

Εξαγωγή ορολογίας από διαβικικούς συνδέσμους

Σπύρος Δόικας

Περίληψη

Ένας μεγάλος αριθμός άρθρων της Βικιπαίδειας (σε κάθε μια από τις 299 γλώσσες για τις οποίες υπάρχουν υλοποιήσεις) διαθέτει τους λεγόμενους *διαβικικούς συνδέσμους* (interwiki links) προς το «ομότιτλο» άρθρο σε τρίτες γλώσσες. Εν προκειμένω, η χρήση των εισαγωγικών συνιστά και τον βασικό προβληματισμό για την άντληση ορολογικών ισοδυναμών. Κατά την αξιολόγηση ενός μεγάλου αριθμού άρθρων της Βικιπαίδειας βρέθηκαν στοιχεία ανισοδυναμίας και γλωσσικής ακαταλληλότητας που κατατάσσονται στις παρακάτω (ενδεικτικές και όχι εξαντλητικές) κατηγορίες:

1. Ακρόνυμο/Πλήρης μορφή όρου
2. Ενικός/Πληθυντικός
3. Διαφορετική έννοια
4. Μερικό/Γενικό (Υπερώνυμο/Υπώνυμο)
5. Συνώνυμο
6. Αντώνυμο
7. Ορολογικό σφάλμα
8. Γραμματικό/Τυπογραφικό σφάλμα
9. Εσφαλμένη μεταγραφή ονόματος
10. Λατινικός όρος/εντόπιος όρος
11. Αμετάφραστο
12. Ιδιότητα/Άτομο
13. Παραπομπή σε υποενότητα άρθρου

Για την αξιολόγηση δημιουργήθηκε ένα σενάριο ενεργειών (script) το οποίο αντλεί τους διαβικικούς συνδέσμους από το ογκώδες αρχείο XML Wikidata (40+GB) βάσει συγκεκριμένων παραμέτρων. Ένα από τα προβλήματα των δεδομένων είναι ο μεγάλος βαθμός «θορύβου» ιδίως με τη μορφή αντιστοιχιών χαμηλού ή μηδενικού γλωσσικού ενδιαφέροντος όπως π.χ. προσωπονιμίων και τοπωνυμίων. Επίσης, βρέθηκε σημαντικός πλούτος εν είδει συνωνύμων στο κυρίως σώμα των άρθρων (και κατά κύριο λόγο στην πρώτη παράγραφο) ο οποίος δεν ήταν δυνατόν να αξιοποιηθεί (ημι)αυτόματα με τη μέθοδο εξαγωγής που υιοθετήθηκε.

Περαιτέρω, για κάποια ζεύγη γλωσσών έγινε μετεπεξεργασία και εισαγωγή τους σε διαδικτυακή εφαρμογή ανοιχτής αναζήτησης.

Terminology extraction from Wikipedia's interlanguage links

Spiros Doikas

Abstract

Many Wikipedia articles have a list of links of the same article to Wikipedias in other languages. They can be found further down on the left column. These links offer potential multilingual information which could be useful as a lexicographic resource.

Before 2012, these links were stored in each separate language Wikipedia database. Exporting this type of information required running a query on the database of a specific language Wikipedia (that is the page title table and the interlanguage links table) and it was limited to a small number of languages since the query could take hours to run or even time out.

Since 2012, these links are stored in a huge mediawiki xml file which is exported from the wikidata.org site, counting over 14 million data items. Parsing this file in order to extract the interlanguage links required the creation of a custom script.

However, due to the non-lexicographic nature of these links and their varied relationships in terms of equivalence, as well as a number of other issues affecting their quality, they are not immediately usable as multilingual terminological entries. This paper is an attempt to analyze and classify these issues as well as evaluate the lexicographic potential of this resource.

The classification was largely based on the relationships between a term and its “translation”:

1. Initialism — fully expanded form
2. Singular — plural
3. Different concept
4. Hypernym — hyponym
5. Synonym
6. Antonym
7. Terminological error
8. Grammatical error/Typo
9. Erroneous proper name transcription
10. Latin term/Local term
11. Non-translated
12. Quality/Person

13. Link to article section

Apart from the linguistic information found in interlanguage links per se, a great wealth of lexicographically useful information can be harvested by means of synonym extraction; synonyms are commonly found in the first paragraph of an article, usually in bold.

The terms and examples mentioned deal mostly with Greek, German and Romance languages..

0. Εισαγωγή

Υπάρχουν εκδόσεις της Βικιπαίδειας σε πολυάριθμες γλώσσες. Για την ακρίβεια, κατά τη συγγραφή της παρούσας εργασίας ο αριθμός φτάνει τις 299, με 13 γλώσσες να έχουν πάνω από 1 εκατομμύριο άρθρα και 59 γλώσσες να έχουν πάνω από 100 χιλιάδες άρθρα. Η ελληνική κατατάσσεται στην 49^η θέση από πλευράς αριθμού άρθρων, αριθμώντας 133.142 (18/7/2017).

Πολλά από αυτά τα άρθρα διαθέτουν τους λεγόμενους *διαβικικούς συνδέσμους* (interwiki links) ή πιο συγκεκριμένα *διαγλωσσικούς συνδέσμους* (interlanguage links) μεταξύ των διάφορων εκδόσεων βίκι.

Αυτοί οι σύνδεσμοι εμφανίζονται σε μια ενότητα με τίτλο «In other languages» στο κάτω μέρος της αριστερής στήλης. Σε αυτήν την ενότητα εμφανίζονται σύνδεσμοι με τα ονόματα γλωσσών που διαθέτουν ένα αντίστοιχο άρθρο. Αρκεί να πατήσουμε σε έναν από αυτούς τους συνδέσμους για να μεταβούμε στο άρθρο.

Δεδομένου του αριθμού των άρθρων («λημμάτων») που υπάρχουν στη Βικιπαίδεια (που καλύπτουν ευρύτατο επιστημονικό και γνωσιολογικό φάσμα), καθώς και του αριθμού των γλωσσών («αποδόσεων όρων»), έχουμε δυνητικά έναν πολύ μεγάλο και πολυκατευθυντικό ορολογικό πόρο που θα μπορούσαμε να αξιοποιήσουμε.

1.0 Δομή των διαγλωσσικών συνδέσμων

Μέχρι το 2012, οι διαγλωσσικοί σύνδεσμοι αποθηκεύονταν στη βάση δεδομένων κάθε διαφορετικής γλωσσικής έκδοσης της Βικιπαίδειας. Για τη δημιουργία γλωσσικών αντιστοιχιών μεταξύ λίγων μόνο γλωσσών, έπρεπε κανείς να εκτελέσει μια εντολή MySQL που να περιλαμβάνει στοιχεία από τους πίνακες **page_title** και **langlinks** και να περιμένει ώρες μέχρι την ολοκλήρωσή της λόγω του τεράστιου αριθμού των δεδομένων. Παράδειγμα ερωτήματος στην ελληνική βάση για δημιουργία αντιστοιχιών μεταξύ ελληνικών και πορτογαλικών:

```
SELECT a.page_title, b.ll_title
FROM PAGE a
LEFT JOIN langlinks b ON a.page_id = b.ll_from
WHERE b.ll_lang = 'pt'
```

Το 2012 άλλαξε ο τρόπος αποθήκευσης διαγλωσσικών συνδέσμων με τη δημιουργία ενός κοινού αποθετηρίου αντιστοιχιών (wikidata.org) από το οποίο αντλούσαν όλες οι διαφορετικές γλωσσικές εκδόσεις. Για παράδειγμα, μπορούμε να δούμε την καταχώριση για την έννοια «calculator»¹. Στο κάτω μέρος της καταχώρισης βρίσκουμε τις αντίστοιχες σελίδες της έννοιας σε διαφορετικές γλώσσες (93 εν προκειμένω):

¹ <https://www.wikidata.org/wiki/Q31087>

Wikipedia (93 entries)  edit	
af	Sakrekenaar
am	ካልኪሊተር
ar	آلة حاسبة
atj	Akitasopitcikan
az	Kalkulyator
be	Калькулятар
bg	Калкулатор
bn	ক্যালকুলেটর
bs	Kalkulator
ca	Calculadora
chy	Hoéstoné'ó
ckb	ژمێره‌ر
cs	Kalkulačka
cv	Калькулятор
cy	Cyfrifiannell
da	Lommeregner
de	Taschenrechner
el	Αριθμομηχανή

Εικόνα 1. Τμήμα καταχώρισης «calculator» από Wikidata

Τα δεδομένα που υπάρχουν στο wikidata μπορούν να εξαχθούν με τη μορφή ενός αρχείου xml με μέγεθος γύρω στα 50 GB ή εναλλακτικά σε μορφή JSON ή RDF². Η μορφή αποθήκευσης των χαρακτήρων σε μη λατινογενείς γλώσσες ενδέχεται να είναι σε ακολουθίες διαφυγής Unicode (Unicode escape sequences)²: βλέπε για παράδειγμα παρακάτω το ελληνικό τμήμα της καταχώρισης New York (μετά το «title») σε μορφή RDF.

```
"elwiki": {
  "site": "elwiki",
  "title": "\u0394\u03ad\u03b1
\u03a5\u03cc\u03c1\u03b1\u03b7",
```

² https://www.wikidata.org/wiki/Wikidata:Database_download

2.0 Σενάριο ενεργειών για εξαγωγή διαγλωσσικών συνδέσμων

Για να αξιοποιηθούν τα δεδομένα του αρχείου xml, δημιουργήθηκε ένα σενάριο ενεργειών (script) που πραγματοποιεί εξαγωγή των δεδομένων σε μορφή csv (καθώς και μετατροπή των χαρακτήρων με μορφή ακολουθιών διαφυγής Unicode σε τυπικούς χαρακτήρες Unicode) όπου κάθε στήλη αντιστοιχεί σε μια γλώσσα και υπάρχει και μια επιπλέον στήλη για επιπλέον παραμέτρους που βοηθούν στο φιλτράρισμα των δεδομένων. Αυτές οι παράμετροι μπορούν για παράδειγμα να βοηθούν στον διαχωρισμό των εννοιών που αφορούν ονόματα προσώπων (καθώς από ορολογική άποψη δεν είναι ιδιαίτερα σημαντική πληροφορία). Το παραγόμενο αρχείο csv, μόνο με τους ελληνικούς όρους που περιέχουν αντιστοιχίες σε άλλες γλώσσες, φτάνει τα 100 MB (στην πρώτη στήλη οι κωδικοί των γλωσσών):

el→	en→	ab→	ace→	af→	als→	am→	an→	ang→	ar→	arc→	ε
Επιστήμη υπολογιστών→	Computer science→							Rekenaarwetenskap→			1
Ελληνικός→	Greek→				Grieks (dubbelsinnig)→						-
Αθλητισμός→	Sport→				Sport→		ስፖርት→	Esporte→	Indryhtegamen→		1
Οξιτανική γλώσσα→	Occitan language→						Bahsa Usitanyo→	Oksitaans→			0
Ελευθέριος Βενιζέλος→	Eleftherios Venizelos→										-
Κύπρος→	Cyprus→		Siprus→	Siprus→	Republik Zypern→	சீπருசு	Chipre→		قبرص→		1
Γεωγραφία→	Geography→		Γεωγραφία→			Geografie→	Geografie→				0
Λευκωσία→	Nicosia→		Nikosia→	Nicosia→		Λεσίνια→	Nicosia→		نيقوسيا→		1
Ομηρος→	Homer→			Homeros→	Homer→	Ὅμηρος→	Homero→		هومروس→		1
Ελληνική γλώσσα→	Greek language→		Барзен бызшаа→			Grieks→	Griechische Sprache→				0
Μυθολογία→	Mythology→				Mitologie→		Mythologie→				1
Κράτος→	State (polity)→				Staat→		Estato→		دولة→		1
Φυσική→	Physics→		Fisika→	Physik→	የተፈጥሮ ስነ-επιστήμη→	Fisica→			فيزياء→		0
Τηλεοπτικοί σταθμοί→	Television→										-
Mega_Channel→	Mega Channel→										-
Αλβανία→	Albania→	Арнауттәтәл	Albania→	Albanië→	Albanien→		ላላባያ→	Albania→	Albania→		1
Γλώσσα→	Language→			Taal→	Sprache→	ἄῤῥῆ→	Luengache→				1
Μαθηματικά→	Mathematics→				Wiskunde→		Mathematik→		ትምህርት ሃሳብ		1
Κατάλογος χωρών→	List of sovereign states→				Атәылакæи атерриториакæи алфавиттæ рыхтæ						1
Κατάλογος νησιών→	List of islands→										1
Ευρωπαϊκή Ένωση→	European Union→		Европатаи Аиллһыла				Europese Unie→	Europäisc			1
Κατάλογος κωδικών κλήσεων χωρών→	List of country calling codes→								Lys van l		1
Πληροφορική→	Computer science→										-

Εικόνα 2 Πολύγλωσσο αρχείο csv με αντιστοιχισμένα άρθρα των διαφορετικών γλωσσικών εκδόσεων της Βικιπαίδειας

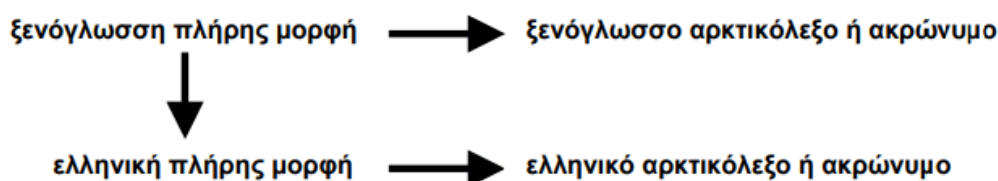
3.0 Επισκόπηση αντιστοιχιών

Οι σελίδες της Βικιπαίδειας έχουν τη δική τους δομή και κανόνες σχηματισμού οι οποίοι μπορεί να μη συνάδουν με αυτή ενός ορολογικού πόρου. Στη συνέχεια θα επιδιωχθεί μια κατηγοριοποίηση και εξέταση κάποιων χαρακτηριστικών παραδειγμάτων. Χάριν συντομίας, η γλώσσα για την οποία θα γίνεται η αναφορά θα είναι της μορφής του κωδικού γλώσσας που χρησιμοποιείται στην εκάστοτε έκδοση Βικιπαίδειας, π.χ. [en], [fr], [it], κ.ο.κ. Μπορεί έτσι να συναγάγει κανείς την ακριβή διεύθυνση URL του άρθρου με τον τύπο: [https://\[κωδικός γλώσσας\].wikipedia.org/wiki/\[τίτλος άρθρου\]](https://[κωδικός γλώσσας].wikipedia.org/wiki/[τίτλος άρθρου]) όπου «[τίτλος άρθρου]» είναι αυτό που αναφέρω ως «όρο» και εμφανίζεται με έντονη γραφή.

Το θεωρητικό υπόβαθρο για την αξιολόγηση των αντιστοιχιών από ορολογικής σκοπιάς είναι αυτό των διαγλωσσικών αρχών σχηματισμού όρων³, ήτοι «διαφάνεια, συνέπεια, καταλληλότητα, οικονομία, παραγωγικότητα, ορθότητα, εντοπιότητα».

3.1 Ακρώνυμο/Πλήρης μορφή όρου

Ο όρος **near field communication** [en] παραπέμπει στο **NFC** [el], δηλαδή έχουμε έναν αγγλικό όρο πλήρους μορφής ο οποίος αποδίδεται στα ελληνικά με το ακρώνυμο του αγγλικού όρου. Αν και σε πολλές περιπτώσεις υπάρχουν διεθνοποιημένα ακρώνυμα, δεν θα λέγαμε ότι έχουμε να κάνουμε με μια ορολογικά αποδεκτή λύση. Το ιδανικό θα ήταν ⁴:



Αν διαβάσουμε όμως την πρώτη παράγραφο του ελληνικού άρθρου, θα βρούμε με έντονη γραφή την ελληνική πλήρη μορφή του όρου με την αγγλική πλήρη μορφή και το αγγλικό ακρώνυμο να έπονται: «**Η επικοινωνία κοντινού πεδίου (near field communication, NFC) αποτελεί μια πρότυπη τεχνολογία συνδεσιμότητας...**»

Εξετάζοντας τις υπόλοιπες γλώσσες, διαπιστώνουμε τρεις βασικές προσεγγίσεις: 1) απόδοση με το αγγλικό ακρώνυμο όπως παραπάνω, π.χ. **NFC** [el, et, tr, vi]· 2) απόδοση με την αγγλική πλήρη μορφή, π.χ. **near field communication** [es, de, it, pt, ru, nl, lv, hu, pl, uk]· 3) απόδοση με εντόπια πλήρη μορφή, π.χ. **communication en champ proche** [fr], **nærfeltskommunikasjon** [no], **närfältskommunikation** [sv], **komunikacija bliskog polja** [hr].

Μια σπάνια υποκατηγορία αυτής της περίπτωσης είναι η μορφή: **πλήρης μορφή όρου → πλήρης μορφή όρου και ακρώνυμο σε παρένθεση**. Για παράδειγμα, για τον όρο **anti-thyroid autoantibodies** [en] η αντιστοιχία ήταν **αντιθυρεοειδικά αντισώματα (ΑΘΑ)** [el]. Βέβαια, δεν υφίσταται με αυτή τη μορφή πλέον καθώς μετακίνησα τη σελίδα σε μορφή χωρίς το ακρώνυμο, ωστόσο, θα διαπιστώσει κανείς το πέρασμά της βλέποντας το ιστορικό της σελίδας ή τη σημείωση ανακατεύθυνσης που θα εμφανιστεί σε περίπτωση αναζήτησης της μορφής του όρου μαζί με το ακρώνυμο.

3.2 Ενικός/Πληθυντικός

Ο όρος **primary color** [en] παραπέμπει στο **βασικά χρώματα** [el]. Ως εκ τούτου έχουμε αναντιστοιχία αριθμού. Ομοίως, ο όρος **advanced glycation end-product** [en] παραπέμπει στο **конечные продукты гликирования** [ru] ο οποίος είναι σε πληθυντικό αριθμό.

³ Κώστας Βαλεοντής, Ορόγραμμα (Διμηνιαίο περιοδικό της ΕΛΕΤΟ), Φύλλο αρ.102, Μάιος-Ιούνιος 2010, <https://www.translatum.gr/forum/index.php?topic=82421.0>

⁴ Για εκτενέστατη ανάλυση ακρωνύμων και αρκτικόλεξων βλέπε: Κώστας Βαλεοντής, *Ελληνικά αρκτικόλεξα και ακρώνυμα υπό το πρίσμα των σύγχρονων επικοινωνιακών αναγκών*, 4^ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία», 2003

Υπάρχουν, ωστόσο, ορισμένες περιπτώσεις όπου είναι απολύτως αποδεκτό ένας όρος σε πληθυντικό αριθμό να αποδίδεται σε ενικό σε μια άλλη γλώσσα όπως για παράδειγμα όροι που έχουν μόνο πληθυντικό αριθμό (pluralia tantum / πληθυντικόκλιτα). Στα αγγλικά ο όρος «trousers» (πληθυντικός) αποδίδεται ως «παντελόνι» (ενικός).

3.3 Διαφορετική έννοια

Ο όρος **European Association of Libraries and Information Services on Addictions** [en] παραπέμπει στο **Ευρωπαϊκή Ένωση Ιατρικών Βιβλιοθηκών και Πληροφόρησης** [el] οι οποίοι είναι δύο εντελώς διαφορετικοί οργανισμοί, πράγμα που είναι σαφές από την πρώτη παράγραφο του ελληνικού όρου: «(αγγλ. European Association for Health Information and Libraries - EAHIL)».

Ο όρος **number needed to harm** [en] παραπέμπει εσφαλμένα στο **número (de pacientes) que es necesario tratar** [es]. Η ορθή ισπανική απόδοση του αγγλικού όρου θα ήταν «número necesario para dañar» και η ορθή αγγλική απόδοση του ισπανικού όρου θα ήταν «number needed to treat».

3.4 Μερικό/Γενικό (Υπερώνυμο/Υπώνυμο)

Ο όρος **sadistic personality disorder** [en] παραπέμπει στα **sadisme** [fr], **sadismo** [es, it, io, pt], **Sadismus** [de], **сади́зм** [ru], κ.ο.κ. Αν θεωρήσουμε τον όρο «σαδισμός» υπερώνυμο, τότε η «σαδιστική διαταραχή της προσωπικότητας» είναι υπώνυμό του, ως τυποποιημένος χαρακτηρισμός ψυχολογικής διαταραχής. Χαρακτηριστικό είναι ότι σε πολλές γλώσσες αναφέρεται στη συνέχεια, συχνά σε έντονη γραφή μάλιστα, το υπώνυμο, π.χ. «un autre terme, celui de **trouble de la personnalité sadique**...» [fr], «Il **disturbo sadico della personalità** (comprendente anche il sadismo non sessuale)...» [it], κ.ο.κ.

Ο όρος **urinary incontinence** [en] παραπέμπει στο **недержание мочи у женщин** [ru] το οποίο σημαίνει «ακράτεια ούρων στις γυναίκες» και αποτελεί ξεκάθαρα υπώνυμο του όρου «ακράτεια ούρων».

Ο όρος **low molecular weight heparin** [en] παραπέμπει στα **eparina** [it], **heparina** [es], **heparine** [nl], **Heparin** [de]. Η «ηπαρίνη» είναι το υπερώνυμο και ένα από τα υπώνυμά της είναι η «ηπαρίνη χαμηλού μοριακού βάρους» (άλλο υπώνυμο είναι η «μη κλασματοποιημένη ηπαρίνη»).

Ο όρος **Bassae** [en] παραπέμπει στο **Ναός Επικούριου Απόλλωνα** [el]. Οι Βάσες είναι μια αρχαιολογική τοποθεσία στη Φιγαλεία μέρος της οποίας είναι και ο εν λόγω ναός. Στην πρώτη παράγραφο του αγγλικού άρθρου αναφέρεται σαφέστατα: «It is famous for the well-preserved mid- to late-5th century BC **Temple of Apollo Epicurius**.» Ως εκ τούτου, δεν μπορούμε να θεωρήσουμε τους τίτλους των σελίδων στις διαφορετικές γλώσσες ως έγκυρες ορολογικές αντιστοιχίες, παρά πρέπει να τις αντλήσουμε από το κυρίως σώμα του κειμένου.

3.5 Συνώνυμο

Η συγκεκριμένη κατηγορία είναι ίσως η πολυπληθέστερη και με τις περισσότερες παραλλαγές. Ιδιαίτερα ενδιαφέρον παρουσιάζει το πλήθος των συνωνύμων που μπορεί κανείς να βρει συνήθως στην πρώτη παράγραφο των άρθρων με έντονη γραφή, και που συνιστούν ιδιαίτερα πολύτιμη γλωσσική πληροφορία.

Ο όρος **Mediterranean diet** [en] παραπέμπει στο **Kreta-Diät** [de], δηλαδή «κρητική διατροφή». Στην πρώτη πρόταση της γερμανικής έκδοσης αναφέρεται και η πλήρως ισοδύναμη αντιστοιχία: «Die **Kreta-Diät**, oft auch **Mittelmeer-Diät** genannt...».

Ο όρος **abdominojugular test** [en] παραπέμπει στα **reflux hépato-jugulaire** [fr], **reflujo hepatojugular** [es]. Με μια πρώτη μάτια διαπιστώνουμε ότι διαφέρουν δύο οροστοιχεία. Στα αγγλικά υπάρχει το οροστοιχείο «test» (ενώ στις άλλες γλώσσες τη θέση του παίρνει το «reflux»/«reflujo») και ενώ το δεύτερο συνθετικό είναι κοινό, το πρώτο διαφέρει: στα αγγλικά «abdomino-» («κοιλιο-») ενώ στις άλλες γλώσσες «hépato»/«hepato» («ηπατο-»). Αν διαβάσουμε το αγγλικό άρθρο θα διαπιστώσουμε ότι πρόκειται για συνώνυμα: «The **abdominojugular test** or **abdominojugular reflux** (AJR) (historically known by the term **hepatojugular reflux**)». Ένα μάλιστα των συνωνύμων θεωρείται παλαιότερος όρος. Αν θα θέλαμε να είμαστε ορολογικά ακριβείς, οι αποδόσεις που θα ακολουθούσαμε θα ήταν οι εξής:

abdominojugular test [en]	κοιλιοσφαγιτιδική δοκιμασία
abdominojugular reflux (AJR) [en]	κοιλιοσφαγιτιδική παλινδρόμηση
hepatojugular reflux [en] reflux hépato-jugulaire [fr] reflujo hepatojugular [es]	ηπατοσφαγιτιδική παλινδρόμηση

Μια ενδιαφέρουσα περίπτωση συνωνύμων είναι τα κύρια ονόματα που ενδέχεται να έχουν αρκετές παραλλαγές, όπου κάποιες από αυτές να είναι πιο δημοφιλείς σε ορισμένες γλώσσες.

Το κύριο όνομα **James, brother of Jesus** [en] παραπέμπει στο **Ιάκωβος ο αδελφόθεος** [el], **Jacques le Juste** [fr], **Giacomo il Giusto** [it]. Ενώ υπάρχει αρκετά κοντινή αντιστοιχία μεταξύ αγγλικών και ελληνικών (θα ήταν απόλυτη αν αντί για «αδελφόθεος» ήταν «αδελφός του Ιησού»), βλέπουμε σημαντική διαφορά στις δύο άλλες γλώσσες όπου αποδίδεται «Ιάκωβος ο Δίκαιος». Ανατρέχοντας στην πρώτη παράγραφο του ελληνικού άρθρου διαβάζουμε: «Ο **Ιάκωβος** ήταν σημαντική προσωπικότητα της εποχής των αποστόλων. Αποκαλείται επίσης ως **Ιάκωβος ο Δίκαιος** και **Ιάκωβος ο αδελφόθεος**.» Το κείμενο αυτό προσθέτει μία ακόμη παραλλαγή, τη μονωνυμική («Ιάκωβος»), ενώ πάνω από την εικόνα διαβάζουμε «Άγιος Ιάκωβος ο Δίκαιος» και κάτω «Μάρτυς, Αδελφόθεος» (δύο ακόμη συνώνυμα).

3.6 Ανώνυμο

Ο όρος **combination therapy** [en] παρέπεμπε στο **monothérapie** [fr] ο οποίος σημαίνει ακριβώς το αντίθετο. Έχει διορθωθεί έκτοτε και πλέον παραπέμπει στο **polythérapie**.

3.7 Ορολογικό σφάλμα

Σε αυτήν την κατηγορία εμπίπτουν σφάλματα που έχουν να κάνουν με τη γλωσσική ορθότητα (linguistic correctness). Σύμφωνα με αυτήν την αρχή «προτιμώνται σχηματισμοί όρων που συμμορφώνονται με τα μορφολογικά, μορφοσυντακτικά και φωνολογικά πρότυπα της αντίστοιχης γλώσσας»⁵.

Ο (μετέπειτα διορθωθείς από τον γράφοντα) όρος **Περιβαλλοντολογικό Πρόγραμμα των Ηνωμένων Εθνών** [el] είναι εσφαλμένος. Το ορθό είναι «περιβαλλοντικό...» καθώς το επίθετο προκύπτει από το **περιβάλλον** και όχι την **περιβαλλοντολογία**⁶.

⁵ Κώστας Βαλεοντής, Ορόγραμμα: Φύλλο αρ.102

Ο (μετέπειτα διορθωθείς από τον γράφοντα) όρος **ενθρογενή** [el] είναι εσφαλμένος. Στην απόδοση αυτού του ελληνογενούς αγγλικού ουσιαστικού **entheogen** γίνεται το ίδιο λάθος με την απόδοση του επιθέτου **seismogenic** για τις **σεισμογενείς περιοχές** ενώ το σωστό είναι **σεισμογόνες περιοχές**. Η ορθή ελληνική απόδοση του όρου είναι **ενθρογόνο** (δεν παράγεται, αλλά παράγει/γεννά θεόν εντός) όπως παράγουν/γεννούν και τόσα άλλα **-gens** (oxygen → οξυγόνο, hydrogen → υδρογόνο, halogen → αλογόνο)⁷.

Ο όρος **οιστριονική διαταραχή προσωπικότητας** [el] είναι εσφαλμένος καθώς το «histrionic» δεν έχει καμία σχέση με τον «οίστρο», αλλά ετυμολογείται από το **histrion** που σημαίνει ηθοποιός στα λατινικά. Παλαιότερη ονομασία της διαταραχής ήταν το «υστερική», και ως τέτοιο καταγράφεται σε κάποιες γλώσσες στη Βικιπαίδεια: **истерическое расстройство личности** [ru], «le trouble de la personnalité histrionique (TPH ; anciennement hystérique)» [fr]. Άρα ορθότερο θα ήταν το **ιστριονική** (ιστρίωνες στην αρχαία Ρώμη ήταν ηθοποιοί που κατά την πλειονότητά τους προέρχονταν από τους απελευθέρους)⁸.

3.8 Γραμματικό/Τυπογραφικό σφάλμα

Τα γραμματικά/τυπογραφικά σφάλματα εντάσσονται επίσης στην ευρύτερη κατηγορία της «γλωσσικής ορθότητας».

Ο όρος **ταξίδι στο χρόνο** [el] είναι εσφαλμένος σύμφωνα με την τελευταία ενημέρωση του κανόνα της γραμματικής για τη χρήση του τελικού –ν: «Το τελικό ν της αιτιατικής ενικού του αρσενικού γένους του οριστικού και του αόριστου άρθρου (τον/στον, έναν), καθώς και της προσωπικής αντωνυμίας (αυτόν, τον) διατηρείται στον γραπτό λόγο πάντοτε»⁹.

Ο (μετέπειτα διορθωθείς από τον γράφοντα) όρος **Αλτζερνον Τσαρλς Σουϊνμπορν** [el] είναι εσφαλμένος διότι το γιώτα δεν χρειάζεται διαλυτικά στη λέξη «Σουϊνμπορν» καθώς δεν υπάρχει περίπτωση διαφορετικής προφοράς του.

Ο (μετέπειτα διορθωθείς από τον γράφοντα) όρος **μουικό σύστημα** [el] είναι εσφαλμένος διότι η λέξη «μουικός» χρειάζεται διαλυτικά στο γιώτα έτσι ώστε τα δύο φωνήεντα να προφέρονται ξεχωριστά.

3.9 Εσφαλμένη μεταγραφή ονόματος

Το κύριο όνομα **Penélope Cruz** [es] αποδίδεται ως **Πενέλοπε Κρουζ** [el] αντί για **Πενέλοπε Κρουθ** (είναι ισπανίδα και όχι όνομα πυραύλου!). Σε περίπτωση που η καταγωγή της ήταν από τη Λατινική Αμερική, τότε η προφορά πάλι δεν θα ήταν **Κρουζ**, αλλά **Κρους**.

Το κύριο όνομα **Percy Williams Bridgman** [en] αποδίδεται ως **Πέρσι Γουίλλιαμς Μπρίντγκμαν** [el] αντί για **Μπρίτζμαν** που είναι η ορθή προφορά του ονόματος (πρβ.: bridge → μπριτζ). Σε περίπτωση αυτού του είδους των αποριών μπορεί κανείς να συμβουλευτεί ιστοτόπους με ηχητικά αρχεία προφοράς όπως το forvo.com¹⁰. Αντίστοιχα, το ορुकτό «bridgmanite» το οποίο ονομάστηκε κατ' αυτόν

⁶ <https://www.translatum.gr/forum/index.php?topic=4013.msg103268#msg103268>

⁷ <https://www.translatum.gr/forum/index.php?topic=47581.msg193447#msg193447>

⁸ Πάπυρος – Μέγα Λεξικό της Ελληνικής Γλώσσας

⁹ Γραμματική Νέας Ελληνικής Γλώσσας Α', Β', Γ' Γυμνασίου, ΟΕΔΒ, 2008, σελ. 21. Βλέπε επίσης:

<https://www.translatum.gr/forum/index.php?topic=3439.0>

¹⁰ https://forvo.com/word/percy_williams_bridgman/

τον τρόπο προς τιμήν του προαναφερθέντος επιστήμονα, θα πρέπει να αποδίδεται ως «μπριτζμανίτης» και όχι ως «μπρινγκμανίτης».

Ο ιδιαίτερα γνωστός στην Ελλάδα **Henry Kissinger** [en] μεταγράφεται ως **Χένρυ Κίσινγκερ** [el]. Ωστόσο, όπως φαίνεται τόσο από το σχόλιο προφοράς στην αγγλική Βικιπαίδεια, όσο και από ηχητικές εγγραφές στον ιστότοπο forvo¹¹, η προφορά του ονόματός του στον αγγλόφωνο κόσμο είναι **Κίσιντζερ**. Ίσως συνδράμει σε αυτήν την ελληνική προτίμηση η γερμανοεβραϊκή καταγωγή του (το αρχικό του όνομα ήταν Heinz Alfred Kissinger).

Το κύριο όνομα **Vladimir Putin** [en] μεταγραφόταν (μέχρι που διορθώθηκε από τον γράφοντα) ως **Βλαντιμίρ Πούτιν** [el] αντί για το ορθό **Βλαντίμιρ Πούτιν** (κατά το «Βλαδίμηρος» και όχι «*Βλαδιμήρος»). Σε μια αναζήτηση βρέθηκε ιδιαίτερη ποικιλομορφία όσον αφορά τον τονισμό και την μεταγραφή ή ελληνοποίηση στις αποδόσεις ονομάτων που περιείχαν το «Vladimir»:

Vladimir Lossky	Βλαντιμίρ Λόσκι
Vladimir Jirinovski	Βλαντιμίρ Ζιρινόφσκι
Vladimir Kramnik	Βλαντιμίρ Κράμνικ
Vladimir Voronin	Βλαντιμίρ Βορόνιν
Vladimir Ier	Βλαδίμηρος Α' του Κιέβου
Vladimir Nabokov	Βλαντιμίρ Ναμπόκοφ
Vladimir Maïakovski	Βλαντίμιρ Μαγιακόβσκι
Vladimir II Monomaque	Βλαδίμηρος Μονομάχος
Vladimir Vyssotski	Βλαντίμιρ Βισότσκι

Το κύριο όνομα **Joseph Capgras** [fr] μεταγραφόταν (μέχρι που διορθώθηκε από τον γράφοντα) ως **Ζοζέφ Κάπγκρας** [el] αντί για το ορθό **Ζοζέφ Καπγκρά**. Πιθανολογώ ότι αιτία ήταν η μετάφραση από το αγγλικό άρθρο και από άτομο που δεν γνώριζε γαλλικά ή βασικά στοιχεία γαλλικής προφοράς.

3.10 Λατινικός όρος/εντόπιος όρος

Σε ορισμένες γλώσσες χρησιμοποιείται αντί για του εντόπιου όρου, ο λατινικός όρος. Ουσιαστικά, κάτι τέτοιο παραπέμπει σε σφάλμα γλωσσικού ζεύγους. Το φαινόμενο είναι ιδιαίτερα συχνό σε ονομασίες ειδών του φυτικού και ζωικού βασιλείου.

Ο όρος **zebra finch** [en] παραπέμπει σε **Taeniopygia guttata** [es, it, la, ast, ceb, eu, war], στον λατινικό όρο δηλαδή. Σε αρκετές γλώσσες αναγράφεται ο εντόπιος όρος στην πρώτη παράγραφο. Για παράδειγμα, στο ισπανικό άρθρο: «El **diamante cebra de Timor** (*Taeniopygia guttata*), también conocido como **diamante mandarín** o **pinzón cebra**», αναγράφονται τρεις εντόπιοι όροι.

3.11 Αμετάφραστο

Σύμφωνα με την αρχή της εντοπιότητας «ακόμα και αν ο άμεσος δανεισμός από άλλες γλώσσες είναι αποδεκτή μορφή δημιουργίας όρων, θα πρέπει, έναντι των άμεσων δανείων, να δίνεται προτεραιότητα στη γλώσσα του τόπου όπου θα χρησιμοποιηθεί ο όρος»¹². Εν προκειμένω ο τίτλος της ενότητας «Αμετάφραστο» είναι καταχρηστικός καθώς οι όροι δεν «μεταφράζονται» αλλά αποδίδονται, αλλά χρησιμοποιήθηκε για λόγους έμφασης και οικονομίας.

¹¹ <https://forvo.com/search/Kissinger/en/>

¹² Κώστας Βαλεοντής, Ορόγραμμα: Φύλλο αρ.102

Ο όρος **reduced instruction set computer** [en] έχει την ίδια απόδοση σε [el, de, it, nl] ενώ σε κάποιες γλώσσες εμφανίζεται με την παραλλαγή **reduced instruction set computing** [es, fr]. Σε κάποιες περιπτώσεις, όπως στα ελληνικά, υπάρχει απόδοση του όρου με έντονη γραφή στην πρώτη παράγραφο: «Η συντομογραφία του είναι RISC (προφέρεται ρισκ) και σημαίνει **υπολογιστής περιορισμένου συνόλου εντολών**».

Ο όρος **bullying** [en] έχει διατηρηθεί ως έχει σε κάποιες γλώσσες [el, pt, ro]. Για τα ελληνικά τουλάχιστον μάλλον αδικαιολόγητα καθώς υπάρχει καθιερωμένη απόδοση, η οποία εμφανίζεται και στην αρχή του άρθρου με έντονη γραφή: «Ο **εκφοβισμός** (αγγλική γλώσσα: bullying γαλλική γλώσσα: chicane) αναφέρεται στη σωματική και ψυχολογική κακοποίηση ή μείωση ασθενέστερων ατόμων σε μια ομάδα». Η αναφορά για «chicane» στη γαλλική γλώσσα δεν αντιστοιχεί με το αντίστοιχο άρθρο της γαλλικής Βικιπαίδειας («brigade»). Σε παλαιότερη έκδοση του ελληνικού άρθρου (βλέπε ιστορικό) αναφερόταν: «Το νταηλίκι (τουρκική: dayilik) είναι μια μορφή κακοποίησης, πιο συγκεκριμένα εκφοβισμού ή εξαναγκασμού» οπότε βλέπουμε ότι υπήρχε μια ορολογική... αναταραχή όσον αφορά την απόδοση του συγκεκριμένου όρου, ίσως και γι' αυτό προτιμήθηκε ο αγγλικός όρος στον τίτλο.

3.12 Ιδιότητα/Άτομο

Ο όρος **autodidacticism** [en] («αυτοδιδασκισμός», «αυτομόρφωση») αποδίδεται με πλήρως αποδεκτή αντιστοιχία ως **autoaprendizaje** [es], **autoformation** [fr] αλλά σε κάποιες άλλες γλώσσες χρησιμοποιείται η απόδοση «αυτοδίδακτος», δηλαδή το άτομο που ενσαρκώνει την ιδιότητα του αγγλικού όρου: **autodidata** [pt], **Autodidakt** [de].

Ο όρος **false self-employment** [en] («ψευδής αυτοαπασχόληση») παραπέμπει στο **faux indépendant** [fr] («ψευδοαπασχολούμενος», «εργαζόμενος υπό καθεστώς ψευδούς αυτοαπασχόλησης»), δηλαδή έχουμε στην πρώτη περίπτωση την ιδιότητα και στη δεύτερη το άτομο που την ενσαρκώνει.

3.13 Παραπομπή σε υποενότητα άρθρου

Ενώ η απόλυτη πλειονότητα των αντιστοιχιών έχουν να κάνουν με τίτλους άρθρων μεταξύ των διαφορετικών εκδόσεων, υπάρχουν κάποιες περιπτώσεις όπου η παραπομπή δεν γίνεται σε τίτλο άρθρου μιας άλλης γλωσσικής έκδοσης, αλλά σε υποενότητα τίτλου άρθρου. Π.χ. ο όρος **Judgement of Paris** [en] παραπέμπει στην υποενότητα **Η «Κρίση του Πάρι»** του άρθρου **Πάρις** [el].

4. Εισαγωγή αντιστοιχιών στο Translatum Forum

Αρκετές αντιστοιχίες, κατόπιν επεξεργασίας, έχουν εισαχθεί σε διάφορα γλωσσικά ζεύγη στο Translatum Forum. Για παράδειγμα, σχεδόν κατ' αποκλειστικότητα, το γλωσσικό ζεύγος Ελληνικά→Ιαπωνικά περιέχει όρους που έχουν εξαχθεί από τη Βικιπαίδεια. Η ιδιαιτερότητα του εν λόγω ζεύγους είναι ότι καθώς οι δύο γλώσσες έχουν ριζικά διαφορετικούς χαρακτήρες, ακόμη και πληροφορίες που θα είχαν μηδαμινό γλωσσικό ενδιαφέρον μεταξύ συγγενικών γλωσσών (π.χ. Αγγλικά→Γερμανικά), αποκτούν εν προκειμένω άλλη βαρύτητα. Για παράδειγμα τα κύρια ονόματα:

Απόστολος Παύλος	パウロ
Αγιος Παύλος	セントポール
Νήσος Άγιος Παύλος	サンポール島
Πατριάρχης Βελιγραδίου Παύλος	パヴレ (セルビア総主教)
Παύλος ντε Γκρέτσια	パウロス (ギリシャ王太子)
Πάπας Παύλος Α'	パウルス1世 (ローマ教皇)
Πάπας Παύλος Γ'	パウルス3世 (ローマ教皇)
Πάπας Παύλος ΣΤ'	パウロ6世 (ローマ教皇)
Παύλος Α' της Ρωσίας	パーヴェル1世
Λεύκιος Αιμίλιος Παύλος Μακεδονικός	ルキウス・アエミリウス・パウルス・マケドニクス
Παύλος Α' της Ελλάδας	パウロス1世 (ギリシャ王)
Πάπας Ιωάννης Παύλος Α'	ヨハネ・パウロ1世 (ローマ教皇)
Πάπας Ιωάννης Παύλος Β'	ヨハネ・パウロ2世 (ローマ教皇)

Εικόνα 3 Αναζήτηση «Παύλος» Ελληνικά→Ιαπωνικά στο Translatum Forum

5. Επίλογος

Η Βικιπαίδεια καθώς ολοένα εμπλουτίζεται με υλικό σε πολυάριθμες γλώσσες, συνιστά έναν ιδιαίτερα πλούσιο και εν πολλοίς αναξιοποίητο ορολογικό πόρο. Ωστόσο, ένεκα της δομής των άρθρων της, θα πρέπει κανείς να έχει υπόψη του ότι έχει να κάνει με ακατέργαστα δεδομένα που πρέπει να υποβάλλονται σε κριτική εξέταση και διερεύνηση πριν την οποιαδήποτε αποδοχή τους.

Εδώ, σημαντικός είναι ο ρόλος του κατ' εξοχήν κριτικού αναγνώστη αυτής της μορφής των δεδομένων: του επαγγελματία ορολόγου, ο οποίος μπορεί να αναλάβει να ξεχωρίσει την ήρα από το σιτάρι και να επιμεληθεί το υλικό για την παρουσίασή του σε μορφή που να είναι καταλληλότερη για σκοπούς ορολογικής χρήσης και επεξεργασίας.

[Σπύρος Δόικας](#)

Ιδρυτής της μεταφραστικής πύλης www.translatum.gr