



HAL
open science

How proper names enter the lexicon – From the Central Intelligence Agency to CIA: the reduction of full denominative expressions interpreted as a means to achieve (greater) lexicalization

Elise Mignot, Manon Philippe

► To cite this version:

Elise Mignot, Manon Philippe. How proper names enter the lexicon – From the Central Intelligence Agency to CIA: the reduction of full denominative expressions interpreted as a means to achieve (greater) lexicalization. *Lexis. Journal in English Lexicology*, 2022, Proper names and the lexicon, Elise Mignot and Manon Philippe (eds.), 20, 10.4000/lexis.6672 . hal-03808847

HAL Id: hal-03808847

<https://hal.science/hal-03808847>

Submitted on 1 Mar 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

How proper names enter the lexicon – From *the Central Intelligence Agency* to *CIA*: the reduction of full denominative expressions interpreted as a means to achieve (greater) lexicalization

Manon Philippe and Élise Mignot



Electronic version

URL: <https://journals.openedition.org/lexis/6672>

DOI: [10.4000/lexis.6672](https://doi.org/10.4000/lexis.6672)

ISSN: 1951-6215

Publisher

Université Jean Moulin - Lyon 3

Electronic reference

Manon Philippe and Élise Mignot, "How proper names enter the lexicon – From *the Central Intelligence Agency* to *CIA*: the reduction of full denominative expressions interpreted as a means to achieve (greater) lexicalization", *Lexis* [Online], 20 | 2022, Online since 28 December 2022, connection on 04 January 2023. URL: <http://journals.openedition.org/lexis/6672> ; DOI: <https://doi.org/10.4000/lexis.6672>

This text was automatically generated on 4 January 2023.



Creative Commons - Attribution-ShareAlike 4.0 International - CC BY-SA 4.0

<https://creativecommons.org/licenses/by-sa/4.0/>

How proper names enter the lexicon – From *the Central Intelligence Agency* to *CIA*: the reduction of full denominative expressions interpreted as a means to achieve (greater) lexicalization

Manon Philippe and Élise Mignot

Introduction

- 1 Proper names are different from the rest of the lexical units, or so it seems on the surface. With their capitalised initials and apparent lack of visible determination, for most of them, and their constraints when it comes to pluralisation and modification, they come across as “less than” or “far from” perfect nouns. Many proper names are not even nouns, morphologically speaking – like *Hired* (name of an online marketplace), *Lost* (name of a TV show), *Grumpy* (name of a dwarf in *Snow White*) which are verbs and/or adjectives; sometimes, they look shorter than a noun, or any word-level unit, as is the case for *T* (the first name of a person in the essay *Evicted*, by Matthew Desmond); very often, they contain more than one word and may be analysed in terms of syntactic sequences as well as, or instead of, a complex word – *United Kingdom* (name of a political institution, a common noun and its deverbal pre-modifier), *I’m Still In* (name of an anti-Brexit association), *Crossing the Rubicon* (name of an art installation) which can both be analysed as having the internal organisation of a clause, even when inserted into a noun phrase and a larger clause, e.g., *I’m Still In organised a series of protests across the United Kingdom last week that culminated with the display of Crossing the Rubicon, an artwork by British artist deCaires Taylor*, invented example). Add to that their preferred referential profiling of unique entities dependent on context, compared to the

predicative logic of common nouns, which signify categories of objects (outside context) rather than referring to individuals, and it is no wonder that grammars and dictionaries have been treating proper names as a separate linguistic category (Gary-Prieur [1991]).

- 2 This would explain why proper names appear in a different section of dictionaries, separated from the rest of the words, when they are not excluded from dictionaries altogether and relegated to encyclopaedias. The move is significant, since the former deal with linguistic units, considered insofar as they are part of the available lexicon in a given language, whereas the latter compile descriptions of cultural objects. Nevertheless, however contextually bound or culturally relevant the referents of proper names are, proper names themselves should be given a lexical characterisation, or at least their lexical status should be discussed (Vaxelaire [2006: 3]). There are of course lexicographic dictionaries of proper names, in which lemmatization and linguistic information (morphosyntax, spelling variants, etymology and diachronic evolution) have pride of place, but they are used to widen the gap between proper names and the rest of the lexicon:

The concept of a proper noun as an independent lexical unit is important, since it enables onomasticians to recognize it, especially for lemmatization purposes, as a lexical item with attributes that are different from all other lexical items, and thus to be treated separately from the specific grammatical rules of a language. [Möller 1995].

- 3 Beyond lemmas, there remains the question of treating them as lexemes, along with the rest of the lexical units in the “common” dictionaries, that is, dictionaries that do not specialize in the onomasticon but include common nouns, verbs, adjectives, and other non-proper-name units.¹ Some proper names do make it into these dictionaries, and we expect to find them, either with a fully assumed proper name status among “common units”, like a necessary exception, or as a deproprialised unit (see Section 1.). What we are more interested in are clues that perhaps, some proper names are already integrated, or on their way to being integrated, into the lexicon (as is represented by these dictionaries) but without as much conscious effort to retain the proper name status or to “force” it to behave like an already established lexical category.
- 4 The aim of this article is to examine one entry point that may lead to the integration of proper names into the lexicon: when morpho-phonological reduction is involved. We will argue that the reduction process, which does not solely concern proper names, helps them and other units to reach the first step to being accepted as lexemes. We use the lexicographic treatment of abbreviated units in the dictionary as a starting point for a truly lexicological reflection on the extent to which the morphological reduction of a proper name may or may not lead to the creation of a new lexeme, included in the usual lexicon.
- 5 We start with the context of the research and the method we used to define our object of study – non-semantically modified proper names occurring as part of the *Concise Oxford English Dictionary* (Section 1.). We then confirm the presence and profile of these reduced proper names that entered the dictionary under ambiguous pretenses, the *abbr.* label (Section 2.). We offer the hypothesis that reduction is a key factor enabling the integration of proper names into the lexicon (not just that particular dictionary) *via* word-ification, opacification and syntactic differentiation, which we confirm with a

case study on the distinction between *Central Intelligence Agency* (full form, proper name) and *CIA / C.I.A.* (reduced forms, abbr., new lexical items) in the *COCA* (Section 3.).

1. Reduction and proper names in the lexicon: scope and relevance of the study, corpus, method and research questions

- 6 Following previous research (Mignot [2017]) based on the *Concise Oxford English Dictionary Twelfth Edition* [2011] [COED], we chose it as our corpus of study and manually retrieved 2835 entries labelled “abbr.” where a part-of-speech would be for other items in the dictionary. We then annotated the proper name status (PrN) of the corresponding full forms when relevant. For each entry, proper name or not, we analysed the type of reduction process or reduction category the abbreviated form illustrated.
- 7 Our two main exploratory questions were:
- To what extent are proper names inserted into the lexicon (represented by the list of lexemes present in a dictionary) along with the rest of abbreviated units?
 - To what extent are some reduction processes preferred by, reserved for or available to proper names?
- 8 The rest of this section discusses the relevance of abbreviations in the analysis of proper names and the questions that general observations raised, the corpus chosen, and the criteria used to annotate the retrieved abbreviations.

1.1. Why focus on abbr. units in a study on proper names: the place of proper names in dictionaries, abbreviations and lexical status

- 9 Whether or not to include proper names is a real question lexicographers have to ask themselves when constructing their dictionaries. Atkins & Rundell [2008: 179] describe a shift in lexicographic practices from two separate headword lists (meaning two parts in the dictionary, with proper names separated and at the end) to a single headword list (meaning the inclusion of some proper names in the main part of the dictionary), sacrificing “a ‘purer’ headword list” in order to help users navigate the dictionary.
- 10 In practice though, not all lexicographers choose to include proper names, and the reasons why they do include some and not others involve the analysis of the grammatical structure of the resulting lexeme and the referential profile of proper names. In a series of semi-conducted interviews with 23 lexicographers, Bozkurt [2019] showed that a fourth of the respondents did not include proper names in their headword lists, while the remaining lexicographers included some proper names if they were:
- Derivatives: n. and adj. derived from a proper name, such as gentiles, religious doctrines and members of religious communities, characteristic of / relating to proper names: *American, Catholic, Shakespearian*...
 - Part of a compound or an idiom: *Adam’s apple, Aladdin’s cave, Brussels sprouts, Dear John letter, Tay-Sachs disease*...
 - Very common or culturally salient: names of legendary or mythical characters, names of famous places (countries, important cities / major cities, counties, provinces, regions,

rivers), metonyms (*White House, Big Apple*), unique objects (*Sun, Earth, Moon* can have common noun or proper name uses), religious / characteristic uses (*Virgin Mary*), brand names / trademarks when they illustrate “deproprialisation” (loss of unique reference and application to the set of objects sold under the brand’s name).

- 11 The list of criteria and categories could go on (see Atkins & Rundell [2008: 186–189]. Suffice it to say that as long as the proper name is seen as either less proper in itself (“deproprialisation” through derivation or brand names, proper name antonomasia – metaphor: *he is a real Shakespeare*, metonymy: *Bordeaux*, for a type of wine –, so-called generic names: *Tom, Dick and Harry*) and/or not standing on its own and therefore part of a bigger lexeme (derivatives, compounds, idioms), it could be justified to include it within the common lexicon. More subjective or pragmatic criteria such as cultural relevance may lead to the inclusion of prototypical proper names but these will stand out as acknowledged exceptions, or they will surprise the linguist if they are labelled as nouns (n.) like the rest of common nouns. In any case, these will spark or have already sparked the attention of lexicographers and linguists alike.
- 12 However, reduced forms have not, to our knowledge, been given pride of place. Atkins & Rundell [2008: 120] only ask which form (full or reduced) should be headword in case the reduced form is the most used (*BBC, CIA*) and, more relevant to our study, they mention that an abbreviation is as much a lexical item as a noun or a verb or anything in a full form. But a question remains: while full forms attract attention because they need to be analysed and given a part of speech, which triggers discussion, reduced forms may in some dictionaries be only labelled as abbreviations without discussing their POS, which allows quite a few headwords to fly under the radar and be integrated into the lexicon without further discussion of their lexical status. This is all the more interesting since other units, which are reduced forms, are given a POS (usually n. noun) while their abbreviated status is given in the definition’s additional information; or since the same reduced form can stand for full forms that would not belong to the same syntactic category. In order to illustrate the unstable treatment of abbreviated forms in the *COED*, two examples of homonymous forms, *GI/Gi* and *AA* and their definitions, are reproduced in Table 1.

Table 1. Entries *GI1*, *GI2*, *gi* and *AA* in the *COED*

<p>GI¹ ● n. (pl <i>GIs</i>) a private soldier in the US army.</p> <p>ORIGIN 1930s (orig. denoting equipment supplied to the US forces): abbrev. of government (or general) issue.</p> <p>GI² ● abbr. glycaemic index.</p> <p>Gi /gi:/ ● n. the loose white jacket worn in judo.</p> <p>ORIGIN: from Japanese.</p>	<p>AA ● abbr. 1 Alcoholics Anonymous. 2 anti-aircraft. 3 Automobile Association</p> <p>1 is a proper name</p> <p>2 is a common noun</p> <p>3 is a descriptor, a descriptive part of a proper name</p>
--	--

- 13 To determine whether this apparent confusion in lexicographic conventions was common practice or if it was restricted to the *COED*, we observed what was done in other dictionaries.

1.2. Comparison with other dictionaries: beyond lexicographic conventions and inconsistencies, questioning lexicality and n. status

- 14 It turns out that we could have come up with the same interrogations using another dictionary. We initially chose to work with the *COED* because our research project was meant as a follow-up on previous research on nominal units (Mignot [2017]), for which the general framework for data collection and analysis had already been set. Collecting data from the *COED* suited our research aims, being a self-proclaimed “accessible, convenient dictionary of the current language” [*COED* 2011: ix] drawing on actual use but designed to answer the needs of the average English speaker (compared to the complete *OED*). As such, it was supposed to be less specialised and lexicographically complex than research-oriented lexicographic work, but slightly more nuanced and complete than advanced learner’s dictionaries whose aim is to filter the English language in order to give more importance to explanations of context, idiomaticity and writing tips to non-native speakers of English. In all cases, conventions used regarding the inclusion of proper names or abbreviations is first and foremost a question of choices made to fully accept this or that unit as a noun, verb, adjective, etc., or to leave the question unanswered: whatever dictionary is used, what matters is the (in)consistency behind these choices and what the differences in treatment reveal about the linguistic interpretation of these units.
- 15 We compared the treatment of some units labelled *abbr.* in the *COED* in five other dictionaries:
- the *Oxford Advanced Learner’s Dictionary of Current English* [*OALDCE*],
 - the *Penguin Concise English Dictionary* [*PCED*],
 - the *Macmillan English Dictionary for Advanced Learners* [*MEDAL*],
 - the *Oxford Advanced Learner’s Dictionary* [*OALD*],
 - and the *Collins COBUILD Advanced Learner’s Dictionary* [*COBUILD*]
- 16 The *OALDCE* seems to follow the older conventions of not including (too many) proper names or non-prototypical lexical units in the main part of the dictionary. A selection of Appendices at the end of the book lists irregular verbs (forms), common abbreviations, affixes, numerical expressions, weights and measures, geographical names, common forenames, Shakespeare’s works, punctuation and books of the Bible (no definition, just the list, phonological transcription and typography when necessary). However, fully formed proper names do make it into the main part of the dictionary, either as headwords within an idiom (*Achilles* n. The heel of ~; ~’s heel (fig) small but weak and vulnerable point, e.g., in sb’s character: Spelling is my ~’s heel.), or additional information within an entry (*bailey* /'beɪlɪ/ n outer wall of a castle; courtyard of a castle enclosed by strong walls. Old B~, London Central Criminal Court.). Other proper names with metaphoric or metonymic meaning appear (*Amazon*, *Armageddon*, *Babel*, *Bordeaux*...) as well as full-fledged proper names such as *America*, *Britain*, *Argus*, etc., all marked n. The other dictionaries included abbreviated terms within the dictionary so the logic was slightly different. The *Penguin Concise* and *Oxford Learner’s Dictionary* seem to consistently use *abbr* as a label just before giving the full form (and explanation if needed), whereas the *Macmillan English Dictionary* and the *COBUILD Learner’s Dictionary* sometimes indicate the POS of the full form (e.g., *Asap* adv; *A&E* noun [C/U]; *A&E* N-UNCOUNT) while inconsistently indicating in the definition that it

is an abbreviation, and sometimes fail to label the lexical unit altogether, providing the full form without necessarily mentioning the abbreviated status. When the *MacMillan* includes reduced proper names, the likelihood that there be no label or POS or mention of abbreviation is higher than for other reduced forms (see Appendix 1 for a comparison of entries in these dictionaries for various types of reduced units). Interestingly, the size of the dictionary did not influence the inclusion of reduced proper names: in the sample of reduced units compared, the *Penguin Concise*, the smallest dictionary, included reduced proper names more frequently than the other (learner's) dictionaries, which however included more written abbreviations (*Asst, Ave.*) to guide the learner's reading / writing skills. Overall, they appear to contain much fewer abbreviations than the *COED*, while being more specific about the role the abbreviation will fulfil (POS: word in a sentence; abbr. or mention of abbreviation: clues to understanding written conventions, not to be used in a normal sentence).

- 17 These comparisons confirm three main research questions that this paper would like to address:
- the tendency, whenever abbr. is used, to suspend any discussion on the syntactic category of the reduced unit, which leaves room for interpretation of a unit as fully lexical because it looks like an existing word (e.g., *GI, GI, gi*);
 - the tendency to label the reduced form n. by default when it is not clear that usage makes it a verb / adverb / adjective, meaning n. becomes the norm, regardless of how many nominal criteria the reduced unit complies with;
 - the need to understand the different roles of abbreviations in a dictionary and a discussion of what it means to be a lexical unit.
- 18 Our first step was to determine whether the reduced units we retrieved were proper names or not.

1.3. Proper name status in the corpus study: criteria

- 19 Each reduced entry was annotated, in the proper name column: yes OR no OR PrN-adjacent, as reproduced in Table 2.

Table 2. Excerpt from our corpus, annotating the proper name status of the reduced units *RC*

RC	abbr.	PrN-adjacent	(in cycling) <i>racing club</i>
RC	abbr.	yes	<i>Red Cross</i>
RC	abbr.	no	<i>reinforced concrete</i>
RC	abbr.	no	<i>resistance/capacitance (or resistor/capacitor)</i>
RC	abbr.	PrN-adjacent	<i>Roman Catholic</i>

- 20 In order to be annotated as a proper name, the full form must refer to a unique individual (animate or inanimate: person and juridical person – institution / organization / association / union / etc., place, time / event / period, some products....) and will in all likelihood bear an initial capital letter on each of its lexical elements:

Alberta (Al.), Drug Enforcement Administration (DEA), February (Feb.), Hampshire (Hants), Oregon (OR, Ore.), Red Cross (RC), World Trade Organisation (WTO). We have included the reductions of names of days and months in what we considered to be proper names, even though their full form is included in the dictionary with the POS n. In this article, we keep to the expression “proper names” because many reduced units are longer than proper nouns and may not contain proper nouns at all, but what we describe is not a full noun phrase, so we ought to call them “proper nominals” following Philippe [2020].

- 21 The “PrN-adjacent” status means the unit does not qualify as a prototypical proper name for closely related reasons:

- Because the linguistic unit does not form a proper name alone

The unit *racing club* is a descriptor (Huddleston & Pullum [2002: 519]) which can be considered either part of a proper name or an embellishment of the proper name depending on whether or not it can be omitted. It remains that outside the proper name, *racing club*, *Const.* or *PC (Police Constable)*, or even *Ave. (Avenue)*, *Cllr (Councillor)*, *FC (Football Club)*, *Hon (Honourable)*, *RH (Right Honourable)*, *Jun (Junior)*, etc. are not proper names themselves so we did not treat them as such.

- Because the reference potential leans more towards the categorising logic of common nouns and properties of verbal / adjectival predicates than the uniquely-identifying logic of proper names

Previously mentioned *Const.* or *PC*, *Cllr*, being descriptors, suggest a role that can be filled by countless different individuals. In that respect, full forms that denote a role / title / membership using a “description”, even if it is institutionalised, even if it is meant to be filled by one person at a given time, have been regarded as lacking an essential attribute of prototypical proper names: *Cmdr (Commander)*, *POTUS (President of the United States)*... This means that if there is a reduced form for the name of the institution and one for its member, the former has been labelled “yes” and the second “PrN-adjacent”: *RCP (Royal College of Physicians: yes, proper name)* vs. *MRCP (Member of the Royal College of Physicians: “PrN-adjacent”)*.

- 22 Following the same logic, full forms of units denoting establishments that are meant to be duplicated are “PrN-adjacent” but not quite proper names (they are either descriptors in or larger proper names or descriptions of “a type of...”): *LEA (Local Education Authority)*, *SCR (Senior Common Room)*, *SSSI (Site of Special Scientific Interest)*. The same goes for descriptions of diplomas and degrees (*Bachelor of...*), medals and rewards (*VC – Victoria Cross*), internet or software protocols that can be and are duplicated in as many recipients. The last *RC* in Table 2 stands for *Roman Catholic*, which combines all these aspects: it denotes a property of an institution and can be used as a descriptor (e.g., *Roman Catholic Church*, name of the institution). Its adjectival ending also makes it a derivative that would be accepted, in its full form, in most dictionaries now, probably because the derivation has achieved enough “depropriation” of the unit.
- 23 We retained the “PrN-adjacent” annotation to account for the fact that there is a significant link with a proper name, but that the distance is too great: while we understand that there could be degrees of properhood, and that these, along with all derivatives (*American, Shakespearean, Roman Catholic...*) could be understood as secondary proper names in Algeo [1981]’s terms, we prefer to separate them from the rest of the more prototypical proper names, simply because they are less problematic and more likely to be accepted into the lexicon.

- 24 All other units annotated “no” are expressions with no connection to proper names, such as *reinforced concrete*, *resistance/capacitance (RC)*, *rest in peace (RIP)*, *rolling on the floor (RO(T)F)*, *round the world (RTW)*, *singular (sing.)*, *edited by (ed.)*, etc.
- 25 We then proceeded to annotate the type of reduction illustrated in proper name and non-proper name reduced units.

1.4. Reduction processes

- 26 The units were annotated as follows:

- Initialism

If the reduced form contains the initial letter of a number of words / bases / syllables of the full form and that each of the initials is pronounced separately: *a/c* /,eɪ'si:/ *account*, *HFC* /,eɪtʃeɪ'si:/ *hydrofluorocarbon*, *WTO* /,dʌbəlju:ti:'əʊ/ *World Trade Organisation...*

- Acronym

If the reduced form contains the initial letter of a number of words / bases / syllables of the full form and the whole unit is pronounced as if it were a simple word: *DoS* /'dɒs/ *denial of service*, *HAART* /'hɑ:t/ *highly active antiretroviral therapy*, *NAFTA* (also *Nafta*) /'næftə/ *North American Free Trade Agreement*, *Waaf* /'wæf/ (*member of the*) *Women's Auxiliary Airforce...*

- Clipping

If the reduced form contains consecutive phonological parts of the full form: *Med* /'med/ *Mediterranean Sea*, *nor* /'nɔ:/ *north* (especially in compounds: *nor'west*); we also include units where clipping occurred in several elements of the full form (clipped compounds (Bauer [2012])): *Amex* /'æmeks/ *American Stock Exchange*, /'eɪmeks/ *American Express*, *COMINT* /'kɒmɪnt/ *communications intelligence*, *Lib Dem* /,lɪb'dem/ *Liberal Democrat*, *Ofcom* /'ɒfkɒm/ *Office of Communications*. It may also include clipping of the original base (*Oxon*, from *Oxoniensis*, meaning: (*of or from*) *Oxford University*).

When the reduced form illustrated one reduction process or the other, or two reduction processes at the same time, they were labelled accordingly, e.g., *ZANU-PF* /'zɑ:nu:pi:,ef/ *Zimbabwe African National Union (Patriotic Front)*: acronym + initialism, *SRAM* /,es'ræm/ *static random-access memory*: initialism + acronym, *Ofgem* /'ɒfdʒem/ *Office of Gas and Electricity Markets*: clipping + acronym, *BMus* /,bi:'mʌz/ *Bachelor of Music*: initialism + clipping, *MusB* /,mʌz'bi:/ *music bachelor*: clipping + initialism, *UFO* /'ju:fəʊ/ or /,ju:ef'əʊ/ *unidentified flying object*: acronym or initialism. They will be referred to as “combinations of reduction processes” in the rest of this paper.

- Orthographic shortenings

If the reduced form is only meant to appear in a written form for conventions, formulas, symbols, to save space, etc., without a phonological counterpart (read aloud, the full form would be heard): *AL* and *Ala.* (*Alabama*, the former: in official postal use), *Sqn Ldr* (*Squadron Leader*), *v* (*volt(s)*, *verse*, *verb*, *verso*, *very*, (in textual references) *vide*), *Warks.* (*Warwickshire*), *Wash.* (*Washington*), *Wed.* or *Weds.* (*Wednesday*), *WLTM* (*would like to meet*); *Zech.* (*Zechariah* (in biblical references))...

- 27 Some shortenings look very much like orthographic clippings or initialisms; however, since they cannot illustrate loss of phonological material (due to their lack of phonological counterpart), we decided to separate them from what we labelled “clipping”, “initialism”, “acronym” (following Plag [2003] rather than Devereux [1984]). The *COED* does not provide a phonological transcription for all abbr. units, so we

checked the pronunciation of the reduced form for each definition in pronunciation dictionaries (*Longman Pronunciation Dictionary*, *Cambridge Dictionary online*), in the *Wiktionary*, on *Youghlish* and in *Youtube* videos when necessary. When no phonological counterpart could be found, they were automatically labelled “orthographic shortenings”, so there is a slight chance this category is a little inflated in case of false negatives (if an existing phonological counterpart was not encountered). This category shows units that are very similar (in spelling) to other units labelled *symp.* in the *COED*, which represent chemical elements and various variables in sciences mostly, e.g., *C* (carbon), *Ca* (calcium), *Cd* (cadmium), *Cf* (californium), *Cs* (caesium); *d* (diameter or denoting a small increment in a given variable); *e* (the transcendental number that is the base of Napierian or natural logarithms), *v* (voltage)...

- 28 We were then able to draw a panorama of reduction processes and compare them for proper name and non-proper name full forms.

2. Results: panorama of reduction processes in the *COED*

2.1. Reduction processes

- 29 We obtain a total of 2828 items labelled as abbreviations in the *COED*. The following table indicates the respective numbers of proper names, non-proper names, and “PrN-adjacent” items.

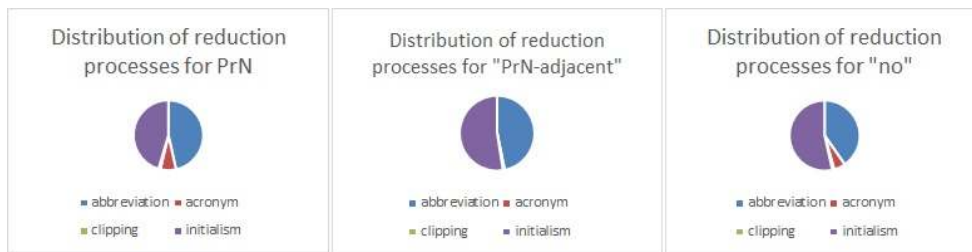
Table 3. Number of proper names, non-proper names, and PrN-adjacent items, in abbreviations listed in the *COED*

	proper name	no	PrN-adjacent	Total
Orthographic shortening	385	631	188	1204
acronym	61	84	1	146
clipping	7	14	2	23
initialism	375	838	210	1423
acronym or initialism	2	3	0	5
Combinations of reduction processes:				
acronym + initialism	1	0	0	1
initialism + clipping	0	0	16	16
initialism + acronym	0	4	0	4
clipping + initialism	1	1	2	4
clipping + acronym	2	0	0	2

Total	834	1575	419	2828
-------	-----	------	-----	------

- 30 It has to be acknowledged, however, that the dictionary also includes some acronyms, clippings, or initialisms that are not labelled as abbreviations, but are rather classified under a part-of-speech label such as adjective, noun, or verb, and are therefore treated as regular lexemes in all respects. For instance, *lab* is described as a noun when it is short for *laboratory*, but as an abbreviation when it stands for *Labour*, or *Labrador*. Similarly, the lexeme *fab* is labelled as a noun in its ‘chip fabrication plant’ meaning, and as a verb when it is a reduced form of ‘produce (a microchip) in a chip fabrication plant’. The difference between these examples and the items of our list sometimes seems slight, especially as there does not seem to be any consistent and reliable printing convention associated with the label “abbreviation”. Some abbreviations are noted with a dot (*approx.* for *approximately*, *Ariz.* for *Arizona*, *arr.* for *arrives*, *Berks.* for *Berkshire*, *Apoc.* for *Apocalypse*), but that is not consistently the case, as testified by *res* for *resolution* in *high-res images*, which results from a clipping process and is labelled as an abbreviation. Despite these apparent inconsistencies, we will use our list of abbreviations as a set of data and basis for our reflection, in so far as it provides us with information as to how (in which form?) and why (for what reason? why those?) proper names enter the (common) lexicon.
- 31 As mentioned before, the list of abbreviations does indeed include proper names – which, again, are not typically listed in a dictionary. Out of the initial 2 828 items of the list, 834, i.e., 29,5%, almost a third, are proper names; a surprisingly high proportion, especially as this figure only takes into account the most prototypical proper names, and excludes items that could be described as being “PrN-adjacent”.
- 32 In the table, the items are classified on the basis of their formation process: orthographic shortening (*CA* for *California*, *dr.* for *Doctor*), acronym (*NATO* for *North Atlantic Treaty Organisation*), clipping (*Met* for *The Metropolitan Police* in London), initialism (*CIA* for *Central Intelligence Agency*). We found that a few entries may be alternatively pronounced as either acronyms or initialisms (*LOL* for *lots of laugh*, *RAF* for *Royal Air Force*, *TIFF* for *tagged image file format*, *UFO* for *Unidentified Flying Object*, *WRAFF* for *Women’s Royal Air Force*). They have accordingly been described as such and placed in a separate category (“acronym or initialism”). Other entries illustrate a combination of formation processes, namely initialism and acronymy: *CD-ROM* starts with an initialism and ends with an acronym, while *ZANU-PF* (*Zimbabwe African National Union - Patriotic Front*) illustrates the opposite combination (acronym followed by initialism). Other hybrid formation processes involve clipping: *Ofgem* (for *Office of Gas and Electricity Markets*) is an example of clipping followed by an acronym, while *Litt D* (short for *Litterarum Doctor*, i.e., *Doctor of Letters*) results from a clipping followed by an initialism.
- 33 Initialism comes out as the most represented formation process, closely followed by orthographic shortenings. This holds both for proper names and items other than proper names: we also note that the proportions do not vary significantly depending on whether or not the entry is a proper name. This suggests that proper names are not treated any differently than the other linguistic units that are abbreviated, as illustrated in Figure 1.

Figure 1. Distribution of reduction processes for “proper names”, “no” and “PrN-adjacent” items



In all three cases, initialisms and orthographic shortenings represent the vast majority of cases, in comparable (at least, not significantly different) proportions. However, we need to exclude orthographic shortenings from our results, and will now explain why.

2.2. Exclusion of orthographic shortenings from the results

- 34 Orthographic shortenings and initialisms are not created for the same reasons. As a consequence, while the second type allows us to discuss the insertion of proper names in the lexicon, the first one does not.
- 35 The probable reason why orthographic shortenings are inserted in the dictionary is that they may need decoding, being abbreviated variants of another form – yet only from a graphical (not phonological) point of view. In that sense, they cannot be taken to be distinct lexemes, not having a pronunciation of their own, in the way that the pronunciation of the abbreviated form *CIA* differs from that of the long version *Central Intelligence Agency*.
- 36 Most of them represent place names (*AR* for *Arkansas*, *AZ* for *Arizona*, *Berks* for *Berkshire*), and their form corresponds to national or international conventions, meant to be used, in many cases, in postal addresses, or on vehicle registration plates. They are created in order to fit in the limited space of an envelope or a plate, and, in this view, are just a printer’s solution to a space issue. Others are used to provide textual references (*Mk* for *the Gospel of Mark*), or to indicate measure units (*GHz* for *giga herz*, *mil* for *millilitre*), or are used as abbreviated titles (*Dr.* instead of *Doctor* on the heading of a prescription or on a plate signalling the presence of a doctor’s surgery in a building, *Pr.* for *Professor*). In these cases, again, the point is to gain space, and perhaps to reduce the amount of space that the unabbreviated form would take in order to background the nature of, e.g., the measure unit (which in most contexts is probably fairly obvious, but has to be indicated for rigour or clarity) in favour of the actual figure which it follows. The reasons behind this type of reduction therefore differ from those that yield acronyms, initialisms and clippings, which enable speakers to insert a word-like unit in a larger syntactic unit, be it a phrase or a clause, i.e., in connected speech. Contrary to acronyms, clippings, and initialisms, the orthographic shortenings that we are discussing remain largely outside of syntax, at least of the syntax of texts: they are simply orthographic variants, or allographs, restricted to the written medium.
- 37 They do not just differ from the other abbreviations in their use, but also from a formal point of view, in that the reduction processes that they illustrate are more varied than those observed in the other abbreviations. In addition to the reduction processes used in other formations, they may retain the extremities of the original word (*Middx* for *Middlesex*, *utd* for *united*), and/or keep just the consonants, to the detriment of vowels (e.g., *MB* for *Manitoba*). Many actually make use of both reduction means (retaining the

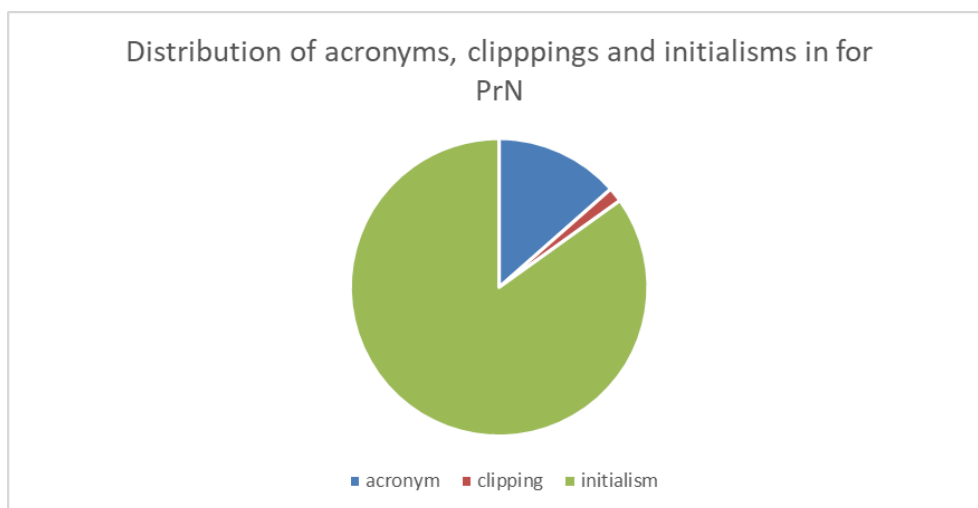
extremities, and selecting just the consonants), as is the case for *GHz* for *gigahertz*, *Mgr* for *Manager* or *Monseigneur*, *tbsp* for *tablespoon*, *ltd* for *limited*). Another notable difference with acronyms and initialisms (but not clipping) is that the reduction processes at play operate to a large extent independently of morphemic boundaries. The letters *M* and *B* of *MB* (for *Manitoba*) do not correspond to either lexemes or morphemes, i.e., units of meaning, as would the letters forming the initialism *CIA*, where each letter stands for a particular lexeme associated with a particular meaning. On this account, the orthographic shortenings that we are dealing with here seem closer to the items noted as “symbols” in the dictionary, as is the case for *ac*, representing the chemical element actinium. It is typical that they belong to a specialised scientific domain, chemistry, which is characterised by international norms.

- 38 Considering, overall, that the orthographic shortenings that we have just addressed are created with very specific purposes, which are not those of acronyms and initialisms, we may exclude them from our discussion, and now focus on the other types – more specifically on the items that are proper names.

2.3. Proper names and initialisms / acronyms

- 39 The first thing we may note concerning abbreviations that are also proper names is the overwhelming proportion of initialisms, which by far outnumber the other reduction processes, as illustrated in Figure 2.

Figure 2. Proportion of acronyms, clippings and initialisms among proper names listed in the *COED* with the label *abbr.*



The largest part of the chart, in green, represents initialisms. The other significant category (in blue) is that of acronyms, which is worth noting, as acronymy is the formation process closest to initialism. Let us remember that in both cases, the initial letter is retained, the difference between the two being that acronyms go one step further in the reduction process, in that the new unit is ultimately pronounced as a word, rather than as a string of individual letters.

- 40 The unity of the category of abbreviations that are proper names is also striking from a semantic and referential point of view. The formations in question consistently denote institutions, i.e., groups of people, that are also treated as forming a higher

individualized unit: *BAFTA*: British Academy of Film and Television Arts; *DEA*: Drug Enforcement Admistration; *DVLA*: Driver and Vehicle Licensing Agency; *BAOR*: British Arry of the Rhine; *RA*: Royal Artillery; *SFA*: Scottish Football Association; *ASA*: Advertising Standards Authority; *IBRD*: International Bank for Reconstruction and Development; *CAB*: Civil Aeronautics Board; *CAB*: Citizens' Advice Bureau; *NEC*: National Exhibition Centre; *RAC*: Royal Automobile Club; *RCA*: Royal College of Art; *EU*: European Commission; *ICRC*: International Committee of the Red Cross; *CIS*: Commonwealth of Independent States; *EC*: European Community; *NBC*: National Broadcasting Company; *CRE*: Congress of Racial Equality; *IFC*: International Finance Corporation; *RAMC*: Royal Army Medical Corps; *WBC*: World Boxing Council; *ICC*: International Criminal Court; *CID*: Criminal Investigation Department; *DC*: District of Columbia; *IAAF*: International Amateur Athletics Federation; *RAF*: Royal Air Force; *G20*: Group of 20; *HC*: House of Commons; *LI*: Light Infantry; *SDI*: Strategic Defense Initiative; *MIT*: Massachusetts Institute of Technology; *ICE*: Institution of Civil Engineers; *NFL*: National Football League; *MOD*: Ministry of Defense; *RN*: Royal Navy; *CNN*: Cable News Network; *EPO*: European Patent Office; *LPO*: London Philharmonic Opera; *LSO*: London Symphony Orchestra; *WTO*: World Trade Organization; *EP*: European Parliament; *SNP*: Scottish National Party; *RCMP*: Royal Canadian Mounted Police; *PQ*: Province of Quebec; *UDR*: Ulster Defence Regiment; *GDR*: German Democratic Republic; *LSE*: London School of Economics; *NHS*: National Health Service, *HHS*: Health and Human Services; *NSPCC*: National Society for the Prevention of Cruelty to Children; *CBS*: Columbia Broadcasting System; *PCT*: Primary Care Trust; *USSR*: Union of Soviet Socialist Republics; *OU*: Open University.

- 41 It is easy to see the semantic unity behind the numerical diversity of all these nouns. They denote (or presuppose the existence of) both a group of people and an institution, that is, at the same time, plurality and unity. Many of them license both singular and plural agreement and pronominal anaphora: *The DEA (Drug Enforcement Administration) has/have classified yet another substance as carcinogenic – it/they hold the view that precautions should be taken*). In some (less frequent) cases, it is just the plural marking that conveys the notion of grouping and creating a new unity out of a plurality, as in *RM* for *Royal Marines*, or *RE* for *Royal Engineers* (that may be compared with *ICE*, *Institution of Royal Engineers*). And in yet other cases, the representation of a unity is not morphologically marked, but is encapsulated in the semantic make-up of the unit, as is the case for *FT (Financial Times)* – it may be assumed that the paper is taken as a collective product, i.e., it has involved many contributors.
- 42 Considering the remarkable semantic unity of these initialisms and acronyms that are also proper names, it is tempting to suggest that there may be an element of iconicity at play, whereby complex entities, made up of a number of individuals, are “reduced” to simple ones. The reduced form is apt to foreground the unity over the plurality of the individuals that make up the group.
- 43 Now we may wonder why all names of institutions are not reduced to the same degree. It may initially be hypothesized that this has to do with a property of the referent. However, it is more likely that the reason for a further reduction has to do with the properties of the linguistic material.
- 44 Our main result is therefore that when proper names enter the lexicon (if in this case entering the dictionary can be interpreted as entering the (common) lexicon), this

triggers a process of reduction. We will now discuss this relation between reduction and insertion of a proper name in the lexicon.

3. Discussion: the lexicalization of reduced proper names

45 We will address two main questions in this discussion:

- How do we account for the differences in reduction processes for proper names (initialisms and acronyms)?
- How do we account for the fact that there are reduced proper names in the dictionary?

In other words, why is reduction so effective in creating lexemes and what are the differences between a full form proper name and a reduced proper name?

46 The answer we give is a simple one: whenever a unit looks like a word and proves useful as an autonomous unit of syntax, its original linguistic category is very likely to undergo reanalysis as a new “word”. Similar yet not completely identical to their original longer forms, reduced proper names can do the same as their longer counterparts and more, and their morphophonological and functional differences are enough to treat them as new, distinct units (contrary to orthographic shortenings, which are just written variants) which do not have to follow the rules for proper names anymore and may therefore follow the rules of the other “common” units instead.

3.1. Different reduction processes for different stages towards word-ification

47 Let us first of all recall the resemblance between initialism and acronymy, and state again that they may be regarded as representing different phases of a unique formation process. An alternative view may of course be that they are distinct, unrelated to each other, other than being the result of retaining only the initials of a longer form. Speakers would then have two options. Option A would consist in pronouncing each initial letter, and option B in pronouncing what you see, as a word. We will, however, postulate a gradient between the two, or, in other words, chronological steps, which lead from one to the other – provided that is possible of course, depending on the potential phonological structure of the reduced unit.

48 The mechanism is as follows. The unit *ASAP* (*as soon as possible*), for instance, is initially pronounced as an initialism: each letter is spelt out. But one also perceives a graphic unit that potentially contains two syllables (because of the two vowels), which makes it possible to pronounce it as an acronym. The process may culminate in a third stage, where the unit is not perceived as an acronym any more, but as a word, which belongs to a specific part of speech, such as noun, adjective, verb, etc. This is the case for *laser*, *POTUS* (*President of the United States*), or *MIPS* (*a unit of computing speed equivalent to a Million Instructions Per Seconds*), which are pronounced as regular words (not initialisms) and entered in the dictionary as nouns, not abbreviations. From a graphic point of view, the counterpart of this gradual process of word-ification is the shift from upper case to lower case (*ECU* or *ecu* – for *European Currency Unit*, the former name of the euro), with sometimes yet another intermediary stage where just the initial capital letter is retained (*Nicam*, or *NICAM*, for *near instantaneously companded audio multiplex*).

Note, in this respect, that *ecu* has lost even the initial capital letter, so that it may be taken to have reached the last stage in the word-ification process, which may have to do with the fact that both from a formal and semantic point of view it resembled a pre-existing word in some European languages (*écu* in French).

- 49 Whatever the stage reached by a reduced unit, the general mechanism accounts for the existence of hybrids. Indeed, if some items are 100% initialisms, and others are 100% acronyms, others (e.g., *UFO*) may be pronounced either way, as an initialism or an acronym. It also explains why we encounter reductions that are similar in writing but not in pronunciation (homographs but not homophones). For instance, *SAT* is pronounced as an initialism when it means *Scholastic Aptitude Test* (US) and as an acronym when it means *Standard Assessment Task* (UK).
- 50 Indeed, the fact that not all initialisms eventually become acronyms, and all acronyms do not become “regular” lexemes, should not be taken as an indication that the general mechanism does not apply. Let us recall that, in order for an initialism to become an acronym, it must be possible to identify a potential syllabic structure. Besides, it may be deemed necessary to avoid homonymy between two identical reductions, and therefore to keep one as an initialism, and the other as an acronym. Recreating a contrast between two units may block the evolution towards the next step of wordification (Aronoff [1988: 767]). This does not, however, undermine the hypothesis of a general mechanism tending towards “wordification”.

3.2. Similarity of forms, backgrounding of former stages and opacity of form / origin / meaning

- 51 We saw that initialisms and acronyms were ordered on a cline toward more “wordedness” and illustrated a process which we called “wordification”. This target word corresponds to what Hohenhaus [2005: 366] calls a “potential listeme”, i.e., “a possible word purely from the point of view of the lexicon” as opposed to a “possible word” created in performance and found in use without being listed in the lexicon. Compared to other proper names which can be lexicalized through loss of semantic features (modified proper names), reduced proper names achieve their gradual integration into the lexicon by taking advantage of the gradual blurring between the 3 types of naming units: a) simple and complex lexemes: *bug*, *bus-driver*; b) expressions (noun phrases, collocations, definite descriptions): *friendly fire*, *the President (of the USA)*, *Gulf War II* and c) proper names (person names, place names, eponyms): *George W. Bush*, *Camp David*, *Alzheimer*, *Poinsettia*, *Bakelite* (list and examples from Lipka *et al.* [2004: 1–2]). Indeed, the reduced forms we study here are proper names and/or expressions (depending on the number of elements in the full form), and reduction makes them look like monomorphemic (simple, because unanalysable) lexemes. Indeed, lexicalization through reduction leads to the unit being treated as a single syntactic and semantic unit (what Hilpert [2019] calls “holistic processing”), something that is already at stake when syntactic structures and common nominals are lexicalized (Mignot [2017]) but also when proper names are created and automatically nominalized (Philippe [2020]). With reduced proper names, the process goes a step further as it leads to a visually single unit, one orthographic block that does not exceed two phonological syllables on average, which makes it easier to assimilate it with already existing simple/complex “common” lexemes, in written and/or phonological form: *ACT*

(*advanced corporation tax*) vs. *ACT* (*Australian Capital Territory*) vs. *act* (n./v.); *AI* (*Amnesty International*) vs. *AI* (*artificial intelligence*) vs. *ai* (n.); *ECU* (*European Currency Unit*) vs. Fr. *écu*; *GI* (*Government issue; glycaemic index*) vs. *gi* (n.); *SALT* (*Strategic Arms Limitation Talks*) vs. *salt* (n./v.), etc.

- 52 The reduced proper name cuts ties with its original full form *via* a double opacification effect: if initialisms retain some motivated link with their full form by keeping the initials of the original elements, it is however possible to use the reduced form without knowing, or remembering (accurately) what the initials stand for (Mignot [2017]), e.g., who remembers that *CNN* stands for *Cable News Network*? Moreover, with reduction and the creation of this new form comes a renewed possibility of homonymy which dilates the connection to the original form even more, like *AAA* for both *Amateur Athletic Association* and *American* (or *Australian*) *Automobile Association*. This means that the reduced form has to be integrated into the mental lexicon of the speakers (partly) on its own, since it is not unequivocally tied to the full form. With acronyms, there go the last remains of the connection to the full form when initialisms stop being capitalized and pronounced separately: Lipka *et al.* [2004] describe this as an example of demotivation leading to lexicalization.
- 53 Therefore, reduced proper names illustrate both Lipka *et al.* [2004]’s *institutionalization* (“the integration of a lexical item, with a particular form and meaning, into the existing stock of words as a generally acceptable and current lexeme”) and *lexicalization* (“the process by which complex lexemes tend to become a single unit with a specific content, through frequent use. In this process, they lose their nature as a syntagma, or combination [of smaller units], to a greater or lesser extent”, Lipka [1992: 107], quoted in Lipka *et al.* [2004: 7]).
- 54 According to them, these institutionalized and lexicalized units:
- neither belong fully to the level of language – not yet entirely, which explains the variants in initialisms with or without dots, or the hesitation between acronyms and initialisms, or even the hesitation between the abbr. label and a POS;
 - nor fully to the level of parole – not anymore, even though lexicalization is brought about by frequency of use in discourse –
 - but to the intermediary level of norm defined as the “collective realisation of the language system” (Lipka *et al.* [2004: 3]), which helps speakers know what can and cannot be done in morphology, phonology and syntax.
- 55 This leads to one last property that we believe to be key in their treatment as a lexical unit distinct from proper names and more likely to be integrated fully into the lexicon: their ability to enter in syntactic constructions and semantic associations more freely than their full forms, for example the pre-modifying position within a noun phrase which, while open to long nominals such as full form proper names, is in practice mostly restricted to simple / complex lexemes like adjectives and nouns. By looking like a lexeme (the main naming unit allowed in dictionaries) and by behaving in discourse like “common” lexemes more easily than the full form they gradually detach from, reduced proper names achieve a deep transformation into an autonomous unit that can be integrated into the lexicon regardless of where it came from.
- 56 We propose to confirm this autonomy (or loss of dependence on the full form) of reduced proper names in discourse with a case study, comparing the uses of *CIA/C.I.A.* with those of the full form *Central Intelligence Agency* in the *Corpus of Contemporary*

American English, to show that their distinct syntactic and semantic profiles call for a distinct lexical treatment.

3.3. Case study: *Central Intelligence Agency* vs. *CIA / C.I.A.* in the *COCA*

3.3.1. Frequency

- 57 The *COCA* confirms that the reduced forms *CIA/C.I.A.* are much more frequent in discourse than their full form counterpart: a frequency research yields 25 416 results for *CIA*, 1 358 results for *C.I.A.*, but only 1 121 results for *central intelligence agency*.
- 58 We selected the first 300 results for the reduced forms (200 *CIA*, 100 *C.I.A.*) and for the full form for further analysis: 299 reduced forms and 277 full forms were retained after cleaning false positives (*central intelligence agency*: common nominal, *Korean Central Intelligence Agency*: not the right referent) and frequent occurrence duplications.
- 59 We also noted that the full form often did not occur alone in non-specialized discourse: barring law texts and reviews, the reduced forms *CIA/C.I.A.* are found either in *Central Intelligence Agency*'s immediate co-text (between parentheses, on the same line) or in the larger co-text provided by the *COCA*, 110 times out of 277 (39.71%). On the contrary, the reduced forms will mostly occur without the full form without impeding comprehension.

3.3.2. Profile part 1: head vs. modifier, types of modified CNs, semantic agency

- 60 The analysis of the context of occurrence for these forms revealed that *CIA* and *C.I.A.* were respectively 118/199 and 78/100 times heads of a noun phrase and 78/199 and 37/100 times pre-modifiers within a noun phrase headed by a common noun (with 1 subject complement and 1 n.a), which means that even though being a head noun is their main context of occurrence, they do appear in pre-modifying position in more than a third of cases (115/299: 38.46%).
- 61 The contrast with the full form is striking: 234/277 occurrences of *Central Intelligence Agency* are as heads of an NP and only 37/277 occurrences are in pre-modifying position (13.36%). This low result has to be qualified and brought even further down, since 17 of those pre-modifying occurrences are due to their being embedded in three official proper names for laws or sections of those laws: *Central Intelligence Agency Act of 1949*; *The freedom of Information Act: Central Intelligence Agency Exemptions*; *Central Intelligence Agency Exemptions in the Privacy Act of 1974*. The full form could never be reduced here, since it is part of the laws' names. Another dozen pre-modifying occurrences can be questioned since they are part of a title or a pseudo-title mainly in close appositions (Philippe [2018]): *Central Intelligence Agency d/Director* (First Name + Last Name), *Central Intelligence Agency chief* (First Name + Last Name). We argue that the title using *Director*, at least, favours the use of the full form because it is a very close re-arranging of the official title *Director of the Central Intelligence Agency* (very often found in the sampling due to the selection of a law website as a source), which cannot be reduced since it is part of the official statutes of the institution. All in all, unambiguous cases of pre-modification by *Central Intelligence Agency* amount to 8 occurrences with the following head CNs: *black-op* (1), *contractor* (1), *official* (1), *operative* (2), *people* (1), *personnel* (1),

program (1). All in all, pre-head modification by the full form *Central Intelligence Agency* in discourse only amounts to 8–20/277 occurrences, i.e., 2.89 – 7.22% of the sampling.

62 Not only are the common nouns modified by the full form very limited, they are also not exclusive to the full form. The reduced forms *CIA/C.I.A.* can modify these common nouns along with a great variety of other common nouns:

- CNs modified by *CIA* (78 occ.): *agent* (6), *airline* (1), *analyst* (2), *A/annex* (5), *asset* (1), *base* (1), *chain of command* (1), ***chief*** (2), *complicity in drug dealing* (1), *component* (1), ***contracted thug*** (1), *coup* (1), ***director*** (23), *External Advisory Board* (1), *failure to (...)* (1), *IG report* (1), *memo* (2), *mole in the Time and Fate mission* (1), *narrative* (1), *Office of Security* (1), *officer* (2), ***ops*** (1), ***operative*** (6), *operatives and contractors* (1), *operation* (3), *data-base operations* (1), ***personnel*** (1), *prison* (2), *report* (2), *rumors* (1), *screw ups* (1), *security team* (1), *torture* (1), *World Factbook* (1).
- CNs modified by *C.I.A.* (37 occ.): *account* (1), *base* (1), *brainstorms* (1), ***chief*** (2), *communications gear* (1), *counterterrorism operations* (1), *detention centers* (1), ***director*** (9), *efforts* (1), *etal* (1), *female agent* (1), ***interrogation program*** (1), *interrogations and new lingerie* (1), *involvement* (1), *job* (1), *officer* (4), ***official*** (3), ***operative*** (1), *soap opera* (1), *veteran* (3), *World Factbook* (1).

63 Frequency aside, the variety of extralinguistic contexts the reduced form is associated with through modification of a CN confirms that they are the usual form in non-specialised discourse. Whereas the full form is somehow restricted to the reference to the top of the hierarchy, to a few individuals or partners mentioned in the media and tied to the Agency for the purpose of one article, with expected words in the semantic field of an intelligence agency, the reduced form accompanies the reference to people of all statuses connected to the Agency, enables reference to inanimate entities (objects, actions and activities) relating to the institution but also events that connect the institution with the rest of the (non-intelligence) world in a more dynamic way. If the lexicon really is “a collection of stored associations among fragments of disparate representations” (Jackendoff [1997: 108], in Hohenhaus [2005: 358]), then it is clear that *CIA/C.I.A.* corresponds to a slightly different set of mental associations which could justify its treatment as a separate unit of the lexicon rather than simple allonym of *Central Intelligence Agency*.

64 This difference in terms of semantic dynamism is corroborated by the syntactic functions and semantic roles of the NPs headed by *Central Intelligence Agency* and *CIA/C.I.A.*

Exactly half of the reduced forms have syntactic functions associated with more agency (90/180): subjects (66), agentive complements (9), preposition complements (11) / genitive determiners (4) linked with the “agent” semantic role.

65 On the contrary, 183/234 occurrences of the full form as head of an NP are preposition complements without any trace of agency (in the title *Director of the Central Intelligence Agency*, or *at/from/within the Central Intelligence Agency*, etc.). Only 6 agent complements and 10 preposition complements/genitive determiners with agency are found, while there are just 18 occurrences in subject position that do not imply much agency: lexical BE and passive voice (*is gripped by anxiety*, *are being sued*, *is responsible for...*), processes that have to do with the cognition or information seeking and sharing (*obtained information*, *is reportedly investing*, *told*, *revealed*, *denied...*) or negated processes (*withheld*, *never managed*). On the contrary, reduced forms tended to denote more agentive subjects (*did*, *created*, *brought*, *was transferring*, *tortured*, *contributed*, *collated and computerized...*).

- 66 The full form is referentially more constrained and pragmatically restricted to first mentions of the Agency, whereas the reduced form is the one to use to describe the world and interactions with the world: as such, it is the most relevant unit for the speakers and it explains the necessity to validate its lexical status, so that more English speakers may use it where the full form fails.

3.3.3. Profile part 2: reduced form and overall syntactic acceptability, reduction and syntactic / lexicological reanalysis

- 67 In addition to what the reduced forms achieve referentially and pragmatically, they also have an impact on the fabric of discourse itself, and how easy they come to the speakers and are decoded by the co-speakers. The lexical status of *CIA/C.I.A.* might also rest on its ability to conform to the expectations of a “common” lexical item, such as easy insertion in phrases and co-occurrence with other sub-phrasal units.
- 68 Unsurprisingly, the reduced forms *CIA/C.I.A.* co-occur more frequently with other pre-head modifiers as well as more complex head nominals than the full form *Central Intelligence Agency*: only 6 occurrences of other pre-modifiers for the full form (*United States* (1), *former* (4) + title/pseudo-title, *relentless [...] black-ops in Khuzestan and Balochistan* (1)) against 39 occurrences for the reduced forms. The nouns the reduced forms modify may take genitive and numeral determiners (*two CIA officers*, *his C.I.A. account*), something which was not found when the full form was a modifier. As heads of a noun phrase, the reduced forms’ modifiers are again more varied: *former* (16), *acting*, *‘black-site’*, *career*, *disgruntled*, *great*, *important*, *inevitable*, *large*, *largest*, *notorious Vietnam-era*, *other top*, *undercover*, *secret*. Finally, the reduced forms were found in several patterns of sub-phrasal coordination, either as head or modifier: *a small and select group of CIA and other top government officials*; *transcripts of black-site C.I.A. interrogations and new lingerie*.
- 69 When the full form is used, as a modifier (but also as a head), it forces the reader or co-speaker to process each word one after the other, which foregrounds the institution for as much time as needed for the decoding process. On the contrary, shorter units such as the reduced form *CIA* are less likely to attract the decoder’s attention, or for a considerably shorter amount of time, which reduces the foregrounding of the institution or even backgrounds it in case of easy and frequent recognition of the unit in discourse. This de-foregrounding or backgrounding is completely coherent with the pre-head position in an NP, which is almost exclusively reserved for reduced forms such as *CIA*. Syntactically, too, a shorter, unanalysable unit such as *CIA* (where the internal syntactic organisation in Adj + mod-N + head-N is hidden, opacified) does not slow down or prevent the apprehension of the NP’s internal structure and hierarchy. Compare: *a small and select group of CIA and other top government officials* with *a small and select group of Central Intelligence Agency and other top government officials*. This explains why full forms tend to be relegated to the end of units, before an apposition or at the end of a clause / sentence, because of its syntactic complexity and phonological weight that obscure the rest of the sentence.
- 70 Lastly, we noticed that even though the full form *Central Intelligence Agency*, being a “weak proper name” (Huddleston & Pullum [2002: 517]), is supposed to be determined by the definite article *the* when it heads a noun phrase, the same constraint does not entirely apply to the reduced form *CIA/C.I.A.*: 17 occurrences out of 180 were

determined by Ø: according to Ø CIA and other sources; Ø CIA, FBI and many other organisations have been doing it for years without getting caught; a group well versed in bringing about regime change within Ø CIA itself; After the general had a statement issued that Ø C.I.A. issued no Stand Down instructions in the night of 9/11 in Libya, etc. We regard this as a clue that the speakers reanalysed the unit, giving it more independence from the original full form. Syntactically, it is no longer the reduction of the full form (otherwise, the same syntactic constraints which do not involve end-weight would apply, such as determination): it has become something else, a new unit, with new syntactic properties (such as the possibility to be determined by zero) and therefore the potential to acquire a new, different, lexical status.

Conclusion

- 71 This paper analysed the integration of proper names into the lexicon when they are reduced to initialisms, acronyms and clippings. We confirmed their apparently abnormal inclusion in dictionaries, despite their not being semantically or referentially modified compared to other proper names present in dictionaries, and analysed their distribution compared to that of other reduced units, marked *abbr.* in the *COED* (our corpus of reference). We noticed a strong tendency to include reduced names of institutions (hybrid between animate and inanimate reference, mostly collective). We then discussed the reasons for their inclusion in dictionaries, which we took to be a sign of their lexicalization, and concluded it was an extreme illustration of opacification, demotivation and wordification, all of which achieved via the reduction of these units. Beyond the simple orthographic and phonological form these reduced proper names assumed, we also showed in a case study that their syntactic and semantic profile differs slightly from that of their full forms, which further justifies their differentiated lexical treatment. This comparison between reduced and full forms ought to be replicated at a greater scale. The *COED* labelled them *abbr.*, along with a good many different types of abbreviations, while other dictionaries either labelled “common” abbreviations according to the part-of-speech of their full forms and did not label reduced proper names at all or labelled them *abbr. too*. We take it to mean that these reduced proper names have achieved partial integration into the lexicon: they are considered lexical units *a minima*, from a lexicographic perspective (a lexical item is any unit “which can figure in a dictionary [...] as the ‘target’ of some form of lexicographic description” Atkins & Rundell [2008: 163]) but also from a lexicological perspective. Their only description is their full form, and their lack of a syntactic part-of-speech speaks to the room to grow until they are fully recognised in language. We leave open the question of the part-of-speech most suitable to describe these units (proper name? n.?) and what it would entail.

BIBLIOGRAPHY

- ARONOFF Mark, 1988, “Review of: Di Sciullo/Williams (1987)”, *Language* 64, 766–770.
- ATKINS B. T. Sue & RUNDELL Michael, 2008, *The Oxford Guide to Practical Lexicography*, Oxford: Oxford University Press.
- BOZKURT Ferdi, 2019, “Proper Names in General (Purpose) Dictionaries: Necessity”, *Uluslararası Eğitim ve Dil Dergisi – International Journal of Education and Language* 12(1), 10–20.
- DEVEREUX Robert, 1984, “Shortenings, Blends and Acronyms”, *Word Ways* 17(4), 210–215.
- GARY-PRIEUR Marie-Noëlle, 1991, « Le nom propre constitue-t-il une catégorie linguistique ? », *Langue française* 92(1), 4–25.
- HILPERT Martin, 2019, “Lexicalization in Morphology”, *Oxford Research Encyclopedia of Linguistics*, Oxford University Press online.
- HOHENHAUS Peter, 2005, “Lexicalization & Institutionalization”, in ŠTEKAUER Pavol & LIEBER Rochelle (Eds.), *Handbook of Word-Formation*, Dordrecht: Springer, 353–374.
- HUDDLESTON Rodney, & PULLUM Geoffrey (Eds.), 2002, *The Cambridge Grammar of the English Language*, Cambridge: Cambridge University Press.
- LIPKA Leonhard, 1992, “Lexicalization and Institutionalization in English and German. Or: *Piefke, Wendehals, smog, perestroika, AIDS* etc.”, *Zeitschrift für Anglistik und Amerikanistik* 40, 101–111.
- LIPKA Leonhard, HANDL Susanne & FALKNER Wolfgang, 2004, “Lexicalization & Institutionalization: The State of the Art in 2004”, 1–18.
- MIGNOT Élise, 2017, *La morphologie du nom en anglais : vers une sémantique des parties du discours*, Habilitation monograph, Paris : Sorbonne Université.
- MÖLLER Lucie A., 2008, “Methods and Problems in Proper Name Lexicography”, in EICHLER Ernst, HILTY Gerold, LÖFFLER Heinrich, STEGER Hugo & ZGUSTA Ladislav (Eds.), *1. Halbband: Ein internationales Handbuch zur Onomastik*, Berlin/New York: De Gruyter Mouton, 324–328.
- PHILIPPE Manon, 2018, “‘The Scottish novelist William Black’: Close Appositions and the Modification of Proper Names”, *Anglophonia* 26, available at <http://journals.openedition.org/anglophonia/1722>.
- PHILIPPE Manon, 2020, *Le Nominal propre. Étude du nom propre en anglais*, Doctoral Thesis, Paris : Sorbonne Université.
- PLAG Ingo, 2003, *Word-Formation in English*, Cambridge: Cambridge University Press.
- VAXELAIRE Jean-Louis, 2006, « Le nom propre en contexte – Une approche lexicologique », in BLAMPAIN Daniel, THOIRON Philippe & VAN CAMPENHOUDT Marc (Eds.), *Mots, termes et contextes : Actes des septièmes journées scientifiques du réseau de chercheurs ‘Lexicologie Terminologie Traduction’*, Bruxelles, Belgique, 8, 9 et 10 septembre 2005, Paris : Éditions des archives contemporaines, 591–599.
- Dictionaries, corpora and databases**
- ALLEN Robert (Ed.), 2002, *Penguin Concise English Dictionary*, London: Penguin.
- BARHEN DAN, 2019, *YouGlish*. Available online at <https://youglish.com/>.

DAVIES Mark, 2008, *The Corpus of Contemporary American English (COCA)*. Available online at <https://www.english-corpora.org/coca/>.

HORNBY Albert Sydney (Ed.), 1981, *Oxford Advanced Learner's Dictionary of Current English* (14th Edition).

HORNBY Albert Sydney, TURNBULL Joanna, LEA Diana & PARKINSON Dilys (Eds.), 2010, *Oxford Advanced Learner's Dictionary* (8th Edition), Oxford: Oxford University Press.

RUNDELL Michael (Ed.), 2012, *Macmillan English Dictionary for Advanced Learners* (2nd Edition), Oxford: Macmillan.

STEVENSON Angus & WAITE Maurice (Eds.), 2011, *Concise Oxford English Dictionary* (12th Edition), Oxford: Oxford University Press.

SINCLAIR John (Ed.), 2018, *Collins COBUILD Advanced Learner's Dictionary* (9th Edition), Collins CoBUILD.

Wikimedia Foundation. *Wiktionary*. Available at <https://www.wiktionary.org/>.

Youtube. Available at <https://www.youtube.com/>.

NOTES

1. We would like to thank a *Lexis* reviewer for their observations regarding the classification of dictionaries, collective nouns and a few stress patterns, which helped make this a more rigorous paper. All remaining errors or imprecisions are our own.

ABSTRACTS

This paper analyses the integration of proper names into the lexicon when they are reduced to initialisms, acronyms and clippings. We confirm their apparently abnormal inclusion in “common” dictionaries, despite their not being semantically or referentially modified, and analyse their distribution compared to that of other reduced units, marked abbr. in the *Concise Oxford English Dictionary* (our corpus of reference). We notice a strong tendency to include reduced names of institutions (hybrid between animate and inanimate reference, mostly collective). We then discuss the reasons for their inclusion in dictionaries, which we take to be a sign of their lexicalization, and conclude to an extreme illustration of opacification, demotivation and wordification, all of which achieved via reduction. Beyond the simple orthographic and phonological form these reduced proper names assume, we also show via a case study that their syntactic and semantic profile differs slightly from that of their full forms, which further justifies their differentiated lexical treatment.

Cet article analyse l'intégration des noms propres dans le lexique dès lors qu'ils ont subi une réduction de type initialisme, acronymie ou troncation. Nous confirmons leur inclusion, en apparence anormale, dans les dictionnaires « communs », alors même qu'il ne s'agit pas de noms propres modifiés sémantiquement ou référentiellement ; nous comparons leur distribution par

rapport aux autres unités abrégées, marquées abbr. dans notre corpus de référence, le *Concise Oxford English Dictionary*. Nous relevons une forte tendance à inclure les noms propres réduits d'institutions (unités intermédiaires entre la référence animée et inanimée, très souvent collective). Nous discutons ensuite les raisons permettant leur inclusion dans les dictionnaires, que nous interprétons comme le signe de leur lexicalisation, et concluons à leur très grande opacification, démotivation et lexème-isation, permises par la réduction. Au-delà de leur forme graphique ou phonique, nous montrons via une étude de cas que le profil syntaxico-sémantique de ces noms propres réduits diffère légèrement de celui des formes pleines, ce qui justifie leur traitement lexical différencié.

INDEX

Mots-clés: noms propres, réduction, abréviation, lexicalisation, dictionnaires

Keywords: proper names, reduction, abbreviation, lexicalization, dictionaries

AUTHORS

MANON PHILIPPE

Rennes 2 University, ACE (Anglophonie : communautés, écritures)

manon.philippe@univ-rennes2.fr

ÉLISE MIGNOT

Sorbonne University, CELISO (Centre de Linguistique en Sorbonne)

elise.mignot@sorbonne-universite.fr