

디지털 성서학 플랫폼: 성서 데이터베이스 구축 현황과 디지털 성서학 연구 플랫폼 개발의 방향성

김경식(감신대)

1. 들어가는 말

ChatGpt는¹ 인공지능(Artificial Intelligence)의 특이점을 열었다고 평가될 정도로 오랜 역사를 가진 인공지능 기술의 분기점이 되고 있다. 인공지능 기술의 역사는 매우 길고,² 그동안 획기적인 발전도 이루어져 왔지만, ChatGpt 이전의 인공지능은 근본적인 ‘언어 능력의 한계’를 갖고 있었다.³ 따라서 컴퓨터가 인간의 언어를 완벽히 이해하고, 인간이

- 1 OpenAI가 개발한 ChatGpt는 Gpt 버전 3.5와 4가 탑재된 챗봇(chatbot)으로 사용자가 인공지능 프로그램과 직접적인 언어로 소통할 수 있도록 만들어진 것이다. “ChatGpt”, <https://chat.openai.com/> (2023.4.14. 접속).
- 2 인공지능(Artificial Intelligence)이라는 학문 분야는 1956년 다트머스 회의(Dartmouth Workshop)에서 사용되기 시작하였다. 이와 관련하여 다음을 참고하라. Nils J. Nilsson, *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge: Cambridge Univ. Press, 2010), 52-55.
- 3 Gpt는 대규모 언어 모델(Large Language Model)이라 불리는 방법으로 이 문제를 상당 부분 해결했다고 여겨지고 있는데, 정확하게 공개하고 있지는 않지만 수천억 개에서 조

이해할 수 있는 자연스러운 언어를 구사하게 하기까지 컴퓨터에 수많은 문헌 자료를 입력시키고 인간 언어의 일정한 패턴과 규칙을 학습시키기 위한 언어 분석 기술의 개량과 발전은 필수적이었다.⁴ 그 결과 소위 ‘자연어 처리’(Natural Language Processing) 분야가 획기적으로 발전하게 되었고, 이 분야의 하위 범주로 형태소 분석(Morphological Analysis), 구문 분석(Parsing, Syntax Analysis), 의미 분석(Semantic Analysis), 그리고 감정 분석(Sentiment Analysis) 등과 같은 다양한 언어 분석 기술이 발전을 거듭해 왔다. 이러한 분석 기술은 단순히 컴퓨터가 인간의 언어를 학습하는 과정을 효율적으로 만들었을 뿐만 아니라, 매우 탁월한 컴퓨터의 연산 능력을 통해 인간 스스로 언어를 구사하거나 문헌을 기록할 때 의식하지 못했던 다양한 패턴과 규칙을 밝혀내 주기도 하였다. 컴퓨터를 활용한 이러한 언어 분석 기술은 문헌을 주된 연구 대상으로 삼는 인문학에도 영향을 미쳐 소위 “디지털 인문학”(Digital Humanities)을 태동시키고 발전시키는 계기가 되었다.⁵

단위의 문서를 학습시킨 것으로 알려져 있다. 그 결과 지금까지 만들어진 인공지능 모델 가운데 가장 자연스럽게 인간의 언어를 구사하게 되었다. Gpt 모델을 기반으로 한 챗봇 서비스(ChatGpt)는 다음의 주소에서 사용해 볼 수 있다. <https://chat.openai.com/> (2023.4.12. 접속).

- 4 1950년 앨런 튜링(Alan Turing)이 고안한 “튜링 테스트”에 따르면 컴퓨터와 대화하는 사람이 상대방을 컴퓨터가 아니라 사람으로 생각할 때, 그 컴퓨터는 인공지능의 중요한 조건을 갖춘 것으로 판단된다. 지금도 이 튜링 테스트는 다양한 방식으로 인공지능 평가에 활용되고 있다. 앨런 튜링, “계산 기계와 지능” (김지홍 옮김), http://aitimes.org/wp-content/uploads/2017/02/Allan_turing_Paper_1950_한국어번역.pdf (2023.4.14. 접속), 원제는 A. M. Turing, “Computing Machinery and Intelligence”, *Mind* 59 (1950), 433-460
- 5 디지털 인문학의 시초는 로베르토 부사(Roberto Busa)가 천공 카드로 토마스 아퀴나스의 저작을 디지털화하기 시작했던 토마스티쿠스 색인 프로젝트(Robert Busa, “Index Thomasticus”, <https://www.corpusthomaticum.org/it/index.age> (2023.4.14. 접속))라 여겨지고 있다. 그리고 1990년대 인터넷 웹 기술이 본격적으로 발전하면서 데이터 접근성이 향상되었고 컴퓨터 성능의 비약적인 발전과 함께 대규모 데이터 분석이 가능해지기 시작했기 때문에 이때부터 진정한 의미에서의 디지털 인문학이 시작되었다고 할 수 있다. 그리고 이 분야는 21세기에 들어와 오픈 데이터, 클라우드 컴퓨팅, 인공지능과 기계

사실 ‘디지털 인문학’이라는 용어와 그에 대한 정의는 이미 현저히 상투화된 표현인 ‘4차산업혁명’만큼이나 모호하고 손에 잘 잡히지 않는다. 일례로 “What Is Digital Humanities”라는 웹사이트는 디지털 인문학에 대한 800여 개의 정의를 무작위로 출력해 주고 있다.⁶ 그만큼 ‘디지털 인문학’ 자체를 직관적으로 정의하는 일은 결코 쉬운 일이 아니다. 전통적인 시각에서 보면 결코 어울릴 것 같지 않은 ‘디지털’과 ‘인문학’은 앞서 언급한 컴퓨터 연산 능력을 활용한 다양한 방식의 언어 분석을 통해 매우 놀라운 시너지 효과를 이루어 내고 있으며 현대 인문학의 지평을 넓히고 있다.

성서학 역시 성서 문헌을 일차적인 연구 대상으로 삼는다는 점에서 문헌을 연구하는 인문학의 한 갈래이며, 다른 인문학 분야만큼 아직 본격적이라 할 수는 없는 수준이지만, 해외 학계에서는 디지털 인문학의 방법론을 성서학 분야에 적용하는 연구들이 적지 않게 진행되고 있다.⁷ 이 논의에서는 디지털 인문학 방법론을 도입한 성서학 연구 방법론을 편의상 ‘디지털 성서학’이라 명명하고자 한다.⁸

학습 등의 기술이 발전함에 따라 최근 비약적으로 발전하고 있다. 디지털 인문학의 개괄적인 역사에 대해서는 다음을 참고하라. Susan Hockey, “The History of Humanities Computing”, Susan Schreibman/Ray Siemens/John Unsworth (eds.), *A Companion to Digital Humanities* (Malden: Blackwell Publishing, 2004), 3-19; 국내에서의 디지털 인문학의 현황과 과제를 정리한 다음의 논문 또한 참고하라. 김현, “디지털 인문학”, 『인문콘텐츠』 29 (2013), 9-26.

- 6 “What Is Digital Humanities?”, <http://whatisdigitalhumanities.com/> (2023.4.14. 접속).
- 7 대표적인 예가 브릴(Brill) 출판사가 출간한 “디지털 성서학”(Digital Biblical Studies) 시리즈의 저서들이다(Claire Clivaz, Ken M. Penner 편집). 이 시리즈에는 성서학과 고대 근동학, 그리고 사본학 등과 같은 분야에 있어 디지털 인문학을 적용한 연구물들이 수록되어 있다.
- 8 이 용어는 각주 7에서 언급한 저작물의 이름을 차용한 것이다. 물론, ‘디지털 성서학’(Digital Biblical Studies)라는 용어가 보편적으로 사용되는 것은 아니지만 간결한 표현을 통해 그 의미가 효과적으로 전달될 수 있으므로 이 용어를 사용하고자 한다.

www.kci.go.kr

현재 디지털 인문학을 위한 다양한 도구들이 개발되어 있지만,⁹ 인문학의 각 영역에 적용되는 방법론이나 연구 목표와 방향성이 다양한 만큼, 각 영역에 특화된 도구들의 개발이 중요하다고 볼 수 있다. 그러나 아쉽게도 국내에는 성서학에 특화된 디지털 인문학 도구의 개발이 상대적으로 활발히 진행되지 못하는 상황 가운데 있다.¹⁰ 본 연구에서는 이러한 상황 가운데 디지털 인문학의 방법론을 성서학에 접목하기 위한 연구 플랫폼 개발의 필요성과 그 방향성에 대해 논의하려고 한다. 특히, 디지털 성서학 연구 플랫폼 개발을 위해 필수적이라 할 수 있는 성서 본문 데이터와 이를 이용한 연구 플랫폼의 구조적인 방향성에 대해 논의할 것이다. 히브리성서 본문에 기반한 연구 플랫폼 개발, 그리고 그 활용과 관련한 선행 연구는 다소 제한적이기는 하지만 참고해 볼 만한 몇 가지 중요한 논의들이 있다. 디지털 문헌 자료 플랫폼과 언어 분석 도구 개발을 위한 성서 데이터의 구조와 이를 활용한 성서 소프트

9 예를 들면 키워드 빈도 분석, 워드클라우드, n-그램 분석 등의 다양한 기능을 제공하는 “보안트 툴즈”(“Voyant Tools”, <https://voyant-tools.org/> (2023.4.14. 접속), 말뭉치 분석 도구인 “앤티 컨”(“AntConc”, <https://www.laurenceanthony.net/software/antconc/> (2023.4.14. 접속)) 등이 있다.

10 국내에도 성서학의 관점에서 인공지능을 논의하는 연구나, 메타버스와 성서학을 접목하려는 연구들이 일부 있다. 최근 「구약논단」을 통해 발표된 주요 논문들은 다음과 같다. 김창주, “인공지능과 구약성서: 무슨 상관이 있는가(욥 21:21)”, 「구약논단」 28/4 (2022), 187-214; 소형근, “포스트휴머니즘 시대에 인간과 인공지능 이해하기: ‘강한 인공지능’의 등장에 어떻게 대처할 것인가?”, 「구약논단」 28/4 (2022), 247-273; 이유미, “유용한 해석적 도구인 ‘비블리카버스’(bilibiverse) 구축을 위한 제언-ID drama의 아가 연구 가상체험기”, 「구약논단」 28/2 (2022), 324-354. 국내에 소개된 논문 가운데 보다 진정한 의미에서 디지털 인문학 방법론을 성서학에 접목시킨 연구의 예는 다음과 같다. 진규상, “Analysis of Participants in the First Major Division of Numbers”, 「구약논집」 22 (2022), 8-59; “Analysis of Participants’ Agent Role in the Two Major Divisions of Leviticus”, 「성서원문연구」 52 (2023), 101-138. 이 연구는 암스테르담 자유 대학의 부설 연구소 ETCBC에서 개발한 BHSA 데이터를 활용하여 진행되었는데, 이 데이터에 대해서는 아래에서 논의할 예정이다. 이 외에도 권성달 교수는 통계 프로그램을 활용한 히브리어 언어 관련 연구물들을 출판하였다. 예를 들어 다음을 참고하라. 권성달, “성서히브리어 명사문장의 어순에 관한 연구”, 「성경원문연구」 35 (2014), 170-194.

웨어(SHEBANQ) 개발의 예를 제시한 디르크 루르다(Dirk Roorda)의 연구와¹¹ 디지털 성서학 도구가 성서학의 다양한 연구 분야에 어떻게 적용될 수 있을지를 구체적으로 제시한 빈터-닐슨(Nicolai Winther-Nielsen)의 연구 등을 참고할 만하다.¹² 그리고 동시에 디지털 도구를 사용하여 문헌 연구에 접근했을 때, 어떤 과정으로 진행되는지 모른 채 버튼만 눌러 결과만 보려 하는 연구 방식의 문제점을 제기한 마리엔버그-밀리코우스키(Marienberg-milikowsky)의 연구는 균형 있는 디지털 성서학 연구 플랫폼 개발 과정에서 놓치지 말아야 할 중요한 논의를 담고 있다.¹³

이제 디지털 성서학 연구를 위해 필수적이라 할 수 있는 성서 본문 데이터의 종류와 특징을 살펴보고자 한다. 본 연구의 연구 범위는 지면의 한계상, 구약성서 히브리어 본문에 국한하고 있음을 밝힌다.

2. 구약성서 히브리어 본문 데이터

컴퓨터를 활용한 성서 본문 분석을 위해서는 분석의 대상이 되는 문헌 데이터가 우선적으로 갖추어져야 하는데, 특정한 연구 목적을 위해 다양한 형태로 가공할 수 있는 데이터가 필수적이다. 현재 연구자들

11 Dirk Roorda, "The Hebrew Bible as Data: Laboratory - Sharing - Experiences", Jan Odijk and Arjan van Hessen (eds.), *CLARIN in the Low Countries* (London: Ubiquity Press, 2017), 217-230.

12 Nicolai Winther-Nielsen, "Interactive Tools and Tasks for the Hebrew Bible: From Language Learning to Textual Criticism", *Journal of Data Mining and Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages (2017), <https://doi.org/10.46298/jdmdh.4003> (2023.5.20. 접속).

13 Itay Marienberg-milikowsky, "It Functions, and That's (Almost) All": Tagging the Talmud", Domenico Fiormonte et al. (eds.), *Global Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2022), 141-150.

이 사용할 수 있는 다양한 성서 본문 데이터가 구축되어 있지만, 신뢰성 있는 연구를 위해서는 각 데이터가 어떠한 배경과 특징을 지니고 있는지 이해해 볼 필요가 있다. 여기에서는 구약성서 본문과 관련된 주요 데이터들을 유형별로 소개해 보고자 한다.

1) 구약성서 히브리어 본문 (Text)

구약성서 히브리어 본문 데이터는 문헌 분석을 위한 가장 기초적인 데이터라 할 수 있다. 온라인상에서 다양한 히브리어 성서 본문 데이터를 찾아볼 수 있지만¹⁴ 본 연구에서는 디지털 인문학 방법론을 적용할 수 있는 데이터 포맷으로 만들어진 디지털 자료를 다루어 보고자 한다. 먼저, 웨스트민스터 신학 대학(Westminster Theological Seminary)의 그로브스 센터(Groves Center)에서 진행하는 프로젝트이다. 이 연구소에서 진행하는 프로젝트는 3가지인데,¹⁵ 그중 하나가 바로 히브리어 성서 원문을 디지털화하는 작업인 “The Westminster Leningrad Codex”(WLC)이다. 이 데이터는 1983년에 출간된 BHS 본문을 1987년에 디지털화하는 작업으로 시작되었고, 알란 그로브스(Alan Groves)의 주도하에 레닌그라드 코덱스(Leningrad Codex, Firkovich B19A) 원문 내용에 가깝게 지속적인 수정 작업을 거쳐 만들어졌다.¹⁶ 이 원문 데이터는 같은 연구소에서 만들어진 형태소 데이터베이스(Westminster Hebrew Morphology)와 구문 데이

14 예를 들면, 독일성서공회에서 제공하는 BHS 본문이 있다. “Biblia Hebraica Stuttgartensia (BHS)”, <https://www.academic-bible.com/en/online-bibles/biblia-hebraica-stuttgartensia-bhs/read-the-bible-text/> (2023.4.14. 접속).

15 “J. Alan Groves Center”, <https://students.wts.edu/resources/alangroves/grovesprojects.html> (2023.4.14. 접속); <https://www.grovescenter.org/> (2023.4.14. 접속) 등을 참고하라.

16 이 데이터는 공공 자산(public domain)으로 공개되어 있어 누구나 자유롭게 사용할 수 있다. 앞서 언급한 Open Scriptures Hebrew Bible 프로젝트 역시 이 데이터를 활용하였다. 원문 내용은 다음의 사이트에서 확인할 수 있다. “Tanach”, <https://tanach.us/> (2023.3.23. 접속).

터베이스(Westminster Hebrew Syntax)의 기반이 되었고, 또 다른 공개 형태소 자료인 “Open Scriptures Hebrew Bible”에 사용되었다. 이 데이터는 XML의 형식으로 공개되어 누구나 사용할 수 있다.¹⁷

이 외에도 아래에서 소개할 형태소(Morphology)와 구문(Syntax) 데이터들에는 기본적으로 히브리어 원문 데이터가 탑재되어 있고, 히브리어 원문 정보를 제공하는 성서 소프트웨어나 웹사이트는 상당히 많으므로 히브리어 원문 데이터에 접근하는 것은 그리 어렵지 않다.

2) 본문 + 형태소 데이터베이스(Text + Morphology Database)

히브리어 본문 데이터만을 가지고는 디지털 성서학 연구를 시작할 수는 없다. 기본적인 본문 데이터에 언어와 문법에 관련된 정보들이 추가되어야 한다. 일단 히브리어 성서 본문 데이터가 준비되면 히브리어 본문을 이루는 문법적인 구성 요소의 단위를 세분화하여 각 층에 대한 정보를 단계적으로 부가해야 한다. 이를테면 성서 본문 구조는 <그림 1>과 같은 계층구조로 구분될 수 있고, 각 구성 요소는 고유한 기능을 가지면서, 서로 유기적으로 연결되어 있다. 특별히 본문의 의미나 구조 파악을 위해 중요한 요소는 문장(sentence), 절(clause), 구(phrase), 그리고 단어(word) 등의 요소이다. 이 가운데 단어 요소와 관련하여 중요한 데이터는 단어의 변화형 정보

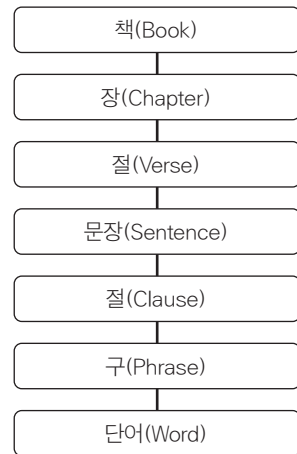


그림 1. 성서 본문의 계층 구조

17 이 원문 데이터는 공공재(Public Domain) 라이선스로 공개되어 있다. 다음의 사이트를 통해 다운로드 할 수 있다. “Westminster Leningrad Codex Hebrew Old Testament”, <https://find.bible/bibles/HEBWLC/> (2023.4.14. 접속).

를 담은 형태소 데이터, 그리고 의미론적 정보를 포함한 사전 데이터 등이다.

먼저 형태소 정보와 관련하여 주목해볼 만한 데이터베이스는 오픈 소스의 형태(CC BY 4.0)로 공개된 “Open Scriptures Hebrew Bible”(OSHB)이다.¹⁸ 이 본문 데이터는 앞서 언급한 WLC 히브리어 본문에 형태소(morphology) 정보를 추가한 것이다. 이 데이터는 OSIS XML 방식으로¹⁹ 구조화되어 있는데 데이터 형태는 <그림 2>와 같다.

```

<div type="book" osisID="Gen">
  <chapter osisID="Gen.1">
    <verse osisID="Gen.1.1">
      <w lemma="b/7225" n="1.0" morph="HR/Ncfsa" id="01xeN">בְּרֵאשִׁית</w>
      <w lemma="1254 a" morph="HVqp3ms" id="01Nvk">אֱלֹהִים</w>
      <w lemma="430" n="1" morph="HNcempa" id="01TyA">עָשָׂה</w>
      <w lemma="853" morph="HTo" id="01vuQ">לְמַעַן</w>
      <w lemma="d/8064" n="0.0" morph="HTd/Ncempa" id="01TSc">אֶתְיָ</w>
      <w lemma="c/853" morph="HC/To" id="01k5P">לְעֹשֶׂה</w>
      <w lemma="d/776" n="0" morph="HTd/Ncbsa" id="01nPh">וְיָרֵא</w><seg type="x-sof-pasuq">.</seg>
    </verse>
    ...
  </chapter>
  ...
</div>

```

그림 2. OSHB 형태소 데이터 구조

18 “The Open Scriptures Hebrew Bible”, <https://hb.openscriptures.org/> (2023.4.14, 접속) 본문 데이터의 소스코드는 다음의 주소를 참고하라. “Open Scriptures (Github)”, <https://github.com/openscriptures/morphhb/tree/master/wlc/> (2023.4.14, 접속).

19 XML이란 웹문서에 부가적인 주석을 추가하는 마크업(Markup) 표기 방식의 일종으로 <chapter></chapter>와 같은 태그로 기본적인 속성을 표시한다. 그리고 아래의 예에서 볼 수 있듯이 태그(tag) 안에 부가적인 속성값들(lemma, morph)을 추가할 수 있다. 다음을 참고하라. “XML이란 무엇인가요?” <https://aws.amazon.com/ko/what-is/xml/> (2023.3.23, 접속). 특별히 이 데이터는 성서 데이터의 통일성 있는 구축을 위해 미국 성서공회(American Bible Society)에서 제안한 XML 표준 포맷인 OSIS(Open Scriptural Information Standard) XML 포맷을 활용하여 문법 정보가 부가되어 있다. OSIS XML

위 데이터는 앞서 언급한 계층구조 일부를 반영하고 있는데, 책(book), 장(chapter), 절(verse)/단어(w)의 태그로 구분되어 상위 계층이 하위 계층을 위아래로 둘러싸고 있는 형식으로 구조화되어 있다. 탭(tab)으로 들여쓰기가 되어 있는 요소가 하위 요소가 된다. 그리고 각 단어 요소마다 형태소 속성, 그리고 사전(lexicon) 정보와의 연결을 위해 필요한 스트롱 번호(Strong's Number)가²⁰ 추가되어 있다. 데이터 구조상으로 각 단어의 스트롱 번호와 어형 정보는 각각 lemma와 morph와 같은 속성값을 통해 표시된다. 위의 예에서 lemma는 각 히브리어 단어의 스트롱 번호를 나타내고, morph는 형태소 정보를 나타내는데, 이는 축약어로 표시된다.²¹ 위와 같은 형태의 XML 파일을 활용하기 위해서는 연구 목적에 맞게 데이터를 적절한 형태로 표현하는 파서(parser)가 필요한데, 이 데이터를 활용한 구약성서 히브리어 본문은 프로젝트 사이트에서 확인할 수 있다.²²

이 데이터는 표준화된 XML 방식을 채택하여, 이를 이용한 프로그램 개발을 용이하게 했다는 데에 의미가 있다. 그리고 이 프로젝트의 제목에서 볼 수 있듯이 누구나 이 프로젝트에 참여하여 데이터의 수정·

의 데이터 구조와 자세한 정보에 대해 다음의 문서를 참고하라. Patrick Durusau, *Osiris Users Manual (OSIS Schema 2.1.1)*, <https://crosswire.org/osis/OSIS%202.1.1%20User%20Manual%2006March2006.pdf> (2023.4.14. 접속).

- 20 스트롱 번호(Strong's Number)란 19세기 제임스 스트롱(James Strong)이 편찬한 사전(James Strong, *The Exhaustive Concordance of the Bible* (Cincinnati: Jennings & Graham., 1890))에서 사용된 번호 체계이다. 이 체계에는 성서 속의 모든 히브리어, 아람어, 그리스어 단어에 고유한 번호를 부여하여 용이하게 각 단어의 의미를 참조하기 위한 목적으로 만들어졌다. 이 스트롱번호는 현재까지도 사용되고 있다.
- 21 예를 들어 Ncfsa는 다음의 의미가 있다. N=명사, c=보통명사, f=여성, s=단수, a=절대형. 이 데이터가 사용하는 축약어는 다음을 참고하라. "Hebrew Morphology Codes (The Open Scriptures Hebrew Bible)", <https://hb.openscriptures.org/parsing/HebrewMorphologyCodes.html> (2023.4.14. 접속).
- 22 "The Open Scriptures Hebrew Bible - Read", <https://hb.openscriptures.org/read/> (2023.4.14. 접속)

www.kci.go.kr

보완이 가능하다는 것이 장점이다.²³ 그리고 이 데이터를 활용한 프로젝트들도 다수 보인다.²⁴ 다만, 이 프로젝트의 관리가 불특정 다수에 의해서 진행되고 있다는 점은 긍정적인 요소인 동시에 부정적인 요소가 될 수도 있다. 히브리성서 데이터는 결국 정확하고 충분한 언어 정보가 매우 중요하다고 볼 수 있다. 이 데이터는 비교적 최근에 완성되었기 때문에 지속적인 데이터 검증이 필요하며, 구문 속성 등과 같은 보다 다양한 언어 정보가 보완될 필요가 있다.

두 번째로 살펴볼 히브리어 성서 형태소 데이터베이스는 앞서 언급한 웨스트민스터 신학 대학의 그로브스 센터에서 진행하는 두 번째 프로젝트로, WLC 원문 데이터를 기반으로 하여, 형태소 정보가 추가된 데이터베이스인 “Westminster Hebrew Morphology”(WHM)이다. WHM은 성서 연구자들에게 이미 익숙한 바이블웍스(BibleWorks)와²⁵ 어코던스바이블(Accordance Bible Software), 그리고 로고스 바이블(Logos Bible) 등의 전문 성서 소프트웨어를 통해 제공되어 왔다. WHM의 기본적인 데이터 구조와 표현 방식은 <그림 3>과 같다.²⁶

23 이러한 방식은 소위 오픈 소스(open source) 생태계에서 빈번하게 활용되는 방식이다. 물론 이 점이 장점으로 작용하려면 철저한 검증 체계와 전문화된 자원봉사자들이 함께 모여 커뮤니티를 지속적으로 이루어가는 것이 중요하다. 그렇지 않으면 오히려 데이터가 훼손될 위험성도 있다. 이런 작업 방식과 관련하여 다음의 글을 참고하라. Jesse Griffin, “Morphology For the Masses by the Masses”, <https://etcbc.nl/methodology/morphology-for-the-masses-by-the-masses/> (2023.4.14. 접속).

24 Jesse Griffin, “Morphology for the Masses by the Masses.”

25 바이블웍스는 지난 2018년 6월 15일, 버전 10을 끝으로 종료되어 더 이상 새로운 버전이 개발되지 않고 있다. 이 프로그램의 개발이 중단된 데에는 불법 복제 문제와 개발 인력이 충원되지 못했던 문제가 주요하게 작용했던 것으로 알려져 있다. 이 문제는 앞으로 성서 소프트웨어가 어떻게 관리되어야 하는가에 대한 방향성에 중요한 참고가 될 수 있을 것이다. 이 문제는 아래에서 다시 언급할 것이다.

26 WHM 형태소 정보와 OSHB의 형태소 정보 표기 방식을 비교해 보면 상당 부분 그 구조가 유사하다는 것을 발견하게 된다. 나중에 만들어진 OSHB가 WHM의 형태소 표기 방식을 상당부분 참고한 것으로 여겨진다. WHM은 OSHB와는 달리 정형화된 틀을 제시하고 있다. OSHB는 결여된 요소는 완전히 생략하고 있는데 WHM은 x 기호로 표기

(@Pp+SxxxExHxNxxRx) ק +(@ncfsa+SxxxExHxNxxRx) תִּשְׁאֵל
 (@vqp3ms+SxxxxJxCxExHaNxxRx) שָׁמַר
 (@ncmpa+SxxxExHxNxxRx) מִתְּחִלָּה
 (@Po+SxxxExHaNxxRx) תִּשְׁאֵל
 (@Pa+SxxxExHxNxxRx) ק +(@ncmpa+SxxxExHxNxxRx) מִתְּחִלָּה
 (@Pc+SxxxExHxNxxRx) ק +(@Po+SxxxExHaNxxRx) תִּשְׁאֵל
 (@Pa+SxxxExHxNxxRx) ק +(@ncfsa+SxxxExHxNxxRx) תִּשְׁאֵל

그림 3. WHM의 형태소 데이터 구조

위 구조에서 볼 수 있듯이 WHM은 전치사나 접속사와 같이 독립적인 문법 요소이면서, 형태적으로 다음의 단어와 결합하는 접사들(enclitic)을 완전히 구별하여 분리시키고 있다. OSHB는 빗금 기호(/)로 구분하고 있기는 하지만 위와 같이 완전히 분리된 구조보다는 구문 검색 활용에 있어 효율성이 떨어질 수 있다. 그리고 데이터 구조에 있어 WHM의 중요한 특징은 히브리어 명사와 동사의 경우 본문 형태가 아닌 ‘원형’을 입력해 놓았다는 점이다. 본문 상에 나타난 히브리어 단어의 변화형이 다양하므로 특정한 단어의 용례를 검색하기 위해서는 각 단어의 기준이 되는 ‘원형’ 정보가 매우 중요하다. OSHB의 경우에는 원형 정보를 스트롱 번호가 대신하고 있는데, ‘자료의 직관성’이라는 측면에서의 한계가 있다. 그리고 스트롱코드는 만들어진 지 100년이 훌쩍 넘은 구체제라 할 수 있는데, 이 체제를 도입하지 않는 사전 자료와의 연결성이 어려울 수 있다는 문제점이 있다. 이 점에서 원형 데이터를

한다. 이러한 자료 구조는 결여된 형태소까지 고려하여 검색 조건에 추가할 수 있는 보다 발전된 검색에 용이한 구조라 할 수 있다. 형태소 정보에 대한 약어는 다음의 자료에서 찾아볼 수 있다. “The WTM Hebrew Morphological Coding Scheme”, <http://www.clavmon.cz/ultranet/bw/bwCodingHeb.pdf> (2023.4.14. 접속).

www.kci.go.kr

제공하고 있는 WHM은 검색 도구를 위한 기반 데이터로서의 강점이 있다.

이 프로젝트는 웨스트민스터 신학 대학 산하의 연구소를 중심으로 오랫동안 관리되어 왔기 때문에 정보의 정확성이나 안정성이 보장되어 있다고 볼 수 있으며, 기존의 전문 성서 프로그램을 통해 비교적 어렵지 않게 본 데이터에 접근하고 활용할 수 있다는 이점이 있다. 그러나 이 프로젝트에는 한계점도 있다. WHM 데이터는 기본적으로 아스키 파일(단순 텍스트 파일 포맷)로 만들어져 있다.²⁷ 이는 데이터 포맷 상으로 XML과 같이 표준화된 형식을 갖추지 않았음을 의미하는 것이다. 그리고 이쉽게도 본 데이터는 누구나 사용할 수 있는 형태로 공개되어 있지 않고,²⁸ 바이블웍스, 어코던스, 그리고 로고스바이블 등과 같은 상용 성서 소프트웨어를 통해 사용할 수 있다.

3) 원문 + 형태소 + 구문 데이터베이스 (Text + Morphology + Syntax Database)

위에서 디지털 성서학 연구를 위해 활용될 수 있는 성서 본문과 형태소 데이터베이스에 대해 살펴보았다. 그런데 보다 체계적이고 심화된 히브리어 본문 분석을 위해서는 구문 정보(syntax)가 매우 중요하다. 성서 본문의 형태소 정보가 개별 단어의 어형과 품사 등과 같은 단어 단위의 정보를 담고 있다면, 구문 데이터는 개별 단어들이 조합되어 형성된

27 "A Reference Guide to the Westminster Hebrew Morphology Database", https://library.mibckerala.org/lms_frame/eBook/MORPHmanual.pdf (2023.4.14. 접속)

28 그로브스(Groves) 연구소 홈페이지에 등록된 2019년의 글에 따르면 연구소에서 만든 데이터를 공유가 가능한 라이선스(CC BY-NC-ND 4.0)로 배포할 예정이라고 했지만, 사용자가 자유롭게 사용할 수 있는 데이터는 제공되어 있지 않다. "The Groves Center will be at the Society of Biblical Literature 2019", <https://www.grovescenter.org/2019/06/18/the-groves-center-will-be-at-the-society-of-biblical-literature-2019/> (2023.4.14. 접속)

더 큰 문법 단위인 구(phrase), 절(clause), 그리고 문장(sentence) 등의 ‘기능’과 ‘형태’에 대한 정보를 제공한다. 따라서 구문 데이터베이스를 통해 특정한 단어나 구가 어떠한 성격의 문장에서 ‘주어’, ‘서술어’, ‘목적어’ 등의 역할을 하는지 볼 수 있고, 특정한 범위 내에서 동사 문장(verbal sentence)이나 명사 문장(nominal sentence) 등과 같은 주요한 패턴의 히브리어 문장들이 어떤 비율과 형태로 나타나는지 등을 파악할 수 있다.

구문 데이터베이스에 있어 가장 핵심이 되는 것은 구문 구조를 파악하기 위한 구문 트리(parse tree)이다. 구문 트리는 문장의 구조와 개별 문법 요소들의 관계를 시각화하는 계층적인 트리 형태의 도식인데, 문장 내에 속해 있는 단어와 구, 절 등의 구성 요소를 노드(node, ‘접속점’)로 표현하고, 이들 사이의 문법적인 관계를 선으로 연결하여 나타낸다. 이러한 구문 트리는 문장을 이루는 각 요소의 계층구조와 상호 관계를 분석하고 시각화함으로써 문법 규칙이나, 특정한 의미를 도출해내는 언어적인 원리를 파악하는 데 도움을 준다. 그런데 이러한 구문 트리는 기반이 되는 구문 이론에 따라 다른 방식으로 표현될 수 있다. 이러한 표현 방식상의 차이는 아래에 소개할 몇 가지 데이터베이스의 구문 트리의 예를 통해 보게 될 것이다. 이처럼 특정한 구문 데이터베이스가 포함하고 있는 구문 정보와 구문 트리 체계는 고유한 형식 구문 이론에 기인한 것이기 때문에, 데이터베이스가 기반하고 있는 이론을 이해하고 접근하는 것이 중요하다.²⁹

29 형식 구문 이론에 대한 개론적인 이해를 위해 참고해 볼 만한 개론적인 자료들은 다음과 같다. Dominique Sportiche, Hilda Koopman, and Edward Stabler, *An Introduction to Syntactic Analysis and Theory* (Malden: Wiley-Blackwell, 2014); Maggie Tallerman, *Understanding Syntax* (Abingdon: Routledge, 2015); David Adger, *Core Syntax: A Minimalist Approach* (Oxford: Oxford Univ. Press, 2003); 아래에 소개할 몇 가지 성서 데이터베이스는 대체로 노암 촘스키(Noaam Chomsky)의 생성 문법(Generative Grammar) 이론에 근거하고 있다. 이 이론은 무한한 언어 표현을 생성해 내는 인간의 ‘보편적이고 공통된’ 형식 문법이 존재한다는 가정을 토대로 언어 구조를 분석한다. 생성 문법에 기

구약성서 히브리어 성서 본문의 형태소 데이터베이스 구축 작업은 꽤 오랜 역사를 갖고 있지만, 그에 비해 구문 데이터베이스가 본격적으로 만들어지기 시작한 것은 비교적 최근의 일이다. 이제 소개할 히브리어 성서 본문의 구문 데이터베이스는 히브리어 본문의 구문 기능에 대한 정보를 담고 있을 뿐만 아니라, 문장을 구성하는 문법적 요소들의 계층 관계에 대한 정보도 담고 있어, 구문 트리 시각화가 가능하다.

먼저 소개할 구문 정보(syntax) 데이터베이스는 “Andersen-Forbes Phrase Marker Analysis”(A-F)이다.³⁰ 1970년대부터 히브리 성서 구문 데이터를 만들기 시작한 프랭크 앤더슨(Frank Anderson)과 딘 포브스(Dean Forbes)는 각기 과학과 물리학 학위를 취득하였으며, 또한 고대 근동학과 성서 언어에 관한 관심을 두고 오랫동안 연구를 진행해 왔다. 특별히 이들은 성서 언어에 대한 연구 과정을 통해 형성한 구조적인 구문 이론을 적용한 데이터를 만들었으며, 자신들의 고유한 구문론을 적용한 히

반한 형식 구문 분석은 단순한 선형이 아닌, 상하 계층구조로 표현되어 문장을 이루는 각 구성 요소 간에 나타나는 의존 구조를 통해 표현된다. 이를 통해 문장의 복잡도에 상관없이 모든 문장이 보편적인 원칙에 의해 구성되어 있음을 보여 주는 동시에, 언어의 구조적인 변환이 언어 표현상의 다양성을 나타내고 있음을 보여 준다. 다음을 참고하라. Noam Chomsky, *Syntactic Structures* (The Hague: Mouton, 1957); *Aspects of the Theory of Syntax*. (Cambridge: MIT Press, 1965); *The Minimalist Program*. (Cambridge: MIT Press, 1995). 이러한 이론에 근거하여 구문 구조를 시각화하여, 각 문장의 구조를 분석함으로써 표현상의 특징이나, 전체적인 문체, 그리고 특정한 의미를 도출하기 위해 어떤 방식으로 문장을 구성하는지 살펴볼 수 있다. 성서 히브리어와 아람어에 있어 다양한 언어 이론의 배경에 대한 개괄이 다음의 저서에 잘 소개되어 있다. 벤자민 J. 누난, 『성서 히브리어와 아람어 연구』 (신철호 옮김), (서울: 감은사, 2022), 47-78, 원제는 Benjamin J. Noonan, *Advances in the Study of Biblical Hebrew and Aramaic* (Grand Rapids: Zondervan Academic, 2020).

30 이 데이터는 현재 로고스바이블과 어코던스에서 사용할 수 있다. 이 데이터 프로젝트에 관한 웹페이지 주소는 다음과 같다. “Anderson-Forbes Resources”, <http://www.andersen-forbes.org/> (2023.4.14. 접속). 그러나 현재는 정보가 업데이트되고 있지는 않고 있다; 이 데이터베이스에 대한 개괄적 소개는 다음을 참고하라. Dean A. Forbes/Francis I. Anderson, “The Andersen-Forbes Computational Analysis of Biblical Hebrew Grammar”, *Journal for Semitics* 27/1 (2018) <https://doi.org/10.25159/1013-8471/2936>.

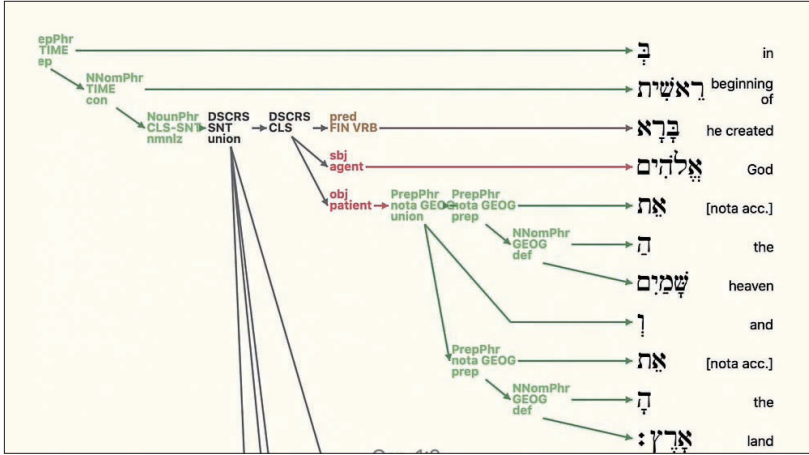


그림 4. A-F 구문트리(Accordance 캡처)

브리어 문법서를 출간하기도 했다.³¹ 이 책에서 제시하고 있는 구문 트리의 예를 제시하면 <그림 4>와 같다. 이 도식에서 볼 수 있듯이 이들은 히브리어 구문 이해에 있어 문장을 구성하는 각 구문의 의미와 기능을 중요시하고 있기 때문에, 구문 단위의 기능(이를테면, 시간 표시 구문, 위치 표시 구문)을 세부적으로 정의하고 있다.

그리고 세 번째로 대표적인 생성문법론자들인 로버트 홀스테드(Robert D. Holmstedt)와 마틴 아백(Martin G. Abegg Jr.) 등은 “Accordance Hebrew Syntax Database”(AHSD)라는 데이터베이스를 새롭게 만들어 출시했다.³² 홀스테드 등은 이 데이터가 촘스키의 생성 문법 이론에 기

31 이들은 자신들을 “생성문법론자”(generativists)라 명명하지만, 촘스키의 핵심 문법 이론인 “보편 문법”이나 “변형 이론”보다는 말뭉치 언어학(corpus linguistics)의 영향을 많이 받았음을 말하고 있다. 이들은 심층적인 의미와 질적인 차원보다는 표면적이고 양적인 차원, 그리고 형식보다는 기능주의 측면에서의 문법을 중요하게 여기고 있음을 밝히고 있다. Francis I. Andersen/A. Dean Forbes, *Biblical Hebrew Grammar Visualized* (Winona Lake: Eisenbrauns, 2012), 5.

32 이 데이터베이스에 대한 개괄적 소개는 다음을 참고하라. Robert D. Holmstedt/John

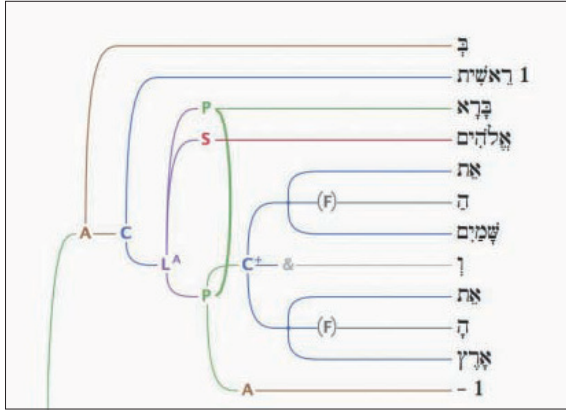


그림 5. AHSD 구문트리(Accordance 캡처)

초한 구문 이론을 바탕으로 하고 있음을 밝히고 있다.³³ 이들은 A-F 데이터베이스가 구문의 기능을 한정적으로 정의하는 것에 반대하며 ‘순수한’ 형식 구문 정보만을 담아야 한다는 의견을 피력하고 있다.³⁴ 그래서 <그림 5>와 같이 A-F 데이터베이스보다는 간소화된 형태의 구문 트리 구조를 볼 수 있다.

위에서 소개한 데이터 중, A-F 데이터베이스는 로고스와 어코던스 소프트웨어를 통해 사용 가능하고, AHSD는 오직 어코던스를 통해

A. Cook, “The Accordance Hebrew Syntactic Database Project”, *Journal for Semitics* 27/1 <https://doi.org/10.25159/1013-8471/3010>.

33 이들의 이론과 관련하여 다음을 참고하라. 벤자민 J. 누난, 『성서 히브리어와 아람어 연구』, 64-67. 이들은 보다 ‘순수한’ 생성문법론자들로 A-F 데이터베이스가 다수 수정된 구문 이해를 가지면서도 스스로를 생성문법론자들로 칭하는 것에 대한 비판적인 입장을 지니고 있다. Robert D. Holmstedt/John A. Cook, Martin G. Abegg/Roy B. Brown, “A Brief User’s Guide for the Accordance Hebrew Syntax Database”, https://www.accordancebible.com/wp-content/uploads/2020/10/BriefUserGuideHebrewSyntaxDatabase_2016.pdf (2023.4.14. 접속), 3의 각주 3번. 이 가이드 문서는 AHSD 데이터베이스의 이론적 배경을 잘 설명해 주고 있다.

34 Robert D. Holmstedt/John A. Cook, Martin G. Abegg/Roy B. Brown, “A Brief User’s Guide for the Accordance Hebrew Syntax Database”, 2.

서만 사용할 수 있다. 이미 다양한 기능들이 갖추어진 상용 소프트웨어를 통해 이 데이터들을 활용할 수 있다는 것은 장점이 되기도 하지만, 활용성이나 확장성 면에서는 제한될 수 있다. 다시 말해, 이 데이터베이스를 토대로 새로운 연구 모델이나 앱을 만드는 것이 불가능하다. 설명 재가공이 가능한 포맷으로 변환하여 데이터베이스 내의 데이터를 출력(export)해서 사용하는 방법도 있기는 하겠지만 상업용 라이선스 제한으로 인한 법적 문제가 있을 수 있기 때문에, 이 또한 문제가 된다. 다음에 소개할 두 데이터베이스는 데이터의 신뢰성과 전문성을 담보하면서도 위와 같은 라이선스 문제를 해결할 수 있는 대안적인 모델이 된다.

위에서 언급한 두 개의 상용 데이터베이스보다 자유롭게 활용이 가능한 첫 번째 구문 데이터베이스는 앞에서 언급했던 웨스트민스터 신학 대학 산하 그로브스 연구소에서 만든 구문 데이터베이스인 Westminster Hebrew Syntax(WHS)이다. WHS 데이터는 성서 번역 단체인 “클리어 바이블”(Clear Bible)의 성서 원문 데이터 저장소에 포함되어 있다.³⁵ 현재 상용 소프트웨어를 통해서만 사용할 수 있는 형태소 데이터베이스(WHM)와 달리 WHS 데이터베이스는 오픈 라이선스(CC BY 4.0)로 공개되어 있다.³⁶ 이 데이터는 XML 형식으로 만들어져 있는데, 이를 구문 트리로 시각화하기 위해서는 별도의 파서(parser)를 만들어야 한다. 필자가 파이썬(Python)으로 간단하게 만든 파서를 이용하여 WHS의 창세기 1장 1절의 구문 트리를 시각화한 결과는 <그림 6>과 같다.³⁷

35 “Clear Bible / macula-hebrew (github), <https://github.com/Clear-Bible/macula-hebrew> (2023.4.14. 접속). 이 소스 저장소 안에는 앞서 언급한 OSHB를 자체적으로 보완, 수정한 버전도 업로드되어 있다.

36 XML 포맷으로 만들어진 본 데이터는 다음의 주소에서 접근할 수 있다. <https://github.com/Clear-Bible/macula-hebrew/tree/main/sources/GrovesCenter/nodes> (2023.4.14. 접속) 형태소 정보가 빠져 있는 이 데이터의 정식 명칭은 “The Westminster Hebrew Syntax without Morphology”이다.

37 이 데이터를 시각화하기 위한 코드와 이 코드에 대한 설명은 다음의 주소에서 확인할

Biblia Hebraica Stuttgartensia (4th edition)에 기초하여 암스테르담 자유 대학(Vrije Universiteit Amsterdam)의 부설 연구소인 ‘에프 탈스트라 성서와 컴퓨터 연구소’(Eep Talstra Centre for Bible and Computer)에서 1977년 도부터 프로젝트를 시작하여 지금까지 관리하고 있는 “Biblia Hebraica Stuttgartensia Amstelodamensis”(BHSA) 데이터베이스이다.³⁸ 이 데이터는 지금까지 수차례 보정·보완되어 왔는데, 그 데이터 파일 형태는 MQL, tf 등의 포맷으로 변화해 왔다. 이는 앞에서 언급한 WHS나 OSHB와 같은 XML 표준 포맷이 아닌, ETCBC에서 자체적으로 개발한 소프트웨어를 통해 사용할 수 있는 포맷이다. 이 포맷은 ETCBC에서 개발한 언어 분석 시스템에 최적화된 형태이며,³⁹ 표준 데이터 포맷인 XML이나 CSV⁴⁰ 등으로도 변환하여 출력할 수 있다.⁴¹

이 데이터베이스는 성서 본문, 형태소 자료, 그리고 구문 자료 등이 모두 하나의 데이터베이스 체계 안에 통합되어 있으며 “text-fabric”이라는 언어 분석 소프트웨어⁴²에 최적화된 형태로 구성되어 있다. 이 데

38 “ETCBC”, <https://etcbc.github.io/bhsa/> (2023.4.14. 접속). 이 데이터베이스에 대한 개괄적인 설명은 다음을 참고하라. Wido Van Peursen, “A Computational Approach to Syntactic Diversity in the Hebrew Bible,” 『성경원문연구』 44 (2019), 237-253; Wido Van Peursen/Cody Kingham, “The ETCBC Database of the Hebrew Bible”, 『*Journal for Semitics*』 27/1 (2018) <https://doi.org/10.25159/1013-8471/2974>.

39 MQL은 “EMDROS”, <https://emdros.org/> (2023.4.14. 접속)를 통해, 그리고 tf는 text-fabric이라고 하는 파이썬 패키지(<https://github.com/annotation/text-fabric>)를 통해 사용할 수 있다. MQL 포맷을 이용한 검색 방식의 경우, 다음의 웹사이트를 통해 사용해 볼 수 있다. “SHEBANQ”, <https://shebanq.ancient-data.org/> (2023.4.14. 접속) MQL와 tf 형식의 데이터는 파이프라인(pipeline)으로 불리는 데이터 연결 방식을 통해 상호 변환이 가능하다(<https://github.com/ETCBC/pipeline> (2023.4.14. 접속) 참조).

40 Comma-Separated Values의 약자로 콤마(.) 기호로 구분된 문서 파일을 가리킨다. 이는 엑셀과 같은 스프레드시트의 데이터 구조로 활용할 수 있다.

41 다음을 참고하라. <https://github.com/ETCBC/bhsa/blob/master/programs/bigTables.ipynb> (2023.4.14. 접속). 이 페이지는 tf 데이터를 R과 같은 통계 프로그램에서 사용할 수 있는 포맷으로 출력하는 예를 보여주고 있다.

42 BHSA 데이터베이스 파일 자체는 단순 텍스트 파일(plain text) 포맷으로 되어 있으며 각

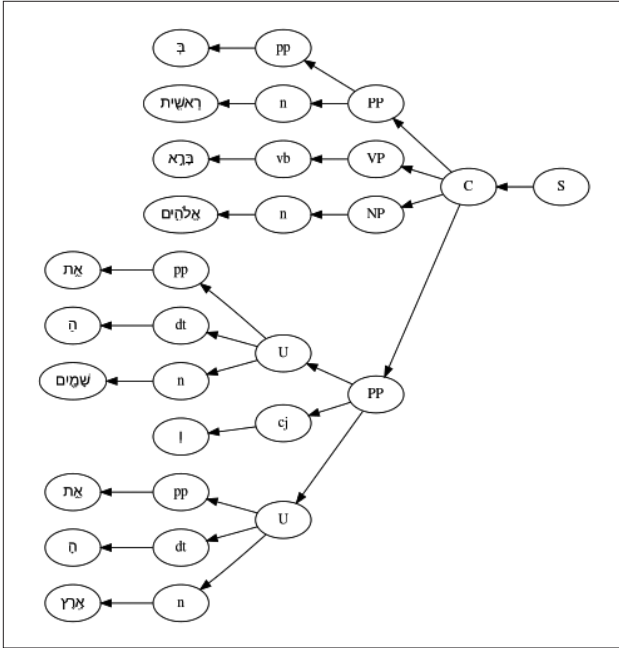


그림 7. BHSA 구문 트리 시각화

이터 방식의 특징은 소위 ‘파이프라인’(pipeline)이라고 불리는 ‘연결과 확장성’에 적절한 체계로 구축되어 있다는 점이다.⁴³ 이러한 시스템하에서 ETCBC 연구소는 성서 데이터베이스뿐만 아니라, 사해 사본, 아

단어, 구, 절, 구절, 장, 책 단위에 노드(node) 번호가 부여되고 있고, 이 노드 번호들을 계층적인 구조로 분류되어 있어, 어떤 단어나 구가 어떤 계층의 위치에 있는지 노드 번호를 통해 확인할 수 있고, 이를 더 손쉽게 검색하게 하는 소프트웨어가 text-fabric이다. 이 소프트웨어는 파이썬 패키지의 형태로 되어 있으며, 오픈소스로 관리되고 있다. <https://github.com/annotation/text-fabric> (2023.4.14. 접속).

43 이에 대해서는 다음을 참고하라. Cody Kingham, “Data Creation”, <https://www.etcbc.nl/datacreation/#intro> (2023.4.14. 접속). 이 페이지는 문헌 데이터 파일이 어떤 과정으로 생성되는지 보여 주고 있다. 이러한 언어 데이터 생성 방식은 언어를 막론하고 모든 문서에 공통적으로 적용될 수 있다는 것이 장점이다.

카드어 문헌 등을 분석할 수 있는 데이터를 구축해 놓고 있다.⁴⁴

이 데이터 역시 구문 정보를 포함하고 있기 때문에 구문 트리를 다양한 방식으로 시각화할 수 있다. 앞의 <그림 7>은 필자가 간단하게 만든 파서와 시각화 도구 패키지를 이용하여 창세기 1장 1절의 BHSA 구문 트리를 시각화한 예이다.⁴⁵ 위 도식에는 드러나 있지는 않지만, A-F 데이터베이스와 같이 구문 기능에 대한 정보도 본 데이터베이스 내에 포함되어 있다. 즉, 자신이 선호하는 구문 이론에 따라 시각화의 형태를 조정할 수 있다는 것이 이 데이터베이스의 장점이라 할 수 있다.

앞서 언급한 바와 같이 각 성서 본문 데이터는 기반이 되는 구문 이론에 근거하면서도, 히브리어 구문 구조의 특성에 따라 일부 수정하여 적용하고 있다. 그렇기 때문에 각 데이터가 어떤 이론에 기반하여 만들어졌는지 이해하고 사용할 필요가 있다.

지금까지 디지털 성서학 연구를 위해 가장 필수적이라 할 수 있는 대표적인 성서 데이터베이스의 종류와 그 성격에 대해 살펴보았다. 이제 우리가 생각해야 할 과제는 이러한 데이터베이스를 어떻게 디지털 성서학 연구에 활용할 수 있는가이다.

3. 성서 데이터를 활용한 디지털 성서학 플랫폼 개발의 방향성

위에서 언급한 데이터베이스를 있는 그대로 사용하는 것은 많은 제약이 따른다. 왜냐하면 데이터베이스의 자료를 연구목적에 맞게 가

44 다양한 문헌 데이터는 다음의 페이지에서 확인할 수 있다. <https://annotation.github.io/text-fabric/tf/about/corpora.html> (2023.4.14. 접속).

45 코드는 다음을 참고하라. <https://github.com/alphalef/digital-biblical-studies/blob/main/notebooks/text-fabric-trees.ipynb> (2023.4.14. 접속).

공하고, 시각화하기 위해서는 어느 정도 프로그래밍에 대한 지식과 기술이 요구되기 때문이다. 따라서 많은 연구자는 상용 성서 소프트웨어를 통해 이러한 데이터를 사용할 수밖에 없는 것이 현실이다. 그러나 여기에는 몇 가지 문제가 있다. 먼저, 성서 본문 데이터의 활용 범위가 상용 소프트웨어에서 제공하는 기능에 한정되어 있다는 한계가 있다. 기존의 성서 소프트웨어는 사용자들로 하여금 보다 쉽게 데이터에 접근하게 하는 편리한 도구가 되기도 하지만, 연구자의 고유한 연구 목적을 위해 데이터를 새로운 방식으로 가공하거나 보완하는 작업 등과 같은 커스터마이징 작업에 커다란 장벽으로 작용하기도 한다. 또 다른 문제는 다소 현실적인 측면의 문제인데 경제적인 부분에서의 진입장벽이다. 위에서 언급한 성서 데이터는 일반적으로 어코던스나 로그스 바이블 등과 같은 상용 소프트웨어에 추가 모듈 형식으로 되어 있어, 이 데이터를 사용하기 위해서는 적지 않은 비용이 소요된다. 이는 모든 학문이 지향해야 하는 형평성과 보편성의 가치를 침해할 수 있는 요인이 된다.

이러한 문제의식을 느끼면서, 앞으로의 디지털 성서학 연구를 위한 디지털 성서학 플랫폼 개발의 방향성에 대해 논의하고자 한다.

1) 사용자 주도의 콘텐츠 생성과 기능성 확장이 가능한 플랫폼 구축

먼저 필자는 ‘플랫폼’으로서의 디지털 성서학 도구 개발의 방향성을 제안한다. 여기서 ‘플랫폼’(platform)이라는 개념을 사용하는 이유는 다음과 같다. ‘소프트웨어’는 특정한 목적을 위해 디자인된 하나의 ‘디지털 완제품’을 의미한다. 사용자는 소프트웨어의 개발자가 이미 결정한 가이드라인에 따라 데이터베이스에 접근하거나 해당 소프트웨어에서 일방적으로 제공하는 도구만을 사용할 수 있을 뿐이다. 반면, 플랫폼은 사용자가 플랫폼상에 구축된 기반 데이터베이스나 기본적인 콘텐츠 생성 도구를 사용하여 직접 콘텐츠를 만들거나, 특정한 연구 목적에 적

www.kci.go.kr

합한 확장 기능을 만들 수 있게 지원하는 기반 시스템이라 할 수 있다. 디지털 인문학의 경우, 인문학자와 디지털 기술 사이에 존재하는 틈은 바로 이러한 플랫폼을 통해 좁혀져야 한다. 그런데 여기서 중요한 것은 사용자의 편의성과 자율성 사이에서 플랫폼으로서의 균형점을 어디에 잡아야 하는가이다. 성서 본문 데이터베이스에 직접 접근하여 완전한 자율성을 보장하기 위해서는 컴퓨터 프로그래밍이나 통계학 등과 같은 전문적인 기술과 지식이 요구되기 때문에, 사용자의 진입장벽이 높아질 수밖에 없다. 반면, 진입장벽을 낮추기 위해 사용자 편의 부분에만 집중한다면 기존의 상용 소프트웨어에서 볼 수 있듯이 사용자의 자율성과 확장성이 다소 제한될 수 있다. 그렇기에 기본적으로 사용자가 데이터베이스에 접근할 수 있게 하면서, 자율성과 확장성이 보장된 시스템을 구축하는 것이 중요하다 할 수 있다.

아래에 실례로 제시할, 두 가지 프로젝트는 공개된 성서 데이터베이스를 통해 구축될 성서 본문 연구와 문헌 분석을 위한 플랫폼 개발의 방향성을 위해 참고할 수 있는 모델이 된다.

연구자가 기존의 데이터베이스를 이용하여 자신의 콘텐츠를 직접 생성·공유하는 데이터베이스의 활용 예는 “세파리아”(Sefaria)에서 찾아볼 수 있다.⁴⁶ 세파리아는 히브리성서와 유대 문헌을 디지털화하고, 각 문헌 자료가 유기적으로 연결된 시스템을 갖춘 유대학 연구 플랫폼이다. 여기에는 사용자에게 ‘시트’(Sheet)라는 기능을 제공하고 있는데, 사용자는 특정 성서 본문이나 주석과 관련된 글을 작성할 수 있고, 이 글은 해당 본문 데이터와 연결성을 갖게 된다. 그리고 이 시트 내에는 기존의 문헌 데이터를 삽입할 수 있는 기능을 제공하여 사용자의 데이터 접근성을 높이고 있다. 그런데 세파리아는 ‘도서관’(library)의 역할을 표

46 “Sefaria”, <https://www.sefaria.org/> (2023.4.14. 접속)

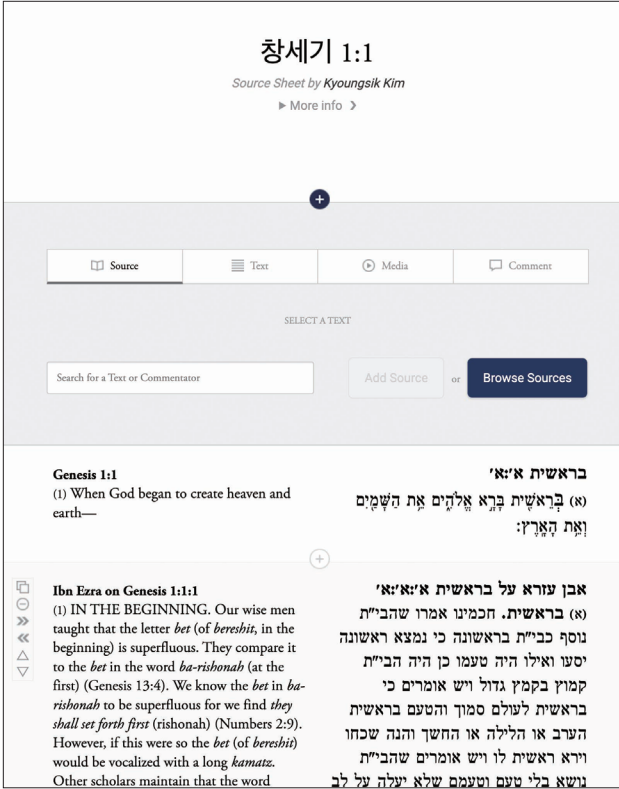


그림 8. 세파리아의 시트

방하고 있는 만큼, 히브리어 본문의 언어적이거나 문법적인 정보에 대해서는 접근할 수는 없다.

우리가 주목해 볼 만한 또 다른 성서 연구 플랫폼은 이스라엘의 바르일란 대학(Bar-Ilan Univ.)의 컴퓨터 공학자들과 성서학자들이 함께 진행하고 있는 “딕타”(Dicta) 프로젝트이다.⁴⁷ 이 웹사이트는 기존

47 이 프로젝트는 바르일란 대학의 컴퓨터 공학 교수 모세 코펠(Moshe Koppel) 박사의 주도하에 이루어지고 있다. 자연언어처리와 다양한 디지털 언어 분석 방법들을 이용하여 다양한 도구들을 개발하고 있는데, 구약 성서의 경우 ETCBC의 BHS 데이터틀을 사용하

의 성서 소프트웨어에서는 사용할 수 없는 실험적인 언어 분석 도구들을 사용해 볼 수 있다. 이 가운데 사용해 볼 만한 흥미로운 도구는 문체 분석기(Stylistic Segmentation)와 티베리아 성서 본문 분류기(Tiberias : Bible Classification)이다. 이 분석 도구들은 자연 언어 처리(Natural Language Processing) 알고리즘을 통해 히브리어 문장의 형태와 문체를 구별해 내어, 특정한 본문 범위 내에 어떻게 서로 다른 문체가 존재하는지, 혹은 서로 다른 본문들 사이의 문체적인 유사성을 도출해낼 수 있는 도구들이다. 이러한 기능은 다분히 ‘실험적’인 성격이 강하기 때문에, 기존의 상용 소프트웨어에 통합되기 위해서는 복잡한 절차와 긴 시간이 요구될 수밖에 없을 것이다. 그러나 자유롭게 사용할 수 있는 데이터베이스에 기반을 둔 연구자 중심의 유연한 플랫폼 체계 내에서는 다양한 분석 도구들이 얼마든지 빠른 시일 내에 개발되고 추가될 수 있다.

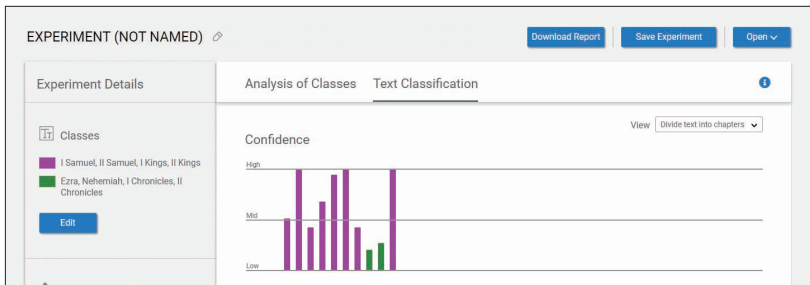


그림 9. 티베리아 성서 본문 분류기는 에스더서 대부분이 사무엘서-열왕기서의 문체와 유사하지만 8-9장은 에스라, 느헤미야, 역대기의 문체와 유사하다는 결과를 보여줌.

고 있다. 이 프로젝트에 대한 소개와 다양한 도구들과 관련하여 다음의 주소를 참고하라. “DICTA”, <https://dicta.org.il/> (2023.4.14. 접속).

www.kci.go.kr

위 두 모델은 성서학 연구 플랫폼 개발에 있어 중요한 힌트가 된다. 세파리아는 사용자들이 능동적으로 참여하여 콘텐츠를 생산하여 공유할 수 있는 콘텐츠 플랫폼으로서의 형태를 보여 주고 있으며, 딕타 프로젝트는 성서 본문 데이터를 활용하여 연구자가 직접 새로운 분석 도구를 개발하고 다른 연구자들과 공유하는 좋은 사례라 할 수 있다.

2) 알고리즘과 소스의 공개

위에서 살펴본 도구들 가운데 딕타는 성서 히브리어 언어/문체 분석의 새로운 장을 연다고 볼 수 있을 정도로 흥미로운 기능을 제시하고 있다. 그러나 여기에는 한계가 있다. 그것은 딕타에서 제공하는 문체 분석의 기능이 어떤 원리와 알고리즘으로 작동하는지 공개되어 있지 않다는 점이다.⁴⁸ 물론, 문체 분석의 상세 결과를 열면 어떤 문법적인 특성이 결과에 우선적으로 영향을 미쳤는지 보여 주고 있기는 하지만 알고리즘 전체를 파악하지 못하기 때문에 학술적인 논의에서 근거 자료로 사용하기가 크게 제한된다고 볼 수 있다.

소스를 공개한다는 것은 프로그램 기법을 공개하는 것 이상의 의미가 있다. 소스 공개는 다른 연구자들도 프로그램 개선에 참여할 수 있음을 나타낸다. 그리고 원 개발자가 더 이상 주도적으로 프로그램을 개발할 수 없는 상황 속에서도 다른 개발자가 이어서 프로그램을 유지하고 개선해 나갈 수 있다는 것이 오픈소스 프로젝트의 강점이다. 수년 전 바이블웍스 개발과 유지가 중단될 수밖에 없었던 데에는 기존 상용 소프트웨어의 폐쇄성의 한계 때문이라고 할 수 있다. 만일 이 프로젝트가

48 모세 코펠 박사는 한 영상에서 본래 딕타 프로젝트를 통해 만들어진 분석 도구들의 소스 코드를 오픈 소스(open source)로 공개할 계획이었지만, 소스 관리의 문제가 복잡해진다는 이유로 쉽지 않은 상황이라 언급한 바가 있다(<https://www.youtube.com/watch?v=IggKr6QWL30>).

오픈소스로 전환될 수 있었다면 틀림없이 누군가에 의해 소스가 관리되고, 지속적인 사용이 가능했을 것이다.

그리고 앞서 구문 데이터베이스와 관련한 논의에서 언급했듯이, 각 데이터베이스는 서로 다른 언어 이론을 배경으로 하고 있다. 따라서 각 연구자가 자신이 생각하는 방식대로 ‘수정 가능’해야 데이터베이스를 사용하는 의미가 있는 것이다. 자신이 지지하는 언어 이론과는 다르고, 이에 따라 수정할 수 없는 데이터베이스의 활용 가치는 그만큼 떨어질 수밖에 없다. 따라서 수정할 수 있는 데이터베이스의 소스 공개가 매우 중요하다. 이러한 차원에서 WHS나 BHSA 같은 공개된 데이터베이스는 그만큼 중요한 가치를 지니고 있다고 볼 수 있다.

따라서 지속 가능하고 신뢰성 있는 성서 연구 플랫폼 개발을 위해서는 플랫폼 그 자체도 오픈 소스의 형태로 관리되는 것이 중요하다.

3) 오픈 플랫폼으로서의 알파알렘 성경 개발의 예

필자는 디지털 성서학 도구를 국내에 도입하기 위한 시도와 연구를 해 왔으며, 기초적인 작업의 결과물인 “알파알렘 성경”을 만들고 공유하고 있다.⁴⁹ 여기에는 위에서 언급한 디지털 성서학 연구 플랫폼으로서의 방향성에 대한 고민이 담겨 있다.

첫째는, 디지털 성서학 플랫폼의 개발을 위해서는 신뢰성 있고, 지속적으로 보완 가능한 히브리어 성서 본문 데이터가 핵심이다. 위에서 언급한 다양한 데이터가 있지만, 알파알렘 성경은 데이터 포맷의 변경이 자유롭고, 필요하다면 데이터를 손쉽게 수정하거나 개선할 수 있는 ETCBC의 BHSA 데이터베이스를 기반으로 하고 있다. 이 데이터는 앞

49 “알파알렘성경”, <https://app.alphalef.com/> (2023.4.12 접속). 알파알렘 성경은 위에서 소개한 ETCBC의 BHSA 데이터에 기반하여 만들어졌다.

서 언급했던 것처럼 30년 이상 검증되어 왔고, 지금도 특정 대학을 초월하여 여러 연구자가 협업하여 관리되고 있기 때문에 데이터베이스의 신뢰성 또한 높은 수준이라 할 수 있다.

둘째로 다양한 콘텐츠와 기능을 추가할 수 있는 확장성 있는 플랫폼의 구조를 구성하는 것이 중요하다. 플랫폼은 고정된 틀로 존재하는 것이 아니라 계속해서 변화하고 확장될 수 있는 유연한 시스템의 구조를 갖추어야 한다. 필자가 지향하는 디지털 성서학 플랫폼의 기본 구조는 <그림 10>과 같다.

디지털 성서학 플랫폼에 있어 문법적인 정보나 의미론 정보 등을 포함한 다양한 메타 데이터가 부가된 성서 본문 데이터를 중심으로 관련 기능과 콘텐츠들이 추가될 수 있는 시스템 개발이 이루어져야 한다. <그림 10>에서 성서 본문 주변에 결합 되어 있는 도구나 자료는 “모듈”이라 지칭된다. 마치 블록을 끼워 넣고 확장하듯이 이러한 모듈을 성서 본문에 연결함으로써 다양하게 본문 데이터를 가공하고 재구성하여 성



그림 10. 디지털 성서학 플랫폼의 구조

확대표시 • 도구모음 • 번역본 •

이전 장 다음 장

창세기 12

지도 위성

Google

*위 지도에 표시된 빨간 점들은 본문 장에 기록되어 있는 지역을 표시한 것입니다. 위 점을 클릭하시면 지명과 인공된 성경구절을 보실 수 있습니다. (번역본 장절 기준)
 *본 장에 별도로 기록된 지명이 없다면 아무런 지역 표시가 되지 않습니다.
 *Copyright ©2017 by Bible Geocoding

.1

C.Wayqtol-K clause

P.분치사구.보어	P.고위명사구.주어	P.동사구.서술어	P.접속구.결어미
אֱלֹהִים	יְהוָה	אָמַר	וַ

C.Zero-Imperativemull clause

P.분치사구.보어	P.분치사구.보어	P.분치사구.보술어	P.동사구.서술어
אֱלֹהִים	מֵאֵל	לֵךְ	לְךָ

C-yiqtol-null clause

P.동사구.목적어.접미어가 있는 서술어	P.접속구.결어미
אֲשֶׁר	אֲשֶׁר

주석 원본식

12:1 여호와께서 아브람에게 이르시되 너는 너의 본토 친척 아버지를 떠나 내가 네게 지시할 땅으로 가라

그림 11. 본문을 중심으로 연결된 지도와 문법/구문 정보

서 본문과 관련된 다양한 자료를 다양하게 시각화 하거나 필요한 형식으로 출력할 수 있다. <그림 11>은 특정한 본문에서 언급되는 지명을 표시하는 지도(지리 정보), 해당 본문을 잘 이해할 수 있게 해 주는 구문 단위 정보, 그리고 개별 단어의 사전적이고 문법적인 정보를 한 번에 출력해 주고 있다. 즉, 이처럼 유기적인 연결성을 제공해 주는 플랫폼을 통해 연구자는 성서 본문에 대한 통전적인 정보를 얻을 수 있다. 물론, 이

제공하는 것이다. 앞의 〈그림 12〉는 학습자가 특정 히브리어 본문을 학습할 때 필요한 내용(본문, 단어 분석, 단어 리스트)을 뽑아서 출력해 주는 기능의 예이다. 이를 응용하여 발전시키면 언어 학습 단계와 필요로 하는 난이도에 따라 본문을 선별해 주거나, 교수자들이나 학습자들 스스로가 원하는 방식의 데이터를 출력할 수 있게 하는 기능을 추가할 수도 있다. 이처럼 주어진 데이터를 재가공하고 새로운 기능을 자유롭게 구현할 수 있는 도구들이 지속적으로 추가될 수 있는 틀과 시스템을 갖춰 나갈 수 있는 개발 방향성이 마련되어야 한다. 알파알렘 성경은 바로 이러한 시스템을 갖출 계획으로 개선되는 가운데 있다.

그리고 디지털 성서학 연구 플랫폼 개발에 있어 중요한 세 번째 방향성은 자유로운 모듈 개발과 확장을 위한 공개적인 시스템 구축이다. 앞서 언급한 바와 같이 디지털 성서학 플랫폼은 오픈소스의 형태로 유지되어야 하고, 새롭게 개발되고 추가되는 모듈의 소스 코드 또한 개방되어, 그 안에서 작동하는 알고리즘과 로직을 파악할 수 있어야 한다. 이렇게 투명하게 공개된 플랫폼 환경은 신뢰성 있고, 지속 가능한 연구 환경을 제공해 줄 수 있을 것이다.

알파알렘 성경은 오픈 소스의 형태로 유지되고 있으며,⁵⁰ 앞으로 협업을 가능케 하기 위한 개발 문서 정보 체계를 마련할 예정이다.

50 알파알렘 성경의 소스코드는 다음의 주소에서 확인할 수 있다. <https://github.com/kungsik/kimsbible> (*초기 개발 이름이 kimsbible이기 때문에 소스 저장소의 이름이 이렇게 되어 있음); 그리고 자유 소프트웨어(free software) 철학에 입각한 개발 방향성에 대해 논의한 성서 애플리케이션 개발과 관련하여 다음의 글을 참고하라. Kyoungsik Kim, "ETCBC Data for the Libre Bible Software", <https://etcbc.nl/uncategorized/etcbc-data-for-the-libre-bible-software/> (2023.4.14. 접속).

4. 결론과 앞으로의 과제

지금까지 디지털 성서학 플랫폼 개발을 위해 필요한 성서 본문 데이터베이스와 이 데이터베이스를 기반으로 한 플랫폼 개발의 방향성에 대해 논의해 보았다. 본 연구는 먼저 디지털 인문학의 방법론을 활용한 성서 연구, 소위 ‘디지털 성서학’ 연구에 있어 가장 중요하고 핵심적인 성서 본문, 형태소, 그리고 구문 데이터베이스의 종류와 특성에 대해 살펴보았다. 특히 BHSA나 OSHB 등과 같이 개별 연구자가 자유롭게 활용할 수 있도록 공개된 성서 데이터베이스는 컴퓨터를 활용한 다양한 방식의 연구를 진행할 수 있게 해주고, 때로 새로운 통찰을 제공하기도 한다. 필자는 이러한 디지털 성서학 분야가 본격적으로 국내에 도입되고, 잘 활용될 수 있게 하기 위해서, 공개된 성서 데이터베이스를 기반으로 한 디지털 성서학 플랫폼의 개발 방향성을 제안하였다. 보다 창의적이고 자유로운 연구를 도울 수 있는 디지털 성서학 플랫폼이 되기 위해서는 연구자 스스로 콘텐츠를 생성하고, 새로운 기능을 추가할 수 있으며, 누구나 새로운 기능 개발에 참여할 수 있게 하는 방향성을 갖고 개발되어야 한다.

‘디지털 성서학’은 아직 국내 성서학계에서는 낯선 분야이다. 그러나 새로운 지식의 발견이라는 학문의 근본적인 목적을 달성하기 위해 반드시 이해해야 할 연구 분야라고 생각한다. 디지털 성서학의 지속적인 발전을 위해 필수적으로 요구되는 가치는 바로 ‘협업’이다. 디지털 성서학 분야가 태동한 이래, 이를 적용한 다양한 연구들이 꾸준히 진행되어 올 수 있었던 것은 수십 년에 걸쳐 만들어진 성서 본문 데이터베이스와 이를 공개하기로 한 연구자들과 개발자들의 결정 덕분이라 할 수 있다. 디지털 성서학이라는 거대한 산에 오르기 위해서는 인문학, 성서학, 그리고 공학 등의 다양한 학문적 배경을 가진 연구자들의 협업

이 있어야 가능하다. 이러한 과제를 갖고 필자는 알파알렙 성경이라는 모델을 만들어 ‘협업’과 ‘공유’의 가치 안에서 이를 발전시켜 나가려고 노력하고 있다.

이 논문에서는 디지털 성서학 연구를 위한 기초적인 개념과 방법론에 대한 논의가 중심이었다면, 다음의 과제는 이를 어떻게 구체적으로 성서학 연구에 접목할 수 있는가에 대한 예시를 구체적으로 보여 주는 것이라 할 수 있겠다. 모쪼록 이 논의가 앞으로의 국내 디지털 성서학 연구에 있어 작은 기여가 될 수 있기를 기대한다.

참고문헌

- 권성달, “성서히브리어 명사문장의 어순에 관한 연구”, 『성경원문연구』 35 (2014), 170-194.
- 김창주, “인공지능과 구약성서: 무슨 상관이 있는가(욥 21:21)”, 『구약논단』 28/4 (2022), 187-214.
- 김현, “디지털 인문학”, 『인문콘텐츠』 29 (2013), 9-26.
- 누난, 벤자민 J., 『성서 히브리어와 아람어 연구』 신철호 옮김, (서울: 감은사, 2022).
원제 Noonan, Benjamin J., *Advances in the Study of Biblical Hebrew and Aramaic* (Grand Rapids: Zondervan Academic, 2020).
- 소형근, “포스트휴머니즘 시대에 인간과 인공지능 이해하기: ‘강한 인공지능’의 등장에 어떻게 대처할 것인가?”, 『구약논단』 28/4 (2022), 247-273.
- 이유미, “유용한 해석적 도구인 ‘비블리카버스’(biblicaverse) 구축을 위한 제언-ID drroma의 아가 연구 가상체험기”, 『구약논단』 28/2 (2022), 324-354.
- 진규상, “Analysis of Participants’ Agent Role in the Two Major Divisions of Leviticus”, 『성서원문연구』 52 (2023), 101-138.
- _____, “Analysis of Participants in the First Major Division of Numbers”, 『구약논집』 22 (2022), 8-59.
- 앨런 튜링, “계산 기계와 지능” (김지홍 옮김), http://aitimes.org/wp-content/uploads/2017/02/Allan_turing_Paper_1950_한국어번역.pdf (2023.4.14. 접속). 원제 Turing, A. M., “Computing Machinery and Intelligence”, *Mind* 59 (1950), 433-460.

- Adger, David. *Core Syntax: A Minimalist Approach* (Oxford: Oxford Univ. Press, 2003).
- Andersen, Francis I./A. Dean Forbes, “The Andersen–Forbes Computational Analysis of Biblical Hebrew Grammar”, *Journal for Semitics* 27/1 (2018) <https://doi.org/10.25159/1013-8471/2936>.
- _____, *Biblical Hebrew Grammar Visualized* (Winona Lake: Eisenbrauns, 2012).
- Busa, Robert, “Index Thomasticus”, <https://www.corpusthomaticum.org/it/index.age> (2023.4.14. 접속).
- Chomsky, Noam, *Syntactic Structures* (The Hague: Mouton, 1957).
- _____, *Aspects of the Theory of Syntax*. (Cambridge: MIT Press, 1965).
- _____, *The Minimalist Program*. (Cambridge: MIT Press, 1995).
- Durusau, Patrick, “Osis Users Manual (OSIS Schema 2.1.1)”, <https://crosswire.org/osis/OSIS%202.1.1%20User%20Manual%20006March2006.pdf> (2023.4.14. 접속).
- Griffin, Jesse, “Morphology for the Masses by the Masses”, <https://etcbc.nl/methodology/morphology-for-the-masses-by-the-masses/> (2023.4.1. 접속).
- Hockey, Susan, “The History of Humanities Computing”, Susan Schreibman/Ray Siemens/John Unsworth(eds.), *A Companion to Digital Humanities* (Malden: Blackwell Publishing, 2004), 3-19.
- Holmstedt, Robert D./John A. Cook, “The Accordance Hebrew Syntactic Database Project”, *Journal for Semitics* 27/1 <https://doi.org/10.25159/1013-8471/3010>.
- Holmstedt, Robert D./John A. Cook, Martin G. Abegg/Roy B. Brown, “A Brief User’s Guide for the Accordance Hebrew Syntax Database”, https://www.accordancebible.com/wp-content/uploads/2020/10/BriefUserGuideHebrewSyntaxDatabase_2016.pdf (2023.4.14. 접속).
- Kim, Kyoungsik, “ETCBC Data for the Libre Bible Software”, <https://etcbc.nl/uncategorized/etcbc-data-for-the-libre-bible-software/> (2023.4.14. 접속).
- Kingham, Cody, Cody Kingham, “Data Creation”, <https://www.etcbc.nl/datacreation/#intro> (2023.4.14. 접속).
- Marienberg-milikowsky, Itay, “It Functions, and That’s (Almost) All’: Tagging the Talmud”, Domenico Fiormonte et al.(eds.), *Global Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2022), 141-150.

- Nilsson, Nils J., *The Quest for Artificial Intelligence: A History of Ideas and Achievements* (Cambridge: Cambridge Univ. Press, 2010).
- Roorda, Dirk, "The Hebrew Bible as Data: Laboratory - Sharing - Experiences", Jan Odijk and Arjan van Hessen(eds.), *CLARIN in the Low Countries* (London: Ubiquity Press, 2017), 217-230.
- Sportiche, Dominique, Hilda Koopman, and Edward Stabler, *An Introduction to Syntactic Analysis and Theory* (Malden: Wiley-Blackwell, 2014).
- Strong, James, *The Exhaustive Concordance of the Bible* (Cincinnati: Jennings & Graham., 1890).
- Tallerman, Maggie. *Understanding Syntax*. (Abingdon: Routledge, 2015).
- Van Peursen, Wido, "A Computational Approach to Syntactic Diversity in the Hebrew Bible", 『성경원문연구』 44 (2019), 237-253.
- Van Peursen, Wido/Cody Kingham, "The ETCBC Database of the Hebrew Bible", *Journal for Semitics* 27/1 (2018) <https://doi.org/10.25159/1013-8471/2974>.
- Winther-Nielsen, Nicolai, "Interactive Tools and Tasks for the Hebrew Bible: From Language Learning to Textual Criticism", *Journal of Data Mining and Digital Humanities*, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages (2017), <https://doi.org/10.46298/jdmdh.4003> (2023.5.20. 접속).
- "XML이란 무엇인가요?" <https://aws.amazon.com/ko/what-is/xml/> (2023.3.23. 접속).
- "Anderson-Forbes Resources", <http://www.andersen-forbes.org/> (2023.4.14. 접속).
- "AntConc", <https://www.laurenceanthony.net/software/antconc/> (2023.4.14. 접속).
- "A Reference Guide to the Westminster Hebrew Morphology Database", https://library.mibckerala.org/lms_frame/eBook/MORPHmanual.pdf (2023.4.14. 접속)
- "Biblia Hebraica Stuttgartensia (BHS)", <https://www.academic-bible.com/en/online-bibles/biblia-hebraica-stuttgartensia-bhs/read-the-bible-text/> (2023.4.14. 접속).
- "ChatGpt", <https://chat.openai.com/> (2023.4.14. 접속).
- "Clear Bible / macula-hebrew (github)", <https://github.com/Clear-Bible/macula-hebrew> (2023.4.14. 접속).
- "DICTA", <https://dicta.org.il/> (2023.4.14. 접속).
- "ETCBC", <https://etcbc.github.io/bhsa/> (2023.4.14. 접속).

“EMDROS”, <https://emdros.org/> (2023.4.14. 접속)

“Hebrew Morphology Codes (The Open Scriptures Hebrew Bible)”, <https://hb.openscriptures.org/parsing/HebrewMorphologyCodes.html> (2023.4.14. 접속).

“J. Alan Groves Center”, <https://students.wts.edu/resources/alangroves/grovesprojects.html> (2023.4.14. 접속); <https://www.grovescenter.org/> (2023.4.14. 접속).

“Open Scriptures (Github)”, <https://github.com/openscriptures/morphhb/tree/master/wlc/> (2023.3.23. 접속).

“Sefaria”, <https://www.sefaria.org/> (2023.4.14. 접속).

“SHEBANQ”, <https://shebanq.ancient-data.org/> (2023.4.14. 접속) .

“Tanach”, <https://tanach.us/> (2023.3.23. 접속).

“The Groves Center will be at the Society of Biblical Literature 2019”, <https://www.grovescenter.org/2019/06/18/the-groves-center-will-be-at-the-society-of-biblical-literature-2019/> (2023.4.14. 접속).

“The Open Scriptures Hebrew Bible”, <https://hb.openscriptures.org/> (2023.4.14. 접속).

“The WTM Hebrew Morphological Coding Scheme”, <http://www.clavmon.cz/ultranet/bw/bwCodingHeb.pdf> (2023.4.14. 접속).

“Voyant Tools”, <https://voyant-tools.org/> (2023.4.14. 접속).

“Westminster Leningrad Codex Hebrew Old Testament”, <https://find.bible/bibles/HEBWLC/> (2023.4.14. 접속).

“What Is Digital Humanities?”, <http://whatisdigitalhumanities.com/> (2023.4.14. 접속).

검색어

디지털인문학, 디지털성서학 연구 플랫폼, 성서데이터베이스, 오픈소스, 알파알렘성경

[ABSTRACT]

Digital Biblical Studies Platform: Current Status of Biblical Database Construction and Directions for Developing a Digital Biblical Studies Research Platform

Kyoungsik Kim

Methodist Theological University

This article investigates the potential pathways for creating a digital research framework aimed at biblical studies, incorporating digital humanities methodologies, an area identified as “digital biblical studies.” The discourse emphasizes two fundamental elements: firstly, pinpointing the pivotal and core biblical databases — related to text, morphology, and syntax — imperative for digital biblical studies; and secondly, scrutinizing the architecture and course of action for fashioning a digital biblical studies platform grounded in these databases. Several academic institutions have created a wealth of biblical databases. Some of these databases can only be accessed through proprietary software (such as Andersen-Forbes Phrase Marker Analysis, Accordance Hebrew Syntax Database, Westminster Hebrew Morphology), whereas others are distributed as open-source tools, accessible freely (examples include Open Scriptures Hebrew Bible, Westminster Hebrew Syntax, Biblia Hebraica Stuttgartensia Amstelodamensis). When considering the design of a versatile and scalable digital biblical studies platform, it is beneficial to employ those open-source databases, as they permit free customization

www.kci.go.kr

and enhancements. Given that every bible database is predicated on its unique linguistic theory, it is vital to select or modify a database that suits one's research trajectory. The intended digital biblical studies platform will be rooted in these bible databases, but it must be designed to be dynamic and open, enabling users to produce their own content and integrate new functionalities.

The digital biblical studies platform should not be the sole product of a single person; rather, it should be a cooperative effort between scholars from various academic domains. With this developmental strategy, my ambition is to design a digital biblical studies platform that can be operational in Korea, leveraging the “AlphaAlef Bible” that is currently being developed.

key words

digital humanities, digital biblical studies platform, biblical database, open source, AlphaAlef Bible

투고일 : 2023년 04월 15일

심사일 : 2023년 04월 30일

게재 확정일 : 2023년 05월 17일

www.kci.go.kr