: John Benjamins, 121-130.
- Kurz, Ingrid (2001) 'Conference Interpreting: Quality in the Ears of the User', *Meta* 46(2): 394-409.
- Lee, Jieun (2008) 'Rating Scale for Interpreting Performance', *The Interpreter and Translator Training* 2(8): 172-173.
- Moore, Alan D. and Suzanne Young (1997) 'Clarifying the Blurred Image: Estimating the Interrater Reliability of Performance Assessment', paper presented at the annual meeting of the Northern Rocky Mountain Educational Research Association. Eric Document No. 414319, 1-19.
- NAATI (n.d.) 'Certified Interpreter Test Assessment Rubrics'. Available at www.naati.com.au/https://www.naati.com.au/certification/certification-testing/
- Pöschhacker, Franz (2001) 'Quality Assessment in Conferences and Community', *Meta* 46(2): 410-425.
- Riccardi, Alessandra (2002) 'Evaluation in Interpretation: Macrocriterias and Microcriterias', in Eva Hung (ed.) *Teaching Translation and Interpreting 4: Building Bridges*, Amsterdam & Philadelphia: John Benjamins, 115-26.
- Schjoldager, Anne (1996) 'Assessment of Simultaneous Interpreting', in Cay Dollerup and Vibeke Appel (eds) *Teaching Translation and Interpreting 3: New Horizons*, Amsterdam & Philadelphia: John Benjamins, 187-95.
- Stansfield, Charles, W. and William E. Hewitt (2005) 'Examining the Predictive Validity of a Screening Test for Court Interpreters', *Language Testing* 22(4): 438-462.
- Tiselius, Elisabet (2009) 'Revisiting Carroll's Scales', in Claudia V. Angelelli and Holly E. Jacobson (eds) *Testing and Assessment in Translation and Interpreting Studies: A Call for Dialogue between Research and Practice*, Amsterdam & Philadelphia: John Benjamins, 95-121.

[Abstract]

Court Interpreting Assessment Using Rating Scales

Lee, Jieun·Lee, Yoo-Jin Choi, Hyo-eun
(Ewha Womans University·HUFS·Ewha Womans University)

For defendants and witnesses who do not speak the official language of the court, the provision of interpreting and the quality of court interpreting are equally important. This paper aims to develop a rating scale for the assessment of the competence of court interpreters and court interpreting quality. Based on the literature review, we have developed a rating scale with the three criteria of accuracy, target language quality, and delivery. The ten-point rating scale has four bands, 'excellent', 'good', 'mediocre', and 'poor', for each criterion of quality, and points assigned to each band: 10 points for 'excellent', 8 to 9 points for 'good', 6 to 7 points for 'mediocre', and 5 points and below for 'poor'.

The audio-recorded interpreting data and Korean court record of the examinations, which were made available for this study, consisted of dialogue interpreting during fourteen courtroom examinations of witnesses and defendants from non-Korean speaking backgrounds. The data of 10 hour and 34 minute long interpreting involved eleven interpreters in seven languages, including English, Chinese, Russian, Japanese, Mongol, Thai, and Uzbek. Each language interpreting was assessed by two raters based on the rubric provided by the researchers. While English, Chinese, Russian and Japanese raters had at least MA level training in translation and interpreting and included at least one rater who was familiar with court interpreting, Mongol, Thai, and Uzbek raters were language instructors or fluent bilinguals, lacking professional translator or interpreter training. The results indicate that the raters generally understood the

criteria and the band descriptors and came up with similar results. The inter-rater reliability measured by a percent agreement between raters proved to be high. In terms of overall band scores, the rater agreement was 100% in six languages except for English, which was still strong (75%). The findings support the practicality and effectiveness of using rating scales in the assessment of court interpreting in different contexts, including certification examination. Despite the limitations of the current research, the rating scale proved to be a useful tool that can be employed by non-professional interpreter raters. The assessment of eleven interpreters performances revealed that the overall interpreting competence of the eleven interpreters was mediocre.

▶ Key Words: court interpreting, assessment, quality, criteria, rating scale

▶ 주제어: 사법통역, 평가, 품질, 평가 기준, 평가 척도

이지은

이화여대 통역번역대학원 통역학과 교수

jieun.lee@ewha.ac.kr

관심분야: 사법통역, 커뮤니티 통역, 통번역 교육, 담화분석

이유진

한국외대 통번역대학원 한중과 강사

feifei99@hanmail.net

관심분야: 사법통역, 제도번역, 커뮤니티 통역, 통역 교육

최효은

이화여대 통역번역대학원 통역학과 강사

cutedinojr@naver.com

관심분야: 법률번역, 특허번역, 사법통역, 번역 교육

논문투고일: 2019년 4월 29일

심사완료일: 2019년 5월 25일

게재확정일: 2019년 5월 28일