

Research Article

Attila Imre*

Categorizing and translating abbreviations and acronyms

<https://doi.org/10.1515/opli-2022-0204>

received May 24, 2022; accepted August 26, 2022

Abstract: The popularity of various types of abbreviations makes it necessary to rediscuss their categorization and possible disambiguation. We rely on categorization types applied in cognitive linguistics, also confronting definitions and forms stemming from both linguistic and software-based approaches. A major distinction is observed between abbreviations resulting from one-word and multi-word sequences, leading to various subtypes with prototypical, less central, and hybrid cases. Although guidelines offer advice on their use, these rules should be re-evaluated in specific settings, such as subtitling and translation. While previous research on the topic focused on journal articles, we have collected a database of nearly 13,000 abbreviations and acronyms from five American TV series with the help of a specially designed algorithm. Our research also highlights the importance of punctuation, exemplifying some of the most frequent ones with alternative versions (with or without period) and discusses the Romanian and Hungarian translations of a well-known American agency. The concluding remarks mention that even if subtitle conventions are not severely regulated, a database of acronyms may significantly improve quality, especially in the case of TV series.

Keywords: audiovisual translation, abbreviation, TV series, algorithm, database

1 Introduction

Using various sorts of abbreviations is a topical issue, and scholars face various challenges regarding them, ranging from overlapping definitions to possible translations.

Their topicality is mentioned by recent studies, as abbreviations, acronyms, and initialisms are “gaining ground in every language because they accelerate communication as they are clear and time saving” (Panajotu 2010, 160). Recent research proves that acronym use has steadily increased over time, even if “only 0.2% is used regularly,” while “the majority (79%) fewer than 10 times” (Barnett and Doubleday 2020, 1). However, there is a considerable fluctuation in using abbreviations and acronyms, and the same authors mapped again the most popular acronyms during the height of the COVID-19 pandemic, offering the most recent data. Thus, we know that *COVID* (← “coronavirus disease”) has become the sixth most popular acronym in journal titles over a period of five decades (1950–2020), which is rather noteworthy, as it had more occurrences than *AIDS* (← “acquired immunodeficiency syndrome”), *PCR* (← “polymerase chain reaction”), or *MRI* (← “magnetic resonance imaging”) “in just one year” (Barnett and Doubleday 2021, 6128).

The figures prove that “human languages are very prone to the creation of acronyms” (Sánchez and Isern 2011, 311), for which the authors offer a few explanations: the typical graphic image of an acronym captures the attention by relying on full uppercase letters, the shorter form is easier to remember,

* **Corresponding author: Attila Imre**, Department of Applied Linguistics, Sapientia Hungarian University of Transylvania, Cluj-Napoca, 400112, Romania, e-mail: attilaimre@ms.sapientia.ro
ORCID: Attila Imre 0000-0003-3886-7091

redundancy is avoided, and ultimately acronyms save space. Yet, others tend to focus on the negative aspects, as abbreviations and acronyms can be alienating or ostracizing (Hales et al. 2017, 22), difficult to read (Thomas 2021, 467), and disambiguation is often needed even when experts are involved, as one single acronym may have more similar expanded forms. For instance, *CCC* may have as many as 472 possible “definitions,” that is, expanded versions, split into various categories (Information Technology 39, Military and Government 98, Science and Medicine 60, Organizations, Schools, etc. 274, Business and Finance 99, Slang, Chat, and Pop culture 14).¹ These are difficult to interpret, as some might think about the *Christchurch City Council* in New Zealand, while others recall the *CCC* shoe stores company in Poland, and the context by itself may not be relevant enough.

2 Defining and categorizing abbreviations

Before discussing issues and possible options for disambiguation, we should list and define types of abbreviations, starting from the most widespread term, then enlarging the category with possible subtypes.

Abbreviation may be considered the umbrella term for all cases when the original word or phrase is shortened. Thus, the original (expanded) version becomes “shortened,” “truncated,” or “contracted” (cf. Nicoll 2016, 2, Thomas 2021, 467). However, the majority of scholars do not make a clear distinction between one-word shortening and multi-word shortening; hence, many ambiguous definitions stem from this.

We propose to create two main categories, one of them containing all the available processes to shorten a single word, resulting in *abbreviation/shortening*, *clipping*, *truncation*, and *contraction* within a word. This distinction is also supported by Mattiello (2013, 72), whereas others overlooking this initial separation seem to provide ambiguous terms and definitions when all their definitions are confronted (Cintas and Remael 2020, 137, HaCohen-Kerner et al. 2004, 59, Kasprowicz 2010, Koelsch 2016, 359, Kuzmina et al. 2015, 550, López Rúa 2004, 116, Trumble and Stevenson 2002, 3).

While categorizing terms, we are also aware of the impossibility to establish clear-cut categories, which is natural since the advent of cognitive linguistics, whose promoters describe fuzzy boundaries (cf. Brugman 1988, Kövecses 2002, Lakoff 1987, Rosch 1975). Thus, we primarily focus on prototypical instances accepting that exceptions and hybrid cases also exist.

A prototypical *abbreviation* is “a shortened form of a term” (Soyer 2018, 589), but these types of abbreviations only appear in written form (Deme 1955, 399), and they are still pronounced as a full word. A good example is *p.* (← “page”), in which case only the first letter is preserved, and the period replaces the rest of the word. Although *shortening* is a similar term, some considering that it should replace *abbreviation* (Cannon 1989, 106, Kortmann 2020, 60, López Rúa 2004, 124), it is still a less circulated term.

Truncation is a specific sub-type of abbreviation, during which only the first part of the word remains (*teach* ← “teacher”), which is often used as a synonym for *clipping*, which signals a loss of the middle part (*flu* ← “influenza”) or the beginning of the word (*burger* ← “hamburger”). While more scholars discuss these two types of abbreviations (Carter and McCarthy 2006, 482, Ludányi 2013, 154–5, Quirk et al. 1985, 1580), little effort is made to clearly distinguish them, and the majority of people use the two terms as synonyms. Moreover, the definition of *truncation* still overlaps with the definition of *abbreviation*, and *Prof.* (← “professor”) may be labeled as both an abbreviation and a truncation.

Finally, single-word *contraction* is the last type we consider to be among *abbreviations*, as this refers to the situations when only the initial and final letter(s) remain, occasionally preserving a middle letter as well: *Dr./Dr.* (← “doctor”),² *km* (← “kilometer”), or postal abbreviations, such as *Bld.* (← “boulevard”).

While a few abbreviations are considered to have a professional air, the majority of scholars agree that they are rather informal, belonging to the “conversational use” (Soyer 2018, 589). Further remarks

¹ <https://www.acronymfinder.com/CCC.html>, April 22, 2022.

² The abbreviation may be used with or without a final period (US/UK).

concerning the abbreviations are possible, yet they are less relevant, as little agreement is found when statements are confronted (e.g., article use and lack or presence of period).

To sum up, the term *abbreviation* proves to be a successful cover for prototypical *abbreviation* or *shortening*, *truncation*, *clipping*, and *contraction* within a single word. On the other hand, few people consider it important to clearly distinguish them, enabling marginal and hybrid cases as well, such as examples without a final period (most notably units of measurements and chemical symbols) or abbreviations disregarding the original letters (*lb* ← “pound”).

However, the term *abbreviation* encapsulates various shortenings of multi-word sequences as well, leading to somewhat puzzling terms, such as “graphic abbreviations” (Mattiello 2013, 72) or “pseudo-acronyms” (López Rúa 2004, 118), mostly used in social media (*IH8U* ← “I hate you.”). More commonly accepted terms for abbreviations deriving from multi-word sequences are *blend* and *contraction* of the second word.

Blends preserve various parts of multi-word sequences, most typically the first letters or syllables (*Interpol* ← “International Police”) or “the initial part of one word with the last part of the second word” (Kuzmina et al. 2015, 550).

The *contraction* of the second word usually implies an auxiliary word dropping its first letter(s), replaced by an apostrophe, and joining it to the preceding word: *we’re* (← “we are”), *you’ve* (← “you have”). Similarly to single-word contractions, this type is not recommended for “formal writing” either (Soyer 2018, 589).

The third batch associated with the term *abbreviation* is *acronym*, which is also used as an umbrella term for abbreviations deriving from multi-word sequences “that include capital letters” (Barnett and Doubleday 2020, 4), most notably *acronyms*, *initialisms*, and *alphabetisms*. Although *initialism* is the older term, “it has never caught on in wider usage” (Zimmer 2010); hence, *acronym* may be considered the covert term for the entire category.

Interestingly, acronyms have captured the attention of computer programmers as well, who designed various codes (algorithms) to detect acronyms in texts, together with their possible expanded forms, as it was observed that these are usually found within the range of less than 20 words distance from the acronym. It is also known that the IT industry uses the term *acronym* for both acronyms and abbreviations (Cannon 1989, 106, Koelsch 2016, 80). Although some state that “acronym and initialism are often used as synonyms” (Caon 2016, 11), others deny that (Péter 2003, 125–6). As explained, acronym is “often misused to refer to any arrangement of letters that stand in for full words” (Hales et al. 2017, 22), including initialisms, also used in algorithm-driven approaches.

These clashes signal that “acronyms were a nuisance – words that almost never appeared in dictionaries but, of course, were known to be valid strings” (Taghva and Gilbreth 1999, 191) and the term *acronym* “has remained maddeningly ill-defined for its entire existence” (Zimmer 2010), characterized by “overlap,” “vagueness,” and “lack of agreement” on its scope (López Rúa 2004, 110), and explanatory dictionaries only add to the confusion as they do not clarify that *abbreviation* is used as a general term for abbreviations, acronyms, or initialisms as well (Trumble and Stevenson 2002).

Prototypical *acronyms* derive from multi-word sequences starting with initial uppercase letters, all preserved in the acronym; thus, a new word comes into being, which is pronounced as a word (*POTUS* ← “President of the United States”). Prototypical *initialisms* only differ from prototypical acronyms that they cannot be pronounced as words, the only option being a letter-by-letter pronunciation (*BBC* ← “British Broadcasting Corporation”). This is the only relevant difference between these two categories, being distinguished as “orthoepic, or letter-sounding” and “alphabetic, or letter-naming” (Kreidler 2000, 957). Similar distinctions are formulated by other scholars as well (Jacobs et al., 2018, 517, Koelsch 2016, 359, Quirk et al. 1985, 1582, Thomas 2021, 467, Trumble and Stevenson 2002, 21, Yule 2010, 58, Zahariev 2004, 366).

Alphabetism may be defined as the “use of initials as a signature or assumed indication of authorship” (Trumble and Stevenson 2002, 61), and the first part of the definition is used by scholars dealing with word-shortening processes. Thus, some scholars state that *alphabetism* is a synonym of *initialism* (López Rúa 2004, 118), being uttered as “sequences of letters” (Quirk et al. 1985, 1581). However, others consider that *alphabetism* is the hyperonym for both *initialism* and *acronym* (Mattiello 2013, 67, Scarpa 2020, 66). Based

on the majority of sources, we consider that the term *alphabetism* may be ignored and preserve *initialism* instead.

Definitions stemming from algorithm developers go further and try to set the minimum and maximum number of letters forming a prototypical acronym or initialism, ranging from two or three to nine or ten uppercase letters the most. However, they also consider non-alphabetic characters in the string, most commonly digits and specific signs and symbols, such as period, hyphen, slash, and ampersand (Barnett and Doubleday 2020, 4, Cannon 1989, 116, Dannewitz Linder 2016, 253, Park and Byrd 2001, 131, Sánchez and Isern 2011, 313, Taghva and Gilbreth 1999, 192), observing punctuation differences between British and American (Milinković 2019, 156) or stating that punctuation changes over time (Thomas 2021, 469). As such, full uppercase acronyms should have neither space nor period between them (CE 2014, 19, Cintas and Remael 2020, 137, Thomas 2021, 467, Wallwork 2014, 106), although this is no more than a recommendation, as – for instance – *The Washington Post* consistently uses *U.S.* and not *US* (← “the United States of America”).

The created acronyms or initialisms are considered to be the “most peripheral to word formation” (Carter and McCarthy 2006, 482), scholars even describing them as a sort of “nonword with meaning” (Izura and Playfoot 2012, 864) or similar to “irregular” words (Laszlo and Federmeier 2007, 1161), and they gain an “iconographic independence” (Alonso 2008, 16). As such, they are either “highly familiar to the language user” (Izura and Playfoot 2012, 862) or “often used without our knowing what the letters stand for” (Quirk et al. 1985, 1582), and their frequent use in “scientific communication” is “mostly unnecessary” as “they can confuse and alienate unfamiliar audiences” (Hales et al. 2017, 22) and are often viewed as an “obstacle in reading” (Thomas 2021, 467). Less prototypical or hybrid cases may pose further challenges, for instance acronyms containing both uppercase and lowercase letters (*CoS* ← “Chief of Staff”), acronyms created not only from the initials of the expanded version (*ASPEN* ← “Automated Survey Processing Environment”),³ or when the acronym is pronounced as a mixture of initialism and acronym (HaCohen-Kerner et al. 2013, 2133), as in the case of *JPEG* (← “Joint Photographic Experts Group”), where the first character is pronounced as a letter, and the rest of the string as a word.

3 Guidelines for using abbreviations

As this very brief presentation of abbreviations already foreshadows possible misunderstandings or spelling errors connected to acronyms, the logical question arises: Why not completely ignore them? We tend to think that the most succinct answer is brevity. All sources dealing with them highlight that they save space (when printed) and time (when uttered), and extra features are provided by their visual appearance: all uppercase letter words capture our attention, they can be more easily remembered, and they reduce the awkwardness of communication by “summarizing” redundant parts of a long string of characters (*DARPA* of *DOD* ← “the Defense Advanced Research Projects Agency of the United States Department of Defense”). Naturally, these advantages of acronyms have been fully employed, and English is “very prone to using and creating new words like acronyms” (Cintas and Remael 2020, 138) to the extent that editors of scientific journals have been forced to implement restrictions on their use, albeit accepting the obvious benefits. Thus, authoritative guidelines (especially medical ones) formulate the following:

- *AMA Manual of Style*: “Author-invented abbreviations should be avoided” (Bauchner and the JAMA Network 2020, 556);
- “It is best to avoid making up your own acronyms.” (Nicoll 2016, 8);
- “More than one new neologism or novel abbreviation per paper burdens the reader” (Bloom 2000, 4);
- *APA Publication Manual*: “To maximize clarity, use abbreviations sparingly” (APA 2019, 175);

³ <https://aspe.hhs.gov/common-acronyms>, April 24, 2022.

- “Write it in full at least for the first time... followed immediately by the abbreviated form in parentheses. Full form of the abbreviation shall be repeated, if it is used again in the document after a lengthy gap” (Thomas 2021, 467);
- “If you use a lot of uncommon acronyms, provide a glossary.” (Wallwork 2014, 105).

Arguably, the most ardent opponent of acronyms is Colin B. Begg, editor of *Clinical Trials*, who finishes his editorial this way: “I am announcing henceforth our ZeTA policy – Zero Tolerance for Acronyms. Please be aware of this if you are contemplating a submission.” (Begg 2017, 562). However, the majority of editors are aware that this is rather untenable, so it is worth considering a standard recommendation. *Open Linguistics* restricts them in the instruction connected to the title and abstract of the article: “Avoid specialist abbreviations and non-standard acronyms.”⁴ Then, a separate section is dedicated to their use elsewhere:

Please use standard abbreviations. Ensure consistency of abbreviations throughout the article. Non-standard abbreviations should not be used unless they appear at least three times in the text. List all abbreviations, acronyms and symbols in alphabetical order, along with their expanded form, at the end of the text. Define them as well upon first use in the text.

While it is beyond doubt that journals will not accept manuscripts disregarding the instructions, we are interested whether these guidelines may be applied in a specific setting, namely subtitling, belonging to the audiovisual industry. As known, this is concerned with the entertainment of their customers; hence, no whatsoever impediment should occur while enjoying the video, and the supporting subtitle should enhance comprehension.

While few sources discuss the use of abbreviations and acronyms in subtitles, the *BBC Guidelines* specifies briefly that “Never use symbols for units of measurement. Abbreviations can be used to fit text in a line, but if the unit of measurement is the subject do not abbreviate.” (Williams 2009, 34). Others explain that “most companies recommend the avoidance of abbreviations” in subtitles, unless “they should be known by the target audience or not cause any confusion” (Cintas and Remael 2020, 137). In this respect, subtitles are supported by mass media and social media, which popularize many abbreviations and acronyms on a daily basis. Some of these become commonly known, especially connected to current affairs in politics, economics, or health, which can easily appear in subtitles as well.

4 Translating acronyms

As there are multiple issues with acronym use, this logically entails that attempts to render them in other languages amplify the problems.

Our hypothesis is that source text acronyms challenge translators in the majority of cases, the only exception being when the source term is also used in the target language community, so pure borrowing is possible, as in the case of *BBC* or *CIA*, which need no further disambiguation.

Another satisfactory case scenario is when the source term has an established equivalent in the target text, in which case a dictionary or an online search can help easily (*UN* ← “United Nations,” *OVN* in German ← “Organisation der Vereinten Nationen”; *ONU* in Romanian ← “Organizația Națiunilor Unite,” or *ENSZ* in Hungarian ← “Egyesült Nemzetek Szervezete”).

A rather frequent option is to offer the translation of the expanded form. In these cases, no exact match is possible: *DOD* ← “Department of Defense” becomes *ministerul apărării* in Romanian (Cojocaru 1976, 185). Although “Ministerul Apărării Naționale” is the institution with a similar function in Romania (→ *MApN*), the acronym cannot be used, as it refers to a different country. However, this is still a fortunate case, when a similar institution exists in the target culture, but acronyms may be replaced “with an explanation” (Cintas and Remael 2020, 138) when a similar institution is not present in the target country.

⁴ <https://www.degruyter.com/journal/key/opli/html>, Instructions for authors, April 24, 2022.

Character growth may become a serious issue given the character limit of subtitle lines; thus, subtitlers may rely on disambiguation on its first occurrence and then re-use the original acronym on subsequent occasions. We tend to believe that this is the most frequent strategy, as “the abbreviation can be used in its original form every time it appears in the subtitles” (Cintas and Remael 2020, 138), mitigating the risk of lengthy lines or leaving the readers in the dark. However, this is not an option for all languages, as Arabic or Hebrew have no uppercase letters (al-Qinai 2007, 368–9). In more fortunate cases, translators can preserve many political, financial, technical, or medical acronyms. Yet, a few scholars think that giving the abbreviation form “should be the exception” as “[a]bbreviations are mentally taxing on a reader and can incidentally alienate an audience” (Hales et al. 2017, 24), and – in their view – improving the communication involves “replacing them [the abbreviations] with the words that they stand for.”

A further, seemingly clever solution is to omit the acronym with the help of grammar. We have in mind replacing the acronym with a pronoun (*the DOD ~ they*) or rephrasing the sentence in passive voice to eliminate the subject (*The DOD stated ~ It was stated*).

A final method is complete omission, without leaving any trace of the source text acronym, which ultimately leads to text impoverishment if translators resort to it too often.

5 A newly created subtitle database

While the presented theoretical background may offer guidelines on acronym use, including their translation, we are interested whether these are applied “in reality,” so we have created a database of acronyms of highly popular TV series.

The database contains the English, Romanian, and Hungarian subtitles of 587 episodes, mostly focusing on US politics, crime, and law: *House of Cards* (73 episodes), *Designated Survivor* (53 episodes), *The West Wing* (156 episodes), *24* (205 episodes), and *Blindspot* (100 episodes). The main reason to select these series is that they are aired on the most popular on-demand streaming services (e.g., *Netflix*), thus they set certain standards when official multi-language subtitles join the video material.

We have also used a specific software created for personal use by Nándor Makkai, a professional computer programmer, to detect and extract *acronyms* from the English original subtitles, obtaining 12,916 instances. Although this is a large number, this means in reality that around 22 capitalized abbreviations or acronyms appear in a single episode of about 40 min (around one acronym per every other minute). It is known that the 26 English characters may result in 17,576 ($= 26^3$) possibilities, and an ambitious research states that 94% of them was actually used in scientific texts, although 30% of them occurs only once, and half of them only between two and ten times (Barnett and Doubleday 2020, 2). They also found that only 0.2% is used very frequently, and three-letter acronyms are the most popular.

It is important to mention that our algorithm was implemented to detect any string of at least two uppercase letters, knowing that “[a]cronyms are detected using capitalization heuristics” (Sánchez and Isern 2011, 312).

Although our definition of “any string of at least two uppercase letters” seems simplistic, not being able to detect non-prototypical acronyms in the regular way, such as mixtures of uppercase and lowercase letters or combinations of alphanumeric characters and various signs and symbols (apostrophe, ampersand, slash, or hyphen), the bulk of acronyms was found. However, we also know that digits and symbols must never break the acronyms into two separate lines, which is why the collected acronym is completed with the entire line in which it is detected, including the source and timing, illustrated with a print screen fragment of *House of Cards* (Figure 1).

TRIPLE	HOC [1x12] en.srt	00:06:27,387	My sources at the D-TRIPLE-C
BRAC	HOC [1x12] en.srt	00:06:56,708	From the BRAC commission.

Figure 1: SRT Manager, acronym with data.

As for the detection of period between uppercase letters, both versions were coded to be found, thus *US* and *U.S.* are collected if found. The software also provides further statistics, such as the option to save the entire list of findings into a separate file to be handled later by a spreadsheet file, and a separate list of unique values with the total number of occurrences for each entry, illustrated in Figure 2. Subsequent data processing is needed, as Figure 2 already shows preliminary issues: there are variants with and without periods (*DC* and *D.C.*) as well as entries for a possible ignore list (*TV* and *OK*), which are international words. Hence, their translation poses no challenge whatsoever in the languages in question (Romanian and Hungarian).

Save as File	Statistics
^	180 unique values
	ICO 87
	TV 84
	OK 79
	FBI 53
	US 50
	UN 40
	DC 30
	NSA 29
	FEMA 26
	D.C. 23
	VP 18
	CIA 15
	EKG 14
	U.S. 14
	DOJ 12
	D.N.C. 12
	EPA 11
	GOP 11
	PAC 11
	CNN 11
	NRA 10
	C.W.I. 10

Figure 2: SRT Manager, unique values and total number of occurrences.

A further important remark is that the majority of episodes start with a flashback, reminding the viewer of the previous events, which may contain acronyms, so the collected number of occurrences may be higher than in reality, but we have left the subtitles unaltered, containing the flashback texts as well.

The next stage was to save the obtained data and feed them into a spreadsheet table with various categories, during which we have detected an initial issue due to the text type. Subtitles have limited options to highlight various pieces of information, so a rather logical option is to capitalize entire words when wishing to draw the attention of readers upon particular speakers (e.g., over the telephone), names (e.g., *JACK*), jobs used instead of names (e.g., *POLICEMAN*), elevated voice, signs (e.g., *NO ENTRY*), directions, labels, word plays, puns (*CALL 1-800-BITE ME*), language, company names, etc., which are “false

alarms” for software detection. Similarly, a few Roman numbers may cause issues, such as *II*, *III*, *VII*, and *IX* (instances detected in all series). However, *IV* needs utmost care, as it may refer either to the Roman number or the *intravenous* medical procedure, commonly abbreviated as *IV* (← “intravenous”) in English, which should be translated as *i.v.* (← “intravenos”) into Romanian and *iv.* (← “intravénás”) into Hungarian, as prescribed by the European Medicines Agency.⁵ Thus, the only option is to check every single instance. Yet, the worst-case scenario is when the entire subtitle is fully capitalized, which – however unacceptable – happens in the case of the English subtitle of *House of Cards*, season 1.

The embedded deny list of the software resulted in a considerable change in figures and quality of results; for instance, in the case of *House of Cards*, the unfiltered run returned 978 results with 180 unique values, while activated deny list returned 804 results (17.8% drop) with 156 unique values (13.4% drop).

After having collected the acronyms from all the English subtitles, we have selected the top occurrences of each series, which may offer a hint to which acronyms should be rendered most consistently.

Table 1 illustrates the inconsistency regarding period use even among the top occurrences as well, in which cases the dominant version is listed first. Inconsistent period use (e.g., US/U.S., V.P./VP) is a first sign that subtitles are not consistent regarding the spelling of acronyms, not even within the same TV series, which may be explained with either ignorance or changing subtitlers. This entails that the translated subtitles should present similar cases, completed with results from the previously discussed possibilities.

Table 1: Top occurrences in the series

#	House of Cards	No.	Designated Survivor	No.	The West Wing	No.	24	No.	Blindspot	No.
1	ICO	87	FBI	191	U.S., US	280	CTU	1513	F.B.I., FBI	800
2	US, U.S.	64	US, U.S.	88	FBI	137	FBI	304	CIA	298
3	DC, D.C.	53	DC, D.C.	76	V.P., VP	130	I.D., ID	164	ZIP	151
4	FBI	53	NATO	42	MS	92	L.A., LA	147	US, U.S.	130
5	UN	40	VP, V.P.	40	U.N., UN	83	U.S., US	132	I.D.	86
6	NSA	29	SEAL	33	CNN	66	NSA	131	NYPD	71
7	FEMA	26	CIA	31	NASA	62	U.N., UN	110	NSA	64
8	VP, V.P.	21	POTUS	28	DNC	60	DOD	90	HCI	62
9	CIA	15	NSA	27	D.C., DC	57	CIP	87	DNA	48
10			NASA	26	CIA	50	CIA	75	I.D.	86

As the analysis of the Romanian and Hungarian versions is still in process, we have selected a term that appears in four TV series: *House of Cards* (26), *Designated Survivor* (2), *The West Wing* (19), and *Blindspot* (1), totaling 48 occurrences in the English version. Its Romanian and Hungarian renditions are listed in Table 2.

Although the Romanian translators offered variants for all the occurrences of *FEMA*, 17 translations cannot be listed, as the Romanian subtitles for *The West Wing* were untraceable, except for season 1. The Hungarian translators avoided the term four times, which is only 8.33% of all cases.

The number of variants reflects the struggle to render the original acronym, leaving space for improper variants as well, marked with an asterisk. The multiple versions clearly signal that there is not an established variant for the American acronym, thus similar words try to explain it. The original acronym was used in nearly half of the Romanian and almost 75% of the Hungarian cases, although no explanatory version was found during the first occurrence in neither language. Although the equivalent Romanian acronym is *DSU* (← “Departamentul pentru Situații de Urgență”)⁶ and the Hungarian is *OKF* (← “Országos Katasztrófavédelmi Főigazgatóság”),⁷ the translators avoid using these terms for two possible

⁵ https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/tables-non-standard-abbreviations-be-used-summary-product-characteristics_en.pdf, April 25, 2022.

⁶ <http://www.dsu.mai.gov.ro/>, April 25, 2022.

⁷ <https://www.katasztrofavedelem.hu/2/bemutakozas>, April 25, 2022.

Table 2: Translations of FEMA

FEMA (← Federal Emergency Management Agency)	
Romanian	
1. FEMA (22)	
2. fonduri FEMA (3)	← “FEMA funds”
3. fonduri de urgență (3)	← “emergency funds”
4. fondul pentru dezastre al FEMA (1)	← “FEMA’s funds for disasters”
5. Agenția pentru Urgențe (1)	← “Emergency Agency”
6. Crucea Roșie* (1)	← “Red Cross”
Hungarian	
1. FEMA (34)	
2. Katasztrófavédelem (4)	← “Emergency/Disaster management”
3. katasztrófaelhárítók (2)	← “disaster management team” / “emergency response team”
4. FEMA katasztrófavédelmi alap (1)	← “FEMA’s disaster management funds”
5. katasztrófavédelmi alap (1)	← “disaster management fund”
6. Tűzoltóság* (1)	← “The fire department”
7. Vöröskereszt* (1)	← “Red Cross”
8. - (4)	

reasons: on the one hand, the Romanian and Hungarian terms are not connected to the US setting, while the other reason might be risk aversion, as it is questionable whether the target audience is familiar with these acronyms. Furthermore, there is also the *Romanian General Inspectorate for Emergency Situations* (→ *IGSU*)⁸ with similar attributes.

The findings show that the larger the database, the more inconsistencies are to be expected referring to terminology, which is not necessarily a negative remark, if equally acceptable solutions are found. However, readers might be puzzled on facing various renditions for one original term, which draws the attention away from following the storyline. Subtitle lines should contain fewer than 43–45 characters per line (depending on various sources), and no more than two lines are recommended, not to obscure the image (Carroll and Ivarsson 1998, Williams 2009).

6 Conclusions

Although various subtitle guidelines offer recommendations for character number per line and specific subtitling software counts the number of characters (e.g., *CaptionHub* for TED Translators or *Subtitle Edit*), these cannot guarantee that rules are followed. As explained, “[s]ubtitling conventions are not set in stone” (Cintas 2005, 16).

Whenever possible, both the Romanian and Hungarian subtitlers rely on the original acronym (e.g., *FBI* and *CIA*); in other cases, they offer variants: *DC* may be preserved, rendered as “the capital” (Ro. *capitală*), completed with *Washington*, or even changed to *Washington*. While explicitation results in more characters, it is the best strategy against “alienating” the audience (Hales et al. 2017, 23), knowing that the uttermost function of any subtitle is to offer an aid to following and better understanding the storyline. Hence, a particular challenge for subtitlers is the popularity of TV series, which may contain hundreds of acronyms, all of which should be saved into a database and re-used over and over again, or consciously develop viable alternatives, even if “it is very difficult to construct a general and up-to-date database of acronym-definition repository” (Sánchez and Isern 2011, 312).

⁸ <https://www.igsu.ro/>, April 25, 2022.

At present, it is questionable whether the second subtitler commissioned to create the subtitles of season 2 of any TV series considers the solutions in season 1, which will be instantly observed by avid fans and frowned upon. Very popular TV series aired on mainstream TV channels or streaming services set the new standards in subtitling as well; thus, quality assurance on their behalf is a must. Yet, it is visible that both transcripts and subtitles need much more improved versions, as their quality is often behind expectations. There is no place for fully capitalized subtitles, and careful proofreading should eliminate the types of mistakes spotted in the case of *FEMA*.

While we have tried to define and categorize abbreviations, acronyms, and other word-shortening processes that visually differ from the “normal” word by being full uppercase letters (at least the prototypical ones), it is visible that algorithms developed to track them have limited functionality, as relatively many “false alarms” (e.g., capitalized names, jobs, signs, etc.) are detected. However, a careful recheck may eliminate them by creating special deny lists for each TV series separately to offer statistically more relevant cases.

As both categorization and translation of abbreviations are challenging, further research is needed to collect data from other areas as well, such as mass and social media, which both popularize and create a large proportion of recently circulating abbreviations.

Acknowledgments: This research would have been impossible without the constant support of Nándor Makkai, who designed, developed, and perfected the software to cover all the real needs. Furthermore, I am indebted to Krisztina Makkai, who helped in testing the software in all the initial stages.

Conflict of interest: The author states no conflict of interest.

Data availability statement: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

- Alonso, Pedro Ignacio. 2008. “Acronym.” *AA Files*, 57, 13–16.
- al-Qinai, Jamal. 2007. “Abbreviation and acronymy in English Arabic translation.” *Meta* 52(2), 368–75.
- APA. 2019. *Publication Manual of the American Psychological Association*. 7th edition. Washington, DC: American Psychological Association.
- Barnett, Adrian, and Zoe Doubleday. 2020. “The growth of acronyms in the scientific literature.” *eLife* 9, 1–10.
- Barnett, Adrian, and Zoë Doubleday. 2021. “Demonstrating the ascendancy of COVID-19 research using acronyms.” *Scientometrics* 126, 6127–30.
- Bauchner, Howard, and the JAMA Network. (eds). 2020. *AMA Manual of Style*. 11th edition. New York: Oxford University Press.
- Begg, Colin B. 2017. “Zero tolerance for acronyms.” *Clinical Trials (London, England)* 14(6), 561–62.
- Bloom, D. A. 2000. “Acronyms, abbreviations and initialisms.” *BJU International* 86(1), 1–6.
- Brugman, Claudia M. 1988. *The Story of “Over”: Polysemy, Semantics and the Structure of the Lexicon*. 1st ed. Garland.
- Cannon, Garland. 1989. “Abbreviations and acronyms in English word-formation.” *American Speech* 64(2), 99–127.
- Ivarsson, Jan and Carroll, Mary. 1998. *Subtitling*. Simrishamn: TransEdit.
- Caon, Martin. 2016. “Abbreviations, initialism and acronyms: their use in medical physics (THUMP).” *Australasian Physical & Engineering Sciences in Medicine* 39(1), 11–12.
- Carter, Ronald, and Michael McCarthy. 2006. *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge England; New York: Cambridge University Press.
- Comisia Europeană. 2014. *Ghid practic al Departamentului de limba română din Direcția Generală Traduceri a Comisiei Europene*. Bruxelles.
- Cintas, Jorge Díaz, and Aline Remael. 2020. *Subtitling: Concepts and Practices*. London: Routledge.
- Cojocaru, Leonid. 1976. *Dicționar militar englez-român*. București: Editura Militară.
- Dannewitz Linder, Mats. 2016. *SDL Trados Studio 2017. The Manual*. Norrtälje: Nattskift Konsult.
- Deme, László. 1955. “Betűszavaink használatához.” *Magyar Nyelvőr* 79(4), 397–403.
- Díaz-Cintas J. 2005. “Back to the future in subtitling.” *MuTra* 2005, 16–32.

- HaCohen-Kerner Y., A. Kass, and Peretz A. 2004. "Baseline methods for automatic disambiguation of abbreviations in Jewish law documents." In *Advances in Natural Language Processing, 4th International Conference, ESTAL*, edited by J. L. Vicedo, P. Martínez-Barco, R. Muñoz, and M. S. Noeda, p. 58–69, Alicante: Springer.
- HaCohen-Kerner, Yaakov, Kass Ariel, and Peretz Ariel. 2013. "Initialism disambiguation: man versus machine." *Journal of the American Society for Information Science and Technology* 64(10), 2133–48. doi: 10.1002/asi.22909.
- Hales, Andrew H., Kipling D. Williams, and Joel Rector. 2017. "Alienating the audience: how abbreviations hamper scientific communication." *Observer* 30(2), 22–4.
- Izura Cristina, and David Playfoot. 2012. "A normative study of acronyms and acronym naming." *Behavior Research Methods* 44(3), 862–89.
- Jacobs, Kayla, A. Itai, and S. Wintner. 2018. "Acronyms: identification, expansion and disambiguation." *Annals of Mathematics and Artificial Intelligence* 88, 517–32.
- Jan Ivarsson, and Carroll, Mary. 1998. "Subtitling." Simrishamn: TransEdit. *The Code of Good Subtitling Practice*. <https://esist.org/wp-content/uploads/2016/06/Code-of-Good-Subtitling-Practice.pdf>.
- Kasproicz, Małgorzata. 2010. "Handling abbreviations and acronyms in medical translation." *Translation Journal* 14(2). <https://benjamins.com/online/etsb/publications/19949>.
- Koelsch, George. 2016. *Requirements Writing for System Engineering*. Herndon: Apress.
- Kortmann, Bernd. 2020. *English Linguistics: Essentials*. 2nd revised, updated and enlarged edition. Berlin Heidelberg: J.B. Metzler.
- Kövecses, Zoltán. 2002. *Metaphor: A Practical Introduction*. Oxford: Oxford University Press.
- Kreidler, Charles W. 2000. "Clipping and Acronymy." In *Morphologie/Morphology: An International Handbook on Inflection and Word-Formation*. Vol. 1, edited by G. E. Booij, C. Lehmann, J. Mugdan, W. Kesselheim, and S. Skopeteas, p. 956–63. Berlin/New York: Walter de Gruyter.
- Kuzmina, Olga D., Anna D. Fominykh, and Natalia A. Abrosimova. 2015. "Problems of the English abbreviations in medical translation." *Procedia – Social and Behavioral Sciences* 199, 548–54.
- Lakoff, George. 1987. *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: The University of Chicago Press.
- Laszlo, Sarah, and Kara D. Federmeier. 2007. "The acronym superiority effect." *Psychonomic Bulletin & Review*, 14(6), 1158–63.
- López Rúa, Paula. 2004. "Acronyms & Co.: a typology of typologies." *Estudios Ingleses de La Universidad Complutense* 12, 109–29.
- Ludányi, Zsófia. 2013. *A mai orvosi helyesírás*. Budapest: Eötvös Loránd Tudományegyetem. PhD Thesis.
- Mattiello, Elisa. 2013. *Extra-grammatical Morphology in English: Abbreviations, Blends, Reduplicatives, and Related Phenomena*. Berlin/Boston: Walter de Gruyter.
- Milinković, Slađana. 2019. "Translating degrees and academic titles abbreviations: challenges and perspectives." In *A diszciplínák találkozása: nyelvi közvetítés a XXI. században*, edited by S. Szoták, p. 146–67. Budapest: OFFI Zrt.
- Nicoll, Leslie H. 2016. "Abbreviations, initialisms, and acronyms: guidance for authors." *Nurse Author & Editor* 26(4), 1–9.
- Panajotu, Kosztasz. 2010. "Abbreviations and acronyms in military English." *AARMS* 9(1), 159–65.
- Park, Youngja, and Roy J. Byrd. 2001. "Hybrid text mining for finding abbreviations and their definitions." In *Proceedings of Conference on Empirical Methods in Natural Language Processing, EMNLP (Empirical Methods in Natural Language Processing)*, edited by L. Lee and D. Harman, p. 126–33. Pittsburgh: Intelligent Information Systems Institute.
- Péter, László. 2003. "Levélszekrény." *Magyar Nyelv* 99(1), 125–6.
- Quirk, Randolph, Greenbaum Sidney, Leech Geoffrey, and Svartvik Jan. 1985. *A Comprehensive Grammar of the English Language*. London; New York: Pearson Longman.
- Rosch, Eleanor. 1975. "Cognitive representation of semantic categories." *Journal of Experimental Psychology General* 104, 192–233.
- Sánchez, David, and David Isern. 2011. "Automatic extraction of acronym definitions from the web." *Applied Intelligence* 34(2), 311–27.
- Scarpa, Federica. 2020. *Research and Professional Practice in Specialised Translation*. UK: Palgrave Macmillan.
- Soyer, P. 2018. "Acronyms, initialisms, and abbreviations." *Diagnostic and Interventional Imaging* 99(10), 589–90.
- Taghva, Kazem, and Jeff Gilbreth. 1999. "Recognizing acronyms and their definitions." *International Journal on Document Analysis and Recognition* 1999(1), 191–8.
- Thomas, C. George. 2021. "Punctuation marks and abbreviations." In *Research Methodology and Scientific Writing*, edited by C. G. Thomas, p. 455–73. Cham: Springer International Publishing.
- Trumble, William R., and Angus Stevenson. (eds). 2002. *Shorter Oxford English Dictionary*. Vol. I. A-M. 5th ed. Oxford: Oxford University Press.
- Wallwork, Adrian. 2014. *User Guides, Manuals, and Technical Writing: A Guide to Professional English*. New York: Springer.
- Williams G. F. (ed.). 2009. *BBC's Online Subtitling Editorial Guidelines V1.1*.
- Yule, George. 2010. *The Study of Language*. Cambridge: Cambridge University Press.
- Zahariev, Manuel. 2004. "A linguistic approach to extracting acronym expansions from text." *Knowledge and Information Systems* 2004(6), 366–73.
- Zimmer, Ben. 2010. "Acronym." *The New York Times* (December 17).

Webography

<http://www.dsu.mai.gov.ro/>, April 25, 2022.

<https://aspe.hhs.gov/common-acronyms>, April 24, 2022.

<https://www.acronymfinder.com/CCC.html>, April 22, 2022.

<https://www.degruyter.com/journal/key/opli/html>, Instructions for Authors, April 24, 2022.

https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/tables-non-standard-abbreviations-be-used-summary-product-characteristics_en.pdf, April 25, 2022.

<https://www.igsu.ro/>, April 25, 2022.

<https://www.katasztrofavedelem.hu/2/bemutakozas>, April 25, 2022.