

Levente Zoltán Király **Parallelizing Bible texts**

Developing the database of the Unified Bible Reader online application¹

Abstract

The Unified Bible Reader (UBR), developed by the Károli Gáspár University of the Reformed Church in Hungary, aims to facilitate the parallel reading of various Bible versions, including original language editions and antique as well as modern translations. This paper outlines the creation of a parallelized database of Bible texts through a four-step process: digitization, data cleaning, integration, and parallelization. The paper discusses the inherent challenges of the digitization phase which involves the conversion of non-digital texts to digital format, primarily using optical character recognition (OCR) technology. Data cleaning focuses on correcting OCR errors and resolving format inconsistencies. Integration, that is, importing text into a unified database, addresses issues with rare cases of text division. Parallelization involves aligning textual units across different versions, accounting for variations in verse numbering and verse boundaries. The development of software tools enabled semi-automatic and manual alignment. The application may serve as a basis for more refined text splitting and the addition of annotations to support advanced textual analysis and translation comparisons. The UBR project marks a significant step towards comprehensive, accessible parallel Bible reading and analysis.

Keywords: Bible, database design, digital humanities, Bible translations, text parallelization

Introduction

The Unified Bible Reader (UBR)² is a project implemented by a research team of the Károli Gáspár University of the Reformed Church in Hungary, under the leadership of Prof. Tibor Fabiny. The project had a twofold goal. On the one hand it aimed to enable the closely parallel reading of different versions (mostly translations) of Bible passages – so-called “original language” editions (modern reconstructions of the antique Hebrew, Aramaic and Greek texts), as well as translations from both antiquity and the modern era. On the other hand the project set out to create a digital interface on which the Bible use of daily devotionals of various Hungarian denominations can be closely compared. The present study is only concerned with the creation of the database of parallelized Bible text versions.

¹ A Hungarian version of this study has been published as (Király 2021).

² Unified Bible Reader, Károli Gáspár University of the Reformed Church in Hungary, ebo.kre.hu (accessed 31 May 2024).

While there exists a plethora of online Bible reading tools, the UBR is unique in multiple ways. First, while keeping in mind the usage of the tool as a free aid of devotional practice,³ the selection of texts put equal emphasis on historical (late medieval and early modern) translations in order to promote scholarly research of translations.⁴

Second, the database has been made so that the parallelization of the Bible text versions maintains the chapter and verse numberings which traditionally differ in most Bible editions to a greater or lesser extent. To achieve this last goal, a system of identification had to be invented on the level of the smallest textual units which makes independent parallelization possible.

The process of building a database of parallelized Bible texts follows steps as:

- digitization,
- data cleaning,
- integration,
- parallelization.

The present study is going to elaborate on these, with examples describing the development of what the current version of the UBR is.

Digitization

The texts relevant to us were almost exclusively created non-digitally; the few exceptions are translations made in the last ten years. Therefore, we must first assure that the non-digital texts are transferred to digital media.

There are basically two approaches to this problem: photographic image scanning and character-based digitization. Since the database of our project has been designed solely for text-based corpora, here we do not deal with the issues of graphical aligning of texts contained in images. Character-based digitization can be achieved by manual typing – a famous example is the 1989 edition of the King James Version (KJV) by the Project Gutenberg (Project Gutenberg 1989), or an unfinished, community-based amateur attempt to digitize the first complete Hungarian Bible translation, the 1590 Vizsoly Bible, and make it available online.⁵ By the advancement of optical character recognition (OCR) technology it became the norm to extract the text automatically from the scanned images of the original, and proofread it.

Although at the beginning of our project it was not entirely known, almost all of the texts we have chosen had already been digitized, and they could be either obtained from publicly available sources being in the public domain (some distributed in other Bible study software), or by agreement with the copyright holders. This however did not bar out errors in the texts.

³ This is the obvious and most popular aim of countless online applications like the Bible Gateway (biblegateway.com), the Bible Hub (biblehub.com), YouVersion (bible.com) and specific-to-universal digital libraries and Bible study tools like the SWORD Project (crosswire.org/sword), the e-Sword (e-sword.net), or the Logos Bible Study Platform which alone now contains over 100,000 theological resources, including almost the entire library of the Christian tradition from antiquity to today enriched by professional tools for linguistic analysis (logos.com).

⁴ Such is the free online Parallel Bible Reader (parallelbible.nytud.hu), developed by the Research Institute for Linguistics at the Hungarian Academy of Science, almost synchronously with the UBR. The Parallel Bible Reader was based on a recently built Old Hungarian Corpus, and its purpose was to further the historical study of the Hungarian language. Also worth mentioning is the paywalled Paratext (paratext.org), a professional software for Bible translation with a strong emphasis on automation features.

⁵ Szabó, Bálint, <https://biblia.szeroczei.hu> (accessed 31 May 2024).

Both digitization and format conversion are the sources of several types of errors which are spread by digital sharing.

1Ki 15:28 Bása így megölte Azának, Júda királyának 3. esztendejében, s helyette ő lett a király.
 1Ki 15:29 Amikor királlyá lett, Jerobeám egész házát kiirtotta, nem hagyott Jerobeám (házából) senkit élve, mind kipusztította az Úr szavai szerint,
 1Ki 15:30 amelyeket szolgálja, a silói Achija által hallatott Jerobeám bűnei miatt, amelyeket elkövetett, s amelyekbe Izraelt is belevitte, haragra ingerelve az Urat, Izrael Istenét.
 1Ki 15:31 Nadab történetének többi részét, amit csak végbevitt, mind följegyezték Izrael királyainak könyvében.
 1Ki 15:32 1Ki 15:33 Azának, Júda királyának 3. esztendejében Achija fia, Bása lett Izrael királya Tircában huszonnégyszáz esztendőre.
 1Ki 15:34 Azt tette, ami gonosznak számít az Úr szemében, mert Jerobeám nyomába lépett, a bűn útjára, ahová az Izraelt is vezette.
 1Ki 16:1 Akkor az Úr ezeket a szavakat intézte Bása ellen Hanani fiához, Jehuhoz:

Figure 1. In a version of the “SZIT” Bible text shared online⁶ the verse in 1Kings 15:32 (which was deliberately left empty by the translators) has been filled with the contents of verse 33 together with its locus marker. The error probably emerged during exporting the corpus, caused by the empty contents of verse 32.

One type of errors emerging during OCR is when the algorithm does not recognize the printed letter, or recognizes it mistakenly. OCR software has usually been equipped with a dictionary to associate the graphic data with meaningful words, but since a machine without artificial intelligence does not understand the text as thoroughly as a human, it may eventually choose the wrong word.

1Pe 2:8 de egyszersmind ütköző s botránykő. Mert azok, amire különben rendeltettek is, nem hallgatva az igére, megütköznek.
 1Pe 2:9 Ti pedig választott nemzetség, királyi papság, szent nemzet, tulajdon nép vagytok, a végre, hogy hirdessétek annak hatalmas dolgait, aki a sötétségből hívott ki titeket az ő csodálatos világosságára,
 1Pe 2:10 "titeket, akik egykoron nem **név**", most azonban „Isten népe”, egykoron „meg nem szánt”, most azonban „megszánt népe” vagytok.”
 1Pe 2:11 Szeretteim, kértem-kérlek, hogy mint jövevények és vándorok e földön, tartózkodjatok a lelket ostromló testi kívánságoktól.
 1Pe 2:12 A pogányok között példás életet éljete, hogy jótetteiteknek láttára, azért dicsőítsék az Istent az ítélet napján, amiért most titeket, mint gonosztevőket megszólalnak.

Figure 2. The word “nép” (‘people’) in 2Peter 2:10 was misrecognized as “név” (‘name’). This example also shows a technical error where the verse got surrounded by quotation marks, and the quotations marks inside the text got doubled. Text: (Masznyi 1917).

Another typical issue is formatting which OCR had trouble recognizing consistently for a long time. Markings of emphasis, splitting the text in paragraphs and longer units, even subtitles and notes can equally be part of a text. Very often it was a deliberate decision of the digitizing project team to ignore these during the recognition process. Often the texts were preserved in

⁶ A former blog of the Protestant Theological Institute of Cluj-Napoca hosted on proteo.cj.edu.ro, accessed 02 January 2021, removed since then.

pure text format, breaking one Bible verse in each line, thus losing all formatting which the text originally had.

With files from earlier times it is usually necessary to convert the data from an obsolete format to one used today. When acquiring texts, one encounters a wide variety of file formats from HTML and MS Word to pure text files with different encodings. Integrating them in a unified database can be time consuming, as each data source may have even left its peculiar technical notation in the data. These may cause conversion errors, but even parts of a text may be lost in this phase.

Data cleaning

Semi-automatic correction of the OCR errors is usually possible, i. e. finding meaningless words with the aid of grammar checking tools. Such programs will not identify and correct meaningful but incorrectly recognized words though, neither formatting errors. This can only be done via human labour.

Correction of some types of errors emerging from data format conversion can be automated. Files originating from different or obsolete operating systems may be coded in a peculiar “code page”. E.g. Hungarian texts in the 1990’s used ‘ő’ and ‘ű’ for ‘ó’ and ‘ü’, letters which were not represented in the extended ANSI character set of the time. Converting the text from a different code page is a simple task, but requires human insight and decision.

1Pe 2:20 " ember vallásos meggyőződéséért ártatlanul szenvedve, tûri a bántalmat. Hiszen mi dicsőség, ha vétkeitek miatt veretve tûrtök? Ha jótettek miatt szenvedve tûrtök, az kedves Istennél;"

*Figure 3. Hungarian text showing in Western (Latin-1 or ISO-8859-1) code page
(Hungarian New Testament by Endre Masznyik, 1 Peter 2:20)*

Digital representation of texts may cause other types of problems, too. Exporting a text from a database, or converting a text to a text-based format may cause ambiguity when meta-characters (technical markers of the encoding) occur inside the text, too. An improper conversion includes these meta-characters in the text itself. Filtering and eliminating these can often be automated, but may eventually cause further errors. E.g., a machine may not be able to decide if quotation marks denoted the technical boundaries of a textual unit, or it is the part of the text itself.

chapter	verse	subverse	text	refere
1	20	0	Miközben ezen gondolkodott, álmában megjelent az Úr angyala és így szólt: \"József, Dávid fi...	NULL
1	21	0	Fiút fog szülni, akit te Jézusnak nevezel, mert ő szabadítja meg népét bűneitől.\"	NULL
1	23	0	\"Íme a szűz méhében fogan és fiút szül, És Emmanuelnek fogják hívni.\" Ez azt jelenti: Velün...	NULL
2	2	0	\"Hol van a zsidók Újszülött királya? Láttuk csillagát napkeleten és eljöttünk, hogy hódoljunk n...	NULL
2	5	0	Azok azt válaszolták: \"A judeai Betlehemben, mert így jövendölte a próféta:	NULL
2	6	0	Te Betlehem Júda földje, Éppen nem vagy a legkisebb Júda fejedelmi városai között, Mert bel...	NULL

Figure 4. Querying text units that contain an unnecessary backslash before quotation marks

Integration

Integration in this concept means loading, or importing, the text into the unified database. Normally this is a simple one-step operation, but this may also shed light on hitherto unknown errors.

Sometimes this step makes it evident that the database was not prepared for a set of rare cases in the data. Databases often impose necessary limitations on the stored data for technical reasons. Such is the length of one textual unit. Setting this limit to 1024 (2^{10} chosen randomly with the conjecture that it is a sufficiently large number) characters may initially seem sufficient for a database of Bible verses, yet occasionally there may come a text which, on an unusual place contains a textual unit that is longer. If changing the text is not an option, then the database settings must be adjusted.

In some cases it only becomes evident during importing that the file employs a different inner logic than the rest of the database. Such was the case of importing the book of Esther in a Hungarian Catholic translation (Szent István Társulat 1973) that includes the deuterocanonical extensions organized each as if it were one huge Bible verse. This made it necessary to edit the file manually before importing, to split the extensions into sub-verses.

Parallelization

Having imported the selected texts into the database, aligning follows, so that each unit of one text may correspond to the unit of the other with the same content. At first glance this seems an easy task, since the text of the Bible has been divided into numbered chapters and verses since the Middle Ages to make referring to any specific locus possible. However, the automatization of this aligning is met with a number of various obstacles.

Different content

Automatized parallelization must be prepared to handle Bibles with different corpora of texts. Some Bibles are “complete”, some only contain the Old Testament (the “Hebrew Bible”), some only the New Testament. Although we did not plan to integrate Bibles with a different selection of biblical books, we may mention such containing e.g. “The New Testament and the Psalms” etc. Catholic Bibles contain the so-called deuterocanonical Old Testament books, which are usually omitted in Protestant Bibles.

Similarly we can find differences on the smaller scale. Deuterocanonical materials in the book of Daniel take the form of extra chapters, while in the book of Esther, extra verses. Both of these are presented as separate books in the 1545 Lutherbibel. Critical editions are missing selected verses where the editors decided that a verse of the received text should not be part of the main text. Translations made from such critical editions inherit these omissions, too.

Slipping verse numbers

Although the chapter and verse numbering is intended to be universal, there are multiple cases for alternate verse numberings, usually emerging at alternate chapter divisions. Among popular Hungarian translations the “revised Károli” (Károli 1908) and the “new Protestant translation” (NP) (Magyar Bibliatanács 1974) differ on more than 50 places in verse numbering. The

numbering usually slips only 1 verse, but there is an example of a text block being numbered 8 verses away from the other translation.

There are even bigger differences between other Bible versions. A well-known problem is the numbering of Psalms, which, in Catholic Bibles, traditionally followed the numbering of the Vulgate (Deutsche Bibelgesellschaft 2007), splitting some psalms, joining some others as compared to the Biblica Hebraica and the Protestant translations made therefrom. E.g. Psalm 9:39 in the Vulgate is equivalent to Psalm 10:18 in the NP.

Parallelizing such loci requires a new, unified numbering which is independent of any chapter and verse numbering versions. Obviously this technical numbering needs not to be visible to the end users of the application. The algorithm starts by numbering the (sub)verses from 1 on in the order of the integration into the database, and when it meets a verse that has already been integrated in a different version, it applies its unified identifier instead of giving it a new id.

Different verse boundaries

In Bible verse numbering not only chapter boundaries, but verse boundaries may change. In simpler cases a section is treated as one verse in one edition, while another splits it in multiple verses. The revision of the NP, called the “revised new Protestant translation” (RNP) (Magyar Bibliatársulat 2014) created 10 such differences between the old and the new versions. These all occur at the end of a chapter, so they do not cause verse numbers to slip on long sections.

However, there are more sophisticated differences between numberings. Hebrew poetry has an inherent structure that numbering usually follows, but sometimes digresses from. Thus occasionally we even see differences in Psalm verse divisions in a “brick pattern”, as on Figure 5.

RNP Psalm 73

21 Ha keseregne a szívem,
és sajogna a bensőm,
22 akkor olyan ostoba és tudatlan volnék
veled szemben, mint az állat.
23 De én mindig veled leszek,
mert te fogod jobb kezemet.
24 Tanácsoddal vezetsz engem,
és végül dicsőségbe fogadsz.

VUL Ps 72

21 quia contractum est cor meum
et lumbi mei velut ignis fumigans
22 et ego insipiens et nescius
23 quasi iumentum factus sum apud te
et eram semper tecum
24 et tenebas manum dexteram meam
in consilium tuum deduces me
et postea in gloria suscipies me

Figure 5. “Brick pattern” in psalm verse divisions

As we can see in this example, aligning the textual units of these Bibles is still impossible by creating a universal verse numbering. It is also necessary to split all overlapping units along their intersection boundaries into smaller units, so that the resulting sub-verses might be parallelized without ambiguity. In the above example, the verses of the RNP and VUL have to be split so that pairing 22a = 22, 22b = 23a, 23a = 23b, 23b = 24a, 24 = 24b would be possible.

Speeding up the work with software

The splitting process illustrated above is extremely tedious, and the amount of work grows in quadratic proportion to the number of texts integrated into the database. Going for an automatized pairing of equivalent textual units, we have equipped the numbering algorithm with an application that analyzes the text to be imported, and makes it possible to fine-tune the alignment on an intuitive graphical interface.

When the first text is imported into the database, there is nothing to parallelize it with, therefore the Bible verses of this text simply receive an automatically incremented number. In every subsequent text only those verses receive a new identifier which have no previous text to parallelize with. Thus, having N texts in the database, when text $N + 1$ is loaded into the application, it is aligned to a specific text already integrated into the database (text $k \in [1 .. N]$). Text k must be chosen so that it is the closest variant to new $N + 1$, in order to avoid textual units not present in text k to be numbered as previously unknown. On the illustrations text $N + 1$ can be seen waiting to be aligned on the left side, and text k on the right side.

In case of an extra verse or a missing verse the alignment can be slipped to keep the contents parallel.

12	15	o	A kígyó torkából olyan vízsz	=	30977	12	15	o	A kígyó pedig a szájából tor
12	16	o	De a föld megsegítette az a	=	30978	12	16	o	De segített a föld az asszon
12	17	o	Erre a sárkány haragra lobb	=	30979	12	17	o	Megharagudott a sárkány a
				>	30980	12	18	o	és odaállt a tenger fővenyé
13	1	o	Akkor láttam, hogy egy vad	=	30981	13	1	o	°És láttam, hogy a tengerbő
13	2	o	Ez a vadállat, melyet láttam	=	30982	13	2	o	Ez a fenevad, amelyet láttam
13	3	o	(Úgy tűnt föl), mintha egyik	=	30983	13	3	o	Láttam azt is, hogy az egyik
13	4	o	Az egész föld csodálattal n	=	30984	13	4	o	imádták a sárkányt, mivel á
13	5	o	Nagyvadász és kéremlő ez a	=	30985	13	5	o	És adott neki nagyokat m

Figure 6. In the translation BD (left) the textual unit that would correspond to Revelations 12:18 in the NP (right) is missing, therefore the text must be slipped one unit to align correctly

Different numberings must be acknowledged by a human, since the algorithm is unable to decide what kind of action is required when the numbers differ. Cases of different numberings may mean slipping verse numbers (as above), acknowledging different chapter numbering, or splitting text units to make aligning them possible.

19	39	o	Ha meg más valami vitás ke	=	27696	19	39	o	Ha pedig ezenkívül van vala		
19	40	o	Még abba a veszélybe kerü	=	27697	19	40	o	Mert így is az a veszély fen		
✓	✓	19	41	o	Beszéde végeztével pedig t	≠	35887	19	40	1	Ezeket mondva feloszlatta a
20	1	o	A fölfordulás megszűntével	=	27698	20	1	o	°Miután megszűnt a zavarg		
20	2	o	Végigjárta ennek a kerülete	=	27699	20	2	o	Bejárta annak vidékeit, szár		
20	3	o	Mikor ezután Szíriába készü	=	27700	20	3	o	Három hónapig maradt ott		

19	39	o	Ha meg más valami vitás ke	=	27696	19	39	o	Ha pedig ezenkívül van vala
19	40	o	Még abba a veszélybe kerü	=	27697	19	40	o	Mert így is az a veszély fen
19	41	o	Beszéde végeztével pedig t	≠	35887	19	40	1	Ezeket mondva feloszlatta a
20	1	o	A fölfordulás megszűntével	=	27698	20	1	o	°Miután megszűnt a zavarg
20	2	o	Végigjárta ennek a kerülete	=	27699	20	2	o	Bejárta annak vidékeit, szár
20	3	o	Mikor ezután Szíriába készü	=	27700	20	3	o	Három hónapig maradt ott

Figure 7. In the translation BD (left) Acts 19:41 corresponds to the second half of verse 40 in NP. To align these texts, verse 40 in NP must be split into subverses 40/0 and 40/1. Then the difference in the numbering of BD Acts 19:41 and NP Acts 19:40/1 must be manually acknowledged by the operator of the software. The unresolved difference is highlighted by a reddish background as seen on the first image; the acknowledgment of a difference is marked by dark green as seen on the second.

It should be possible to split a verse in both texts k and $N + 1$.

6	50	o	Itt a mennyből alászállott kenyér, hogy aki eszik belőle, meg r	=	26379	6	50	o	De ez az a kenyér, amely a mennyből szállt le, hogy aki eszik t
6	51	o	Én vagyok a mennyből alászállott élő kenyér. Aki e kenyérből	=	26380	6	51	o	Én vagyok az az élő kenyér, amely a mennyből szállt le:

Eltolás 1
Felosztás
Explicit pár
Eltolás 1

SplitTextUnit

Szövegegység felosztása

Felosztás!
Mégsem

bd1951 J 6, 51 /0

Én vagyok a mennyből alászállott élő kenyér. Aki e kenyérből eszik, örökké él. A kenyér, melyet adni fogok, az én testem a világ életéért.

rúf J 6, 51 /0 [ESZ #26380]

Én vagyok az az élő kenyér, amely a mennyből szállt le:

✓	✓	6	52	o	Vita támadt erre a zsidók közt: "Hogyan adhatja ez testét elec	≠	31201	6	51	1	ha valaki eszik ebből a kenyérből, élni fog örökké, mert az a ke
6	53	o	Jézus megerősítette: "Bizony, bizony mondom nektek: Ha nen	≠	26381	6	52	o	°A zsidók vitatkozni kezdtek egymás között, és ezt kérdezték: t		
6	54	o	Aki eszi az én testem és issza az én vérem, annak örök élete v	≠	26382	6	53	o	Jézus így szólt hozzájuk: Bizony, bizony, mondom nektek: ha n		

Figure 8. Choosing which text unit is to be split. The unit of text $N + 1$ is in the above field, the currently parallel unit of text k is below. Clicking the corresponding button on the right opens the editor where the user can split the verse into two, as in a word processor.

Splitting a verse in the “new text” is self-evident. Two text inputs appear with the full text of the verse in the first one. The user has to move the second part into the second one, then to save the change.

6

50

Itt a mennyből alászállott kenyér, hogy aki eszik belőle, meg r

=

26379

6

50

De ez az a kenyér, amely a mennyből szállt le, hogy aki eszik t

6

51

Én vagyok a mennyből alászállott élő kenyér. Aki e kenyérből

=

26380

6

51

Én vagyok az az élő kenyér, amely a mennyből szállt le:

Eltolás 1

Felosztás

Explicit pár

Eltolás 1

SplitTextUnit
Felosztás!
Mégsem

bd1951 J 6, 51 /0

Én vagyok a mennyből alászállott élő kenyér. Aki e kenyérből eszik, örökké él.

A kenyér, melyet adni fogok, az én testem a világ életéért.

rúf J 6, 51 /0 [ESZ #26380]
 Én vagyok az az élő kenyér, amely a mennyből szállt le:

✔ 6 52 o Vita támadt erre a zsidók közt: "Honnan adhatja ez testét elér ✖ 31201 6 51 i ha valaki eszik ebből a kenyérből, élni fog örökké, mert az a ke

Figure 9. Splitting a unit of text $N + 1$ by copying the second part into the second input field

Splitting a verse in text k is more sophisticated, since each corresponding unit in all texts $1 \dots N$ must be split simultaneously. The application makes this easier by showing all parallel sections of these texts in editable input fields, so the user can do all of them instantly, and save these changes in one step.

Levente Zoltán Király:
Parallelizing Bible texts
Argumentum 20 (2024), 438–449
Debreceni Egyetemi Kiadó
DOI: 10.34103/ARGUMENTUM/2024/26

The screenshot displays a Bible text editor interface. At the top, there is a toolbar with various editing tools. Below the toolbar, a list of verses is shown, each with a reference number and a text snippet. The text is organized into columns, with some verses highlighted in orange. At the bottom, there is a section titled "Szöveg egység felosztása" (Text unit splitting). This section contains a list of text units, each with a reference number and a text snippet. A "Split" button is visible next to each unit. The interface is designed to allow users to split text units into smaller segments for better management and editing.

Figure 10. Splitting a unit of previously added texts. In the present case, all 7 texts, imported and parallelized earlier, must be synchronously split to preserve parallelization. In each case the operator manually cuts and pastes the second part of the text units into the corresponding input fields.

After splitting there may be the need to acknowledge the emerging difference in the numbering, but this way the parallelization along alternate text unit boundaries can be achieved.

13	37	o	De azt, akit Isten (halottaiból)	=	27471	13	37	o	Akit azonban Isten feltámaszt
13	38	o	Vegyétek tudomásul, testvé	=	27472	13	38	o	Vegyétek tehát tudomásul,
13	38	i	Mindattól, amitől Mózes tör	≠	27473	13	39	o	és mindabból, amiből Móze
13	39	o	Ő általa megigazul minden	≠	31203	13	39	i	őáltala mindenki megigazul
13	40	o	Ügyeljetez ezért, hogy be n	=	27474	13	40	o	Vigyázzatok tehát, hogy be
13	41	o	Ide nézzetek, ti önteltek, Cs	=	27475	13	41	o	„Lássátok meg, ti gúnyolód
13	42	o	Kifelé mentek megkérték őke	=	27476	13	42	o	Amikor pedig mentek kifelé

Figure 11. In the translation BD (left) the boundary between verses Acts 13:38 and 39 is elsewhere than in the NP. This requires splitting both verse 38 in BD and verse 39 in the earlier parallelized text units, as well as acknowledging the differing verse-subverse numbers caused by the splitting.

Unfortunately there are even more sophisticated differences between biblical texts: verses and sub-verses in different order, or e.g. the last seven chapters of Jeremiah in the BHS (Deutsche Bibelgesellschaft 1997) can be found in the middle of this book in the LXX (Deutsche Bibelgesellschaft 2005). To develop a software for an automatic alignment of such differences may require more work than the manual recording of this single alignment in the database.

One of the main goals of the UBR project was to make it easy to lay different Bible translations and text versions side-by-side. We achieved this by making the transition possible among different systems of reference, by displaying the texts parallelly alongside each other, and by highlighting the equivalent textual units. Thus the reader is visually aided to see the precise parallels of a passage, even if parts of it exist in different sections in other texts, as Figure 12 shows.

KJV Psalm 73

²¹ Thus my heart was grieved,
and I was pricked in my reins.
²² So foolish was I, and ignorant:
I was as a beast before thee.
²³ **Nevertheless I am continually with thee:**
thou hast holden me by my right hand.
²⁴ Thou shalt guide me with thy counsel,
and afterward receive me to glory.

VUL Ps 72

²¹ quia contractum est cor meum
et lumbi mei velut ignis fumigans
²² et ego insipiens et nescius
²³ quasi iumentum factus sum apud te
et eram semper tecum
²⁴ **et tenebas manum dexteram meam**
in consilium tuum deduces me
et postea in gloria suscipies me

Figure 12. The exact parallel text of KJV Psalm 73:23 is highlighted in the VUL despite the different verse boundaries.

Perspectives

Currently the texts are split only as little chunks as parallelization needed them to be. However, in the future this approach can be improved in several ways. On one hand, the texts could be split everywhere by their natural inner boundaries – along the line breaks of Hebrew verse, on

the boundaries of inserted quotations, or other inner units. On the other hand, the texts could be extended with an annotation that represents the connection between equivalent units of the same texts (inner quotations, units of the synoptic gospels etc.), or between the parts of the texts that have the same linguistic function. This would not only make it possible to parallelize different texts side-by-side, but also to analyze the dynamic listing of translation alternatives for words or phrases. The parallelization of texts in the UBR can be the first, humble step on a road promising such perspectives.

References

- Deutsche Bibelgesellschaft (1997⁵): *Biblia Hebraica Stuttgartensia*. Stuttgart: Deutsche Bibelgesellschaft.
- Deutsche Bibelgesellschaft (2005²): *Septuaginta*. Stuttgart: Deutsche Bibelgesellschaft.
- Deutsche Bibelgesellschaft (2007⁵): *Sacra Vulgata*. Stuttgart: Deutsche Bibelgesellschaft.
- Károli Gáspár et al. (translators) (1908): *Szent Biblia Azaz Istennek Ó És Új Testamentomában Foglaltatott Egész Szent Írás*. Brit és Külföldi Bibliatársulat.
- Király, L. Z. (2021): Bibliaszövegek digitalizálása és párhuzamosítása. In: Fabiny, T. & M. Pintér, T. (eds.): *πῶς ἀναγινώσκεις; Hogyan olvasod? Felekezeteket összekötő Egyesített Bibliaolvasó felé*. Budapest: Hermeneutikai Kutatóközpont, 49–60.
- Magyar Bibliatanács (1974): *Biblia*. Budapest: Magyar Bibliatanács.
- Magyar Bibliatársulat (2014): *Revised New Translation*. Budapest: Magyar Bibliatársulat.
- Maszyk, Endre (1917): *Új Testámentom azaz: Istennek Új Szövetsége a mi Urunkban a Jézus Krisztusban*, Bratislava: private edition.
- Project Gutenberg (1989): *The King James Version of the Bible*, Project Gutenberg. <https://www.gutenberg.org/ebooks/10> (accessed 31 May 2024).
- Szent István Társulat (1973): *Biblia. Ószövetségi és Újszövetségi Szentírás*. Budapest: Szent István Társulat.

Levente Zoltán Király
 Károli Gáspár University of the Hungarian Reformed Church
 Unified Bible Reader research team, external collaborator
 lev.z.kiraly@gmail.com